





Article

Comparing Early Transcriptomic Responses of 18 Soybean (*Glycine max*) Genotypes to Iron Stress

Daniel R. Kohlhase¹, Chantal E. McCabe^{2,†}, Asheesh K. Singh¹, Jamie A. O'Rourke^{2,*} 
and Michelle A. Graham^{2,*} 

¹ Department of Agronomy, Iowa State University, Ames, IA 50011, USA; kohlhase@iastate.edu (D.R.K.); singhak@iastate.edu (A.K.S.)

² U.S. Department of Agriculture (USDA)—Agricultural Research Service (ARS), Corn Insects and Crop Genetics Research Unit, Ames, IA 50011, USA; McCabe.Chantal@mayo.edu

* Correspondence: jamie.orourke@usda.gov (J.A.O.); michelle.graham@usda.gov (M.A.G.)

† Present address: Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN 55901, USA.

Abstract: Iron deficiency chlorosis (IDC) is an abiotic stress that negatively affects soybean (*Glycine max* [L.] Merr.) production. Much of our knowledge of IDC stress responses is derived from model plant species. Gene expression, quantitative trait loci (QTL) mapping, and genome-wide association studies (GWAS) performed in soybean suggest that stress response differences exist between model and crop species. Our current understanding of the molecular response to IDC in soybeans is largely derived from gene expression studies using near-isogenic lines differing in iron efficiency. To improve iron efficiency in soybeans and other crops, we need to expand gene expression studies to include the diversity present in germplasm collections. Therefore, we collected 216 purified RNA samples (18 genotypes, two tissue types [leaves and roots], two iron treatments [sufficient and deficient], three replicates) and used RNA sequencing to examine the expression differences of 18 diverse soybean genotypes in response to iron deficiency. We found a rapid response to iron deficiency across genotypes, most responding within 60 min of stress. There was little evidence of an overlap of specific differentially expressed genes, and comparisons of gene ontology terms and transcription factor families suggest the utilization of different pathways in the stress response. These initial findings suggest an untapped genetic potential within the soybean germplasm collection that could be used for the continued improvement of iron efficiency in soybean.

Keywords: *Glycine max*; soybean; iron deficiency chlorosis; abiotic stress; RNA-seq; comparative transcriptomics



Citation: Kohlhase, D.R.; McCabe, C.E.; Singh, A.K.; O'Rourke, J.A.; Graham, M.A. Comparing Early Transcriptomic Responses of 18 Soybean (*Glycine max*) Genotypes to Iron Stress. *Int. J. Mol. Sci.* **2021**, *22*, 11643. <https://doi.org/10.3390/ijms222111643>

Academic Editor: Anna M. Mastrangelo

Received: 23 September 2021

Accepted: 25 October 2021

Published: 28 October 2021

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



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

1. Introduction

Iron deficiency chlorosis (IDC) in soybean (*Glycine max* [L.] Merr.) is characterized by interveinal chlorosis, stunted growth, and yield loss. IDC is typically found in soybeans grown throughout the North Central U.S., where a high pH (>7.2) and calcareous soils limit iron availability, resulting in IDC development [1]. Soil properties and genetic differences between lines create a variability in iron stress tolerance [2–5]. Froehlich and Fehr (1981) demonstrated the genotypic variability of the IDC response among 15 soybean varieties, finding that each one point change on the IDC visual rating scale (1–5) correlated to an approximately 20% yield loss at the end of the season [3]. Using the 2020 median price of soybean, the estimated economic loss due to IDC in the North Central U.S. would be approximately 117 million USD [1]. Due to the high potential for yield loss associated with IDC, we must improve our understanding of iron stress responses in order to keep economic losses to a minimum.

A collective effort to improve our ability to breed for iron efficiency has resulted in a strong research foundation addressing the genetics of iron utilization and crop stress adaptations. Weiss (1943) was the first to suggest a single dominant gene underlying the

efficiency of iron utilization in soybean [6]. Cianzio and Fehr (1980) justified the variation in iron stress responses by suggesting that modifying genes accompany major quantitative trait loci (QTL) [7]. Since then, multiple genetic studies have provided more evidence supporting the idea of multiple genes controlling iron efficiency [8–12]. Diers et al. [13] first mapped an iron efficiency QTL using an early soybean genetic map. Later, Lin et al. [9] mapped an iron efficiency QTL using two different mapping populations: in one population, a number of minor effect QTL were associated with iron efficiency, whereas, in the other population, 68–73% of variance associated with iron efficiency was mapped to a single QTL. Following the publication of the soybean genome, Severin et al. [14] narrowed the location of this major QTL on soybean chromosome Gm03 using an introgression mapping of near-isogenic lines (NILs) Clark (iron stress tolerant) and IsoClark (iron stress susceptible), and the iron inefficiency donor T203 (iron stress susceptible). Peiffer et al. [15] used introgression and QTL mapping to narrow the QTL within the introgressed region even further. Recently, Assefa et al. [12] performed a genome-wide association study, characterizing IDC tolerance in 460+ soybean lines using multiple phenotyping methods and timepoints to evaluate IDC symptoms in the field and in hydroponics. This analysis split this historical QTL into four discrete linkage blocks, each containing candidate iron stress responsive genes. It is still unknown if different combinations of these linkage blocks can be associated with differences in the stress response and gene expression patterns of soybeans responding to iron stress.

Continued improvements of sequencing technologies allow for scientists to examine genome-wide expression differences in response to stress within any soybean genotypes of interest. Initial soybean gene expression studies using Clark and IsoClark identified responses after 14 days of iron stress, including genes involved in general stress responses, iron uptake/homeostasis, and DNA repair/replication [16–18]. Moran Lauter et al. [19] utilized RNA-seq to study the early transcriptional response (1 h and 6 h after iron stress) in the leaf and root tissue of Clark. Similar to the previous gene expression studies, Moran Lauter et al. found genes involved in what are now considered the hallmarks of the Clark iron stress response: a defense response, iron homeostasis, and DNA replication/methylation. Recently, Moran Lauter et al. [20] found Clark responds to iron stress as early as 30 min after stress. In addition, a shift in gene expression from root to shoot was observed 30–120 min after stress onset. This shift was attributed to the movement of a novel stress signal. Atencio et al. [21] examined Clark and IsoClark responses two and ten days after iron stress, observing an ebb and flow in the gene expression across these same pathways. Remarkably, significant differences in the leaf chlorophyll content could be detected at two days of iron stress.

Thus far, the majority of iron stress studies in both model and crop species have been restricted to a few genotypes of interest. Soybean, however, has evolved around the world in different environments and soil conditions. In order to identify and characterize the full breadth of the soybean iron stress response, including novel iron stress tolerance mechanisms, we need to characterize iron stress responses in multiple genotypes. In this study, the objective was to compare early differential gene expression patterns of soybeans with varying iron efficiencies. We selected 18 lines from the Assefa et al. [12] GWAS panel for iron stress response gene expression analyses. We used RNA sequencing on leaf and root tissue collected 60 min after growing plants in either iron-deficient or iron-sufficient hydroponic solutions. After testing for differential expression, we found a rapid and varied response to iron stress across genotypes. While genome-wide association studies and transcriptomic studies have separately proven to be very useful tools in identifying stress or developmental genes and regions of interest, we must leverage both research tools simultaneously to increase our understanding of IDC responses in soybean. Leveraging diversity found in the soybean germplasm collection can be used to enhance breeding efforts and develop a greater tolerance to nutrient stresses, such as IDC.

2. Results

2.1. Clustering Genotypes into Efficient and Inefficient Classes Based on Phenotypic Data

Phenotypic data provided by Assefa et al. [12] were first clustered using all years, growth stages, and environments (Figure 1, Supplementary File S1). The dendrogram from hierarchical clustering shows a distinct break between two clusters, each containing nine genotypes (Figure 1a). To visualize SPAD readings and IDC ratings in the same heatmap, we standardized the phenotypic values in a given growth stage and environment by using z-scores, converting raw values to standard deviations from the mean. Using z-scores, opposite phenotypic ratings for each trait were easily distinguished into two clusters. The cluster with genotypes G1, G2, G8, G10, G12, G14, G15, G16, and G17 (Clark) generally received low IDC ratings and high SPAD readings and will be denoted as the iron-efficient (EF) group. The cluster with genotypes G3, G4, G5, G6, G7, G9, G11, G13, and G18 (IsoClark) generally received high IDC ratings and low SPAD readings and will be denoted as the iron-inefficient (INF) group. Within each cluster, two subgroups could be identified. In the EF group, G2, G8, G12, and G16 were the best performing lines, whereas, in the INF group, G11 and G18 were the worst performing lines. Interestingly, a greater variation of phenotypic scores between years and environments was seen among the INF group, specifically, G5, G7, and G9. The only two genotypes that consistently showed similar phenotypic scores were G11 and G18 of the INF group. Principal component analysis (PCA) was also used to cluster the genotypes (Figure 1b). The first two principal components explained 90.6% of the variance, (83.3% and 7.3%, respectively). In the PCA plot, the genotypes clustered into the same two groups defined using the hierarchical clustering. Again, we saw that the INF group contained more variation than the EF group based on the distribution of the genotypes on the PCA plot.

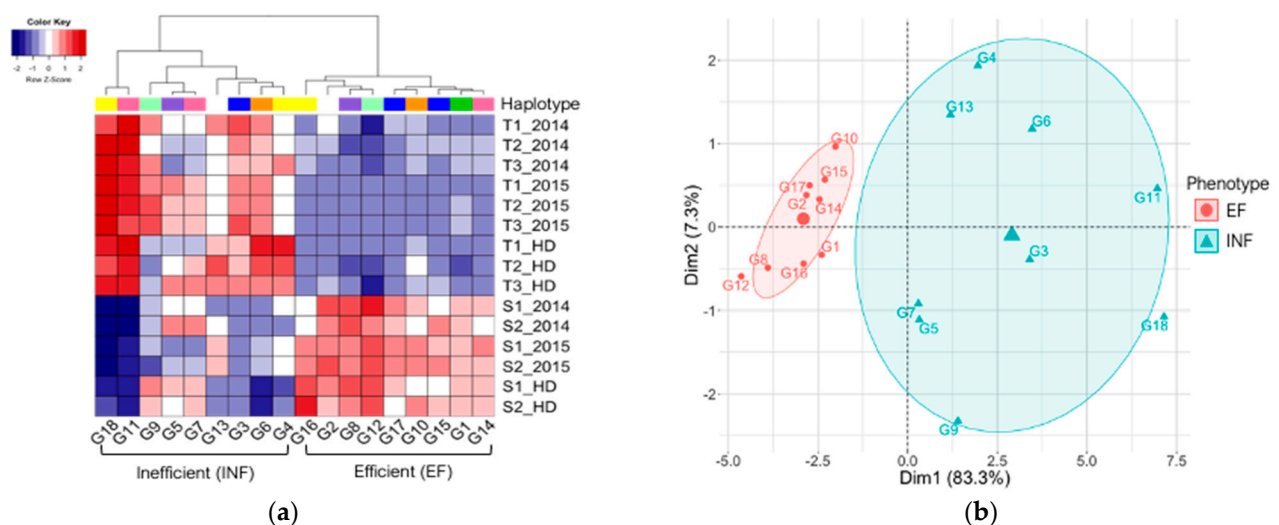


Figure 1. Clustering of 18 soybean genotypes based on phenotypic response to iron stress. (a) Iron deficiency chlorosis (IDC) ratings and soil plant analysis development (SPAD) readings from Assefa et al. (2020) were used to generate a heatmap using row z-scores (standard deviation from the mean within a row). IDC ratings were collected at three growth stages (T1, T2, T3) and SPAD readings were collected at two growth stages (S1, S2) across two field seasons (2014, 2015) and in greenhouse hydroponics (HD). Genotypes ordered based on hierarchical clustering revealed two major groups of soybean genotypes, iron-efficient (EF) and iron-inefficient (INF). Haplotypes were identified from sequences within the historic iron deficiency chlorosis (IDC) quantitative-trait loci (QTL) on chromosome Gm03. Matching haplotypes are indicated in the color bar above the heatmap. (b) Principal component analysis (PCA) was performed on the same data used to generate the heatmap. The two major groups of soybean genotypes revealed by hierarchical clustering, EF and INF, were also seen in the PCA and are shown in red and blue, respectively.

When selecting the genotypes used in this study, we intentionally selected genotypes with eight haplotype combinations for the three IDC QTL linkage blocks identified by

Assefa et al. [12] that corresponded to the narrowed introgressed region identified by Peiffer et al. [15]. For seven of the eight combinations, we selected the two genotypes with the highest and lowest iron stress tolerance, based on three timepoints for visual ratings and two timepoints for SPAD readings [12]. For the eighth combination, only seeds from a single EF genotype was available. Clark (G17, EF) and IsoClark (G18, INF) were then added to the correct haplotype combinations as internal controls. Finally, for one haplotype group, a third genotype was added, representing the average IDC ratings for the group. We found that genotypes clustered by IDC phenotype, not by haplotype, suggesting other genomic regions must impact the IDC tolerance.

2.2. Identification of Differentially Expressed Genes in Early Response to IDC Stress

From the 216 purified RNA samples (eighteen genotypes, two tissue types, two iron treatments, three replicates) that were sent to the Iowa State DNA sequencing facility, approximately 6.2 billion raw reads were produced. The sequences were filtered and mapped to the soybean reference genome, as outlined in the materials and methods. The number of mapped reads varied from 5644 to 186,296,039, with nine samples (eight in leaves and one in roots) containing fewer than five million mapped reads (Supplementary File S2). Using FastQ Screen [22] to examine the quality of the reads, along with the unusually low numbers of mapped reads for some samples, raised concerns about the global coverage and depth of sequencing for nine samples. Two genotypes (G3, G15) each had two replicates with fewer than five million mapped reads in leaf tissue samples under sufficient iron conditions. Similarly, genotype (G9) had three replicates with fewer than five million mapped reads in leaf tissue samples under sufficient iron conditions. Due to the lack of replication and the inability to make treatment comparisons, the three genotypes were completely removed from further analyses in the leaf tissue. The other two samples (corresponding to genotype G8 leaves and genotype G16 roots) identified with fewer than five million mapped reads were in different genotypes and tissue types, leaving at least two replicates after removal. Sample removal resulted in 15 and 18 genotypes used in downstream leaf and root tissue analyses, respectively.

Following the edgeR workflow, we tested the treatment effect of iron deficiency by comparing the expression of genes in deficient conditions against sufficient conditions within each genotype. The number of differentially expressed genes (DEGs, FDR < 0.05) varied considerably across genotypes in both tissue types (Supplementary Table S1, Supplementary Files S3 and S4). The total number of DEGs ranged from 1 to 6747 in leaves and from 16 to 1611 in roots. Plotting the number of DEGs by tissue type across genotypes clearly demonstrated the variability in numbers of DEGs (Figure 2). Within both the EF and INF groups, we identified distinct patterns of DEG numbers. In the EF group, genotypes G1, G2, and G8 had higher DEG counts in both leaves and roots relative to other genotypes in the group. In genotypes G10, G12, G16, and G17, we identified few DEGs from leaves, but many from roots (<100 in leaves and >200 in roots), and genotype G14 had DEG counts <50 in both leaves and roots. This suggests differences in iron stress responses among the EF group. In the INF group, all genotypes aside from G4 had DEG counts <100 in leaves and a range of DEG counts in the roots.

2.3. Comparison of Differentially Expressed Genes between Genotypes

Searching for similar DEGs between individual pairs of genotypes, we compared overlapping DEGs in all pairwise combinations of genotypes (Supplementary Figure S1). The number of overlapping DEGs in a pair of genotypes ranged from 0 to 2837 in leaves and 0 to 135 in the roots. Most comparisons made in the leaf tissue resulted in very few to no overlapping genes. This was not surprising considering fewer than 15 DEGs were found in at least half of the genotypes. However, comparing the three EF genotypes with DEG counts >500, G1 and G8 had 2837 DEGs in common and G1 and G2 had 305 genes in common, whereas G2 and G8 only had 215 DEGs in common. Comparisons in the root tissue resulted in over half of the pairs containing at least 15 overlapping DEGs. Within

the EF group, G1 and G2 had the most DEGs in common (57). These results again suggest differences in iron stress responses across genotypes.

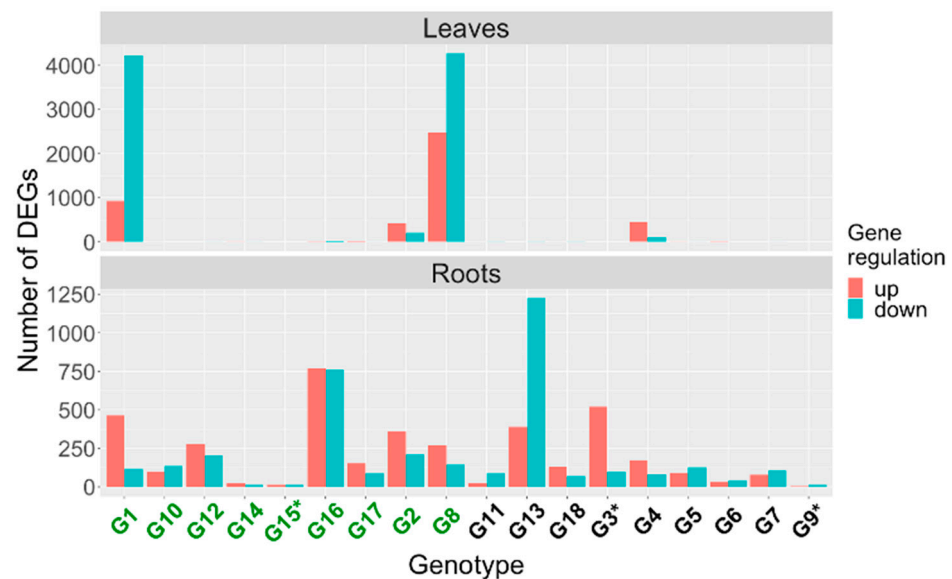


Figure 2. Numbers of differentially expressed genes (DEGs) responding to iron stress across 18 soybean genotypes. Significant DEGs (deficient vs. sufficient; FDR < 0.05) were identified 60 min after iron stress in leaf and root tissue. Genes up-regulated and down-regulated in response to iron stress are shown in red and blue, respectively. Previous hierarchical cluster analysis based on iron stress phenotypic measurements revealed two major clusters of soybean genotypes, iron-efficient and iron-inefficient, shown in green and black, respectively. Three genotypes that were omitted from leaf tissue due to sample removal during sequence processing are indicated with an asterisk.

2.4. Comparisons across Genotypes

2.4.1. Differentially Expressed Genes

In order to identify conserved stress response genes in soybean, we identified DEGs that were most common to all genotypes (Supplementary Table S2, Supplementary File S5). Comparing all genotypes in the leaves, the highest overlap was two DEGs shared by five genotypes, followed by 24 DEGs shared by four genotypes, 192 DEGs shared by three genotypes, and 2992 DEGs shared by two genotypes. Genes that were identified in two or more genotypes were typically found in various combinations between G1, G2, G4, and G8. Three of these genotypes are EF (G1, G2, G8) and one genotype is INF (G4). Most of the overlap between two genotypes occurred between G1 and G8. For the overlap of three genotypes, 97% (187/192) of the genes overlapped with some combination of those four genotypes, and for the overlap of four genotypes, 83% (20/24) of the genes overlapped with those four genotypes. These groupings suggest that some core stress mechanisms may be conserved between these four genotypes.

The two genes, *Glyma.11G190200* and *Glyma.18G104400*, shared across five genotypes, encode a UDP-D-apiose/UPD-D-xylose synthetase and a citrate synthase, respectively. Ahn et al. [23] found silencing UDP-D-apiose/UPD-D-xylose synthetase expression caused changes in plant growth, cell death, and leaf yellowing, similar to IDC. This suggests an increased expression of *Glyma.11G190200* could help to alleviate IDC symptoms. López-Millán et al. [24] demonstrated that iron deficiency caused an increase in the activity of multiple citric acid cycle enzymes, including citrate synthase. An overexpression of the *Malus xiaojinensis Citrate Synthase 1 (MxCS1)* increases the iron stress tolerance in tobacco [25]. The 24 genes found across four genotypes were involved in growth and various hormone responses. For example, *Glyma.06G102100* is homologous to the *AtEXO* gene, which responds to a brassinosteroid stimulus and is required for cell expansion in leaves [26]. Lisso et al. [27] found that *AtEXO* modifies the sugar responsiveness during

seedling growth. Moran Lauter et al. [19] identified eight EXO homologs, including *Glyma.06G102100*, that were repressed in response to 60 min of iron stress in Clark leaves. *Glyma.18G030200* is homologous to the *COI1* gene, which is involved in jasmonate signaling and can inhibit growth and induce defense-related processes [28]. Both *Glyma.06G102100* and *Glyma.18G030200* were down-regulated in response to iron stress in the four genotypes.

In roots, we identified 24 genes found in six or more genotypes, including *Glyma.19G016400*, shared by 17 genotypes. *Glyma.19G016400* is a member of the ATP-binding cassette (ABC) transporter superfamily. This gene family has been associated with many functions of plant development and response, such as the transportation of auxin and secondary metabolites [29]. *Glyma.03G160100*, shared across 14 genotypes, is most homologous to *AtCYP_{94B1}*, which is involved in apoplastic barrier formation in the roots and confers salt tolerance [30]. Remarkably, the 24 genes shared across six or more genotypes were significantly overrepresented with the GO terms' cellular response to potassium ion (GO:0035865) and response to hypoxia (GO:0001666). These genes include two homologs of *AtRAP2.1* (*Glyma.19G026000* and *Glyma.13G060600*, shared in 14 and seven genotypes, respectively) and two homologs of *AtADH1* (*Glyma04G,240800* and *Glyma.06G122600*, shared in ten and eight genotypes, respectively). *AtRAP2.1* is a negative regulator of abiotic stress responses [31], whereas *AtADH1* confers resistance to biotic and abiotic stress [32]. More overlapping genotypes were observed in the roots than in the leaves, suggesting a more uniform recognition and response of the stress in root tissue.

To help examine the response of significant DEGs across all genotypes, we generated heatmaps for the 218 and 349 significant DEGs shared by three or more genotypes in leaves and roots, respectively (Figure 3). The hierarchical clustering of the log₂ fold-change (logFC) allowed us to organize groups of genes and genotypes by similar response profiles. In the leaves, the genotypes clustered into two major clades, but in the roots, the genotypes clustered into three major clades. Interestingly, Clark (G17) was clustered with IsoClark (G18) in roots but not in the leaf tissue. Regardless of the tissue type, there was a combination of EF and INF genotypes within each clade. We found an overlap of genotype groups between tissue types (G13, G16, G18), but were limited in this comparison because of the three genotypes that were removed from the leaf analysis. The clusters found in both tissue types suggest that there are at least two iron stress response mechanisms represented in our mini panel. The hierarchical clustering of the DEGs resulted in two and four major clades in leaves and roots, respectively. Using GO enrichment analysis, we examined the biological processes associated with each DEG clade. In leaves, no GO terms were significantly overrepresented. However, in one nested genotypic clade in leaves (G7, G8, G10, G12, G4, G6, G1, G2), the DEGs were easily grouped by the direction of expression. In the roots, two of the four DEG clades (clade 2 [green] and clade 3 [blue]) contained significantly overrepresented GO terms. Clade 2 was overrepresented with DEGs associated with a response to hypoxia (GO:0001666) and the negative regulation of the ethylene-mediated signaling pathway (GO:0010105). Ethylene is involved in multiple stress signaling, including iron stress [33] and hypoxia [34]. Clade 3 was overrepresented with DEGS associated with a response to hypoxia (GO:0001666) and the sulfolipid biosynthetic process (GO:0046506). The membrane lipid composition can change under stresses, including hypoxia and phosphate starvation [35,36]. Interestingly, Thimm et al. [37] also found expression changes in anaerobic-related genes while using hydroponics but attributed this response to attempted glycerol production rather than a hypoxia response.

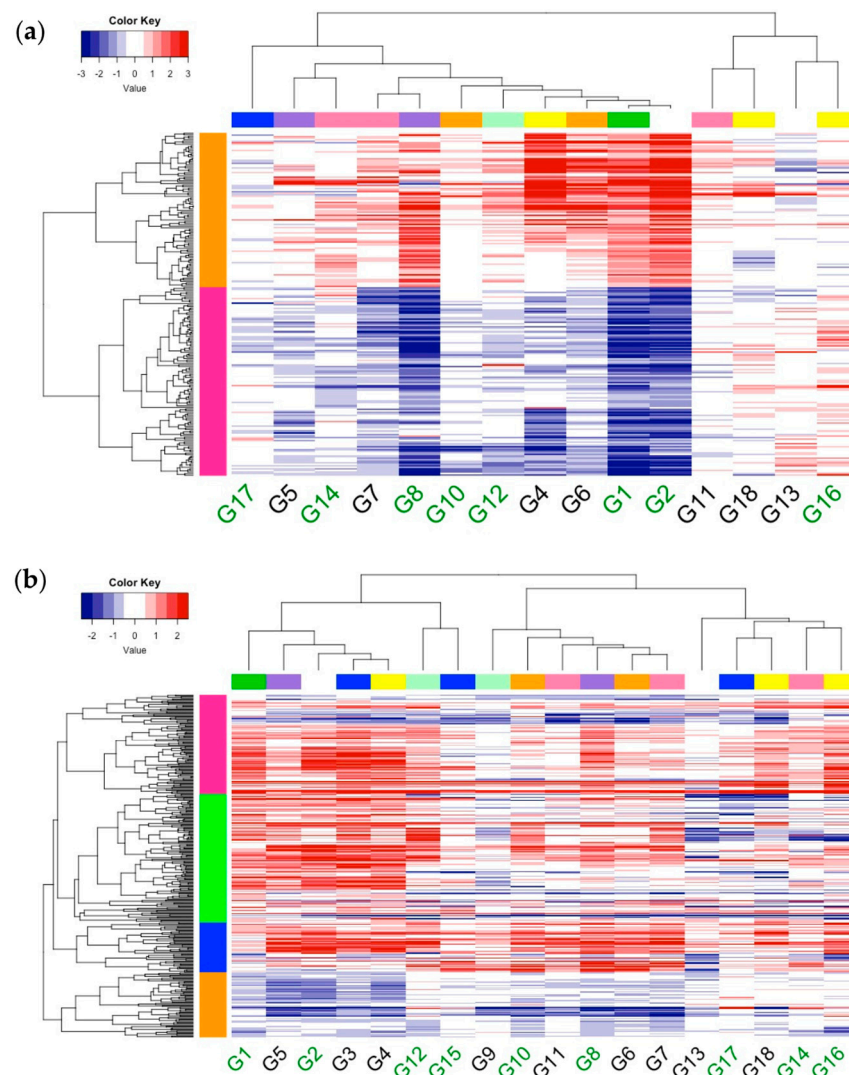


Figure 3. Heatmap of differentially expressed genes across 18 soybean genotypes. Differentially expressed genes (DEGs) in response to 60 min of iron stress response were identified in 18 soybean genotypes. Genes significant in at least three genotypes were identified and then log₂ fold-change values were plotted across (a) 15 genotypes in leaves and (b) 18 genotypes in roots. Three genotypes were omitted from leaf tissue due to sample removal during sequence processing. Haplotypes were identified from sequences within the historic iron deficiency chlorosis (IDC) quantitative trait loci (QTL) on chromosome Gm03. Matching haplotypes are indicated in the color bar above the heatmap. Previous hierarchical cluster analysis based on iron stress phenotypic measurements revealed two major clusters of soybean genotypes, iron efficient and iron inefficient, shown in green and black font, respectively.

2.4.2. Enriched Biological Process Terms

Our analyses thus far suggest that the genotypes used in this study had largely unique responses to iron stress. Of the 9718 DEGs identified in leaves, 97.7% were only found in one or two genotypes. Similarly, in the roots, 93.8% of DEGs were found in one or two genotypes. However, soybean is a paleopolyploid with multiple genome duplication events in its history [38]. Therefore, it is possible that the gene-specific differences we observed could be due to homeologous genes performing similar functions. To address this issue, we took advantage of GO term enrichment. Homeologous genes should have the same best *Arabidopsis* hit and would be assigned the same GO terms, allowing us

to compare biological processes involved in the response to iron stress across genotypes (Supplementary Tables S3 and S4, Supplementary File S6).

In leaves, we found 106 non-repetitive overrepresented GO terms across genotypes. Only two genotypes (G1, G8) had overlapping significant GO terms (corrected p -value < 0.05). The lack of DEGs found across genotypes likely contributed to the lack of shared overrepresented GO terms. However, within these two genotypes, were GO terms directly associated with iron (iron–sulfur cluster assembly [GO:0016226] and iron ion homeostasis [GO:0055072]). Additionally, other GO terms related to the defense response and photosynthesis were shared by these two genotypes.

In the root tissue, we found 54 non-repetitive overrepresented GO terms across all genotypes, with up to seven genotypes that were significant for the same overrepresented GO term. Although no GO terms were found with the iron specifically mentioned in the description, multiple GO terms were associated with the signaling or response to stress. The GO term with the highest number of overlapping genotypes (five) was the response to hypoxia (GO:0001666). GO terms that were significant in four genotypes were the hydrogen peroxide metabolic process (GO:0010310), systemic acquired resistance, and the salicylic-acid-mediated signaling pathway (GO:0009862). Hydrogen peroxide (H₂O₂) is produced by plants under waterlogged conditions and is associated with stress signaling [39]. Moran Lauter et al. [20] identified GO:0009862 (systemic acquired resistance, salicylic-acid-mediated signaling pathway) as one of the top ten overrepresented GO terms in one cluster of DEGs responding to iron stress in the root.

2.5. Comparing Differentially Expressed Genes between Iron Efficiency Groups

In order to understand the relationships between EF and INF genotypes, we compiled a list of total unique DEGs for each tissue type (9720 unique DEGs in leaves and 5633 unique DEGs in roots). We then identified DEGs unique to either EF or INF genotypes and quantified the number of genotypes the DEG considered significantly differentially expressed (Supplementary Table S2, Supplementary Files S7 and S8). Identifying EF- and INF-specific DEGs would help to separate specific iron stress responses from general stress responses.

We identified 9141 and 2848 DEGs unique to EF leaves and roots, respectively (Supplementary File S7). In EF leaves, 6160 were unique to a given genotype, 2859 were shared by two genotypes, 123 were shared by three genotypes, and one was shared by four genotypes. While looking at *Arabidopsis* homologs of DEGs shared by three or more EF genotypes, we found many genes related to stress signaling and tolerance. The only gene unique to four EF genotypes was *Glyma.13G155200*, with the greatest homology to *AtTIL*, a temperature-induced lipocalin. Charron et al. [40] found that an overexpression of *AtTIL* enhanced stress tolerance, delayed flowering, and maintained leaf greenness. Interestingly, this gene was down-regulated in all four EF genotypes under iron-deficient conditions. GO term enrichment analyses of the 9141 DEGs unique to EF leaves identified 93 significant terms, including numerous terms associated with photosynthesis, defense, cell division, iron homeostasis, and the response to iron. In EF roots, 2677 DEGs were unique to a given genotype, 159 DEGs were common to two genotypes, and 12 DEGs were common to three EF genotypes. Among the genes shared by the three EF genotypes in roots were genes responding to stress-related hormones, including: *Glyma.05G222400*, with the greatest homology to *AtACO1*, involved with ethylene biosynthesis and the induction of iron acquisition genes [41]; *Glyma.15G062400*, with the greatest homology to *AtPRB1*, a basic pathogenesis-related protein induced by the defense signaling molecules methyl jasmonate and ethylene [42]; and *Glyma.20G248100*, a bHLH transcription factor (*AtAKS2*) negatively regulated by the stress-related hormone abscisic acid [43]. GO term enrichment analyses of the 2848 DEGs unique to EF roots identified only five significant GO terms: RNA methylation (GO:0001510), nucleosome assembly (GO:0006334), chalcone biosynthesis (GO:0009715), ribosome biogenesis (GO:0042254), and the response to gravity (GO:0009629).

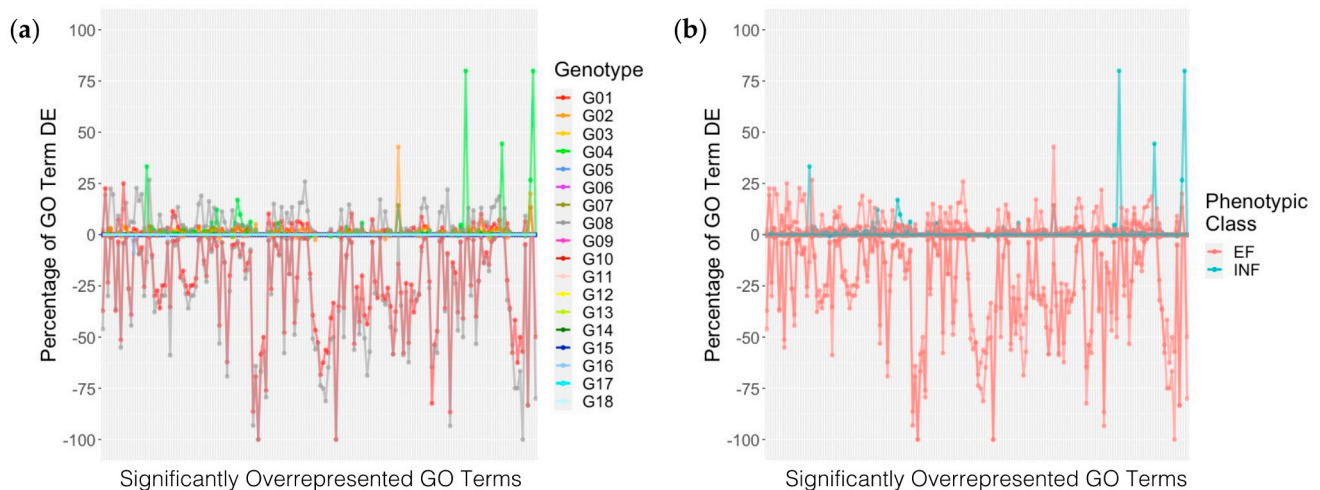
We identified 354 and 2026 DEGs unique to INF leaves and roots, respectively (Supplementary File S8). For INF leaves, 350 DEGs were unique to a given genotype and four DEGs were shared by two genotypes. These encoded a PHD transcription factor, a fatty acid desaturase, and two Kunitz trypsin inhibitors. *AtKT11*, with the greatest homology to *Glyma.08G342100*, regulates plant cell death in plant pathogen interactions [44]. GO enrichment identified 12 significantly overrepresented GO terms, largely associated with hormones (jasmonic acid biosynthesis [GO:0009695], signaling [GO:0009867], response [GO:009753] and metabolism [GO:0009694], ethylene biosynthesis [GO:0009693], and abscisic acid signaling [GO:0009738]) and defense and wounding responses (response to fungus [GO:0009620], regulation of insect defense response [GO:2000068], and response to wounding [GO:0009611]). In INF roots, 1939 DEGs were unique to a given genotype, 77 DEGs were shared by two genotypes, nine DEGs were shared by three genotypes, and one DEG was shared by four and five genotypes. Among the ten genes shared by three or more genotypes, were three genes involved in abiotic stress responses. *Glyma.18g064900*, shared by three genotypes, is most homologous to *AtPTR1* which transports dipeptides as a nitrogen source in the roots [45]. *Glyma.17g162000*, shared by five genotypes, is most homologous to *AtLBD38*, which negatively regulates genes involved in nitrate uptake and assimilation, including members of the peptide transporter family [46]. *Glyma.14g171700*, shared by three genotypes, shares the greatest homology with *AtTPPI*, which regulates stomatal apertures to enhance drought tolerance and water use efficiency [47]. GO term enrichment analyses of the DEGs unique to INF roots identified 16 significant terms associated with defense and wounding responses, including numerous cell wall modifications (secondary cell wall [GO:0009834] and cell wall biogenesis [GO:0009832], cell wall macromolecule metabolism [GO:0044036], lignin [GO:0009809], coumarin [GO:0009805], cellulose [GO:0030244] and glucuronoxylan [GO:0010417] biosynthesis, glucuronoxylan metabolism [GO:0010413], and lignin catabolism [GO:0046274]).

2.6. Characterization of DEG Expression Trends within Biological Processes Terms

Identifying biological relevance in GO terms in the leaves and roots was challenging due to the lack of statistically significant overrepresentation across genotypes. When reviewing the compiled list of GO terms, many GO terms that were not significantly overrepresented still contained significant DEGs across genotypes. For example, in the GO analysis above, in leaves, 15 GO terms associated with hormone signaling and the defense response were only significant in G4. We found comparable numbers of DEGs associated with these GO terms, but not significant, in G1 and G8. In roots, GO:0009862 (systemic acquired resistance, salicylic-acid-mediated signaling pathway) and GO:0010310 (the regulation of hydrogen peroxide metabolic process) were significantly overrepresented in four genotypes (G2, G3, G4, G5) but have DEGs associated with these GO terms in 17 genotypes. This raised the possibility that our results were impacted by the timing differences between genotypes, and not mechanistic differences. To address this, we needed to visualize how GO terms responded to iron stress across genotypes. We separated DEGs within genotypes based on the direction of expression (induced by iron stress vs. repressed by iron stress). We used GO term enrichment on each DEG list and then compiled a nonredundant list of significant biological process terms. We then used this list to determine how many DEGs were identified in each genotype for each GO term. We identified 168 and 90 significant GO terms in leaves and roots, respectively. We then plotted this data to visualize the expression patterns of different biological process across genotypes in leaves and roots; plots were colored by the genotype and by the iron stress phenotype (Supplementary Figure S2, Supplementary File S9). To adjust for GO terms with small genome counts that would not be easily identified in the plot, we calculated the percentage of genes in the GO term that were significantly differentially expressed relative to the GO term genome count for all GO terms and genotypes (Figure 4, Supplementary File S9). If differences between genotypes were simply timing, all genotypes would have varying peaks under the same GO term. We would expect this pattern for genotypes

G1 and G8, which had a number of overlapping DEGs and GO terms in the leaves. If differences between the genotypes were mechanistic, we would see differences in the GO terms associated with different genotypes. For ease of interpretation, we focused on GO terms and genotypes where the number of DEGs was greater than 2% of all genes assigned to the GO term.

GO BP terms significantly differentially expressed in response to iron stress in leaves (166 terms)



GO BP terms significantly differentially expressed in response to iron stress in roots (90 terms)

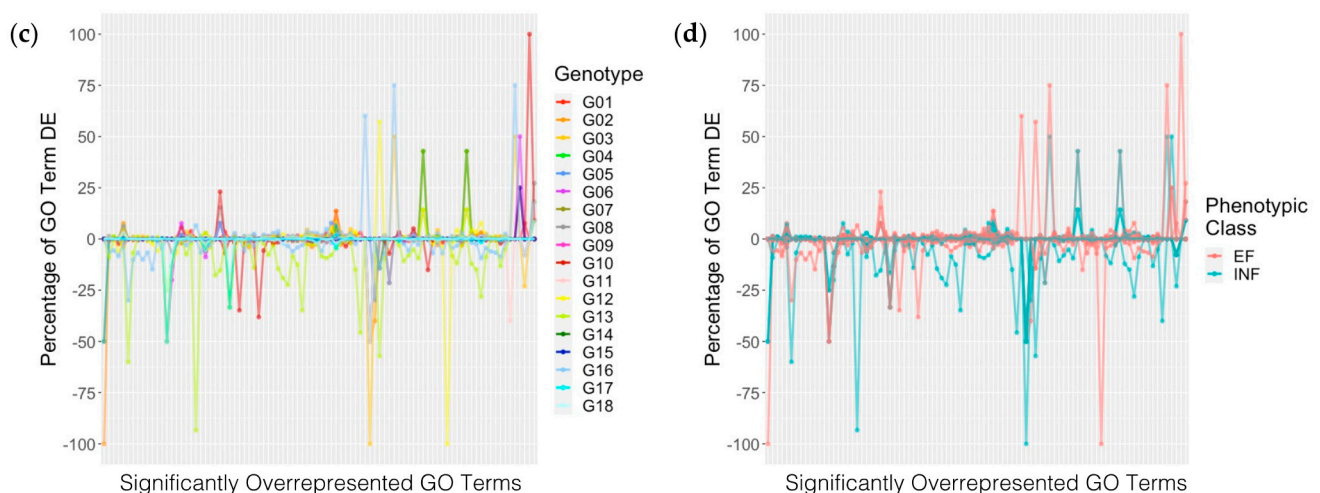


Figure 4. Percentage of differentially expressed genes (DEGs) associated with select gene ontology (GO) terms in leaf tissue of 18 soybean genotypes. GO term enrichment analysis was used on DEGs that were up-regulated or down-regulated in response to 60 min of iron stress in (a,b) leaf and (c,d) root tissue of each genotype. DEG numbers for GO terms that were significant in at least one genotype were compiled across genotypes. The percentage of DEGs expressed relative to the total count of that GO term across the genome was calculated for each genotype and plotted with up-regulated genes shown as positive values and down-regulated genes shown as negative values. Genotypes G3, G9, and G15 have values of zero in the leaves due to sample removal during sequence processing. (a,c) Each genotype is represented by a unique color. (b,d) Previous hierarchical cluster analysis based on iron stress phenotypic measurements revealed two major clusters of soybean genotypes, iron-efficient and iron-inefficient, shown in red and blue, respectively. Additional data available in Supplementary File S9.

For the vast majority of GO terms plotted, the response was largely genotype-specific. One or two genotypes had similar GO term expression patterns, whereas the remaining genotypes had background DEG expression levels that were less than 2%. In the roots, G13

(INF) and G16 (EF) had similar expression trends among 27 GO terms repressed in both genotypes. Interestingly, 17 additional GO terms were induced in G16, but repressed in G13, suggesting differences in timing between the two genotypes, and not different iron stress mechanisms. G2 (EF) also shared six repressed GO terms with G13 (INF), which were induced in G16 (EF). These GO terms included four signaling hormones involved in the stress response: ethylene (GO:0009873), salicylic (GO:0009862 and GO:0009863), and jasmonic acid (GO:0009753) [48,49]. In the leaves, genotypes G1 (EF) and G8 (EF) shared similar expression trends for 135 GO terms among repressed DEGs and 54 GO terms among induced DEGs. Genotypes G2 (EF) also shared 28 GO terms with genotypes G1 and G8, but only among induced DEGs (expression > 2%). This suggests that genotypes are using different strategies to cope with iron stress.

Comparing phenotypic groups in leaves, we found three EF genotypes (G1, G2, G8) with induced gene expression and two EF genotypes with repressed gene expression (G1, G8). However, in leaves, only one INF genotype responded (G4). Remarkably, G4 only induced gene expression (>2%). If we examine the 168 GO terms identified in leaves, 141 are specific to EF groups (INF expression < 2%), 3 are specific to INF genotypes (EF expression < 2%), and 24 are common to the EF and INF genotype. The three GO terms unique to INF (G4) included the regulation of sequence-specific DNA binding transcription factor activity (GO:0051090), regulation of transcription from RNA polymerase II promoter in response to oxidative stress (GO:0043619), and regulation of the defense response to insects (GO:2000068). Eight GO terms involved in jasmonic acid processes and defense response were expressed in G1 (EF), G4 (INF), and G8 (EF). EF-specific terms were associated with a range of processes, including photosynthesis, methylation, defense, iron homeostasis, the regulation of transcription and translation, and growth. A lack of expression of other INF genotypes in the leaves suggests that either INF genotypes are not able to respond to iron stress in the leaves, or that INF roots are unable to signal iron stress to the leaves, which could be an important distinction between EF and INF genotypes. Additionally, the range of responses found in EF leaves suggests a cascade of iron stress responses, whereas the response of INF leaves seems to be a more general defense response.

We saw induced and repressed GO terms in the root for EF and INF genotypes. If we examine each GO term in the root, 23 were specific to EF groups (expression < 2% across INF genotypes), 3 are specific to INF groups (expression < 2% across EF genotypes), and 64 could be found in EF and INF genotypes. INF-specific GO terms were associated with nucleotide–sugar metabolism (GO:0009225), the response to fructose (GO:0009750), and chaperone-mediated protein folding (GO:0061077). EF-specific terms were associated with stress, defense, DNA replication, cell division, and methylation. Interestingly, two genotypes (G14, G15) had little to no overlap of GO terms in roots, suggesting distinct iron stress responses.

2.7. Characterization of Differentially Expressed Transcription Factors

In order to identify regulators of potential pathways of interest, we identified DEGs annotated as transcription factors (Supplementary Table S5, Supplementary File S5). Log₂ fold-change values of differentially expressed transcription factors (TFs) grouped by the transcription factor family (TFF) were plotted for each genotype × tissue type (Figure 5). In leaves, we identified 897 TFs belonging to 56 TFFs. Most (92%) of the TFs were unique to EF genotypes, 43 TFs (5%) were unique to INF genotypes, and only 25 (3%) of TFs were found in at least one genotype of each phenotypic group. Of the 56 TFFs identified in leaves, 16 TFFs were found in both phenotypic groups, 40 TFFs were unique to EF in leaves, and no TFFs were unique to INF in leaves. In roots, we identified 569 TFs belonging to 49 TFFs. Almost half of the TFs (47%) were unique to EF, fewer TFs were unique to INF (36%), and only 17% of TFs were found in at least one of each phenotypic group. Similar to leaves, all TFFs identified in INF genotypes were identified in EF genotypes, whereas 12 TFFs were unique to EF in roots. Interestingly, 71% and 78% of the TFs were unique to a single genotype in the leaves and roots, respectively. An overlap of TFF between phenotypic

groups could suggest similar target pathways for a general stress response, with additional target pathways that distinguish the EF genotypes.

The expression patterns in TFs were similar to the expression patterns of total DEGs. We found that EF genotypes (G1, G2, G8) had relatively strong numbers in the leaves and roots. Most of the other EF genotypes (G10, G12, G16, G17) had consistent numbers of TFs in the roots, but little to no TFs in the leaves. The remaining EF genotypes (G14, G15) had few TFs in either the leaves or roots. Most TFs for INF genotypes were identified in the roots. Only one INF genotype (G4) had a noticeable number of TFs in the leaves. Two INF genotypes (G6 and G13) had only one and two TFs identified in the leaves, respectively. Higher numbers of TFs observed in the leaves of EF genotypes suggests a more active response to the iron stress.

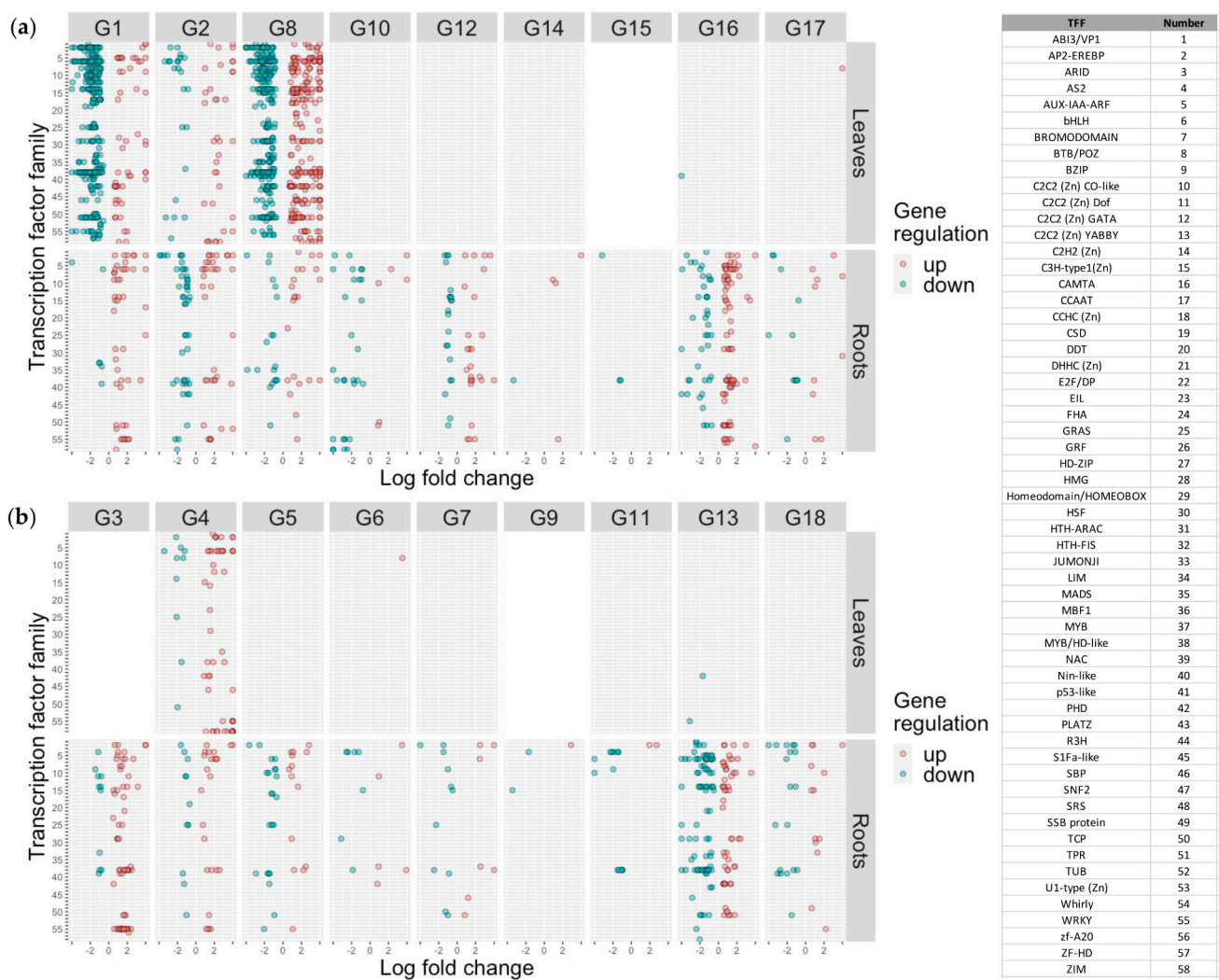


Figure 5. Expression values of differentially expressed transcription factors of 18 soybean genotypes. Differentially expressed genes were identified in leaf and root tissue in response to 60 min of iron stress. Transcription factors were identified in each DEG list and then plotted by transcription factor family using the \log_2 fold-change of the stress response. Genes considered up- or down-regulated are shown in red or blue, respectively. Genotypes are divided by iron efficiency ((a) iron-efficient, (b) iron-inefficient) based on hierarchical clustering of phenotypic data. Genotypes G3, G9, and G15 contain blank cells in leaf tissue due to sample removal during sequence processing.

We performed TF enrichment analyses to identify overrepresented transcription factor families (TFF; Supplementary Figures S3 and S4, Supplementary File S10). Surprisingly, we did not find many significantly overrepresented TFFs, suggesting a range of processes that are targeted during the stress response. In leaves, we identified four overrepresented

TFFs: C2C2 [Zn] CO-like (G1 and G8), MYB/HD-like (G1), SBP (G8), and ZIM (G2 and G4). In roots, we identified three overrepresented TFFs: AS2 (G6 and G11), WRKY (G3), and ZIM (G10). Most overrepresented TFFs in leaves overrepresented TFs belonged to EF genotypes, and, in roots, overrepresented TFs belonged mostly to INF genotypes.

2.8. Characterization of Differentially Expressed Genes across IDC QTL

The increasing body of IDC literature has identified numerous regions across the soybean genome that are associated with the iron stress response. Assefa et al. [12] identified 69 regions of interest as a result of a GWAS using 460 diverse lines, including the lines in our mini-panel. The 69 regions correspond to 278 candidate genes, some supported with previous RNA-seq data from our group, which were restricted to genotypes Clark (G17, [19,20]) and IsoClark (G18, [50]). Cross referencing our lists of overlapping DEGs in both tissue types (Supplementary File S5), we identified 67 candidate DEGs corresponding to 43 of these 69 regions. A total of 49 genes were unique to leaves, 12 genes were unique to roots, and 6 genes were common to both tissue types. Some DEGs are specific for iron-stress responses, whereas others are for more general stress responses. For example, *Glyma.02G075100* is homologous to *AtSUC2*, a sucrose transporter gene that increases iron deficiency tolerance when overexpressed in *Arabidopsis* [51]. Moran Lauter et al. [19] found that *Glyma.16g157100*, which is homeologous to *Glyma.02G075100*, is induced in Clark (G17) leaves six hours after iron stress. *Glyma.05G000300* encodes an iron-sulfur cluster containing ferredoxin-thioredoxin reductase enzymes and was identified as a candidate gene for IDC tolerance by Butenhoff [52] using the Fiskeby III x Mandarin (Ottawa) mapping population. *Glyma.06G056400*, homologous to *AT2G26330*, encodes a leucine-rich repeat receptor-like kinase. Shanmugam et al. [53] overexpressed a truncated dominant-negative *Arabidopsis* ERECTA gene in soybean and observed a decrease in plant development and an increase in stress response. *Glyma.14G031700* is homologous to *AtWDR26*, a WD-40 repeat containing protein. An overexpression of *AtWDR26* induced gene expression across a range of processes, including hormone, light, and abiotic stress [54]. Among the six genes common to both tissue types was *Glyma.03G144500*, which is homologous to the *FAD2* gene. Yuan et al. [55] found that *FAD2* is involved in the plant response to phytohormones and abiotic stress.

Based on the location of the Gm03 QTL defined by Assefa et al. [12], we further explored DEGs within this region. We identified 10 DEGs unique to leaves, four unique to roots, and one gene significant in both tissue types. Surprisingly, DEGs that were located within the Gm03 QTL were only identified in five genotypes, not including Clark or IsoClark, between both tissue types. In leaves, only G1 and G8 had DEGs identified in the Gm03 QTL, two of which were significant in both genotypes (*Glyma.03G128900* and *Glyma.03G130300*). Assefa et al. [12] identified *Glyma.03G128900*, homologous to *AtLCY*, as a high priority candidate gene in region one of the Gm03 QTL. The transformation of β -lycopene cyclase genes from *Salicornia europaea* L. into both *Arabidopsis* and tobacco increased carotenoid retention and improved oxidative and salt stress tolerance [56]. O'Rourke et al. [57] also identified *Glyma.03G130300* as differentially expressed in leaves 24 h after iron stress and in roots after multiple exposures to iron and phosphate stress. Another gene of interest was *Glyma.03G128300*, which is homologous to the glutamate synthase, *AtGLU1*. Knock-down *Arabidopsis* mutants showed large transcriptional changes to various pathways, including photosynthesis and stress response [58], while Cui et al. [59] found *AtGLU1* to be involved in iron homeostasis. In roots, three genotypes (G2, G13, G16) had DEGs in the Gm03 QTL. Remarkably, two genes (*Glyma.03G131200* and *Glyma.03G131400*) were annotated as the same protein, but were differentially expressed in different genotypes (G13 and G2, respectively). Both genes were homologous with members of the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily. These non-heme enzymes utilize ferrous iron as a co-factor, catalyze a wide range of reactions, and are potentially involved in sensing the iron status [60,61]. Moran Lauter et al. [19] identified *Glyma.07g150900*, also a member of the 2OG-Fe(II)-dependent oxygenase superfamily,

as differentially expressed in Clark (G17) roots one hour after iron stress. *Glyma.03G130200* was identified in leaves (G1) and roots (G16) and is homologous with a strictosidine synthase-like (SSL) protein. Sohani et al. [62] demonstrated that members of the SSL gene family are involved in plant defense mechanisms.

Zhang et al. [63] used image analysis and machine learning to rate iron deficiency chlorosis. In a GWAS using the image analysis output, they identified seven QTL associated with iron deficiency across the genome. Within an 847 kb region on Gm03 (overlapping the historic IDC QTL on Gm03), they identified seven candidate genes. One of the seven candidate genes located on Gm03 (*Glyma.03G128300*) was identified in the leaves (G8) and two (*Glyma.03G131200* and *Glyma.03G131400*) were identified in the roots (G13, G2). All three genes on Gm03 were highlighted in the previous paragraph. An additional 2OG-Fe(II)-dependent oxygenase (*Glyma.18G111000*) 41.4 kb downstream from another variant found on Gm18 was also identified in the leaves (G8). These findings highlight the utility of leveraging early gene expression studies with GWAS field studies to identify candidate genes controlling agronomically important traits.

2.9. Single Linkage Clustering

We used single linkage clustering to group iron-stress-responsive DEGs (13,980) by shared sequence homology (TBLASTX, $E < 10^{-20}$) or individual genes shared across multiple genotypes, tissues, or expression patterns. Of the 13,980 unique DEGs identified in our experiment, 12,138 DEGs clustered into 2922 clusters. Clusters ranged in size from one DEG to 2136 DEGs, and represented up to 18 genotypes (Supplementary Figure S5). Of the 2922 clusters, 1763 and 50 were specific to EF and INF genotypes, respectively. On average, EF clusters contained 2.28 DEGs (STD = 1.9), whereas INF clusters contained 2.02 DEGs (STD = 1.37). Similarly, EF clusters represented 2.28 genotypes (STD = 0.65), whereas INF clusters represented 1.48 genotypes (STD = 0.58). The limited number of genotypes found on average in each cluster again suggests that most genotypes respond very differently to iron stress.

3. Discussion

Soybean is a major cash crop grown in the Midwest; due to various soil properties, soybeans grown in this geographic region of the United States have a higher chance of developing the nutrient stress, iron deficiency chlorosis. Although many studies have contributed to the current knowledge of the molecular response of soybean to IDC, no study has investigated the variation of the molecular response across a wide breadth of the germplasm collection. Similarly, studies in model species have largely focused on one or two main genotypes. Therefore, we sought to compare the early responses to IDC across a diverse panel of soybean genotypes to identify both differences in the stress response across genotypes and novel IDC tolerance mechanisms to exploit in the future.

3.1. Soybean Responds Rapidly to Iron Stress

Plants have the ability to rapidly respond to changes in environmental conditions in scales of seconds and minutes [64]. Buckhout et al. [65] examined the early iron stress response of *Arabidopsis* in a time-course analysis by examining differential expression at 0, 0.5, 1, 6, and 24 h after iron stress. While their analysis shows differential expression at all timepoints, they concluded the initiation of the iron deficiency stress response to be sometime between 1 and 6 h after stress. They interpreted that DEGs identified at the first three timepoints were not iron-specific because they were only identified at a single timepoint. In soybean, Atencio et al. [21] compared Clark (G17) iron stress responses observed by Moran Lauter et al. (30, 60, 120 min after iron stress) [20] and O'Rourke et al. (24 h after iron stress) [57] to their own study (two and ten days after iron stress). Of the 9102 and 15,881 DEGs unique to leaves and roots, respectively, approximately 60% were unique to a single time point. While the majority of genes were specific to a given timepoint, they included the hallmarks of the Clark (G17) iron stress response: genes involved iron

homeostasis, defense response, and DNA replication/methylation [18–20,57]. In this study, 67% and 82% of DEGs identified in leaves and roots, respectively, were unique to a single genotype. This suggests that the majority of soybean genotypes in our panel, and not just Clark (G17), are able to recognize and respond to iron stress within 60 min.

Khan et al. [66] examined expression levels of the canonical *Arabidopsis* genes *OPT3*, *FIT*, and *IRT1* and detected expression at 4, 8, and 12 h after iron stress, respectively. Since *OPT3* was detected earliest in the leaves, and *FIT* and *IRT1* were detected later in the roots, they suggested that leaves sense changes in iron availability more quickly than roots. In contrast, Moran-Lauter et al. [20] found higher numbers of DEGs in Clark (G17) roots than in the leaves at the earliest timepoint of 30 min after stress, suggesting that roots respond more quickly than leaves to iron stress in soybean. Examining GO terms across timepoints and tissues revealed that the same GO terms were affected, first in the roots, then in the leaves, suggesting a root-to-shoot signal in soybean. Here, we identified varying numbers of DEGs in the leaf and root tissue across 18 soybean genotypes. For the majority of genotypes, more DEGs were identified in the roots than the leaves, supporting early root-to-shoot signaling in soybean. Only four genotypes had more DEGs identified in the leaves than the roots. Interestingly, three of the four genotypes with more DEGs in leaves than roots were EF (G1, G2, G8), suggesting that these lines respond faster than Clark (G17), where leaf expression was just beginning at 60 min [20]. Future gene expression studies using a variety of soybean genotypes would benefit by including multiple timepoints to enhance our understanding of the timing and movement of the stress signal across genotypes.

3.2. Diversity of Iron Stress Responses Found within the Soybean Germplasm Collection

Many studies across plant species have utilized RNA sequencing (RNA-seq) to identify genes, pathways, and networks that are triggered in response to stress. Due to their expense, early RNA-seq studies focused on one or two genotypes with contrasting stress responses. Recently, studies have begun to increase the number and diversity of genotypes used with RNA-seq to identify novel genes and pathways associated with a trait or stress response [67–70]. Stein and Waters [71] and Waters et al. [72] compared the iron stress response from the root and rosette tissues of five *Arabidopsis* ecotypes. Their conclusions highlighted a small handful of ‘core’ iron stress response genes overlapping between ecotypes. The differentially expressed genes not shared between ecotypes were thought to represent genotype × environment interactions, and not primary Fe-responsive genes. However, genotype by environment (G×E) interactions are critically important for crop improvement. A recent review by Cooper and Messina [73] highlighted the importance of leveraging cross-disciplinary approaches in order to both understand G×E interactions and accelerate crop improvement. Within soybean, traditional genetic studies demonstrate the existence of multiple iron stress tolerance mechanisms. Lin et al. [9] used two mapping populations to study the IDC response in soybean. One population (Pride B216 × A15) found a minor effect QTL on six linkage groups, and the other population (Anoka × A7) found a single major effect QTL, suggesting that there are at least two distinct mechanisms that control the IDC response in soybean. Butenhoff [52] and Merry et al. [74] used the same mapping population (Fiskeby III × Mandarin [Ottawa]) and identified QTL on three chromosomes. Both studies found a QTL on Gm05, and Merry et al. [74] additionally identified QTL on Gm03 (same as previously identified IDC QTL [9,15]) and Gm06. Merry et al. [74] suggested that the QTL on Gm05 contains significant variation for future breeding efforts due to low minor allele frequencies of the iron-inefficient alleles on Gm03 and Gm06 among elite breeding lines. In this study, we identified DEGs in the same regions on all three chromosomes defined by Merry et al. [74] and for 43 GWAS regions identified by Assefa et al. [12], suggesting that these regions, identified in different genotypes and studies, contain important genes for iron stress responses in soybean. In Figure 1, the EF genotypes clearly cluster by the phenotype. In Figure 4, the EF genotypes have largely distinct expression patterns and mechanisms from themselves and INF genotypes. We

believe that these differences represent novel resources to enhance the iron stress tolerance in soybean.

3.3. Identifying Targets for Future Analyses

We have cross referenced the DEGs identified in this study with Clark gene expression studies conducted by Moran Lauter et al. [19], Moran Lauter et al. [20], Atencio et al. [21], and O'Rourke et al. [57]. Of the 9718 and 5632 unique DEGs identified in the leaves and roots of this study, 5491 (56.5%) and 3493 (62.0%) were identified in at least one tissue sample of the Clark studies above (Supplementary Tables S6–S8). Remarkably, 4227 and 2140 DEGs from leaves and roots remain unique to this study. Of these, 1247 DEGs from leaves and 289 DEGs from roots were identified in at least one other genotype. In Supplementary File S5, we provide the DEGs identified in this study and the corresponding genotype information (total Genotypes, EF Genotypes and IN Genotypes). We have cross referenced the DEGs with the previously identified Clark iron-stress-responsive DEGs, we have identified DEGs falling within GWAS QTL identified by Assefa et al. [12], and we have provided multiple annotation sources. It is our hope that we and others can use this information to prioritize candidate genes for future functional characterization in soybean and other crop species.

To demonstrate novel ways that these data sets could be leveraged, we focused on the 25 largest EF-specific clusters identified with single linkage clustering (Supplementary File S11). In order to investigate if the EF clusters might interact, we took the 308 DEGs corresponding to the 25 EF-specific clusters and identified their best Arabidopsis homolog (120 total unique proteins). We then used STRING (ver. 11.5, [75]) to visualize interactions among the clusters (Figure 6).

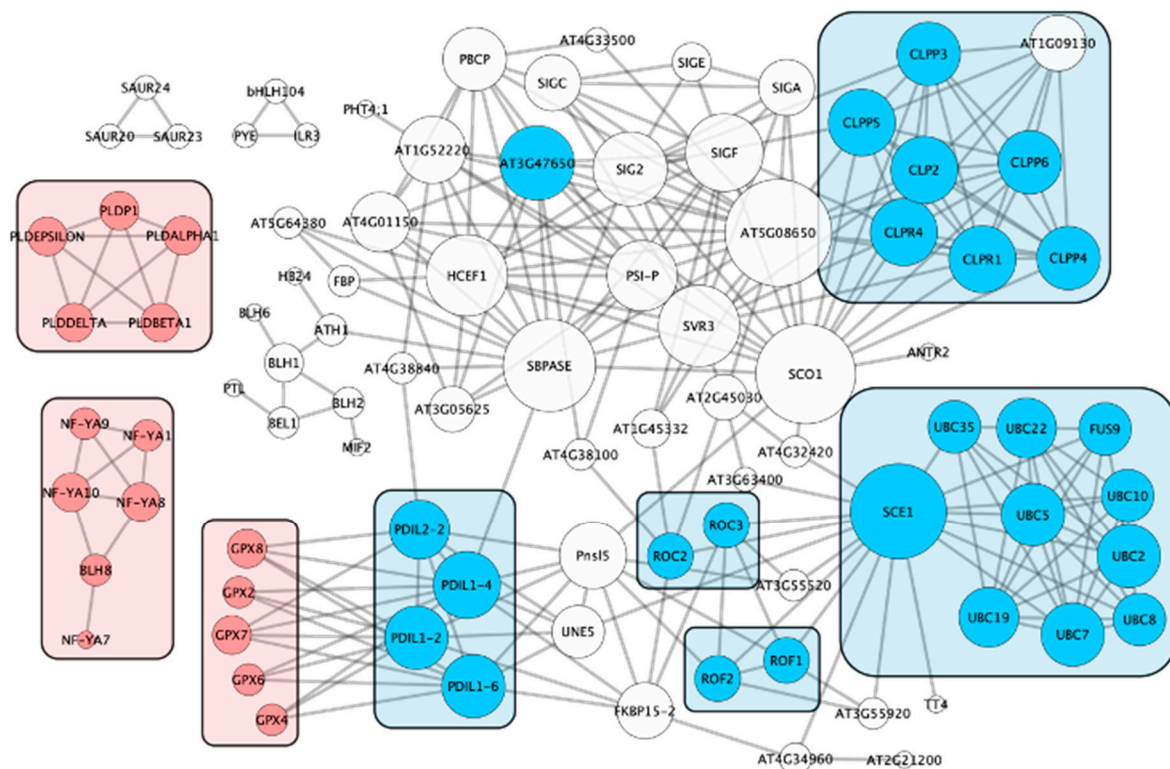


Figure 6. Interactions of Arabidopsis homologs of differentially expressed soybean genes. Differentially expressed genes (DEGs) were identified across 18 soybean genotypes and two tissue types (leaves and roots) 60 min after iron stress. Single linkage clustering was used to identify DEGs with shared sequence homology. Previous hierarchical cluster analysis based on iron stress phenotypic measurements revealed two major clusters of soybean genotypes, iron-efficient (EF) and

iron-inefficient (INF). Arabidopsis homologs were identified for the 25 largest EF-specific clusters and used with STRING (version 11.5) to identify protein interactions of the Arabidopsis homologs. Cytoscape (version 3.7.2) was used to visualize the interaction network of proteins with at least one interaction. Six soybean clusters, highlighted in blue, were associated with protein regulation, including quality control (cluster 606), folding (clusters 258, 334, 406, and 842), and modification (cluster 392). Three soybean clusters, highlighted in red, were associated with endoplasmic reticulum (ER) stress (clusters 218, 556, and 689).

The majority of clusters were associated with a single large network. Within the network, six clusters could be directly associated with protein regulation, including quality control (cluster 606), folding (clusters 258, 334, 406, 842) and modification (cluster 392). CLP proteases (cluster 606) degrade misfolded proteins [76]. Peptidyl-prolyl cis-trans isomerases (clusters 258 and 406, [77]), protein disulfide isomerases (Cluster 334, [78]), HSP40s (see review [79]), and other chaperones (cluster 842, [80]) catalyze protein folding in the endoplasmic reticulum (ER). Ubiquitination targets proteins for degradation [81], whereas SUMOylation can regulate protein function (cluster 392, [82]). In plants, abiotic and biotic stress can result in misfolded proteins, which accumulate in the ER and cause ER stress, toxicity, and programmed cell death (see review [83]). In order to maintain ER homeostasis, cells activate the unfolded protein response, upregulating genes involved in preserving the protein quality and quantity [84]. While genes involved in the unfolded protein response were not statistically overrepresented in our study, they were significantly overrepresented in Clark roots at 30 min after iron stress [20]. Since this study focused on 60 min after iron stress, it suggests that we are observing downstream stages of the unfolded protein response, and not the initiation. The analysis of overrepresented terms within STRING supports this hypothesis: DEGs associated with protein quality control (GO:0006515) and protein folding (GO:0006457) are significantly overrepresented.

We also identified other clusters within the STRING network that could be associated with ER stress. In plants, phospholipase D (cluster 218), is associated with tolerance to osmotic and temperature stress, plant pathogen defense, phosphate and nitrogen deficiencies, and heat stress memory (see review [85,86]). However, recent work in mammalian systems has demonstrated that the inhibition of phospholipase D results in ER stress [87]. Similarly, ER stress activates glutathione-related enzymes, including glutathione peroxidases (Cluster 556, [88]). NF-Y transcription factors (Cluster 689) form a transcriptional complex with BZIP60 to bind an ER stress response element located in the promoter of genes involved in unfolded protein responses (see review [83]).

The unfolded protein response is broadly conserved across eukaryotes [89] and responds to a variety of abiotic and biotic stresses, including heat, cold, salinity, drought, flooding, high light, heavy metals, and pathogens [90]. Stress signaling is important for striking a balance between survival and continued growth and development. To our knowledge, the unfolded protein response has not been tied to iron stress responses in any plant species. Remarkably, while the response is well conserved, in our study, it is largely limited to iron-efficient genotypes G1 and G8. Of the DEGs associated directly or indirectly with the unfolded protein response above, 83% came from G1 and G8. This suggests that these genotypes are able to exploit the unfolded protein response through novel signaling mechanisms. This is just one example of how the data from this study can be exploited for improving soybean iron stress responses.

4. Conclusions

Utilizing a variety of genotypes, with a range of phenotypic responses to IDC, can help to improve our understanding of the diverse iron stress responses found in the soybean germplasm. We used RNA-seq to analyze patterns and early changes of gene expression across 18 soybean genotypes in response to iron stress. Changes in gene expression profiles across the genotypes at 60 min after stress demonstrate that the rapid response to iron stress is not limited to the Clark genotype. Variation in the differentially expressed genes and biological processes identified in the early response demonstrated that multiple response mechanisms and potential differences in the response time to iron stress exist in soybean

germplasm. While this study highlights differences between genotypes at 60 min after stress, future work would benefit with the inclusion of additional timepoint(s) to analyze changes in signaling pathways and to develop an iron stress response curve for multiple genotypes over time. Ultimately, this study reveals the utility of expanding gene expression studies to include a variety of genotypes.

5. Materials and Methods

5.1. Phenotypic Clustering

Visual iron deficiency chlorosis (IDC) ratings and soil plant analysis development (SPAD; Spectrum Technologies, Aurora, IL, USA) measurements were collected at multiple growth stages from plants grown in the field in 2014, and in the field and hydroponics in 2015 as described by Assefa et al. [12]. The 'stats' package in RStudio [91,92] was used to calculate a distance matrix using the Euclidean method, and then genotypes were clustered using Ward's method. The same package was used for the principal components analysis (PCA).

5.2. Plant Materials

Eighteen diverse plant introduction (PI) lines were selected from a genome wide association study (GWAS) panel used by Assefa et al. [12]. The selected lines comprised eight haplotypes, corresponding to four linkage blocks spread across a historical IDC quantitative trait loci (QTL) on soybean chromosome Gm03 (Supplementary File S1). Based on the field and hydroponic phenotypic data reported by Assefa et al. [12] genotypes from the same haplotype but with contrasting severity ratings in response to IDC were selected. The goal of including contrasting severity ratings was to assess the impact of other genomic locations on IDC tolerance. Two near-isogenic lines (NILs) historically used to study IDC responses, iron-efficient Clark (PI 548533), and iron-inefficient IsoClark (PI 547430), were included to serve as internal controls. Additional distinguishing characteristics that were maintained for our records were the country of origin and maturity group.

5.3. Tissue Collection

Seeds of each genotype were germinated on paper in a growth chamber set at 24 °C with a 14 h day length. After seven days in the growth chamber, seedlings were transferred to eight hydroponic buckets, where each bucket contained one seedling of each genotype. All buckets were setup using an iron-sufficient (100 µM Fe[NO₃]₃•9H₂O) hydroponic system in a single growth chamber, as outlined by O'Rourke et al. [17], with nutrient solutions described by Chaney et al. [93] adjusted for 10 L buckets. Hydroponic systems using these nutrient solutions have been used to identify and validate soybean iron stress tolerance QTL found in field conditions [12,94]. After nine days of growth in the hydroponic solution (16 days after germination), seedlings grown in the same bucket were transferred to a new 10 L bucket with either iron-sufficient or iron-deficient conditions (100 µM Fe[NO₃]₃•9H₂O and 50 µM Fe[NO₃]₃•9H₂O, respectively), resulting in four biological replicates of each genotype in each iron condition. During transfer, the group of seedlings was carefully rinsed in solution of the same iron condition as the destination bucket. Moran Lauter et al. [20] observed a shift in root-to-shoot differential gene expression in Clark over the course of 30–120 min, with an inflection point at 60 min after the onset of iron stress. Therefore, we decided to collect tissue samples 60 min after iron stress; this would allow us to capture stress responses in both roots and leaves from genotypes with faster and slower responses relative to Clark. Sixty minutes after transferring the seedlings to new iron conditions, leaflet tissue from the first trifoliolate and whole root tissue were harvested, frozen in liquid nitrogen, and then maintained at –80 °C. All tissue was collected and stored in individual 50 mL Falcon[®] tubes (Thermo Fisher Scientific, Waltham, MA, USA). Three biological replicates were collected from each genotype and iron condition. The remaining biological replicate for each iron condition was grown for two more weeks

to validate phenotypic responses, especially of Clark and IsoClark under iron-deficient conditions (data not shown).

5.4. RNA Isolation and Sequencing

Frozen tissue was crushed with an inverted pestle in the 50 mL Falcon® tubes used in tissue collection. One full microspatula scoop (approximately 100 mg) of crushed tissue was transferred to a 2 mL Safe-Lock™ microcentrifuge tube (Eppendorf, Hamburg, Germany), and then ground with a 5 mm stainless steel bead for one minute at 30 Hz using the Qiagen TissueLyser II (Qiagen, Germantown, MD, USA). RNA was extracted following the RNeasy® Plant Mini Kit protocol. Extracted RNA was DNase treated in 50 µL reactions using the Ambion® TURBO DNA-free™ Kit (Thermo Fisher Scientific, Waltham, MA, USA) and further purified using an RNeasy® MinElute® Cleanup Kit (Qiagen, Germantown, MD, USA). Final RNA concentration and quality was measured using a NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

RNA samples were sequenced at the Iowa State University DNA Facility. Prior to sequencing, the DNA facility validated the quality of each RNA sample using an Agilent® 2100 Bioanalyzer™ (Agilent®, Santa Clara, CA, USA). After quality confirmation, sequences were generated on the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA) using normal output mode with 150 base pair, single-end sequencing. A total of 216 samples were run on 19 lanes across three eight-lane flow cells (two full and one partial). Each lane was assigned one rep of six genotypes from one tissue type from both iron conditions (sufficient and deficient).

5.5. Identification of Differentially Expressed Genes in Response to Iron Stress

Sequencing adaptors were removed using the program Scythe (version 0.981, [95]), the first 15 bases were removed using the program fastx_trimmer (version 0.0.14, http://hannonlab.cshl.edu/fastx_toolkit, released on 5 January 2014), and bases with quality scores below 20 were removed using the program Sickle (version 1.2, [96]). Cleaned fastq files were sorted and mapped to the soybean reference genome (*Glycine max* Wm82.a2.v1, Phytozome version 12) using TopHat2 (version 2.1.1, [97]). SAMtools (version 1.6, [98]) was used to filter for and reliably sort mapping reads. The resulting binary alignment/map (BAM) files were used for downstream differential expression analyses. BAM files and raw fastq files generated by this study were deposited in the National Center for Biotechnology Short Read Archive (NCBI SRA BioProject accession PRJNA706999).

RStudio [91,92] was used for statistical analyses. Samples with fewer than five million mapped reads were removed from further analysis. If two of the three reps within iron treatment and tissue type contained fewer than five million mapped reads, the genotype within that tissue type was removed from further analysis. The “edgeR” package [99] was used to identify differentially expressed genes (DEGs). Genes with counts per million (cpm) of one or more ($\text{cpm} \geq 1$) in at least three samples were considered expressed and used for further analyses. Library sizes were normalized across all samples within tissue type using the trimmed mean of M-values (TMM) method [100]. We fit a negative binomial generalized log-linear model to the normalized count data with genotype x iron condition groups as the factor in our design matrix. Individual contrast statements were made between iron conditions (deficient versus sufficient) of a given genotype within tissue type. The likelihood ratio test was used with each contrast to test for differential expression of the treatment effect by genotype. Genes with a false discovery rate of less than 0.05 ($\text{FDR} < 0.05$) were considered differentially expressed.

5.6. Gene Annotation

All DEGs were annotated using the *Glycine max* Wm82.a2.v1 ‘Gene Annotation Lookup’ under the SoyBase Toolbox tab (<https://soybase.org/genomeannotation/>, released June 2015) [101]. This annotation tool returns the BLASTP ($E < 10^{-6}$, [102]) results of the top hit for the Uniref100 database [103], the most descriptive hit from the Uniref100

database, and the top *Arabidopsis* hit from the TAIR10 database [104]. Additionally, gene ontology descriptions and IDs for biological processes, molecular function, and cellular components associated with the top *Arabidopsis* hit are included. To identify transcription factors within our DEGs, we took advantage of the SoyDB transcription factor database [105]. The SoyBase ‘Gene Model Correspondence Lookup’ (<https://soybase.org/correspondence/>, released June 2015) was used to update transcription factors to *Glycine max Wm82.a2.v1* gene calls.

5.7. Identification of Overrepresented Gene Ontology (GO) Terms and Transcription Factors

A Fisher’s exact test [106] with a Bonferroni correction (corrected p -value < 0.05 , [107]) was used to test for enriched GO terms associated with a DEG list of interest compared to all genes in the soybean genome. The same approach was used to identify significantly overrepresented transcription factor families.

5.8. Single Linkage Clustering

To identify gene families that might play a role in iron stress adaptations and single genes important across multiple genotypes, we used a single linkage clustering approach, as described by Graham et al. [108] and O’Rourke et al. [17]. Custom perl scripts were used to generate a FASTA file of DEGs for each genotype \times treatment \times expression combination. To each DEG identifier, we added genotype and tissue information (L01-L18 [leaves] or R01-R18 [roots]) and direction of expression (+ induced by iron stress,—repressed by iron stress). For example, Glyma.10G027100 became LG01+_Glyma.10G027100. Note that a particular DEG could be found in different genotype FASTA files, but with different genotype, tissue, and expression information added to the name. Genotype FASTA files were then combined into a single FASTA file, generating a database of iron-stress-responsive genes. TBLASX [102] was used to compare the database against itself using an E-value cutoff of $E < 10^{-20}$. Perl scripts generated by Graham et al. [108] were then used to assign homologous sequences to a cluster.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms222111643/s1>.

Author Contributions: Conceptualization, methodology, and resources, A.K.S., J.A.O. and M.A.G.; investigation, C.E.M., J.A.O. and M.A.G.; formal analysis, data curation, and writing—original draft preparation, D.R.K. and M.A.G.; writing—review and editing, D.R.K., C.E.M., A.K.S., J.A.O. and M.A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financed in part by funds provided by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) CRIS Project 3625-21220-006-00D, USDA CRIS Project IOW04714 to A.K.S., and the North Central Soybean Research Program Project FAR0024859 to J.A.O. and M.A.G.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw fastq and cleaned and sorted BAM files generated during the current study are available in the National Center for Biotechnology Small Reads Archive (NCBI SRA, <http://ncbi.nlm.nih.gov/sra>, accessed September 2021) under BioProject PRJNA706999. Datasets generated or analyzed during this study are included in this published article and its additional files.

Acknowledgments: Appreciation is extended to Jiaoping Zhang for providing haplotype data and to Lori Lincoln for technical assistance. The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U. S. Department of Agriculture.

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

Abbreviations

CPM	count per million
DEGs	differentially expressed genes
EF	iron-efficient
FDR	false discovery rate
GO	gene ontology
GWAS	genome-wide association study
IDC	iron deficiency chlorosis
INF	iron-inefficient
log ₂ FC	log ₂ fold-change
NCBI SRA	National Center for Biotechnology Small Reads Archive
PI	plant introduction line
QTL	quantitative trait loci
ROS	reactive oxygen species
RNA-seq	RNA sequencing
SNP	single nucleotide polymorphism
SPAD	soil plant analysis development
TAIR10	The Arabidopsis Information Resource version 10
TF	transcription factor
TFF	transcription factor family
TMM	trimmed mean of M-values

References

- Hansen, N.C.; Jolley, V.D.; Naeve, S.L.; Goos, R.J. Iron deficiency of soybean in the North Central U.S. and associated soil properties. *Soil Sci. Plant Nutr.* **2004**, *50*, 983–987. [[CrossRef](#)]
- Brown, J.C.; Weber, C.R.; Caldwell, B.E. Efficient and inefficient use of iron by two soybean genotypes and their isolines. *Agron. J.* **1967**, *59*, 459–462. [[CrossRef](#)]
- Froehlich, D.M.; Fehr, W.R. Agronomic performance of soybeans with differing levels of iron deficiency chlorosis on calcareous soil. *Crop Sci.* **1981**, *21*, 438–441. [[CrossRef](#)]
- Inskeep, W.P.; Bloom, P.R. A comparative study of soil solution chemistry associated with chlorotic and nonchlorotic soybeans in western Minnesota. *J. Plant Nutr.* **1984**, *7*, 513–531. [[CrossRef](#)]
- Hansen, N.C.; Schmitt, M.A.; Andersen, J.E.; Strock, J.S. Iron deficiency of soybean in the upper midwest and associated soil properties. *Agron. J.* **2003**, *95*, 1595–1601. [[CrossRef](#)]
- Weiss, M.G. Inheritance and physiology of efficiency in iron utilization in soybeans. *Genetics* **1943**, *28*, 253–268. [[CrossRef](#)]
- Cianzio, S.R.; Fehr, W.R. Genetic control of iron deficiency chlorosis in soybeans. *Iowa State J. Res.* **1980**, *54*, 367–375.
- Cianzio, S.R.; Fehr, W.R. Variation in the inheritance of resistance to iron deficiency chlorosis in soybeans. *Crop Sci.* **1982**, *5*, 433–434. [[CrossRef](#)]
- Lin, S.; Cianzio, S.; Shoemaker, R. Mapping genetic loci for iron deficiency chlorosis in soybean. *Mol. Breed.* **1997**, *3*, 219–229. [[CrossRef](#)]
- Mamidi, S.; Chikara, S.; Goos, R.J.; Hyten, D.L.; Annam, D.; Moghaddam, S.M.; Lee, R.K.; Cregan, P.B.; McClean, P.E. Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. *Plant Genome* **2011**, *4*, 154–164. [[CrossRef](#)]
- Mamidi, S.; Lee, R.K.; Goos, J.R.; McClean, P.E. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PLoS ONE* **2014**, *9*, e107469. [[CrossRef](#)]
- Assefa, T.; Zhang, J.; Moran Lauter, A.N.; Singh, A.; O'Rourke, J.A.; Graham, M.A.; Singh, A.K. Deconstructing the genetic architecture of iron deficiency chlorosis in soybean using genome-wide approaches. *BMC Plant Biol.* **2020**, *20*, 42. [[CrossRef](#)]
- Diers, B.W.; Cianzio, S.R.; Shoemaker, R.C. Possible identification of quantitative trait loci affecting iron efficiency in soybean. *J. Plant Nutr.* **1992**, *15*, 2127–2136. [[CrossRef](#)]
- Severin, A.J.; Peiffer, G.A.; Xu, W.W.; Hyten, D.L.; Bucciarelli, B.; O'Rourke, J.A.; Bolon, Y.-T.; Grant, D.; Farmer, A.D.; May, G.D.; et al. An integrative approach to genomic introgression mapping. *Plant Physiol.* **2010**, *154*, 3–12. [[CrossRef](#)]
- Peiffer, G.A.; King, K.E.; Severin, A.J.; May, G.D.; Cianzio, S.R.; Lin, S.F.; Lauter, N.C.; Shoemaker, R.C. Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. *Plant Physiol.* **2012**, *158*, 1745–1754. [[CrossRef](#)]
- O'Rourke, J.A.; Graham, M.A.; Vodkin, L.; Gonzalez, D.O.; Cianzio, S.R.; Shoemaker, R.C. Recovering from iron deficiency chlorosis in near-isogenic soybeans: A microarray study. *Plant Physiol. Biochem.* **2007**, *45*, 287–292. [[CrossRef](#)]
- O'Rourke, J.A.; Charlson, D.V.; Gonzalez, D.O.; Vodkin, L.O.; Graham, M.A.; Cianzio, S.R.; Grusak, M.A.; Shoemaker, R.C. Microarray analysis of iron deficiency chlorosis in near-isogenic soybean lines. *BMC Genom.* **2007**, *8*, 476. [[CrossRef](#)]

18. O'Rourke, J.A.; Nelson, R.T.; Grant, D.; Schmutz, J.; Grimwood, J.; Cannon, S.; Vance, C.P.; Graham, M.A.; Shoemaker, R.C. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genom.* **2009**, *10*, 376. [[CrossRef](#)] [[PubMed](#)]
19. Moran Lauter, A.N.; Peiffer, G.A.; Yin, T.; Whitham, S.A.; Cook, D.; Shoemaker, R.C.; Graham, M.A. Identification of candidate genes involved in early iron deficiency chlorosis signaling in soybean (*Glycine max*) roots and leaves. *BMC Genom.* **2014**, *15*, 702. [[CrossRef](#)] [[PubMed](#)]
20. Moran Lauter, A.N.; Rutter, L.; Cook, D.; O'Rourke, J.A.; Graham, M.A. Examining short-term responses to a long-term problem: RNA-seq analyses of iron deficiency chlorosis tolerant soybean. *Int. J. Mol. Sci.* **2020**, *21*, 3591. [[CrossRef](#)] [[PubMed](#)]
21. Atencio, L.; Salazar, J.; Moran Lauter, A.N.; Gonzales, M.D.; O'Rourke, J.A.; Graham, M.A. Characterizing short and long term iron stress responses in iron deficiency tolerant and susceptible soybean (*Glycine max* L. Merr.). *Plant Stress* **2021**, *2*, 100012. [[CrossRef](#)]
22. Wingett, S.W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control [version 2; referees: 4 approved]. *F1000Research* **2018**, *7*, 1338. [[CrossRef](#)] [[PubMed](#)]
23. Ahn, J.W.; Verma, R.; Kim, M.; Lee, J.Y.; Kim, Y.K.; Bang, J.W.; Reiter, W.D.; Pai, H.S. Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. *J. Biol. Chem.* **2006**, *281*, 13708–13716. [[CrossRef](#)]
24. López-Millán, A.F.; Morales, F.; Gogorcena, Y.; Abadía, A.; Abadía, J. Metabolic responses in iron deficient tomato plants. *J. Plant Physiol.* **2009**, *166*, 375–384. [[CrossRef](#)] [[PubMed](#)]
25. Han, D.; Wang, L.; Wang, Y.; Yang, G.; Gao, C.; Yu, Z.; Li, T.; Zhang, X.; Ma, L.; Xu, X.; et al. Overexpression of *Malus xiaojinensis* CS1 gene in tobacco affects plant development and increases iron stress tolerance. *Sci. Hortic.* **2013**, *150*, 65–72. [[CrossRef](#)]
26. Schröder, F.; Lisso, J.; Lange, P.; Müssig, C. The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biol.* **2009**, *9*, 20. [[CrossRef](#)]
27. Lisso, J.; Schröder, F.; Müssig, C. EXO modifies sucrose and trehalose responses and connects the extracellular carbon status to growth. *Front. Plant Sci.* **2013**, *4*, 1–18. [[CrossRef](#)]
28. Katsir, L.; Schillmiller, A.L.; Staswick, P.E.; Sheng, Y.H.; Howe, G.A. COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7100–7105. [[CrossRef](#)]
29. Hwang, J.U.; Song, W.Y.; Hong, D.; Ko, D.; Yamaoka, Y.; Jang, S.; Yim, S.; Lee, E.; Khare, D.; Kim, K.; et al. Plant ABC transporters enable many unique aspects of a terrestrial plant's lifestyle. *Mol. Plant* **2016**, *9*, 338–355. [[CrossRef](#)]
30. Krishnamurthy, P.; Vishal, B.; Ho, W.J.; Lok, F.C.J.; Lee, F.S.M.; Kumar, P.P. Regulation of a cytochrome P450 gene CYP94B1 by WRKY33 transcription factor controls apoplastic barrier formation in roots to confer salt tolerance. *Plant Physiol.* **2020**, *184*, 2199–2215. [[CrossRef](#)] [[PubMed](#)]
31. ChunJuan, D.; JinYuan, L. The Arabidopsis EAR-motif-containing protein RAP2.1 functions as an active transcriptional repressor to keep stress responses under tight control. *BMC Plant Biol.* **2010**, *10*, 47.
32. Shi, H.; Liu, W.; Yao, Y.; Wei, Y.; Chan, Z. Alcohol dehydrogenase 1 (ADH1) confers both abiotic and biotic stress resistance in Arabidopsis. *Plant Sci.* **2017**, *262*, 24–31. [[CrossRef](#)]
33. Lucena, C.; Porras, R.; García, M.J.; Alcántara, E.; Pérez-Vicente, R.; Zamarreño, Á.M.; Bacaicoa, E.; García-Mina, J.M.; Smith, A.P.; Romera, F.J. Ethylene and phloem signals are involved in the regulation of responses to Fe and P deficiencies in roots of strategy I plants. *Front. Plant Sci.* **2019**, *10*, 1–14. [[CrossRef](#)] [[PubMed](#)]
34. Xie, Z.; Nolan, T.M.; Jiang, H.; Yin, Y. AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in Arabidopsis. *Front. Plant Sci.* **2019**, *10*, 1–17. [[CrossRef](#)] [[PubMed](#)]
35. Taran, N.; Okanenko, A.; Musienko, N. Sulpholipid reflects plant resistance to stress-factor action. *Biochem. Soc. Trans.* **2000**, *28*, 922–924. [[CrossRef](#)]
36. Klecker, M.; Gasch, P.; Peisker, H.; Dörmann, P.; Schlicke, H.; Grimm, B.; Muströph, A. A shoot-specific hypoxic response of Arabidopsis sheds light on the role of the phosphate-responsive transcription factor PHOSPHATE STARVATION RESPONSE1. *Plant Physiol.* **2014**, *165*, 774–790. [[CrossRef](#)] [[PubMed](#)]
37. Thimm, O.; Essigmann, B.; Kloska, S.; Altmann, T.; Buckhout, T.J. Response of Arabidopsis to iron deficiency stress as revealed by microarray analysis. *Plant Physiol.* **2001**, *127*, 1030–1043. [[CrossRef](#)]
38. 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](#)]
39. Garcia, N.; Da-Silva, C.J.; Cocco, K.L.T.; Pomagualli, D.; de Oliveira, F.K.; da Silva, J.V.L.; de Oliveira, A.C.B.; do Amarante, L. Waterlogging tolerance of five soybean genotypes through different physiological and biochemical mechanisms. *Environ. Exp. Bot.* **2020**, *172*, 103975. [[CrossRef](#)]
40. Charron, J.-B.F.; Ouellet, F.; Houde, M.; Sarhan, F. The plant Apolipoprotein D ortholog protects Arabidopsis against oxidative stress. *BMC Plant Biol.* **2008**, *8*, 86. [[CrossRef](#)]
41. García, M.J.; Suárez, V.; Romera, F.J.; Alcántara, E.; Pérez-Vicente, R. A new model involving ethylene, nitric oxide and Fe to explain the regulation of Fe-acquisition genes in Strategy I plants. *Plant Physiol. Biochem.* **2011**, *49*, 537–544. [[CrossRef](#)]
42. Santamaria, M.; Thomson, C.J.; Read, N.D.; Loake, G.J. The promoter of a basic PR1-like gene, AtPRB1, from Arabidopsis establishes an organ-specific expression pattern and responsiveness to ethylene and methyl jasmonate. *Plant Mol. Biol.* **2001**, *47*, 641–652. [[CrossRef](#)] [[PubMed](#)]

43. Takahashi, Y.; Ebisu, Y.; Kinoshita, T.; Doi, M.; Okuma, E.; Murata, Y.; Shimazaki, K.I. bHLH transcription factors that facilitate K⁺ uptake during stomatal opening are repressed by abscisic acid through phosphorylation. *Sci. Signal.* **2013**, *6*, ra48. [[CrossRef](#)]
44. Li, J.; Brader, G.; Palva, E.T. Kunitz trypsin inhibitor: An antagonist of cell death triggered by phytopathogens and fumonisin B1 in *Arabidopsis*. *Mol. Plant* **2008**, *1*, 482–495. [[CrossRef](#)]
45. Komarova, N.Y.; Thor, K.; Gubler, A.; Meier, S.; Dietrich, D.; Weichert, A.; Grottemeyer, M.S.; Tegeder, M.; Rentsch, D. AtPTR1 and AtPTR5 transport dipeptides in planta. *Plant Physiol.* **2008**, *148*, 856–869. [[CrossRef](#)] [[PubMed](#)]
46. Rubin, G.; Tohge, T.; Matsuda, F.; Saito, K.; Scheible, W.R. Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in *Arabidopsis*. *Plant Cell* **2009**, *21*, 3567–3584. [[CrossRef](#)]
47. Lin, Q.; Wang, S.; Dao, Y.; Wang, J.; Wang, K. *Arabidopsis thaliana* trehalose-6-phosphate phosphatase gene TPPI enhances drought tolerance by regulating stomatal apertures. *J. Exp. Bot.* **2020**, *71*, 4285–4297. [[CrossRef](#)]
48. Hindt, M.N.; Gueriot, M. Lou Getting a sense for signals: Regulation of the plant iron deficiency response. *Biochim. Biophys. Acta-Mol. Cell Res.* **2012**, *1823*, 1521–1530. [[CrossRef](#)] [[PubMed](#)]
49. Shen, C.; Yang, Y.; Liu, K.; Zhang, L.; Guo, H.; Sun, T.; Wang, H. Involvement of endogenous salicylic acid in iron-deficiency responses in *Arabidopsis*. *J. Exp. Bot.* **2016**, *67*, 4179–4193. [[CrossRef](#)] [[PubMed](#)]
50. Atwood, S.E.; O'Rourke, J.A.; Peiffer, G.A.; Yin, T.; Majumder, M.; Zhang, C.; Cianzio, S.R.; Hill, J.H.; Cook, D.; Whitham, S.A.; et al. Replication protein A subunit 3 and the iron efficiency response in soybean. *Plant Cell Environ.* **2014**, *37*, 213–234. [[CrossRef](#)]
51. Lin, X.Y.; Ye, Y.Q.; Fan, S.K.; Jin, C.W.; Zheng, S.J. Increased sucrose accumulation regulates iron-deficiency responses by promoting auxin signaling in *Arabidopsis* plants. *Plant Physiol.* **2016**, *170*, 907–920. [[CrossRef](#)]
52. Butenhoff, K.J. QTL Mapping and GWAS Identify Sources of Iron Deficiency Chlorosis and Canopy Wilt Tolerance in the Fiskeby III X Mandarin (Ottawa) Soybean Population. Ph.D. Thesis, University of Minnesota Digital Conservancy, Minneapolis, MN, USA, 2015.
53. Shanmugam, S.; Zhao, S.; Nandy, S.; Srivastava, V.; Khodakovskaya, M. Modification of soybean growth and abiotic stress tolerance by expression of truncated ERECTA protein from *Arabidopsis thaliana*. *PLoS ONE* **2020**, *15*, e0233383. [[CrossRef](#)]
54. Chuang, H.W.; Feng, J.H.; Feng, Y.L.; Wei, M.J. An *Arabidopsis* WDR protein coordinates cellular networks involved in light, stress response and hormone signals. *Plant Sci.* **2015**, *241*, 23–31. [[CrossRef](#)] [[PubMed](#)]
55. Yuan, S.W.; Wu, X.L.; Liu, Z.H.; Luo, H.B.; Huang, R.Z. Abiotic stresses and phytohormones regulate expression of FAD2 gene in *Arabidopsis thaliana*. *J. Integr. Agric.* **2012**, *11*, 62–72. [[CrossRef](#)]
56. Chen, X.; Han, H.; Jiang, P.; Nie, L.; Bao, H.; Fan, P.; Lv, S.; Feng, J.; Li, Y. Transformation of β -lycopene cyclase genes from *Salicornia europaea* and *Arabidopsis* conferred salt tolerance in *Arabidopsis* and tobacco. *Plant Cell Physiol.* **2011**, *52*, 909–921. [[CrossRef](#)]
57. O'Rourke, J.A.; McCabe, C.E.; Graham, M.A. Dynamic gene expression changes in response to micronutrient, macronutrient, and multiple stress exposures in soybean. *Funct. Integr. Genom.* **2019**, *20*, 321–341. [[CrossRef](#)] [[PubMed](#)]
58. Kissen, R.; Winge, P.; Tran, D.H.T.; Jørstad, T.S.; Størseth, T.R.; Christensen, T.; Bones, A.M. Transcriptional profiling of an Fd-GOGAT1/GLU1 mutant in *Arabidopsis thaliana* reveals a multiple stress response and extensive reprogramming of the transcriptome. *BMC Genom.* **2010**, *11*, 190. [[CrossRef](#)] [[PubMed](#)]
59. Cui, M.; Gu, M.; Lu, Y.; Zhang, Y.; Chen, C.; Ling, H.Q.; Wu, H. Glutamate synthase 1 is involved in iron-deficiency response and long-distance transportation in *Arabidopsis*. *J. Integr. Plant Biol.* **2020**, *62*, 1925–1941. [[CrossRef](#)] [[PubMed](#)]
60. Farrow, S.C.; Facchini, P.J. Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front. Plant Sci.* **2014**, *5*, 1–15. [[CrossRef](#)] [[PubMed](#)]
61. Viganì, G.; Morandini, P.; Murgia, I. Searching iron sensors in plants by exploring the link among 2'-OG-dependent dioxygenases, the iron deficiency response and metabolic adjustments occurring under iron deficiency. *Front. Plant Sci.* **2013**, *4*, 1–7. [[CrossRef](#)] [[PubMed](#)]
62. Sohani, M.M.; Schenk, P.M.; Schultz, C.J.; Schmidt, O. Phylogenetic and transcriptional analysis of a strictosidine synthase-like gene family in *Arabidopsis thaliana* reveals involvement in plant defence responses. *Plant Biol.* **2009**, *11*, 105–117. [[CrossRef](#)]
63. Zhang, J.; Naik, H.S.; Assefa, T.; Sarkar, S.; Reddy, R.V.C.; Singh, A.; Ganapathysubramanian, B.; Singh, A.K. Computer vision and machine learning for robust phenotyping in genome-wide studies. *Sci. Rep.* **2017**, *7*, 44048. [[CrossRef](#)] [[PubMed](#)]
64. Kollist, H.; Zandalinas, S.I.; Sengupta, S.; Nuhkat, M.; Kangasjärvi, J.; Mittler, R. Rapid responses to abiotic stress: Priming the landscape for the signal transduction network. *Trends Plant Sci.* **2019**, *24*, 25–37. [[CrossRef](#)] [[PubMed](#)]
65. Buckhout, T.J.; Yang, T.J.W.; Schmidt, W. Early iron-deficiency-induced transcriptional changes in *Arabidopsis* roots as revealed by microarray analyses. *BMC Genomics* **2009**, *10*, 147. [[CrossRef](#)]
66. Khan, M.A.; Castro-Guerrero, N.A.; McInturf, S.A.; Nguyen, N.T.; Dame, A.N.; Wang, J.; Bindbeutel, R.K.; Joshi, T.; Jurisson, S.S.; Nusinow, D.A.; et al. Changes in iron availability in *Arabidopsis* are rapidly sensed in the leaf vasculature and impaired sensing leads to opposite transcriptional programs in leaves and roots. *Plant Cell Environ.* **2018**, *41*, 2263–2276. [[CrossRef](#)]
67. Lin, F.; Zhao, M.; Baumann, D.D.; Ping, J.; Sun, L.; Liu, Y.; Zhang, B.; Tang, Z.; Hughes, E.; Doerge, R.W.; et al. Molecular response to the pathogen *Phytophthora sojae* among ten soybean near isogenic lines revealed by comparative transcriptomics. *BMC Genom.* **2014**, *15*, 18. [[CrossRef](#)] [[PubMed](#)]
68. Koenig, D.; Jiménez-Gómez, J.M.; Kimura, S.; Fulop, D.; Chitwood, D.H.; Headland, L.R.; Kumar, R.; Covington, M.F.; Devisetty, U.K.; Tat, A.V.; et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2–9. [[CrossRef](#)] [[PubMed](#)]

69. Guan, S.; Xu, Q.; Ma, D.; Zhang, W.; Xu, Z.; Zhao, M.; Guo, Z. Transcriptomics profiling in response to cold stress in cultivated rice and weedy rice. *Gene* **2019**, *685*, 96–105. [CrossRef]
70. Gao, H.; Wang, Y.; Li, W.; Gu, Y.; Lai, Y.; Bi, Y.; He, C. Transcriptomic comparison reveals genetic variation potentially underlying seed developmental evolution of soybeans. *J. Exp. Bot.* **2018**, *69*, 5089–5104. [CrossRef]
71. Stein, R.J.; Waters, B.M. Use of natural variation reveals core genes in the transcriptome of iron-deficient *Arabidopsis thaliana* roots. *J. Exp. Bot.* **2012**, *63*, 1039–1055. [CrossRef]
72. Waters, B.M.; McInturf, S.A.; Stein, R.J. Rosette iron deficiency transcript and microRNA profiling reveals links between copper and iron homeostasis in *Arabidopsis thaliana*. *J. Exp. Bot.* **2012**, *63*, 5903–5918. [CrossRef]
73. Cooper, M.; Messina, C.D. Can we harness “enviromics” to accelerate crop improvement by integrating breeding and agronomy? *Front. Plant Sci.* **2021**, *12*. [CrossRef]
74. Merry, R.; Butenhoff, K.; Campbell, B.W.; Michno, J.-M.; Wang, D.; Orf, J.H.; Lorenz, A.J.; Stupar, R.M. Identification and fine-mapping of a soybean quantitative trait locus on chromosome 5 conferring tolerance to iron deficiency chlorosis. *Plant Genome* **2019**, *12*, 190007. [CrossRef] [PubMed]
75. 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.* **2021**, *49*, D605–D612. [CrossRef] [PubMed]
76. Lu, M.; Hellmann, H.A.; Liu, Y.; Wang, W. Editorial: Protein quality controlling systems in plant responses to environmental stresses. *Front. Plant Sci.* **2018**, *9*, 1–3. [CrossRef]
77. Singh, H.; Kaur, K.; Singh, M.; Kaur, G.; Singh, P. Plant cyclophilins: Multifaceted proteins with versatile roles. *Front. Plant Sci.* **2020**, *11*, 1–30. [CrossRef]
78. Yuen, C.Y.L.; Matsumoto, K.O.; Christopher, D.A. Variation in the subcellular localization and protein folding activity among *Arabidopsis thaliana* homologs of protein disulfide isomerase. *Biomolecules* **2013**, *3*, 848–869. [CrossRef] [PubMed]
79. Rajan, V.B.; D’Silva, P. *Arabidopsis thaliana* J-class heat shock proteins: Cellular stress sensors. *Funct. Integr. Genom.* **2009**, *9*, 433–446. [CrossRef] [PubMed]
80. Tominaga, J.; Muroya, M.; Shirakami, N.; Takahashi, S.; Higuchi, C.; Muranaka, A.; Takami, T.; Watanabe, S. Overexpression of BUNDLE SHEATH DEFECTIVE 2 improves the efficiency of photosynthesis and growth in *Arabidopsis*. *Plant J.* **2020**, *102*, 129–137. [CrossRef]
81. Wang, L.; Wen, R.; Wang, J.; Xiang, D.; Wang, Q.; Zang, Y.; Wang, Z.; Huang, S.; Li, X.; Datla, R.; et al. *Arabidopsis* UBC13 differentially regulates two programmed cell death pathways in responses to pathogen and low-temperature stress. *New Phytol.* **2019**, *221*, 919–934. [CrossRef] [PubMed]
82. Skelly, M.J.; Malik, S.I.; Le Bihan, T.; Bo, Y.; Jiang, J.; Spoel, S.H.; Loake, G.J. A role for S-nitrosylation of the SUMO-conjugating enzyme SCE1 in plant immunity. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17090–17095. [CrossRef] [PubMed]
83. Nawkar, G.M.; Lee, E.S.; Shelake, R.M.; Park, J.H.; Ryu, S.W.; Kang, C.H.; Lee, S.Y. Activation of the transducers of unfolded protein response in plants. *Front. Plant Sci.* **2018**, *9*, 1–10. [CrossRef] [PubMed]
84. Howell, S.H. Evolution of the unfolded protein response in plants. *Plant Cell Environ.* **2021**, *44*, 2625–2635. [CrossRef]
85. Deepika, D.; Singh, A. Plant phospholipase D: Novel structure, regulatory mechanism, and multifaceted functions with biotechnological application. *Crit. Rev. Biotechnol.* **2021**, 1–19. [CrossRef]
86. Urrea Castellanos, R.; Friedrich, T.; Petrovic, N.; Altmann, S.; Brzezinka, K.; Goroka, M.; Graf, A.; Bäurle, I. FORGETTER2 protein phosphatase and phospholipase D modulate heat stress memory in *Arabidopsis*. *Plant J.* **2020**, *104*, 7–17. [CrossRef]
87. Nakagawa, H.; Hazama, K.; Ishida, K.; Komori, M.; Nishimura, K.; Matsuo, S. Inhibition of PLD1 activity causes ER stress via regulation of COPII vesicle formation. *Biochem. Biophys. Res. Commun.* **2017**, *490*, 895–900. [CrossRef] [PubMed]
88. Uzilday, B.; Ozgur, R.; Sekmen, A.H.; Turkan, I. Endoplasmic reticulum stress regulates glutathione metabolism and activities of glutathione related enzymes in *Arabidopsis*. *Funct. Plant Biol.* **2018**, *45*, 284–296. [CrossRef]
89. Hollien, J. Evolution of the unfolded protein response. *Biochim. Biophys. Acta-Mol. Cell Res.* **2013**, *1833*, 2458–2463. [CrossRef]
90. Depaepe, T.; Hendrix, S.; Janse van Rensburg, H.C.; Van den Ende, W.; Cuypers, A.; Van Der Straeten, D. At the crossroads of survival and death: The reactive oxygen species–ethylene–sugar triad and the unfolded protein response. *Trends Plant Sci.* **2021**, *26*, 338–351. [CrossRef]
91. RS Team. *RStudio: Integrated Development Environment for R*; RStudio, Inc.: Boston, Mass, USA, 2021.
92. RC Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2021.
93. Chaney, R.L.; Coulombe, B.A.; Bell, P.F.; Angle, J.S.; Coulombe, B.A.; Bell, P.F.; Angle, J.S. Detailed method to screen dicot cultivars for resistance to Fe-chlorosis using FeDTPA and bicarbonate in nutrient solutions. *J. Plant Nutr.* **1992**, *15*, 2063–2083. [CrossRef]
94. Lin, S.F.; Baumer, J.; Ivers, D.; Cianzio, S.; Shoemaker, R. Nutrient solution screening of Fe chlorosis resistance in soybean evaluated by molecular characterization. *J. Plant Nutr.* **2000**, *23*, 1915–1928. [CrossRef]
95. Buffalo, V. Scythe—A Bayesian Adapter Trimmer. Available online: <https://github.com/ucdavis-bioinformatics/scythe> (accessed on 24 October 2021).
96. Joshi, N.; Fass, J. Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33). Available online: <https://github.com/ucdavis-bioinformatics/sickle> (accessed on 24 October 2021).
97. Dobin, A.; Gingeras, T.R. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **2013**, *14*, 1–13. [CrossRef]

98. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, 1–4. [[CrossRef](#)]
99. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26*, 139–140. [[CrossRef](#)]
100. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11*, R25. [[CrossRef](#)]
101. Grant, D.; Nelson, R.T.; Cannon, S.B.; Shoemaker, R.C. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **2009**, *38*, 843–846. [[CrossRef](#)]
102. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PS I-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
103. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115–D119. [[CrossRef](#)]
104. Huala, E.; Dickerman, A.W.; Garcia-Hernandez, M.; Weems, D.; Reiser, L.; LaFond, F.; Hanley, D.; Kiphart, D.; Zhuang, M.; Huang, W.; et al. The Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **2001**, *29*, 102–105. [[CrossRef](#)]
105. Wang, Z.; Libault, M.; Joshi, T.; Valliyodan, B.; Nguyen, H.T.; Xu, D.; Stacey, G.; Cheng, J. SoyDB: A knowledge database of soybean transcription factors. *BMC Plant Biol.* **2010**, *10*, 14. [[CrossRef](#)]
106. Fisher, R.A. *The Design of Experiment*, 8th ed.; Oliver and Boyd: Edinburgh, UK; London, UK, 1966.
107. Bonferroni, C.E. Il calcolo delle assicurazioni su gruppi di teste. *Studi Onore Del Profr. Salvatore Ortu Carboni* **1935**, 13–60.
108. Graham, M.A.; Silverstein, K.A.T.; Cannon, S.B.; VandenBosch, K.A. Computational identification and characterization of novel genes from legumes. *Plant Physiol.* **2004**, *135*, 1179–1197. [[CrossRef](#)]