



Inference of Transcription Regulatory Network in Low Phytic Acid Soybean Seeds

Neelam Redekar^{1†}, Guillaume Pilot², Victor Raboy³, Song Li^{1*} and M. A. Saghai Maroof^{1*}

¹ Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, ² Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech, Blacksburg, VA, United States, ³ National Small Grains Germplasm Research Center, Agricultural Research Service (USDA), Aberdeen, ID, United States

OPEN ACCESS

Edited by:

Diego Mauricio Riaño-Pachón,
Institute of Chemistry, University of
São Paulo, Brazil

Reviewed by:

Jedrzej Jakub Szymanski,
Weizmann Institute of Science, Israel
Sudip Kundu,
University of Calcutta, India
Maria Katherine Mejia Guerra,
Cornell University, United States

*Correspondence:

Song Li
songli@vt.edu
M. A. Saghai Maroof
smaroof@vt.edu

† Present Address:

Neelam Redekar,
Department of Crop and Soil Science,
Oregon State University, Corvallis, OR,
United States

Specialty section:

This article was submitted to
Plant Systems and Synthetic Biology,
a section of the journal
Frontiers in Plant Science

Received: 24 July 2017

Accepted: 14 November 2017

Published: 30 November 2017

Citation:

Redekar N, Pilot G, Raboy V, Li S and
Saghai Maroof MA (2017) Inference of
Transcription Regulatory Network in
Low Phytic Acid Soybean Seeds.
Front. Plant Sci. 8:2029.
doi: 10.3389/fpls.2017.02029

A dominant loss of function mutation in *myo*-inositol phosphate synthase (*MIPS*) gene and recessive loss of function mutations in two multidrug resistant protein type-ABC transporter genes not only reduce the seed phytic acid levels in soybean, but also affect the pathways associated with seed development, ultimately resulting in low emergence. To understand the regulatory mechanisms and identify key genes that intervene in the seed development process in low phytic acid crops, we performed computational inference of gene regulatory networks in low and normal phytic acid soybeans using a time course transcriptomic data and multiple network inference algorithms. We identified a set of putative candidate transcription factors and their regulatory interactions with genes that have functions in *myo*-inositol biosynthesis, auxin-ABA signaling, and seed dormancy. We evaluated the performance of our unsupervised network inference method by comparing the predicted regulatory network with published regulatory interactions in *Arabidopsis*. Some contrasting regulatory interactions were observed in low phytic acid mutants compared to non-mutant lines. These findings provide important hypotheses on expression regulation of *myo*-inositol metabolism and phytohormone signaling in developing low phytic acid soybeans. The computational pipeline used for unsupervised network learning in this study is provided as open source software and is freely available at <https://lilabatvt.github.io/LPANetwork/>.

Keywords: phytic acid, soybean seed development, *myo*-inositol metabolism, unsupervised machine learning, gene regulatory network

INTRODUCTION

Seed development is a complex metabolic process, which involves both synthesis and breakdown of macromolecules for growth and maintenance of the embryo (Weber et al., 2005; Le et al., 2007). During seed development, glucose-6-phosphate is converted to *myo*-inositol, an intracellular signaling molecule, which is phosphorylated several times to form phytic acid (Raboy, 1997). Seeds with reduced phytic acid content are commercially more valuable because consumption of low phytic acid seeds by monogastric animals alleviates mineral deficiency and reduces phosphorus pollution from animal waste (Raboy, 2007). Mutations that block the phytic acid biosynthesis

Abbreviations: *lpa*, Low phytic acid; MIPS, *myo*-inositol phosphate synthase; MRP, Multi-drug resistance protein; DEG, Differentially expressed gene; RFO, Raffinose family oligosaccharide; ABA, Abscisic acid.

pathway have been shown to alter the seed metabolite levels in soybean, rice, maize, and other plant species (Wilcox et al., 2000; Shi et al., 2003, 2005; Stevenson-Paulik et al., 2005; Raboy, 2007; Glover, 2011; Jervis et al., 2015). For example, a mutation in *myo*-inositol phosphate synthase (*MIPS*) gene results in reduced phytic acid, stachyose, raffinose, and elevated sucrose, and low seed emergence in soybean (Hitz et al., 2002; Saghai Maroof and Buss, 2008). Other non-biosynthetic pathway genes such as multi-drug resistance protein (*MRP*) genes encoding ATP-binding cassette transporters that are believed to be involved in the transport of phytic acid to storage vacuoles, are also known to regulate phytic acid levels and affect seed emergence (Shi et al., 2007; Nagy et al., 2009; Saghai Maroof et al., 2009; Xu et al., 2009; Jervis et al., 2015).

Transcriptome analysis is a valuable tool for the characterization of the regulatory networks that mediate this complex interaction. The expressions of genes involved in the metabolic activities in seeds are tightly regulated by the synergistic action of many transcription factors and other regulatory genes (Weber et al., 2005; Le et al., 2007). Two independent studies, one with barley low phytic acid (*lpa*) mutant, and another with soybean *mips1/mrp-l/mrp-n* (*3mlpa*) triple mutant have reported the effect of *lpa* mutations on the transcriptomic profiles of developing seeds (Bowen et al., 2007; Redekar et al., 2015). The differential expression of transcription factor genes such as WRKY and CAMTA (Calmodulin-binding Transcription Activator), was linked to phytic acid biosynthesis pathway, suggesting a complex regulatory mechanism (Redekar et al., 2015). Association of WRKY transcription factors and Ca²⁺ binding activity with inositol metabolism was also confirmed by another independent study with maize low phytic acid breeding line Qi319 (Zhang et al., 2016). Zhang et al. (2016) also identified ABC transporter gene candidates associated with low phytic acid phenotype in maize using co-regulatory network. In this article, we focus on the discovery of transcription regulatory networks to further investigate the inositol metabolism in soybean.

One type of the widely applied method of network inference is the use of Pearson Correlation Coefficient or related methods for data analysis (Langfelder and Horvath, 2008; Bassel et al., 2011; Li et al., 2012). Although such correlation analyses can cluster genes with similar functions, the resultant networks do not predict the direction of gene regulation. Many other approaches, such as mutual information (Faith et al., 2007), partial correlation (Faith et al., 2007), random forest (Huynh-Thu et al., 2010), and least angle regression (LARS) (Haury et al., 2012), have been developed to perform inference of directed gene regulatory networks. For well-characterized model organisms such as *Arabidopsis*, known interactions from ChIP-chips or ChIP-seq experiments can be used as prior knowledge in supervised machine learning approaches (Maetschke et al., 2014; Ni et al., 2016). However, for biological systems where little prior information is available, such as in soybeans, unsupervised methods have to be used for network inference. In particular, three methods including co-expression analysis (Bassel et al., 2011), decision trees (Zhu et al., 2013) and mutual information (Gonzalez-Morales et al., 2016) have been successfully applied to

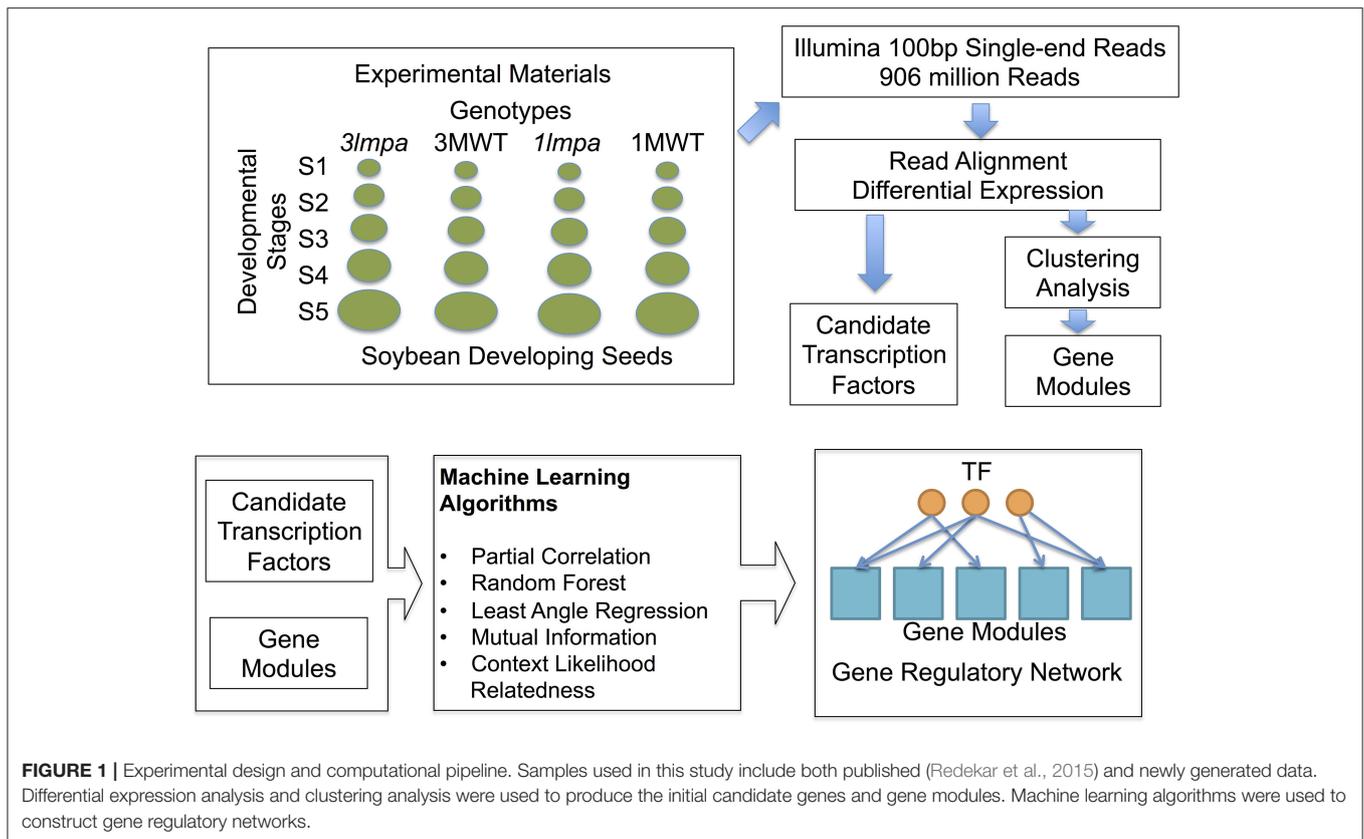
identify functional networks in *Arabidopsis* and soybean seeds. With numerous inference methods available, it has been found that congregating the prediction results from multiple methods (so-called “community-based method”) improves the prediction accuracy as compared to any individual method (Marbach et al., 2012).

In this study, we performed computational inference of gene regulatory networks in low phytic acid mutants and the corresponding non-mutant soybean seeds from time course transcriptomic data. In addition to previously published RNA-seq data (Redekar et al., 2015), we generated new RNA-seq data of developing seeds using a pair of soybean isogenic lines, one carrying the *mips1* mutation (*1mlpa*) and the other the corresponding wild type allele. We implemented a computational pipeline for unsupervised gene regulatory network inference using five different methods: ARACNE (Margolin et al., 2006), Random Forest (Huynh-Thu et al., 2010), LARS (Haury et al., 2012), partial correlation (Schafer and Strimmer, 2005b), and context likelihood relatedness (CLR) (Faith et al., 2007). To improve computational efficiency and interpretability of the inferred network, we adopted the widely used module network approach by which genes were grouped into co-expression modules and inference of gene regulation was performed between transcription factors and gene modules (Segal et al., 2003). We found that many gene modules included genes with meaningful biological functions and some gene modules showed genotype-specific expression patterns. We identified several transcription factors that were differentially expressed between developmental stages and some of the inferred regulatory interactions were specifically found in mutants or non-mutants. Genes involved in phytic acid metabolism and related metabolic processes were found in multiple modules and were predicted to be regulated by different transcription factors. For validation, the predicted interactions were compared with known regulatory interactions observed in the model plant species, *Arabidopsis thaliana*. These findings provide important hypotheses on expression regulation of *myo*-inositol metabolism, and phytohormone signaling in developing *lpa* soybeans. The computational method for inferring regulatory networks is freely available at <https://lilabvt.github.io/LPANetwork/>. This method can be used to perform network inference using time series data from soybean or any other crop species.

MATERIALS AND METHODS

Genetic Materials

Four soybean experimental lines designated as: (i) *3mlpa*, (ii) 3MWT, (iii) *1mlpa*, and (iv) 1MWT were used in this study (Figure 1). The *lpa* mutant line, *3mlpa*, carrying three mutations *mips1/mrp-l/mrp-n*, and its non-mutant sibling line with normal phytic acid, 3MWT, were derived from crossing of ‘CX-1834’ (*lpa* line with two *mpr-l/mrp-n* mutations on soybean chromosomes 19 and 3, respectively) with ‘V99-5089’ (*lpa* line with single *mips1* mutation) (Saghai Maroof et al., 2009). The low phytic acid causing mutations in the parental lines have been mapped to genes Glyma.11G238800 (*MIPS1*),



Glyma.19g169000 (MRP-L), and Glyma.03g167800 (MRP-N) (Saghai Maroof et al., 2009). Another *lpa* line, ‘*1mlpa*’, carrying a single *mips1* mutation on soybean chromosome 11, and its isogenic sibling line with normal phytic acid, ‘1MWT’, were derived from crossing of ‘Essex’ (a normal phytic acid line with no mutations) with V99-5089 (Saghai Maroof and Buss, 2008; Glover, 2011).

Plant Growth and Tissue Sampling

Four seeds from each of experimental lines—*3mlpa*, 3MWT, *1mlpa*, and 1MWT, were planted in each of 12 pots containing Metro-Mix[®] 360 (Sun Gro) media topped with GardenPro ULTRALITE soil (Redekar et al., 2015). These plants (48/line) were grown in growth chambers with 14 /10 h photoperiod, 24°C/16°C temperature, 300–400 μ E light and 50–60% relative humidity. Developing seeds were sampled in triplicates for each experimental line based on seed lengths corresponding to 2–4 mm (S1), 4–6 mm (S2), 6–8 mm (S3), 8–10 mm (S4), and 10–12 mm (S5), respectively. Samples were flash frozen using liquid nitrogen and stored at -70°C . High-quality total RNA (RIN 9–10) was extracted from frozen samples using RNeasy Plant Mini Kit, with on-column DNase digestion (QIAGEN). Total of 60 mRNA libraries were prepared from total RNA samples and sequenced as 100SE using HiSeq2000 at the Genome Quebec Innovation Center, Canada.

Sequence Data Processing and Differential Gene Expression

Reads were aligned to the latest soybean reference genome (‘Williams 82’ Wm82.a2.v1, downloaded from Phytozome¹ with STAR (version 2.4.2) and number of reads mapped to each gene was counted using featureCount (version 1.4.6). Differential gene expression was analyzed using DESeq2 (version 1.8.2) in R (version 3.2.4). Four genotypes, *3mlpa*, 3MWT, *1mlpa*, and 1MWT, were analyzed in this data. For each pair of mutant and corresponding non-mutant, stage-wise comparison was performed to identify differentially expressed genes for each stage (Supplementary Figure 1A) (Redekar et al., 2015). For each genotype, between-stage comparisons were performed to identify differentially expressed genes between adjacent developmental stages (Supplementary Figure 1B). These analyses were performed using DESeq2 with default parameters. Between-stage comparisons and stage-wise comparisons address different type of biological question. Between-stage comparisons find genes that change between stages, but do not directly identify genes that are affected by mutations. Stage-wise comparison, on the other hand, directly finds genes that change between mutant and non-mutant lines, but does not find genes with interactions. Differentially expressed genes were the genes with FDR adjusted $p < 0.01$ and \log_2 fold change > 1 . Differentially expressed genes and their \log_2 fold changes are provided as Supplementary Tables

¹<https://phytozome.jgi.doe.gov/>

1–3. RNA-Seq data used in this study have been deposited into the NCBI Gene Expression Omnibus (GEO) repository under accession number GSE101692.

Inference of Gene Regulatory Networks Expression Clustering, Gene Ontology, and Gene Function Analysis

Gene expression levels for each gene were normalized using DESeq2 and summarized as FPKM (Fragments Per Kilo-base pair per Million reads) values. The gene expression levels (FPKM values) were averaged across replicates and only differentially expressed genes were used in the clustering analysis. K-means clustering (Sherlock, 2000) was performed using R packages, and the number of clusters (K) was determined using the minimum Bayesian Information Criteria (BIC) method (Ramsey et al., 2008). In brief, K was set to be an integer number from 20 to 100 with an incremental step size of 5. For each K-value, k-means clustering was performed and BIC statistics were computed. The minimum BIC was achieved with $K = 60$ (Supplementary Figure 2). Gene Ontology (GO) annotation of all soybean genes was downloaded from Soybase² GO enrichment analyses were performed for each gene module. Significantly enriched GO categories were selected using Fisher's exact test with FDR <0.05 (Supplementary Table 4). Transcription factor annotation was downloaded from plant TFDB (Jin et al., 2015, 2017). Metabolic pathway genes were downloaded from the SoyCyc 7.0³ database from the Plant Metabolic Network⁴ website.

Network Inference Methods

To infer regulatory networks, we adopted the methods of module networks (Segal et al., 2003). First, genes were grouped into modules using the k-means clustering method. Second, differentially expressed transcription factors were used as putative regulators for network inference. In our data, we found 60 clusters (gene modules) and 1245 transcription factors that were differentially expressed in at least one comparison. The mean expression profile for each of the 60 modules was computed and the expression levels of 1245 transcription factors were included to construct an expression matrix with 1305 rows (genes) and 20 columns (five developmental stages for four experimental lines). Five distinct network inference algorithms: ARACNE (Margolin et al., 2006), Random Forest (Huynh-Thu et al., 2010), LARS (Haury et al., 2012), partial correlation (Schafer and Strimmer, 2005b), and CLR (Faith et al., 2007), were applied to this expression matrix to infer putative regulatory interactions between each transcription factor and gene modules. These methods were chosen because they represent a diverse set of computational methods for gene network inference. These methods were selected also because they were ranked as top performers in a recently published benchmark of network inference methods (DREAM challenge) (Marbach et al., 2012). Details of each method, statistical

analysis of network and network validation are provided as supplementary text.

RESULTS

Summary of Differential Gene Expression Analysis

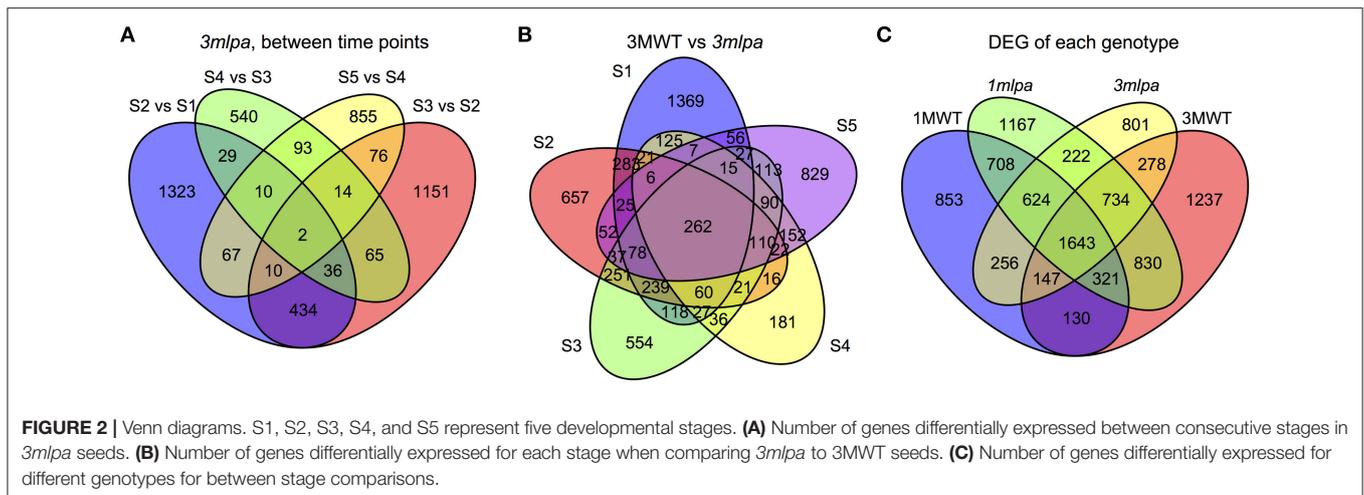
Transcriptome sequencing data from five developing seed stages of four soybean lines (*3mlpa*, 3MWT, *1mlpa*, and 1MWT) were analyzed (Figure 1). Stage-wise comparisons were performed for mutants (*3mlpa*, *1mlpa*) and their corresponding non-mutants (3MWT, 1MWT), to determine the number of genes affected by the mutation at each stage (Figure 2). For stage-wise comparisons, when *1mlpa* was compared with 1MWT, we found fewer than 250 differentially expressed genes in all time points (Supplementary Figure 2). However, we found more than 4000 differentially expressed genes between *3mlpa* and 3MWT (Supplementary Figure 2). It is expected to have higher number of differentially expressed genes for comparison between recombinant inbred lines (*3mlpa* vs. 3MWT) as opposed to that between near-isogenic lines (*1mlpa* vs. 1MWT). Few genes are differentially expressed in all five stages when comparing *3mlpa* vs. 3MWT (Figure 2B). These results suggest that genes affected by mutations are unique at each developmental stage. Between-stages comparisons were performed for each genotype separately (Figure 1, Supplementary Figure 1). Results of the stage-wise and between-stages differential expression analyses are provided as Supplementary Tables 1, 2, respectively. For between-stages comparisons (Supplementary Figure 1B), we found that hundreds of genes are differentially expressed when comparing adjacent developmental stages (Supplementary Figure 2). However, there are few genes differentially expressed across all stages. For example, in *3mlpa* mutant, only 2 genes are differentially expressed between any two adjacent stages (Figure 2A). We found 1643 genes that are differentially expressed between developmental stages and found in all four genotypes in this study (Figure 2C), suggesting these genes are a core set of genes that change expression between developmental stages and are not affected by the genotypes. We determined the number of differentially expressed genes in each of the comparisons (Supplementary Figure 2). Interestingly, the highest numbers of differentially expressed genes were found in two comparisons: between stage 1 and stage 2 for 1MWT and for 3MWT (3MWT S2 vs. S1 and 1MWT S2 vs. S1) suggesting that non-mutant plants have a high number of differentially expressed genes in the early stages of seed development than mutants.

The two sets of low-phytic acid causing mutations (*mips1* and *mrp-l/mrp-n*) interrupt the phytic acid biosynthesis and transmembrane transporter activity, ultimately reducing the seed emergence potential in our mutant lines. We were, therefore, interested in studying the behavior of genes associated with phytic acid metabolism, abscisic acid (ABA) and auxin signaling and metabolism, and transmembrane transport. The log₂ fold change for significantly differentially expressed genes that belonged to this category is summarized in Supplementary Table 3.

²<https://soybase.org/>

³<http://www.plantcyc.org/databases/soycyc/7.0>

⁴<http://www.plantcyc.org/>



Co-expression Modules Represent Distinct Functional Categories

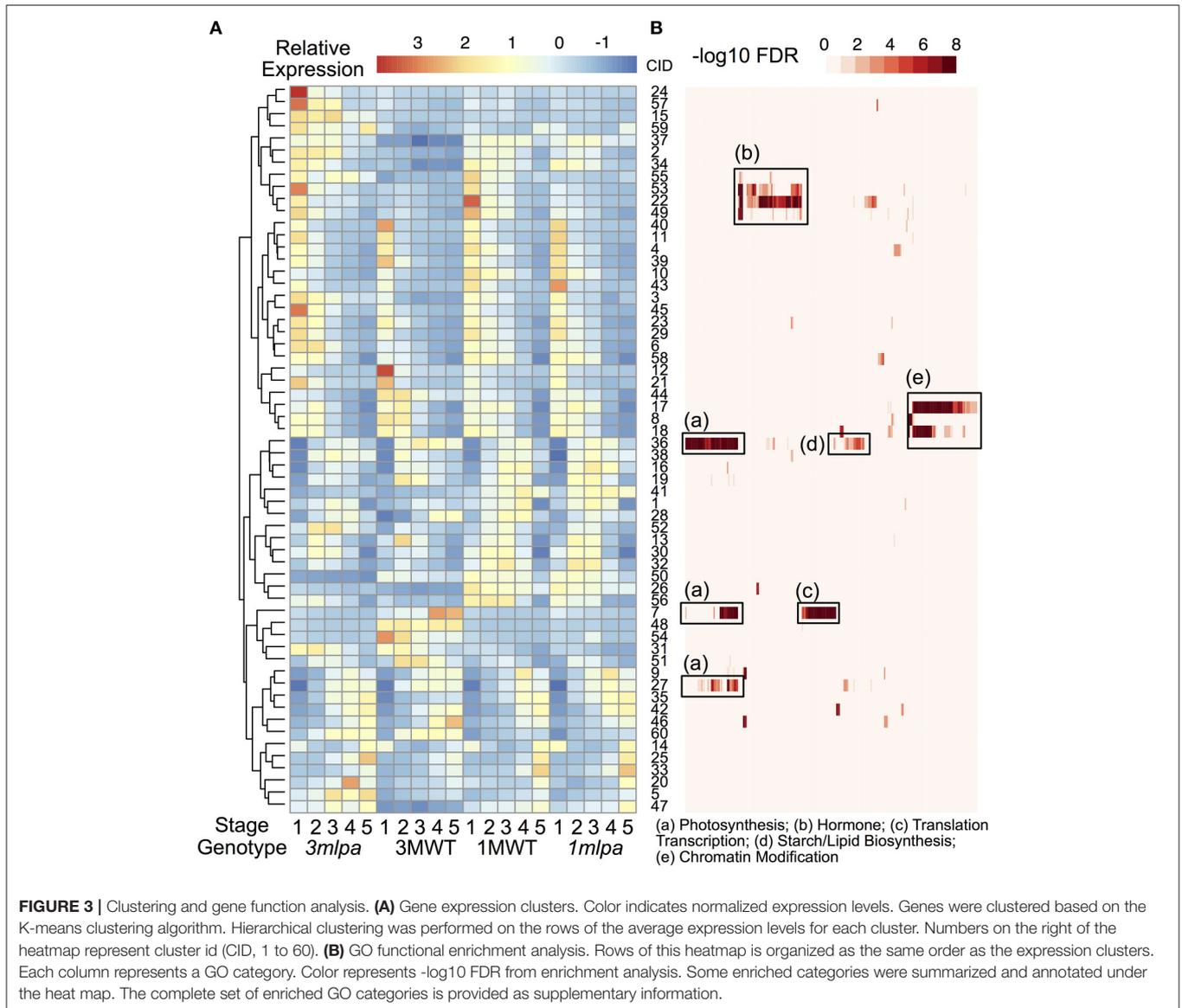
We identified 12998 genes that are differentially expressed in at least one of the comparisons and these genes were used for K-means clustering analysis to identify co-expressed gene modules (Figures 1, 3). We found that the optimal number of clusters (modules) is 60 based on BIC (Supplementary Figure 3). Average expression levels of all genes in each module were used to generate a heat map (Figure 3A) and GO enrichment analyses were performed for each of the modules (Figure 3B, Supplementary Table 4). The expression modules can be approximately classified into three main patterns. In pattern 1, there are 29 modules highly expressed at the early stage of seed development as shown in upper half of the heat map (Figure 3A, cluster 24 to cluster 18). In pattern 2, there are 11 modules highly expressed in the later stage of seed development as shown by the lower portion of the heat map (Figure 3A, modules 9 to 47, except for module 14, which shows high expression at both first and last developmental stages). In pattern 3, 12 modules showed high expression in the middle of the developmental stages but low expression levels in the early and late developmental stages. Genotype-specific expression patterns were also found by clustering analysis. For example, modules 24, 57, and 15 are highly expressed in *3mlpa* at the first time point, whereas the expression levels are not high in the other three genotypes. The modules identified in near isogenic lines (*1mlpa* and 1MWT) showed highly similar expression patterns than those identified in recombinant inbred lines (*3mlpa* and 3MWT).

GO enrichment analyses showed that many gene modules are enriched with genes in specific functional categories (Supplementary Table 4). For example, genes in modules 22, 49, 53, and 55 are highly expressed during the early stages of seed development and are enriched with genes with functions in hormone signaling and responses (Figure 3B, box b). Of these, several genes showed increased expression in *3mlpa* (S1) while decreased in *1mlpa* (S1) when compared with respective non-mutant lines at stage 1 (Table 1). We also

found that genes in module 42 are highly expressed at the last stage of seed development and this module is enriched with genes functioning in seed dormancy, seed germination and lipid storage (Figure 3B). Genes in module 36 are highly expressed in the middle stages of seed development and are enriched with genes in starch and lipid biosynthesis (Figure 3B, box d). The functional enrichment of genes in these modules indicates that our clustering analysis can find genes representing biological functions that are known to be active at different stages of seed development. We also found that some modules have genotype-specific expression patterns. For example, module 7 is highly expressed only in 3MWT, which is enriched with genes functioning in photosynthesis, translations, and transcription (Figure 3B, boxes a,c). This result shows some gene modules have genotype-specific expression pattern and are enriched with specific functional genes.

Regulatory Network Interactions

Five different network inference algorithms were used to infer putative regulatory interactions between regulators and their targets (Figures 1, 4). Fifty-four interactions between 54 transcription factors and 32 modules were predicted by all five algorithms (Supplementary Table 5). Some modules were predicted to be regulated by more than one transcription factor and no transcription factor was predicted to regulate two modules. The identified interactions represent highly stringent predictions and are a very conservative estimation of all possible interactions since only 0.06% of all 74,700 possible interactions were found to be significant by all five computational methods. Four hundred six interactions between 348 transcription factors and 60 modules were supported by four or more methods (Supplementary Figure 4, Supplementary Table 6), representing a larger number for predicted regulatory interactions. The network figure includes both directed and undirected edges (Figure 4). Each directed edge connects a transcription factor with its targeted gene module. Such edge represents predicted regulatory interactions. Each undirected



edge connects a transcription factor to the module this factor belongs to. These undirected edges reflect the fact that each transcription factor is also co-expressed with other genes in the genome and can be assigned to a specific gene module. The regulatory interactions (directed edges) are further classified based on the differential expression pattern of the regulatory TFs. We found 10 TF-module interactions (black arrows) in which the TFs are differentially expressed between stage comparisons (Supplementary Figure 1B) in both mutants (*3mlpa* and *1mlpa*) and both non-mutants (3MWT and 1MWT). These interactions are not affected by the mutations or genetic backgrounds. We found 10 interactions (Figure 4, green arrows) in which the TFs are differentially expressed between stage comparisons for non-mutants but not in mutants. These interactions are potentially lost due to the mutations. We also found nine interactions (blue arrows) that are not present in the non-mutants, but

are present in either one of the mutants, suggesting that these interactions are gained in the mutants. Finally, we found 14 interactions (red arrows) that are not present in either the mutants or the non-mutants, but the TFs are differentially expressed when comparing *3mlpa* to 3MWT at one or more developmental stages. These interactions are altered in *3mlpa*/3MWT and do not change the trajectory of gene expression between stages but affect gene expression within specific stages. We found that many TFs are also connected to the target modules by undirected edges, indicating that these genes are co-expressed with their target genes. We also found several cases where a TF does not regulate its own module but is regulating other modules, suggesting our method can find non-linear interactions between TF and target modules. All the predicted regulatory interactions are provided as a (Supplementary Tables 5, 6).

TABLE 1 | Differentially expressed genes from co-expression modules in early stages of seed development.

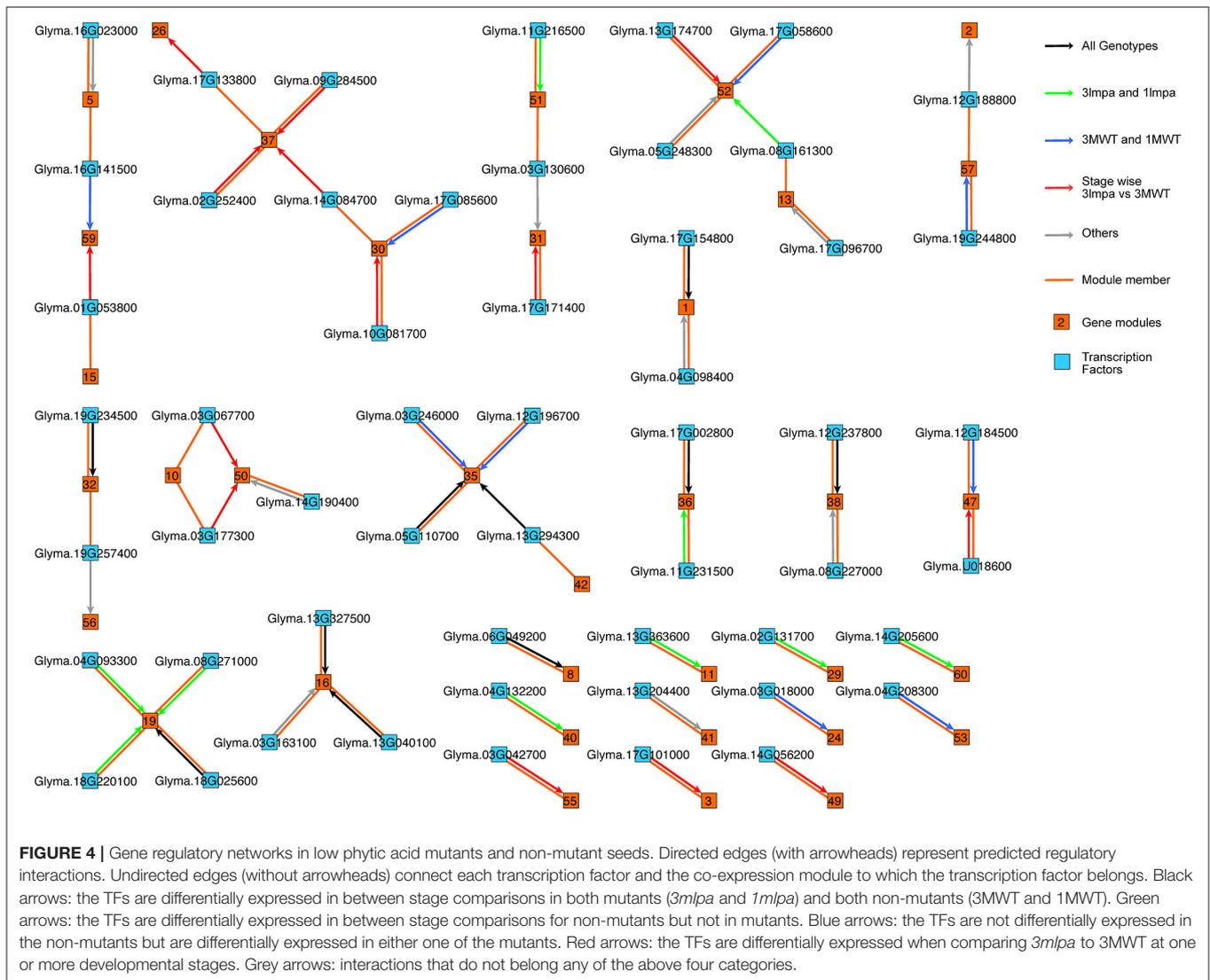
Cluster ID	GO Term	Gene Name	Log2 fold change		Arabidopsis homolog	Gene Symbol	Description
			3mlpa vs. 3MWT-S1	1mlpa vs. 1MWT-S1			
22	GO:0048838: release of seed from dormancy	Glyma.01G153300	3.13805	-2.84401	AT4G19230	<i>CYP707A1</i>	Cytochrome P450, family 707, subfamily A, polypeptide 1
		Glyma.12G073000	2.67451	-2.55671	AT3G45640	<i>MPK3</i>	Mitogen-activated protein kinase 3
	GO:0009738: abscisic acid-activated signaling pathway	Glyma.U021800	2.5946	-2.50453	AT3G45640	<i>MPK3</i>	Mitogen-activated protein kinase 3
		Glyma.07G023300	2.96832	-2.24850	AT1G80840	<i>WRKY40</i>	WRKY DNA-binding protein 40
	Glyma.10G082400	1.40473	-1.55483	AT3G11820	<i>SYP121, SYR1, PEN1</i>	Syntaxin of plants 121	
	Glyma.08G271900	2.04156	-1.54559	AT1G32640	<i>RD22BP1, JAI1, JIN1, MYC2, ZBF1</i>	Basic helix-loop-helix (bHLH) DNA-binding family protein	
	Glyma.03G138000	0	-1.46586	AT3G57530	<i>CPK32, CDPK32</i>	Calcium-dependent protein kinase 32	
	Glyma.05G194100	0	-1.41822	AT3G52430	<i>PAD4</i>	Alpha/beta-Hydrolases superfamily protein	
Glyma.17G126400	1.18340	-1.31950	AT3G48360	<i>BT2</i>	BTB and TAZ domain protein 2		
42	GO:0009845: seed germination	Glyma.17G086400	1.56707	0	AT3G01570	-	Oleosin family protein
49	GO:0009738: abscisic acid-activated signaling pathway	Glyma.15G078600	0	-1.84111	AT1G28480	<i>GRX480, roxy19</i>	Thioredoxin superfamily protein
		Glyma.14G066400	0	-1.10638	AT3G11410	<i>PP2CA, AHG3</i>	Protein phosphatase 2CA
53	GO:0009737: response to abscisic acid	Glyma.13G119500	1.60968	-2.02755	AT3G17510	<i>CIPK1, SnRK3.16</i>	CBL-interacting protein kinase 1
		Glyma.12G186800	2.75115	-1.95635	AT2G27310		F-box family protein
	GO:0009738: abscisic acid-activated signaling pathway	Glyma.12G186800	2.75115	-1.95635	AT2G27310		F-box family protein
		GO:0009733: response to auxin	Glyma.13G094900	4.14408	-2.26796	AT5G57560	<i>TCH4, XTH22</i>

Validation of Predicted Regulatory Networks

To validate the results from the computational inference, we compared the regulatory network as predicted by our methods with published regulatory interactions observed in the model plant species, *Arabidopsis*. We combined three recently published *Arabidopsis* genome-wide gene regulatory networks (Sparks et al., 2016), including 2,914 regulatory interactions between 578 regulators and 717 targets. We mapped soybean genes to *Arabidopsis* genome and searched for predicted regulator-module interactions, which are also found in *Arabidopsis*. Among 54 interactions we predicted, five TF-module interactions and 13 TF-gene interactions from this dataset are also found in the *Arabidopsis* regulatory networks (Supplementary Table 7), providing support for the predicted gene regulatory networks. This small overlapping is expected because the current *Arabidopsis* interaction network (2914 interactions) is

only a tiny fraction of true interactions that happen *in-vivo* (see discussion).

Transcription factors regulate their target through binding of sequence specific motifs in the promoter regions of the target genes. To further validate the predicted regulatory networks, we performed promoter motif search. Among many programs that are available for motif discovery, the MEME suite contains the most comprehensive sets of programs that allows users to perform motif discovery, motif search and motif comparison. We used the MEME program to identify motifs in the promoter regions of genes in each gene module. We tested the enrichment of these newly discovered motifs and identified 101 motifs that are enriched in 36 modules, with each module having one or more enriched motifs. In our predicted interaction networks, there are 54 transcription factors that regulate 32 modules. We found that 21 out of these 32 modules contain one or more enriched motifs (Supplementary Table 8). These motifs are



putative binding sites of the 54 transcription factors that are regulating genes in these modules.

To further validate the results of the motif search, we compared our newly discovered motifs to a database of motifs generated by direct sequencing of binding sites of over 400 *Arabidopsis* TFs. The pattern of these motifs is represented by position specific weight matrices (PSWM) (Bailey et al., 2009). This comparison aims at testing whether any of the enriched motifs are similar to the binding motifs of the genes in the same gene family in *Arabidopsis*. For example (Table 2), we found that a bZIP transcription factor (Glyma.19G244800, whose most similar *Arabidopsis* gene is AT5G28770) is regulating module 57 in our predicted network. Our analysis found a GCCACGT motif enriched in the promoter regions of module 57 ($p < 1.21e-3$). This motif is highly similar to the binding motif (GCCACGT) of an *Arabidopsis* bZIP transcription factor ($p < 1.82e-10$). Four such examples are shown in Table 2. Among the 21 modules with enriched motifs in our predicted regulatory networks, we found

that 17 motifs are highly similar to the corresponding *Arabidopsis* motifs in the same TF gene family (Supplementary Table 8), providing strong support to the validity of our predictions.

Regulatory Network Changes and Genes in Phytic Acid Metabolic Pathways

To understand the connections between transcription regulation and metabolic pathways that were altered in the *3mlpa* and *1mlpa* mutant lines, we downloaded the metabolic pathway annotation from the SoyCyc 7.0 database. We mapped genes in myo-inositol metabolism, stachyose metabolism and sucrose metabolism to different gene modules, because these metabolites are altered as a result of *lpa* mutations (Supplementary Table 3). We found that there were 64 genes involved in these metabolic pathways, which mapped to 34 gene modules (Supplementary Table 3). Twelve of the 64 genes were involved in stachyose (or raffinose family oligosaccharides (RFOs)) biosynthesis and eight were associated with inferred gene regulatory networks

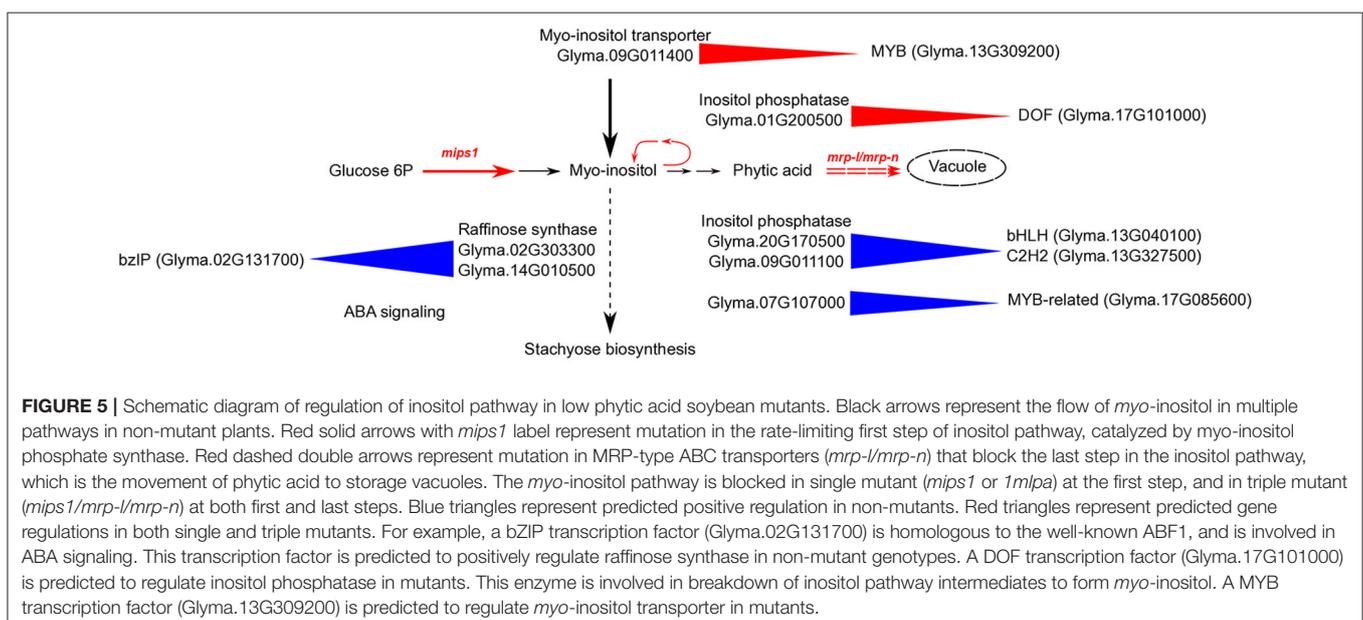
(Figure 5). In module 29, which is up-regulated in the first stage of seed development, two genes (Glyma.02G303300 and Glyma.14G010500, both encoding raffinose synthase / seed imbibition protein 1) were found to be regulated by a bZIP transcription factor (Glyma.02G131700). This bZIP transcription factor is homologous to *ABF1* (ABA response element-binding factor 1) in *Arabidopsis*, providing a potential connection between ABA response and stachyose biosynthesis (Figure 5). Some of the enzymes involved in *myo*-inositol biosynthesis are regulated similarly in mutant and non-mutant lines. For example, inositol-polyphosphate 5-phosphatase (Glyma.20G170500) and inositol-phosphate phosphatase (Glyma.09G011100) are found in module 16. Genes in this module are up-regulated during mid-stages of seed development. Module 16 is predicted to be regulated by two transcription

factors, a bHLH transcription factor (Glyma.13G040100) and a C2H2 transcription factor (Glyma.13G327500). The bHLH transcription factor is homologous to SPCH, which regulates stomatal lineage specification during embryo development (Danzer et al., 2015). Some regulatory interactions are only found in non-mutants. For example, genes from module 40 are predicted to be regulated by Glyma.17G085600, a MYB-related transcription factor, and this module contains two genes related to *myo*-inositol biosynthesis (Glyma.07G107000 encoding inositol-polyphosphate 5-phosphatase and Glyma.05G180600 encoding inositol-1-phosphate synthase). This module is highly expressed in the early stage of seed development and the MYB-related transcription factor is similar to *RSM1* in *Arabidopsis*, which has been found to be related to auxin signaling in early morphogenesis (Hamaguchi et al., 2008).

TABLE 2 | Motifs enrichment analysis.

TF name	TF family	Target module name	ATH TF name	Motif enrich adj p-value	MEME motif	DAP-seq motif	Motif similarity p-value
Glyma.19G244800	bZIP	57	AT5G28770	1.21E-03			1.82E-10
Glyma.03G042700	WRKY	55	AT2G38470	1.14E-03			4.14E-05
Glyma.05G110700	bHLH	35	AT4G37850	6.20E-03			8.09E-05
Glyma.17G101000	Dof	3	AT3G47500	3.79E-04			8.87E-05

"ATH TF name" is *Arabidopsis* transcription factor gene homologous to the soybean transcription factor gene. "Motif enrich adj p-value" is the enrichment p-value for each of the motif found in the promoter regions of predicted gene modules. "DAP-seq motif" is the most similar motif from DAP-seq database. "Motif similarity p-value" represents the significance of similarity between the motifs found in promoters in the gene modules and the DAP-seq motifs. All significantly enriched motifs and their corresponding DAP-seq motifs are provided in Supplementary Table 8.



Several target genes associated with regulation of phytic acid biosynthesis pathway matched with those discovered in co-expression network of developing maize kernel with low phytic acid content (Zhang et al., 2016). These predictions provide genotype-specific, testable hypotheses that may connect gene expression patterns with putative regulatory TFs and hormone regulations during different seed developmental stages.

DISCUSSION

In this study, we performed computational inference of gene regulatory networks using data from developing soybean seeds from two mutants (*3mlpa* and *1mlpa*), with *lpa*-causing mutations, and the respective non-mutant siblings (3MWT and 1MWT). We identified co-expression gene modules with distinct and genotype specific expression patterns. These gene modules are also enriched with genes with various functional categories that are related to different stages of seed development. We identified transcription factors and their putative targets using supervised machine learning methods. Some of these transcription factors are differentially expressed only in non-mutants or only in mutants, suggesting that their regulations are lost or gained due to mutations. Many genes that encode enzymes in the metabolic processes of phytic acid, myo-inositol, sucrose, and stachyose and related oligosaccharides are found in these gene modules. Overall, our analysis provides a framework to connect transcription factors with genes in biological processes such as phytic acid metabolism, auxin-abscisic acid signaling and seed dormancy.

Importance of Regulatory Network Inference

The predicted interactions provide a testable hypothesis for experimental validation using transgenic plants or ChIP-seq experiments. The results from this analysis can also be used to guide interpretation of other genomic mapping experiments such as genome wide association studies or quantitative trait loci analyses and to provide guidance for refining candidate gene lists. Soybean has a complex genome, which encodes over 4,500 putative transcription factors and over 46,000 coding genes. Although transcriptome data have become widely accessible for research in soybean and other crop species, it is still challenging to identify mechanistic connections between the observed transcriptome data with underlying regulatory networks. Differential gene expression analyses can be used to identify candidate genes that change under certain conditions or in specific mutants. However, in most situations, one still faces the problem of interpreting a long list of differentially expressed genes.

Our approach provides one alternative solution to this problem using well-developed machine learning methods to infer regulatory interactions. Our method implements the “community approach,” which has been shown to provide better performance than any individual method alone. The five computational methods are based on fundamentally different

statistical and mathematical formulations thus complement each other and provide a list of highly confident prediction results. Our approach successfully reduces the total number of candidate genes from over 10,000 genes that change in at least one comparison to 54 transcription factors, providing a much shorter list of key genes that can be focused on for validation experiments. However, combining five methods also limited the total number of predicted interactions (54 predicted interactions), because each method predicts some interactions that are not predicted by other methods.

We validated the predicted interactions using *Arabidopsis* gene regulatory networks. *Arabidopsis* is the model plant species, which provides most gene regulatory network information among all plant species. Although soybean and *Arabidopsis* diverged over 120 million years ago, key genes in metabolic pathways and signaling networks are conserved in both species (Jung et al., 2000; Hegeman et al., 2001; Le et al., 2010; Xu et al., 2011; Leite et al., 2014; Gerrard Wheeler et al., 2016), which is expected for a physiological process as conserved as seed development. Therefore, one can expect some regulatory interactions to be conserved between these two species. In fact, 13 interactions predicted by our method are also found in the *Arabidopsis* gene regulatory networks (Supplementary Table 7). The *Arabidopsis* genome contains approximately 2200 TFs and more than 27000 genes (Jin et al., 2015, 2017; Cheng et al., 2016). There are more than 59 million possible interactions between these TF and genes. Although the number of biologically active interactions is probably <1% of all possible interactions, the total number of true interactions is still far more than the 2914 interactions that were used in this comparison. Therefore, it is not surprising that we found a small overlap between the predicted interactions and those identified in *Arabidopsis*. As more interactions will be identified in both plant species, we would expect such overlap to increase. In our analysis, 17 motifs from 21 enriched modules (Supplementary Table 8) similar to the motifs identified in *Arabidopsis* orthologous genes, indicating that the interactions identified in this study are likely to be conserved between the two species. The actual number of conserved interactions is likely to be underestimated, because the regulatory network from *Arabidopsis* is far from complete. Nevertheless, our results provide an important first step toward characterizing gene regulatory networks in soybeans and other crop species.

One would expect more transcription factors being active during the seed development process. To observe a larger number of predicted interactions, we also provide results that are predicted by four out of five methods (Supplementary Table 6). This can be further extended to include predictions from fewer methods. Although aggregating multiple methods has been shown to outperform individual methods, some predictions by a specific method can represent interactions that cannot be detected by other approaches. If a specific target gene or specific function is of interest, one can also use our method to generate a ranked list of all predictions for the target of interest and use predicted regulators as candidate genes.

Regulation of *myo*-Inositol Metabolism in *3mlpa* Mutant Line

Myo-inositol is an essential signaling molecule with multifunctional properties including gene regulation, chromatin modeling, mRNA transport, signal transduction, cell death, pathogen resistance, vesicle trafficking, plasma membrane, and cell wall formation (Martin, 1998; Stevenson et al., 2000; Chen and Xiong, 2010). It is synthesized in a two-step pathway (Figure 5), where glucose-6-phosphate is first converted to inositol monophosphate (IMP), a rate-limiting step catalyzed by MIPS1 enzyme, followed by dephosphorylation of IMP to form *myo*-inositol (Loewus and Murthy, 2000). Upon synthesis, *myo*-inositol is utilized in and recycled from multiple metabolic pathways such as biosynthesis of phytic acid and RFOs (such as stachyose and raffinose), inositol and phosphatidylinositol intermediates, auxin-inositol conjugates and glucuronic acid.

The *mips1* mutation in *Arabidopsis* is associated with reduction in cellular *myo*-inositol levels and defects in early embryogenesis (Meng et al., 2009; Chaouch and Noctor, 2010; Chen and Xiong, 2010; Donahue et al., 2010). The *mips1* mutation in soybean (as in parent V99-5089, *1mlpa* of this study) is associated with decreased levels of phytic acid and RFOs such as raffinose and stachyose, increased levels of sucrose and low emergence. The soybean *mips1* mutants also displayed normal RFO phenotype upon application of exogenous *myo*-inositol (Hitz et al., 2002). It is likely that, similar to *Arabidopsis*, the *mips1* mutation in soybean reduces *myo*-inositol levels, preventing RFO biosynthesis in parent V99-5089 (or in *1mlpa*), as *myo*-inositol is one of the starting substrates in this pathway. Since *myo*-inositol is not consumed by synthesis of RFO (it is a necessary intermediate, but recycled along the pathway), it is possible that the concentration of *myo*-inositol is reduced to such a low level such that synthesis of RFOs is greatly reduced. Sucrose is consumed in stachyose synthesis; it is therefore possible that inhibition of this pathway by the absence of *myo*-inositol is causing the increase in sucrose levels in V99-5089, due to accumulation of unused sucrose substrate.

The *mrp-1/mrp-n* mutations resulted in reduced seed phytic acid and low emergence, but did not alter the RFOs composition. In addition, the *myo*-inositol content in *mrp-1/mrp-n* mutant increases during the seed development phase prior to maturation (Israel et al., 2011). This suggests that in the presence of *mrp-1/mrp-n* mutations, the *mips1*-associated decrease in RFOs composition is restored, despite the reduced *myo*-inositol production due to the *mips1* mutation. This supports the prevalent hypothesis that the lack of transporters in the *mrp-1/mrp-n* mutant may trigger hydrolysis of cytoplasmic phytic acid to form inositol intermediates and *myo*-inositol, hence elevating the *myo*-inositol levels and preventing phytic acid synthesis, diverting the metabolic pathways to *myo*-inositol production (Figure 5). Another possibility is that lack of intracellular *myo*-inositol or phytic acid triggers feedback regulation that up-regulates transporters which can import inositol from the outside of the seed, hence elevating *myo*-inositol levels. This is in agreement with increase in the expression of inositol transporter genes during early stages of seed development in *3mlpa* mutant (Redekar et al., 2015).

In the present study, the inositol transporter gene (Glyma.09G011400) is also up-regulated in *1mlpa*, a single *mips1* mutant. This common up-regulation of inositol transporter in *mips1* and in *mips1/mrp-1/mrp-n* mutants suggests that these mutations are triggering the same signaling pathway (Figure 5). This inositol transporter gene belongs to module 5 and is predicted to be regulated by a MYB transcription factor (Glyma.13G309200), which represents a promising target for experimental validation (Figure 5). These target genes are in agreement with those identified in developing maize kernel with low phytic acid content (Zhang et al., 2016).

We identified over 30 genes associated with *myo*-inositol metabolism in this network analysis (Supplementary Table 3). Some of these genes belong to the modules that are only regulated in mutants. For example, an inositol-polyphosphate 5-phosphatase (Glyma.01G200500) is predicted to be regulated by a DOF transcription factor (Glyma.17G101000). This enzyme catalyzes an intermediate step in converting phytic acid to *myo*-inositol. The DOF transcription factor is differentially expressed when comparing *3mlpa* with 3MWT at stages 1, 3, and 5 and is not differentially expressed in any other comparisons. This observation is consistent with our hypothesis that in *3mlpa* mutant, phytic acid is recycled to produce *myo*-inositol, which participates in RFO synthetic pathways.

Regulation of Auxin-ABA Signaling and Seed Dormancy-Related Genes in *lpa* Mutants

A main goal of this work was to elucidate the molecular basis of negative downstream impacts of *lpa* mutations. Here we found one possible component of this: that in soybean these mutations greatly impact phytohormone pathways. In addition to *myo*-inositol, the importance of phytohormones such as auxin and ABA, in seed development is well-documented (Locascio et al., 2014). Auxin is a key hormone through all phases of seed development including embryogenesis, organ differentiation, endosperm formation, and seed maturation, whereas ABA is involved in onset and maintaining seed dormancy, and is active in the seed maturation and desiccation phases (Locascio et al., 2014). Cross signaling between auxin and ABA are involved in regulating seed dormancy and hence germination (Liu et al., 2013). Hormone signal transduction is controlled by regulating their biosynthesis, accumulation and distribution in different sections and stages of developing seeds by factors such as *myo*-inositol. The *Arabidopsis mips1* mutants have demonstrated defects in cotyledon development and reduced expression of basipetal auxin efflux carriers such as *PIN1* and *PIN2* (Chen and Xiong, 2010; Luo et al., 2011). The factors regulating cellular *myo*-inositol levels might therefore also regulate cross signaling of auxin and ABA. However, it is still unclear to what extent auxin-ABA cross signaling is disturbed in *lpa*-mutants.

In this study, nearly the entire spectrum of the auxin signaling pathway was identified and was differentially regulated in *1mlpa* and *3mlpa* mutants as compared to the corresponding wild types. We identified 188 auxin-related genes differentially expressed in one or more comparisons tested in this study

(Supplementary Table 3). These included genes involved in auxin metabolism, transport, signal transduction, and transcription regulation. We also observed 36 genes from the ABA signaling pathway differentially expressed in one or more comparisons. The *1mlpa* vs. 1MWT comparison identified down regulation of two ABA catabolism genes in *1mlpa* (Glyma.01G153300 and Glyma.09G218600, which remove active ABA), whereas, in *3mlpa* vs. 3MWT comparison, these two genes are up-regulated.

In summary, we identified potential candidate genes that may play a role in regulating inositol metabolism, auxin-ABA signaling and seed maturation-dormancy in low phytic acid soybean during seed development. Although follow up experiments are required to validate these findings, the comprehensive regulatory network and the computational analysis pipeline of this study has set the necessary groundwork for future hypothesis driven investigations.

AUTHOR CONTRIBUTIONS

NR and MASM designed the experiment. NR performed the sequencing experiments. NR and SL analyzed the data. SL

developed the machine-learning pipeline. NR, SL, GP, VR and MASM wrote the paper.

FUNDING

This project was funded by Bio-design and Bioprocessing Research Center, John Lee Pratt Fellowship Program, Agricultural Experiment Station Hatch Program, and Open Access Subvention Fund—all at Virginia Tech.

ACKNOWLEDGMENTS

We would like to thank Virginia Tech's Advanced Research Computing servers and Translational Plant Sciences' MAGYK server for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02029/full#supplementary-material>

REFERENCES

- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Bassel, G. W., Lan, H., Glaab, E., Gibbs, D. J., Gerjets, T., Krasnogor, N., et al. (2011). Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9709–9714. doi: 10.1073/pnas.1100958108
- Bowen, D. E., Souza, E. J., Guttieri, M. J., Raboy, V., and Fu, J. (2007). A low phytic acid barley mutation alters seed gene expression. *Crop Sci.* 47, S-149. doi: 10.2135/cropsci2006.07.0456tpg
- Chaouch, S., and Noctor, G. (2010). Myo-inositol abolishes salicylic acid-dependent cell death and pathogen defence responses triggered by peroxisomal hydrogen peroxide. *New Phytol.* 188, 711–718. doi: 10.1111/j.1469-8137.2010.03453.x
- Chen, H., and Xiong, L. (2010). myo-Inositol-1-phosphate synthase is required for polar auxin transport and organ development. *J. Biol. Chem.* 285, 24238–24247. doi: 10.1074/jbc.M110.123661
- Cheng, C.-Y., Krishnakumar, V., Chan, A., Schobel, S., and Town, C. D. (2016). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *bioRxiv*.
- Danzer, J., Mellott, E., Bui, A. Q., Le, B. H., Martin, P., Hashimoto, M., et al. (2015). Down-regulating the expression of 53 soybean transcription factor genes uncovers a role for speechless in initiating stomatal cell lineages during embryo development. *Plant Physiol.* 168, 1025–1035. doi: 10.1104/pp.15.00432
- Donahue, J. L., Alford, S. R., Torabinejad, J., Kerwin, R. E., Nourbakhsh, A., Ray, W. K., et al. (2010). The *Arabidopsis thaliana* Myo-inositol 1-phosphate synthase1 gene is required for Myo-inositol synthesis and suppression of cell death. *Plant Cell* 22, 888–903. doi: 10.1105/tpc.109.071779
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., et al. (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* 5:e8. doi: 10.1371/journal.pbio.0050008
- Gerrard Wheeler, M. C., Arias, C. L., Righini, S., Badia, M. B., Andreo, C. S., Drincovich, M. F., et al. (2016). Differential contribution of malic enzymes during soybean and castor seeds maturation. *PLoS ONE* 11:e0158040. doi: 10.1371/journal.pone.0158040
- Glover, N. (2011). *The Genetic Basis of Phytate, Oligosaccharide Content, and Emergence in Soybean*. Doctor of Philosophy, Virginia Tech.
- Gonzalez-Morales, S. I., Chavez-Montes, R. A., Hayano-Kanashiro, C., Alejo-Jacunde, G., Rico-Cambron, T. Y., de Folter, S., et al. (2016). Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5232–E5241. doi: 10.1073/pnas.1610985113
- Hamaguchi, A., Yamashino, T., Koizumi, N., Kiba, T., Kojima, M., Sakakibara, H., et al. (2008). A small subfamily of Arabidopsis RADIALIS-LIKE SANT/MYB genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. *Biosci. Biotechnol. Biochem.* 72, 2687–2696. doi: 10.1271/bbb.80348
- Haury, A. C., Mordelet, F., Vera-Licona, P., and Vert, J. P. (2012). TIGRESS: trustful inference of gene regulation using stability selection. *BMC Syst. Biol.* 6:145. doi: 10.1186/1752-0509-6-145
- Hegeman, C. E., Good, L. L., and Grabau, E. A. (2001). Expression of D-myo-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. *Plant Physiol.* 125, 1941–1948. doi: 10.1104/pp.125.4.1941
- Hitz, W. D., Carlson, T. J., Kerr, P. S., and Sebastian, S. A. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiol.* 128, 650–660. doi: 10.1104/pp.010585
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* 5:12776. doi: 10.1371/journal.pone.0012776
- Israel, D. W., Taliencio, E., Kwanyuen, P., Burton, J. W., and Dean, L. (2011). Inositol metabolism in developing seed of low and normal phytic acid soybean lines. *Crop Sci.* 51, 282–289. doi: 10.2135/cropsci2010.03.0123
- Jervis, J., Kastl, C., Hildreth, S. B., Biyashev, R., Grabau, E. A., Saghai-Marooof, M. A., et al. (2015). Metabolite profiling of soybean seed extracts from near-isogenic low and normal phytate lines using orthogonal separation strategies. *J. Agric. Food Chem.* 63, 9879–9887. doi: 10.1021/acs.jafc.5b04002
- Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., et al. (2015). An Arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. *Mol. Biol. Evol.* 32, 1767–1773. doi: 10.1093/molbev/msv058
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982

- Jung, W., Yu, O., Lau, S. M., O'Keefe, D. P., Odell, J., Fader, G., et al. (2000). Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nat. Biotechnol.* 18, 208–212. doi: 10.1038/72671
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8063–8070. doi: 10.1073/pnas.1003530107
- Le, B. H., Wagmaister, J. A., Kawashima, T., Bui, A. Q., Harada, J. J., and Goldberg, R. B. (2007). Using genomics to study legume seed development. *Plant Physiol.* 144, 562–574. doi: 10.1104/pp.107.100362
- Leite, J. P., Barbosa, E. G., Marin, S. R., Marinho, J. P., Carvalho, J. F., Pagliarini, R. F., et al. (2014). Overexpression of the activated form of the AtAREB1 gene (AtAREB1DeltaQT) improves soybean responses to water deficit. *Genet. Mol. Res.* 13, 6272–6286. doi: 10.4238/2014.August.15.10
- Li, S., Pandey, S., Gookin, T. E., Zhao, Z., Wilson, L., and Assmann, S. M. (2012). Gene-sharing networks reveal organizing principles of transcriptomes in Arabidopsis and other multicellular organisms. *Plant Cell* 24, 1362–1378. doi: 10.1105/tpc.111.094748
- Liu, W., Mazarei, M., Peng, Y., Fethe, M. H., Rudis, M. R., Lin, J., et al. (2014). Computational discovery of soybean promoter cis-regulatory elements for the construction of soybean cyst nematode-inducible synthetic promoters. *Plant Biotechnol. J.* 12, 1015–1026. doi: 10.1111/pbi.12206
- Liu, X., Zhang, H., Zhao, Y., Feng, Z., Li, Q., Yang, H. Q., et al. (2013). Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15485–15490. doi: 10.1073/pnas.1304651110
- Locascio, A., Roig-Villanova, I., Bernardi, J., and Varotto, S. (2014). Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. *Front. Plant Sci.* 5:412. doi: 10.3389/fpls.2014.00412
- Loewus, F. A., and Murthy, P. N. (2000). myo-Inositol metabolism in plants. *Plant Sci.* 150, 1–19. doi: 10.1016/S0168-9452(99)00150-8
- Luo, Y., Qin, G., Zhang, J., Liang, Y., Song, Y., Zhao, M., et al. (2011). D-myo-inositol-3-phosphate affects phosphatidylinositol-mediated endomembrane function in Arabidopsis and is essential for auxin-regulated embryogenesis. *Plant Cell* 23, 1352–1372. doi: 10.1105/tpc.111.083337
- Maetschke, S. R., Madhamsheetiwar, P. B., Davis, M. J., and Ragan, M. A. (2014). Supervised, semi-supervised and unsupervised inference of gene regulatory networks. *Brief. Bioinformatics* 15, 195–211. doi: 10.1093/bib/bbt034
- Marbach, D., Costello, J. C., Kuffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., et al. (2012). Wisdom of crowds for robust gene network inference. *Nat. Methods* 9, 796–804. doi: 10.1038/nmeth.2016
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7
- Martin, T. F. (1998). Phosphoinositide lipids as signaling molecules: common themes for signal transduction, cytoskeletal regulation, and membrane trafficking. *Annu. Rev. Cell Dev. Biol.* 14, 231–264. doi: 10.1146/annurev.cellbio.14.1.231
- Meng, P. H., Raynaud, C., Tcherkez, G., Blanchet, S., Massoud, K., Domenichini, S., et al. (2009). Crosstalks between Myo-inositol metabolism, programmed cell death and basal immunity in Arabidopsis. *PLoS ONE* 4:e7364. doi: 10.1371/journal.pone.0007364
- Nagy, R., Grob, H., Weder, B., Green, P., Klein, M., Frelet-Barrand, A., et al. (2009). The Arabidopsis ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexakisphosphate transporter involved in guard cell signaling and phytate storage. *J. Biol. Chem.* 284, 33614–33622. doi: 10.1074/jbc.M109.030247
- Ni, Y., Aghamirzaie, D., Elmarakeby, H., Collakova, E., Li, S., Grene, R., et al. (2016). A machine learning approach to predict gene regulatory networks in seed development in Arabidopsis. *Front. Plant Sci.* 7:1936. doi: 10.3389/fpls.2016.01936
- O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 166, 1280–1292. doi: 10.1016/j.cell.2016.04.038
- Raboy, V. (1997). "Accumulation and storage of phosphate and minerals," in *Cellular and Molecular Biology of Plant Seed Development*, eds B. A. Larkins, I. K. Vasil (Dordrecht: Kluwer Academic Publishers), 441–477
- Raboy, V. (2007). The ABCs of low-phytate crops. *Nat. Biotechnol.* 25, 874–875. doi: 10.1038/nbt0807-874
- Ramsey, S. A., Klemm, S. L., Zak, D. E., Kennedy, K. A., Thorsson, V., Li, B., et al. (2008). Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLoS Comput. Biol.* 4:e1000021. doi: 10.1371/annotation/e14ad837-e5ff-4bd5-a5f2-f32e784d75a2
- Redekar, N., Biyashev, R., Jensen, R., Helm, R., Grabau, E., and Maroof, S. M. A. (2015). Genome-wide transcriptome analysis of developing seeds from low and normal phytic acid soybean lines. *BMC Genomics* 16, 1074. doi: 10.1186/s12864-015-2283-9
- Saghai Maroof, M. A., and Buss, G. R. (2008). *Low Phytic Acid, Low Stachyose, High Sucrose Soybean Lines*. Google Patents.
- Saghai Maroof, M. A., Glover, N. M., Biyashev, R. M., Buss, G. R., and Grabau, E. A. (2009). Genetic basis of the low-phytate trait in the soybean line CX1834. *Crop Sci.* 49, 69. doi: 10.2135/cropsci2008.06.0362
- Schafer, J., and Strimmer, K. (2005a). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21, 754–764. doi: 10.1093/bioinformatics/bti062
- Schafer, J., and Strimmer, K. (2005b). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4, Article32. doi: 10.2202/1544-6115.1175
- Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., et al. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* 34, 166–176. doi: 10.1038/ng1165
- Sherlock, G. (2000). Analysis of large-scale gene expression data. *Curr. Opin. Immunol.* 12, 201–205. doi: 10.1016/S0952-7915(99)00074-6
- Shi, J. R., Wang, H. Y., Schellin, K., Li, B. L., Faller, M., Stoop, J. M., et al. (2007). Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nat. Biotechnol.* 25, 930–937. doi: 10.1038/nbt1322
- Shi, J., Wang, H., Hazebroek, J., Ertl, D. S., and Harp, T. (2005). The maize low-phytic acid 3 encodes a myo-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. *Plant J.* 42, 708–719. doi: 10.1111/j.1365-313X.2005.02412.x
- Shi, J., Wang, H., Wu, Y., Hazebroek, J., Meeley, R. B., and Ertl, D. S. (2003). The maize low-phytic acid mutant *lpa2* is caused by mutation in an inositol phosphate kinase gene. *Plant Physiol.* 131, 507–515. doi: 10.1104/pp.014258
- Sparks, E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., et al. (2016). Establishment of expression in the shortroot-scarecrow transcriptional cascade through opposing activities of both activators and repressors. *Dev. Cell* 39, 585–596. doi: 10.1016/j.devcel.2016.09.031
- Stevenson, J. M., Perera, I. Y., Heilmann, I., Persson, S., and Boss, W. F. (2000). Inositol signaling and plant growth. *Trends Plant Sci.* 5, 252–258. doi: 10.1016/S1360-1385(00)01652-6
- Stevenson-Paulik, J., Bastidas, R. J., Chiou, S.-T., Frye, R. A., and York, J. D. (2005). Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12612–12617. doi: 10.1073/pnas.0504172102
- Taylor-Teeple, M., Lin, L., de Lucas, M., Turco, G., Toal, T. W., Gaudinier, A., et al. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575. doi: 10.1038/nature14099
- Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201
- Wilcox, J. R., Premachandra, G. S., Young, K. A., and Raboy, V. (2000). Isolation of high seed inorganic P, low-phytate soybean mutants. *Crop Sci.* 40: 1601. doi: 10.2135/cropsci2000.4061601x
- Xu, F., Meng, T., Li, P., Yu, Y., Cui, Y., Wang, Y., et al. (2011). A soybean dual-specificity kinase, GmSARK, and its Arabidopsis homolog, AtSARK, regulate leaf senescence through synergistic actions of auxin and ethylene. *Plant Physiol.* 157, 2131–2153. doi: 10.1104/pp.111.182899

- Xu, X. H., Zhao, H. J., Liu, Q. L., Frank, T., Engel, K. H., An, G., et al. (2009). Mutations of the multi-drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic acid in rice seeds. *Theor. Appl. Genet.* 119, 75–83. doi: 10.1007/s00122-009-1018-1
- Zhang, S., Yang, W., Zhao, Q., Zhou, X., Jiang, L., Ma, S., et al. (2016). Analysis of weighted co-regulatory networks in maize provides insights into new genes and regulatory mechanisms related to inositol phosphate metabolism. *BMC Genomics* 17:129. doi: 10.1186/s12864-016-2476-x
- Zhu, M., Dahmen, J. L., Stacey, G., and Cheng, J. (2013). Predicting gene regulatory networks of soybean nodulation from RNA-Seq transcriptome data. *BMC Bioinformatics* 14:278. doi: 10.1186/1471-2105-14-278

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

Copyright © 2017 Redekar, Pilot, Raboy, Li and Saghai Maroof. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.