Comparative Functional Genomics Characterization of Low Phytic Acid Soybeans and Virus Resistant Soybeans

Lindsay C. DeMers

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Crop and Soil Environmental Sciences

M. A. Saghai Maroof - Chair
Richard F. Helm
Guillaume Pilot
Song Li

April 27, 2020
Blacksburg, Virginia

Keywords: Transcriptomics, Gene regulatory network, Metabolomics, Lipids, Seed germination, Phytic acid, *Soybean mosaic virus*, *Rsv3* resistance

Comparative Functional Genomics Characterization of Low Phytic Acid Soybeans and Virus Resistant Soybeans

Lindsay C. DeMers

ABSTRACT

The field of functional genomics aims to understand the complex relationship between genotype and phenotype by integrating genome-wide approaches, such as transcriptomics, proteomics, and metabolomics. Large-scale "-omics" research has been made widely possible by the advent of high-throughput techniques, such as next-generation sequencing and mass-spectrometry. The vast data generated from such studies provide a wealth of information on the biological dynamics underlying phenotypes. Though functional genomics approaches are used extensively in human disease research, their use also spans organisms as miniscule as mycoplasmas to as great as sperm whales. In particular, functional genomics is instrumental in agricultural advancements for the improvement of productivity and sustainability in crop and livestock production. Improvement in soybean production is especially imperative, as soybeans are a primary source of oil and protein for human and livestock consumption, respectively. The research presented here employs functional genomics approaches – transcriptomics and metabolomics – to discern the transcriptional regulation and metabolic events underlying two economically important agronomic traits in soybean: seed phytic acid content and *Soybean mosaic virus* resistance. At normal levels, seed phytic acid content inhibits mineral absorption in humans and livestock, acting as an antinutrient and contributing to phosphorus pollution; however, the development of low phytic acid soybeans has helped mitigate these issues, as their seeds increase nutrient bioavailability and reduce environmental impact. Despite these desirable qualities, low phytic acid soybeans exhibit poor seed performance, which negatively affects germination rates and yield and has prevented their large-scale commercial production. Thus, part of the focus of this

research was investigating the effects of mutations conferring the low phytic acid phenotype on seed germination. Comparative studies between low and normal phytic acid soybean seeds were carried out and revealed distinct differences in metabolite profiles and in the transcriptional regulation of biological pathways that may be vital for successful seed germination. The final part of this research concerns *Rsv3*-mediated extreme resistance, a unique mode of resistance that is effective against the most virulent strains of *Soybean mosaic virus*. The molecular mechanisms governing this type of resistance are poorly characterized. Therefore, the research presented here attempts to elucidate the regulatory elements responsible for the induction of the *Rsv3*-mediated extreme resistance response. Utilizing a comparative transcriptomic time series approach on *Soybean mosaic virus*-inoculated *Rsv3* (resistant) and *rsv3* (susceptible) soybean lines, this final study provides gene candidates putatively functioning in the regulation of biological pathways demonstrated to be crucial for *Rsv3*-mediated resistance.

GENERAL AUDIENCE ABSRACT

Soybeans are a crop of great economic importance, being a primary source of oil and protein for human and livestock consumption, respectively. Increasing demand for soybean calls for improvement in its production. An emerging field that has had tremendous impact on this endeavor is the field of functional genomics. Functional genomics approaches generate large-scale biological data that can aid in discerning how specific processes are regulated and controlled in an organism. The research presented in this work utilizes functional genomics approaches to elucidate the biological mechanisms underlying two economically important traits in soybean: seed phytic acid content and *Soybean mosaic virus* resistance. Phytic acid is a compound found in soybean seeds that causes nutrient deficiencies and phosphorus pollution. Soybeans with reduced to phytic acid content have been developed to mitigate these problems; they have poor seed germination and emergence. The studies in this work employ functional genomics approaches to compare unique sets of low and normal phytic acid soybeans to help establish the relationship between seed phytic acid content and seed performance. These studies resulted in new and promising hypotheses for future studies on investigating the low phytic acid trait. The final focus of this work used a functional genomics approach to discern the molecular mechanisms underlying a unique mode of resistance to *Soybean mosaic virus*. The study identified genes in soybean that are potentially critical to resistance against *Soybean mosaic virus*.

To my Mom,
For giving me a beautiful life,
Opening every door,
And upholding every dream.

ACKNOWLEDGEMENTS

The success of my research would not have been possible without the support of several people, and the words below are not nearly enough to express my gratitude for each of them.

I would like to thank my professor, Dr. M. A. Saghai Maroof. I first met him during my freshman year as an undergraduate. He was a guest lecturer for an introductory course, and I remember being inspired and fascinated by his research. Not only did he give me the opportunity to do undergraduate research with him, but he also believed in me enough to give me the opportunity to pursue a PhD in his lab. In doing research in Dr. Maroof's lab, I have grown and learned more than I ever could have imagined, and for that, I am incredibly thankful.

I would like to thank Dr. Richard Helm for helping, teaching, and guiding me through the complicated world of metabolomics. I had absolutely no background in metabolomics and was initially quite worried about having to do a project on the subject, thinking I would never be able to figure it out; however, I have learned it and have even come to enjoy looking at chromatograms and spectra for clues as to the identity of a feature. I would also like to thank my other committee members, Dr. Guillaume Pilot and Dr. Song Li. Their expert advice, support in obtaining results and navigating their meaning, and always giving me new ideas on how to improve my projects or look at problem have been an invaluable part of shaping my research.

I would like to thank Dr. Ruslan Biyashev, Elizabeth Clevinger, Dr. Neelam Redekar, and Dr. Colin Davis from the Maroof lab for their training, assistance, and advice on all my projects; I am lucky to have learned and worked with all of them and to call each a friend. From the Helm lab, I would like to thank Dr. Sherry Hildreth and Jody Jervis for their friendship, laughs, and endless entertainment; without them and their technical expertise and support, I would still be scratching my head over metabolomics. I would like to thank Dr. Stephen

Rigoulot, a wonderful mentor and friend. I would also like to thank Dr. Sue Tolin, Dr. Elizabeth Grabau, Dr. Gregory Welbaum, and Dr. Victor Raboy for their advice and valuable feedback on my research.

I would like to thank my family and friends for their love, support, and encouragement. I would like to thank my mom for the sacrifices she made, for being my rock, and for her unwavering belief in me. I would like to thank my dad for his inspiring determination and encouragement to keep learning, and I would like to thank my sister, Grace, for all her little acts of thoughtfulness and making me laugh harder than anyone else. Lastly, I would like to thank Grant for being the best surprise and making me smile.

TABLE OF CONTENTS

ABSTRACT ..... **ii**

GENERAL AUDIENCE ABSTRACT ..... **iv**

ACKNOWLEDGEMENTS ..... **vi**

LIST OF FIGURES ..... **xi**

LIST OF TABLES ..... **xiii**

ABBREVIATIONS ..... **xiv**

ATTRIBUTIONS ..... **xvii**

CHAPTER 1 LITERATURE REVIEW **1**
INTRODUCTION 1
Overview of functional genomics 1
Importance of soybean research 4
Functional genomics research in soybeans 4
PHYTIC ACID 6
Phytic acid and its ecological impact 6
Phytic acid biosynthesis and its role in seeds 7
The development, significance, and drawbacks of low phytic acid crops 9
Lpa soybeans and the genetic material used in this research 10
GENETIC RESISTANCE TO SOYBEAN MOSAIC VIRUS 11
Soybean mosaic virus 11
SMV resistance loci – Rsv1, Rsv3, and Rsv4 13
Mechanism of Rsv3-mediated resistance 14
RESEARCH OBJECTIVES 15
REFERENCES 16

CHAPTER 2 NETWORK INFERENCE OF TRANSCRIPTIONAL REGULATION IN GERMINATING LOW PHYTIC ACID SOYBEAN SEEDS **34**
ABSTRACT 35
INTRODUCTION 36
MATERIALS AND METHODS 39
Genetic material 39
Seed germination and sampling 40
Transcriptomics data processing and differential gene expression 41
Transcriptional network construction and inference 41
Validation of network inferred interactions 42
RESULTS 43
Differential expression analysis 44
Co-expression analyses reveal altered phosphate ion homeostasis activity and

stress responses in lpa lines 45
Biological processes enriched in both developing and germinating seeds 49
Gene regulatory networks 49
DISCUSSION 51
Regulation of phosphate ion homeostasis in lpa lines 51
Downstream effects of perturbed myo-inositol metabolism in mips1 mutants 52
Myo-inositol metabolism and seed storage proteins in mrp-1/mrp-n mutant 54
Altered regulation in auxin and ABA signaling in lpa seeds 55
CONCLUSION 58
REFERENCES 60

CHAPTER 3 ANALYSIS OF LOW AND NORMAL PHYTIC ACID SOYBEAN (*GLYCINE MAX*) SEED LIPIDS AND SEED EXUDATES REVEALS DISTINCT CHEMOTYPES **81**
ABSTRACT 82
INTRODUCTION 84
MATERIALS AND METHODS 86
Plant material 86
Preparation of the lipid extracts 87
Liquid chromatography-mass spectrometry 87
Data processing and analysis 89
Seed electrolyte conductivity testing 90
Analysis of exuded proteins 90
Analysis of exuded metabolites 91
RESULTS AND DISCUSSION 92
Overview of seed lipid profiles 92
Mutations in phytic acid synthesis and transport have little effect on lipids in positive ion mode 93
Significant differences in lipid profiles between low and normal phytic acid lines in negative ion mode 94
Conductivity testing indicates lpa seeds have compromised membranes 97
2mlpa has unique protein and metabolite seed exudate profiles 98
CONCLUSION 99
REFERENCES 101

CHAPTER 4 A TRANSCRIPTIONAL REGULATORY NETWORK OF *RSV3*-MEDIATED EXTREME RESISTANCE AGAINST *SOYBEAN MOSAIC VIRUS* **117**
ABSTRACT 118
INTRODUCTION 119
MATERIALS AND METHODS 122
Soybean mosaic virus inoculations, leaf sampling, and RNA extraction 122

Sequence data processing and differential gene expression 123
Inference of gene regulatory networks 124
Expression clustering and gene function annotation 124
Network inference methods 124
Validation of inferred network interactions 125
RESULTS AND DISCUSSION 126
Fate of SMV-induced susceptibility or resistance in soybean is determined between 4 to 8 hours post-inoculation 127
Biological processes associated with Rsv3-mediated resistance in soybean show differential hormone responses 128
Suppression of MYC2 transcription factor expression is important for Rsv3-mediated ER 132
Modular regulation of abscisic acid signaling and suppression of jasmonic acid signaling are features of Rsv3-mediated ER 136
CONCLUSION 138
REFERENCES 140

CHAPTER 5 CONCLUSION **158**
RESEARCH SUMMARY 158
FUTURE DIRECTIONS 161

APPENDICES **164**

LIST OF FIGURES

2.1. Venn diagrams of differentially expressed genes (DEGs) 73

2.2. Mips class gene co-expression modules and significantly enriched biological processes 75

2.3. MRP class gene co-expression modules and significantly enriched biological processes 76

2.4. Mips-MRP class gene co-expression modules and significantly enriched biological processes 77

2.5. Consensus GRN of Mips and Mips-MRP genotypic subsets 79

3.1. Principal component analyses of the two genotypic classes in positive ion mode 108

3.2. Base peak ion chromatogram of complete master mix in positive ion mode showing enrichment in triacylglyercide content 109

3.3. Principal component analyses of the two genotypic classes in negative ion mode..... 110

3.4. Volcano plots of EMRTs found in negative ionization mode for each genotypic class . 111

3.5. Electrolyte conductivity of low and normal phytic acid soybean seeds from each genotypic class..... 114

3.6. SDS-PAGE of exudate from seeds belonging to the MRP genotypic class..... 115

3.7. Principal component analyses of seed exudate from NILs belonging to MRP genotypic class 116

4.1. Number of differentially expressed genes between soybean cultivars L29 and Williams82 at 2, 4, 6, and 8 hours post inoculation with *Soybean mosaic virus* strain G7 151

4.2. Co-expression gene modules and their biological functions 152

4.3. Comparison of normalized gene expression profiles of validated TFs in L29 and Williams82 154

LIST OF TABLES

2.1. Characteristics and classification of parental and experimental soybean lines 72

2.2. Differentially expressed genes in phytic acid biosynthesis pathway 74

2.3. Putative candidate transcription factors shared between genotypic subsets' gene regulatory networks 78

2.4. Putative target genes in the MRP subset with annotations for observed significant GO categories validated by *Arabidopsis* DAP-seq dataset and motif sequence analysis 80

3.1. Characteristics and classification of parental and experimental soybean lines 107

3.2. Significantly different EMRTs between low and normal phytic acid lines in the Mips subset from negative ion mode 112

3.3. Significantly different EMRTs between low and normal phytic acid lines in the MRP subset from negative ion mode 113

4.1. *A. thaliana* and motif validated interactions 153

4.2. TF target genes in module-2 related to ABA and auxin processes and defense responses 155

4.3. TF target genes in module-4 related to JA processes and defense responses 157

ABBREVIATIONS

1mlpa: a low phytic acid line with one mutation (*mips1*)

1MWT: a wild-type, normal phytic acid line with no mutations (MIPS1)

2mlpa: a low phytic acid line with two mutations (*mrp-l/mrp-n*)

2MWT: a wild-type, normal phytic acid line with no mutations (MRP-L/MRP-N)

2MWT-L: a normal phytic acid line with one mutation (MRP-L/*mrp-n*)

2MWT-N: a normal phytic acid line with one mutation (*mrp-l*/MRP-N)

3mlpa: a low phytic acid line with three mutations (*mips1/mrp-l/mrp-n*)

3MWT: a wild-type, normal phytic acid line with no mutations (MIPS1/MRP-L/MRP-N)

ABA: abscisic acid

ABC: ATP-binding cassette

ARACNE: Algorithm for the Reconstruction of Accurate Cellular Networks

bHLH: basic/helix-loop-helix

BIC: Bayesian Information Criteria

CK: cytokinin

CLR: context likelihood of relatedness

DAP-seq: DNA affinity purification sequencing

DEG: differentially expressed gene

ER: extreme resistance

ERF: ethylene responsive factor

EMRT: exact mass-retention time pair

FDR: false discovery rate

GRN: gene regulatory network

GO: gene ontology

hpi: hours post inoculation

HR: hypersensitive response

ICL: integrated complete-data likelihood

JA: jasmonic acid

LARS: least angle regression

LC-MS: liquid chromatography-mass spectrometry

lpa: low phytic acid

MIPS: *myo*-inositol phosphate synthase

ML: machine learning

MRP: multidrug resistance-associated protein

MYB: myeloblastosis oncogene

MYC2: myelocytomatosis proto-oncogene, bHLH 6

NAC: NAM, ATAF1/2, and CUC

NGS: next-generation sequencing

NIL: near isogenic line

P: phosphorus

PA: phosphatidic acid

PE: phosphatidylethanolamine

Pi: phosphate

PI: phosphatidylinositol

PCA: principal component analysis

RSM: RADIALIS-LIKE SANT/MYB

Rsv: resistance to SMV

SA: salicylic acid

SMV: *Soybean mosaic virus*

TAG: triacylglycerol

TF: transcription factor

ATTRIBUTIONS

The manuscripts presented in this dissertation, i.e. chapters 2, 3, and 4 have multiple co-authors. Contributions of all co-authors are described as follows:

Dr. M. A. Saghai Maroof: Professor, School of Plant and Environmental Sciences at Virginia Tech. Participated in research proposal preparation, experimental design, and review of all manuscripts.

Dr. Richard F. Helm: Professor, Department of Biochemistry at Virginia Tech. Participated in research proposal preparation, experimental design, conduct of experiments, and review of phytic acid manuscripts.

Dr. Song Li: Assistant Professor, School of Plant and Environmental Sciences at Virginia Tech. Participated in experimental design and review of transcriptomics manuscripts.

Dr. Neelam R. Redekar: Research Associate, Department of Crop and Soil Science at Oregon State University. Participated in experimental design and review of transcriptomics manuscripts.

Dr. Sherry B. Hildreth: Reasearch Associate, Department of Biochemistry and Biology at Virginia Tech. Participated in experimental design, conduct of experiments, data analysis, and review of metabolomics manuscript.

Jody Jervis: Lab Specialist, Department of Biochemsitry at Virginia Tech. Participated in conduct of metabolomics experiments.

Dr. Victor Raboy: Research Geneticist, National Small Grains Germplasm Center at Agricultural Research Service (USDA). Participated in review of phytic acid manuscripts.

Dr. Sue A. Tolin: Professor Emerita, Plant Pathology, Physiology, and Weed Science at Virginia Tech. Participated in review of *Rsv3* manuscript.

Dr. Aardra Kachroo: Professor, Department of Plant Pathology at University of Kentucky. Participated in review of *Rsv3* manuscript.

CHAPTER 1

Literature Review

INTRODUCTION

Overview of functional genomics

Functional genomics is a field of molecular biology that takes advantage of the vast wealth of available genomics data to discern gene functions and interactions. Rather than focusing on the static features of genomic information, such as DNA sequence and structures, functional genomics is more concerned with the dynamic features of "-omics" related research, such as transcriptomics, proteomics, metabolomics, and protein-protein interactions, which enables the large-scale study of gene transcription, translation, and gene expression regulation. A distinguishing aspect of functional genomics studies is their genome-wide approach, which typically requires high-throughput approaches instead of the more traditional "gene-by-gene" approach [1].

Since their introduction in 2005, next-generation sequencing (NGS) technologies have had a tremendous impact on advancing high-throughput functional genomics research of DNA and RNA. NGS methods have allowed millions to trillions of nucleic acid observations to be

made in parellel during a single intrsument run, working considerably faster and more cheaply than the older method of Sanger sequencing [2, 3]. The ability to generate gigabase (Gb)-sized sequences in just a few days or hours has enabled not only whole-genome sequencing (WGS) for the construction of *de novo* draft genome sequences, but NGS has also allowed for whole transcriptome shotgun sequencing (WTSS) (more commonly known as RNA sequencing (RNA-seq)), whole-exome sequencing (WES), targeted (TS) or candidate gene sequencing (CGS), methylation sequencing (MeS), and ChIP-seq [4-11]. RNA-seq can be used to determine total transcriptional activity (coding and noncoding) or a select subset of targeted RNA transcripts in a given sample, providing a more accurate and sensitive measurement of gene expression than microarrays [6, 12]. WES permits the analysis of protein-coding regions (CDS) of the genome and the identification of coding variants [7, 13]. Investigation of the methylome by MeS aids in the discovery of active methylation sites and epigenetic markers that regulate gene expression, epistructural base variations, imprinting, development, differentiation, and the epigenetic state [11, 14-16]. With ChIP-seq, chromatin immunoprecipitation (ChIP) is succeded by NGS sequencing and facilitates genome-wide profiling of DNA-binding proteins and histone and nucleosome modifications [11]. Each of these types of sequencing requires an NGS platform. Common platforms include HiSeq and MiSeq (Illumina), SOLiD (Life Technologies), Ion Torrent, PacBio RS II and SMRT (Pacific Biosciences), and Nanopore (Oxford Nanopore Technologies). When deciding on which sequencing platform to use for a project, certain factors should be considered, such as read length, time per run, cost per base, and raw error rate. The main features and performances of these platforms are reviewed here [4].

Following genomics and transcriptomics, proteomics and then metabolomics are the next level of study of biological systems and constitute an important component of functional

genomics research, as they provide a closer interface to the cellular phenotype [17, 18]. Proteomics typically refers to the large-scale study of proteins and proteomes – the complete set of proteins produced or modified by an organism [19, 20]. Metabolomics is the study of small-molecule metabolite profiles produced by cellular processes, and the metabolome is the entire set of metabolites (generally defined as <1.5 kDa) in a biological cell, tissue, organ, or organism [18, 21]. Both proteomics and metabolomics provide more information on the physiological state of an organism than transcriptomics and genomics. In genomics, genomes are relatively static, whereas proteomes and metabolomes are transient, fluctuating from cell to cell and from time to time [22, 23]. With transcriptomics, it was found that there is no correlation between mRNA content and protein or metabolite content [24, 25]. To obtain the added the levels of information provided by proteomics and metabolomics, the most common high-throughput analysis methods are mass-spectrometry (MS)-based, with a preceding separation step, such as gas chromatography (GC), high performance liquid chromatography (HPLC), or capillary electrophoresis (CE), for metabolomics samples. It should be noted that there is not yet a single analytical method that is able to capture the entire metabolome within a sample [26].

The primary botttleneck in genomics, transcriptomics, proteomics, and metabolomics studies is processing and analyzing the vast amount of data generated by the high-throughput technologies. This has sparked major growth in the rapidly evolving field of bioinformatics. For each type of "-omics" study, information on best practices for wet-lab procedures, computational pipelines, and software tools can be reviewed here [23, 27-32]. Together, the different levels of functional genomics – genomics, transcriptomics, proteomics, and metabolomics – help us connect genotype to phenotype, providing a more complete picture and improving our understanding of the dynamic properties that make up and govern an organism.

Importance of soybean research

Domesticated soybeans (*Glycine max* (L.) Merr.) are one of the most extensively grown crops in the world, with the United States being the number one producer, planting 90 million acres in 2018 for an estimated economic value of nearly $40 billion [33]. Soybeans are primarily grown as protein and oil sources to be used in feed and food products. They account for roughly 90% of US oilseed production [33], but the predominant end use of soybeans is actually soymeal for livestock feed production. Soymeal, the residue remaining after oil extraction, is a major metabolizable energy source and the number one protein source for livestock industries throughout the world [34]. Still yet, soybeans are also used in cosmetics, pharmaceuticals, biodiesels, and other industrial products such as adhesives, printing inks, building materials, and lubricants. Because of the rising demand for soybean products, climate change, and the expanding global population, soybean production must increase in order to meet our needs. The continued development of high-performing cultivars with desirable agronomic traits, such as high yield, stress tolerance, pest resistance, and enhanced nutritional composition and seed perfomance, is perhaps best achieved through genetic research. At the forefront, making major advancements in soybean research, has been the implementation of functional genomics approaches.

Functional genomics research in soybean

The release of the soybean reference genome sequence of cultivar "Williams82" in 2010 helped facilitate the integration of massive volumes of genetic, phenotypic, and genomic data. This has aided in the creation of websites like "Soybase" (https://soybase.org), "SoyKB" (http://soykb.org), and "Phytozome" (https://phytozome.jgi.doe.gov/pz/portal.html), which

enable users to access data by sequence, gene name, marker, trait of interest, expression pattern, and homology to genes in other species [35-37]. Recently, genomic sequences from 106 soybean accessions were published [38]. Represented are wild, landrace, and elite lines, allowing a landscape analysis of genome-wide genetic variation and an association study of major soybean domestication and agronomic traits. The re-sequenced lines permitted the discovery of more than 10 million SNPs, 159 putative domestication sweeps, and novel alleles for major traits such as oil and protien content. The genomic information from this landmark study is a valuable resource for studying genetic diversity and thus advancing the genetic improvement of soybean. At the transcriptome level, the "RNA-Seq Atlas" is a comprehensive, high-resolution, gene expression resource, providing expression data on seven unique soybean tissues and seven stages of soybean seed development (https://soybase.org/soyseq/) [39]. This resource serves as a model for future RNA-seq studies and for evaluating gene model annotations in the soybean reference genome, as well as providing a means for evaluating differential gene expression beteween differing tissues and developmental stages, allowing insights to be gained on gene functions and biological processes in distinct tissue types. The most comprehensive soybean proteomics resource is the "Soybean Proteome Database" [40]. The focus of this database is majorly on soybean seedling responses to flooding, but it also contains data on drought and salt stress and several organs, tissues, and organelles. This database can aid in functional analyses for advacing soybean research and is available at http://proteome.dc.affrc.go.jp/Soybean/. Lastly, at the metabolome level is "SoyMetDB" (http://soymetdb.org/), a metabolomics database for soybean [41]. It aids in integrating, mining, and visualizing soybean metabolomics data, as well as identifying metabolites and measuring their expression. SoyMetDB also integrates metabolite information from other public metabolomics databases and includes a pathway enrichment tool for highly

expressed metabolites. Such a database is a valuable resource for the soybean community, where still much progress is needed in the field of metabolomics. Besides the soybean functional genomics resources given above, the advent of high-throughput functional genomics technologies has spurred a wealth of other soybean "-omics" studies, and accordingly, our knowledge on soybean is rapidly growing, thus enabling the crop's continued improvement.

PHYTIC ACID

Phytic acid and its ecological impact

Soybean seeds are rich in a compound called phytic acid (phytate, *myo*-inositol-(1,2,3,4,5,6)-hexa*kis*phosphate, $InsP_6$). It is the primary storage form of seed phosphorus (P), sequestering approximately 75% of total seed P [42, 43]. Typically in the form of a mixed salt called phytin, phytic acid chelates nutritionally important, positively charged minerals such as Fe^{3+}, K^{+}, Ca^{2+}, Mg^{2+}, and Zn^{2+} [43-46]. Because phytic acid binds up these elements, the bioavailability of P and other important minerals is reduced. This is a problem in monogastric livestock (swine, poultry, fish) and humans, which have minimal or no phytase activity in their digestive tract, rendering phytic acid indigestible to them [44]. Resultantly, nutrient deficiencies can be found among monogastric livestock and human populations that heavily rely on staple crops such as maize, rice, and wheat as their primary source of nutrition [47-54]. In fact, phytic acid is considered the most important antinutritional factor limiting the availability of minerals like zinc, calcium, and iron [55, 56]. It is estimated that over one billion people suffer from iron deficiency, and hundreds of millions suffer from zinc and other mineral deficiencies [57]. In respect to seed total P, only about 20-30% is available (non-phytic acid P); however, this is insufficient for meeting nutritional needs [43]. Thus, in the case of livestock and in order to

optimize productivity, feed is supplemented with microbial phytase (an enzyme that degrades phytic acid) or an available form of P to release phosphate (Pi) for absorption [44]. Nonetheless, the nutritional disadvantages of phytic acid are not its only shortcomings. Phytic acid is also considered one of the leading causes of P pollution [58]. Since it is indigestible, it is excreted into the environment in the form of animal manure, where it contributes to the build up of P in soil and water. This excess P can lead to pollution of water systems and result in eutrophication [59, 60].

Phytic acid biosynthesis and its role in seeds

Phytic acid biosynthesis takes place in the cytoplasm during seed development. The first substrates required for biosynthesis are Pi and *myo*-inositol (Ins) [43]. Pi is supplied by P uptake at the root-rhizosphere interface. P is then transported and localized at the endosperm of the developing seed, as seeds usually store more P than is needed to fulfill basic cellular processes [43, 61]. Ins, a 6-carbon cyclic alcohol and the backbone of phytic acid, is produced from the conversion of D-glucose-6-P to Ins(3)P_1. This reaction is catalyzed by *myo*-inositol-3-monophosphate synthase (MIPS), and it is the sole source of the Ins ring [57, 62]. Ins monophosphatase hydrolyzes Ins(3)P_1 to Ins and Pi. Ins is the first substrate required for two different, subsequent pathways for phytic acid biosynthesis, the lipid-dependent pathway and the lipid-independent pathway [63, 64]. The main difference between these pathways is in their early intermediate steps in converting Ins to Ins trisphosphates. Though the lipid-dependent pathway is the primary mechanism for phytic acid biosynthesis in the eukaryotic cells, the main pathway for phytic acid accumulation in seeds is the lipid-independent pathway [57, 64-67]. The first step in this pathway is the phosphorylation of Ins to $InsP_1$ to $InsP_2$ via Ins kinase and Ins

monophosphate kinase; these early steps could be unique to phytic acid biosynthesis in seeds. Ins polyphosphate kinases catalyze subsequent sequential phosphorylations to $InsP_3$, $InsP_4$, $InsP_5$, and ultimately $InsP_6$, i.e. phytic acid [43, 57, 66, 68]. After synthesis, phytic acid is transported as the mixed salt, phytin, and deposited as inclusion bodies (referred to as "globoids") found in protein storage vacuoles [63, 69]. Transport and storage occurs via a multidrug resistance-associated protein (MRP), a type of ATP-binding cassette (ABC) transporter [43, 70, 71]. Phytic acid deposition only takes place in cells that remain alive during the quiescent phase of seed development [45]. In dicots, this deposition occurs in the endosperm and cotyledons; in monocots, it occurs in the endosperm and aleurone layer [72]. Throughout seed development and even into seed maturation, phytic acid continues to accumulate in the seed. Once at the dry seed stage, the seeds contain phytase potential, which is activated upon germination. During germination, phytase activity increases, and phytic acid is eventually degraded to release Ins and useful minerals for seedling growth [44, 73, 74].

In addition to P and mineral storage, phytic acid and the intermediates in its biosynthesis, such as Ins, have fundamental roles in various metabolic, developmental, and signaling pathways critical to plant function and productivity [57]. These include a variety of basic cellular housekeeping activities, such as DNA repair, chromatin remodeling, RNA editing and export, ATP and cell wall polysaccharide synthesis, and regulation of gene expression, guard cells, basal defense, and cell death [62, 67, 75-81]. Furthermore, the phytic acid pathway has roles in P homeostasis and signal transduction for stress responses, development, and Pi sensing [57, 63, 80]. Lastly, phytic acid itself acts as an antioxidant during seed germination, chelating heavy metals [82].

The development, significance, and drawbacks of low phytic acid crops

Because of phytic acid's antinutritive properties and negative environmental impacts, the development of low phytic acid (*lpa*) crops has been a widespread pursuit. In fact, this trait has already been engineered by targeting enzymes in the phytic acid biosynthesis and transport pathways. Enzymes that have been targeted and resulted in reduced phytic acid content include MIPS, multidrug resistance-associated protein ATP-binding cassette (ABC) transporters (henceforth referred to as MRPs), and Ins and polyphosphate kinases [70, 83-85]. So far, *lpa* barley, maize, rice, soybean, and wheat have all been developed via random mutagenesis and phenotype screening [70, 83-89]. These *lpa* crops exhibit increased seed Pi content, while maintaining the same total seed P content. Depending on the mutation conferring the *lpa* trait, phytic acid reductions can be as high as 90% [88]. Multiple animal nutrition studies with poultry, swine, and fish have shown that *lpa* seeds increase P bioavailability and reduce P waste, helping satisfy dietary requirements and decrease water pollution [65, 90]. In the case of human nutritional studies, the bioavailability of iron, zinc, and calcium increased 30-50% using foods prepared from *lpa* crops [53, 91, 92]. The cases presented here illustrate there would be clear advantages with commercial production of *lpa* crops – increased nutritional value of feed and food supplies and reduced P pollution. Nevertheless, there is a drawback to *lpa* crops which has prevented their large-scale use. Many studies have demonstrated that perturbations in phytic acid metabolism are detrimental to seed viability and performance, causing low germination and emergence rates and ultimately reducing yield. As well, other observations in *lpa* crops include reduced seed dry weight accumulation and stress and desiccation tolerance and increased disease susceptibility and oxidative damage, the latter of which results in premature aging [57, 88, 93-95]. The severity of these issues associated with the *lpa* trait is heavily influenced by

environment, being exacerbated in subtropical climates [94, 96]. However, it is ill understood why reducing phytic acid content has such damaging consequences, thus calling for more research into the genetic and molecular basis of seed and seedling performance in relation to phytic acid content.

Lpa soybeans and the genetic material used in this research

Various mutant soybean lines with the *lpa* phenotype have been characterized, such as "LR33," "M766," "M153," "CX-1834," "V99-5089," "*Gm-lpa*-TW-1," and "*Gm-lpa*-ZC-2" [83, 89, 97-101]. The sets of experimental lines used in this research were developed from crosses using CX-1834, V99-5089, and "Essex" (PI 548667), a high-yielding cultivar developed at Virginia Tech with normal phytic acid and sugar content [102].

CX-1834 is a *G. max* cultivar developed by the USDA/Purdue University [89]. It produces *lpa* seeds that retain normal sugar content. This phenotype is controlled by two alleles that must both be homozygous recessive [98]. Quantitative trait locus (QTL) mapping revealed two epistatically interacting loci on chromosomes 3 (linkage group (LG)-N) and 19 (LG-L) to be responsible for the *lpa* phenotype in CX-1834 [97]. The mutations in both of these loci are single nucleotide polymorphisms (SNPs) in MRP genes – a G to A SNP on LG-L, resulting in the substitution of an arginine amino acid residue with a lysine, and an A to T SNP on LG-N, resulting in the replacement of an arginine residue with a stop codon [100]. These two MRP genes on LG-L and -N will hereafter be referred to as MRP-L and MRP-N, respectively.

V99-5089, a patented *G. max* cultivar developed at Virginia Tech, produces seeds that are not only low in phytic acid but are also low in stachyose and high in sucrose [101]; these additional changes in sugar content are desirable, as they provide more metabolizable energy.

The QTL responsible for this *lpa*/low stachyose/high sucrose phenotype mapped to chromosome 11 and is the result of a C to G SNP in the coding region of the MIPS1 gene [101].

From three unique genetic crosses using CX-1834, V99-5089, and Essex as parents, the Maroof Lab at Virginia developed several experimental lines in order to study the molecular and genetic basis of phytic acid in relation to seed and seedling performance. Of these experimental lines, eight were used in this research. One of these genetic crosses was Essex x V99-5089 from which the following near isogenic lines (NILs) were developed: an *lpa* line, designated as *1mlpa* (*mips1*/MRP-L/MRP-N), and a normal phytic acid line, designated as 1MWT (MIPS1/MRP-L/MRP-N). In another genetic cross, CX-1834 x V99-5089, four NILs were developed: one *lpa* line, designated as *2mlpa* (MIPS1/*mrp-l/mrp-n*), and three normal phytic acid lines, designated as 2MWT (MIPS1/MRP-L/MRP-N), 2MWT-L (MIPS1/MRP-L*/mrp-n*), and 2MWT-N (MIPS/*mrp-l*/MRP-N). Finally, from the same genetic cross (CX-1834 x V99-5089), the last set of genetic crosses includes: an *lpa* line, designated as *3mlpa* (*mips1/mrp-l/mrp-n*), and a normal phytic acid line, designated as 3MWT (MIPS1/MRP-L/MRP-N).

For the studies presented here, seeds from all eight experimental lines (*1mlpa*, 1MWT, *2mlpa*, 2MWT, 2MWT-L, 2MWT-N, *3mlpa*, 3MWT) were harvested in 2017 from a field in Blacksburg, VA. Their genotypes were verified by allelic discrimination with SNP genotyping using KASPar^{TM} assays (LGC Biosearch Technologies, Hoddesdon, UK).

GENETIC RESISTANCE TO SOYBEAN MOSAIC VIRUS

Soybean mosaic virus

Predominantly originating in South and East Asia, *Soybean mosaic virus* (SMV, genus *Potyvirus*, family *Potyviridae*) is one of 39 viruses in the *Bean common mosaic virus* (BCMV)

lineage of potyviruses [103]. It is a widespread viral pathogen of *G. max* (cultivated soybean) and *G. soja* (wild soybean), being found in all soybean-growing regions of the world and causing significant damage to seed yield and quality. It is transmitted via aphid species and seeds, the latter being the most prevalent form of transmission. Once infected, foliar symptoms range from moderate to severe leaf mottling and distortion, necrosis, stunting, and even plant death. In seeds, infection symptoms include seed coat mottling, reduced size, weight, and viability, and altered chemical compostion [104]. With these damages, yield losses associated with SMV are typically more than 30% but can be as high as 94% [105]. The severity of losses is highly dependent on host genotype, virus strain, infection incidence, and plant growth stage at the time of infection [106-112]. In the case of virus strain, seven strain groups (G1 to G7) were classified in the United States based on disease reactions of 98 collected SMV isolates on a series of differential soybean cultivars [113]. The same classification system was used in Korea, leading to the identification of SMV strains G5H, G6H, and G7H [114-117]. However, in Japan and China, a different series of soybean cultivars were used in disease differenials, and SMV isolates collected in these countries were classified into five (A to E) and 21 (SC1 to SC21) strains, respectively [118-121].

The SMV genome consists of a monopartite, single-stranded, positive-sense RNA approximately 9.6 kilobases long [122]. Its single open reading frame (ORF) is translated into polyprotein, which is cleaved by three SMV protienases to yield 11 multifunctional proteins – P1 (protein 1), HC-Pro (helper component-protease), P3 (protein 3), 6K1 (first 6KDa peptide), CI (cylindrical inclusion), 6K2 (second 6KDa peptide), NIa (nuclear inclusion “a”-protease), NIb (nulcear inclusion “b”-replicase), CP (coat protein), and P3-PIPO (pretty interesting potyviruses ORF); self-cleavage of NIa produces VPg (virus genome-linked protein, covalently attached at

the 5' end) and a protease domain [123-125]. The functions of these proteins include symptom development, host adaptation, aphid and seed transmission, suppression of gene silencing, virus movement, plasmodesmata targeting, virulence and pathogenicity, protein cleavage, viral genome replication, and virion assembly [104]. Sometimes these viral proteins are detected by host disease resistance (R) proteins, which triggers a host defense response [126, 127]. Taking advantage of genes encoding R proteins has been one of the most effective strategies for managing SMV.

SMV resistance loci – Rsv1, Rsv3, and Rsv4

R proteins are encoded by resistance (*R*) genes. Through extensive research on the SMV-soybean pathosystem, three dominant genes, *Rsv1*, *Rsv3*, and *Rsv4*, have been identified and characterized as conferring strain-specific resistance to the seven US SMV strains [128-132]. The *Rsv1* locus was mapped to soybean chromosome 13 and contains at least 10 alleles [104, 133]. Within this region, a cluster of nucleotide-binding leucine-rich repeat (NB-LRR)-type *R* gene candidates were found. Several of these genes were demonstrated to condition unique resistance responses to SMV strains [134]. Soybeans with the *Rsv1*-genotype exhibit extreme resistance (ER) to strains G1-G3, while strains G4-G7 result in necrotic or mosaic symptoms [135]. With this type of resistance, i.e. ER, the virus is asymptomatic and undetectable in inoculated leaves [136, 137]. The *Rsv4* locus was mapped to chromosome 2, and eleven candidate genes were identified [128, 138, 139]. *Rsv4* confers resistance to G1-G7 strains [132, 140]; however, limited replication in inoculated leaves is still detected, and systemic movement can be observed later on. Thus *Rsv4*-mediated resistance may be overcome and result in late susceptibility symptoms [104]. Finally, the *Rsv3* locus was mapped to a 154 kilobase region on

chromosome 14 containing a cluster of five highly similar *R* genes encoding coiled-coil nucleotide binding-leucine rich (CC-NB-LRR) proteins [141, 142]. Comparative sequence analysis of these five genes from resistant and susceptible lines indicated that Glyma.14G38533 is the most likely candidate gene for *Rsv3* [143]. This was later confirmed by cloning [144]. Like *Rsv1*, *Rsv3* conditions ER-type resistance; however, ER is instead conferred to the most virulent SMV strains (G5-G7), while G1-G4 strains result in susceptibility [145].

Mechanism of Rsv3-mediated resistance

Only a few studies have investigated the mechanism of *Rsv3*-mediated resistance; therefore, very little is known on how *Rsv3* induces an ER response upon SMV inoculation. Thus far, research has shown that *Rsv3*-mediated resistance is triggered by a CC-NB-LRR R protein detecting the viral CI protein [143, 146]; however, the signaling pathway following this is unclear. The proposed model for *Rsv3*-mediated resistance, based on findings at 8, 24, and 54 hours post-inoculation (hpi), is that recognition of CI by the R protein Rsv3 causes abscisic acid (ABA) accumulation and thus activates the ABA signaling pathway [147]. The portion of the ABA signaling pathway that is linked to *Rsv3*-mediated resistance transcriptionally up-regulates a subset of the type 2C protein phosphatase (*PP2C*) genes. These genes function as positive regulators of the *Rsv3*-mediated signaling and stimulate callose deposition, which inhibits viral cell-to-cell movement and restricts virus accumulation at the initially infected cells [147]. Despite the proposed model, there remains a disconnect regarding how the ABA signaling pathway is triggered and regulated. It should also be noted that ABA signaling and callose deposition are not the sole mechanism by which *Rsv3*-mediated resistance is conditioned, as many changes in gene expression have also been found in the autophagy, small interfering (si)

RNA, and jasmonic acid (JA) pathways [148]; these findings were also from data collected at 8, 24, and 54 hpi. However, ER responses are rapid, working faster than the more typical resistance mediated by a hypersensitive response (HR) [149]. Thus, many of the early events induced by *Rsv3* are likely unexplored. Consequently, because of *Rsv3*'s uniqueness and ill-characterized mode of activation, one of the studies presented in this work focuses on identifying regulatory genes in biological processes that may be essential to the dynamics of *Rsv3*-mediated resistance.

RESEARCH OBJECTIVES

1) Network inference of transcriptional regulation in germinating low phytic acid soybean seeds
2) Analysis of low and normal phytic acid soybean (*Glycine max*) seed lipids and seed exudates reveals distinct chemotypes
3) A transcriptional regulatory network of *Rsv3*-mediated extreme resistance against *Soybean mosaic virus*

REFERENCES

1. Kaushik S, Kaushik S, Sharma D. Functional Genomics. Encyclopedia of Bioinformatics and Computational Biology. 2019; 2:118-133.
2. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. Genomics. 2008; 92(5):255-64.
3. Metzker ML. Sequencing technologies—the next generation. Nature Reviews Genetics. 2010; 11(1):31-46.
4. Kulski JK. Next-generation sequencing—an overview of the history, tools, and "Omic" applications. Next Generation Sequencing–Advances, Applications and Challenges. 2016:3-60.
5. Lam HY, Clark MJ, Chen R, Chen R, Natsoulis G, O'huallachain M, et al. Performance comparison of whole-genome sequencing platforms. Nature Biotechnology. 2012; 30(1):78-82.
6. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009; 10(1):57-63.
7. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. Journal of Human Genetics. 2014; 59(1):5-15.
8. Leo V, Morgan N, Bem D, Jones M, Lowe G, Lordkipanidzé M, et al. Use of next-generation sequencing and candidate gene analysis to identify underlying defects in patients with inherited platelet function disorders. Journal of Thrombosis and Haemostasis. 2015; 13(4):643-50.
9. Kulski JK, Suzuki S, Ozaki Y, Mitsunaga S, Inoko H, Shiina T. In phase HLA genotyping by next generation sequencing-a comparison between two massively parallel

sequencing bench-top systems, the Roche GS Junior and ion torrent PGM. HLA and Associated Important Diseases Croatia: Intech. 2014:141-81.

10. Pelizzola M, Ecker JR. The DNA methylome. FEBS letters. 2011; 585(13):1994-2000.

11. Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. Molecular Systems Biology. 2013; 9(1):640-53.

12. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nature Reviews Genetics. 2011; 12(2):87-98.

13. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. BMC Bioinformatics. 2014; 15(1):247-57.

14. Chang G, Gao S, Hou X, Xu Z, Liu Y, Kang L, et al. High-throughput sequencing reveals the disruption of methylation of imprinted gene in induced pluripotent stem cells. Cell Research. 2014; 24(3):293-306.

15. Ekram MB, Kim J. High-throughput targeted repeat element bisulfite sequencing (HT-TREBS): genome-wide DNA methylation analysis of IAP LTR retrotransposon. PloS One. 2014; 9(7):e101683.

16. Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schönegger A, Klughammer J, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Reports. 2015; 10(8):1386-97.

17. Karr T. Application of proteomics to ecology and population biology. Heredity. 2008; 100(2):200-6.

18. Fiehn O. Metabolomics—the link between genotypes and phenotypes. Functional Genomics: Springer; 2002. p. 155-71.

19. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998; 19(11):1853-61.

20. Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. Trends in Biotechnology. 1999; 17(3):121-7.

21. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the human metabolome database. Nucleic Acids Research. 2007; 35(suppl_1):D521-D6.

22. Kosmides AK, Kamisoglu K, Calvano SE, Corbett SA, Androulakis IP. Metabolomic fingerprinting: challenges and opportunities. Critical Reviews™ in Biomedical Engineering. 2013; 41(3):205-21.

23. Graves PR, Haystead TA. Molecular biologist's guide to proteomics. Microbiology and Molecular Biology Reviews. 2002; 66(1):39-63.

24. Rogers S, Girolami M, Kolch W, Waters KM, Liu T, Thrall B, et al. Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. Bioinformatics. 2008; 24(24):2894-900.

25. Cavill R, Jennen D, Kleinjans J, Briedé JJ. Transcriptomic and metabolomic data integration. Briefings in Bioinformatics. 2015; 17(5):891-901.

26. Riekeberg E, Powers R. New frontiers in metabolomics: from measurement to insight. F1000Research. 2017; 6:1148-57.

27. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biology. 2016; 17(1):13-31.

28. Ekblom R, Wolf JB. A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications. 2014; 7(9):1026-42.

29. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. Nature Reviews Genetics. 2018; 19(6):329-46.

30. Sinitcyn P, Rudolph JD, Cox J. Computational methods for understanding mass spectrometry–based shotgun proteomics data. Annual Review of Biomedical Data Science. 2018; 1:207-34.

31. Gorrochategui E, Jaumot J, Lacorte S, Tauler R. Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: overview and workflow. TrAC Trends in Analytical Chemistry. 2016; 82:425-42.

32. Nalbantoglu S. Metabolomics: Basic Principles and Strategies. In Molecular Medicine 2019 Aug 7. IntechOpen.

33. USDA Economic Research Service. ERS Datasets. Economic Research Service, U.S. Department of Agriculture; 2018.

34. Stein HH, Berger LL, Drackley JK, Fahey Jr G, Hernot DC, Parsons CM. Nutritional properties and feeding values of soybeans and their coproducts. In: Johnson LA, White PJ, Galloway R, editors. Soybeans. Urbana, IL: AOCS Press; 2008. p. 613-60.

35. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research. 2009; 38(suppl_1):D843-D6.

36. Joshi T, Patil K, Fitzpatrick MR, Franklin LD, Yao Q, Cook JR, Wang Z, Libault M, Brechenmacher L, Valliyodan B, Wu X. Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. In BMC Genomics 2012 Dec 1 (Vol. 13, No. S1, p. S15). BioMed Central.

37. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research. 2011; 40(D1):D1178-D86.

38. Valliyodan B, Qiu D, Patil G, Zeng P, Huang J, Dai L, et al. Landscape of genomic diversity and trait discovery in soybean. Scientific Reports. 2016; 6:23598-607.

39. Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, et al. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biology. 2010; 10(1):160-75.

40. Ohyanagi H, Sakata K, Komatsu S. Soybean Proteome Database 2012: update on the comprehensive data repository for soybean proteomics. Frontiers in Plant Science. 2012; 3:110-5.

41. Joshi T, Yao Q, Levi DF, Brechenmacher L, Valliyodan B, Stacey G, et al., editors. SoyMetDB: the soybean metabolome database. 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2010: IEEE.

42. Erdman J. Oilseed phytates: nutritional implications. Journal of the American Oil Chemists' Society. 1979; 56(8):736-41.

43. Raboy V. Seed total phosphate and phytic acid. In: Kriz AL, Larkins BA, editors. Molecular Genetic Approaches to Maize Improvement, Biotechnology in Agriculture and Forestry. Berlin, Germany: Springer Berlin Heidelberg; 2009. p. 41-53.

44. Brinch-Pedersen H, Sørensen LD, Holm PB. Engineering crop plants: getting a handle on phosphate. Trends in Plant Science. 2002; 7(3):118-25.

45. Lott JNA, Greenwood JS, Batten GD. Mechanisms and regulation of mineral nutrient storage duringseed development. In: Kigel J, Galili G, editors. Seed development and germination. New York: Marcel Dekker, Inc; 1995. p. 215–235.

46. Weaver CM, Kannan S. Phytate and mineral bioavailability. Food Phytates. 2002; 2002:211-23.

47. Adeola O, Lawrence B, Sutton A, Cline T. Phytase-induced changes in mineral utilization in zinc-supplemented diets for pigs. Journal of Animal Science. 1995; 73(11):3384-91.

48. Bohn L, Meyer AS, Rasmussen SK. Phytate: impact on environment and human nutrition. A challenge for molecular breeding. Journal of Zhejiang University Science B. 2008; 9(3):165-91.

49. Brown K, Solomons N. Nutritional problems of developing countries. Infectious disease clinics of North America. 1991; 5(2):297-317.

50. Cromwell GL. Overview of nutritional and environmental benefits of phytases. Journal of Animal Science. 2002; 80(Suppl 1):54.

51. Cromwell G, Traylor S, White L, Zavier E, Lindemann M, Sauber T, et al. Effects of low-phytate corn and low-oligosaccharide, low-phytate soybean meal in diets on performance, bone traits, and phosphorus excretion by growing pigs. Journal of Animal Science. 2000; 78:72.

52. Hurrell RF. Influence of vegetable protein sources on trace element and mineral bioavailability. The Journal of Nutrition. 2003; 133(9):2973S-7S.

53. Mendoza C. Effect of genetically modified low phytic acid plants on mineral absorption. International Journal of Food Science & Technology. 2002; 37(7):759-67.

54. Spencer J, Allee G, Frank J, Sauber T. Relative phosphorus availability and retention of low-phytate/low-oligosaccharide soybean meals for growing pigs and chicks. Journal of Animal Science. 2000; 78(Suppl 2):72.

55. Bouis HE. Improving human nutrition through agriculture: the role of international agricultural research. Conference summary and recommendations. Food and Nutrition Bulletin. 2000; 21(4):550-67.

56. Ravindran V. Phytates: occurrence, bioavailability and implications in poultry nutrition. Poultry and Avian Biology Reviews. 1995; 6:125-43.

57. Raboy V. Approaches and challenges to engineering seed phytate and total phosphorus. Plant Science. 2009; 177(4):281-96.

58. Cromwell G, Coffey R. Phosphorus-a key essential nutrient, yet a possible major pollutant-its central role in animal nutrition. Biotechnology in the Feed Industry. 1991:133-45.

59. Daverede I, Kravchenko A, Hoeft R, Nafziger E, Bullock D, Warren J, et al. Phosphorus runoff from incorporated and surface-applied liquid swine manure and phosphorus fertilizer. Journal of Environmental Quality. 2004; 33(4):1535-44.

60. Sharpley AN, Chapra S, Wedepohl R, Sims J, Daniel TC, Reddy K. Managing agricultural phosphorus for protection of surface waters: Issues and options. Journal of Environmental Quality. 1994; 23(3):437-51.

61. Raboy V, Young KA, Dorsch JA, Cook A. Genetics and breeding of seed phosphorus and phytic acid. Journal of Plant Physiology. 2001; 158(4):489-97.

62. Loewus FA, Murthy PP. myo-Inositol metabolism in plants. Plant Science. 2000; 150(1):1-19.

63. Raboy V. myo-Inositol-1, 2, 3, 4, 5, 6-hexakisphosphate. Phytochemistry. 2003; 64(6):1033-43.

64. Stevenson-Paulik J, Bastidas RJ, Chiou S-T, Frye RA, York JD. Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. Proceedings of the National Academy of Sciences. 2005; 102(35):12612-7.

65. Raboy V. Seed phosphorus and the development of low-phytate crops. In: Turner BL, Richardson AE, Mullaney EJ, editors. Inositol phosphates: Linking agriculture and the environment. Oxfordshire, UK: CAB International; 2006. p. 111-132.

66. Stephens L, Irvine R. Stepwise phosphorylation of myo-inositol leading to myo-inositol hexakisphosphate in Dictyostelium. Nature. 1990; 346(6284):580-3.

67. York JD, Odom AR, Murphy R, Ives EB, Wente SR. A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. Science. 1999; 285(5424):96-100.

68. Brearley CA, Hanke DE. Metabolic evidence for the order of addition of individual phosphate esters in the myo-inositol moiety of inositol hexakisphosphate in the duckweed Spirodela polyrhiza L. Biochemical Journal. 1996; 314(1):227-33.

69. Lott J. Accumulation of seed reserves of phosphorus and other minerals. Seed Physiology. 1984; 1:139-50.

70. Shi J, Wang H, Schellin K, Li B, Faller M, Stoop JM, et al. Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nature Biotechnology. 2007; 25(8):930-7.

71. Bentsink L, Yuan K, Koornneef M, Vreugdenhil D. The genetics of phytate and phosphate accumulation in seeds and leaves of Arabidopsis thaliana, using natural variation. Theoretical and Applied Genetics. 2003; 106(7):1234-43.

72. O'Dell BL, De Boland AR, Koirtyohann SR. Distribution of phytate and nutritionally important elements among the morphological components of cereal grains. Journal of Agricultural and Food Chemistry. 1972; 20(3):718-23.

73. Raboy V, Dickinson DB. The timing and rate of phytic acid accumulation in developing soybean seeds. Plant Physiology. 1987; 85(3):841-4.

74. Ogawa M, Tanaka K, Kasai Z. Phytic acid formation in dissected ripening rice grains. Agricultural and Biological Chemistry. 1979; 43(10):2211-3.

75. Laussmann T, Pikzack C, Thiel U, Mayr GW, Vogel G. Diphospho-myo-inositol phosphates during the life cycle of Dictyostelium and Polysphondylium. European Journal of Biochemistry. 2000; 267(8):2447-51.

76. Shears SB. Assessing the omnipotence of inositol hexakisphosphate. Cellular Signalling. 2001; 13(3):151-8.

77. Lemtiri-Chlieh F, MacRobbie EA, Brearley CA. Inositol hexakisphosphate is a physiological signal regulating the K+-inward rectifying conductance in guard cells. Proceedings of the National Academy of Sciences. 2000; 97(15):8687-92.

78. Donahue JL, Alford SR, Torabinejad J, Kerwin RE, Nourbakhsh A, Ray WK, et al. The Arabidopsis thaliana myo-inositol 1-phosphate synthase1 gene is required for myo-inositol synthesis and suppression of cell death. The Plant Cell. 2010; 22(3):888-903.

79. Shen X, Xiao H, Ranallo R, Wu W-H, Wu C. Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. Science. 2003; 299(5603):112-4.

80. Safrany ST, Caffrey JJ, Yang X, Shears SB. Diphosphoinositol polyphosphates: the final frontier for inositide research? Biological Chemistry. 1999; 380(7-8):945-51.

81. Murphy AM, Otto B, Brearley CA, Carr JP, Hanke DE. A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens. The Plant Journal. 2008; 56(4):638-52.

82. Graf E, Empson KL, Eaton JW. Phytic acid. A natural antioxidant. Journal of Biological Chemistry. 1987; 262(24):11647-50.

83. Hitz WD, Carlson TJ, Kerr PS, Sebastian SA. Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiology. 2002; 128(2):650-60.

84. Shi J, Wang H, Hazebroek J, Ertl DS, Harp T. The maize low-phytic acid 3 encodes a myo-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. The Plant Journal. 2005; 42(5):708-19.

85. Shi J, Wang H, Wu Y, Hazebroek J, Meeley RB, Ertl DS. The maize low-phytic acid mutant lpa2 is caused by mutation in an inositol phosphate kinase gene. Plant Physiology. 2003; 131(2):507-15.

86. Larson S, Young K, Cook A, Blake T, Raboy V. Linkage mapping of two mutations that reduce phytic acid content of barley grain. Theoretical and Applied Genetics. 1998; 97(1-2):141-6.

87. Larson SR, Rutger JN, Young KA, Raboy V. Isolation and genetic mapping of a non-lethal rice (Oryza sativa L.) low phytic acid 1 mutation. Crop Science. 2000; 40(5):1397-405.

88. Raboy V, Gerbasi PF, Young KA, Stoneberg SD, Pickett SG, Bauman AT, et al. Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiology. 2000; 124(1):355-68.

89. Wilcox JR, Premachandra GS, Young KA, Raboy V. Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Science. 2000; 40(6):1601-5.

90. Cichy K, Raboy V. Evaluation and development of low-phytate crops. In: Krishnan H, editor. Modification of Seed Composition to Promote Health and Nutrition, Agronomy Monograph 51. American Society of Agronomy and Crop Science Society of America; 2008. p. 177-200.

91. Hambidge KM, Huffer JW, Raboy V, Grunwald GK, Westcott JL, Sian L, et al. Zinc absorption from low-phytate hybrids of maize and their wild-type isohybrids. The American Journal of Clinical Nutrition. 2004; 79(6):1053-9.

92. Hambidge KM, Krebs NF, Westcott JL, Sian L, Miller LV, Peterson KL, et al. Absorption of calcium from tortilla meals prepared from low-phytate maize–. The American Journal of Clinical Nutrition. 2005; 82(1):84-7.

93. Bregitzer P, Raboy V. Effects of four independent low-phytate mutations on barley agronomic performance. Crop Science. 2006; 46(3):1318-22.

94. Meis SJ, Fehr WR, Schnebly SR. Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. Crop Science. 2003; 43(4):1336-9.

95. Oltmans SE, Fehr WR, Welke GA, Raboy V, Peterson KL. Agronomic and Seed Traits of Soybean Lines with Low–Phytate Phosphorus. Crop Science. 2005; 45(2):593-8.

96. Anderson BP, Fehr WR. Seed source affects field emergence of low-phytate soybean lines. Crop Science. 2008; 48(3):929-32.

97. Walker D, Scaboo A, Pantalone V, Wilcox J, Boerma H. Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Science. 2006; 46(1):390-7.

98. Oltmans SE, Fehr WR, Welke GA, Cianzio SR. Inheritance of low-phytate phosphorus in soybean. Crop Science. 2004; 44(2):433-5.

99. Gillman JD, Pantalone VR, Bilyeu K. The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome. 2009; 2(2):179-90.

100. Saghai Maroof MA, Glover NM, Biyashev RM, Buss GR, Grabau EA. Genetic basis of the low-phytate trait in the soybean line CX1834. Crop Science. 2009; 49(1):69-76.

101. Saghai Maroof MA, Buss GR. Low phytic acid, low stachyose, high sucrose soybean lines. Google Patents; 2011.

102. Smith T, Camper H. Registration of Essex soybean (reg. no. 97). Crop Science. 1973; 13(4):495.

103. Gibbs AJ, Trueman J, Gibbs MJ. The bean common mosaic virus lineage of potyviruses: where did it arise and when? Archives of Virology. 2008; 153(12):2177-87.

104. Hajimorad M, Domier LL, Tolin S, Whitham S, Saghai Maroof MA. Soybean mosaic virus: a successful potyvirus with a wide distribution but restricted natural host range. Molecular Plant Pathology. 2018; 19(7):1563-79.

105. Hartman GL, Rupe JC, Sikora EJ, Domier LL, Davis JA, Steffey KL, editors. Compendium of Soybean Diseases and Pests. St. Paul, MN: American Phytopathological Society; 2015.

106. Goodman R, Oard J. Seed transmission and yield losses in tropical soybeans infected by

soybean mosaic virus. Plant Disease. 1980; 64(10):913-4.

107. Hill J, Bailey T, Benner H, Tachibana H, Durand D. Soybean mosaic virus: Effects of primary disease incidence on yield and seed quality. Plant Disease. 1987; 71:237–239.

108. Pfeiffer TW, Peyyala R, Ren Q, Ghabrial SA. Increased soybean pubescence density. Crop Science. 2003; 43(6):2071-6.

109. Ren Q, Pfeiffer T, Ghabrial S. Soybean mosaic virus incidence level and infection time: interaction effects on soybean. Crop Science. 1997; 37(6):1706-11.

110. Ren Q, Pfeiffer G, Ghabrial S. Soybean mosaic virus resistance improves productivity of double-cropped soybean. Crop Science. 1997;37(6):1712-8.

111. Song YP, Li C, Zhao L, Karthikeyan A, Li N, Li K, et al. Disease spread of a popular Soybean mosaic virus strain (SC7) in Southern China and effects on two susceptible soybean cultivars. Formerly The Philippine Agriculturist. 2016; 99(4):355-384.

112. To J. Effect of different strains of soybean mosaic virus on growth, maturity, yield, seed mottling and seed transmission in several soybean cultivars. Journal of Phytopathology. 1989; 126(3):231-6.

113. Cho E-K, Goodman RM. Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars. Phytopathology. 1979; 69(5):467-70.

114. Cho E, Choi S, Cho W. Newly recognized soybean mosaic virus mutants and sources of resistance in soybeans. Res. Rep. ORD (SPMU). 1983; 25(10):18-22.

115. Cho E, Chung K. Strains of soybean mosaic virus causing soybean necrotic disease in Korea. Korean Journal of Breeding. 1986; 18(2):150-153.

116. Kim Y-H, Kim O-S, Lee B-C, Moon J-K, Lee S-C, Lee J-Y. G7H, a new Soybean mosaic virus strain: its virulence and nucleotide sequence of CI gene. Plant Disease. 2003; 87(11):1372-5.

117. Cho E, Chung B, Lee SH. Studies on identification and classification of soybean virus diseases in Korea. II. Etiology of a necrotic disease of Glycine max. Plant Disease Reporter. 1976; 61(4):313-7.

118. Saruta M, Kikuchi A, Okabe A, Sasaya T. Molecular characterization of A 2 and D strains of Soybean mosaic virus, which caused a recent virus outbreak in soybean cultivar Sachiyutaka in Chugoku and Shikoku regions of Japan. Journal of General Plant Pathology. 2005; 71(6):431-5.

119. Li K, Yang Q, Zhi H, Gai J. Identification and distribution of soybean mosaic virus strains in southern China. Plant Disease. 2010; 94(3):351-7.

120. Ma F-F, Wu X-Y, Chen Y-X, Liu Y-N, Shao Z-Q, Wu P, et al. Fine mapping of the Rsv1-h gene in the soybean cultivar Suweon 97 that confers resistance to two Chinese strains of the soybean mosaic virus. Theoretical and Applied Genetics. 2016; 129(11):2227-36.

121. Wang XQ, Gai JY, Pu ZQ. Classification and distribution of strains of Soybean mosaic virus middle and lower Huang-Huai and Changjiang Valleys. Soybean Science. 2003; 22:102-7.

122. Jayaram C, Hill JH, Miller WA. Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the Rsv resistance gene. Journal of General Virology. 1992; 73(8):2067-77.

123. Berger P, Adams M, Barnett O, Brunt A, Hammond J, Hill J, et al. Potyviridae In: Virus taxonomy: classification and nomenclature of viruses. Eighth report of the International Committee on the Taxonomy of Viruses. C. Fauquet ed. San Diego, California: Elsevier Academic Press; 2005. p. 819e-41e.

124. Chung BY-W, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the Potyviridae. Proceedings of the National Academy of Sciences. 2008; 105(15):5897-902.

125. Cui X, Chen X, Wang A. Detection, Understanding and Control of Soybean Mosaic Virus. In: Sudaric A, editor. Soybean: Molecular Aspects of Breeding. Rijeka, Croatia:IntechOpen; 2011. p. 335-54.

126. Dangl JL, Horvath DM, Staskawicz BJ. Pivoting the plant immune system from dissection to deployment. Science. 2013; 341(6147):746-51.

127. Flor HH. Current status of the gene-for-gene concept. Annual Review of Phytopathology. 1971; 9(1):275-96.

128. Hayes AJ, Ma G, Buss GR, Saghai Maroof MA. Molecular marker mapping of Rsv 4, a gene conferring resistance to all known strains of soybean mosaic virus. Crop Science. 2000; 40(5):1434-7.

129. Kiihl RA, Hartwig E. Inheritance of Reaction to Soybean Mosaic Virus in Soybeans 1. Crop Science. 1979; 19(3):372-5.

130. Buzzell R, Tu J. Inheritance of a soybean stem-tip necrosis reaction to soybean mosaic virus. Journal of Heredity. 1989; 80(5):400-1.

131. Buss G, Ma G, Kristipati S, Chen P, Tolin S, editors. A new allele at the Rsv3 locus for resistance to soybean mosaic virus. Proc World Soybean Res Conf VI, Chicago, IL; 1999.

132. Buss G, Ma G, Chen P, Tolin S. Registration of V94-5152 soybean germplasm resistant to soybean mosaic potyvirus. Crop Science. 1997; 37(6):1987-8.

133. Saghai Maroof MA, Tucker DM, Tolin SA. Genomics of viral–soybean interactions. In: Stacey, G, editor. Genetics and Genomics of Soybean. New York, NY: Springer; 2008. p. 293-319.

134. Hayes A, Jeong S, Gore M, Yu Y, Buss G, Tolin S, et al. Recombination within a nucleotide-binding-site/leucine-rich-repeat gene cluster produces new variants conditioning resistance to soybean mosaic virus in soybeans. Genetics. 2004; 166(1):493-503.

135. Chen P, Buss G, Roane C, Tolin S. Inheritance in soybean of resistant and necrotic reactions to soybean mosaic virus strains. Crop Science. 1994; 34(2):414-22.

136. Bendahmane A, Kanyuka K, Baulcombe D. High-resolution genetical and physical mapping of the Rx gene for extreme resistance to potato virus X in tetraploid potato. Theoretical and Applied Genetics. 1997; 95(1-2):153-62.

137. Hämäläinen J, Watanabe K, Valkonen J, Arihara A, Plaisted R, Pehu E, et al. Mapping and marker-assisted selection for a gene for extreme resistance to potato virus Y. Theoretical and Applied Genetics. 1997; 94(2):192-7.

138. Ilut DC, Lipka AE, Jeong N, Bae DN, Kim DH, Kim JH, et al. Identification of haplotypes at the Rsv4 genomic region in soybean associated with durable resistance to soybean mosaic virus. Theoretical and Applied Genetics. 2016; 129(3):453-68.

139. Saghai Maroof MA, Tucker DM, Skoneczka JA, Bowman BC, Tripathy S, Tolin SA. Fine mapping and candidate gene discovery of the soybean mosaic virus resistance gene, Rsv4. The Plant Genome. 2010; 3(1):14-22.

140. Ma G, Chen P, Buss G, Tolin S. Genetic characteristics of two genes for resistance to soybean mosaic virus in PI486355 soybean. Theoretical and Applied Genetics. 1995; 91(6-7):907-14.

141. Suh SJ, Bowman BC, Jeong N, Yang K, Kastl C, Tolin SA, et al. The Rsv3 locus conferring resistance to soybean mosaic virus is associated with a cluster of coiled-coil nucleotide-binding leucine-rich repeat genes. The Plant Genome. 2011; 4(1):55-64.

142. Jeong S, Kristipati S, Hayes A, Maughan P, Noffsinger S, Gunduz I, et al. Genetic and sequence analysis of markers tightly linked to the soybean mosaic virus resistance gene, Rsv 3. Crop Science. 2002; 42(1):265-70.

143. Redekar N, Clevinger E, Laskar M, Biyashev R, Ashfield T, Jensen R, et al. Candidate gene sequence analyses toward identifying Rsv3-type resistance to soybean mosaic virus. The Plant Genome. 2016; 9(2):1-12.

144. Tran P-T, Widyasari K, Seo J-K, Kim K-H. Isolation and validation of a candidate Rsv3 gene from a soybean genotype that confers strain-specific resistance to soybean mosaic virus. Virology. 2018; 513:153-9.

145. Gunduz I, Buss G, Chen P, Tolin S. Characterization of SMV resistance genes in Tousan 140 and Hourei soybean. Crop Science. 2002; 42(1):90-5.

146. Seo J-K, Lee S-H, Kim K-H. Strain-specific cylindrical inclusion protein of Soybean mosaic virus elicits extreme resistance and a lethal systemic hypersensitive response in two resistant soybean cultivars. Molecular Plant-Microbe Interactions. 2009; 22(9):1151-9.

147. Seo J-K, Kwon S-J, Cho WK, Choi H-S, Kim K-H. Type 2C protein phosphatase is a key regulator of antiviral extreme resistance limiting virus spread. Scientific Reports. 2014; 4:5905-12.

148. Alazem M, Tseng K-C, Chang W-C, Seo J-K, Kim K-H. Elements Involved in the Rsv3-Mediated Extreme Resistance against an Avirulent Strain of Soybean Mosaic Virus. Viruses. 2018; 10(11):581-96.

149. Bendahmane A, Kanyuka K, Baulcombe DC. The Rx gene from potato controls separate virus resistance and cell death responses. The Plant Cell. 1999; 11(5):781-91.

CHAPTER 2

Network Inference of Transcriptional Regulation in Germinating Low Phytic Acid Soybean Seeds

Lindsay C DeMers[1], Victor Raboy[2], Song Li[1], MA Saghai Maroof[1*]

[1]School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia, United States of America. [2]National Small Grain Germplasm Research Center, Agricultural Research Service (USDA), Aberdeen, Idaho, United States of America. [*]Corresponding author: smaroof@vt.edu

This chapter is to be submitted for publication in *Frontiers in Plant Science*.

ABSTRACT

The low phytic acid trait in soybeans can be conferred by loss-of-function mutations in genes encoding *myo*-inositol phosphate synthase and two epistatically interacting genes encoding multidrug-resistance protein ABC transporters. However, perturbations in phytic acid biosynthesis are associated with poor seed vigor. Since the benefits of the low phytic acid trait, in terms of end-use quality and sustainability, far outweigh the negatives associated with poor seed performance, a fuller understanding of the molecular basis behind the negatives will assist crop breeders and engineers to successfully deal with them. The gene regulatory network for developing low and normal phytic acid soybean seeds was previously constructed and inferred with genes modulating a variety of processes pertinent to phytic acid metabolism and seed viability being identified. In this study, a comparative time series analysis of low and normal phytic acid soybeans was carried out to investigate the transcriptional regulatory elements governing the transitional dynamics from dry seed to germinated seed. Gene regulatory networks were reverse engineered from time series transcriptomic data of three distinct genotypic subsets composed of low phytic acid soybean lines and their normal phytic acid sibling lines. Using a robust unsupervised network inference scheme, putative regulatory interactions were inferred for each subset of genotypes. These interactions were further validated by published regulatory interactions found in *Arabidopsis thaliana* and motif sequence analysis. Results indicate that low phytic acid seeds have increased sensitivity to stress, which could be due to changes in phytic acid levels, disrupted phosphate ion homeostasis, and altered *myo*-inositol metabolism. Putative regulatory interactions were identified for the latter two processes. Changes in abscisic acid signaling candidate transcription factors putatively regulating genes in this process were identified as well. Analysis of the gene regulatory networks reveal altered regulation in processes

that may be affecting the germination of low phytic acid soybean seeds. Therefore this work contributes to the ongoing effort to elucidate molecular mechanisms underlying altered seed viability, germination and field emergence of low phytic acid crops, understanding of which is necessary in order to mitigate these problems.

KEYWORDS

Phytic acid, *myo*-inositol phosphate synthase, multidrug-resistance protein ABC transporter, seed germination, transcriptomics, gene regulatory network, unsupervised machine learning, abscisic acid signaling, phosphate homeostasis

INTRODUCTION

The development and commercialization of low phytic acid (*lpa*) crops could represent one approach to enhanced management of phosphorus (P) in animal agriculture and to addressing mineral deficiency in humans. Seed phytic acid (*myo*-inositol-(1,2,3,4,5,6)-hexa*kis*phosphate) represents about 75% of seed total P. In the intestinal tract of non-ruminant animals seed-derived dietary phytic acid chelates divalent cations, and the resulting salts are excreted. This can contribute to mineral deficiencies in monogastric animals and leads to high levels of excreted phosphorus which pollute water systems [1-6]. *Lpa* barley, maize, rice, soybean, and wheat lines have been developed, and their seeds are shown to increase phosphorus availability in poultry and swine and reduce phosphorus pollution from the subsequent waste [7-13]. Despite these advantages, *lpa* crops have not been commercialized, as they often exhibit poor seed and seedling vigor, low stress tolerance, and reduced germination and emergence rates

[14-17]. Therefore, a fuller understanding of the molecular basis behind these negatives will assist crop breeders and engineers in successfully handling them.

Some *lpa* crops, such as barley *lpa1-1* and common bean *lpa*-280-10, exhibit good seed emergence and yield, demonstrating that development of *lpa* crops without adverse agronomic effects is possible [16, 18]. Furthermore, selection within an *lpa* line might yield progeny with improved germination and field emergence [19]. For example, a soybean *lpa* mutation termed TW-1 had reduced field emergence and reduced viability following seed storage; however, these negative effects were reduced in a single-plant derived line isolated in TW-1 progeny termed TW-1-M [20, 21]. Limiting the wide-scale development of high performing *lpa* crops is a poor understanding of the molecular basis of seed phytic acid content in relation to seed vigor, *i.e.* the properties defining a seed's potential performance during germination and emergence [22]. Previous studies with barley and soybean have investigated the effect of the *lpa* trait on developing seeds and found differences in energy metabolism and phytohormone signaling, as well as regulatory components that may be responsible for these variations [23-25]. Transcriptomic and proteomic analyses of germinating seeds were used to understand the molecular basis of the improvement of field emergence and seed viability observed in the soybean mutant TW-1-M as compared with its parental line TW-1 [20, 21]. These studies revealed changes in gene transcripts and proteins involved in energy metabolism, phytohormone pathways, oxidation-reduction processes, and stress responses [20, 21].

Because seed germination is recognized as the most vulnerable period in a plant's life cycle [26], it is important to understand how the process of germination is regulated. During germination, seeds undergo a massive metabolic transition in order to prepare for seedling growth; this is a highly coordinated and complex process, involving regulatory control over

cellular and metabolic events [26-29]. Primary factors found to mediate germination include metabolism, phytohormone signaling, signal transduction components, and notably, transcription factors (TFs) [26, 30, 31]. Thus it is important to consider the influence of regulatory interactions between genes, as a systematic understanding of the processes governing germination can offer insights into seed and seedling vigor. Advancements in high-throughput technology, such as RNA sequencing (RNA-seq), enable the collection and analysis of genome-wide expression data at a systems level [32]. These data can then be used to construct a gene regulatory network (GRN), a graphical or mathematical representation of the causal relationships between genes regulating cellular functions in an organism [33-35]. A GRN's connections, representing interactions between genes, are established by implementing inference methods on transcriptomic data. Hence inferred GRNs consist of computationally predicted directed interactions between TFs and target genes, making GRNs an effective tool for identifying key regulatory and target genes involved in specific biological processes [36, 37].

Previously, a GRN analysis was performed to understand how the *lpa* trait may be affecting seed vigor by comparing the transcriptomes of developing seeds in low and normal phytic acid soybeans [25]. Differences were found in metabolism, defense responses, phytohormone signaling, and candidate TFs putatively regulating some of these processes. However, to construct a more complete profile of low and normal phytic acid seed transcriptomes and to identify differences between their regulatory networks, this study examines RNA-seq data from germinating seeds of low and normal phytic acid soybeans and infers multiple GRNs. The findings offer new information on the transcriptional regulation of germinating *lpa* soybean seeds and the perturbed biological processes in *lpa* seeds that may be important to successful germination.

MATERIALS AND METHODS

Genetic material

In this study, eight experimental lines were used – *1mpa*, 1MWT, *2mlpa*, 2MWT, 2MWT-L, 2MWT-N, *3mlpa*, and 3MWT (Table 2.1). The three *lpa* lines (*1mlpa*, *2mlpa*, and *3mlpa*) contain one, two, or three mutations, respectively, in genes functioning in the phytic acid pathway. These genes are MIPS1 (Glyma.11G238800), encoding *myo*-inositol-3-monophosphate synthase (MIPS), which synthesizes phytic acid, and MRP-L (Glyma.19G169000) and MRP–N (Glyma.03G167800), encoding multidrug resistance-associated protein ATP-binding cassette (ABC) transporters (henceforth called MRPs), which transport phytic acid for storage [38]. The eight experimental lines represent three distinct subsets of genotypes. The first genotypic subset, designated as the "Mips" subset, contains the *lpa* line "*1mlpa*" (*mips1* mutation) and the normal phytic acid line "1MWT" (no mutation). These lines are isogenic and were developed from a cross between the normal phytic acid line "Essex'" (no MIPS1 mutation) and the *lpa* line "V99-5089" (*mips1* mutation) [39]. This *mips1* mutation conferring the *lpa* trait is the result of a point mutation on chromosome 11 [39]. The second genotypic subset, designated as "MRP," contains four near isogenic lines, the *lpa* line "2mlpa" (*mrp-l* and *mrp-n* mutations) and three normal phytic acid lines, "2MWT" (no mutation), "2MWT-L" (*mrp-n* mutation only), and "2MWT-N" (*mrp-l* mutation only). These lines were developed from a cross between the *lpa* lines "CX-1834" (*mrp-1* and *mrp-n* mutations) and the normal phytic acid line V99-5089. The mutations conferring the *lpa* trait in *2mlpa* and CX-1834 are the result of point mutations in the epistatically interacting loci, MRP-L and MRP-N, on chromosomes 19 and 3, respectively [10, 38, 40]. Lastly, the final genotypic subset, designated as "Mips-MRP," is composed of the *lpa* line "*3mlpa*" (*mips1*, *mrp-1*, and *mrp-n* mutations) and the normal phytic acid line "3MWT" (no

mutation). These lines were developed from a cross between CX-1834 and V99-5089. Seeds from all eight lines were harvested in 2017 from a field in Blacksburg, VA and stored at 4°C until experimentation.

Seed germination and sampling

Tissue from each line was sampled at three stages of seed germination in biological triplicate with ten seeds per sample. The germination stages used were mature dry seeds (stage 1), eight hour imbibed seeds (stage 2), and germinated seeds (defined as radicle emergence; stage 3). For stage 1, seeds were ground to a fine powder using a P14 mill (Pulverisette 14, Fritsch) and stored at -80°C until use. Seeds for stages 2 and 3 were sterilized for two minutes with a 10% hypochlorite + Tris solution, washed in DI water three times for five minutes, and dried overnight. The seeds were then germinated on germination plates with filter paper and DI water in the dark at 29°C. Once the appropriate stage was reached, seed coats and radicles were removed, and the tissue was flash frozen with liquid nitrogen and stored at -80°C until use. The tissue from these stages was ground to powder with mortar, pestle, and liquid nitrogen. Total RNA from all stages was extracted using the RNeasy Plant Kit with on-column DNase digestion and RLC buffer (QIAGEN, Hilden, Germany). RNA quality was determined by UV spectrophotometry (260 nm, NanoDrop1000, Thermo Fischer Scientific, Waltham, MA) and RNA integrity numbers (RIN) (BioAnalyzer, Agilent Technologies, Santa Clara, CA). Samples with a RIN value >8.0 and 260/280 ratios >2.0 were submitted to Novogene (Sacramento, CA) for mRNA sequencing. A total of 72 samples (8 lines x 3 germination stages x 3 biological replicates) were sequenced with HiSeq4000 (Illumina, San Diego, CA) to acquire 30 million, 150 PE reads per sample.

Transcriptomics data processing and differential gene expression

Raw reads were trimmed and filtered using Skewer (version 0.2.2) to remove adapter sequences and low quality reads and bases (<Q30) [41]. Using STAR (version 2.5.2b), the cleaned reads were aligned to 'Williams82,' the well-annotated soybean reference genome (Wm82.a2.v1, downloaded from Phytozome) [42, 43]. Transcript abundances were calculated from the mapping results using featureCounts (version 1.5.1) [44]. These results were subsequently used for differential expression analysis with DESeq2 (version 1.22.2) in R (version 3.5.1) [45]. Comparisons were made to identify differentially expressed genes (DEGs) between *lpa* and normal phytic acid lines at each stage within each subset of genotypes. DEGs were defined as those with false discovery rate (FDR)-adjusted p-value < 0.01, $\log_2$ fold change >|1.0|, and base mean >10. The DEGs between *lpa* and normal phytic acid lines at each stage within each subset of genotypes can be found in Table S2.1. The RNA-seq data from this study will be made available at the NCBI Gene Expression Omnibus (GEO) repository.

Transcriptional network construction and inference

For each subset of genotypes, gene expression levels were normalized for all genes using variance-stabilizing transformation in DESeq2 [45]. The normalized expression was averaged across the three replicates, and then the averaged expression of the DEGs was used for clustering. Clustering analysis for each genotypic subset was performed independently. DEGs from each subset were clustered using Gaussian-finite mixture modeling with the R package, mclust (version 5.4.2), and the best performing models were determined using Bayesian Information Criteria (BIC) [46, 47]. For both the Mips and Mips-MRP genotypic subsets, nine clusters were found, and five clusters were found for the MRP genotypic subset. Gene ontology

(GO) enrichment analysis was performed on each gene cluster using GO annotations obtained from Soybase [48]. Significantly enriched GO categories were identified using Fisher's exact test with FDR <0.05 (Table S2.2) [49]. DEGs encoding TFs were annotated with the plant TFDB [50].

For this study, separate network inferences were performed on the three subsets of genotypes using a computational pipeline developed previously [25, 51]. The pipeline implements the module network approach [52], in which genes are clustered into co-expression modules (gene modules) and then gene regulation is inferred between TFs and gene modules. The pipeline also incorporates multiple inference methods for improved robustness. In the Mips subset, 489 differentially expressed TFs were identified. In the MRP subset, 24 differentially expressed TFs were identified, while 340 were identified in the Mips-MRP subset. These TFs were used as putative regulators of the gene co-expression modules. Gene expression was averaged for each module, and the values, along with the TF expression values, were used to build an expression matrix. This matrix was then used to infer putative regulatory interactions between TFs and modules by applying five distinct network inference algorithms: ARACNE, CLR, LARS, partial correlation, and Random Forest [53-57]. These algorithms represent the top-performing, unsupervised network inference methods, according to the DREAM5 challenge found in the benchmark paper by Marbach et al. [58].

Validation of network inferred interactions

The interactions predicted by the five network inference methods were validated by comparison to published interactions observed in *Arabidopsis* using DAP-seq and motif sequence analysis. For validation against the published *Arabidopsis* interactions, each network

inferred TF-module interaction was expanded to TF-gene interactions by matching the TF putatively regulating the module to all genes assigned to the module. The soybean TF-gene interactions were converted to homologous *Arabidopsis* interactions by identifying homologous *Arabidopsis* genes for soybean gene coding sequences. Using BLAST with an E-value threshold of 1e-5, the top *Arabidopsis* gene hit was selected. The resulting homologous *Arabidopsis* interactions were then compared to the published DAP-seq interactions to identify matches. For validation by motif sequence analysis, the motif discovery tool, MEME, from Meme Suite (version 5.0.4) was used to identify enriched motif sequences among the genes in each module using the 1000 bps flanking those genes' 5' end. The enriched motif sequences were then compared to motif sequences found in *Arabidopsis* with DAP-seq by employing the TomTom tool from Meme Suite (version 5.0.4). This allowed for the identification of TFs that may recognize and bind the discovered motifs in each module. The identified TFs were then compared to the TFs predicted to be module regulators.

RESULTS

In this study, RNA-seq was carried out on germinating seeds from low and normal phytic acid soybeans, and GRNs were inferred to identify disparities in transcriptional regulation. A total of eight experimental soybean lines from three genotypic class subsets were used, with each subset containing at least one normal phytic acid line and one *lpa* line. The *lpa* trait was conferred through various combinations of mutant MIPS1, MRP-L, and MRP-N genes.

Differential expression analysis

For each of the three subsets of genotypes (Table 2.1), the *lpa* and normal phytic acid lines were compared to identify genes that were differentially expressed at each stage. In the Mips subset, 5,841 DEGs were found between *1mlpa* (*lpa*) and 1MWT (normal phytic acid). Using a set of four near isogenic lines, the number of DEGs in the MRP subset was limited to just 430. This number was obtained by designating genes as DEGs only if they were differentially expressed in each comparison of *2mlpa* (*lpa*) to the three normal phytic acid lines (2MWT, 2MWT-L, 2MWT-N). Finally, in the Mips-MRP subset, 4,512 DEGs were found between *3mlpa* (*lpa*) and 3MWT (normal phytic acid). The DEGs for each genotypic subset can be found in Table S2.1. For each subset, few genes were differentially expressed in all three germination stages, indicating the mutations affect genes at specific stages (Figure 2.1A, 2.1B, 2.1C). In each subset, numerous genes were strictly differentially expressed at stage 1 - 40% in the Mips subset, 37% in MRP, and 48% in Mips-MRP (Figure 2.1A, 2.1B, 2.1C); this suggests both *mips1* and *mrp-1/mrp-n* mutations considerably affect genes at the dry seed stage. Not as many genes were differentially expressed in the Mips and Mip-MRP subsets at stage 2, indicating the *mips1* mutation may not impact genes as much in imbibed seeds. However, a substantial number were differentially expressed at stage 3 in germinated seeds (Figure 2.1A, 2.1C). Conversely, in the MRP subset, many genes were differentially expressed at stage 2, but few were differentially expressed at stage 3 (Figure 2.1B). This, along with the DEGs at germination stage 1, suggests the two *mrp* mutations have a greater effect on genes during the early stages of germination, at the dry and imbibed seed stages, when metabolism needs to be reinitiated.

When the DEGs from all three subsets were compared to one another, 85 genes were differentially expressed in all three genotypic subsets (Figure 2.1D). Roughly half of the DEGs in each subset remained unique to their particular subsets. However, there was a fair amount of overlap (1692 DEGs) between the Mips and Mips-MRP subsets, which may be a result of shared perturbations in the *myo*-inositol synthesis pathway due to the *mips1* mutation. Less than 13% of the DEGs in the MRP subset were shared with those in the Mips-MRP subset, despite both of the subsets carrying the *mrp-l*/*mrp-n* mutations.

Both the Mips and Mips-MRP subsets had differential expression in genes functioning in the phytic acid biosynthesis pathway (Table 2.2). In each subset, Glyma.11G218500 and Glyma.18G038800, both encoding inositol 1,3,4-trisphosphate 5/6-kinase 4 (ITPK4), had increased expression in the *lpa* lines, *1mlpa* and *3mlpa*, predominantly in germination stage 1, the dry seed stage. In *1mlpa*, increased expression was also observed in two genes encoding inositol 1,3,4-trisphosphate 5/6-kinase 1 (ITPK1) and Glyma.11G238800, which had increased expression in all three germination stages and encodes *myo*-inositol-1-phosphate synthase 2 (MIPS2).

Co-expression analyses reveal altered phosphate ion homeostasis activity and stress responses in lpa lines

To compare the transcriptional regulation governing the dynamics of seed germination in *lpa* and normal phytic acid lines, gene co-expression modules were generated by individually clustering the set of DEGs found in each subset of genotypes. The co-expression modules, defined as sets of genes with similar temporal expression patterns, were created using a model-based clustering approach and BIC criterion. In the Mips subset, nine co-expression modules

were found, five co-expression modules were found in the MRP subset, and nine were found in the Mips-MRP subset (Table S2.1). GO analysis was carried out on each co-expression module for each genotypic subset. For the modules in the Mips subset, 372 instances of GO enrichment were found. In the MRP subset, 30 instances were found, and 162 were found in the Mips-MRP subset. These were narrowed down to focus on biological processes only. Enrichment for all categories (biological and molecular) can be found in Table S2.2.

For the Mips subset co-expression modules, module 2 was enriched for nucleotide biosynthesis (Figure 2.2A), which had lower expression in *1mlpa* (Figure 2.2B). Modules 3, 5, and 9 were each enriched for signaling and stress-related processes, which were especially enriched in module 5 (Figure 2.2A). In all three modules, the genes functioning in the enriched processes had higher expression in *1mlpa* (Figure 2.2B). Modules 3 and 5 also showed enrichment for genes in the ethylene (ET) and salicylic acid (SA) pathways. For the enriched processes in module 4 (Figure 2.2A), many of which are related to photosynthesis, translation, and carbohydrate metabolism, gene expression was reduced in *1mlpa* in germination stages 1 and 2 (Figure 2.2B). In association with the phytic acid pathway, module 6 was enriched for *myo*-inositol hexa*kis*phosphate biosynthesis (GO:0010264), the genes of which had increased expression in stages 1 and 2 in *1mlpa* (Figure 2.2B). Conversely, in module 7, *1mlpa* had decreased expression in stages 1 and 2 in genes functioning in phosphate ion (Pi) homeostasis (GO:0030643) (Figure 2.2B). Module 8 was enriched for genes in the abscisic acid (ABA) signaling pathway and was also strongly enriched for a number of stress-related processes, including response to heat, high light intensity, hydrogen peroxide, water deprivation, protein folding, heat acclimation, oxidative stress, and endoplasmic reticulum stress (Figure 2.2A). The genes in these processes exhibited increased expression in stages 1 and 2 in *1mlpa* (Figure 2.2B).

Each module in the MRP subset had strong gene enrichment in at least one biological process (Figure 2.3A). Figure 2.3B, an expression heatmap of the genes functioning in these processes, is especially interesting because it highlights the utility of the genetic material. That is, between the four isogenic lines, the three normal phytic acid lines (2MWT-L, 2MWT-N, 2MWT) have nearly identical expression patterns, while expression in the single *lpa* line (*2mlpa*) is unique (Figure 2.3B). Module 1 was enriched for genes in fatty acid and cutin transport activities (Figure 2.3A) and at stage 1, had higher expression in *2mlpa* than the three normal phytic acid lines (Figure 2.3B). Module 2 was solely enriched for genes in cellular Pi homeostasis (GO:0030643), the genes of which had particularly reduced expression in germination stages 2 and 3 in *2mlpa*. In module 4, enrichment was found for genes functioning in ABA stimulus response, stress response, lipid storage, and seed maturation (Figure 2.3A). The genes in these processes had increased expression in *2mlpa* in germination stages 1 and 2 (Figure 2.3B). Lastly, module 5 had enrichment for nucleotide-related processes, the genes of which had decreased expression in stages 1 and 2 in *2mlpa* (Figure 2.3A and Figure 2.3B).

For the Mips-MRP subset co-expression modules, module 1 was enriched for genes functioning in response to hypoxia and oxidative stress (Figure 2.4A), which had increased expression in *3mlpa* (*lpa*) in germination stages 2 and 3 (Figure 2.4B). Like the Mips and MRP genotypic subsets, the Mips-MRP subset had enrichment in module 3 for genes in stress responses as well as the glyoxylate cycle and phytohormone pathways involving JA and ET (Figure 2.4A). The genes in these processes had increased expression in *3mlpa* in stages 2 and 3 (Figure 2.4B). In the case of module 6 and similar to the Mips subset, strong enrichment was found for many genes in biological processes associated with photosynthesis, translation, and a number of metabolic pathways (Figure 2.4A). Like the Mips subset, gene expression for these

processes was reduced in *3mlpa* in all three stages as compared to 3MWT (Figure 2.4B). Module-7 was enriched for genes functioning in glycolipid and galactolipid biosynthesis, Pi homeostasis, and response to Pi starvation (Figure 2.4A), which had decreased expression in *3mlpa* especially in stages 2 and 3 (Figure 2.4B). A number of biological processes were also enriched in module 8, such as several stress-related responses, ABA signaling, and other metabolic pathways (Figure 2.4A). Most of which had increased expression in *3mlpa* (Figure 2.4B).

When comparing GO enrichment between the three subsets of genotypes, the one biological process that was common between all of them was cellular Pi homeostasis (GO:0030643). Because *lpa* seeds have increased Pi levels, the significant expression changes observed in Pi homeostasis genes in germinating *lpa* seeds lends support to the putative role of phytic acid biosynthesis as a means of regulating cellular Pi concentration. Just seven genes in the soybean genome are annotated as functioning in cellular Pi homeostasis. The Mips subset had four DEGs from this GO category, the MRP subset had five DEGs, and the Mips-MRP subset had four DEGs. Between the three genotypic subsets, four DEGs were shared – Glyma.05G247900, Glyma.08G056400, Glyma.16G052000, and Glyma.19G098500. Both Glyma.05G247900 and Glyma.08G056400 encode purple acid phosphatase 17 (PAP17), and Glyma.16G052000 and Glyma.19G098500 encode glycerophosphodiester phosphodiesterase (GDPD1). The two PAP17 genes and the two GDPD1 genes were down-regulated in all *lpa* lines at stage 2, and for the most part, they were also all down-regulated at stage 3.

Biological processes enriched in both developing and germinating seeds

In addition, the GO enrichment results for the Mips and Mips-MRP subsets were compared to the earlier enrichment findings in Redekar et al. [25], where RNA-seq was performed on the same four experimental lines during seed development – *1mlpa*, 1MWT, *3mlpa*, and 3MWT. In the Mips subset, 88 GO categories overlapped with the developing seed expression data, and in the Mips-MRP subset, 41 categories overlapped. Both genotypic subsets had overlap in stress-, photosynthesis-, ion-, *myo*-inositol metabolism-, and hormone-related GO categories. Interestingly, overlap was also found in the pentose-phosphate shunt pathway (GO:0006098). In this study, the Mips subset had 64 genes in this pathway that were differentially expressed, and the Mips-MRP subset had 56. In both subsets, the genes were primarily differentially expressed at the dry seed stage (stage 1). This finding is notable because the pentose-phosphate shunt pathway parallels glycolysis, generating NADPH, pentoses (5-carbon sugars), and ribose 5-phosphate (precursors for nucleotide synthesis), but does so by oxidizing glucose-6-phosphate, the same substrate used by the MIPS enzyme in the first step of phytic acid biosynthesis [59].

Gene regulatory networks

Inference of the constructed GRN detected TF-module interactions for each subset of genotypes. For each subset, the interactions were narrowed down to those detected by at least four out of the five inference methods (Table S2.3). The TF-module interactions were then expanded into TF-gene interactions and computationally validated by comparison to the published *Arabidopsis* DAP-seq interactions and motif sequence analysis [60]. For the Mips subset, this resulted in 4,572 TF-gene interactions, consisting of 31 differentially expressed TF

regulators and 2,743 differentially expressed target genes (Table S2.4). For the MRP subset, 154 TF-gene interactions were found, being regulated by five differentially expressed TF genes with 125 differentially expressed target genes (Table S2.4). As for the Mips-MRP subset, 3,757 TF-gene interactions were found, which were regulated by 31 TFs and consisted of 1,998 target genes (Table S2.4). Between the three subsets' GRNs, the one putative TF regulator found in each was DREB1F encoded by Glyma.01G216000 (Table 2.3). This gene had increased expression in all three *lpa* lines, but no putative target genes of this TF were shared by all three subsets. Nonetheless, five other TF genes were identical in the Mips and Mips-MRP GRNs, with most of them sharing some of the same putative targets (Table 2.3). Interestingly, two of these genes (Glyma.04G249000, Glyma.06G114000) encode the same TF, ATAF1, and both had increased expression in the *lpa* lines *1mlpa* and *3mlpa*. Though not shared in their networks, both *1mlpa* and *3mlpa* had an additional *ATAF1* gene with increased expression in their respective GRNs, Glyma.04G208300 in *1mpa* and Glyma.05G195000 in *3mlpa*. The changes in *ATAF1* expression in both *lpa* lines is notable as *ATAF1* is ABA-responsive and regulates ABA biosynthesis [61, 62].

The Mips and Mips-MRP GRNs shared several of the same putative regulatory interactions (Figure 2.5). Some of the shared target genes that stand out include ABA-insensitive5 (*ABI5*) (Glyma.10G071700) and multiple late embryogenesis abundant (*LEA*) genes (Glyma.07G064700, Glyma.08G239400, Glyma.09G112100, Glyma.13G363300, Glyma.16G031300, Glyma.17G040800). *ABI5* is regulated by the ABA pathway and functions to retain embryos in a dormant state [63]. According to both networks, *ABI5* is putatively regulated by ATAF1 encoded by Glyma.06G114000, and in both *1mlpa* and *3mlpa*, the expression of *ABI5* is increased (Figure 2.5). All but one (Glyma.13G363300) of the six *LEA*

genes had increased expression in *1mlpa* and *3mlpa*, and all are putatively regulated by *ATAF1* (Glyma.06G114000) (Figure 2.5). This family of proteins is ABA-induced and reduces desiccation-induced cellular damage in seed tissue [64].

As for the target genes in the MRP subset, seven were found in significant biological GO categories observed in the subset (Table 2.4). Most of these targets are seed storage proteins, which function in lipid storage, seed maturation, and responses to ABA stimulus. These seed storage proteins are putatively regulated by Glyma.05G032200 (MYB-related) and Glyma.07G060400 (bZIP), the latter of which encodes G-box binding factor 3 (GBF3).

DISCUSSION

In the three *lpa* lines used in this study, *1mlpa*, *2mlpa*, and *3mlpa*, phytic acid metabolism is disrupted to increase Pi bioavailability; however, phytic acid and the intermediate compounds in its biosynthesis have fundamental roles in various developmental, metabolic, and signaling pathways critical to plant function [5]. Consequently, blocks in this pathway appear to have numerous downstream effects.

Regulation of phosphate ion homeostasis in lpa lines

Phosphorus is an essential macronutrient vital to cellular metabolism, bioenergetics, and a core component of vital molecules, such as nucleic acids and phospholipids [65]. In order for enzymatic reactions to proceed in a normal manner, it is critical for cytoplasmic Pi concentrations to remain constant regardless of fluctuations in the external environment. However, in all three *lpa* lines, enrichment was found for genes functioning in Pi homeostasis, indicating that the blocks in the phytic acid pathway perturb cellular Pi homeostasis.

In the three *lpa* lines, Pi homeostasis enrichment was due to the same four genes – two genes encoding PAP17 and two genes encoding GDPD1. The two PAP17 and the two GDPD1 genes were down-regulated in all three *lpa* lines at germination stage 2 and for the most part, were also down-regulated at stage 3. PAP proteins are multifunctional proteins induced under Pi starvation and catalyze the hydrolysis of Pi from monoesters and anhydrides for the transport and recycling of Pi [66]. In particular, PAP17 also has peroxidation activity, functioning in the metabolism of reactive oxygen species [67]. GDPD1 hydrolyzes glycerophosphodiesters and is also induced by Pi starvation, during which it likely releases Pi from phospholipids [68]. According to regulatory network inference and motif sequence analysis, all four Pi homeostasis genes are putatively regulated by Glyma.08G092300, which encodes a C2H2 TF. This TF may in part be responsible for down-regulating Pi homeostasis genes in *lpa* lines. Down-regulation of PAP17 and GDPD1 in *lpa* seeds suggests that sufficient, if not more than sufficient, cellular Pi levels are present. The increased Pi levels in *lpa* seeds may perturb Pi homeostasis in such a way that normal cell metabolism is disrupted, inducing cellular stress and ultimately reducing seed viability. Thus, reducing PAP17 and GDPD1 expression in *lpa* seeds may be an attempt to recover Pi homeostasis.

Downstream effects of perturbed myo-inositol metabolism in mips1 mutants

Phytic acid biosynthesis requires a substrate supply of *myo*-inositol and phosphate. The sole source of *myo*-inositol comes from the activity of the enzyme MIPS synthase. Previous studies not just limited to soybean have also shown that loss-of-function mutations in MIPS1 are associated with impaired seed and plant performance [14, 69-72]. Given that *myo*-inositol synthesis via MIPS is considered a part of general housekeeping [5, 73], it is not surprising that

perturbing its expression can be detrimental or even lethal in some cases [5]. In fact, MIPS2 (Glyma.11G238800) expression was increased in *1mlpa* in all three germination stages, perhaps in an attempt to restore the *myo*-inositol pool. In *lpa* lines *1mlpa* and *3mlpa*, increased expression was observed in ITPK1- and ITPK4-encoding genes as well, which interestingly are the ITPKs demonstrated to reduce phytic acid in *Arabidopsis* mutants [74, 75]. Also unique to the *mips1* mutation were significant expression changes in PIP5K encoding genes. These enzymes function in inositol pyrophosphate synthesis by phosphorylating $InsP_7$ (derived from phytic acid) to $InsP_8$. Changes in their gene expression is significant because of inositol pyrophosphates' recognition as "energetic signaling" molecules, with roles in energetic metabolism, hormone signaling and Pi sensing [76].

In the germination GRNs from this study and the developing seed GRN from the previous study [25], the *lpa* lines carrying the *mips1* mutation (*1mlpa* and *3mlpa*) were significantly enriched for numerous stress responses. Genes encoding proteins functioning in these stress responses had increased expression in the *lpa* lines, indicating that *1mlpa* and *3mlpa* seeds have increased stress sensitivity, thus impairing their viability and performance and ultimately reducing germination and emergence. Disruption of the *myo*-inositol metabolic pathway may have a negative effect on seed viability due to *myo*-inositol's multifunctional nature in plant metabolism. In fact, several such effects that were found in this study are in accordance with the roles of *myo*-inositol [77-83]. For example, significant enrichment was found for genes functioning in the auxin pathway, cell death, cell wall metabolism, stress processes, and other carbohydrate metabolic pathways requiring *myo*-inositol as a precursor. In the case of stress, the increased expression of stress-related genes in *1mlpa* and *3mlpa* may in part be due to *myo*-inositol's role as a substrate for the biosynthesis of raffinose, galactopinitol,

and *O*-methyl inositols, which participate in stress-related responses and seed desiccation tolerance [78, 80, 81, 84]. Therefore, changes in the contents of these compounds may alter *lpa* seeds' ability to tolerate stress and desiccation. In additional support, MIPS1 is also required for cell death suppression [72], and 145 and 80 genes involved in cell death were differentially expressed in *1mlpa* and *3mlpa*, respectively, suggesting cell death regulation is abnormal in these lines. The examples presented here demonstrate that depletion of the *myo*-inositol pool impact pathways that may affect seed viability in *1mlpa* and *3mlpa*. Thus, due to its many roles, perhaps MIPS1 is not the best target for conditioning the *lpa* phenotype in crop seeds.

Myo-inositol metabolism and seed storage proteins in mrp-1/mrp-n mutant

Following synthesis, phytic acid and the mineral cations it chelates are transported into protein storage vacuoles [85]. MRP proteins from the ABC transporter family are responsible for phytic acid transport and accumulation, as loss-of-function of these transporters can result in the *lpa* trait [13, 38, 86-88]. Shi et al. [13] hypothesize that phytic acid is not transported for storage but instead hydrolyzed in the cytoplasm by endogenous phytases in such *mrp* mutants, thereby preventing phytic acid accumulation. This is supported by concomitant increases in inositol intermediates and *myo*-inositol content [13, 89]. With elevated *myo*-inositol levels, it would be expected for other metabolic pathways utilizing *myo*-inositol to also be affected. Accordingly, *2mlpa* has increased expression in genes functioning in inositol trisphosphate metabolism, phosphatidylinositol transport, and *myo*-inositol transport. Alterations in inositol trisphosphate metabolism are noteworthy because it implies signal transduction is abnormal in *2mlpa*. One such gene that was up-regulated was Glyma.17G219300 encoding a G-protein coupled receptor 1 (GCR1). G-protein coupled receptors function in the phosphatidylinositol signaling pathway,

ultimately yielding two significant signaling molecules: inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG) [90]. Consequently, the GCR1 encoding gene along with the other differentially expressed genes in inositol trisphosphate metabolism suggest irregular signaling may be a feature of *2mlpa,* all of which is a result of an elevated *myo*-inositol pool.

In *2mlpa*, increased expression was found in a number of genes encoding seed storage proteins. Many of these genes encode vicilin-like seed storage proteins and RmlC-like cupin 12S storage proteins. Seed storage proteins have particular importance because they provide an amino acid reserve for use during germination and seedling growth [91]. Whether an increase in protein storage content is a detriment to seed vigor is unclear. According to this network analysis, two genes were responsible for the up-regulation of these genes' expression – Glyma.05G032200 and Glyma.07G060400, encoding an MYB-related TF and GBF3, respectively. The induction of these TFs and the seed storage protein genes they putatively regulate is discussed further below.

Altered regulation in auxin and ABA signaling in lpa seeds

Both phytic acid and *myo*-inositol are critical for normal auxin signaling. Phytic acid itself is in fact a cofactor of transport inhibitor response 1 (TIR1), an auxin receptor and primary mediator of auxin-regulated responses [92], and *myo*-inositol is essential for proper auxin transport and localization [93]. Hence, it should not be surprising that auxin physiology was affected in the *lpa* mutants *1mlpa* and *3mlpa*, where 163 auxin-related genes were differentially expressed between *1mpa* and 1MWT and 155 differentially expressed between *3mlpa* and 3MWT. This finding is consistent with the previous seed development study using the same soybean lines [25]. Auxin signaling is a requirement for seed dormancy and germination

inhibition but is so because it functions to enhance ABA action [94]. In the current study, genes were identified encoding AUXIN RESPONSE FACTOR 10 (ARF10), an element that mediates crosstalk between auxin and the ABA signaling pathway during germination [94]. These ARF10 genes, Glyma.13G325200 in *1mlpa* and Glyma.12G076200 and Glyma.13G325200 in *3mlpa*, had increased expression, which could affect the branch of the ABA pathway regulating germination in these lines. In fact, 345 and 209 ABA-related genes were differentially expressed in the *1mlpa* and *3mlpa* lines, respectively. Not only was the ABA pathway affected in this study, but it was also affected in transcriptome and proteome studies of germinating *lpa* soybeans also carrying *mips1* mutations [20, 21]. This is significant as ABA is the sole hormone known to trigger and maintain seed dormancy and is a major inhibitor of seed germination [95, 96]. Among the differentially expressed ABA genes identified in this study, Glyma.10G071700 and Glyma.13G153200 had increased expression in *1mlpa* and *3mlpa* and encode the bZIP TF ABI5. ABI5 reactivates late embryogenesis programs and arrests embryo growth during germination, causing the embryo to go into a state of dormancy. This is an adaptive response to environmental stress mediated by ABA with ABI5 functioning to maintain the quiescent state [63]. Thus, increased expression of ABI5 in *1mlpa* and *3mlpa* could promote seed dormancy and interfere with the embryos' ability to resume growth.

The GRN from this study predicted two other TF genes, Glyma.04G24900 and Glyma.06G114000, as putative regulators of ABA-related genes in both the Mips and Mips-MRP genotypic subsets. Interestingly, these two TF genes are paralogs corresponding to the same *Arabidopsis* homolog, ATAF1, a TF whose transcript expression is induced in response to ABA and functions to positively regulate ABA biosynthesis [97]. In *Arabidopsis* ATAF1 overexpression studies, Wu et al. [61] found that increased ATAF1 expression confers ABA

hypersensitivity, oxidative stress hypersensitivity, and interferes with plant development. Both ATAF1-encoding genes were up-regulated in *1mlpa* and *3mlpa*. Therefore, increased ATAF1 expression in *1mlpa* and *3mlpa* could contribute to the induction of ABA-responsive gene expression and thus irregularities in ABA signaling, such as up-regulation of genes encoding ABI1, ABF2, AFP2, and PP2CA and down-regulation of those encoding ABI4. Such expression disparities in prominent genes of the ABA signaling pathway could have serious effects on *1mlpa* and *3mlpa* seeds' potential to complete germination and have normal seedling growth. Consistent with this study's findings, Donahue et al. [72] also observed impaired germination and increased ABA sensitivity during germination in *Arabidopsis mips1* mutants.

Despite carrying mutations affecting a different aspect of the phytic acid pathway, *2mlpa* from the MRP genotypic subset also had enrichment for stress responses and abnormalities in ABA signaling. Several of the ABA-related genes that were differentially expressed in *2mlpa* were seed storage proteins, the content of which is influenced by ABA, with high ABA levels stimulating their induction [98-100]. According to their *Arabidopsis* homologs, ABA stimulus also induces expression of the two TF genes found to regulate the seed storage protein genes [101, 102]. These TF genes, the MYB-related TF encoded by Glyma.05G032200 and GBF3 encoded by Glyma.07G060400, had significantly increased expression in *2mlpa* in all three germination stages. Increased expression of the two TF genes and the storage proteins they putatively regulate suggests ABA levels are increased in the seeds of *lpa* line *2mlpa*. Correspondingly, GCR1 (Glyma.17G219300), which was up-regulated in *2mlpa*, is a regulator of ABA signaling and has demonstrated involvement in seed dormancy according to its *Arabidopsis* homolog (AT1G48270) [103, 104]. Thus like *1mlpa* and *3mlpa*, the observed

expression changes of ABA-related genes in *2mlpa* could promote seed dormancy and thereby inhibit germination.

Though different components of the ABA signaling pathway were affected, it is notable that the pathway was disrupted in all three *lpa* lines and that it was also disrupted in the previous seed development study utilizing the same lines as well as in other germination studies on *lpa* soybeans [20, 21, 25]. Hence, because ABA has a major influence on seed dormancy and germination, its irregular manifestation in *lpa* soybean seeds may significantly contribute to the poor seed germination associated with these mutations. Consequently, the relationship between seed phytic acid content and ABA signaling warrants further investigation.

CONCLUSION

Disruption of phytic acid synthesis and accumulation elicited stress responses in the *mips1* and *mrp-l*/*mrp-n* mutants during seed germination. In addition to direct effects of reduced phytic acid, another origin of this stress is altered Pi homeostasis due to increased cellular Pi content and altered cellular *myo*-inositol content, which is diminished in *1mlpa* and *3mlpa* and increased in *2mlpa*. The downstream implications of these changes in Pi and *myo*-inositol content are manifold and could easily be cause for stress, thereby affecting normal cell functioning, seed viability, and ultimately germination and emergence potential. How this relates to the altered regulation in ABA signaling observed in all three *lpa* lines remains to be seen, but the changes observed in ABA signaling are significant, as ABA is a primary regulator of seed dormancy and inhibits germination. These findings, as well as findings in previous studies, indicate changes in ABA signaling may also interfere with germination potential in *lpa* seeds.

Lastly, to establish how the discovered biological processes are differentially regulated in the *lpa* lines used in this study, the GRNs constructed for each subset of genotypes are publicly available, consisting of interactions between TF genes and the target genes they putatively regulate. These interactions aim to help clarify the differential regulation of germination in *lpa* soybean seeds.

ACKNOWLEDGEMENTS

We would like to thank Virginia Tech's Advanced Research Computing servers for their support and Dr. Richard Helm for use of his laboratory equipment. Funding for this study was provided by the Virginia Soybean Board (VSB), the John Lee Pratt Fellowship, the Virginia Tech School of Plant and Environmental Sciences (SPES), the Virginia Tech Agricultural Experiment Station Hatch Program, and Virginia Tech's Open Access Subvention Fund.

REFERENCES

1. Brown K, Solomons N. Nutritional problems of developing countries. Infectious disease clinics of North America. 1991; 5(2):297-317.
2. Ravindran V. Phytates: occurrence, bioavailability and implications in poultry nutrition. Poultry and Avian Biology Reviews. 1995; 6:125-43.
3. Weaver CM, Kannan S. Phytate and mineral bioavailability. Food Phytates. 2002; 2002:211-23.
4. Bohn L, Meyer AS, Rasmussen SK. Phytate: impact on environment and human nutrition. A challenge for molecular breeding. Journal of Zhejiang University Science B. 2008; 9(3):165-91.
5. Raboy V. Approaches and challenges to engineering seed phytate and total phosphorus. Plant Science. 2009; 177(4):281-96.
6. Raboy V. Seed total phosphate and phytic acid. Molecular Genetic Approaches to Maize Improvement Biotechnology in Agriculture and Forestry. 2009; 63:41-53.
7. Larson S, Young K, Cook A, Blake T, Raboy V. Linkage mapping of two mutations that reduce phytic acid content of barley grain. Theoretical and Applied Genetics. 1998; 97(1-2):141-6.
8. Larson SR, Rutger JN, Young KA, Raboy V. Isolation and genetic mapping of a non-lethal rice (Oryza sativa L.) low phytic acid 1 mutation. Crop Science. 2000; 40(5):1397-405.
9. Raboy V, Gerbasi PF, Young KA, Stoneberg SD, Pickett SG, Bauman AT, et al. Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiology. 2000; 124(1):355-68.

10. Wilcox JR, Premachandra GS, Young KA, Raboy V. Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Science. 2000; 40(6):1601-5.

11. Hitz WD, Carlson TJ, Kerr PS, Sebastian SA. Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiology. 2002; 128(2):650-60.

12. Guttieri M, Bowen D, Dorsch JA, Raboy V, Souza E. Identification and characterization of a low phytic acid wheat. Crop Science. 2004; 44(2):418-24.

13. Shi J, Wang H, Schellin K, Li B, Faller M, Stoop JM, et al. Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. Nature Biotechnology. 2007; 25(8):930.

14. Meis SJ, Fehr WR, Schnebly SR. Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. Crop Science. 2003; 43(4):1336-9.

15. Oltmans SE, Fehr WR, Welke GA, Raboy V, Peterson KL. Agronomic and seed traits of soybean lines with low–phytate phosphorus. Crop Science. 2005; 45(2):593-8.

16. Bregitzer P, Raboy V. Effects of four independent low-phytate mutations on barley agronomic performance. Crop Science. 2006; 46(3):1318-22.

17. Raboy V. Seed phosphorus and the development of low-phytate crops. In: Turner BL, Richardson AE, Mullaney EJ, editors. Inositol phosphates: Linking agriculture and the environment. Oxfordshire, UK: CAB International; 2006. p. 111-132.

18. Campion B, Sparvoli F, Doria E, Tagliabue G, Galasso I, Fileppi M, et al. Isolation and characterisation of an lpa (low phytic acid) mutant in common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics. 2009; 118(6):1211-21.

19. Oltmans SE, Fehr WR, Welke GA, Raboy V, Peterson KL. Agronomic and seed traits of soybean lines with low–phytate phosphorus. Crop Science. 2005; 45(2):593-8.

20. Yu X, Jin H, Fu X, Yang Q, Yuan F. Quantitative proteomic analyses of two soybean low phytic acid mutants to identify the genes associated with seed field emergence. BMC Plant Biology. 2019; 19(1):1-14.

21. Yuan F, Yu X, Dong D, Yang Q, Fu X, Zhu S, et al. Whole genome-wide transcript profiling to identify differentially expressed genes associated with seed field emergence in two soybean low phytate mutants. BMC Plant Biology. 2017; 17(1):16-32.

22. Hampton JT, DM; ISTA Vigour Test Committee. ISTA Handbook of Vigour Test Methods. 3rd ed. Zurich, Switzerland: The International Seed Testing Association; 1995.

23. Bowen DE, Souza EJ, Guttieri MJ, Raboy V, Fu J. A low phytic acid barley mutation alters seed gene expression. Crop Science. 2007; 47(S2):S-149-S-59.

24. Redekar NR, Biyashev RM, Jensen RV, Helm RF, Grabau EA, Saghai Maroof MA. Genome-wide transcriptome analyses of developing seeds from low and normal phytic acid soybean lines. BMC Genomics. 2015; 16(1):1074.

25. Redekar N, Pilot G, Raboy V, Li S, Saghai Maroof MA. Inference of transcription regulatory network in low phytic acid soybean seeds. Frontiers in Plant Science. 2017; 8:1-14.

26. Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, et al. Seed germination and vigor. Annual Review of Plant Biology. 2012; 63:507-33.

27. Bewley JD. Seed germination and dormancy. The Plant Cell. 1997; 9(7):1055-66.

28. Bove J, Jullien M, Grappin P. Functional genomics in the study of seed germination. Genome Biology. 2001; 3(1):1002.1-.5.

29. Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, et al. Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiology. 2006; 142(3):839-54.

30. Howell KA, Narsai R, Carroll A, Ivanova A, Lohse M, Usadel B, et al. Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. Plant Physiology. 2009; 149(2):961-80.

31. Bellieny-Rabelo D, De Oliveira EAG, da Silva Ribeiro E, Costa EP, Oliveira AEA, Venancio TM. Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. Scientific Reports. 2016; 6:36009.

32. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends in Genetics. 2014; 30(9):418-26.

33. Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models—a review. Biosystems. 2009; 96(1):86-103.

34. Li Y, Pearl SA, Jackson SA. Gene networks in plant biology: approaches in reconstruction and analysis. Trends in Plant Science. 2015; 20(10):664-75.

35. Banf M, Rhee SY. Computational inference of gene regulatory networks: approaches, limitations and opportunities. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2017; 1860(1):41-52.

36. Krouk G, Lingeman J, Colon AM, Coruzzi G, Shasha D. Gene regulatory networks in plants: learning causality from time and perturbation. Genome Biology. 2013; 14(6):123-9.

37. Haque S, Ahmad JS, Clark NM, Williams CM, Sozzani R. Computational prediction of gene regulatory networks in plant growth and development. Current Opinion in Plant Biology. 2019; 47:96-105.

38. Saghai Maroof MA, Glover NM, Biyashev RM, Buss GR, Grabau EA. Genetic basis of the low-phytate trait in the soybean line CX1834. Crop Science. 2009;49(1):69-76.

39. Saghai Maroof MA, Buss GR. Low phytic acid, low stachyose, high sucrose soybean lines. Google Patents; 2008.

40. Walker D, Scaboo A, Pantalone V, Wilcox J, Boerma H. Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. Crop Science. 2006; 46(1):390-7.

41. Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014; 15(1):182-93.

42. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010; 463(7278):178-83.

43. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15-21.

44. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013; 30(7):923-30.

45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014; 15(12):550.

46. Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978; 6(2):461-4.

47. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. The R Journal. 2016; 8(1):289-317.

48. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research. 2009; 38(suppl_1):D843-D6.

49. Fisher RA. Statistical methods for research workers: Genesis Publishing Pvt Ltd; 2006.

50. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research. 2017; 45(D1):D1040–D5.

51. DeMers LC, Redekar NR, Kachroo A, Tolin SA, Li S, Saghai Maroof MA. A transcriptional regulatory network of Rsv3-mediated extreme resistance against Soybean mosaic virus. PloS One. 2020; 15(4):e0231658.

52. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nature Genetics. 2003; 34(2):166.

53. Schäfer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics. 2004; 21(6):754-64.

54. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006; 7(1):S7.

55. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biology. 2007; 5(1):e8.

56. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PloS One. 2010; 5(9):e12776.

57. Haury A-C, Mordelet F, Vera-Licona P, Vert J-P. TIGRESS: trustful inference of gene regulation using stability selection. BMC Systems Biology. 2012; 6(1):145-61.

58. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. Nature Methods. 2012; 9(8):796-804.

59. Kruger NJ, von Schaewen A. The oxidative pentose phosphate pathway: structure and organisation. Current Opinion in Plant Biology. 2003; 6(3):236-46.

60. O'Malley RC, Huang S-sC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and epicistrome features shape the regulatory DNA landscape. Cell. 2016; 165(5):1280-92.

61. Wu Y, Deng Z, Lai J, Zhang Y, Yang C, Yin B, et al. Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. Cell Research. 2009; 19(11):1279-90.

62. Jensen MK, S L, De Masi F, Reimer JJ, Nielsen M, Perera V, et al. ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in Arabidopsis thaliana. FEBS open bio. 2013; 3(1):321-7.

63. Lopez-Molina L, Mongrand S, Chua N-H. A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in Arabidopsis. Proceedings of the National Academy of Sciences. 2001; 98(8):4782-7.

64. Blackman SA, Obendorf RL, Leopold AC. Desiccation tolerance in developing soybean seeds: the role of stress proteins. Physiologia Plantarum. 1995; 93(4):630-8.

65. Theodorou ME, Plaxton WC. Metabolic adaptations of plant respiration to nutritional phosphate deprivation. Plant Physiology. 1993; 101(2):339-44.

66. Veljanovski V, Vanderbeld B, Knowles VL, Snedden WA, Plaxton WC. Biochemical and molecular characterization of AtPAP26, a vacuolar purple acid phosphatase up-regulated in phosphate-deprived Arabidopsis suspension cells and seedlings. Plant Physiology. 2006; 142(3):1282-93.

67. Del Pozo JC, Allona I, Rubio V, Leyva A, De La Peña A, Aragoncillo C, et al. A type 5 acid phosphatase gene from Arabidopsis thaliana is induced by phosphate starvation and by some other types of phosphate mobilising/oxidative stress conditions. The Plant Journal. 1999; 19(5):579-89.

68. Cheng Y, Zhou W, El Sheery NI, Peters C, Li M, Wang X, et al. Characterization of the Arabidopsis glycerophosphodiester phosphodiesterase (GDPD) family reveals a role of the plastid-localized AtGDPD1 in maintaining cellular phosphate homeostasis under phosphate starvation. The Plant Journal. 2011; 66(5):781-95.

69. Pilu R, Panzeri D, Gavazzi G, Rasmussen SK, Consonni G, Nielsen E. Phenotypic, genetic and molecular characterization of a maize low phytic acid mutant (lpa241). Theoretical and Applied Genetics. 2003; 107(6):980-7.

70. Nunes AC, Vianna GR, Cuneo F, Amaya-Farfan J, de Capdeville G, Rech EL, et al. RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta. 2006; 224(1):125-32.

71. Obendorf RL, Zimmerman AD, Zhang Q, Castillo A, Kosina SM, Bryant EG, et al. Accumulation of soluble carbohydrates during seed development and maturation of low-raffinose, low-stachyose soybean. Crop Science. 2009; 49(1):329-41.

72. Donahue JL, Alford SR, Torabinejad J, Kerwin RE, Nourbakhsh A, Ray WK, et al. The Arabidopsis thaliana myo-inositol 1-phosphate synthase1 gene is required for myo-inositol synthesis and suppression of cell death. The Plant Cell. 2010; 22(3):888-903.

73. Chiera JM, Finer JJ, Grabau EA. Ectopic expression of a soybean phytase in developing seeds of Glycine max to improve phosphorus availability. Plant Molecular Biology. 2004; 56(6):895-904.

74. Kuo HF, Hsu YY, Lin WC, Chen KY, Munnik T, Brearley CA, et al. Arabidopsis inositol phosphate kinases IPK 1 and ITPK 1 constitute a metabolic pathway in maintaining phosphate homeostasis. The Plant Journal. 2018; 95(4):613-30.

75. Kim S-I, Tai TH. Identification of genes necessary for wild-type levels of seed phytic acid in Arabidopsis thaliana using a reverse genetics approach. Molecular Genetics and Genomics. 2011; 286(2):119-33.

76. Freed C, Adepoju O, Gillaspy G. Can Inositol Pyrophosphates Inform Strategies for Developing Low Phytate Crops? Plants. 2020; 9(1):115-25.

77. Loewus F, editor. Biogenesis of Plant Cell Wall Polysaccharides. New York: Academic Press; 1973.

78. Obendorf RL. Oligosaccharides and galactosyl cyclitols in seed desiccation. Seed Science Research. 1997; 7:63-74.

79. Murthy PP. Inositol phosphates and their metabolism in plants. myo-Inositol Phosphates, Phosphoinositides, and Signal Transduction: Springer; 1996. p. 227-55.

80. Peterbauer T, Puschenreiter M, Richter A. Metabolism of galactosylononitol in seeds of Vigna umbellata Plant and Cell Physiology. 1998; 39:334-41.

81. Peterbauer T, Richter A. Galactosylononitol and stachyose synthesis in seeds of Adzuki bean. Plant Physiology. 1998; 117:165-72.

82. Slovin JP, Bandurski RS, Cohen JD. Control of Hormone Synthesis and Metabolism Chapter 5 Auxins. Hooykaas PJ, Hall MA, Libbenga KR, editors. Amsterdam: Elsevier; 1999.

83. Loewus FA, Murthy PP. myo-Inositol metabolism in plants. Plant Science. 2000; 150(1):1-19.

84. Horbowicz M, Obendorf RL. Seed desiccation tolerance and storability: dependence on flatulence-producing oligosaccharides and cyclitols—review and survey. Seed Science Research. 1994; 4(4):385-405.

85. Lott JNA, Greenwood JS, Batten GD. Mechanisms and Regulation of Mineral Nutrient Storage During Seed Development. Seed Development and Germination. 1995; 41:215.

86. Nagy R, Grob H, Weder B, Green P, Klein M, Frelet-Barrand A, et al. The Arabidopsis ATP-binding cassette protein AtMRP5/AtABCC5 is a high affinity inositol hexakisphosphate transporter involved in guard cell signaling and phytate storage. Journal of Biological Chemistry. 2009; 284(48):33614-22.

87. Xu X-H, Zhao H-J, Liu Q-L, Frank T, Engel K-H, An G, et al. Mutations of the multi-drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic acid in rice seeds. Theoretical and Applied Genetics. 2009; 119(1):75-83.

88. Panzeri D, Cassani E, Doria E, Tagliabue G, Forti L, Campion B, et al. A defective ABC transporter of the MRP family, responsible for the bean lpa1 mutation, affects the regulation of the phytic acid pathway, reduces seed myo-inositol and alters ABA sensitivity. New Phytologist. 2011; 191(1):70-83.

89. Israel DW, Taliercio E, Kwanyuen P, Burton JW, Dean L. Inositol metabolism in developing seed of low and normal phytic acid soybean lines. Crop Science. 2011; 51(1):282-9.

90. Gilman AG. G proteins: transducers of receptor-generated signals. Annual Review of Biochemistry. 1987; 56(1):615-49.

91. Shewry PR, Napier JA, Tatham AS. Seed storage proteins: structures and biosynthesis. The Plant Cell. 1995; 7(7):945-56.

92. Tan X, Calderon-Villalobos LIA, Sharon M, Zheng C, Robinson CV, Estelle M, et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature. 2007; 446(7136):640-5.

93. Luo Y, Qin G, Zhang J, Liang Y, Song Y, Zhao M, et al. D-myo-inositol-3-phosphate affects phosphatidylinositol-mediated endomembrane function in Arabidopsis and is essential for auxin-regulated embryogenesis. The Plant Cell. 2011; 23(4):1352-72.

94. Liu X, Zhang H, Zhao Y, Feng Z, Li Q, Yang H-Q, et al. Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. Proceedings of the National Academy of Sciences. 2013; 110(38):15485-90.

95. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. Abscisic acid: emergence of a core signaling network. Annual Review of Plant Biology. 2010; 61:651-79.

96. Hubbard KE, Nishimura N, Hitomi K, Getzoff ED, Schroeder JI. Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. Genes & Development. 2010; 24(16):1695-708.

97. Jensen MK, Lindemose S, De Masi F, Reimer JJ, Nielsen M, Perera V, et al. ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in Arabidopsis thaliana. FEBS open bio. 2013; 3(1):321-7.

98. Bray EA, Beachy RN. Regulation by ABA of β-conglycinin expression in cultured developing soybean cotyledons. Plant Physiology. 1985; 79(3):746-50.

99. Finkelstein RR, Tenbarge KM, Shumway JE, Crouch ML. Role of ABA in maturation of rapeseed embryos. Plant Physiology. 1985; 78(3):630-6.

100. Kagaya Y, Okuda R, Ban A, Toyoshima R, Tsutsumida K, Usui H, et al. Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. Plant and Cell Physiology. 2005; 46(2):300-11.

101. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, et al. The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Molecular Biology. 2006; 60(1):107-24.

102. Lu G, Paul A-L, McCarty DR, Ferl RJ. Transcription factor veracity: is GBF3 responsible for ABA-regulated expression of Arabidopsis Adh? The Plant Cell. 1996; 8(5):847-57.

103. Chen J-G, Pandey S, Huang J, Alonso JM, Ecker JR, Assmann SM, et al. GCR1 can act independently of heterotrimeric G-protein in response to brassinosteroids and gibberellins in Arabidopsis seed germination. Plant Physiology. 2004; 135(2):907-15.

104. Pandey S, Assmann SM. The Arabidopsis putative G protein–coupled receptor GCR1 interacts with the G protein α subunit GPA1 and regulates abscisic acid signaling. The Plant Cell. 2004; 16(6):1616-32.

Table 2.1 | Characteristics and classification of parental and experimental soybean lines.

Soybean Lines	Genotypic Class Subset	Genotype	Phytic Acid	Emergence	Stachyose	Sucrose	Cross
V99-5089	-	*mips1*/MRP-L/MRP-N	Low	Low	Low	High	Parent
CX-1834	-	MIPS1/*mrp-l*/*mrp-n*	Low	Low	Normal	Normal	Parent
Essex	-	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	Parent
1mlpa	Mips	*mips1*/MRP-L/MRP-N	Low	Low	Low	High	Essex x V99-5089
1MWT	Mips	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	Essex x V99-5089
2mlpa	MRP	MIPS1/*mrp-l*/*mrp-n*	Low	Low	Normal	Normal	CX-1834 x V99-5089
2MWT	MRP	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	CX-1834 x V99-5089
2MWT-L	MRP	MIPS1/MRP-L/*mrp-n*	Normal	Normal	Normal	Normal	CX-1834 x V99-5089
2MWT-N	MRP	MIPS1/*mrp-l*/MRP-N	Normal	Normal	Normal	Normal	CX-1834 x V99-5089
3mlpa	Mips-MRP	*mips1*/*mrp-l*/*mrp-n*	Low	Low	Normal	Normal	V99-5089 X CX-1834
3MWT	Mips-MRP	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	V99-5089 X CX-1834

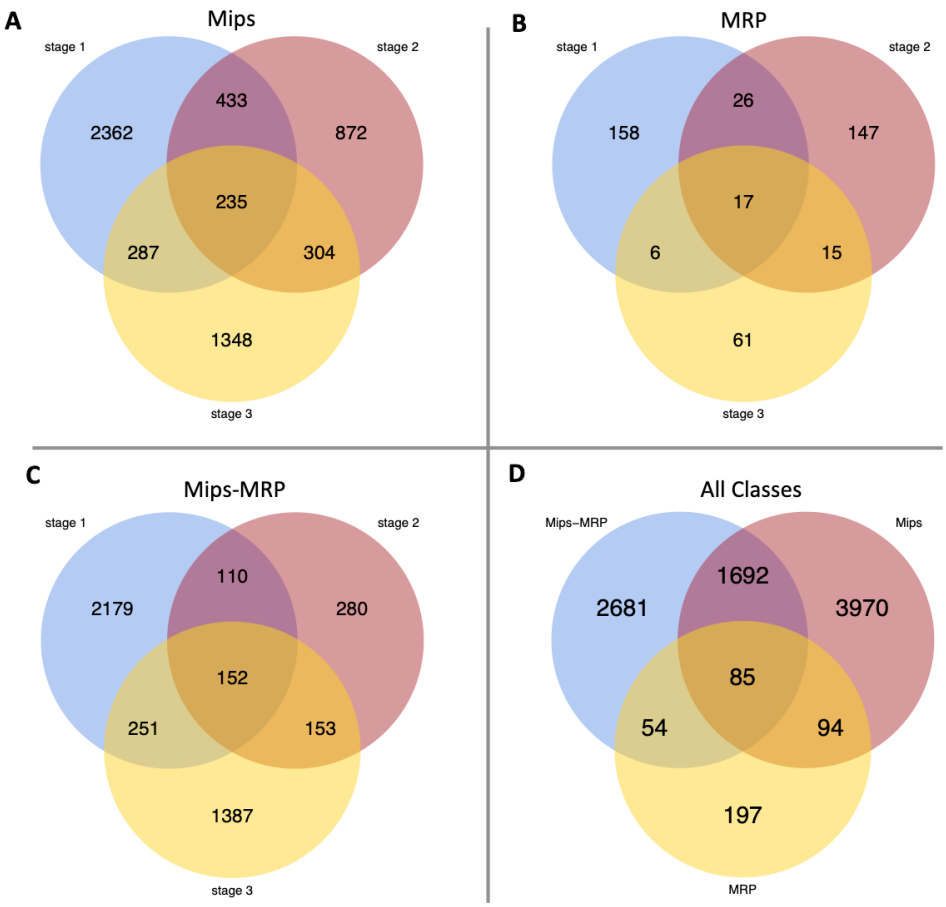


Figure 2.1 | Venn diagrams of differentially expressed genes (DEGs). (**A**) Number of DEGs unique to and shared between each stage in the Mips subset. (**B**) Number of DEGs unique to and shared between each stage in the MRP subset. (**C**) DEGs in the Mips-MRP subset. (**D**) Number of DEGs unique to and shared between all three subsets of genotypes.

Table 2.2 | Differentially expressed genes in phytic acid biosynthesis pathway.

Genotypic Class Subset	Gene ID	*Arabidopsis* Homolog	Log2(FC) (*lpa*/normal)			Gene Symbol	Protein Symbol
			stage 1	stage 2	stage 3		
Mips	Glyma.11G238800	AT2G22240	2.5	1.6	1.2	MIPS2	Myo-inositol-1-phosphate synthase 2
Mips	Glyma.01G016700	AT5G16760	-	-	1.4	ITPK1	Inositol 1,3,4-trisphosphate 5/6-kinase 1
Mips	Glyma.09G206100	AT5G16760	1.0	-	-	ITPK1	Inositol 1,3,4-trisphosphate 5/6-kinase 1
Mips	Glyma.11G218500	AT2G43980	1.6	-	-	ITPK4	Inositol 1,3,4-trisphosphate 5/6-kinase 4
Mips	Glyma.18G038800	AT2G43980	1.3	-	-	ITPK4	Inositol 1,3,4-trisphosphate 5/6-kinase 4
Mips-MRP	Glyma.11G218500	AT2G43980	1.3	-	-	ITPK4	Inositol 1,3,4-trisphosphate 5/6-kinase 4
Mips-MRP	Glyma.18G038800	AT2G43980	2.3	1.1	1.1	ITPK4	Inositol 1,3,4-trisphosphate 5/6-kinase 4

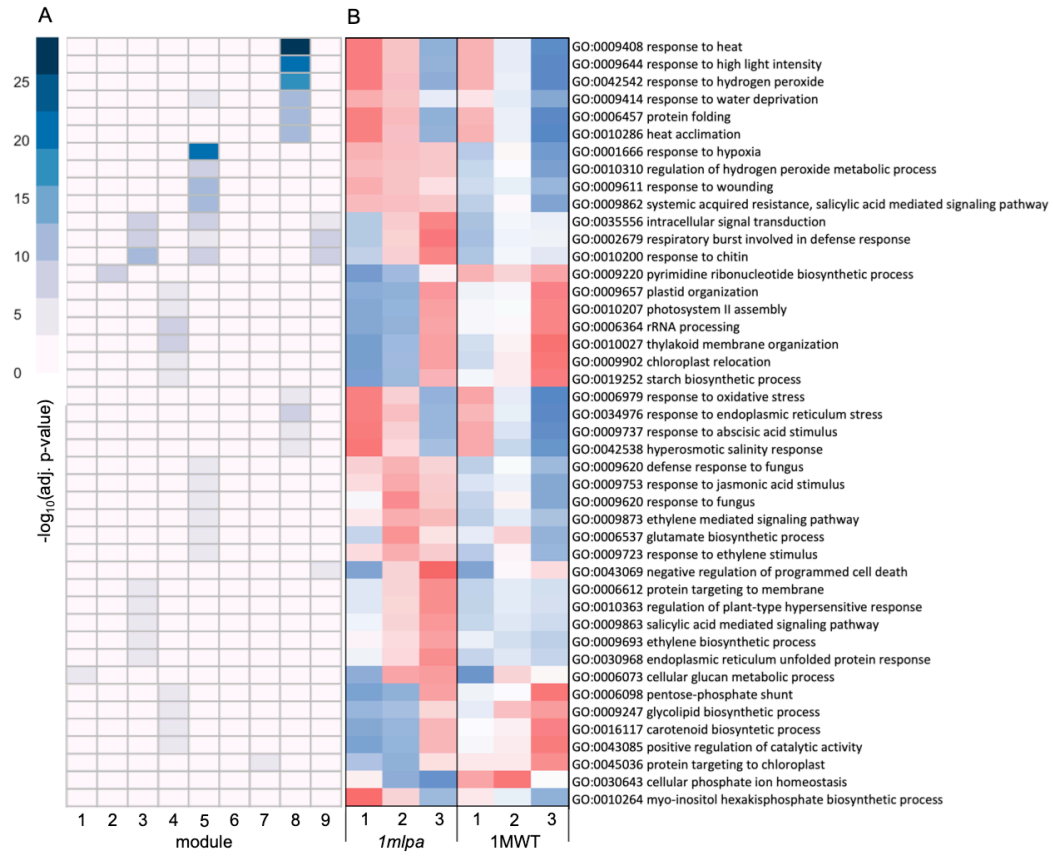


Figure 2.2 | Mips subset gene co-expression modules and significantly enriched biological processes. A module is defined as a group of genes sharing similar expression profiles over time and likely involved in the same biological processes. Rows represent hierarchical clustering of significantly enriched GO categories. Significantly enriched GO categories were defined as those with an FDR <0.05. **(A)** Heatmap of significantly enriched GO biological processes in each gene co-expression module. Columns represent modules. Color represents $-\log_{10}$ adjusted p-value. **(B)** Average scaled expression of genes in significantly enriched biological processes. Columns represent germination stages of *1mlpa* and 1MWT. Red color represents increased expression, and blue color represents decreased expression. Enrichment for all GO categories can be found in Table S2.2.

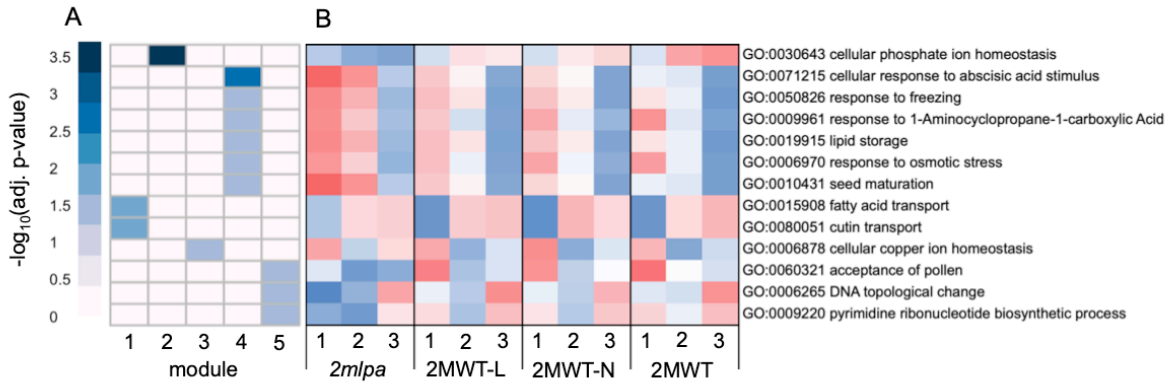


Figure 2.3 | MRP subset gene co-expression modules and significantly enriched biological processes. A module is defined as a group of genes sharing similar expression profiles over time and likely involved in the same biological processes. Rows represent hierarchical clustering of significantly enriched GO categories. Significantly enriched GO categories were defined as those with an FDR <0.05 **(A)** Heatmap of significantly enriched GO biological processes in each gene co-expression module. Columns represent modules. Color represents $-\log_{10}$ adjusted p-value. **(B)** Average scaled expression of genes in significantly enriched biological processes. Columns represent germination stages of *2mlpa*, 2MWT-L, 2MWT-N, and 2MWT. Red color represents increased expression, and blue color represents decreased expression. Enrichment for all GO categories can be found in Table S2.2.

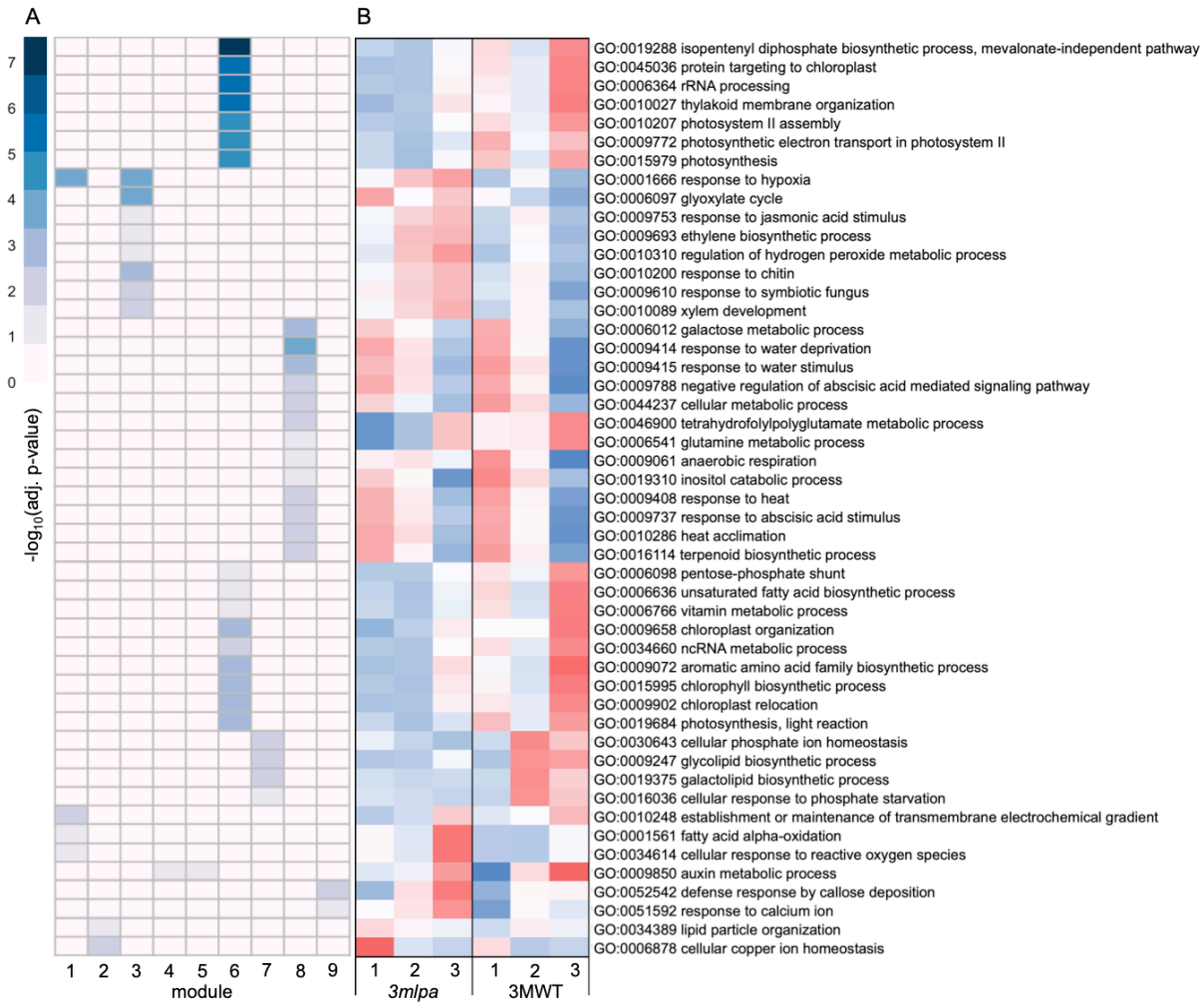


Figure 2.4 | Mips-MRP subset gene co-expression modules and significantly enriched biological processes. A module is defined as a group of genes sharing similar expression profiles over time and likely involved in the same biological processes. Rows represent hierarchical clustering of significantly enriched GO categories. Significantly enriched GO categories were defined as those with an FDR <0.05. **(A)** Heatmap of significantly enriched GO biological processes in each gene co-expression module. Columns represent modules. Color represents $-\log_{10}$ adjusted p-value. **(B)** Average scaled expression of genes in significantly enriched biological processes. Columns represent germination stages of *3mlpa* and 3MWT. Red color represents increased expression, and blue color represents decreased expression. Enrichment for all GO categories can be found in Table S2.2.

Table 2.3 | Putative candidate transcription factors shared between genotypic subsets' gene regulatory networks.

TF Gene	Arabidopsis Homolog	TF Family	Gene Symbol	Subsets Shared With	Log2(FC) (*lpa*/normal)			Number of Putative Targets	Number of Shared Targets
					stage 1	stage 2	stage 3		
Glyma.01G216000	AT1G12610	ERF	DREB1F, DDF2, ERF033	Mips	1.2	2.3	0	22	0
				MRP	0	2.6	2.1	26	
				Mips-MRP	2.5	2.6	1.9	32	
Glyma.02G080200	AT2G33710	ERF	-	Mips	0	0	1.4	103	5
				Mips-MRP	1.3	0	0	99	
Glyma.04G249000	AT1G01720	NAC	ATAF1, NAC2	Mips	0	0	1.8	365	12
				Mips-MRP	0	0	2.1	139	
Glyma.06G114000	AT1G01720	NAC	ATAF1, NAC2	Mips	0	1.2	1.9	158	31
				Mips-MRP	0	0	2.6	228	
Glyma.08G298200	AT2G02820	MYB	MYB88	Mips	0	0	1.1	117	0
				Mips-MRP	-1.1	0	1.3	149	
Glyma.16G167500	AT1G76890	Trihelix	GT-2	Mips	-2.0	0	0	114	11
				Mips-MRP	-1.2	-1.7	-1.3	115	

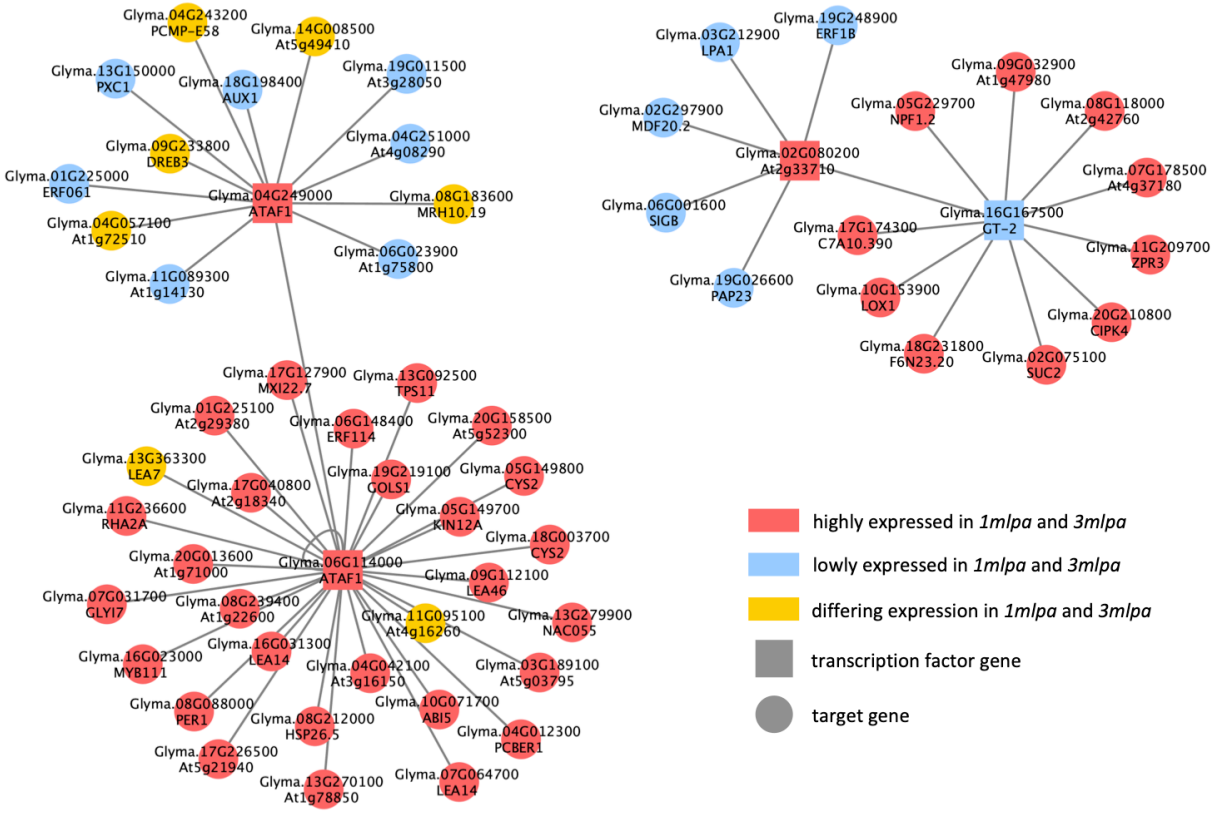


Figure 2.5 | Consensus GRN of Mips and Mips-MRP genotypic subsets. GRN of putative regulatory interactions between differentially expressed TF genes and differentially expressed target genes found in both the Mips and Mips-MRP GRNs. TF genes (square nodes) directly regulate (gray edges) target genes (circular nodes). Nodes in red are genes with increased expression in *1mlpa* and *3mlpa*. Nodes in blue are genes with decreased expression in *1mlpa* and *3mlpa*. Nodes in yellow are genes with differing expression in *1mlpa* and *3mlpa*.

Table 2.4 | Putative target genes in the MRP subset with annotations for observed significant GO categories validated by *Arabidopsis* DAP-seq dataset and motif sequence analysis.

Gene Name	*Arabidopsis* Homolog	Gene Symbol	Description	GO Term	GO Description(s)
Glyma.01G119600	AT2G40170	-	AT2G40170 protein	GO:0019915 GO:0050826	Lipid storage Response to freezing
Glyma.04G085000	AT4G32880	HB8	Homeobox-leucine zipper protein ATHB-8	GO:0010431	Seed maturation
Glyma.07G190100	AT1G61340	-	F-box protein	GO:0009961 GO:0006970	Response to 1-aminocyclopropane-1-carboxylic acid Response to osmotic stress
Glyma.10G037100	AT5G44120	CRA1, 12S STORAGE PROTEIN, CRU1	RmlC-like cupins superfamily protein	GO:0010431 GO:0071215	Seed maturation Cellular response to abscisic acid stimulus
Glyma.13G123500	AT1G03880	CRB, CRU2, CRU3	12S seed storage protein CRB (Cruciferin 2)	GO:0019915 GO:0071215 GO:0010431 GO:0050826	Lipid storage Cellular response to abscisic acid stimulus Seed maturation Response to freezing
Glyma.19G164900	AT5G44120	CRA1, 12S STORAGE PROTEIN, CRU1	RmlC-like cupins superfamily protein	GO:0010431 GO:0071215	Seed maturation Cellular response to abscisic acid stimulus
Glyma.20G148300	AT3G22640	PAP85	Vicilin-like seed storage protein (Glubulin)	GO:0019915 GO:0050826	Lipid storage Response to freezing

CHAPTER 3

Analysis of low and normal phytic acid soybean (*Glycine max*) seed lipids and exudates reveals distinct chemotypes

Lindsay C DeMers[1], Sherry B Hildreth[2], Judith Jervis[3], MA Saghai Maroof[1], and Richard F Helm[3*]

[1]*School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia, United States of America.* [2]*Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, United States of America.* [3]*Department of Biochemistry, Virginia Tech, Blacksburg, Virginia, United States of America.* [*]Corresponding author: helmrf@vt.edu

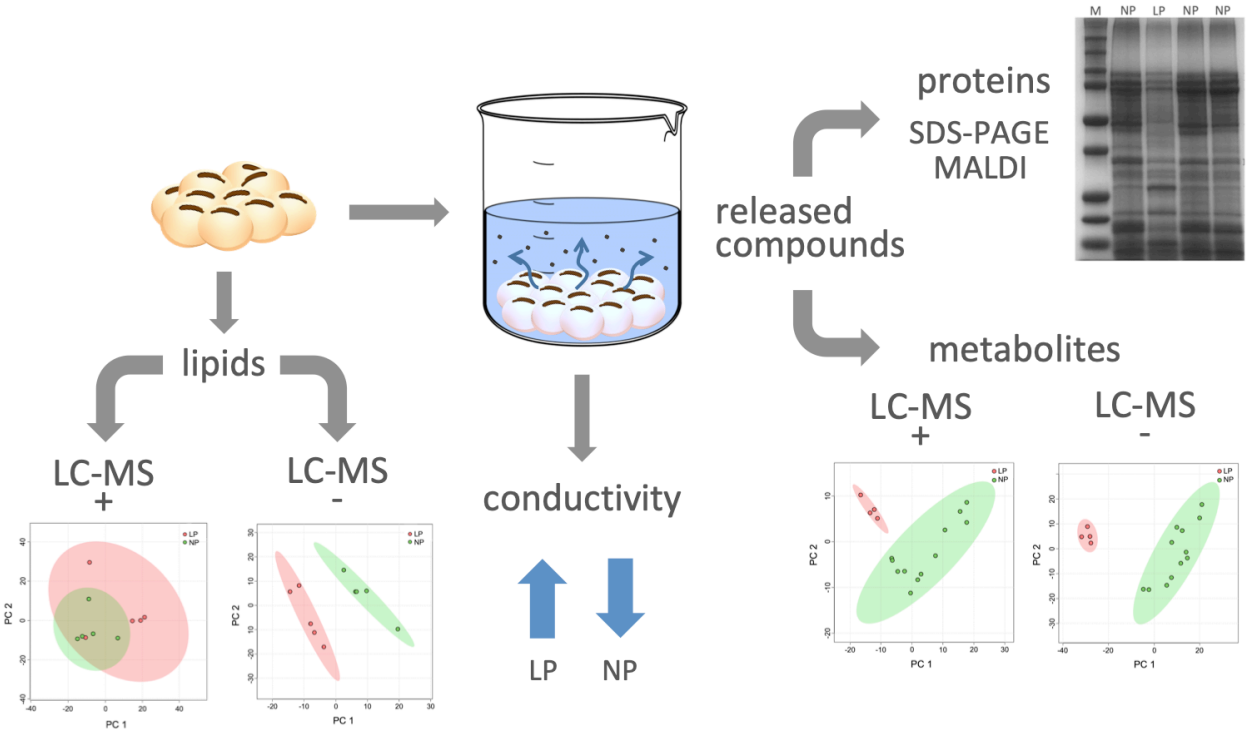
This chapter is to be submitted for publication in *Journal of Agricultural and Food Chemistry*.

ABSTRACT

While reducing phytic acid levels in soybean seeds has significant environmental and nutritional benefits, low phytate soybeans exhibit a poor seedling emergence phenotype. In an effort to better understand the biochemistry related to this phenotype, lipid profiles of low and normal phytic acid soybean seeds, as well as the physiochemical properties of materials released from imbibed seeds were evaluated from two unique genotypic class subsets. Principal component analyses of untargeted LC-MS lipidomic data showed separation of low phytic acid lines from normal phytic acid lines in negative ion mode, though little separation was found in positive ion mode. Differences were found in ceramide, glucose-sitosterol, peroxidized triacylglycerol, and phospholipid contents. Seeds from these lines also underwent electrolyte conductivity testing with the seed exudates analyzed for differences in protein and metabolite contents. Conductivity testing revealed low phytic acid soybean seeds leak significantly more electrolytes than normal phytic acid seeds, while analyses of seed exudates revealed striking differences in protein and metabolite profiles. These results are discussed in relation to the low emergence phenotype.

KEYWORDS

soybean (*Glycine max*), phytic acid, lipidomics, *myo*-inositol phosphate synthase, multidrug-resistance protein ABC transporter, ceramide, phospholipids, sitosterol, seed conductivity, seed vigor, programmed cell death



Graphical Abstract. Workflow of studies performed on low (LP) and normal (NP) phytic acid soybean seeds. Lipids were extracted from dry seeds and profiled using liquid chromatography-mass spectrometry (LC-MS) in both positive and negative ion modes. Seeds underwent conductivity testing to measure the amount of electrolytes released during a water soak. The electrolytes released were then analyzed for protein content and metabolite content by SDS-PAGE followed by MALDI-TOF/TOF and LC-MS, respectively.

INTRODUCTION

Soybean (*Glycine max*) accounted for nearly 90 million acres of planted crops in the United States in 2018, making it the second-most planted field crop after corn [1]. Its use spans many economic sectors, from human and livestock food production to industrial products, such as biodiesels, pharmaceuticals, and building materials. As a food and feed source, soybean seed is especially valuable; it is the number one oilseed crop in the U.S. and the number one protein source in the world for livestock industries [2]. With an increasing demand for soybean and soybean products, the continued development of high-performing cultivars with desirable agronomic traits is necessary in order to meet global demand.

Soybean seeds, along with other grains and legumes, contain an antinutritive compound called phytic acid (phytate, *myo*-inositol-(1,2,3,4,5,6)-hexa*kis*phosphate) that sequesters more than 75% of seed phosphorus, while also chelating cations such as Fe^{3+}, K^{+}, Ca^{2+}, Mg^{2+}, and Zn^{2+} [3-6]. Since phytic acid is indigestible to humans and monogastric livestock, the bioavailability of phosphorus and these cations is reduced in these organisms, which leads to the excretion of excess phosphorus into the environment and potential nutrient deficiency issues [7-10].

Low phytic acid (*lpa*) crops have been developed to mitigate the adverse nutritional and environmental effects of seed phytic acid [11-17]. In soybean, the *lpa* cultivar "V99-5089" carries a mutation in a gene encoding *myo*-inositol 3-phosphate synthase (MIPS1) [18, 19]. This loss-of-function mutation prevents the conversion of glucose-6-phosphate to *myo*-inositol 3-phosphate, the first step in phytic acid biosynthesis. This mutation reduces phytic acid levels while increasing that of inorganic phosphate [20, 21]. "CX-1834," another *lpa* soybean cultivar, carries point mutations on two epistatically interacting loci, chromosome 3 (linkage group N)

and chromosome 19 (linkage group L) [15, 22]. Both mutations are in genes encoding multidrug resistance-associated protein (MRP) ATP-binding cassette (ABC) transporters (henceforth referred to as MRPs), which have been linked to phytic acid transport and storage [18, 23].

Unfortunately, a hallmark feature of *lpa* soybeans (and other *lpa* crops) is reduced seedling emergence [24-31]; however, the biochemical mechanism(s) linking phytic acid metabolism and seedling emergence remains to be established. To understand the basis of the low emergence phenotype, polar metabolomes of normal and *lpa* soybeans were previously profiled using liquid chromatography-mass spectrometry (LC-MS) [32]. While the levels of neutral oligosaccharides were the same for all lines investigated, significant metabolite differences were found in the levels of malonyl isoflavones, soyasaponins, and arginine. In addition, two methanol-soluble polypeptides differed in abundance, one of which was the allergen Gly m 1. Nevertheless, none of these changes could directly explain the low emergence phenotype [32].

Lipidomics, a branch of metabolomics that targets the lipid component of cells or tissues, can provide an additional level of insight into the low and normal phytic acid soybeans [32]. High-throughput lipid profiling has successfully identified lipid metabolic pathways that regulate plant growth and development, environmental stress responses, and cellular processes, such as signal transduction, vesicle trafficking, and cytoskeletal rearrangement [33]. Because lipidomics allows lipid metabolism to be studied in a physiological context, it provides an opportunity to investigate metabolic processes that define the reduced emergence phenotype in *lpa* soybeans. This study used untargeted lipidomics to report novel information regarding the *lpa* soybean lines used previously [32], while including new genotypes as well.

Materials released during the initial stages of seed germination may also provide insights into the low emergence phenotype of *lpa* seeds. Hence, we also evaluated variations in seed conductivity between low and normal phytic acid lines as well as the proteins and metabolites released during a water soak. These results are summarized to provide a working hypothesis on the mechanisms of low emergence in *lpa* soybeans.

MATERIALS AND METHODS

Plant material

The six experimental soybean lines used in this study were *1mlpa*, 1MWT, *2mlpa*, 2MWT, 2MWT-L, and 2MWT-N (Table 3.1). These lines are from two different genotypic class subsets. The first subset, designated as the "Mips" subset, contains the *lpa* line "*1mlpa,*" which carries a homozygous mutant allele for the MIPS1 gene (*mips1* mutation), and the normal phytic acid line "1MWT" with a wild-type MIPS1 allele. The *mips1* mutation responsible for the *lpa* phenotype is the result of a point mutation on chromosome 11 [19]. These lines are isogenic and were developed from a cross between the normal phytic acid line "Essex'" (no MIPS1 mutation) and the *lpa* line "V99-5089" (*mips1* mutation) [19]. The second genotypic subset, designated as the "MRP" subset, contains four near-isogenic lines (NILs) which were used previously [32] - the *lpa* line "*2mlpa,*" which is homozygous for mutations in both MRP genes (*mrp-l/mrp-n* mutations), the normal phytic acid line "2MWT" carrying wild-type alleles for both MRP genes, and two normal phytic acid lines "2MWT-L" and "2MWT-N" with single mutations in MRP-N and MRP-L, respectively. These lines were developed from a cross between the *lpa* lines "CX-1834" (*mrp-1/mrp-n* mutations) and V99-5089 (*mips1* mutation). The mutations conferring the *lpa* phenotype in *2mlpa* and CX-1834 are the result of point mutations in the epistatically

interacting loci, MRP-L and MRP-N, on chromosomes 19 and 3, respectively [15, 18, 22]. The seeds used in this study were harvested in 2017 from a field in Blacksburg, VA and stored at 4 °C until testing.

Preparation of the lipid extracts

The four soybean lines used for lipid profiling included *1mlpa*, 1MWT, *2mlpa*, and 2MWT. These four lines were comprised of five biological replicates with each replicate containing ten randomly selected seeds. In brief, seeds were flash-frozen in liquid nitrogen and finely ground with a P14 mill (Pulverisette 14, Fritsch, Pittsboro, NC) using a 0.5 sieve at 20,000 rpm. The powder was transferred to pre-weighed, ethanol washed 15 mL tubes and then weighed and stored at −80 °C. A portion of this powder (400 mg) was dried overnight on a high-vacuum line with resultant dry weights being used for normalization. The nonpolar components were extracted using dry ethyl acetate, concentrated to an oil, and stored at −80 °C as described previously [32].

Liquid chromatography-mass spectrometry

An Acquity I-class UPLC coupled with a Synapt G2-S HDMS (Waters Corp., Milford, MA) was used for sample analysis in both positive and negative ionization modes. The concentrated oil was dissolved in dichloromethane:methanol:water (60:30:4, v/v) with aliquots combined to create a master mix of each line (*1mlpa*, 1MWT, *2mlpa*, 2MWT) along with a complete master mix composed of all four lines. To condition the column, three blank injections were followed by three complete master mix injections. The samples were run in randomized sets of a single biological replicate from each line with three technical replicates. Each set was

followed by a blank and a complete master mix. Additionally, a set of four complete master mix injections was analyzed in MS^E mode with increasing collision energies to aid in lipid identification (10, 20, 30, 40 V).

Samples were separated with a binary solvent system of acetonitrile:water with 50 mM ammonium acetate (6:4, v/v; Solvent A) and acetonitrile:isopropanol with 50 mM ammonium acetate (1:9, v/v; Solvent B) on an Acquity HSS T3 Column (1.8 μm, 2.1 x 100 mm, Waters Corp., Milford, MA) with a flow rate of 400 μl/min and a 16 minute gradient. The starting gradient conditions was 40% B, then a linear gradient to 100% B (0-10 min), hold at 100% B (10-12 min), return to initial conditions (12-13 min), and isocratic at 40% B (13-16 min). Sample injection volume was 2 μl.

Column eluates were ionized by electrospray ionization and analyzed independently in both positive and negative modes. Data was collected in high resolution, MS^E mode with a scan time of 0.20 sec, and a mass range of 250-1800 *m/z* for low energy function 1 and 50-1800 *m/z* for function 2 with a collision energy ramp from 20-30. The source parameters for positive ion mode were source temperature 125 °C, capillary voltage 3.0, cone voltage 40, source offset 80, desolvation temperature 400 °C, cone gas 60 L/h, desolvation gas 600 L/h, and nebulizer gas 6.0 bar. The source parameters for negative ion mode were source temperature 125^{o}C, capillary voltage 2.4, cone voltage 40, source offset 80, desolvation temperature 400 °C, cone gas 50 L/h, desolvation gas 600 L/h, and nebulizer gas 6.0 bar. A reference sprayer released leucine-enkephalin (200 ng/mL, Waters Corp., MA) continuously at 5 μl/min with a scan frequency of 20 sec.

Data processing and analysis

The raw data was processed using MarkerLynx software (version 4.1, Waters Corp., Milford, MA) with the following parameters for negative ionization mode: retention time range 1.0-10.0 min, retention time window of 0.1 min, mass range 400-1,800 *m/z*, mass window 0.01 *m/z*, noise elimination of level 6, peak intensity threshold of 900, marker intensity threshold of 5000. The raw data from positive ionization mode was processed in two parts due to the abundance of triacylglycerides at later times. The parameters used for the earlier time were: retention time range 1.0-9.0 min, retention time window of 0.1 min, mass range 400-1,800 *m/z*, mass window 0.02 *m/z*, noise elimination of level 10, peak intensity threshold of 1,500, and marker intensity threshold of 10,000. The parameters used for the later time were the same with the exception of the retention time range set at 9.0-10.5 min and a marker intensity threshold of 15,000.

The detected exact mass-retention time pairs (EMRTs) and their raw peak intensities were further processed using the online tool MetaboAnalyst 4.0 [34]. EMRTs with missing peak intensity values in more than 50% of the samples were removed, while remaining missing values for other EMRTs were imputed with small values. EMRTs were then filtered based on interquantile range to remove those EMRTs unlikely to be of use for modeling the data. Lastly, the peak intensities were normalized to the sample median, log transformed, and Pareto scaled. Subsequent statistical analyses were also performed using MetaboAnalyst 4.0 [34]. Significantly different EMRTs between low and normal phytic acid lines were defined as those with a log2 fold change >2.0 and a p-value (*t* test) <0.05. Peak identification was performed using fragmentation patterns obtained from MS^E and MS/MS analysis and databases such as Lipid Maps [35] and published literature.

Seed electrolyte conductivity testing

The soybean lines used for conductivity testing were *1mlpa*, 1MWT, *2mlpa*, 2MWT, 2MWT-L, and 2MWT-N. Conductivity measurements were taken following AOSA guidelines [36]. Seed moisture content was maintained between 10-14%. Each line was tested in quadruplicate, with each replicate comprised of thirty randomly selected seeds. Dry seed bulk weight was measured and recorded for each replicate. Seeds were submerged in 75 mL of distilled water at 20 °C and then covered to prevent evaporation. Conductivity was measured for each replicate 24 hours later using a Multi-Parameter PCSTestr 35 (Oakton®, Vernon Hills, IL) with each measurement being performed in triplicate. Readings were adjusted for distilled water conductivity. Triplicate readings were averaged for each replicate and then weight adjusted using the recorded dry seed bulk weight. The weight adjusted readings were then averaged for each line. Immediately following conductivity measurements, seed exudates were filtered through Miracloth and frozen in -20 °C until use.

Analysis of exuded proteins

Exudate from seeds of the four NILs (*2mlpa*, 2MWT, 2MWT-L, 2MWT-N) in the MRP genotypic subset was analyzed for protein content. The seed exudates from the seed conductivity experiment (4 mL each) were freeze-dried and suspended in MeOH and left overnight at -20 °C. The samples were then centrifuged, with the pellets (protein) taken up in 200 μL of Tris/Glycine SDS-PAGE running buffer, and the MeOH-soluble fraction (metabolites) taken to dryness for use in the metabolite assay described in the next section. Protein concentrations were determined, and the samples were mixed with Laemelli buffer (20 μL) as well as running buffer to normalize protein levels to 15 μg per lane. Prior to gel loading, the samples were then heated

for 5 min at 95 °C and centrifuged (13k x g, 1 min). The resulting supernatants (35 μL) were added to a 10.5-14% SDS-PAGE gel (Bio-Rad Criterion™, Hercules, CA) and separated using constant voltage (150 V). The gels were stained with ProtoBlue Safe for visualization and imaged on a ChemiDoc (Bio-Rad, Hercules, CA). Bands of interest were excised and processed (reduction, alkylation, trypsinization) to generate peptides that were analyzed by MALDI-TOF/TOF mass spectrometry. The resulting data set was queried against the NCBI database using the MASCOT search engine. All proteins reported provided scores well above the identity threshold.

Analysis of exuded metabolites

The MeOH-soluble material obtained after protein precipitation described above was used for this study. Samples were redissolved in 9:1 water:acetonitrile (v/v, containing 0.1% formic acid), sonicated for 10 min and subsequently centrifuged (13k x g, 10 min). Aliqouts were transferred to LC-MS vials for analyses, which utilized an Acquity I-class UPLC coupled with a Synapt G2-S HDMS (Waters Corp., Milford, MA). Analyses were performed in both positive and negative ionization modes. To condition the column, two blank injections were followed by three complete master mix injections, after which the entire sample set was then run in a randomized fashion.

Samples were separated with a binary solvent system of water with 0.1% formic acid (Solvent A) and acetonitrile with 0.1% formic acid (Solvent B) on an Acquity BEH C18 Column (1.8 μm, 2.1 x 50 mm, Waters Corp., Milford, MA) with a flow rate of 200 μl/min and a 10 minute gradient. The starting gradient condition (5% B) was held for 1 minute, then a linear gradient to 70% B (1-7 min), ramp to 95% B (7-7.5 minutes) then hold at 95% B (7.5-8 min),

200 return to initial conditions (8-8.5 min), and isocratic at 5% B (8.5-10 min). Sample injection
201 volume was 1 μl.

202 Column eluates were ionized by electrospray ionization and analyzed independently in
203 both positive and negative modes. Data was collected in high resolution, MS mode with a cycle
204 time of 0.20 sec, and a mass range of 50-1800 *m/z*. The source parameters for positive ion mode
205 were source temperature 125 °C, capillary voltage 3.0, cone voltage 30, source offset 80,
206 desolvation temperature 350 °C, cone gas 50 L/h, desolvation gas 500 L/h, and nebulizer gas 6.0
207 bar. The source parameters for negative ion mode were source temperature 125 °C, capillary
208 voltage 2.2, cone voltage 30, source offset 80, desolvation temperature 350 °C, cone gas 50 L/h,
209 desolvation gas 500 L/h, and nebulizer gas 6.0 bar. A reference sprayer released leucine-
210 enkephalin (200 ng/mL, Waters Corp.) continuously at 5 μl/min with a scan frequency of 20 sec.
211 Data processing was performed as described above using MarkerLynx to generate EMRTs and
212 Metaboanalyst 4.0 for further processing and statistical analyses. Significant EMRTs were
213 defined as those as those with a log2 fold change >2.0 and a p-value (*t* test) <0.05.
214
215 **RESULTS AND DISCUSSION**
216 ***Overview of seed lipid profiles***
217 In this study, we examined the effects of different mutations in the phytic acid synthesis
218 pathway on soybean seed lipid content. Four different soybean lines from two unique genotypic
219 subsets were evaluated – in the Mips genotypic subset, the *lpa* and low emergence line *1mlpa*
220 and its wild-type sibling line 1MWT, and in the MRP subset, the *lpa* and low emergence line
221 *2mlpa* and its wild-type sibling line 2MWT. UPLC-MS analysis was performed on lipids
222 extracted with anhydrous ethyl acetate from freeze-dried soybean powder. Triplicate random

injections in positive and negative ion mode of the five biological replicates from each line produced 120 individual LC-MS runs. Additional runs included 8 LC-MS runs of "master mixes" of each line and 16 LC-MS runs of complete master mixes composed of all four lines. The master mixes of each line and the complete master mix were used as quality checks between runs. The master mixes of each line also underwent fragmentation studies to assist in compound identification using data-independent acquisition mode (DIA, MS^E), yielding 8 more LC-MS runs. In total, 152 LC-MS runs were generated.

Automated feature detection was performed to generate exact mass-retention time pairs (EMRTs) using MarkerLynx (Waters). In the Mips subset, a total of 213 EMRTs were detected in positive mode, and 159 EMRTs were detected in negative mode (Table S1). In the MRP subset, a total of 199 EMRTs were detected in positive mode, and 167 EMRTs were detected in negative mode (Table S2). For each genotypic subset and ionization mode, the detected EMRTs and their raw peak intensities were further filtered and normalized, respectively, using Metaboanalyst 4.0 [34]. In the Mips subset, this resulted in 200 EMRTs in positive mode and 151 EMRTs in negative mode (Table S3). In the MRP subset, it yielded 189 EMRTs in positive mode and 158 EMRTs in negative mode (Table S4). These data sets were used for subsequent statistical analyses, and significantly different EMRTs between low and normal phytic acid lines were identified using significance thresholds with a *p*-value of <0.05 and a log2 fold change of >2.0 or <-2.0; the results of which will follow below.

Mutations in phytic acid synthesis and transport have little effect on lipids in positive ion mode

Principal component analysis (PCA) of the positive ion mode LC-MS data for both genotypic subsets revealed little separation between the *lpa* and normal phytic acid soybean lines

(Figure 3.1). This suggests that the *lpa* mutations have little effect on the triacylglycerol pool, as these metabolites were the predominant species in this ionization mode (Figure 3.2). Such a finding is in agreement with other observations that *lpa* causing mutations have relatively little impact on oil composition [5]. It should be noted that the abundance of triacylglycerols in positive ion mode limited sample injection amounts and could potentially mask other differences between low and normal phytic acid lines. Interestingly, both *lpa* lines, *1mlpa* and *2mlpa*, exhibited more biological variation relative the normal phytic acid lines. This is in accordance with the *lpa* phenotype in that some seeds germinate and emerge normally while others do not.

Significant differences in lipid profiles between low and normal phytic acid lines in negative ion mode

In contrast to positive ion mode, PCA of negative ion mode data for both genotypic subsets revealed more separation between the low and normal phytic acid soybean genotypes (Figure 3.3). This was especially apparent in the Mips subset where there was complete separation of the *1mlpa* and 1MWT clusters (Figure 3.3A). The greater separation found between *1mlpa* and 1MWT, as compared to *2mlpa* and 2MWT, can perhaps be attributed to differences in the *lpa* causing mutations. The loss-of-function mutation in MIPS1, as found in *1mlpa*, affects the beginning of the phytic acid biosynthesis pathway by blocking the synthesis of *myo*-inositol [20, 21]. Not only is *myo*-inositol needed for phytic acid biosynthesis, but it is also a required substrate in a number of other pathways as well [37]; therefore, depletion of the cellular *myo*-inositol pool could have a greater effect on primary metabolism. In contrast, in *2mlpa* from the MRP genotypic subset, the low phytic acid causing mutations in the two MRP transporter genes,

268 is associated with later steps in the phytic acid biosynthesis pathway (storage transport), so
269 metabolism may be disrupted to a lesser degree [38].

270 Among the 159 EMRTs detected in the Mips subset, 19 were significantly different
271 between *1mlpa* and 1MWT. A majority had significantly decreased content in *1mlpa* (Figure
272 3.4A), which could be a result of reductions in *myo*-inositol content. Out of the 167 EMRTs in
273 the MRP subset, 13 were significantly different between *2mlpa* and 2MWT (Figure 3.4B).
274 Interestingly, all but one of these EMRTs had significantly increased content in *2mlpa*; this could
275 be due to increases in the *myo*-inositol pool in *2mlpa,* where other *mrp* mutants, such as maize
276 *lpa1-1* and rice *Os-lpa-XS110-3*, are shown to have elevated *myo*-inositol levels [23, 39].

277 Lipid identification of the significant EMRTs found between *1mlpa* and 1MWT in the
278 Mips genotypic subset showed changes in glucose-sitosterol, peroxidized triacylglycerol (TAG),
279 and phospholipid contents (Table 3.2). Altered phospholipids include phosphoglycerol (PG) and
280 phosphatidylethanolamine (PE), which had significantly increased levels in *1mlpa*, and
281 phosphatidylinositol (PI), which had significantly reduced levels in *1mlpa*. The reduction in PI is
282 in accordance with the depleted *myo*-inositol pool. Given that PI plays an important role in cell
283 signaling and development [40-42], changes in its content could have an adverse effect on seed
284 germination and emergence potential. Alterations in PE content are notable as well considering
285 PE is a primary component of membranes, playing essential roles in membrane architecture and
286 creating structure-forming environments for membrane proteins; moreover, PE is a precursor for
287 several biologically active molecules, such as diacylglycerols, fatty acids, and phosphatidic acid
288 (PA), which function as second messengers [43, 44].

289 Lipid identification of significant EMRTs found between *2mlpa* and 2MWT from the
290 MRP genotypic subset showed changes in ceramide, glucose-sitosterol, peroxidized TAG, and

PA content (Table 3.3). The significantly increased ceramide and PA contents in *2mlpa* are interesting, as ceramides are regulators of programmed cell death in plants [45]. As for PA, Park et al. [46] demonstrated that during stress conditions PA levels increase and induce cell death during stress responses in plants. Further evidence shows PA acts a second messenger with its levels increasing in response to stresses such as oxidative stress, pathogen elicitors, abscisic acid, and wounding [47]. These findings on altered ceramide and PA contents suggest that the regulation of programmed cell death may be askew in *2mlpa*, which could negatively affect seed viability and ultimately seedling emergence.

Both *lpa* lines *1mlpla* and *2mlpa*, especially the former, had significant reduction in peroxidized TAG content. Transcriptomics data on seeds from these lines also showed significant changes in the expression of genes functioning in hydrogen peroxide-related activities, such as regulation of hydrogen peroxide metabolism and response to hydrogen peroxide [48]. Between *1mlpa* and 1MWT, 165 hydrogen peroxide-related genes were significantly differentially expressed, and between *2mlpa* and 2MWT, 13 were significantly differentially expressed. The greater expression change found in *1mlpa* reflects the larger number of changes observed in peroxidized TAGs found in this line.

Both lines also had changes in glucose-sitosterol content; however, in *1mlpa* levels were significantly decreased, while in *2mlpa* they were significantly increased. It is thought that sterol glucosylation modulates free sterol content of the plasma membrane, therefore directly influencing its physical properties. Sitosterols in particular are well known for their importance in membrane architecture and the regulation of membrane fluidity and permeability [49]. Thus changes in glucose-sitosterol content could affect cell membranes in *1mlpa* and *2mlpa*, or this is the result of perturbed membrane compositions.

Conductivity testing indicates lpa seeds have compromised membranes

For conductivity testing, seeds from soybean lines in the Mips and MRP genotypic subsets were soaked in distilled water to measure electrolyte leakage (exudate). Conductivity measurements were taken after 24 hours of soaking (Figure 3.5). In both *lpa* mutants (*1mlpa* and *2mlpa*), electrolyte conductivity was 2-fold higher than in comparison to corresponding normal phytic acid sibling lines. Overall, the MRP genotypic subset exhibited higher electrolyte conductivity than the Mips subset. This may be attributed to differences in genetic backgrounds. The elevated conductivity in the *lpa* mutants indicates an altered membrane composition relative to their normal phytic acid counterparts. The significance of this is that cell membrane integrity is considered the fundamental basis for the success of seed vigor [50]. When a seed rehydrates during early imbibition, cellular membranes must reorganize and repair any damage that occurs; the faster a seed is able to re-establish its membrane integrity, the better its vigor [36]. Compared to their normal phytic acid sibling lines, this ability appears to be impaired in *1mlpa* and *2mlpa* seeds. Previous LC-MS profiling on polar extracts of these seeds revealed significantly reduced levels in *2mlpa* of the allergen Gly m 1 [32], a hydrophobic seed protein (HSP) proposed to be involved in water uptake rates and absorption activities [51]. Whether low levels of HSP affect membrane organization and repair during seed imbibition remains to be seen, but it is clear that maintaining appropriate membrane composition is critical for several biological processes and properties [52]. Hence a compromised membrane could considerably affect normal cell functioning. Regardless, the increased electrolyte leakage observed in *1mlpa* and *2mlpa* indicates a reduction in seed vigor, which ultimately reduces seedling emergence.

337 ***2mlpa* has unique protein and metabolite seed exudate profiles**

338 Vastly different protein profiles, as determined by SDS-PAGE, were found between
339 *2mlpa* and its normal phytic acid sibling lines (Figure 3.6). One of the most striking differences
340 was in the relative levels of glycinin and β-conglycinin (vicilin), both of which had much lower
341 levels in the *2mlpa* exudate based on band intensity. Interestingly, transcriptomics data on dry
342 seeds show significantly increased expression in *2mlpa* for genes encoding vicilin-like seed
343 storage protein (PAP85) [48]; however, the significance of the leakage of these seed storage
344 proteins is unclear. Also, lower in abundance in the *2mlpa* exudate were Kunitz trypsin inhibitor
345 A and 9S-lipoxygenase-3, which is involved in oxylipin biosynthesis. Higher in abundance in the
346 *2mlpa* exudate were 7S globulin, a storage protein with kinase activity, and 24 kDa seed coat
347 protein (SC24), a novel plant defense protein with carboxylate-binding activity induced upon
348 wound stress.

349 To further characterize the seed exudates, untargeted UPLC-MS was used to profile the
350 exuded metabolites of the four NILs from the MRP subset. The exudate profile of the *lpa* line
351 (*2mlpa*) was highly distinct from the normal phytic acid lines (2MWT, 2MWT-L, 2MWT-N),
352 with complete separation found between them in both positive and negative ion mode (Figure
353 3.7A, 3.7B). For positive ion mode, 144 EMRTs were detected (Table S5), with 31 found as
354 significantly different between *2mlpa* and its three normal phytic acid sibling lines (Table S6). In
355 negative ion mode, 392 EMRTs were detected (Table S5), and 92 were significantly different
356 between *2mlpa* and its normal phytic acid sibling lines (Table S6). In both sets of significant
357 EMRTs, a majority were exuded at increased levels from *2mlpa* seeds. Further investigation is
358 needed into the identities of these significant EMRTs. What is known thus far is that a direct
359 correlation exists between the amount of seed materials exuded and seed vigor [53, 54]. Because

low vigor seeds have higher exudate leakage, soil microbial activity is promoted, making seeds more likely to develop secondary infection [55]. In fact, CX-1834, the *lpa* parent of *2mlpa*, is known to be more susceptible to disease infection [31]. These observations suggest the exuded proteins and metabolites of *2mlpa* could stimulate an environment conducive for disease infection, which could also contribute to the reduced emergence phenotype.

CONCLUSION

Untargeted lipidomic analyses enabled examination of the effects of *lpa* causing mutations *mips1* and *mrp-l/mrp-n* on the functional state of cells in soybean seeds. Seeds carrying the *mips1* mutation had major changes in PE content, suggesting altered membrane composition and signaling. Meanwhile, seeds carrying the *mrp-l/mrp-n* mutations had major changes in ceramide and PA contents, both of which are involved in the regulation of programmed cell death. In addition, both sets of mutations altered glucose-sitosterol content, which also has a critical effect on membrane properties. Changes in cell membranes and the regulation of programmed cell death were supported by observations that soybean seeds carrying the *lpa* causing mutations had significantly elevated conductivity, which also resulted in distinctive exudate profiles (the implications of which would require further work). Consequently, the *lpa* causing mutations may cause deviations in lipid profiles that contribute to reducing seed vigor and ultimately seedling emergence. Thus the changes in ceramide, glucose-sitosterol, PA, and PE levels and their relationship to phytic acid and seed vigor are worthy of further investigation.

ACKNOWLEDGEMENTS

Funding for this study was through the Virginia Soybean Board (VSB), as well as the John Lee Pratt Fellowship Program, the Biodesign and Bioprocessing Research Center (BBRC), and Open Access Subvention Fund, all three at Virginia Tech. Additional funding was provided by the Fralin Life Science Institute, as well as the Virginia Tech Agricultural Experiment Station Hatch and McIntire-Stennis Programs.

REFERENCES

1. USDA National Agricultural Statistics Service. USDA-NASS. 2018.
2. Stein HH, Berger LL, Drackley JK, Fahey Jr GC, Hernot DC, Parsons CM. Nutritional properties and feeding values of soybeans and their coproducts. In: Johnson LA, White PJ, Galloway R, editors. Soybeans. Urbana, IL: AOCS Press; 2008. p. 613-660.
3. Erdman J. Oilseed phytates: nutritional implications. Journal of the American Oil Chemists' Society. 1979; 56(8):736-41.
4. Lott JNA, Greenwood JS, Batten GD. Mechanisms and regulation of mineral nutrient storage during seed development. Seed Development and Germination. 1995; 41:215.
5. Raboy V. Approaches and challenges to engineering seed phytate and total phosphorus. Plant Science. 2009; 177(4):281-96.
6. Weaver CM, Kannan S. Phytate and mineral bioavailability. Food Phytates. 2002; 2002:211-23.
7. Brinch-Pedersen H, Sørensen LD, Holm PB. Engineering crop plants: getting a handle on phosphate. Trends in Plant Science. 2002; 7(3):118-25.
8. Brown K, Solomons N. Nutritional problems of developing countries. Infectious disease Clinics of North America. 1991; 5(2):297-317.
9. Cromwell G, Coffey R. Phosphorus-a key essential nutrient, yet a possible major pollutant-its central role in animal nutrition. Biotechnology in the Feed Industry. 1991:133-45.
10. Sharpley AN, Chapra S, Wedepohl R, Sims J, Daniel TC, Reddy K. Managing agricultural phosphorus for protection of surface waters: Issues and options. Journal of Environmental Quality. 1994; 23(3):437-51.

11. Guttieri M, Bowen D, Dorsch JA, Raboy V, Souza E. Identification and characterization of a low phytic acid wheat. Crop Science. 2004; 44(2):418-24.
12. Larson S, Young K, Cook A, Blake T, Raboy V. Linkage mapping of two mutations that reduce phytic acid content of barley grain. Theoretical and Applied Genetics. 1998; 97(1-2):141-6.
13. Larson SR, Rutger JN, Young KA, Raboy V. Isolation and genetic mapping of a non-lethal rice (Oryza sativa L.) low phytic acid 1 mutation. Crop Science. 2000; 40(5):1397-405.
14. Raboy V, Gerbasi PF, Young KA, Stoneberg SD, Pickett SG, Bauman AT, et al. Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiology. 2000; 124(1):355-68.
15. Wilcox JR, Premachandra GS, Young KA, Raboy V. Isolation of high seed inorganic P, low-phytate soybean mutants. Crop Science. 2000; 40(6):1601-5.
16. Cichy K, Raboy V. Evaluation and development of low-phytate crops. Modification of Seed Composition to Promote Health and Nutrition. 2009; 50:177-200.
17. Raboy V. Seed phosphorus and the development of low-phytate crops. Inositol Phosphates: Linking Agriculture and the Environment. 2007:111-32.
18. Saghai Maroof MA, Glover NM, Biyashev RM, Buss GR, Grabau EA. Genetic basis of the low-phytate trait in the soybean line CX1834. Crop Science. 2009; 49(1):69-76.
19. Saghai Maroof MA, Buss GR. Low phytic acid, low stachyose, high sucrose soybean lines. Google Patents; 2008.
20. Chappell A, Scaboo A, Wu X, Nguyen H, Pantalone V, Bilyeu K. Characterization of the MIPS gene family in Glycine max. Plant Breeding. 2006; 125(5):493-500.

452 21. Loewus FA, Loewus MW. Myo-inositol: its biosynthesis and metabolism. Annual
453 Review of Plant Physiology. 1983; 34(1):137-61.
454 22. Walker D, Scaboo A, Pantalone V, Wilcox J, Boerma H. Genetic mapping of loci
455 associated with seed phytic acid content in CX1834-1-2 soybean. Crop Science. 2006;
456 46(1):390-7.
457 23. Shi J, Wang H, Schellin K, Li B, Faller M, Stoop JM, et al. Embryo-specific silencing of
458 a transporter reduces phytic acid content of maize and soybean seeds. Nature
459 Biotechnology. 2007; 25(8):930.
460 24. Anderson BP, Fehr WR. Seed source affects field emergence of low-phytate soybean
461 lines. Crop Science. 2008; 48(3):929-32.
462 25. Bregitzer P, Raboy V. Effects of four independent low-phytate mutations on barley
463 agronomic performance. Crop Science. 2006; 46(3):1318-22.
464 26. Gao Y, Biyashev R, Saghai Maroof MA, Glover N, Tucker D, Buss G. Validation of low-
465 phytate QTLs and evaluation of seedling emergence of low-phytate soybeans. Crop
466 Science. 2008; 48(4):1355-64.
467 27. Guttieri M, Peterson K, Souza E. Agronomic performance of low phytic acid wheat. Crop
468 Science. 2006; 46(6):2623-9.
469 28. Kastl C. Metabolomic Discrimination of Near Isogenic Low and High Phytate Soybean
470 [GLYCINE MAX (L.) MERR. Lines: Virginia Tech; 2014.
471 29. Meis SJ, Fehr WR, Schnebly SR. Seed source effect on field emergence of soybean lines
472 with reduced phytate and raffinose saccharides. Crop Science. 2003; 43(4):1336-9.

30. Pilu R, Landoni M, Cassani E, Doria E, Nielsen E. The maize lpa241 mutation causes a remarkable variability of expression and some pleiotropic effects. Crop Science. 2005; 45(5):2096-105.

31. Spear JD, Fehr WR. Genetic improvement of seedling emergence of soybean lines with low phytate. Crop Science. 2007; 47(4):1354-60.

32. Jervis J, Kastl C, Hildreth SB, Biyashev R, Grabau EA, Saghai Maroof MA, et al. Metabolite profiling of soybean seed extracts from near-isogenic low and normal phytate lines using orthogonal separation strategies. Journal of Agricultural and Food Chemistry. 2015; 63(44):9879-87.

33. Welti R, Wang X. Lipid species profiling: a high-throughput approach to identify lipid compositional changes and determine the function of genes involved in lipid metabolism and signaling. Current Opinion in Plant Biology. 2004; 7(3):337-44.

34. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Research. 2018; 46(W1):W486-W94.

35. Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CR, Shimizu T, et al. Update of the LIPID MAPS comprehensive classification system for lipids. Journal of Lipid Research. 2009; 50(Supplement):S9-S14.

36. AOSA I. Seed vigor testing handbook. Association of Official Seed Analysts Contribution. 1983(32).

37. Loewus FA, Murthy PP. myo-Inositol metabolism in plants. Plant Science. 2000; 150(1):1-19.

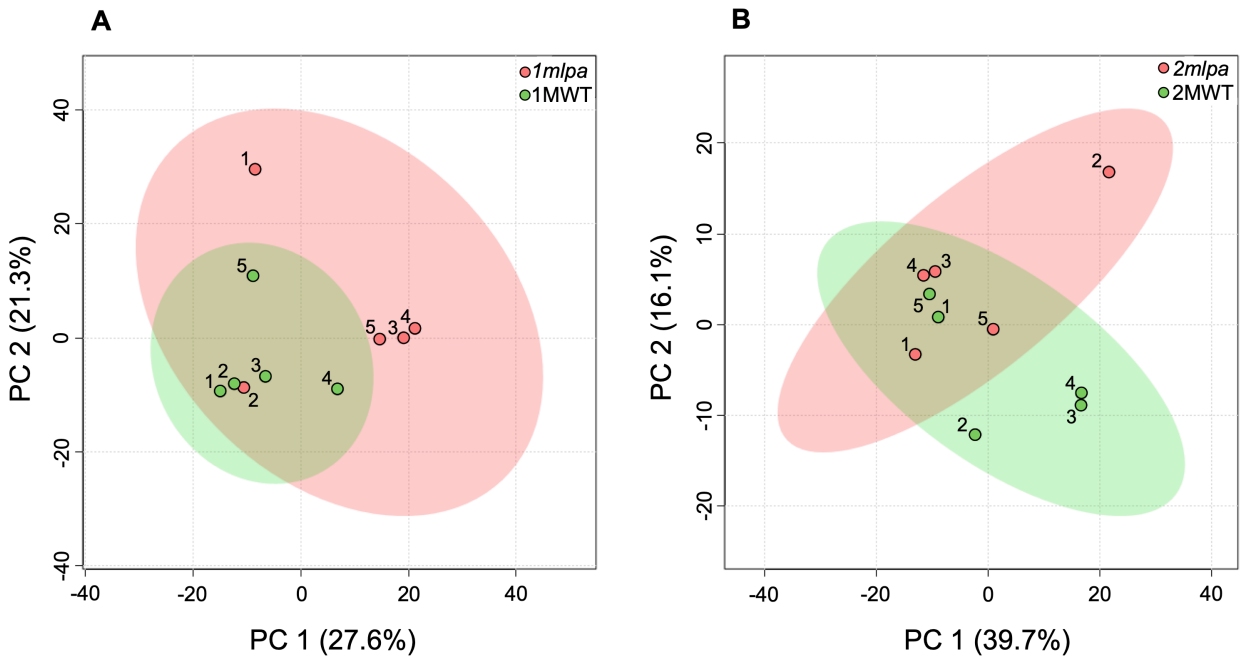
38. Raboy V. The ABCs of low-phytate crops. Nature Biotechnology. 2007; 25(8):874-5.

496 39. Xu X-H, Zhao H-J, Liu Q-L, Frank T, Engel K-H, An G, et al. Mutations of the multi-
497 drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic
498 acid in rice seeds. Theoretical and Applied Genetics. 2009; 119(1):75-83.
499 40. York JD, Odom AR, Murphy R, Ives EB, Wente SR. A phospholipase C-dependent
500 inositol polyphosphate kinase pathway required for efficient messenger RNA export.
501 Science. 1999; 285(5424):96-100.
502 41. Berridge MJ. Inositol trisphosphate and calcium signalling. Nature. 1993; 361(6410):315-
503 25.
504 42. Sasakawa N, Sharif M, Hanley MR. Metabolism and biological activities of inositol
505 pentakisphosphate and inositol hexakisphosphate. Biochemical Pharmacology. 1995;
506 50(2):137-46.
507 43. Gibellini F, Smith TK. The Kennedy pathway—de novo synthesis of
508 phosphatidylethanolamine and phosphatidylcholine. IUBMB Life. 2010; 62(6):414-28.
509 44. Momchilova A, Markovska T. Phosphatidylethanolamine and phosphatidylcholine are
510 sources of diacylglycerol in ras-transformed NIH 3T3 fibroblasts. The International
511 Journal of Biochemistry & Cell Biology. 1999; 31(2):311-8.
512 45. Liang H, Yao N, Song JT, Luo S, Lu H, Greenberg JT. Ceramides modulate programmed
513 cell death in plants. Genes & Development. 2003; 17(21):2636-41.
514 46. Park J, Gu Y, Lee Y, Yang Z, Lee Y. Phosphatidic acid induces leaf cell death in
515 Arabidopsis by activating the Rho-related small G protein GTPase-mediated pathway of
516 reactive oxygen species generation. Plant Physiology. 2004; 134(1):129-36.
517 47. Munnik T. Phosphatidic acid: an emerging plant lipid second messenger. Trends in Plant
518 Science. 2001; 6(5):227-33.

48. DeMers L, Raboy V, Li S, Saghai Maroof MA. Network inference of transcriptional regulation in germinating low phytic acid soybean seeds. Manuscript in preparation. 2020.
49. Hartmann MA. 5 Sterol metabolism and functions in higher plants. Lipid Metabolism and Membrane Biogenesis: Springer; 2004. p. 183-211.
50. Powell AA. Cell membranes and seed leachate conductivity in relation to the quality of seed for sowing. Journal of Seed Technology. 1986:81-100.
51. Gijzen M, Miller SS, Kuflu K, Buzzell RI, Miki BL. Hydrophobic protein synthesized in the pod endocarp adheres to the seed surface. Plant Physiology. 1999; 120:951-9.
52. Orešič M, Hänninen VA, Vidal-Puig A. Lipidomics: a new window to biomedical frontiers. Trends in Biotechnology. 2008; 26(12):647-52.
53. Keeling B. Soybean seed rot and the relation of seed exudate to host susceptibility. Phytopathology. 1974; 64:1445-7.
54. Matthews S, Bradnock W. Relationship between seed exudation and field emergence in peas and French beans. Horticultural Research. 1968; 8:89-93.
55. AOSA. Seed Vigor Testing Handbook: Association of Official Seed Analysts; 2009.

Table 3.1. Characteristics and classification of parental and experimental soybean lines.

Soybean Line	Genotypic Class Subset	Genotype	Phytic Acid	Emergence	Stachyose	Sucrose	Cross
V99-5089	-	*mips1*/MRP-L/MRP-N	Low	Low	Low	High	Parent
CX-1834	-	MIPS1/*mrp-l*/*mrp-n*	Low	Low	Normal	Normal	Parent
Essex	-	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	Parent
1mlpa	Mips	*mips1*/MRP-L/MRP-N	Low	Low	Low	High	Essex x V99-5089
1MWT	Mips	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	Essex x V99-5089
2mlpa	MRP	MIPS1/*mrp-l*/*mrp-n*	Low	Low	Normal	Normal	CX-1834 x V99-5089
2MWT	MRP	MIPS1/MRP-L/MRP-N	Normal	Normal	Normal	Normal	CX-1834 x V99-5089
2MWT-L	MRP	MIPS1/MRP-L/*mrp-n*	Normal	Normal	Normal	Normal	CX-1834 x V99-5089
2MWT-N	MRP	MIPS1/*mrp-l*/MRP-N	Normal	Normal	Normal	Normal	CX-1834 x V99-5089



562

563 **Figure 3.1.** Principal component analyses of the two genotypic subsets in positive ion mode.
564 Points are labeled by sample replicate number. **A.** Mips genotypic subset containing lines *1mlpa*
565 and 1MWT. **B.** MRP genotypic subset containing lines *2mlpa* and 2MWT.
566

567

568

569

570

571

572

573

574

575

576

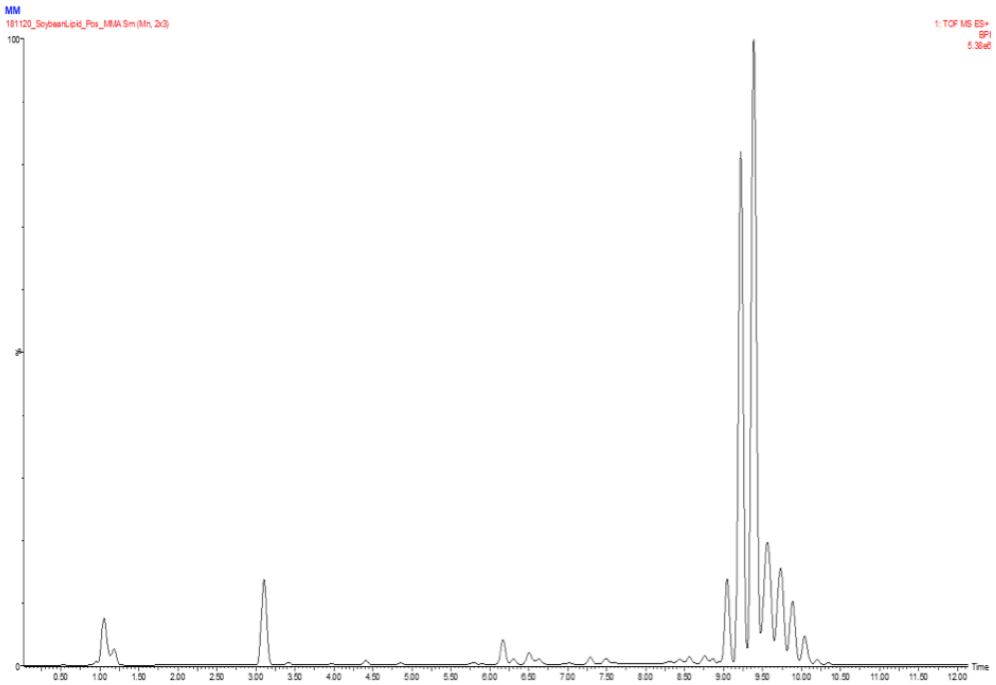


Figure 3.2. Base peak ion chromatogram of complete master mix of all four lines in positive ionization mode showing enrichment in triacylglyercide content.

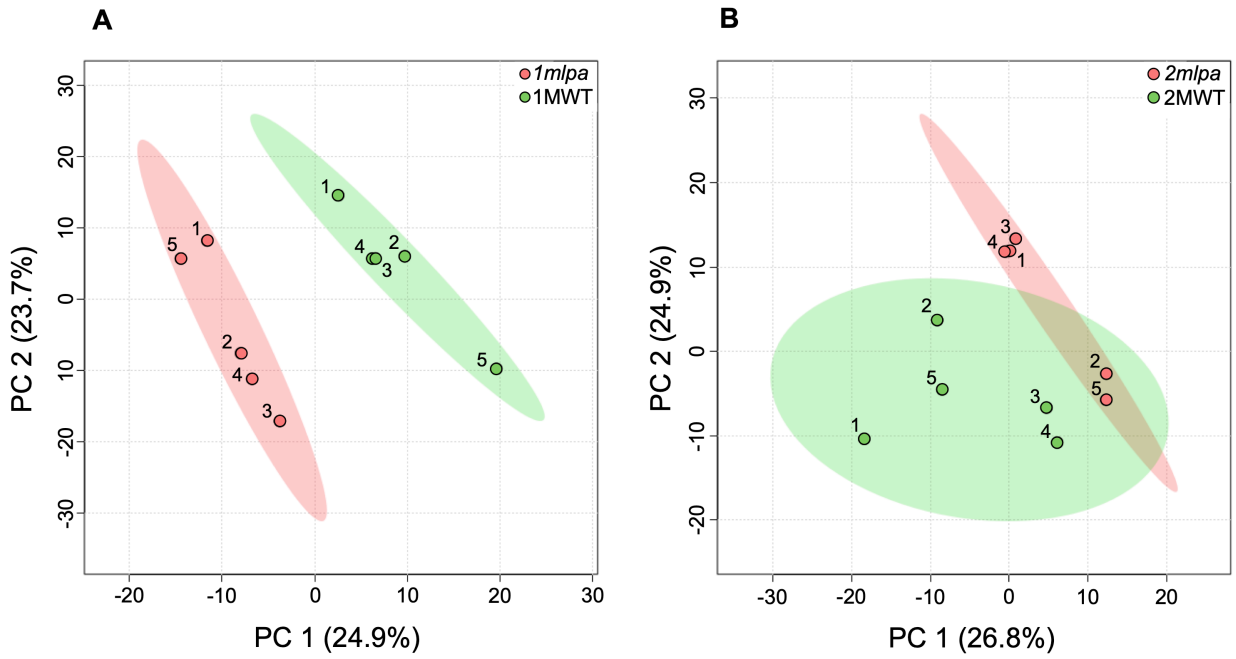
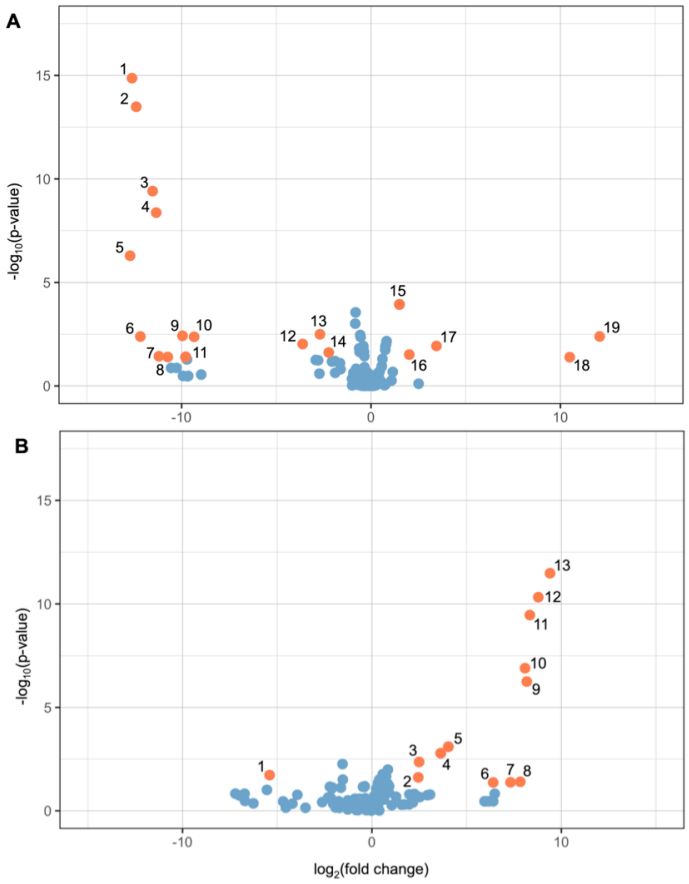


Figure 3.3. Principal component analyses of the two genotypic subsets in negative ion mode. In both genotypic subsets, there is separation between the low and normal phytic acid lines. Points are labeled by sample replicate number. **A.** Mips genotypic subset containing lines *1mlpa* and 1MWT. **B.** MRP genotypic subset containing lines *2mlpa* and 2MWT.



606

607 **Figure 3.4.** Volcano plots of EMRTs found in negative ionization mode for each genotypic
608 subset. **A.** Mips genotypic subset containing lines *1mlpa* and 1MWT. **B.** MRP genotypic subset
609 containing lines *2mlpa* and 2MWT. Orange circles represent significant EMRTs found between
610 low and normal phytic acid lines. Significant EMRTs are labeled with numbers that correspond
611 to EMRTs in Table 3.2 and Table 3.3. The significance threshold was set at a log2 fold change
612 >2.0 and a *p*-value (*t* test) <0.05.
613

614

615

616

617

618

619

Table 3.2. Significantly different EMRTs between low and normal phytic acid lines in the Mips subset from negative ion mode.

Volcano Plot Label	Compound	Retention Time (min)	Mass	Ion	Log2(fc) (low/normal)	*P*-value (*t* test)
1	-	8.67	899.6956	[M + OAc]-	-12.7	1.36E-15
2	-	8.07	879.634	[M + HCOO]-	-12.4	3.30E-14
3	18:2-Glc-Sitosterol	8.57	883.6653	[M + HCOO]-	-11.6	3.89E-10
4	Peroxidized TAG 54:5	7.93	957.7374	[M + HCOO]-	-11.3	4.25E-09
5	Peroxidized TAG 54:6	7.88	955.7221	[M + HCOO]-	-12.7	5.13E-07
6	-	6.1	835.5324	[M - H]-	-12.3	0.0041
7	18:0-Glc-Sitosterol	8.93	901.7111	[M + OAc]-	-11.2	0.0368
8	-	6.55	768.553	-	-10.8	0.0398
9	-	8.42	1722.3315	[2M + HCOO]-	-10.0	0.0040
10	Peroxidized TAG 54:3	8.38	961.7709	[M + HCOO]-	-9.3	0.0042
11	-	7.39	814.6407	[M - H]-	-9.9	0.0397
12	18:3-Glc-Sitosterol	8.2	895.6644	[M + OAc]-	-3.7	0.0094
13	PI 36:4 (18:2/18:2)	5.38	857.5168	[M - H]-	-2.8	0.0032
14	Peroxidized TAG 54:4	8.15	959.7524	[M + HCOO]-	-2.3	0.0242
15	PG 34:2 (16:0/18:2)	5.85	745.5008	[M - H]-	2.0	0.0001
16	PE 34:2 (16:0/18:2)	6.61	1430.0179	[2M - H]-	2.0	0.0305
17	-	7.84	1007.7296	[M - H]-	3.3	0.0117
18	-	5.87	756.5172	[M - H]-	10.4	0.0406
19	PE 36:4 (18:2/18:2)	6.29	1478.0186	[2M - H]-	12.0	0.0040

* "-" indicates unidentified EMRT compound.

Table 3.3. Significantly different EMRTs between low and normal phytic acid lines in the MRP subset from negative ion mode.

Volcano Plot Label	Compound	Retention Time (min)	Mass	Ion	Log2(fc) (low/normal)	*P*-value (*t* test)
1	Peroxidized TAG 54:6	7.88	955.7215	[M + HCOO]-	-5.2	0.0188
2	Cer(t42:1)	8.31	710.6281	[M + HCOO]-	2.6	0.0243
3	Cer(t40:1)	7.94	696.6126	[M + OAc]-	2.5	0.0044
4	18:2-Glc-Sitosterol	8.57	883.6638	[M + HCOO]-	3.8	0.0013
5	18:0-Glc-Sitosterol	8.93	901.7107	[M + OAc]-	4.2	0.0010
6	-	3.63	619.2946	[M - H]-	6.4	0.0429
7	-	3.7	776.4105	[M + HCOO]-	7.4	0.0421
8	-	3.5	746.4001	[M - H]-	7.9	0.0397
9	-	3.17	714.4105	[M - H]-	8.2	5.60E-07
10	18:1-Glc-Sitosterol	8.84	885.6792	[M + HCOO]-	8.0	1.27E-07
11	DGDG 36:6 (18:3/18:3)	5.34	981.5773	[M + HCOO]-	8.4	3.46E-10
12	PA 34:2 (18:2/16:0)	6.1	671.4651	[M - H]-	8.8	4.74E-11
13	PA 36:4 (18:2/18:2)	5.76	695.4642	[M - H]-	9.4	3.31E-12

* "-" indicates unidentified EMRT compound.

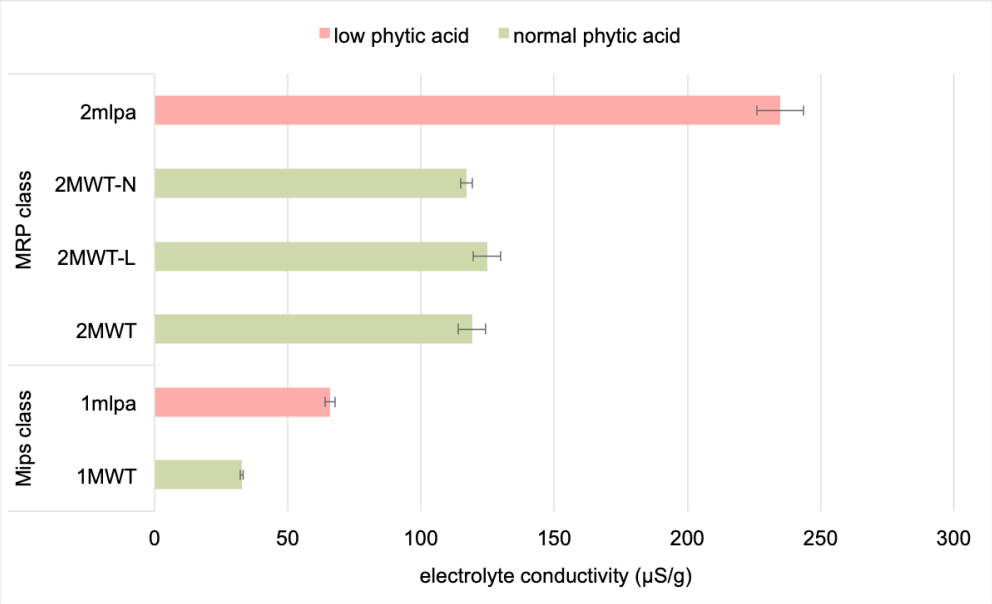


Figure 3.5. Electrolyte conductivity of low and normal phytic acid soybean seeds from each genotypic subset. Conductivity of distilled water containing seeds was measured 24 hours after imbibition. Data are presented as mean ±SE of four replicates, each of which was the mean of three technical measurements.

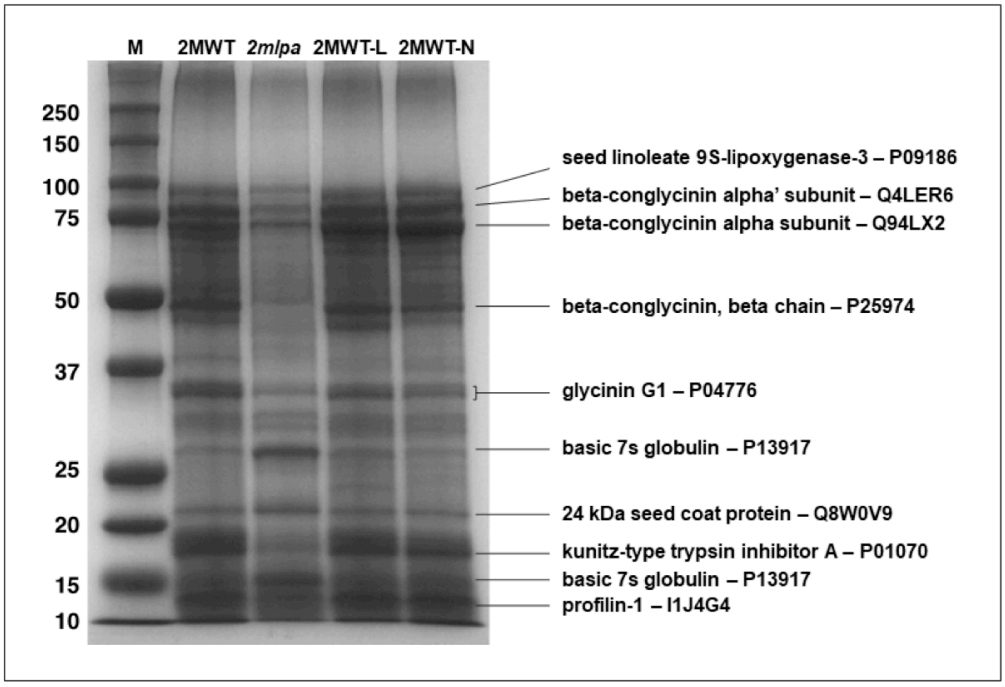


Figure 3.6. SDS-PAGE of exudate from seeds belonging to the MRP genotypic subset. Protein concentrations were determined and normalized prior to gel loading as described in Materials and Methods. The gel was stained with ProtoBlue Safe. Indicated bands were excised and analyzed by MALDI-TOF/TOF mass spectrometry for identification using MASCOT. M: molecular weight marker (kDa).

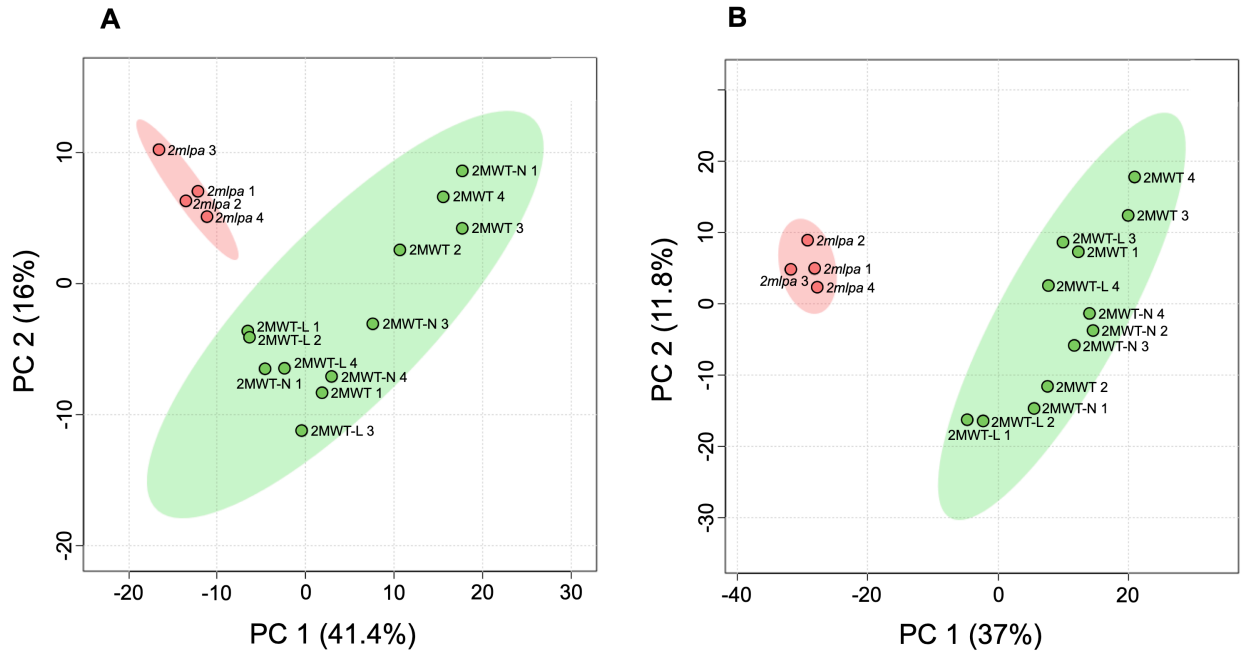


Figure 3.7. Principal component analyses of seed exudate from NILs belonging to MRP genotypic subset. Points are labeled by NIL and sample replicate number. **A.** Positive ion mode. **B.** Negative ion mode.

CHAPTER 4

A transcriptional regulatory network of *Rsv3*-mediated extreme resistance against *Soybean mosaic virus*

Lindsay C DeMers[1], Neelam R Redekar[1,#a], Aardra Kachroo[2], Sue A Tolin[1], Song Li[1], MA Saghai Maroof[1*]

[1] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia, United States of America. [2] Department of Plant Pathology, University of Kentucky, Lexington, Virginia, United States of America. [#a] Current Address: Department of Crop and Soil Science, Oregon State University, Corvallis, Oregon, United States of America. [*]Corresponding author: smaroof@vt.edu

This chapter was published in *PLOS ONE.*

ABSTRACT

Resistance genes are an effective means for disease control in plants. They predominantly function by inducing a hypersensitive reaction, which results in localized cell death restricting pathogen spread. Some resistance genes elicit an atypical response, termed extreme resistance, where resistance is not associated with a hypersensitive reaction and its standard defense responses. Unlike hypersensitive reaction, the molecular regulatory mechanism(s) underlying extreme resistance is largely unexplored. One of the few known, naturally occurring, instances of extreme resistance is resistance derived from the soybean *Rsv3* gene, which confers resistance against the most virulent *Soybean mosaic virus* strains. To discern the regulatory mechanism underlying *Rsv3*-mediated extreme resistance, we generated a gene regulatory network using transcriptomic data from time course comparisons of *Soybean mosaic virus*-G7-inoculated resistant (L29, *Rsv3*-genotype) and susceptible (Williams82, *rsv3*-genotype) soybean cultivars. Our results show *Rsv3* begins mounting a defense by 6 hpi via a complex phytohormone network, where abscisic acid, cytokinin, jasmonic acid, and salicylic acid pathways are suppressed. We identified putative regulatory interactions between transcription factors and genes in phytohormone regulatory pathways, which is consistent with the demonstrated involvement of these pathways in *Rsv3*-mediated resistance. One such transcription factor identified as a putative transcriptional regulator was MYC2 encoded by Glyma.07G051500. Known as a master regulator of abscisic acid and jasmonic acid signaling, MYC2 specifically recognizes the G-box motif ("CACGTG"), which was significantly enriched in our data among differentially expressed genes implicated in abscisic acid- and jasmonic acid-related activities. This suggests an important role for Glyma.07G051500 in abscisic acid- and jasmonic acid-derived defense signaling in *Rsv3*. Resultantly, the findings from our network offer insights into

genes and biological pathways underlying the molecular defense mechanism of *Rsv3*-mediated extreme resistance against *Soybean mosaic virus*. The computational pipeline used to reconstruct the gene regulatory network in this study is freely available at https://github.com/LiLabAtVT/rsv3-network.

KEYWORDS

Soybean mosaic virus, *Rsv3,* extreme resistance, gene regulatory network (GRN), unsupervised machine learning, MYC2, ABA pathway, JA pathway

INTRODUCTION

Soybean is a crop of global importance, and the *Soybean mosaic virus* (SMV)-soybean pathosystem provides an opportunity to study the extreme resistance (ER) response, a type of resistance unique from the typical hypersensitive reaction (HR) response in that it is triggered earlier and cell death is not observed [1]. SMV, a single-stranded RNA virus of the genus *Potyvirus*, considerably reduces seed quality and yield in soybean-growing regions throughout the world. Several SMV isolates recovered from germplasm imported into the United States were classified into seven strain groups, G1 to G7, based on reactions in a set of various soybean genotypes [2]. The most successful management strategies have been the utilization of virus-free seeds and resistant cultivars carrying resistance (*R*) genes. Four dominant *R* genes have been identified - *Rsv1, Rsv3, Rsv4,* and *Rsv5* [3-8]. *Rsv1* and *Rsv3* confer ER against SMV strains G1 to G4 and G5 to G7, respectively [5, 9, 10]. Among these strains, G5 to G7 represent the most virulent SMV strains, making *Rsv3* a particularly interesting gene for functional study. The *Rsv3* locus has been mapped, and the gene responsible for conditioning *Rsv3*-mediated resistance

(Glyma.14g204700; Glyma.Wm82.a2.v1 gene model) has been identified [11-13]. Comparative sequence analysis has revealed that Glyma.14g204700 is highly polymorphic in the LRR domain of soybean lines carrying *Rsv3*. This suggests *Rsv3*-mediated resistance is initiated by the LRR domain's recognition of an effector, the SMV cylindrical inclusion protein (CI) [12, 14]. However, the events directly following recognition remain undefined. It is hypothesized in [15] that the abscisic acid (ABA) signaling pathway is triggered during later stages of the *Rsv3*-mediated ER response. The consequent high ABA levels induce expression of a family of type 2C protein phosphatases, resulting in callose deposition, which impedes viral cell-to-cell movement [15]. Nonetheless, a large gap remains in our understanding of the *Rsv3*-mediated ER response, as the initial molecular events occurring prior to activation of the ABA signaling pathway are still unknown.

One approach to discerning the underlying mechanisms controlling a biological process, such as in *Rsv3*-mediated resistance, is reconstructing and modeling its molecular network. These networks examine complex interactions between genes, proteins, and metabolites. At the gene level, expression is predominantly governed by transcription factors (TFs), which bind to DNA sequence motifs in the regulatory region of their target genes. Improved understanding of gene expression regulation can have considerable scientific impact as many of the biological control mechanisms responsible for certain traits are associated with mutations in regulatory regions or dysfunctional transcriptional regulators [16]. For example, modern-day crops such as maize, rice, and wheat were heavily shaped by alterations in transcriptional regulation [17]; accordingly, elucidation of transcriptional regulation can aid significantly in research. An approach to accomplish this is the utilization of gene regulatory networks (GRNs), the study of which has led to the discovery of important genes and regulatory mechanisms underlying

specific processes in *Escherichia coli*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana* [18-23]. GRNs describe the intricate web of TFs that bind regulatory regions of target genes in order to influence their spatial and temporal expression [24]. Using computational network inference methods, the structure of the gene regulatory interactions that makeup GRNs can be reverse-engineered. That is, causal relationships can be inferred between genes (such as those encoding TFs) directly controlling the expression of other genes [25, 26]. By taking advantage of advancements in high-throughput sequencing technology, GRNs can be reconstructed utilizing genome-wide expression data, such as from RNA sequencing (RNA-seq) [27]. RNA-seq analyses can identify thousands of genes with altered expression in response to virus inoculation and provide more molecular targets to study. Network inference methods can then be applied to the expression data to uncover key genes and regulatory relationships [16]. Thus, the significance of modeling transcriptional regulation is that it provides a means for discerning gene function and important regulators in molecular pathways, such as those involved in mediating the *Rsv3*-mediated ER response.

This study aims to elucidate the key regulatory components involved in the *Rsv3* defense mechanism by constructing a GRN. To do this, we performed a comparative transcriptomic time course analysis of SMV-G7-inoculated cultivars "L29" (*Rsv3*-genotype) and "Williams82" (*rsv3*-genotype) during the early hours post-inoculation. We found differentially expressed genes (DEGs) between L29 and Williams82 at each time point, and among these were several genes belonging to TF families associated with defense. We carried out GRN inference analyses on DEGs utilizing the computational pipeline we developed previously [28]. This pipeline makes use of the well-received module networks method in which GRNs are inferred between TFs and gene co-expression modules. Network inference was performed with unique unsupervised

learning algorithms: ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks), context likelihood of relatedness (CLR), least angle regression (LARS), partial correlation, and Random Forest [29-33]. These algorithms represent the top performing inference methods according to the DREAM5 benchmark challenge [34]. Several of the predicted interactions were validated using published interactions in the model plant species, *A. thaliana,* and by motif sequence analysis [35-37].

MATERIALS AND METHODS

Soybean mosaic virus inoculations, leaf sampling, and RNA extraction

For this study, we used SMV strain G7 (SMV-G7) inoculum originating from [2]. The inoculum was stored in the form of desiccated infected leaves for long-term storage at 5°C or frozen at -80°C. Response of differential cultivars for "trueness to type" was tested periodically as inoculum were activated from storage. In this study, the SMV-G7 strain was maintained on greenhouse-grown soybean cultivar "York" (*rsv3*-genotype "susceptible") prior to the experiment. The SMV-G7 inoculum was prepared from symptomatic trifoliolate leaves of York by crushing in a mortar and pestle with 0.01M sodium phosphate buffer – pH 7.0 (1:10 w/v). The inoculation experiment was performed in greenhouse in the spring of 2014, where temperature, humidity, and light conditions were not artificially controlled. Inoculations were performed by lightly dusting 600-mesh carborundum powder over unifoliolate leaves, and the virus inoculum (see above) was gently rubbed using a pestle onto the two unifoliolate leaves of each plant and followed by a gentle rinsing with tap water. The inoculated unifoliolate leaves were collected at 0, 2, 4, 6, and 8 hours post inoculation (hpi) in biological triplicate, rinsed with DI water, frozen immediately by immersing in liquid nitrogen, and stored at -80°C until RNA extraction. For each

time point, a single biological replicate sample was comprised of six unifoliolate leaves total (= 2 unifoliolate leaves per plant x 3 individual plants within a pot). Thus 15 plants (= 3 plants per time point x 5 time points) were sampled from both cultivars. Total RNA (RIN >7.0) was extracted from frozen samples using RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) with on-column DNase digestion (QIAGEN, Hilden, Germany). A total of 20 mRNA libraries (= 2 cultivars x 5 time points x 2 biological replicates) was prepared from duplicate RNA samples of each virus-inoculated cultivar at each time point and sequenced as 150 PE with Illumina HiSeq4000 (Illumina, San Diego, CA) at Novogene, Sacramento, CA.

Sequence data processing and differential gene expression

Raw reads were filtered using Trimmomatic (version 0.30) to remove adapter sequences (ILLUMINACLIP:<IlluminaAdapters.fa>:2:30:10), trim low quality bases (<Q30, LEADING:30 TRAILING:30), and remove those reads trimmed to less than 50 base pairs (MINLEN:50) [38]. Reads were mapped to the "Williams82" soybean reference genome (Wm82.a2.v1, downloaded from Phytozome) using STAR (version 2.5.3a) with a maximum intron length of 15000 (--alignIntronMax) [39, 40]. The number of reads mapped to each gene was quantified using featureCounts (version 1.5.3) using paired end parameters "-B" and "-p" with features defined as "exons" (-t) being grouped by "gene_id" (-g) [41]. Differential expression analysis was performed with DESeq2 (version 1.22.2) in R (version 3.5.1) with those genes having less than one count being removed [42]. Reference levels were set as the susceptible Williams82 line and 0 hpi, and the DESeq() function "test" parameter was set to "LRT". The resulting output was used to make comparisons between L29 and Williams82 to identify DEGs at each time point by employing the results() function with the "test" parameter set as "Wald". DEGs were defined as

those with a false discovery rate (FDR) adjusted p-value < 0.05, $\log_2$ fold change >|1.0|, and base mean >10. DEGs and their $\log_2$ fold changes can be found in Table S4.1. The RNA-seq data from this study are available at the NCBI Gene Expression Omnibus (GEO) repository under accession number GSE137263.

Inference of gene regulatory networks

Expression clustering and gene function annotation

Gene expression levels for all genes were normalized by variance-stabilizing transformation in DESeq2 and averaged across replicates [42]. Clustering analysis was carried out on DEGs using Gaussian-finite mixture modeling with the R package, mclust (version 5.4.2) using default parameters [43]. The optimal clustering model was determined using Bayesian Information Criteria (BIC) and integrated complete-data likelihood (ICL) criterion [44, 45]. The top performing model identified five gene clusters. Gene ontology (GO) enrichment analysis was performed on each gene cluster using soybean GO annotations from [46]. Significantly enriched GO categories were selected using Fisher's exact test with FDR <0.05 (Table S4.2) Significantly enriched gene families were also analyzed using GenFam online tool, and the results with FDR <0.05 are included (Table S4.2) [47]. DEGs encoding TFs were identified using TF annotations downloaded from PlantTFDB [48].

Network inference methods

Network inference was carried out following the pipeline we developed previously using machine learning methods [28]. Gaussian-finite mixture modeling was used to cluster DEGs, with the best model finding five clusters (gene modules). We identified 131 differentially

883 expressed TFs, which were set as putative regulators of the five modules. The mean expression
884 profile for each module was computed and then constructed into an expression matrix of these
885 values and the expression levels of the 131 TFs. Putative regulatory interactions between each
886 TF and gene module were inferred from the expression matrix by implementing five unique
887 network inference algorithms: ARACNE, CLR, LARS, partial correlation, and Random Forest
888 [29-33]. ARACNE and CLR inference methods were implemented with the R package minet
889 (version 3.40.0) with the "estimator" parameter set as "spearman" and the "eps" parameter set as
890 0.1 for ARACNE and for CLR the "estimator" set as "pearson" [30, 31, 49]. The LARS
891 inference method was implemented with the R package tigress (version 0.1.0) with
892 "nstepsLARS" set at 4 [33]. The partial correlation inference method was implemented with the
893 R package GeneNet (1.2.13) using the "dynamic" shrinkage method [29, 50]. Lastly, the
894 Random Forest inference method was implemented with the R package GENIE3 (version 1.4.3)
895 with all default parameters [32]. Because community-based approaches make for a more robust
896 inference of GRNs, multiple inference methods, based on a diverse set of algorithms, were
897 applied to predict interactions. These methods were among the top performing in the DREAM5
898 challenge [34].
899
900 ***Validation of inferred network interactions***
901 We used two approaches to validate the discovered putative regulatory interactions
902 predicted by the inference methods. The first approach entailed the identification of homologous
903 regulatory interactions in *A. thaliana* using a comprehensive set of published *A. thaliana*
904 interactions observed with DNA affinity purification sequencing (DAP-seq) [35]. This DAP-seq
905 dataset is composed of 2.8 million interactions between 387 TFs and 32,605 genes. For

comparison of our predicted regulatory network with the *A. thaliana* DAP-seq data, we first expanded the TF-module interactions to TF-gene interactions. That is, each TF was set as a putative regulator of all the genes in the modules it was predicted to regulate. Homologous *A. thaliana* interactions for the TF-gene interactions were generated by using BLAST to identify *A. thaliana* homologous genes with soybean gene coding sequences. The best one-to-one BLAST hits were selected, using an E-value of 1e-5 for cut off. The resulting homologous *A. thaliana* interactions were then compared to the DAP-seq dataset and matching interactions identified.

For the second method of network validation, we performed motif sequence analysis using Meme suite (version 5.0.4), which provides a set of tools for motif discovery, enrichment, scanning, and comparison [36]. With this approach, we identified putative TF binding sites in promoter regions (defined as the 1000 bps flanking a gene's 5' end) of the DEGs in each module. These binding sites (motifs) were identified using the motif discovery tool, MEME [37]. The TomTom tool was then used to compare the discovered motif sequences to 872 *A. thaliana* motifs found with DAP-seq and to identify TFs that may bind to those discovered sequences [35, 51].

RESULTS AND DISCUSSION

In this study, we analyzed the transcriptional regulation of the R gene Rsv3, which confers ER against the most virulent SMV strains. This was accomplished by implementing machine learning inference algorithms on a GRN constructed from time course RNA-seq data from leaves of SMV-G7 inoculated resistant and susceptible soybean cultivars, L29 and Williams82, respectively. Our results suggest that an intricate regulatory network is in place modulating the Rsv3-mediated resistance response upon SMV-G7 inoculation.

Fate of SMV-induced susceptibility or resistance in soybean is determined between 4 to 8 hours post-inoculation

To better understand the regulatory mechanism underlying Rsv3-mediated ER, we compared transcriptomic profiles of SMV-G7 inoculated leaves from L29 and Williams82 cultivars at 0, 2, 4, 6 and 8 hpi. Overall, 1128 genes were differentially expressed between two cultivars, at one or more time points between 2 and 8 hpi (Table S4.1); DEGs identified at 0 hpi were excluded, as they were considered effects from differences in genetic backgrounds between the two cultivars. Distribution of the 1128 DEGs found between 2 and 8 hpi is shown in Figure 4.1. The majority of transcriptomic changes occurred between 4 and 8 hpi, suggesting that the large shifts in transcriptional activity during this time frame may be critical to whether a susceptible or defense response is induced. There was a striking increase in the number of DEGs at 6 hpi (859 DEGs), accounting for more than 75% of the total number of DEGs. This was followed by a dramatic drop at 8 hpi to merely 17 DEGs. This likely implies the presence of a tightly defined regulatory system that elicits the *Rsv3*-mediated ER response, suggesting the *Rsv3* pathway is induced very early during the infection process and that a susceptible or resistant response to SMV may be determined by 6 hpi.

At 6 hpi, GO enrichment analyses revealed that the 122 DEGs highly expressed in L29 were involved in cytokinin metabolism and signaling. Also highly expressed was a unique subfamily of MYB-related TFs, the RADIALIS-LIKE SANT/MYBs (RSMs). Up-regulation of six differentially expressed members of this family, specifically at 6 hpi, suggests tight temporal regulation of RSM TFs, which could be important to a process essential in ER-mediated defense. Little is known about the RSM subfamily, but one study showed involvement of RSM1 in auxin

signaling [52]. No other TF family was exclusively highly expressed or had multiple members up-regulated at this time. Interestingly, more than 85% of the DEGs in this time period (4-8 hpi) were expressed at lower levels in L29 as compared to Williams82. At 6 hpi, most of the down-regulated genes were those responsive to water deprivation, light absence, sucrose starvation, genes encoding stress-related proteins, such as multiple glutathione S-transferases, heat shock and LEA (late embryogenesis abundant) chaperones, and proteins related to oxidative stress and signaling, such as transporters, serine/threonine kinases, and receptor kinases. Additionally, a number of genes in the ABA signaling and the salicylic acid (SA) pathways were down-regulated in L29 as well. This finding is unique in that the activation of the SA pathway and exogenous application of SA are both widely recognized as enhancing resistance to viruses [53]. Nevertheless, a few exceptions to this phenomenon have been observed; in inoculated and systemically infected leaves of soybean, SA treatment had no effect on *Bean pod mottle virus* (BPMV) accumulation, and in susceptible pea cultivars, activation of the SA pathway resulted in an increase of *Clover yellow vein virus* virulence [54, 55]. Nonetheless, it remains unclear how SA, in some cases, enhances virulence [53], suggesting that suppression of the SA pathway may be a facet of *Rsv3'*s mechanism for diverting SMV-G7 infection.

Biological processes associated with Rsv3-mediated resistance in soybean show differential hormone responses

In order to study the temporal regulation of the *Rsv3*-mediated ER mechanism, we performed co-expression clustering of DEGs. The 1128 DEGs found between the two cultivars at one or more time points between 2 and 8 hpi were clustered into different co-expressed modules using a model-based clustering approach, where a module is defined as a group of genes sharing

975 similar expression profiles over time and are likely functioning in the same biological processes.
976 Based on BIC and ICL criteria, we identified five modules that optimally explain the observed
977 gene expression pattern; these modules consist of 85 (module-1), 198 (module-2), 383 (module-
978 3), 170 (module-4), and 292 (module-5) DEGs. The expression profile for these modules was
979 determined by averaging the expression levels of DEGs within each module (Figure 4.2A). The
980 expression profiles for module-1, module-4, and module-5 were similar between L29 and
981 Williams82, whereas those for module-2 and module-3 were highly divergent between the two
982 cultivars. This divergence in their expression pattern was noticeable between 4 and 8 hpi, with a
983 peak at 6 hpi. For module-5, despite similar expression patterns, the magnitude of difference
984 between L29 and Williams82 was greater in Williams82 than in L29.

985 GO enrichment analyses of five co-expression modules showed significant enrichment of
986 47 biological processes (shown with asterisk) and molecular functions (Figure 4.2B) (Table
987 S4.2). The co-expression module-2 showed enrichment for several GO terms associated with
988 ABA and auxin biosynthesis and signaling pathways (Figure 4.2B). The expression profile of
989 this module showed a clear contrast between L29 and Williams82, with a maximum (4-fold)
990 difference at 6 hpi, suggesting that ABA- and auxin-related processes were likely down-
991 regulated in SMV-resistant L29 soybean between 4 and 8 hpi (Figure 4.2A). [15] found that
992 ABA-mediated callose deposition in cell walls prevents intercellular virus movement in *Rsv3*-
993 mediated ER in SMV-G5H inoculated L29 after 8 hpi. Callose deposition was not observed in
994 SMV-G7 inoculated L29 (this study); however, Glyma.16152600 and Glyma.03G132700, both
995 encoding beta-1,3-glucanases, were down-regulated at 6 hpi in L29. This is interesting as one of
996 ABA's defense strategies against viruses is inhibition of these proteins, which function to
997 degrade callose [56]. The down-regulation in L29 of genes encoding callose degradation proteins

998 provides further evidence that *Rsv3* begins mounting a defense as early as 6 hpi. Additionally,
999 [15] showed elevated expressions of ABA and ABA responsive genes in SMV-G5H inoculated
1000 L29 leaves after 8 hpi. In contrast, we observed down-regulation of ABA responsive genes in
1001 SMV-G7 inoculated L29 leaves before 8 hpi, indicating changes in ABA signaling begin soon
1002 after inoculation.

1003 Co-expression module-4 showed enrichment of several GO terms associated with
1004 jasmonic acid (JA) biosynthesis and signaling and ethylene (ET) biosynthesis (Figure 4.2B).
1005 Module-4 expression showed similar profiles between the two cultivars but average expressions
1006 were lower in L29 than in Williams82 at 4, 6, and 8 hpi, suggesting JA suppression may be
1007 required for *Rsv3*-mediated ER (Figure 4.2A). Suppression of JA pathway in *Rsv3*-mediated
1008 resistance was also reported in SMV-G5H inoculated L29 cultivar [56]. Though JA's role in
1009 viral defense is not well understood, [43] observed that increased JA levels in soybean enhance
1010 susceptibility to BPMV. Interestingly, co-expression module-5 was enriched with genes
1011 associated with biological processes such as for syncytium formation (GO:0006949), cell wall
1012 modifications (GO:0009828, GO:0009831), cytokinin (CK) degradation (GO:0009823,
1013 GO:0019139), and cell growth (GO:0009826). Enrichment for these processes is indicative of
1014 virus interference with cell growth and metabolism. As for the expression profile of this module,
1015 it fluctuated drastically from 2 hpi to 8 hpi in Williams82 compared to the subtle shifts in L29.
1016 This may indicate greater changes in the activity of these biological processes in Williams82,
1017 which are perhaps associated with soybean susceptibility to SMV and stages of virus replication
1018 occurring as early as 4 hpi (Figure 4.2A).

1019 For the enrichment in CK degradation, multiple genes encoding cytokinin
1020 dehydrogenases were up-regulated in L29 from 2 to 6 hpi, suggesting CK levels were reduced in

1021 L29 relative to Williams82. CKs function to promote cell proliferation and elongation, numerous
1022 developmental processes, and are known to have a role in viral resistance [53]. In Williams82,
1023 the large expression changes in genes involved in membrane activity, syncytium formation, cell
1024 wall loosening, and cell growth and modification are known to be associated with early and
1025 initial stages of the potyvirus infection process in susceptible hosts [57, 58]. In particular,
1026 syncytium formation is a biological process in which virus-infected cells fuse together to form
1027 enlarged multi-nucleated cells called syncytia [59]. The increase in gene products used to form
1028 syncytia, which are not known to occur in cells of potyvirus-infected plants, may reflect the
1029 initiation of virus replication in the susceptible host, Williams82, as it did not occur in L29. After
1030 all, potyviruses are known to form 6K2 membrane-bound vesicles that later form tubular
1031 structures and interact with host endoplasmic reticulum [60]. This response could have been
1032 facilitated by heightened CK levels in Williams82. Interestingly still, CKs can act synergistically
1033 with the SA signaling pathway, triggering its activation [53]. In fact, [61] proposed that CK
1034 levels might aid in determining the amplitude of SA-related immunity. Perhaps in the case of
1035 soybean *Rsv3*-mediated resistance, where it seems suppression of the SA pathway is required,
1036 this suppression is achieved through reduced CK levels.

1037 Only single biological processes such as responses to sucrose starvation and absence of
1038 light were enriched for the co-expression module-1 and module-3, respectively, but the analyses
1039 of these modules will not be included in this study. We also analyzed gene family enrichment
1040 using an online tool, GenFam [47]. We found that some results are in agreement with the GO
1041 analysis. In particular, GenFam found that “Kunitz Trypsin Inhibitor (KTI) gene family” is
1042 enriched in module-2, whereas GO analysis showed (GO:0004866) endopeptidase inhibitor
1043 activity is also enriched in module-2. This result from GenFam is more specific than GO

annotation because KTI is a specific type of endopeptidase inhibitor. Similarly, we also found "Expansin gene family" is enriched in module-5, whereas GO analysis showed (GO:0009828) plant-type cell wall loosening is also enriched in module-5. Although many factors might regulate plant-type cell wall loosening, the results from GenFam enrichment provide a more specific result suggesting expansin genes are the main gene family contributing to cell wall loosening in our experiment.

Suppression of MYC2 transcription factor expression is important for Rsv3-mediated ER

Our network inference analysis identified candidate genes regulating gene expression in each module. Between the five network inference methods, a total of 654 interactions were identified between TF genes and the gene co-expression modules. No interaction was predicted by all five methods, but 56 interactions were predicted by four out of five methods (Table S4.3). These 56 TF-module interactions were regulated by 49 TFs, indicating some TFs regulated more than one module, and all five modules were regulated by more than one TF. Because there could be an unknown number of false negatives (true interactions that were not supported by expression data) and false positives (interactions supported by expression data but not found in biological systems) in the predicted interactions, we chose to use bioinformatics approaches to validate our computational predictions. In the rest of this manuscript, we focused on the predicted interactions that are supported by homologous interactions in the model species, *A. thaliana,* and also analyzed the motif enrichment to compare with known motifs in *A. thaliana*.

When the 56 putative interactions were transformed to homologous *A. thaliana* interactions, comparison to the *A. thaliana* DAP-seq dataset validated 1732 TF-gene interactions, with 21 TFs and 819 genes (Table S4.4). This translates to 25 TF-module interactions found

from the network inferred 56 TF-module interactions (Table S4.5). Further validation by motif sequence analysis discovered 20 enriched motifs in the five modules, with each module containing enrichment of one or more motifs (Table S4.6). The identified motifs represent putative TF binding sites from which TFs can regulate the expression of target genes in each of the modules; this allowed us to identify TF families that may recognize and bind to the enriched motif sequences. From the 25 TF-module interactions validated with the *A. thaliana* DAP-seq data, we found nine interactions further validated by motif sequence analyses (Table 4.1). Still, though the *A. thaliana* DAP-seq dataset is large, it does not represent every interaction; therefore, we included three additional interactions from the inferred 56 TF-module interactions that were validated by motif enrichment only.

Motif sequence analyses showed that co-expressed genes in module-5 are regulated by NAC (NAM, ATAF1/2, and CUC), ERF (ethylene responsive factor) and/or MYB (myeloblastosis oncogene) TFs (Table 4.1). NAC TFs are major regulators of biotic and abiotic stress responses in plants. Several studies have shown the induction of NAC TFs upon virus infection and their essential role in basal defense and the innate plant immune system [62, 63]. This is consistent with the enrichment for genes associated with syncytium formation in module-5. The ERF TFs are well known to be involved in the regulation of disease resistance pathways [64, 65]. Their expression can be altered by pathogen attack and phytohormones like JA, SA, and ET [66]. Only one ERF TF gene (Glyma.17G145300) was found to regulate the JA responsive genes in module-4 (Figure 4.3A) (Table 4.1). The *A. thaliana* homolog of this gene encodes ERF5, which has been implicated as a regulator in the JA-mediated defense pathway [67]. The disparate expression profiles and putative function makes Glyma.17G145300 gene an ideal candidate for the differential regulation of JA-related processes found in module-4, which may

lead to *Rsv3*-mediated ER response in soybean. Some genes in module-4 were also predicted to be regulated by a basic/helix-loop-helix (bHLH) TF (Glyma.17G090500) and a MYB TF (Glyma.08G042100) (Table 4.1). The bHLH TF (Glyma.17G090500) showed contrasting expression profiles between L29 and Williams82, with a two-hour lag in expression changes observed in Williams82 (Figure 4.3A). Another MYB TF (Glyma.04G036700) was also found to regulate genes in module-2, and its expression was significantly down-regulated in L29 at a 6 hpi (Figure 4.3B). MYBs are known to be involved in plant defense and stress responses [65]. In particular, MYB77, encoded by Glyma.04G036700 (the MYB regulating module-2), is associated with stress responses and is a modulator of auxin activity, of which module-2 was enriched with [68, 69].

The module-2 was significantly enriched for the G-box motif ("CACGTG"), which is specifically recognized by the bHLH TF superfamily, and our network happened to predict a bHLH (Glyma.07G051500) regulating module-2 (Table 4.1) [70, 71]. This TF was differentially expressed at 4 hpi with a log2 fold change of -2.30 in L29, showing it was triggered prior to the major transcriptional shift observed at 6 hpi. Comparison of its expression pattern revealed vastly different profiles, with a significant peak in expression in Williams82 (Figure 4.3B**)**. This gene was also identified as a putative resistance gene against a leaf-eating insect, the common cutworm, and similarly, its expression levels were also significantly lower at 4 hpi in the resistant line [72]. This suggests Glyma.07G051500's activity is important in pathogen defense. The *A. thaliana* homolog (AT1G32640) of Glyma.07G051500 encodes a MYC-related transcriptional activator (MYC2) with a bHLH leucine zipper DNA binding domain [73].

MYC2 is reported to condition resistance to insects and regulate ABA signaling, JA-responsive pathogen defense, oxidative stress response genes, and other TFs' expressions, as

well as negatively regulate its own expression [73-79]. Notably, *MYC2* is described as a "master switch" in modulating both positive and negative crosstalk between ABA and JA signaling [80]. As mentioned earlier, we found enrichment for both ABA- and JA-related processes in this study; thus MYC2, encoded by Glyma.07G051500, could be a key regulator in mediating the modular phytohormone responses observed with *Rsv3*-mediated ER. Interestingly, examination of the data from the study using avirulent SMV-G5H and virulent SMV-G7H strains on L29 [56] revealed that the *MYC2* gene Glyma.07G051500 as well as other *MYC2* genes were also exclusively expressed at low levels in L29 during *Rsv3*-mediated resistance. Interesting still, these are not the only instances where suppression of *MYC2* has been shown to promote resistance. In another RNA-seq experiment using near-isogenic soybean lines to study bacterial leaf pustule resistance, three genes encoding MYC2 TFs were expressed at low levels in the resistant line and predicted to be important for conditioning resistance [81]. In an even more striking genome-wide association study (GWAS) on soybean, the same *MYC2* gene (Glyma.07G051500) that was found in this study was identified as a putative resistance gene against the common cutworm where its expression was also significantly down-regulated in the resistant line [72]. Even in tomato, *MYC2* has been shown to regulate immunity via the JA pathway by coordinating a transcriptional cascade [82]. Taken together, these findings indicate that *MYC2* activity may be important in pathogen defense. In particular, it appears that suppression of its activity may in some cases promote resistance, which may be a consequence of its status as a master regulator, allowing it to efficiently suppress expression of targets exploited by pathogens. Because, perhaps by altering a master regulator's expression, the expression of numerous downstream genes (some of which may be targets for pathogen exploitation) can be altered in such a way as to condition resistance. Whatever the case, the function of *MYC2* in

relation to *Rsv3*-mediated ER poses an interesting subject for more research, as it may be responsible for many of the changes observed in ABA and JA signaling that are observed during *Rsv3* resistance [15, 56].

Modular regulation of abscisic acid signaling and suppression of jasmonic acid signaling are features of Rsv3-mediated ER

We examined the gene targets of the *MYC2* (Glyma.07G051500) and *MYB* (Glyma.04G036700) TFs regulating module-2. In particular, we looked at genes involved in ABA, auxin, and defense processes (Table 4.2). All gene targets were down-regulated at 6 hpi in L29. Among the targets were genes encoding ABA and auxin responsive element-binding factors (ABFs, SAUR), ABI five-binding proteins (AFPs), type 2C protein phosphatases (PP2Cs), and MYB-like TFs (RVE1s).

We also examined JA- and defense-related gene targets of the *bHLH* (Glyma.17G090500), *ERF* (Glyma.17G145300), and *MYB* (Glyma.08G042100) TFs regulating the module-4 (Table 4.3). Most genes were expressed at low levels in L29, such as those involved in JA biosynthesis and a number of TFs; however, at 2 hpi, a few genes were up-regulated. These were Glyma.19G164600 encoding an MYB14 TF, and Glyma.12G114100 encoding an L-type lectin receptor kinase, which induces hydrogen peroxide production, cell death, and is required for resistance to oomycetes and fungal pathogens [83, 84]. Lastly, Glyma.11G139500 encoding another PP2C was also up-regulated in L29. This protein family was shown to be an essential signaling component of *Rsv3*-mediated ER against SMV, involved in inducing callose deposition via the ABA signaling pathway [15]. We found that differential

regulation of *PP2C* genes begins as early as 2 hpi, suggesting the *Rsv3* resistance pathway is elicited almost immediately after inoculation.

Between the differential regulation of several TFs and signaling molecules, such as the ABF, AFP, PP2C, and JAZ encoding genes in modules 2 and 4, it appears a complex transcriptional cascade is at work, finely regulating both ABA and JA signaling. Characteristically, ABA and JA are mutually antagonistic in a defense response [74, 85]; however, according to our results, this does not appear to be the case during the early hours of *Rsv3*-mediated resistance. Between 0 and 8 hpi, ABA- and JA-related genes were largely down-regulated in L29, indicating a signaling scheme divergent from the typical antagonistic relationship between ABA and JA. The purpose of this interaction is not clear, but certain components of their signaling pathways, such as ABFs in the ABA pathway, may be targets for viral exploitation and would thus require suppression in order to condition SMV resistance. For example, high *ABF1* expression was observed during *Sonchus yellow net virus* and *Impatiens necrotic spot virus* infection [86]; thus *ABF* suppression may also be important for escaping SMV infection. However, it seems some aspects of the ABA pathway must remain functional, as ABA accumulation was observed in *Rsv3*-mediated ER at 8 hpi and later [15]. This suggests the ABA signaling pathway may be modular in L29, with it first being silenced during the early hours post-inoculation (2-8 hpi) and then later re-activated (8 hpi). Evading viral exploitation may be the case for the JA pathway as well, as genes functioning in this pathway were mostly suppressed (4-8 hpi) in L29. This suppression was also observed in another *Rsv3* RNA-seq study at times even later than 8 hpi [56]. Even more, JA biosynthesis has been shown to increase susceptibility to some viruses in soybean [55]. Consequently, and unlike the modular regulation pattern found with the ABA pathway, it may be critical for the JA pathway to remain suppressed

in order for *Rsv3*-mediated resistance to be conferred; such a condition would be worthwhile to investigate. Regardless, it appears that a finely regulated phytohormone network conditions *Rsv3*-mediated resistance via suppression of the JA pathway and modular regulation of the ABA signaling pathway. This carefully orchestrated network may help explain how *Rsv3*-mediated ER is able to swiftly coordinate a defense against SMV.

CONCLUSION

In conclusion, we compared the transcriptomic response of two soybean varieties exhibiting susceptible and resistant phenotype to SMV-G7 strain and constructed gene regulatory networks to identify key genes and transcription factors that regulate the *Rsv3*-mediated ER mechanism in soybean. Our findings suggest that the *Rsv3*-mediated ER response is initiated early after inoculation once the fate of susceptibility or resistance to SMV is determined. The *Rsv3*-mediated ER response appears to largely involve differential regulation of various phytohormone pathways, suggesting phytohormone signaling to be fundamental in *Rsv3*-mediated resistance. In particular, early suppression of SA, CK, ABA, and JA pathways and the interplay of ABA and JA pathways may be essential. Different TFs, MYC2 in particular, were found to regulate these signaling events possibly via down-regulation of numerous genes to evade viral exploitation in the SMV-resistant cultivar L29 (*Rsv3*-genotype). While experimentation is needed for further confirmation, our analyses predict potential candidate genes for hypothesis-driven experiments. Overall, this study offers new insights into the unique and intricate regulation of the *Rsv3*-mediated ER response to *Soybean mosaic virus*.

ACKNOWLEDGEMENTS

We would like to thank Dr. Colin Davis for his assistance and the Advanced Research Computing and Translational Plant Sciences' MAGYK computing resources at Virginia Tech. This project was funded in part by the Virginia Soybean Board. Additional funding was provided by the Agricultural Experiment Station Hatch Program and Open Access Subvention Fund – both at Virginia Tech. LD was funded in part by the John Lee Pratt Fellowship Program at Virginia Tech.

REFERENCES

1. Bendahmane A, Kanyuka K, Baulcombe DC. The Rx gene from potato controls separate virus resistance and cell death responses. The Plant Cell. 1999;11(5):781-91.
2. Cho E-K, Goodman RM. Strains of soybean mosaic virus: classification based on virulence in resistant soybean cultivars. Phytopathology. 1979;69(5):467-70.
3. Buss G, Ma G, Chen P, Tolin S. Registration of V94-5152 soybean germplasm resistant to soybean mosaic potyvirus. Crop Science. 1997;37(6):1987-8.
4. Hayes AJ, Ma G, Buss GR, Saghai Maroof MA. Molecular marker mapping of Rsv 4, a gene conferring resistance to all known strains of soybean mosaic virus. Crop Science. 2000;40(5):1434-7.
5. Jeong S, Kristipati S, Hayes A, Maughan P, Noffsinger S, Gunduz I, et al. Genetic and sequence analysis of markers tightly linked to the soybean mosaic virus resistance gene, Rsv 3. Crop Science. 2002;42(1):265-70.
6. Saghai Maroof MA, Tucker DM, Skoneczka JA, Bowman BC, Tripathy S, Tolin SA. Fine mapping and candidate gene discovery of the soybean mosaic virus resistance gene, Rsv4. The Plant Genome. 2010;3(1):14-22.
7. Klepadlo M, Chen P, Shi A, Mason RE, Korth KL, Srivastava V, et al. Two tightly linked genes for Soybean mosaic virus resistance in soybean. Crop Science. 2017;57(4):1844-53.
8. Hajimorad M, Domier LL, Tolin S, Whitham S, Saghai Maroof MA. Soybean mosaic virus: a successful potyvirus with a wide distribution but restricted natural host range. Molecular Plant Pathology. 2018;19(7):1563-79.

9. Chen P, Buss G, Roane C, Tolin S. Allelism among genes for resistance to soybean mosaic virus in strain-differential soybean cultivars. Crop Science. 1991;31(2):305-9.

10. Ma G, Chen P, Buss G, Tolin S. Complementary action of two independent dominant genes in Columbia soybean for resistance to soybean mosaic virus. Journal of Heredity. 2002;93(3):179-84.

11. Suh SJ, Bowman BC, Jeong N, Yang K, Kastl C, Tolin SA, et al. The Rsv3 locus conferring resistance to soybean mosaic virus is associated with a cluster of coiled-coil nucleotide-binding leucine-rich repeat genes. The Plant Genome. 2011;4(1):55-64.

12. Redekar N, Clevinger E, Laskar M, Biyashev R, Ashfield T, Jensen R, et al. Candidate gene sequence analyses toward identifying Rsv3-type resistance to soybean mosaic virus. The Plant Genome. 2016;9(2):1-12.

13. Tran P-T, Widyasari K, Seo J-K, Kim K-H. Isolation and validation of a candidate Rsv3 gene from a soybean genotype that confers strain-specific resistance to soybean mosaic virus. Virology. 2018;513:153-9.

14. Seo J-K, Lee S-H, Kim K-H. Strain-specific cylindrical inclusion protein of Soybean mosaic virus elicits extreme resistance and a lethal systemic hypersensitive response in two resistant soybean cultivars. Molecular Plant-Microbe Interactions. 2009;22(9):1151-9.

15. Seo J-K, Kwon S-J, Cho WK, Choi H-S, Kim K-H. Type 2C protein phosphatase is a key regulator of antiviral extreme resistance limiting virus spread. Scientific Reports. 2014;4:5905.

16. Banf M, Rhee SY. Computational inference of gene regulatory networks: approaches, limitations and opportunities. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2017;1860(1):41-52.

17. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. Nature Reviews Genetics. 2013;14(12):840-52.

18. Kaleta C, Göhler A, Schuster S, Jahreis K, Guthke R, Nikolajewa S. Integrative inference of gene-regulatory networks in Escherichia coli using information theoretic concepts and sequence analysis. BMC Systems Biology. 2010;4(1):116-26.

19. Küffner R, Petri T, Tavakkolkhah P, Windhager L, Zimmer R. Inferring gene regulatory networks by ANOVA. Bioinformatics. 2012;28(10):1376-82.

20. Montes RAC, Coello G, González-Aguilera KL, Marsch-Martínez N, De Folter S, Alvarez-Buylla ER. ARACNe-based inference, using curated microarray data, of Arabidopsis thaliana root transcriptional regulatory networks. BMC Plant Biology. 2014;14(1):97-110.

21. Taylor-Teeples M, Lin L, De Lucas M, Turco G, Toal T, Gaudinier A, et al. An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature. 2015;517(7536):571-5.

22. Ikeuchi M, Shibata M, Rymen B, Iwase A, Bågman A-M, Watt L, et al. A gene regulatory network for cellular reprogramming in plant regeneration. Plant and Cell Physiology. 2018;59(4):770-82.

23. Shibata M, Breuer C, Kawamura A, Clark NM, Rymen B, Braidwood L, et al. GTL1 and DF1 regulate root hair growth through transcriptional repression of ROOT HAIR DEFECTIVE 6-LIKE 4 in Arabidopsis. Development. 2018;145(3):dev159707.

24. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-51.

25. Hecker M, Lambeck S, Toepfer S, Van Someren E, Guthke R. Gene regulatory network inference: data integration in dynamic models—a review. Biosystems. 2009;96(1):86-103.

26. Haque S, Ahmad JS, Clark NM, Williams CM, Sozzani R. Computational prediction of gene regulatory networks in plant growth and development. Current Opinion in Plant Biology. 2019;47:96-105.

27. Li Y, Pearl SA, Jackson SA. Gene networks in plant biology: approaches in reconstruction and analysis. Trends in Plant Science. 2015;20(10):664-75.

28. Redekar N, Pilot G, Raboy V, Li S, Saghai Maroof MA. Inference of transcription regulatory network in low phytic acid soybean seeds. Frontiers in Plant Science. 2017;8:1-14.

29. Schäfer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics. 2004;21(6):754-64.

30. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006;7(1):S7.

31. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biology. 2007;5(1):e8.

32. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PloS One. 2010;5(9):e12776.

33. Haury A-C, Mordelet F, Vera-Licona P, Vert J-P. TIGRESS: trustful inference of gene regulation using stability selection. BMC Systems Biology. 2012;6(1):145.

34. Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. Nature Methods. 2012;9(8):796-804.

35. O'Malley RC, Huang S-sC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and epicistrome features shape the regulatory DNA landscape. Cell. 2016;165(5):1280-92.

36. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 2009;37(suppl_2):W202-W8.

37. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology; Menlo Park, CA: AAAl Press; 1994. p. 28-36.

38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

39. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463(7278):178-83.

40. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

41. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013;30(7):923-30.

42. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.

43. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. The R Journal. 2016;8(1):289-317.

44. Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978;6(2):461-4.

45. Biernacki C, Celeux G, Govaert G. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE transactions on pattern analysis and machine intelligence. 2000;22(7):719-25.

46. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research. 2009;38(suppl_1):D843-D6.

47. Bedre R, Mandadi K. GenFam: A web application and database for gene family□based classification and functional enrichment analysis. Plant Direct. 2019;3(12):e00191.

48. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research. 2017;45(D1):D1040–D5.

49. Meyer PE, Lafitte F, Bontempi G. minet: AR/Bioconductor package for inferring large transcriptional networks using mutual information. BMC Bioinformatics. 2008;9(1):461.

50. Schäfer J, Opgen-Rhein R, Strimmer K. Reverse engineering genetic networks using the GeneNet package. The Newsletter of the R Project Volume 6/5, December 2006. 2006;6(9):50.

51. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biology. 2007;8(2):R24.

52. Hamaguchi A, Yamashino T, Koizumi N, Kiba T, Kojima M, Sakakibara H, et al. A small subfamily of Arabidopsis RADIALIS-LIKE SANT/MYB genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. Bioscience, Biotechnology, and Biochemistry. 2008;72(10):2687-96.

53. Alazem M, Lin NS. Roles of plant hormones in the regulation of host–virus interactions. Molecular Plant Pathology. 2015;16(5):529-40.

54. Atsumi G, Kagaya U, Kitazawa H, Nakahara KS, Uyeda I. Activation of the salicylic acid signaling pathway enhances Clover yellow vein virus virulence in susceptible pea cultivars. Molecular Plant-Microbe Interactions. 2009;22(2):166-75.

55. Singh AK, Fu D-Q, El-Habbak M, Navarre D, Ghabrial S, Kachroo A. Silencing genes encoding omega-3 fatty acid desaturase alters seed size and accumulation of Bean pod mottle virus in soybean. Molecular Plant-Microbe Interactions. 2011;24(4):506-15.

56. Alazem M, Tseng K-C, Chang W-C, Seo J-K, Kim K-H. Elements Involved in the Rsv3-Mediated Extreme Resistance against an Avirulent Strain of Soybean Mosaic Virus. Viruses. 2018;10(11):581.

57. Grangeon R, Jiang J, Laliberte J-F. Host endomembrane recruitment for plant RNA virus replication. Current Opinion in Virology. 2012;2(6):683-90.

58. Grangeon R, Jiang J, Wan J, Agbeci M, Zheng H, Laliberté J-F. 6K2-induced vesicles can move cell to cell during turnip mosaic virus infection. Frontiers in Microbiology. 2013;4:351.

59. UniProt. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research. 2018;47(D1):D506-D15.

60. Mäkinen K, Hafrén A. Intracellular coordination of potyviral RNA functions in infection. Frontiers in Plant Science. 2014;5:110.

61. Argueso CT, Ferreira FJ, Epple P, To JP, Hutchison CE, Schaller GE, et al. Two-component elements mediate interactions between cytokinin and salicylic acid in plant immunity. PLoS Genetics. 2012;8(1):e1002448.

62. Puranik S, Sahu PP, Srivastava PS, Prasad M. NAC proteins: regulation and role in stress tolerance. Trends in Plant Science. 2012;17(6):369-81.

63. Nuruzzaman M, Sharoni AM, Kikuchi S. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Frontiers in Microbiology. 2013;4:248.

64. Oñate-Sánchez L, Singh KB. Identification of Arabidopsis ethylene-responsive element binding factors with distinct induction kinetics after pathogen infection. Plant Physiology. 2002;128(4):1313-22.

65. Singh KB, Foley RC, Oñate-Sánchez L. Transcription factors in plant defense and stress responses. Current Opinion in Plant Biology. 2002;5(5):430-6.

66. Gutterson N, Reuber TL. Regulation of disease resistance pathways by AP2/ERF transcription factors. Current Opinion in Plant Biology. 2004;7(4):465-71.

67. Moffat CS, Ingle RA, Wathugala DL, Saunders NJ, Knight H, Knight MR. ERF5 and ERF6 play redundant roles as positive regulators of JA/Et-mediated defense against Botrytis cinerea in Arabidopsis. PloS One. 2012;7(4):e35995.

68. Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, et al. The Arabidopsis transcription factor MYB77 modulates auxin signal transduction. The Plant Cell. 2007;19(8):2440-53.

69. Jung C, Seo JS, Han SW, Koo YJ, Kim CH, Song SI, et al. Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic Arabidopsis. Plant Physiology. 2008;146(2):623-35.

70. Toledo-Ortiz G, Huq E, Quail PH. The Arabidopsis basic/helix-loop-helix transcription factor family. The Plant Cell. 2003;15(8):1749-70.

71. To J. Effect of different strains of soybean mosaic virus on growth, maturity, yield, seed mottling and seed transmission in several soybean cultivars. Journal of Phytopathology. 1989;126(3):231-6.

72. Liu H, Che Z, Zeng X, Zhang G, Wang H, Yu D. Identification of single nucleotide polymorphisms in soybean associated with resistance to common cutworm (Spodoptera litura Fabricius). Euphytica. 2016;209(1):49-62.

73. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. The Plant Cell. 2003;15(1):63-78.

74. Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, et al. Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. The Plant Cell. 2004;16(12):3460-79.

75. Boter M, Ruíz-Rivero O, Abdeen A, Prat S. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and Arabidopsis. Genes & Development. 2004;18(13):1577-91.

76. Lorenzo O, Chico JM, Sánchez-Serrano JJ, Solano R. JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different

jasmonate-regulated defense responses in Arabidopsis. The Plant Cell. 2004;16(7):1938-50.

77. Chini A, Fonseca S, Fernandez G, Adie B, Chico J, Lorenzo O, et al. The JAZ family of repressors is the missing link in jasmonate signalling. Nature. 2007;448(7154):666-71.

78. Dombrecht B, Xue GP, Sprague SJ, Kirkegaard JA, Ross JJ, Reid JB, et al. MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. The Plant Cell. 2007;19(7):2225-45.

79. Schweizer F, Fernández-Calvo P, Zander M, Diez-Diaz M, Fonseca S, Glauser G, et al. Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior. The Plant Cell. 2013;25(8):3117-32.

80. Kazan K, Manners JM. MYC2: the master in action. Molecular Plant. 2013;6(3):686-703.

81. Kim KH, Kang YJ, Kim DH, Yoon MY, Moon J-K, Kim MY, et al. RNA-Seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and-susceptible alleles. DNA Research. 2011;18(6):483-97.

82. Du M, Zhao J, Tzeng DT, Liu Y, Deng L, Yang T, et al. MYC2 orchestrates a hierarchical transcriptional cascade that regulates jasmonate-mediated plant immunity in tomato. The Plant Cell. 2017;29(8):1883-906.

83. Singh P, Chien C-C, Mishra S, Tsai C-H, Zimmerli L. The Arabidopsis LECTIN RECEPTOR KINASE-VI. 2 is a functional protein kinase and is dispensable for basal resistance to Botrytis cinerea. Plant Signaling & Behavior. 2013;8(1):e22611.

84. Wang Y, Cordewener JH, America AH, Shan W, Bouwmeester K, Govers F. Arabidopsis lectin receptor kinases LecRK-IX. 1 and LecRK-IX. 2 are functional analogs in

regulating Phytophthora resistance and plant cell death. Molecular Plant-Microbe Interactions. 2015;28(9):1032-48.

85. Robert-Seilaniantz A, Grant M, Jones JD. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. Annual Review of Phytopathology. 2011;49:317-43.

86. Senthil G, Liu H, Puram V, Clark A, Stromberg A, Goodin M. Specific and common changes in Nicotiana benthamiana gene expression in response to infection by enveloped viruses. Journal of General Virology. 2005;86(9):2615-25.

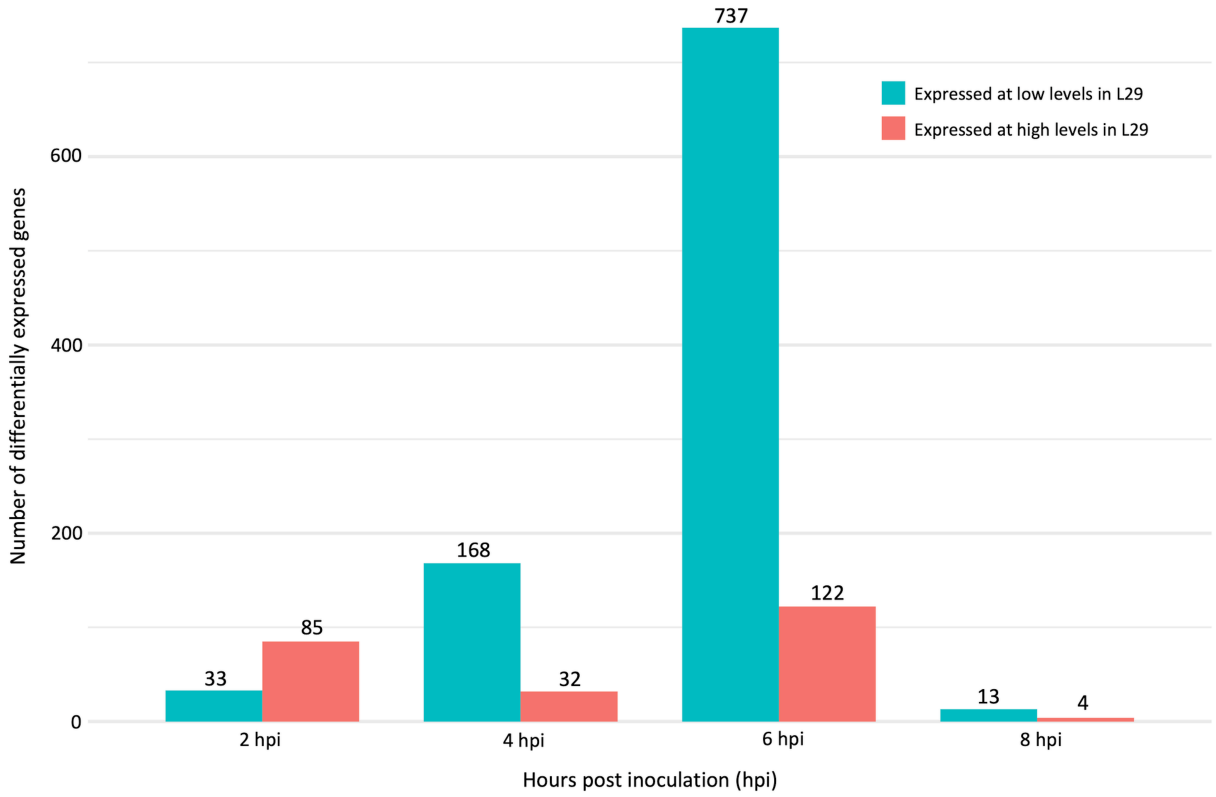


Figure 4.1. Number of differentially expressed genes between soybean cultivars L29 and Williams82 at 2, 4, 6, and 8 hours post inoculation with *Soybean mosaic virus* strain G7. DEGs were defined as those with FDR adjusted p-value < 0.05, $\log_2$ fold change >|1.0|, and base mean >10. High expression or low expression in L29 means the expression of DEG was either higher or lower in L29 as compared to Williams82, respectively. A total of 1128 DEGs were identified between L29 and Williams82 at 2, 4, 6 and 8 hpi. DEGs at 0 hpi were minimal and excluded, being considered effects of differences in genetic backgrounds of the two cultivars and not infection responses.

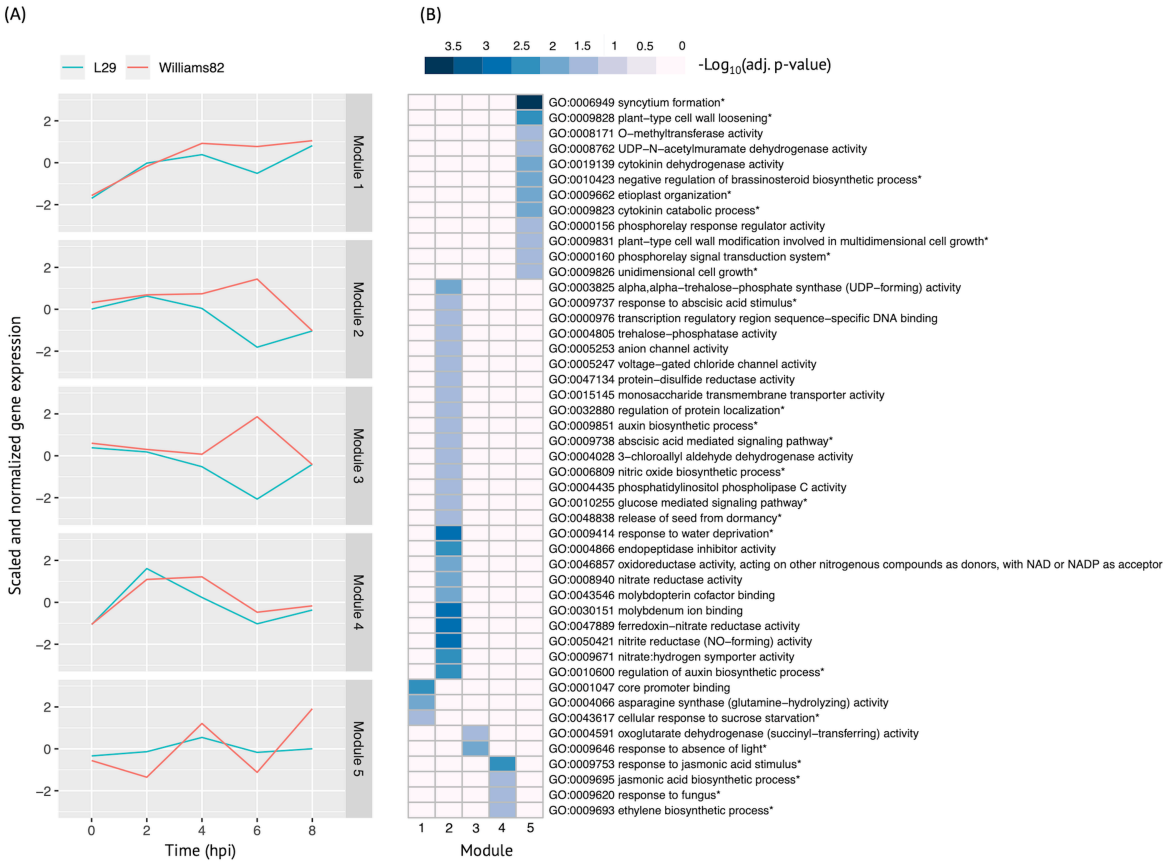
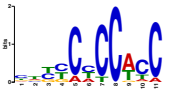
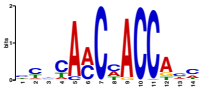
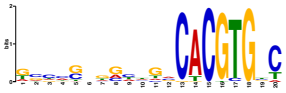
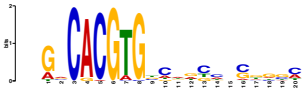
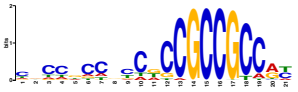
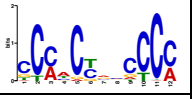
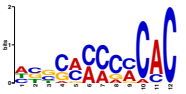
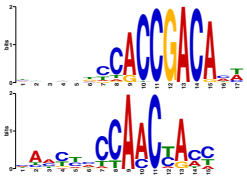
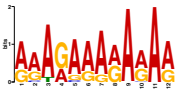
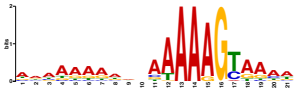


Figure 4.2. Co-expression gene modules and their biological functions. A module is defined as a group of genes sharing similar expression profiles over time and likely involved in the same biological processes. The expression profile for these modules was determined by averaging the expression levels of DEGs within each module. (A) Mean module expression profiles of L29 and Williams82 over time. Normalized expressions of DEGs were used for clustering with Gaussian-finite mixture modeling. (B) Heatmap of GO functional enrichment analyses. Columns represent module groups. Rows represent hierarchical clustering of enriched GO categories; those with an asterisk indicate a biological process, while all others are molecular functions. Color represents $-\log_{10}$ adjusted p-value.

Table 4.1. *A. thaliana* and motif validated interactions.

TF Name	TF Family	Target Module	*A. thaliana* Homolog	MEME Motif Enrichment E-value	MEME Motif	DAP-seq Motif	DAP-seq Motif Similarity p-value
Glyma.07G060400	bZIP	1	AT2G46270	2.00E-20			3.59E-04
Glyma.04G036700	MYB	2	AT3G50060	2.40E-19			8.16E-04
Glyma.07G051500*	MYC2 (bHLH)	2	AT1G32640	9.30E-24			5.58E-05
Glyma.06G092000*	bHLH	3	AT5G65640	6.20E-05			7.62E-05
Glyma.17G090500*	bHLH	4	AT4G20970	2.30E-04			2.22E-04
Glyma.17G145300	ERF	4	AT5G47230	1.60E-02			1.78E-06
Glyma.08G042100	MYB	4	AT1G25340	1.00E-18			1.90E-05
Glyma.02G080200 Glyma.08G216600 Glyma.05G234600 Glyma.08G042100	ERF ERF MYB MYB	5	AT2G33710 AT5G25190 AT1G25340 AT1G25340	2.10E-11			2.89E-04 4.24E-03
Glyma.18G301500	NAC	5	AT5G13180	1.20E-33			5.01E-06

Shown are putative TF-module interactions with their validation results from motif sequence analyses. MEME results show enriched motifs found in each module using promoter sequences of genes belonging to module. *A. thaliana* DAP-seq data was used to find motifs with high similarity to MEME motifs, which enabled identification of TFs that putatively recognize and bind the enriched MEME motifs discovered in each module.

*TFs with asterisks were validated by motif sequence analyses only.

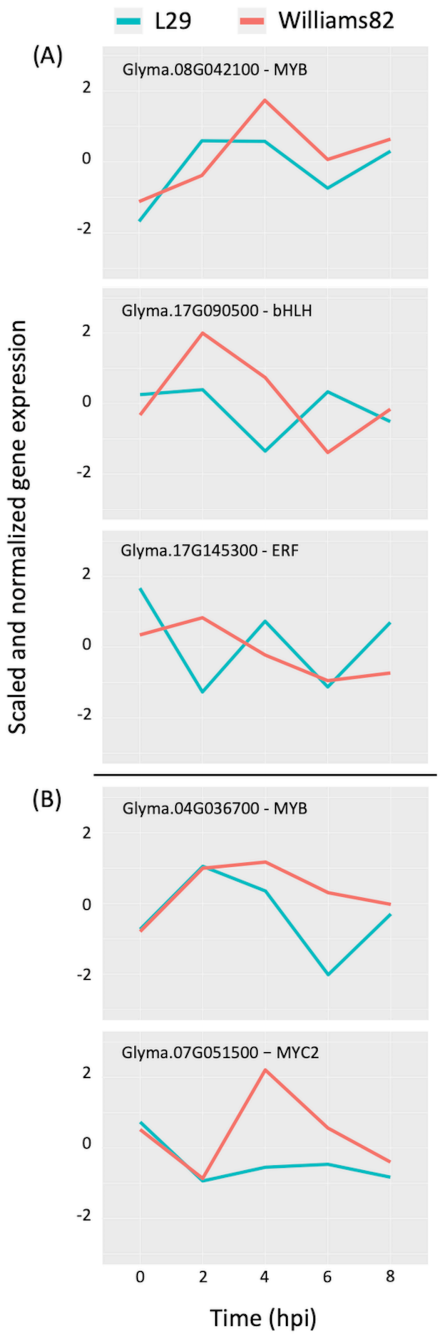


Figure 4.3. Comparison of normalized gene expression profiles of validated TFs in L29 and Williams82. (A) TFs predicted to regulate module-4. (B) TFs predicted to regulate module-2.

Table 4.2. TF target genes in module-2 related to ABA and auxin processes and defense responses.

Target Gene	*A. thaliana* Homolog	Regulator TF	L29 Log2 Fold Change at 6hpi	Gene Symbol	Description
Glyma.07G074400	AT3G61220	MYB	-2.34	SDR1	(+)-neomenthol dehydrogenase
Glyma.09G218600	AT4G19230	MYB	-2.22	CYP707A1	Abscisic acid 8'-hydroxylase 1
Glyma.02G131700	AT1G49720	MYB, MYC2	-1.11	ABF1	Abscisic acid responsive element-binding factor 1
Glyma.06G040400	AT1G45249	MYB	-1.43	ABF2, AREB1	Abscisic acid responsive elements-binding factor 2
Glyma.15G105100	AT5G19140	MYB	-1.04	AILP1, ATAILP1	Aluminum induced protein with YGL and LRDR motifs
Glyma.09G005700	AT1G62300	MYB, MYC2	-1.56	-	At1g62300 protein (Fragment)
Glyma.09G219300	AT5G18050	MYB	-2.23	SAUR22	Auxin-responsive protein
Glyma.04G061500	AT5G25110	MYB, MYC2	-1.39	CIPK25, PKS25, SnRK3.25	CBL-interacting serine/threonine-protein kinase 25
Glyma.06G062100	AT5G25110	MYB	-1.97	CIPK25, PKS25, SnRK3.25	CBL-interacting serine/threonine-protein kinase 25
Glyma.20G241700	AT3G55120	MYB	-1.50	CHI1, CFI, TT5	Chalcone--flavonone isomerase 1
Glyma.16G194600	AT3G05200	MYB	-1.80	ATL6	E3 ubiquitin-protein ligase
Glyma.09G140700	AT3G05200	MYB	-1.72	ATL6	E3 ubiquitin-protein ligase
Glyma.07G060400	AT2G46270	MYB, MYC2	-1.56	GBF3	G-box binding factor 3
Glyma.12G117700	AT2G20570	MYB, MYC2	-1.11	GPRI1, GLK1	GBF's pro-rich region-interacting factor 1
Glyma.02G241000	AT5G17300	MYB, MYC2	-2.11	RVE1	Homeodomain-like superfamily protein
Glyma.13G152300	AT5G17300	MYB	-1.69	RVE1	Homeodomain-like superfamily protein
Glyma.14G210600	AT5G17300	MYB, MYC2	-1.78	RVE1	Homeodomain-like superfamily protein
Glyma.06G319600	AT1G33590	MYB, MYC2	-2.59	-	Leucine-rich repeat (LRR) family protein
Glyma.13G253300	AT1G09970	MYB	-1.39	-	Leucine-rich repeat receptor-like kinase
Glyma.20G054000	AT3G45140	MYB, MYC2	-1.11	LOX2	Lipoxygenase 2
Glyma.02G272700	AT5G20990	MYB	-1.08	-	Molybdopterin biosynthesis CNX1 protein
Glyma.01G060300	AT1G13740	MYB, MYC2	-2.12	AFP2	Ninja-family protein AFP2 (ABI five-binding protein 2)
Glyma.02G118500	AT1G13740	MYB, MYC2	-1.91	AFP2	Ninja-family protein AFP2 (ABI five-binding protein 2)
Glyma.18G267200	AT1G13740	MYB, MYC2	-1.60	AFP2	Ninja-family protein AFP2 (ABI five-binding protein 2)
Glyma.04G014000	AT3G18830	MYB	-1.62	PLT5	Polyol transporter 5
Glyma.13G076700	AT3G20770	MYB	-1.34	EIN3	Protein ETHYLENE INSENSITIVE 3
Glyma.20G051500	AT3G20770	MYB	-1.02	EIN3	Protein ETHYLENE INSENSITIVE 3
Glyma.19G069200	AT1G07430	MYB	-1.55	AIP1	Protein phosphatase 2C 3
Glyma.08G033800	AT4G26080	MYB	-1.09	ABI1	Protein phosphatase 2C 56
Glyma.02G086100	AT1G14790	MYB	-1.87	RDR1, RDRP1	RNA-dependent RNA polymerase 1

Shown are target genes, the TFs putatively regulating them, log2 fold change of target genes, and target genes' functions based on *A. thaliana* homologs.

Table 4.3. TF target genes in module-4 related to JA processes and defense responses.

Target Gene	*A. thaliana* Homolog	Regulator TF	L29 Log2 Fold Change	hpi	Gene Symbol	Description
Glyma.13G361900	AT1G15520	ERF	-1.05	4	ABCG40, PDR12, PDR9	ABC transporter G family member 40
Glyma.01G153300	AT4G19230	bHLH, ERF, MYB	-1.19	4	CYP707A1	Abscisic acid 8'-hydroxylase 1
Glyma.19G044900	AT3G25780	bHLH, ERF, MYB	-1.11	4	AOC3	Allene oxide cyclase 3
Glyma.17G007600	AT4G17230	bHLH	-1.72	4	-	AT4G17230 protein (Fragment)
Glyma.05G082400	AT5G66900	MYB	-2.43	6	MUD21.16	Disease resistance protein (CC-NBS-LRR class) family
Glyma.02G132500	AT4G34410	bHLH, MYB	-1.45	4	ERF109	Ethylene-responsive transcription factor 109
Glyma.15G078600	AT1G28480	bHLH, ERF	-1.08	4	GRXC9, GRX480, ROXY19	Glutaredoxin-C9
Glyma.11G038600	AT1G19180	MYB	-2.61	4	JAZ1	Jasmonate-zim-domain protein 1
Glyma.15G179600	AT1G19180	MYB	-1.69	4	JAZ1	Jasmonate-zim-domain protein 1
Glyma.12G114100	AT4G28350	bHLH, MYB	1.78	2	LECRK72, LECRKD	L-type lectin-domain containing receptor kinase
Glyma.13G030300	AT3G45140	bHLH, MYB	-1.68	6	LOX2	Lipoxygenase 2
Glyma.07G039900	AT1G17420	MYB	-1.13	4	LOX3	Lipoxygenase 3
Glyma.04G226700	AT4G35580	bHLH	-1.05	2	NTL9, CBNAC	NAC transcription factor-like 9
Glyma.06G138100	AT4G35580	bHLH	-1.01	2	NTL9, CBNAC	NAC transcription factor-like 9
Glyma.11G228100	AT2G40000	MYB, ERF	-1.19	6	HSPRO2	Nematode resistance protein-like
Glyma.11G139500	AT1G07630	bHLH, ERF, MYB	1.13	2	PLL5	Protein phosphatase 2C 4
Glyma.01G204400	AT1G74950	bHLH, ERF, MYB	-2.30	4	TIFY10B, JAZ2	Protein TIFY 10B
Glyma.09G145600	AT1G47890	MYB	-2.46	4	RLP7	Receptor-like protein 7
Glyma.07G189300	AT4G21440	bHLH, MYB	-1.62	4	MYB102	Transcription factor MYB102
Glyma.19G164600	AT2G31180	bHLH, MYB	2.57	2	MYB14	Transcription factor MYB14
Glyma.01G128100	AT2G38470	ERF	-2.49	4	WRKY33	WRKY transcription factor 33

Shown are target genes, the TFs putatively regulating them, log2 fold change of target genes, and target genes' functions based on *A. thaliana* homologs.

CHAPTER 5

CONCLUSION

RESEARCH SUMMARY

The research conducted for this dissertation was toward understanding the transcriptional regulation and metabolic events underlying two economically important agronomic traits in soybean: seed phytic acid content and *Soybean mosaic virus* (SMV) resistance. This was achieved through functional genomics approaches such as transcriptomics and metabolomics.

The primary focus of this dissertation is on the characterization of low phytic acid soybeans. Phytic acid, an abundant compound found in seeds, is a chelator of phosphorus and other essential minerals. Accordingly, soybean seed phytic acid is recognized as an antinutrient, causing nutrient deficiencies in humans and monogastric livestock and leading to an accumulation of undigested waste that contributes to phosphorus pollution. The development of low phytic acid soybeans offers a solution to these problems, as they have enhanced nutritional value and reduced environmental impact. Nonetheless, the low phytic acid phenotype is associated with poor seed performance and low seed germination, making these soybeans undesirable for commercial use. Thus, to capitalize on the advantages afforded by low phytic acid soybeans, it is necessary to understand the genetic and molecular basis of seed phytic acid

content in relation to seed performance. The research in this dissertation addresses the issue at hand by applying functional genomics approaches, specifically transcriptomics and metabolomics. Transcriptome profiling was used to investigate the effects of low phytic acid causing mutations on the regulation of seed germination (Chapter 2), and metabolome profiling was implemented to discern metabolic changes in seeds resulting from these mutations (Chapter 3).

The transcriptomics study in Chapter 2 consisted of comparative time series analyses of eight soybean lines from three distinct genotypic classes, each of which contained at least one normal phytic acid soybean line and a unique low phytic acid line. This experimental design enabled the reconstruction of gene regulatory networks (GRN) for each genotypic class, which permitted the examination of transcriptional regulation in germinating low phytic acid soybeans. During the course of this study, differentially expressed genes were identified between low and normal phytic acid soybeans from each genotypic class at three stages of seed germination. Among the differentially expressed genes, several significantly enriched biological processes were discovered; many of which are associated with the phytic acid pathway and could potentially impact germination. A few such processes include phosphate ion homeostasis, *myo*-inositol metabolism, numerous stress associated responses, and ABA signaling. The transcriptional regulation of these processes as well as others was explored by computational inference of GRNs reverse-engineered from the times series transcriptomic data of soybean lines from the three genotypic classes. This allowed for the identification of putative regulatory interactions between transcription factor (TF) genes and target genes in significantly affected biological processes. Such findings provide new information on the molecular mechanisms upsetting the regulation of germination and emergence of low phytic acid soybeans.

The metabolomics study in Chapter 3 specifically focused on the seed lipidomes of low and normal phytic acid soybeans. The effects of low phytic acid causing mutations on seed lipid metabolism were examined using four soybean lines from two distinct genotypic classes with differing low phytic acid causing mutations. Untargeted lipidomic profiling using liquid chromatography-mass spectrometry was used to compare the lipidomes of low and normal phytic acid soybean seeds from both genotypic classes. Profiling analyses revealed that the low phytic acid causing mutations do affect lipid metabolism, with changes being found in ceramide, glucose-sitosterol, phosphatidic acid, phosphatidylethanolamine, and peroxidized triacylglyceride contents. These changes are notable because several are indicative of irregularities in the cell membrane and regulation of programmed cell death. This was supported by the observation that low phytic acid soybean seeds release significantly more electrolytes than their normal phytic acid sibling lines. Elevated electrolyte leakage is an indicator of impaired cell membranes and poor seed vigor. Examination of the protein and metabolite exudates revealed that they too possessed considerable differences between the low and normal phytic acid genotypes. To establish the significance of these changes will require further work, but together, the results from this study provide new hypotheses on the mechanisms of low emergence in low phytic acid soybeans.

The final focus of this dissertation (Chapter 4) is on the characterization of the *Rsv3*-mediated extreme resistance (ER) response against SMV. The *Rsv3* locus, which has been mapped and the resistance gene identified (Glyma.14G38533), confers a unique and rare form of resistance – ER. With ER, pathogen replication and spread are rapidly inhibited, yielding an asymptomatic response. However, little is known on the molecular dynamics behind this type of resistance. Thus, to establish the transcriptional regulatory mechanisms underlying *Rsv3*-

mediated ER, a comparative transcriptomic time series study was performed on SMV-G7-inoculated soybean cultivars 'L29' (*Rsv3*-genotype, resistant) and 'Williams82' (*rsv3*-genoype, susceptible), and a GRN was built to identify putative regulatory interactions. Results indicate that the *Rsv3* defense response is induced as early as 6 hours post-inoculation with many of the transcriptional changes being down-regulation in phytohormone-related genes. This suggests a differentially regulated phytohormone network to be a key feature of *Rsv3*-mediated ER. Perhaps one of the primary regulators found by GRN inference as being potentially responsible for these changes is Glyma.07G051500 encoding an MYC2 TF. MYC2 is widely recognized as a master regulator of ABA and JA signaling, both of which have observed roles in the *Rsv3* defense response. Accordingly, MYC2-mediated regulation may be important in ABA- and JA-derived defense signaling in *Rsv3*-mediated ER. Although functional validation of this is needed, the GRN analysis provides promising candidate genes functioning in biological processes demonstrated to have a role in inducing *Rsv3*-mediated ER.

FUTURE DIRECTIONS

Follow up experiments suggested for phytic acid project

In the transcriptomics study (Chapter 2), expression changes were found in ABA-related genes for each low phytic acid soybean line used. This study, as well as others, suggest ABA content and signaling is altered in low phytic acid soybean seeds, with increased ABA content inhibiting their germination. To determine if this is the case, LC-MS could be used to measure and compare ABA content in low and normal phytic acid soybean seeds. If ABA levels are in fact altered in low phytic acid soybeans, a gibberellic acid treatment could be applied to the low phytic acid seeds in order to test whether the effect of ABA can be antagonized and improve germination.

In the metabolomics study (Chapter 3), the elevated electrolyte leakage found in the low phytic acid soybean seeds suggests these seeds could have impaired cell membranes and increased cell death. These hypotheses are also supported by the findings that low phytic acid seeds have altered glucose-sitosterol, phosphatidylethanolamine, ceramine, and phosphatidic acid contents. To test for impaired membranes, high-resolution microscopy could be used to examine and compare ultrastructural features of low and normal phytic acid soybean seeds. In particular, transmission electron microscopy (TEM) could be implemented, as it is powerful enough to elucidate fine structural details at the nanoscale level and can be used to provide topochemical information as well. Thus, such a study could reveal anomalies in cellular architecture that may be affecting cell functionality in low phytic acid soybean seeds. As for the altered regulation of cell death in low phytic acid soybeans, this could be evaluated by performing a tetrazolium test. This test would allow for the comparison of living versus dead tissue in low and normal phytic acid seeds and would provide another measure of seed viability.

Follow up experiments suggested for Rsv3 project

In the study focused on characterizing *Rsv3*-mediated ER against SMV (Chapter 4), putative regulatory interactions were inferred between TFs and their target genes by GRN analysis. Some of the interactions that were detected could have key roles in regulating the *Rsv3* response. These interactions were validated computationally but should be validated experimentally as well for verification. Experimental validation could be accomplished by using biochemical assays such as ChIP-seq to identify direct targets of the TFs of interest. Additional experiments that could be performed are genetic assays, where a known TF is silenced or overexpressed. This would be particularly interesting to do for functional validation of

Glyma.07G051500 encoding MYC2. The suppression of this gene in this study as well as others suggest its suppression may be important for *Rsv3*-mediated ER given that it is a primary regulator of ABA and JA signaling, both of which are significantly altered during the *Rsv3* response. To test this, two experiments could be performed: (1) overexpression of the gene in a resistant soybean line followed by observation of the response upon virus inoculation, and (2) silencing of the gene in a susceptible line and again observing the response upon virus inoculation. If *MYC2* suppression is required for *Rsv3*-mediated ER, then the first experiment would result in a susceptible response, and the second experiment would result in a resistant response.

APPENDICES

Table S2.1: Log2 fold change for differentially expressed genes between low and normal phytic acid soybean genotypes at each stage.

Table S2.2: Gene ontology enrichment analyses.

Table S2.3: Interactions in each genotypic class subset predicted by four out of five network inference methods.

Table S2.4: Putative TF-gene interactions in each genotypic class subset supported by motif enrichment analysis of co-expression modules and published *Arabidopsis* interactions found by DAP-seq.

Table S3.1: Detected EMRTs in positive ion mode and their raw peak intensities for *1mlpa* and 1MWT in the Mips genotypic subset.

Table S3.2: Detected EMRTs in positive and negative ion mode and their raw peak intensities for *2mlpa* and 2MWT in the MRP genotypic subset**.**

Table S3.3: Filtered EMRTs in positive and negative ion mode and their normalized peak intensities for *1mlpa* and 1MWT in the Mips genotypic subset.

Table S3.4: Filtered EMRTs in positive and negative ion mode and their normalized peak intensities for *2mlpa* and 2MWT in the MRP genotypic subset.

Table S3.5: Detected EMRTs from seed exudates in positive and negative ion mode and their raw peak intensities for four NILs in the MRP genotypic subset.

Table S3.6: Significant seed exudate EMRTs between *2mlpa* and normal phytic acid lines from MRP genotypic subset in positive and negative ion mode.

Table S4.1: Log2 fold change for differentially expressed genes for time pair comparisons.

Table S4.2: Gene ontology enrichment analysis (GO terms with padj <.05 only).

Table S4.3:. Interactions predicted by four out of five network inference methods.

Table S4.4: Putative TF-gene interactions supported by orthologous interactions found in *A. thaliana*.

Table S4.5: Putative TF-module interactions supported by orthologous interactions found in *A. thaliana*.

Table S4.6: Motif enrichment analysis of co-expression modules and transcription factors recognizing motif sequences.