Transcriptome analysis identifies differentially expressed genes involved in the metabolic regulatory network of progenies from the cross of low phytic acid *GmMIPS1* and *GmIPK1* soybean mutants

Hangxia Jin
ZheJiang Academy of Agricultural Sciences
Xiaomin Yu
ZheJiang Academy of Agricultural Sciences
Qinghua Yang
ZheJiang Academy of Agricultural Sciences
Fengjie Yuan (fjyuanhz@126.com)
ZheJiang Academy of Agricultural Sciences
Xujun Fu
ZheJiang Academy of Agricultural Sciences

Research Article

Keywords: low phytic acid, myo-inositol phosphate synthase, inositol 1,3,4,5,6-pentakisphosphate (InsP5) 2-kinase, transcriptome analyses

Posted Date: December 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-112462/v1

Version of Record: A version of this preprint was published at Scientific Reports on April 22nd, 2021. See the published version at https://doi.org/10.1038/s41598-021-88055-4.

Abstract

Lowering the phytic acid (PA) content of crop seeds will be beneficial for improving their nutritional traits. Low phytic acid (*lpa*) crop lines carrying more than one independent mutated gene have been shown to exhibit more pronounced reductions of PA content than mutants with a single *lpa* mutated gene. But little is known about the link between PA pathway intermediates and downstream regulation following mutation of these genes in soybean. Here, we performed a comparative transcriptome analysis using an advanced-generation recombinant inbred line [*2mlpa (mips1/ipk1*)] with low PA and a sibling line with homozygous non-mutant alleles and normal PA [2MWT (MIPS1/IPK1)]. RNA sequencing revealed differential expression levels of numerous genes between seeds of *2mlpa* and 2MWT at five developmental stages. A total of 7,945 differentially expressed genes were identified. 3316 DEGs were in 128 metabolic and signal transduction pathways and 4980 DEGs were classified into 345 function terms associated with biological processes. Genes associated with PA metabolism, photosynthesis, starch and sucrose metabolism, and defense mechanisms were related to low PA in *2mlpa* soybean line. Among these, 36 genes were up/down-regulated in PA metabolic processes, with 22 possibly contributing to the low PA phenotype of *2mlpa*. Most of the genes (81 of 117) associated with photosynthesis were down-regulated in *2mlpa* at the late seed stage. Three genes involved in sucrose metabolism were up-regulated at the late seed stage, which might explain the high sucrose content of *2mlpa* soybeans. Additionally, 604 genes related to defense mechanisms were differentially expressed between *2mlpa* and 2MWT. In this research, the soybean mutant *2mlpa* was found to not only exhibit low PA but also have changes in multiple metabolites and secondary metabolites. The results delineate the regulation of these metabolic events by 2*mlpa*. Many genes associated with PA metabolism would contribute to the drastic reduction of PA and moderate accumulation of InsP3-InsP5 in 2*mlpa* mutant. And other regulated genes found in photosynthesis, starch and sucrose metabolism, and defense mechanisms would give us more insight into the nutritional and agronomic performance of 2*mlpa*.

1. Introduction

Phosphorus (P) is an essential element required for optimal plant growth [1]. In grains, approximately 75% of P is stored in the form of phytic acid (PA, *myo*-inositol 1,2,3,4,5,6-hexakisphosphate) [2]. The poor digestibility of PA by animals and humans causes low mineral bioavailability and phosphate pollution of soil and water. Hence, low-PA (*lpa*) crops are highly desirable for reducing these anti-nutritional and environmental effects [2,3].

To narrow the negative impacts of PA, many low phytic acid mutants have been generated by mutagenesis and biotechnological methods in different crops such as soybean, rice, barley, maize and wheat [4-10] . The reduction level of PA was related to the various mutated gene type. Recently, it has been demonstrated that cross and selection breeding of different soybean mutants and pyramiding different gene mutations could be employed as a strategy to obtain progeny soybean seeds with stable and dramatic reduction in the intended *lpa* traits. Such as, drastic phytic acid reductions ranging from-79% to -88% were observed in double *lpa* soybean mutants with two IPK1 mutation targets on chromosomes 6 and 14 [11]. More recently, the pronounced lower level of phytic acid content (-63%) was found in the progeny resulting from two rice *lpa* mutants（data under review）.

Therefore, The *lpa* soybean line *Gm-lpa*-TW-1 carries a non-lethal mutation in the *MIPS1* gene, which has a 2-bp deletion in the third exon [12,13], and the *lpa* soybean line *Gm-lpa*-ZC-2 carries a G-A point mutation in the *IPK1* gene, were selected to generate a double *lpa* soybean mutant simultaneously carrying the two mutation targets *MIPS* and *IPK1*. MIPS (EC 5.5.1.4) is the first and rate-limiting enzyme in the inositol and PA biosynthesis pathways, and catalyzes the conversion of glucose 6-phosphate to *myo*-inositol 3-phosphate [14]. *IPK1* encodes an enzyme (EC 2.7.1.158) transforming $InsP_5$ into PA [15]. An advanced-generation recombinant inbred line, *2mlpa* (*mips1/ipk1*) and a sibling line designated 2MWT (*MIPS1/IPK1*) with homozygous non-mutant alleles and normal PA were generated, and the reduction in phytic acid contents (up to 87%) observed in the double *lpa* lines *2mlpa* were significantly more pronounced that those expected from the single mutants

(about -50% reduction in phytate content). The *2mlpa* accumulated lower inositol phosphate isomers InsP4 and InsP5, and has a lower level than the mutant which only carrying the *IPK1* mutation [16].

As we known, PA biosynthesis pathway plays a vital role in plant growth, and loss-of-function mutations in MIPS and IPK1 disrupt the PA biosynthesis pathway while dramatically reducing PA concentration. However, the mechanisms by which PA levels and downstream genes are regulated in these *lpa* soybean lines are poorly understood. It is therefore necessary to explore the whole regulatory network of PA metabolism. mRNA-sequencing (RNA-Seq) was used to study the effects of mutations in the *MIPS1* and *IPK1* genes on global changes in the gene expression profiles of developing soybean seeds. The functional enrichment of biological processes and pathways identified in this report further our understanding of the PA regulatory network.

2. Results And Discussion

2.1 Library sequencing and assembly

To obtain a more comprehensive understanding of changes at the transcriptome level in *lpa* soybeans, 30 cDNA libraries (three replicates) were sequenced. A total of 195.7 Gb of 150-bp paired-end clean data was obtained. Each sample yielded up to 6.3 Gb of clean data with a Q20 percentage above 98%. We mapped 85.71% of the clean reads to the annotated soybean reference *Glycine_max*_v2.0. The number of mapped reads ranged from 34,971,326 to 39,249,412 (Table 1). These results suggest that the data are adequate for subsequent analysis. All raw data are published in the NCBI (http://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA522338.

2.2 Sequence annotation and DEGs analysis

RPKM was used to estimate the gene expression and for DEG analysis. Each mapped soybean gene with its RPKM value in each of the 30 libraries was calculated. Principal component analysis showed that the major contributors to variation in the data were seed developmental stage and genotype (Figure 2). At the same time, the biological replicates of each soybean sample clustered together, suggesting there was little variance between replicates.

We identified a total of 7,945 genes with significantly differential expression between the *2mlpa* and 2MWT lines at five seed developmental stages (Figure 3). Most of the DEGs were expressed during the early phase (stage 1) and late phase (stage 5) of seed development. In total, 1,380 DEGs were only differentially expressed at stage 1, while 4,730 DEGs were only differentially expressed at stage 5. There were 96 DEGs that were differentially expressed at all five stages.

2.3 Validation of DEGs through RT-qPCR

The eight randomly selected DEGs associated with the biological processes discussed in the following sections were confirmed through RT-qPCR (Figure 4). In most cases, gene expression trends were similar to those found in the RNA-Seq analysis, although fold changes varied between the RT-qPCR data and RNA-Seq expression data. Glyma.16G065700 (Figure 4C), Glyma.08G109300 (Figure 4E), Glyma.12G232500 (Figure 4F), and Glyma.16G214900 (Figure 4G) showed similar expression profiles in RNA-Seq and RT-qPCR analyses. However, RNA-Seq analysis revealed downregulation of Glyma.11G238800 (Figure 4A) at stage 3, Glyma.14G072200 (Figure 4B) at stage 5, Glyma.12G210600 (Figure 4D) at stage 2, and Glyma.09G073600 (Figure 4H) at stage 3 of seed development; these results were different from those obtained from the expression profiling by RT-qPCR analysis. These results suggest that RT-qPCR expression profiles were in approximate agreement with the RNA-Seq data for 10 of the DEGs.

2.4 Functional enrichment analyses of DEGs

We assigned 4,980 DEGs between *2mlpa* and 2MWT (62.67% of all 7,945 DEGs) to at least one GO term, classifying them by 345 functional terms associated with biological processes at the five seed developmental stages (Supplementary Table S2). Some terms were found at more than one stage while others were stage-specific. For instance, the terms establishment of localization (23% of the total DEGs in stage 1), single-organism metabolic processes (13.39% of the total DEGs in stage 1), and localization (8.05% of the total DEGs in stage 1) were only found at early seed developmental stages (stages 1 and 2), while organic substance metabolic processes (28.95% of the total DEGs in stage 5), macromolecule metabolic processes (21.23% of the total DEGs in stage 5), and responses to abiotic stimulus (4.44% of the total DEGs in stage 5) were only found at the late seed developmental stage (stage 5).

KEGG pathway enrichment analysis is an effective method for elucidating significantly enriched metabolic and signal transduction pathways and the biological functions of DEGs [17]. We found 3,316 DEGs (41.74% of all 7,945 DEGs) in 128 metabolic and signal transduction pathways between the 2*mlpa* and 2MWT lines at five seed developmental stages (Supplementary Table S3). The highly represented pathways, which had more DEGs, included metabolic pathways, plant–pathogen interactions, plant hormone signal transduction, photosynthesis, flavonoid biosynthesis, phenylpropanoid biosynthesis, and starch and sucrose metabolism (Table 2). It has been observed that DEGs were different at each developmental stage. For example, there were no DEGs in stage 2 and 3, but 22 DESs in stage 5. We will discuss them in the following part.

According to GO and KEGG pathway functional enrichment, the enriched DEGs were mainly involved in PA metabolism, photosynthesis, starch and sucrose metabolism, and defense mechanisms. These results were partly consistent with those from other plants, such as Arabidopsis [18,19], rice [20], poplar [21], and sweet potatoes [22]. Low phytate is often associated with undesirable effects on seed development and germination potential, with apoptosis commonly observed in *mips1* mutants [18,19,21]. We have found that the *2mlpa* plants were weaker than the 2MWT plants when grown in the field, consistent with previous reports and related to the biological processes discussed in the following section.

2.5 DEGs involved in PA metabolism

The PA biosynthesis pathway plays a vital role in plant growth. Plants possess two parallel PA biosynthetic pathways, and both pathways start with the production of inositol 3-phosphate (InsP3) and end with PA [23]. In total, 36 genes known to be involved in PA metabolism, as well as in inositol phosphate metabolism (ko00562), displayed differences in expression between *2mlpa* and 2MWT of more than twofold at the late seed developmental stage (stage 5) (Supplementary Table S4). The down-regulated DEG Glyma.11G238800, encoding MIPS1, was known to be mutated in *2mlpa* (Figure 5A), this result was consistent with the previous research that *MIPS1* expression by qRT-PCR revealed a significant reduction in 22 DAF in mutant *Gm-lpa*-TW-1 with single mutated *MIPS1* gene, it means that cross-breeding with mutant of *Gm-lpa*-ZC-2 did not alter the regulated way of MIPS1 [24]. Myo-inositol phosphate synthase (MIPS), are also known to regulate pathways associated with seed and growth development [13,25,26]. Such as, in Arabidopsis (*Arabidopsis thaliana*), the disruption of MIPS severely reduces levels of inositol and PA, resulting in smaller plants, programmed cell death, or hypersensitivity of plants to high-intensity light stress [27-29]. In previous study, mutant *Gm-lpa*-TW-1 showed acceptable agronomic and nutritional traits except low seed field emergence, here, seeds of *2mlpa* also showed a lower field emergence but much better than *Gm-lpa*-TW-1. The improvement of seeds of *2mlpa* would be benefit from the cross-breeding with mutant *Gm-lpa*-ZC-2 which had normal seeds field emergence [12]. Interestingly, two DEGs (Glyma.05G180600 and Glyma.18G018600) also encoding *myo*-inositol 1-phosphate synthase were up-regulated only at the late seed developmental stage. We hypothesized that the DEGs encoding *myo*-inositol 1-phosphate synthase are functionally redundant and, therefore, that

other *MIPS1* genes (Glyma.05G180600 and Glyma.18G018600) compensated for the absence of Glyma.11G238800. The reduction of PA [30] might be due to these DEGs involved in PA metabolism.

The other mutation gene was *IPK1* (Glyma.14G072200), which encodes InsP5 2-kinase and catalyzes the final step in PA biosynthesis [20] ; this gene was down-regulated in *2mlpa* only at the late seed developmental stage (stage 5)，which is not consistent with the RT-PCR results in mutant *Gm-lpa*-ZC-2 , the relative expression levels of ZC-*lpa* were lower than its wild type parent Zhechun No. 3 at the early development stages but higher at later stages [15]. The expression of *IPK1* gene in *2mlpa* might be affected by cross-breeding. In rice, RNAi-mediated seed-specific silencing of the *IPK1* gene leads to a significant reduction in phytate levels and a concomitant increase in the amount of inorganic phosphate (Pi) [20]. Constitutive suppression of such enzymes involved in phytate biosynthesis could be detrimental to plant growth and development [25,31-34]. But in our research, we did not find any defects induced by single IPK1 mutation in *Gm-lpa*-ZC-2; on the contrary, some traits of *Gm-lpa*-TW-1 with single *MIPS1* mutation gene were improved by the cross with *Gm-lpa*-ZC2.

We also observed that some inositol polyphosphate kinase and phosphatase genes involved in PA metabolism were up-regulated only at the late seed developmental stage (stage 5), e.g., two *ITPK1* genes (encoding inositol-1,3,4-trisphosphate 5/6-kinase), two *PI4KA* genes (encoding phosphatidylinositol 4-kinase A),and five *INPP5B/F* genes (encoding inositol polyphosphate 5-phosphatase). These results indicate that the mutations in *2mlpa* mainly affect PA biosynthesis in the late seed developmental stage and induce complex chain reactions in PA metabolism. It has been reported that *Gm-lpa*-ZC-2 (one of the parents of *2mlpa*) and *2mlpa* mutants exhibits significantly increased contents of inositol 3-phosphate (InsP3) and inositol 4-phosphate (InsP4) compared to the wild type [8]. InsP3 and InsP4 are intermediary metabolites in PA metabolism. The up-regulation of the DEGs encoding inositol polyphosphate kinases and phosphatases is consistent with the increase of InsP3 and InsP4 in *lpa* lines. The impacts of single mutated gene on these PA metabolic genes' expression still need to be done in the future research.

2.6 DEGs involved in photosynthesis at the late seed developmental stage

Photosynthesis and photosynthesis-antenna protein (ko00195 and ko00196) pathways were enriched across the five *2mlpa* seed development stages, especially in the late seed developmental stage (stage 5) (Supplementary Table S5). In total, 117 DEGs were associated with photosynthesis and photosynthesis-antenna proteins. Most of the photosynthesis-related DEGs (81 of 94 DEGs) from the late seed developmental stage (stage 5) were down-regulated in the *2mlpa* mutant compared with 2MWT (Figure 5B and Figure 5C). These DEGs encode different subunits in photosystem I, photosystem II, the cytochrome b6/f complex, the photosynthetic electron transport system, F-type-ATPase, and the light-harvesting chlorophyll protein complex. Down-regulation of the photosystem complex building subunits suggests less photosynthesis in the *2mlpa* mutant lines, which has also been found in other *lpa* soybean mutants [35,36]. *Myo*-Inositol biosynthesis is first catalyzed by MIPS (one mutant in *2mlpa*), followed by dephosphorylation of L-*myo*-inositol 1-phosphate to *myo*-inositol. *Myo*-Inositol protects the photosynthesis process in *Mesembryanthemum crystallinum*, which is consistent with the change in *myo*-inositol and down-regulation of photosynthesis-related DEGs in *2mlpa* [8,36].

The photosynthesis-antenna proteins pathway modulates plant defense responses induced by abiotic signals such as light and temperature [20,37,38]. The light-harvesting chlorophyll protein complex (LHC) belongs to the photosynthesis-antenna proteins pathway. The chief function of the LHC is to collect and transfer light energy to photosynthetic reaction centers [38]. In this study, DEGs associated with the LHC were significantly enriched in late soybean seed development (stage 5). In total, 16 DEGs encoding chlorophyll a-b binding proteins related to 9 of 11 LHC subunits were down-regulated only in late seed development (stage 5) in the 2*mlpa* line (Figure 5B). Such diminished expression may reduce light harvesting and the synthesis of plant energy. These results suggest that there is a special relationship between the mutations in *2mlpa* and the

photosynthesis-antenna proteins pathway during late seed development. Furthermore, light provides energy for photosynthesis, and it also functions as a signal to direct plant growth and development. Arabidopsis *mips1* mutants (mutation in *MIPS1* gene of *Arabidopsis thaliana*) display light intensity-dependent cell death [28,39]. We observed that most of the photosynthesis-related DEGs were down-regulated in *2mlpa* in the stage 5, consistent with previous reports. These results suggest that a mutation in *MIPS1* or low PA may decrease photosynthesis during late seed development. These down-regulated genes involved in photosynthesis might contribute to some unfavorable traits in *lpa* mutants, such as lower seed field emergence.

2.7 DEGs involved in starch and sucrose metabolism

In higher plants, sucrose is the major transport form of carbohydrates to meet the carbon and energy needed for growth and the synthesis of storage reserves [40]. Furthermore, sucrose is a signaling molecule affecting plant development processes such as plant growth, plant defense, and the regulation of flowering and the development of storage organs [3,30,40]. The *lpa* soybean *2mlpa* exhibits a high-sucrose phenotype. In total, 67 starch and sucrose metabolism (ko00500)-related DEGs were almost exclusively up-regulated in *2mlpa* compared with 2MWT at the late stage of seed development (Figure 5D, Supplementary Table S6). These DEGs mainly include genes encoding key enzymes involved in starch biosynthesis and degradation, sucrose metabolism, inter-conversion between sugar and starch, and transporters. The *lpa* soybean *2mlpa* showed a series of chain reactions in the starch and sucrose metabolism pathways. *Myo*-Inositol is an important intermediate metabolite in PA biosynthesis and sucrose metabolism. The biosynthesis of raffinosaccharide proceeds by the reversible addition of galactose units from galactinol to sucrose with the help of *myo*-inositol [41]. Variations in *myo*-inositol level might therefore reduce the biosynthesis of raffinosaccharide, increasing sucrose levels [30,41]. Mutation of the *MIPS1* gene results in a reduction in PA level, a series of changes in intermediate metabolites of PA biosynthesis (such as *myo*-inositol), and concomitantly a decrease in raffinosaccharide level, which might result in an increase in sucrose level in *Gm-lpa*-TW-1 [30]. The mutation of the *MIPS1* gene in *2mlpa* might lead to the higher sucrose level observed, and the upregulation of 67 starch and sucrose metabolism (ko00500)-related DEGs.

It has been reported that soybean *lpa* mutants regulate cell wall synthesis [35]. The enzyme β-glucosidase is associated with the cell wall [42,43]. β-Glucosidase in the leaves of Arabidopsis may play a vital role in the breakdown of cell wall polysaccharides, which in turn provide soluble sugar for remobilization and completion of the energy-dependent senescence process [44]. In this study, 13 DEGs encoding β-glucosidase were upregulated in the late stage of seed development in *2mlpa* compared with 2MWT. It is interesting that loss of photosynthesis is accompanied by a significant increase in the activity of cell wall-bound b-glucosidase in Arabidopsis [45], which is consistent with downregulation of DEGs related to photosynthesis and upregulation of DEGs encoding β-glucosidase in *2mlpa* soybean seeds. The clear role of β-glucosidase in *2mlpa* mutants lies outside of the scope of this study; however, it is possible that *2mlpa* mutants may cause the breakdown of cell wall polysaccharides.

DEGs directly associated with sucrose biosynthesis encoded two enzymes: sucrose-phosphate synthase (SPS; EC 2.4.1.14), which catalyzes the biosynthesis of sucrose 6-phosphate from UDP-glucose and fructose 6-phosphate, and sucrose synthase (SuSy; EC 2.4.1.13) [46,47]. There was one up-regulated DEG (Glyma.13G161600) encoding SPS and two up-regulated DEGs (Glyma.09G073600, Glyma.15G182600) encoding SuSy in *2mlpa* at the late seed developmental stage (stage 5). These up-regulated DEGs may result in high sucrose content in 2*mlpa* soybean seeds. Another *lpa* soybean line, V99-5089 (mutation in the *MIPS1* gene), also produces soybean seeds with low PA and high sucrose contents [48]. It is possible that a mutation in *MIPS1* or low PA may improves the sucrose content of soybean seeds, which is consistent with the high-sucrose phenotype of *lpa* mutant seeds [30].

2.8 DEGs involved in defense mechanisms

DEGs related to plant–pathogen interaction (ko04626) and responses to stress (GO:0006950) were enriched in *2mlpa* at the five seed developmental stages, especially in the early (stage 1) and late (stage 5) stages. In total, 604 DEGs were related to defense mechanisms, and most of these were exclusively regulated at one stage (Figure 5E, Supplementary Table S7). For instance, most of the DEGs (105 of 146 DEGs) enriched at early seed developmental stages were not differentially expressed at the late seed developmental stage (stage 5); meanwhile, most of the DEGs (366 of 447 DEGs) enriched at stage 5 were only differentially expressed at this stage. These results suggest that the genes affected by the *lpa* mutations are different at each developmental stage.

PA is a natural antioxidant [49]. In plants, PA is also a key signaling molecule generated in response to the drought stress hormone abscisic acid, which stimulates Ca^{2+} release in guard cells [50]. Furthermore, PA is important for the maintenance of basal resistance to plant pathogens [51]. InsP3 and InsP4 are intermediary metabolites in PA metabolism that act as second messengers to regulate cytosolic Ca^{2+} levels [52]. Hence, the reduction of PA and increase in InsP3 and InsP4 in *lpa* soybean may cause a series of complex chain reactions in defense mechanisms [8,30]. Here, 604 DEGs were related to defense mechanisms, with about a third of these DEGs encoding plant disease resistance (R) proteins including LRR receptor-like serine/threonine-protein kinase, calmodulin, transcription factors, peroxidases, and mitogen-activated proteins kinases (MAPKs).

Plant disease resistance (R) proteins are important components of the genetic resistance mechanism in plants, and several *R* genes confer resistance to a wide spectrum of plant pathogens [53]. In this study, 64 *R* genes were up-regulated in *2mlpa* compared with 2MWT at the late developmental stage. The most prevalent domain of the characterized R proteins is the leucine-rich repeat (LRR), which is the major determinant of pathogen recognition. The LRR receptor-like R protein consists of an extracellular LRR domain and a transmembrane motif [54]. Here, 21 DEGs encoding LRR receptor-like serine/threonine protein kinase FLS2 were up-regulated at the late stage of seed development. In Arabidopsis, FLS2 is a transmembrane receptor kinase that activates antimicrobial defense responses [55]. Brassinosteroid insensitive 1-associated receptor kinase 1 (BAK1) is also an LRR receptor-like protein kinase and is involved in signaling by FLS2 [56]. Plants affected at *BAK1* exhibit reduced responses to various microbial elicitors and cold shock protein [56,57]. In this study, 28 DEGs encoding BAK1 were up-regulated in 2*mlpa* at stage 5. The role of LRR genes in 2*mlpa* soybean seed development is still unknown, and further studies are required to confirm the association between the up-regulation of LRR genes and the *lpa* soybean line 2*mlpa*.

Additionally, some NAC, MYC2, and WRKY transcription factor family genes that regulate defense-related genes were also differentially expressed at the late stage of seed development (stage 5) in the 2*mlpa* mutant. For example, two DEGs encoding WRKY70 were up-regulated. WRKY70 negatively modulates cell wall-associated defenses against pathogens and leaf senescence in Arabidopsis [58-60]. MAPKs are key enzymes mediating adaptive responses to various abiotic and biotic stresses [61]. Here, 10 DEGs encoding MAPKs were differentially regulated between *2mlpa* and 2MWT at the five seed stages. These defense-related DEGs may be affected by changes in PA or its intermediary metabolites. Interactions between plants and pathogens involve bidirectional recognition. Upon pathogen attack, plant defense responses involve modulation of the expression of a large number of genes [62]. In total, 418 DEGs were involved in “plant–pathogen interaction” in *2mlpa* soybean seeds, suggesting that *MIPS1*- and *IPK1*-induced processes share some common features with the defense mechanism against pathogens in plants. Overall, our results verify the correlation between PA and defense mechanisms.

In conclusion, these results contribute to our further understanding of the metabolic regulatory network of the *2mlpa* mutant during seed development.

3. Methods

3.1. Genetic material and background

A homozygous double-mutant soybean line with low PA, designated *2mlpa* (*mips1/ipk1*), and a sibling line with homozygous non-mutant alleles and normal PA, designated 2MWT (MIPS1/IPK1), were produced in the present study. These lines were developed by crossing *Gm-lpa*-TW-1 with *Gm-lpa*-ZC-2. *Gm-lpa*-TW-1 was developed using gamma irradiation of 'Taiwan 75' (a vegetable soybean cultivar widely grown in Zhejiang Province), which produced a 2-bp deletion in *GmMIPS1* [12]. Similarly, the *Gm-lpa*-ZC-2 soybean line carries a G-A point mutation in the *GmIPK1* gene compared with that of its commercial wild-type cultivar, 'Zhechun No. 3' [15]. After crossing, F_1 plants were grown and harvested in individuals. F_2 plants (as well as the original crossing parents) were grown, and genotyped by high inorganic phosphorus testing and high-resolution melting curve (HRM) analysis [12,15], and classified into homozygous wild-type (HWT), homozygous *lpa* mutant (HM), and heterozygous ones. The F_3 seeds from five heterozygous F_2 individual were planted and again genotyped for identification of heterozygous F_3 plants. The F_4 seeds of five individual heterozygous F_3 plants were grown into F_4 populations for identification of homozygous wild-type and lpa mutant plants (Figure 1). The selection process of F_5-F_6 plants was similar to that of F_3-F_4 generation. F_7 populations were developed from five heterozygous F6 individual and segregated, and their genotypes were determined HWT and HM seeds from total 10 F_7 plants were used to produce 5 HWT and 5 HM F_8 lines which were used in our research, the genotypes of which were determined by HRM analysis for the selection of *lpa* and non-*lpa* lines [63]. The PA and inositol phosphate isomers content of seeds of F_8 lines was evaluated and 2.32mg/g and 18.2mg/g PA content were detected in *2mlpa* and 2WMT lines, respectively [16].

3.2. Plant growth and RNA extraction

For each of the experimental F8 lines (*2mlpa* and 2MWT), 50 plants were grown in a greenhouse at 25 °C with a 14/10 h (day/night) photoperiod. Seed length was the criterion for defining five developmental stages of soybean seeds as follows: stage 1, 7 days after flowering (DAF); stage 2, 12 DAF; stage 3, 17 DAF; stage 4, 22 DAF; and stage 5, 27 DAF [15,30]. For each stage, seeds from 50 plants were collected and mixed to create one sample. Three parallel samples collected at each stage were used as biological replicates for RNA extraction. The sampled seeds were immediately ground to a fine powder, and total RNA was extracted using an E.Z.N.A. plant RNA kit (Omega Bio-tek, Inc., USA) according to the manufacturer's protocol. Genomic DNA contamination was eliminated using RQ1 RNase-Free DNase (Promega, USA).

3.3. Library construction and sequencing

The quality of total RNA was assessed using a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) and checked using agarose gel electrophoresis. The final concentration of all RNA samples was adjusted to 500 ng/μl after quantification. High-quality RNA samples were used for library preparation. In total, 30 cDNA libraries were generated using a SMART™ cDNA Library Construction Kit (Takara Bio-tek, Inc., Japan) and tested using an Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System (Thermo Fisher, USA). Raw reads were generated on the Illumina Hi Seq 4000 platform (Illumina, USA) at iGENE Biological Technology (Hangzhou, China).

3.4. Sequence assembly

Raw reads were filtered by removing low-quality sequences and rRNA reads. Assembly of the clean reads was then accomplished using SOAPaligner/SOAP2 [64]. The comparison software Tophat [65] (http://ccb.jhu.edu/software/tophat/index.shtml) was used to compare the clean data from each sample with the reference genome (*Glycine_max_v2.0*). Mismatches of no more than five bases were allowed in the alignment.

3.5. Gene annotation and analysis of differentially expressed genes (DEGs)

All genes were analyzed by sequence alignment to the NR (NCBI non-redundant protein), Swiss-Prot, GO (Gene Ontology), COG (Clusters of Orthologous Groups) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases using BLAST2GO (https://www.blast2go.com/).

Gene expression levels were computed and normalized as reads per kb per million fragments (RPKM). The RPKM method can eliminate the influence of gene length and sequencing quantity on the expression data of genes. The calculated gene expression levels can be used directly to compare differences in gene expression between different products. Differential gene expression was analyzed using the DESeq function in Bioconductor (http://www.bioconductor.org). The false discovery rate (FDR) was obtained through correcting the *P* value using the Benjamini-Hochberg correction method [66]. The screening conditions for significant DEGs were FDR≤0.05 and |log2Ratio|≥1.

The TopGO function in the Bioconductor software was used to perform gene ontology (GO) enrichment analysis of the DEGs with an elimination threshold *P* value of 0.01 [67]. Meanwhile, hypergeometric tests were used to perform KEGG pathway enrichment analysis of the DEGs with 0.05 FDR corrections [68].

3.6. Reverse-transcription quantitative PCR (RT-qPCR)

To validate the data obtained from RNA-Seq, RT-qPCR was performed on eight DEGs. First-strand cDNA was synthesized from high-quality total RNA (see above) using a RevertAid First Strand cDNA Synthesis Kit (Thermofisher Scientific Inc., Rockford, Illinois, USA). Specific primers (Supplementary Table S1) were designed using Primer-Blast from the NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The housekeeping gene *ACT11* was used to normalize expression levels of the selected genes. All primers are shown in Supplementary Table S1. RT-qPCR analysis of the eight genes was conducted in triplicate using SYBR Green Real-time PCR Master Mix (Toyobo, Osaka, Japan) on a LightCycler® 480 Real-Time PCR System (Roche Diagnostics GmbH, Mannheim, Germany). Relative gene expression levels were determined using the corrected relative $2^{-\Delta\Delta CT}$ method.

Data Availability Statement

All raw data are published in the NCBI (http://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA522338. Other data included in this study are available upon request by contact with the corresponding author.

Declarations

Acknowledgements

This research was supported by the Ministry of Science and Technology of People's Republic of China (2017YFE0111000), EUCLEG Horizon 2020 of European Union (No.727312) the National Key R&D Program of China (2017YFD0101500), and the Science and Technology Department of Zhejiang Province (grant no. 2016C02050-10) to FJY. Our heartfelt thanks go to the anonymous reviewers who offered their critical comments for the improvement of this paper.

Author contributions

HJ performed the statistical analysis and drafted the manuscript. XY carried out the seed germination experiment. QY helped for the development of population. XF participated in planting and collecting of materials. FY designed the study and helped draft the manuscript. All authors have read and approved this manuscript.

References

1 Rose, T. J., Liu, L. & Wissuwa, M. Improving phosphorus efficiency in cereal crops: Is breeding for reduced grain phosphorus concentration part of the solution? *Frontiers in plant science* **4**, 444 (2013).

2 Raboy, V. Approaches and challenges to engineering seed phytate and total phosphorus. *Plant Science* **177**, 281-296 (2009).

3 Singh, S. K., Barnaby, J. Y., Reddy, V. R. & Sicher, R. C. Varying Response of the Concentration and Yield of Soybean Seed Mineral Elements, Carbohydrates, Organic Acids, Amino Acids, Protein, and Oil to Phosphorus Starvation and CO2 Enrichment. *Frontiers in plant science* **7**, 1967, doi:10.3389/fpls.2016.01967 (2016).

4 Pilu, R. *et al.* Phenotypic, genetic and molecular characterization of a maize low phytic acid mutant (lpa241). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **107**, 980-987, doi:10.1007/s00122-003-1316-y (2003).

5 Bhati, K. K. *et al.* Differential expression of structural genes for the late phase of phytic acid biosynthesis in developing seeds of wheat (Triticum aestivum L.). *Plant science : an international journal of experimental plant biology* **224**, 74-85, doi:10.1016/j.plantsci.2014.04.009 (2014).

6 Aluru, M. R., Rodermel, S. R. & Reddy, M. B. Genetic modification of low phytic acid 1-1 maize to enhance iron content and bioavailability. *Journal of agricultural and food chemistry* **59**, 12954-12962, doi:10.1021/jf203485a (2011).

7 Thavarajah, P., Thavarajah, D. & Vandenberg, A. Low phytic acid lentils (Lens culinaris L.): a potential solution for increased micronutrient bioavailability. *Journal of agricultural and food chemistry* **57**, 9044-9049, doi:10.1021/jf901636p (2009).

8 Frank, T., Norenberg, S. & Engel, K. H. Metabolite profiling of two novel low phytic acid (lpa) soybean mutants. *Journal of agricultural and food chemistry* **57**, 6408-6416, doi:10.1021/jf901019y (2009).

9 Zhao, N. C. *et al.* Characteristics of Grain Starch Synthesis at Filling Stage and Translocation of Carbohydrates in Leaves and Sheaths for Low Phytic Acid Mutant Rice. *Acta Agronomica Sinica* **34**, 1977-1984 (2008).

10 Suzuki, M., Tanaka, K., Kuwano, M. & Yoshida, K. T. Expression pattern of inositol phosphate-related enzymes in rice (Oryza sativa L.): implications for the phytic acid biosynthetic pathway. *Gene* **405**, 55-64, doi:10.1016/j.gene.2007.09.006 (2007).

11 Vincent, J. A., Stacey, M., Stacey, G. & Bilyeu, K. D. Phytic Acid and Inorganic Phosphate Composition in Soybean Lines with Independent IPK1 Mutations. *Plant Genome* **8** (2015).

12 Yuan, F. J. *et al.* Generation and characterization of two novel low phytate mutations in soybean (Glycine max L. Merr.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **115**, 945-957, doi:10.1007/s00122-007-0621-2 (2007).

13 Yuan, F. *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* **17**, 16, doi:10.1186/s12870-016-0953-7 (2017).

14 Shunmugam, A. S. *et al.* Accumulation of Phosphorus-Containing Compounds in Developing Seeds of Low-Phytate Pea (Pisum sativum L.) Mutants. *Plants***4**, 1-26, doi:10.3390/plants4010001 (2014).

15 Yuan, F. J. *et al.* Identification and characterization of the soybean IPK1 ortholog of a low phytic acid mutant reveals an exon-excluding splice-site mutation. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik***125**, 1413-1423, doi:10.1007/s00122-012-1922-7 (2012).

16 Goßner, S. *et al.* Impact of Cross-Breeding of Low Phytic Acid MIPS1 and IPK1 Soybean (Glycine max L. Merr.) Mutants on Their Contents of Inositol Phosphate Isomers. *Journal of agricultural and food chemistry***67**, 247-257, doi:10.1021/acs.jafc.8b06117 (2019).

17 Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic acids research***32**, D277 (2003).

18 Ho-Seok, L. *et al.* InsP6-Sensitive Variants of the Gle1 mRNA Export Factor Rescue Growth and Fertility Defects of the ipk1 Low-Phytic-Acid Mutation in Arabidopsis. *The Plant cell***27**, 417-431 (2015).

19 Chaouch, S. & Noctor, G. Myo-inositol Abolishes Salicylic Acid-dependent Cell Death and Pathogen Defence Responses Triggered by Peroxisomal Hydrogen Peroxide. *New Phytologist***188**, 711-718 (2010).

20 Ali, N. *et al.* Development of low phytate rice by RNAi mediated seed-specific silencing of inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene (IPK1). *PloS one***8**, e68161 (2013).

21 Zhang, J. *et al.* Overexpression of PeMIPS1 confers tolerance to salt and copper stresses by scavenging reactive oxygen species in transgenic poplar. *Tree physiology*, doi:10.1093/treephys/tpy028 (2018).

22 Zhai, H. *et al.* A myo-inositol-1-phosphate synthase gene, IbMIPS1, enhances salt and drought tolerance and stem nematode resistance in transgenic sweet potato. *Plant biotechnology journal***14**, 592-602, doi:10.1111/pbi.12402 (2016).

23 Raboy, V. Forward genetics studies of seed phytic acid. *Israel Journal of Plant Sciences***55**, 171-181 (2007).

24 Yuan, F. J. Characterization of D-myo-inositol 3-phosphate Synthase Gene Expression in Two Soybean Low Phytate Mutants. *Journal of Nuclear Agricultural Sciences* (2013).

25 Zhao, H. *et al.* Disruption of OsSULTR3;3 reduces phytate and phosphorus concentrations and alters the metabolite profile in rice grains. *The New phytologist***211**, 926-939, doi:10.1111/nph.13969 (2016).

26 Zhang, S. *et al.* Analysis of weighted co-regulatory networks in maize provides insights into new genes and regulatory mechanisms related to inositol phosphate metabolism. *BMC genomics***17**, 129, doi:10.1186/s12864-016-2476-x (2016).

27 Luo, Y. *et al.* D-myo-inositol-3-phosphate affects phosphatidylinositol-mediated endomembrane function in Arabidopsis and is essential for auxin-regulated embryogenesis. *The Plant cell***23**, 1352-1372 (2011).

28 Donahue, J. L. *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***22**, 888-903 (2010).

29 Chen, H. & Xiong, L. M. Myo-inositol-1-phosphate synthase is required for polar auxin transport and organ development. *Journal of Biological Chemistry***285**, 24238-24247 (2010).

30 Yuan, F. J. *et al.* Effects of two low phytic acid mutations on seed quality and nutritional traits in soybean (Glycine max L. Merr). *Journal of agricultural and food chemistry***57**, 3632-3638, doi:10.1021/jf803862a (2009).

31 Marathe, A. *et al.* Exploring the role of Inositol 1,3,4-trisphosphate 5/6 kinase-2 (GmITPK2) as a dehydration and salinity stress regulator in Glycine max (L.) Merr. through heterologous expression in E. coli. *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale***123**, 331-341, doi:10.1016/j.plaphy.2017.12.026 (2017).

32 Zhao, H. J. *et al.* Characterization of OsMIK in a rice mutant with reduced phytate content reveals an insertion of a rearranged retrotransposon. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik***126**, 3009-3020, doi:10.1007/s00122-013-2189-3 (2013).

33 Nunes, A. C. *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***224**, 125-132, doi:10.1007/s00425-005-0201-0 (2006).

34 Denstadli, V., Vestre, R., Svihus, B., Skrede, A. & Storebakken, T. Phytate degradation in a mixture of ground wheat and ground defatted soybeans during feed processing: effects of temperature, moisture level, and retention time in small- and medium-scale incubation systems. *Journal of agricultural and food chemistry***54**, 5887-5893, doi:10.1021/jf0605906 (2006).

35 Redekar, N. R. *et al.* Genome-wide transcriptome analyses of developing seeds from low and normal phytic acid soybean lines. *BMC genomics***16**, 1074, doi:10.1186/s12864-015-2283-9 (2015).

36 Nelson, D. E., Rammesmayer, G. & Bohnert, H. J. Regulation of cell-specific inositol metabolism and transport in plant salinity tolerance. *The Plant cell***10**, 753-764 (1998).

37 Kumar, V. *et al.* Probing Phosphorus Efficient Low Phytic Acid Content Soybean Genotypes with Phosphorus Starvation in Hydroponics Growth System. *Applied biochemistry and biotechnology***177**, 689-699, doi:10.1007/s12010-015-1773-1 (2015).

38 Xia, Y. *et al.* Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (Lhcb1) associated with agronomic traits in barley. *PloS one***7**, e37573, doi:10.1371/journal.pone.0037573 (2012).

39 Ma, L. *et al.* Arabidopsis FHY3 and FAR1 Regulate Light-Induced myo-Inositol Biosynthesis and Oxidative Stress Responses by Transcriptional Activation of MIPS1. *Molecular plant***9**, 541-557, doi:10.1016/j.molp.2015.12.013 (2016).

40 Tognetti, J. A., Pontis, H. G. & Martinez-Noel, G. M. Sucrose signaling in plants: a world yet to be explored. *Plant signaling & behavior***8**, e23316, doi:10.4161/psb.23316 (2013).

41 Karner, U. *et al.* myo-Inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. *Journal of experimental botany***55**, 1981-1987, doi:10.1093/jxb/erh216 (2004).

42 Minic, Z. Physiological roles of plant glycoside hydrolases. *Planta***227**, 723-740 (2008).

43 Lee, E. J., Matsumura, Y., Soga, K., Hoson, T. & Koizumi, N. Glycosyl hydrolases of cell wall are induced by sugar starvation in Arabidopsis. *Plant & cell physiology***48**, 405-413, doi:10.1093/pcp/pcm009 (2007).

44 Mohapatra, P. K. *et al.* Senescence-induced loss in photosynthesis enhances cell wall beta-glucosidase activity. *Physiologia plantarum***138**, 346-355 (2010).

45 Mohapatra, P. K. *et al.* Senescence-induced loss in photosynthesis enhances cell wall beta-glucosidase activity. *Physiologia plantarum***138**, 346-355, doi:10.1111/j.1399-3054.2009.01327.x (2010).

46 Bahaji, A. *et al.* Characterization of multiple SPS knockout mutants reveals redundant functions of the four Arabidopsis sucrose phosphate synthase isoforms in plant viability, and strongly indicates that enhanced respiration and accelerated

starch turnover can alleviate the blockage of sucrose biosynthesis. *Plant science : an international journal of experimental plant biology***238**, 135-147, doi:10.1016/j.plantsci.2015.06.009 (2015).

47 But, S. Y., Khmelenina, V. N., Reshetnikov, A. S. & Trotsenko, Y. A. Bifunctional sucrose phosphate synthase/phosphatase is involved in the sucrose biosynthesis by Methylobacillus flagellatus KT. *FEMS microbiology letters***347**, 43-51, doi:10.1111/1574-6968.12219 (2013).

48 Mozzoni, L., Shi, A. & Chen, P. Genetic Analysis of High Sucrose, Low Raffinose, and Low Stachyose Content in V99-5089 Soybean Seeds. *Journal of Crop Improvement***27**, 606-616 (2013).

49 Graf, E. & Eaton, J. W. Antioxidant functions of phytic acid. *Free radical biology & medicine***8**, 61-69 (1990).

50 Lemtiri-Chlieh, F. *et al.* Inositol hexakisphosphate mobilizes an endomembrane store of calcium in guard cells. *Proceedings of the National Academy of Sciences of the United States of America***100**, 10091-10095 (2003).

51 Alex M, M., Bettina, O., Charles A, B., John P, C. & David E, H. A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens. *The Plant Journal***56**, 638-652 (2010).

52 Cancela, J. M., Gerasimenko, O. V., Gerasimenko, J. V., Tepikin, A. V. & Petersen, O. H. Two different but converging messenger pathways to intracellular Ca(2+) release: the roles of nicotinic acid adenine dinucleotide phosphate, cyclic ADP-ribose and inositol trisphosphate. *The EMBO journal***19**, 2549-2557, doi:10.1093/emboj/19.11.2549 (2000).

53 Wan, H. *et al.* Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. *BMC genomics***13**, 502, doi:10.1186/1471-2164-13-502 (2012).

54 Zhang, H. *et al.* A pair of orthologs of a leucine-rich repeat receptor kinase-like disease resistance gene family regulates rice response to raised temperature. *BMC plant biology***11**, 160, doi:10.1186/1471-2229-11-160 (2011).

55 Sun, W. *et al.* Probing the Arabidopsis Flagellin Receptor: FLS2-FLS2 Association and the Contributions of Specific Domains to Signaling Function. *The Plant cell***24**, 1096-1113, doi:10.1105/tpc.112.095919 (2012).

56 Chinchilla, D. *et al.* A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature***448**, 497-500, doi:10.1038/nature05999 (2007).

57 Heese, A. *et al.* The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proceedings of the National Academy of Sciences***104**, 12217-12222, doi:10.1073/pnas.0705306104 (2007).

58 Zhao, H., Jiang, J., Li, K. & Liu, G. Populus simonii x Populus nigra WRKY70 is involved in salt stress and leaf blight disease responses. *Tree physiology***37**, 827-844, doi:10.1093/treephys/tpx020 (2017).

59 Li, J., Zhong, R. & Palva, E. T. WRKY70 and its homolog WRKY54 negatively modulate the cell wall-associated defenses to necrotrophic pathogens in Arabidopsis. *PloS one***12**, e0183731, doi:10.1371/journal.pone.0183731 (2017).

60 Besseau, S., Li, J. & Palva, E. T. WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in Arabidopsis thaliana. *Journal of experimental botany***63**, 2667-2679, doi:10.1093/jxb/err450 (2012).

61 Pitzschke, A., Schikora, A. & Hirt, H. MAPK cascade signalling networks in plant defence. *Current opinion in plant biology***12**, 421-426 (2009).

62 Abdin, M. Z. *et al.Signal Transduction and Regulatory Networks in Plant-Pathogen Interaction: A Proteomics Perspective*. (Springer New York, 2013).

63 Tan, Y. Y. *et al.* Development of an HRM-based, safe and high-throughput genotyping system for two low phytic acid mutations in soybean. *Molecular Breeding***36**, 1-9 (2016).

64 Gu, S., Fang, L. & Xu, X. Using SOAPaligner for Short Reads Alignment. *Curr Protoc Bioinformatics***44**, 11 11 11-17 (2013).

65 Ghosh, S. & Chan, C. K. K. Analysis of RNA-Seq Data Using TopHat and Cufflinks. *Methods in molecular biology***1374**, 339-361 (2016).

66 Thissen, D., Steinberg, L. & Kuang, D. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. *Journal of Educational & Behavioral Statistics***27**, 77-83 (2002).

67 Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics***22**, 1600 (2006).

68 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics***23**, 257-258 (2007).

Tables

Table 1 Library read analysis statistics

Sample ID	Total clean reads	Total base pairs	Total mapped reads	Perfect match	Total unmapped reads
2mlpa-1-I	47,376,510	7,106,476,500	39,217,355	29,254,240	8,159,155
2mlpa-1-II	46,147,142	6,922,071,300	38,174,762	28,505,756	7,972,380
2mlpa-1-III	45,752,784	6,862,917,600	37,927,606	28,179,059	7,825,178
2mlpa-2-I	47,766,190	7,164,928,500	38,332,152	28,840,283	9,434,038
2mlpa-2-II	46,019,862	6,902,979,300	36,858,472	27,854,925	9,161,390
2mlpa-2-III	47,627,596	7,144,139,400	37,613,372	28,285,902	10,014,224
2mlpa-3-I	45,237,408	6,785,611,200	36,559,719	27,658,430	8,677,689
2mlpa-3-II	46,939,696	7,040,954,400	38,163,913	28,687,412	8,775,783
2mlpa-3-III	47,094,904	7,064,235,600	39,249,412	29,157,151	7,845,492
2mlpa-4-I	46,624,724	6,993,708,600	38,733,400	28,405,148	7,891,324
2mlpa-4-II	45,328,594	6,799,289,100	37,799,964	27,797,006	7,528,630
2mlpa-4-III	46,884,742	7,032,711,300	37,231,517	28,195,905	9,653,225
2mlpa-5-I	46,770,398	7,015,559,700	38,807,109	29,708,501	7,963,289
2mlpa-5-II	45,679,520	6,851,928,000	36,180,381	26,871,446	9,499,139
2mlpa-5-III	45,845,714	6,876,857,100	37,756,564	28,344,742	8,089,150
2MWT-1-I	47,105,886	7,065,882,900	35,450,056	25,998,470	11,655,830
2MWT-1-II	46,892,294	7,033,844,100	35,547,215	25,864,330	11,345,079
2MWT-1-III	47,828,524	7,174,278,600	35,999,905	26,378,660	11,828,619
2MWT-2-I	47,286,614	7,092,992,100	36,019,630	26,110,062	11,266,984
2MWT-2-II	46,048,510	6,907,276,500	35,423,782	26,086,498	10,624,728
2MWT-2-III	46,622,194	6,993,329,100	34,971,326	25,850,610	11,650,868
2MWT-3-I	47,317,516	7,097,627,400	37,157,999	27,267,786	10,159,517
2MWT-3-II	46,063,542	6,909,531,300	36,372,452	26,833,737	9,691,090
2MWT-3-III	47,549,838	7,132,475,700	37,382,695	27,523,661	10,167,143
2MWT-4-I	45,889,736	6,883,460,400	36,197,074	26,470,608	9,692,662
2MWT-4-II	46,699,880	7,004,982,000	37,777,280	27,945,411	8,922,600
2MWT-4-III	47,302,502	7,095,375,300	37,508,685	27,875,751	9,793,817
2MWT-5-I	47,662,516	7,149,377,400	38,108,064	28,405,856	9,554,452
2MWT-5-II	46,925,520	7,038,828,000	38,110,466	28,209,436	8,815,054
2MWT-5-III	46,929,246	7,039,386,900	36,868,656	27,610,343	10,060,590

Table 2 Number of DEGs in enriched KEGG pathways at five seed developmental stages, as determined by pathway enrichment analysis.

Pathway	Pathway ID	Stage 1	*P* value	Stage 2	*P* value	Stage 3	*P* value	Stage 4	*P* value	Stage 5	*P* value
Metabolic pathways	ko01100	361	1.1E-12	108	4.2E-11	153	3.9E-09	138	7.4E-06	794	1.8E-02
Biosynthesis of secondary metabolites	ko01110	210	1.9E-08	42	8.1E-03	66	8.6E-03	65	9.3E-02	374	9.9E-03
Plant–pathogen interaction	ko04626	109	3.9E-02	26	4.2E-02	54	2.0E-02	39	6.1E-02	303	5.6E-02
Plant hormone signal transduction	ko04075	67	9.9E-03	12	9.9E-01	27	9.0E-03	38	2.1E-01	252	4.6E-66
Ribosome	ko03010	27	6.0E-03	18	1.1E-04	22	1.3E-03	15	1.1E-01	252	4.9E-02
Protein processing in endoplasmic reticulum	ko04141	24	9.0E-03	2	9.9E-03	14	2.8E-02	5	9.9E-03	113	2.3E-03
Purine metabolism	ko00230	37	3.3E-02	11	5.2E-02	19	8.8E-03	12	3.3E-02	104	1.1E-03
Pyrimidine metabolism	ko00240	39	8.7E-03	9	1.6E-02	19	6.3E-03	12	2.9E-01	100	1.8E-03
Starch and sucrose metabolism	ko00500	30	4.8E-02	5	8.2E-03	14	2.4E-02	14	2.3E-01	75	8.8E-02
Phenylpropanoid biosynthesis	ko00940	42	2.5E-03	9	1.7E-02	14	1.5E-01	14	1.4E-02	74	6.3E-02
Photosynthesis	ko00195	24	1.2E-08	33	2.2E-35	32	7.3E-27	28	4.1E-22	69	3.5E-25
Flavonoid biosynthesis	ko00941	45	3.3E-08	18	3.4E-07	14	1.2E-02	12	4.8E-02	67	9.4E-07
Oxidative phosphorylation	ko00190	25	1.3E-03	30	3.0E-21	29	2.6E-14	7	2.2E-01	67	2.3E-02
Photosynthesis-antenna proteins	ko00196	1	8.0E-03	0		0		5	3.0E-04	22	1.8E-11

Supplementary Information

This file contains Supplementary Table S1-S7

Figures

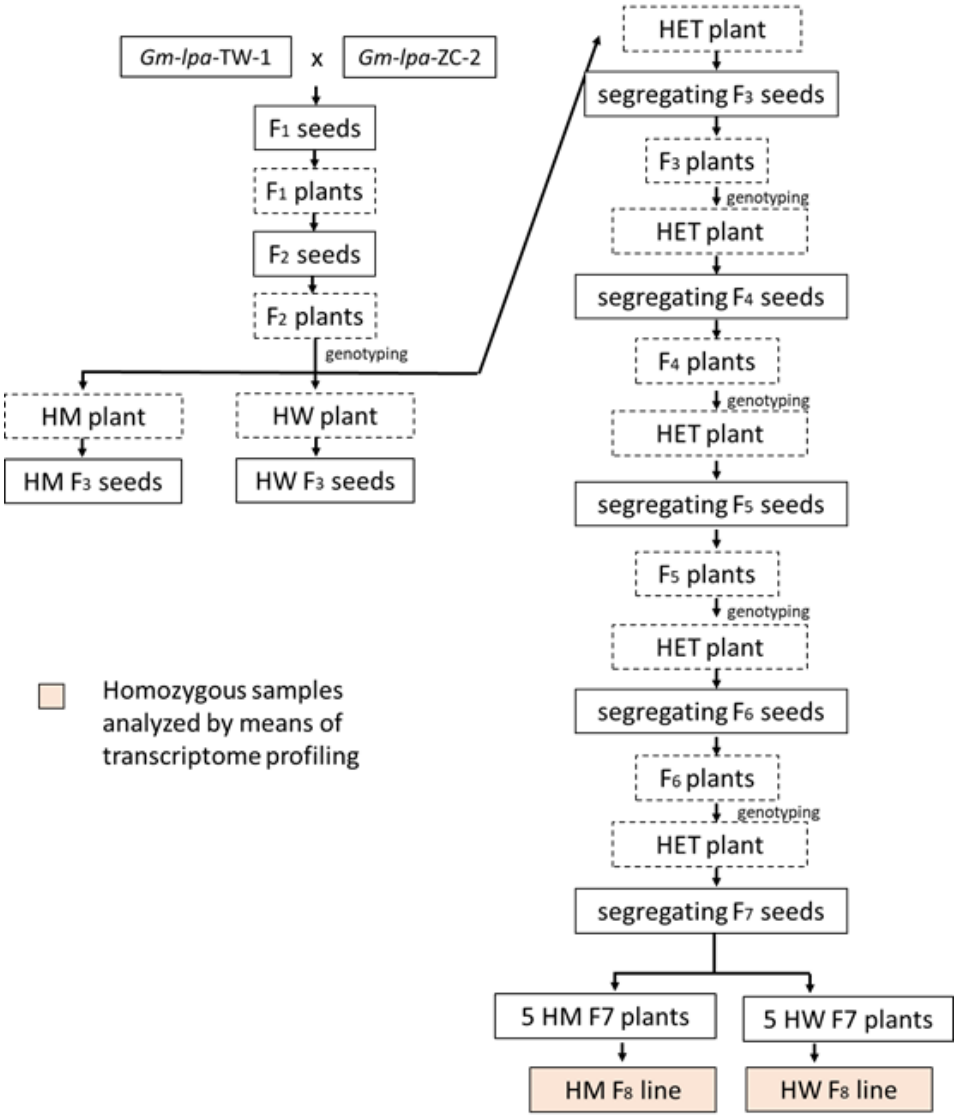


Figure 1

Flowchart of soybean material production by cross-breeding of Gm-lpa-TW-1 and Gm-lpa-ZC-2. HM: homozygous lpa-mutant; HW: homozygous wild-type; HET: heterozygous.

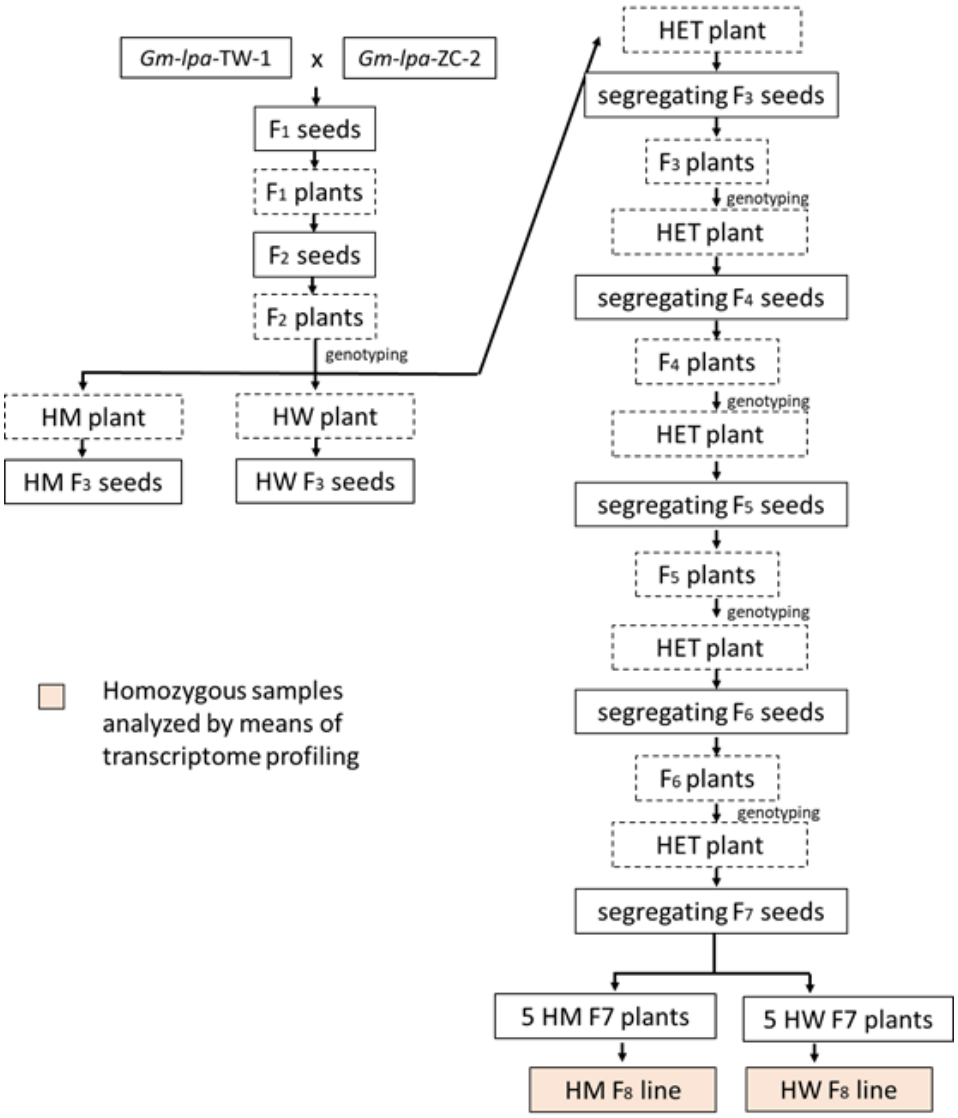


Figure 1

Flowchart of soybean material production by cross-breeding of Gm-lpa-TW-1 and Gm-lpa-ZC-2. HM: homozygous lpa-mutant; HW: homozygous wild-type; HET: heterozygous.

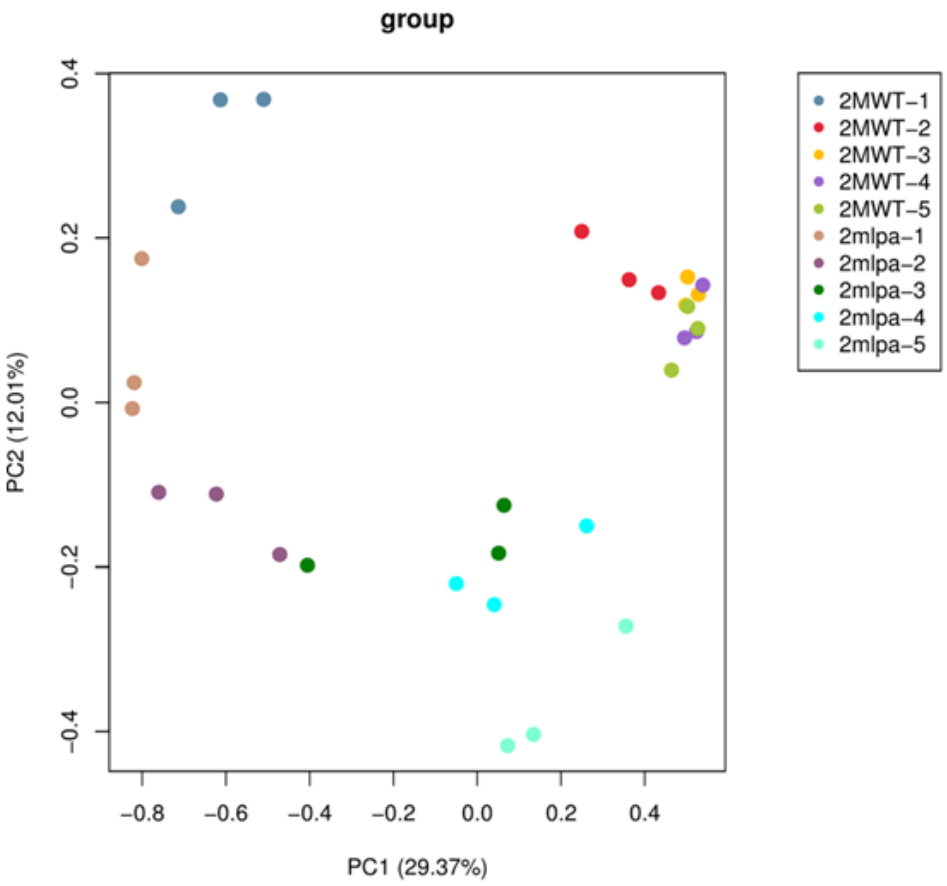


Figure 2

Principal component analysis plots for 30 2mlpa and 2MWT samples.

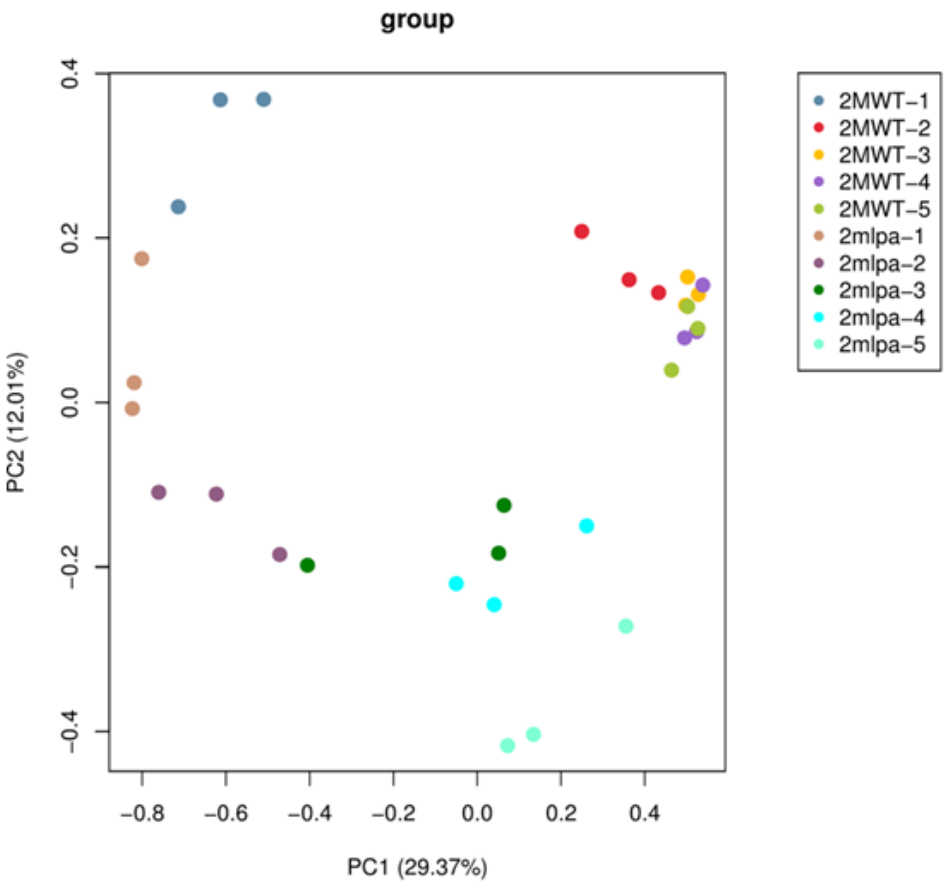
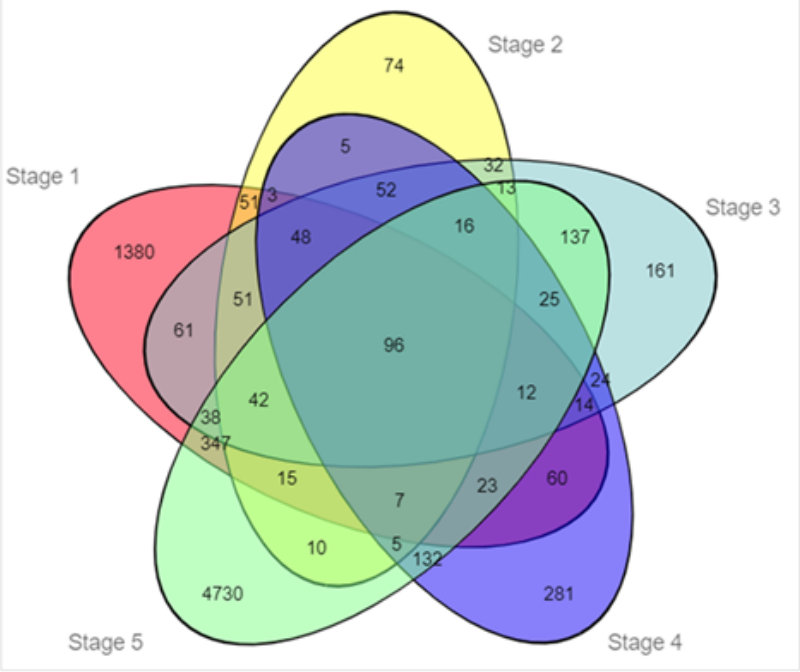


Figure 2

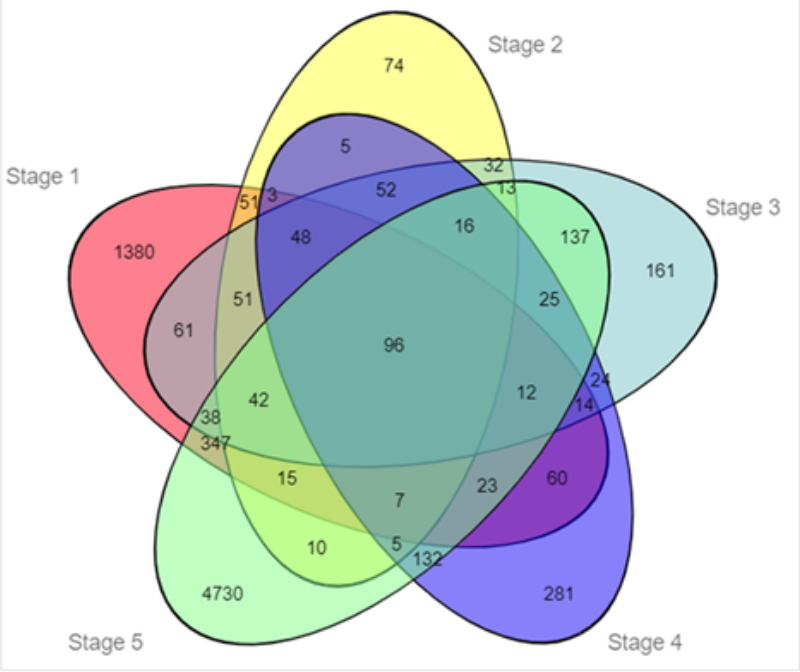
Principal component analysis plots for 30 2mlpa and 2MWT samples.



group	Total	UP	Down
Stage1: 2MWT-1-VS-2mlpa-1	2248	802	1446
Stage2: 2MWT-2-VS-2mlpa-2	520	341	179
Stage3: 2MWT-3-VS-2mlpa-3	822	387	435
Stage4: 2MWT-4-VS-2mlpa-4	803	499	304
Stage5: 2MWT-5-VS-2mlpa-5	5648	2950	2698
Total DEG	7945		

Figure 3

Venn diagram of the co-expressed and uniquely expressed DEGs between 2mlpa and 2MWT at five seed stages, along with the number of DEGs between 2mlpa and 2MWT at the five stages.



group	Total	UP	Down
Stage1: 2MWT-1-VS-2mlpa-1	2248	802	1446
Stage2: 2MWT-2-VS-2mlpa-2	520	341	179
Stage3: 2MWT-3-VS-2mlpa-3	822	387	435
Stage4: 2MWT-4-VS-2mlpa-4	803	499	304
Stage5: 2MWT-5-VS-2mlpa-5	5648	2950	2698
Total DEG	7945		

Figure 3

Venn diagram of the co-expressed and uniquely expressed DEGs between 2mlpa and 2MWT at five seed stages, along with the number of DEGs between 2mlpa and 2MWT at the five stages.

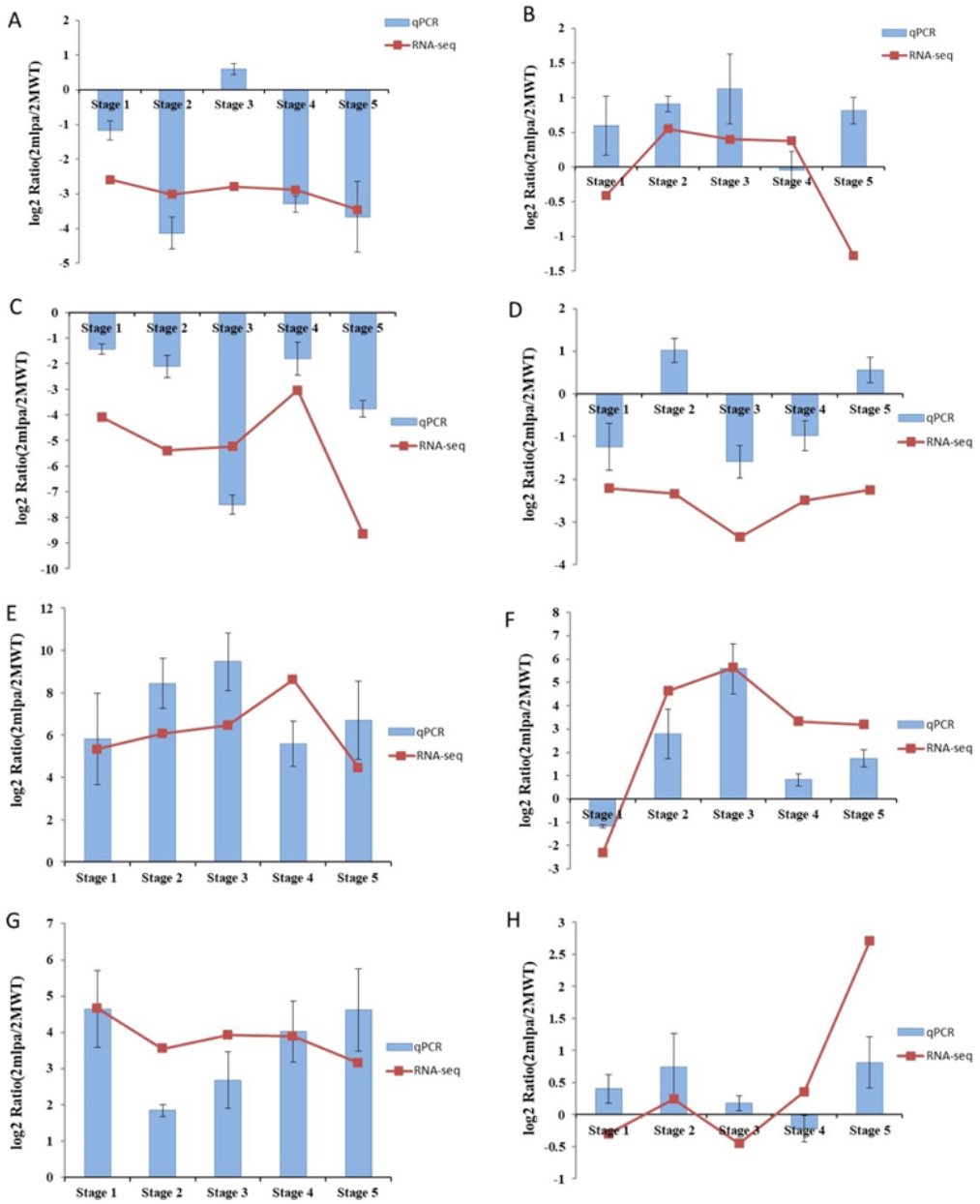


Figure 4

Verification of RNA-Seq results by RT-qPCR. The y-axis represents the log2Ratio (2mlpa/2MWT) in (A) Glyma.11G238800, (B) Glyma.14G072200, (C) Glyma.16G065700, (D) Glyma.12G210600, (E) Glyma.08G109300, (F) Glyma.12G232500, (G) Glyma.16G214900, and (H) Glyma.09G073600 at five seed stages, respectively.

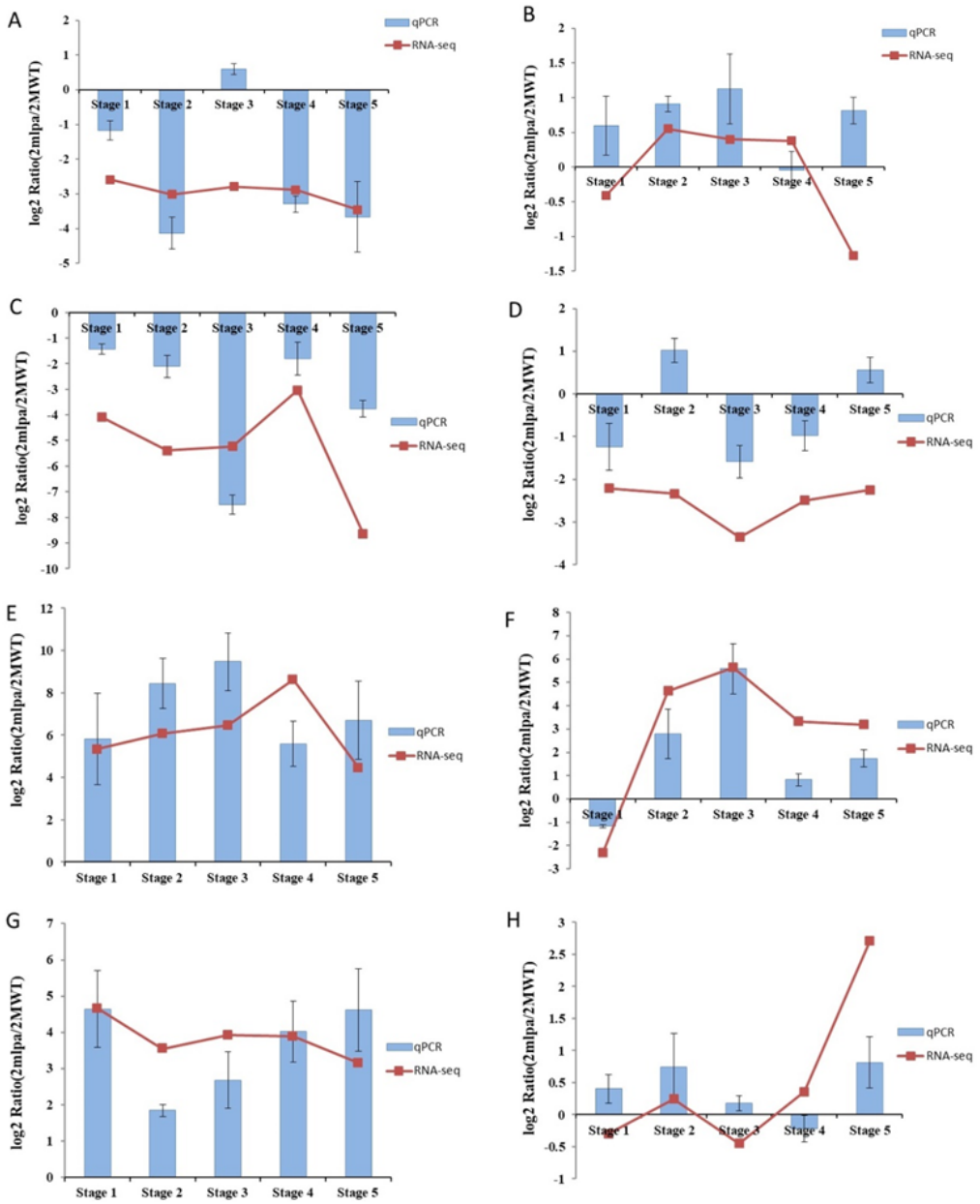


Figure 4

Verification of RNA-Seq results by RT-qPCR. The y-axis represents the log2Ratio (2mlpa/2MWT) in (A) Glyma.11G238800, (B) Glyma.14G072200, (C) Glyma.16G065700, (D) Glyma.12G210600, (E) Glyma.08G109300, (F) Glyma.12G232500, (G) Glyma.16G214900, and (H) Glyma.09G073600 at five seed stages, respectively.

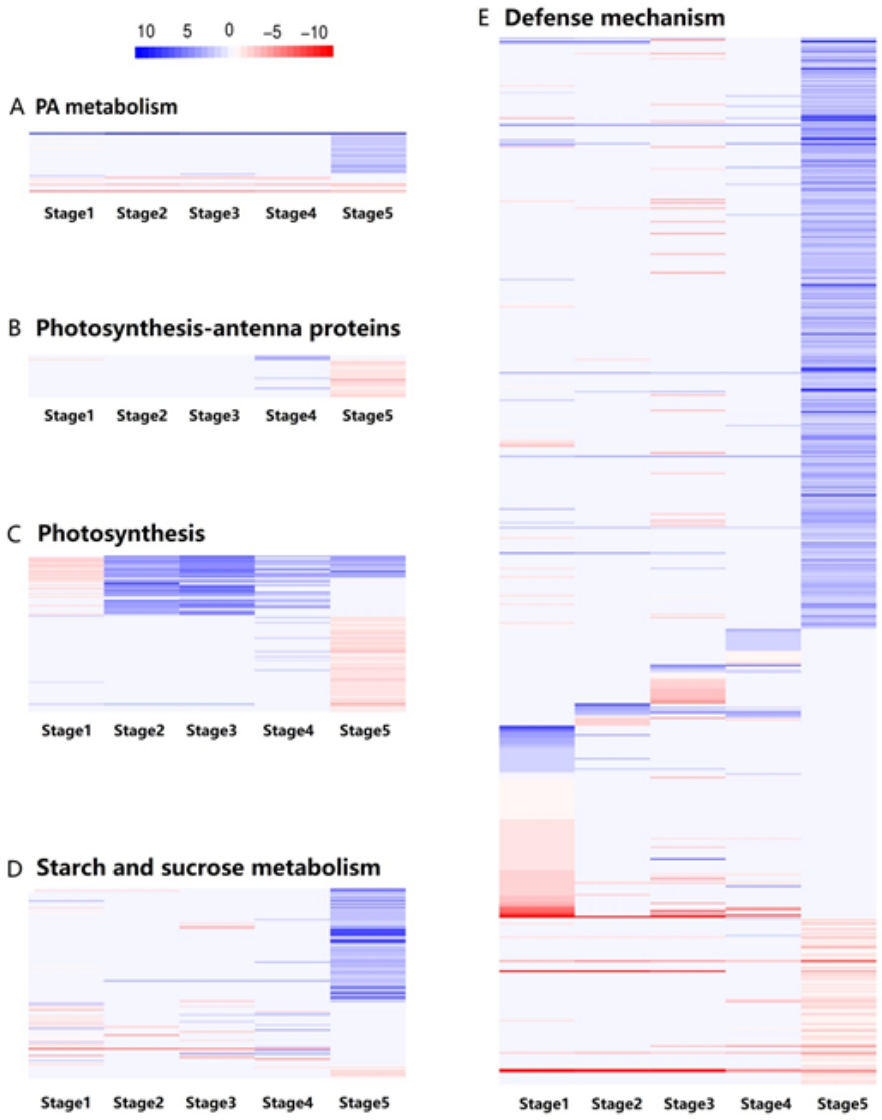


Figure 5

Fold change (2mlpa over 2MWT) ratios of DEGs at five seed developmental stages associated with (A) PA metabolism, (B) photosynthesis-antenna proteins, (C) photosynthesis, (D) starch and sucrose metabolism, and (E) defense mechanisms. Rows represent DEGs between the 2mlpa and 2MWT soybean lines, while columns represent the five seed developmental stages. Blue indicates higher fold change (2mlpa over 2MWT) ratios, and red indicates lower fold change (2mlpa over 2MWT) ratios (downregulated).

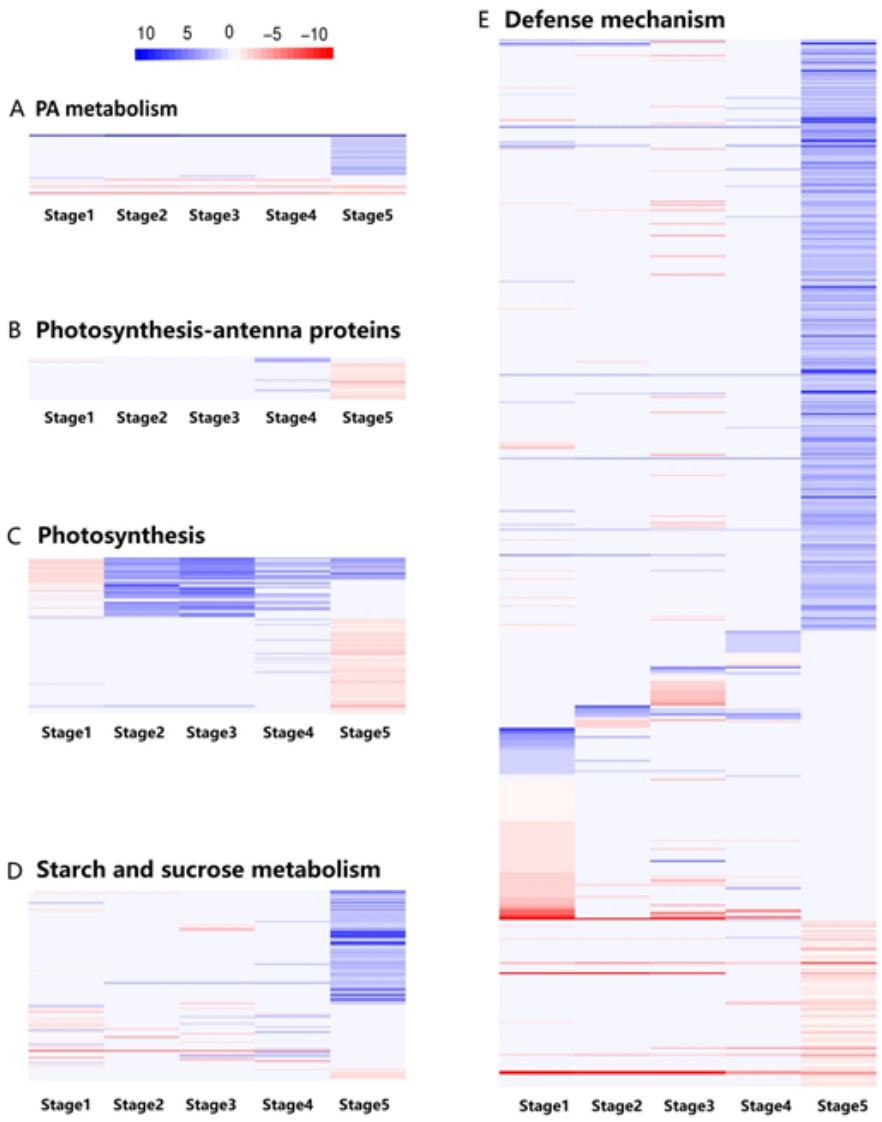


Figure 5

Fold change (2mlpa over 2MWT) ratios of DEGs at five seed developmental stages associated with (A) PA metabolism, (B) photosynthesis-antenna proteins, (C) photosynthesis, (D) starch and sucrose metabolism, and (E) defense mechanisms. Rows represent DEGs between the 2mlpa and 2MWT soybean lines, while columns represent the five seed developmental stages. Blue indicates higher fold change (2mlpa over 2MWT) ratios, and red indicates lower fold change (2mlpa over 2MWT) ratios (downregulated).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.pdf
- SupplementaryTable.pdf
- Additionalfile1.docx
- Additionalfile1.docx
- Additionalfile2.xlsx

- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile3.xlsx
- Additionalfile4.xlsx
- Additionalfile4.xlsx
- Additionalfile5.xlsx
- Additionalfile5.xlsx
- Additionalfile6.xlsx
- Additionalfile6.xlsx
- Additionalfile7.xlsx
- Additionalfile7.xlsx