Genetically Modified Soybean Lines Exhibit Less Transcriptomic Variation Compared to Natural Narieties

Yan Long
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute
Wentao Xu
China Agricultural University
Caiyue Liu
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute
Mei Dong
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute
Xinwu Pei
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute
Rui Chen
Tianjin Academy of Agricultural Sciences
Wujun Jin
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute
Weixiao Liu
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute
Liang Li (liliang@caas.cn)
CAAS BRI: Chinese Academy of Agricultural Sciences Biotechnology Research Institute

Research article

Keywords: Transcriptomic analysis, GM soybean lines, Non-GM soybean lines, Soybean seed tissues, Differentially expressed genes

Posted Date: November 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-112817/v1

Abstract

Background

Genetically modified (GM) soybeans provide a huge amount of food for human consumption and animal feed. However, the possibility of unexpected effects of transgenesis has increased food safety concerns. High-throughput sequencing profiling provides a powerful approach to directly evaluate unintended effects caused by foreign genes.

Results

In this study, we performed transcriptomic analyses to evaluate differentially expressed genes (DEGs) in individual soybean tissues, including cotyledon (C), germ (G), hypocotyl (H), and radicle (R), instead of using the whole seed, from four GM and three non-GM soybean lines. A total of 3,351 DEGs were identified among the three non-GM soybean lines. When the GM lines were compared with their non-GM parents, 1,836 to 4,551 DEGs were identified. Furthermore, Gene Ontology (GO) analysis of the DEGs showed more abundant categories of GO items (199) among non-GM lines than between GM lines and the non-GM natural varieties (166). Results of Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed that most KEGG pathways were the same for the two types of comparisons.

Conclusions

The study successfully employed RNA sequencing to assess the differences in gene expression among four tissues of seven soybean varieties, and the results suggest that transgenes do not induce massive transcriptomic alterations in transgenic soybeans compared with those that exist among natural varieties. This work thus provides important support for safety evaluation of genetically modified soybeans based on seed tissues.

Background

Genetically modified (GM) crops play essential roles in modern agricultural improvement. The first GM plants, including antibiotic-resistant petunia and tobacco, were created independently by three research groups in 1983 [1]. Since then, GM crop production has increased dramatically over the past 20 years, growing more than 100-fold from 1.7 million hectares to 180 million hectares globally [2]. In addition to increasing global yield by 22%, GM crops have also reduced pesticide usage by 37% and environmental impact (insecticide and herbicide use) by 18% [3]. Although the rapid development of GM crop production has resulted in significant economic benefits, there has been continued consumer concern that GM products may lead to unforeseen food and environmental safety issues. Remaining questions include the possibility that the insertion of exogenous DNA fragments into crop genomes might lead to the deletion, insertion or rearrangement of genes, thereby affecting biochemical pathways or resulting in the formation of new biological compounds (such as new allergens or toxins). In the risk assessment (RA) of genetically modified organisms (GMOs), molecular characterization studies play a crucial role in

identifying GM insertions and their target loci. Nucleic acid-level molecular characterization data on GM plants (GMPs) as part of the RA are currently obtained by Southern blot and polymerase chain reaction (PCR) analyses in combination with Sanger sequencing, to determine the precise location of the junction between the transgenic insert and the host genome and to detect the potential presence of backbone sequences from the transformation vector [4]. The development of next-generation sequencing (NGS) technology has introduced an additional or even a replacement tool for the molecular characterization of GMPs [5, 6]. In addition to molecular characterization, NGS facilitates transcriptome profiling [7] and can elucidate the potentially altered expression profiles of genes flanking the transgene insert. Recently, high-throughput sequencing combined with "-omics" profiling technologies, including transcriptomics, proteomics and metabolomics, has been widely used to analyze both target and unexpected effects [8–11].

Soybean (*Glycine max*) is an economically important crop and the area planted with transgenic soybeans now constitutes more than 50% of that covered by all transgenic crops. Different exogenous genes that control important agronomic traits, such as abiotic stress [12, 13], altered growth/yield [14], and insect resistance [15, 16], have been inserted into soybean. Most transgenic modifications in soybean are related to abiotic stress, such as tolerance to herbicides (e.g., MON87705, MON87708, and MON87701 × MON87705) (http://www.isaaa.org/gmapprovaldatabase/default.asp). Therefore, the safety of GM crops must be evaluated, and substantial equivalence is the cornerstone of these safety assessments. Through the development of new methodologies, indicators of substantial equivalence have become increasingly abundant. Innovative profiling techniques (such as genomics, transcriptomics, proteomics, and metabolomics) enable comprehensive measurements and comparisons of the transcripts, proteins and metabolites of organisms and provide detailed insights into any unintended effects.

In this study, we performed a transcriptomic sequencing to compare the gene-expression patterns of four tissues from seven different soybean lines (four GM lines, namely, MON87701 × MON89788, MON87708, MON87705 and FG72, and three non-transgenic soybean cultivars, namely, Zhonghuang13, A3525 and JACK), to expand the depth and breadth of our knowledge of GM and non-GM soybean gene-expression profiles. The four primary aims of the study were (1) to determine gene-expression patterns in different tissues in soybean; (2) to compare the obtained expression patterns of natural genotypic soybean lines; (3) to compare differentially expressed genes (DEGs) between GM soybeans and their donor parents; and (4) to determine whether differences in gene expression among natural genotypic soybean lines were greater than those caused by transgenic modifications.

Results

Statistics of RNA-seq transcript abundance in GM and non-GM plants

To identify the molecular events occurring in different tissues of GM and non-GM plants, 28 RNA-seq libraries were constructed using RNA from ten pooled RNA samples. After Illumina sequencing and the

removal of adaptors and bad-quality reads, approximately 4,216,226 to 14,932,616 reads were obtained for the 28 libraries (Table 1). Clean reads were then mapped to the soybean reference genome, with the mapped ratio ranging from 43.13% to 89.27%. Among the mapped reads, the frequency of unigenes ranged from 41.21% to 87.95% (Table 1).

Table 1
Summary of sequence information for the 28 DEG libraries.

Sample name	Total reads	Total mapped	Multiply mapped	Uniquely mapped	Non-spliced reads	Spliced reads
ZH13_G	4216226	2671012 (63.35%)	95558 (2.27%)	2575454 (61.08%)	2024284 (48.01%)	551170 (13.07%)
ZH13_C	9858652	8797768 (89.24%)	180882 (1.83%)	8616886 (87.4%)	5747294 (58.3%)	2869592 (29.11%)
ZH13_H	7294396	5569502 (76.35%)	133888 (1.84%)	5435614 (74.52%)	4398715 (60.3%)	1036899 (14.22%)
ZH13_R	9377594	7105662 (75.77%)	180228 (1.92%)	6925434 (73.85%)	5715708 (60.95%)	1209726 (12.9%)
A3525_G	6410296	2765020 (43.13%)	123130 (1.92%)	2641890 (41.21%)	2079551 (32.44%)	562339 (8.77%)
A3525_C	12225988	10422694 (85.25%)	168208 (1.38%)	10254486 (83.87%)	7224718 (59.09%)	3029768 (24.78%)
A3525_H	5415494	3291948 (60.79%)	101724 (1.88%)	3190224 (58.91%)	2506364 (46.28%)	683860 (12.63%)
A3525_R	7299716	4275516 (58.57%)	109630 (1.5%)	4165886 (57.07%)	3222708 (44.15%)	943178 (12.92%)
JACK_G	5257778	3299682 (62.76%)	129240 (2.46%)	3170442 (60.3%)	2467896 (46.94%)	702546 (13.36%)
JACK_C	10150190	9061060 (89.27%)	134416 (1.32%)	8926644 (87.95%)	6397850 (63.03%)	2528794 (24.91%)
JACK_H	6325954	3977316 (62.87%)	161790 (2.56%)	3815526 (60.32%)	2864239 (45.28%)	951287 (15.04%)
JACK_R	6545170	5176872 (79.09%)	139338 (2.13%)	5037534 (76.97%)	3852475 (58.86%)	1185059 (18.11%)
MON87705_G	5801422	3121920 (53.81%)	77746 (1.34%)	3044174 (52.47%)	2477649 (42.71%)	566525 (9.77%)

MON87705_C	10040618	8706092 (86.71%)	132424 (1.32%)	8573668 (85.39%)	6366618 (63.41%)	2207050 (21.98%)
MON87705_H	7964642	5135186 (64.47%)	114408 (1.44%)	5020778 (63.04%)	4078999 (51.21%)	941779 (11.82%)
MON87705_R	6089780	3419908 (56.16%)	84256 (1.38%)	3335652 (54.77%)	2778434 (45.62%)	557218 (9.15%)
MON87708_G	6247388	3615618 (57.87%)	108156 (1.73%)	3507462 (56.14%)	2646103 (42.36%)	861359 (13.79%)
MON87708_C	14932616	13280240 (88.93%)	187898 (1.26%)	13092342 (87.68%)	9160767 (61.35%)	3931575 (26.33%)
MON87708_H	6905856	4829608 (69.93%)	120672 (1.75%)	4708936 (68.19%)	3438106 (49.79%)	1270830 (18.4%)
MON87708_R	5449430	4080382 (74.88%)	96646 (1.77%)	3983736 (73.1%)	3029095 (55.59%)	954641 (17.52%)
M0188_G	5518118	2498370 (45.28%)	119158 (2.16%)	2379212 (43.12%)	1911268 (34.64%)	467944 (8.48%)
M0188_C	10115880	8611796 (85.13%)	134268 (1.33%)	8477528 (83.8%)	5871812 (58.05%)	2605716 (25.76%)
M0188_H	6016450	3436352 (57.12%)	129058 (2.15%)	3307294 (54.97%)	2589384 (43.04%)	717910 (11.93%)
M0188_R	6720942	3955694 (58.86%)	129572 (1.93%)	3826122 (56.93%)	2956415 (43.99%)	869707 (12.94%)
FG72_G	13193514	11067818 (83.89%)	243210 (1.84%)	10824608 (82.04%)	7460539 (56.55%)	3364069 (25.5%)
FG72_C	14086268	11984834 (85.08%)	165238 (1.17%)	11819596 (83.91%)	8591902 (60.99%)	3227694 (22.91%)
FG72_H	9768542	7866696 (80.53%)	184778 (1.89%)	7681918 (78.64%)	5334051 (54.6%)	2347867 (24.03%)
FG72_R	13992384	11333216 (81%)	200028 (1.43%)	11133188 (79.57%)	7652831 (54.69%)	3480357 (24.87%)

A PCA was performed on all 28 transcriptomic datasets to obtain a global view of gene expression across the seven soybean lines. The first two principal components (PCs) explained 46.07% (PC1) and 25.46% (PC2) of the total variance (Fig. 1). The two PCs could not separate the GM lines from their non-GM donor parents. For example, MON87705, MON87708 and A3525 were clustered together. The natural soybean lines showed significant genetic distance; for instance, Zhonghuang13 was positioned far from the other two natural lines, JACK, and A3525. Notably, the seven datasets from the cotyledon tissues clustered discretely from the other three tissues.

Analysis of differentially expressed genes in the non-GM lines

Gene-expression data were obtained using the gene-expression formula. Genes that were significantly differentially expressed between different parental lines were identified according to normalized gene-expression levels. Some differentially expressed unigenes were expressed at higher levels in certain lines, whereas others were expressed at lower levels. In total, 7,491 DEGs were detected in the two-group comparisons (non-GM lines/non-GM lines) (Fig. 2A). When using ZH13 was control, a total of 4,384 DEGs were identified between A3525 and ZH13, including 1,888 upregulated genes and 2,496 downregulated genes (Supplementary Table S1), and 6,458 DEGs were identified between JACK and ZH13, including 2,298 upregulated genes and 4,160 downregulated genes (Supplementary Table S2). The combined analysis highlighted 3,351 DEGs that were commonly differentially expressed among the three materials, whereas the remaining 3,107 and 1,033 genes were specifically differentially expressed in the specific material (Fig. 2B).

Analysis of differentially expressed genes in the non-GM and GM lines

The number of DEGs among the four group comparisons (non-GM lines/GM lines) ranged from 1,836 to 4,551, which represented 15.56% to 38.57% of the total number of genes (Fig. 2A). These comparisons revealed 4,551 DEGs between M0188 and A3525, including 1,888 upregulated genes and 2,496 downregulated genes (Supplementary Table S3); 2,518 DEGs were identified between M87705 and A3525, including 1,421 upregulated genes and 1,097 downregulated genes (Supplementary Table S4); 1,836 DEGs were identified between M87708 and A3525, including 1,212 upregulated genes and 624 downregulated genes (Supplementary Table S5); and 2,894 DEGs were present between FG72 and JACK, including 1,697 upregulated genes and 1,197 downregulated genes (Supplementary Table S6). The combined analysis revealed that 587 DEGs were commonly differentially expressed among the four tissues, and that 382, 1,016 and 2,570 genes were specifically differentially expressed in specific comparisons (Fig. 2C).

The numbers of DEGs in each of the two comparisons were then investigated. The total number of DEGs among the three non-GM lines was 7,491 and the total number of DEGs between the GM lines and their donor parents was 6,836. The DEGs were shared by different varieties. The differences in numbers of DEGs among the natural varieties were even larger than those between the GM lines and their donor parents, which is consistent with reports from several previous studies [17, 18].

Gene Ontology (GO) analysis of the identified DEGs

To classify the functions of the potential differentially expressed genes, GO analysis was performed. The majority of the DEGs in the comparison among non-GM lines were assigned to the category biological processes (96/199, 48.24%), followed by molecular functions (70/199, 35.18%) and cellular components (33/199, 16.58%). Within the category of biological processes, “cellular metabolic process” (916, 12.26%) and “macromolecule biosynthetic process” (667, 8.93%) were prominently represented, indicating that important metabolic activities occurred in the analyzed tissues (Fig. 3A, Supplementary Table S7). For the cellular component category, “cellular component” (342, 16.89%) and “intracellular” (177, 8.74%) were the two major classes. The other remaining categories of intracellular part, organelle, intracellular organelle, and cytoplasm, accounted for 27.95% of the DEGs. Within the categories of molecular functions, “nucleic-acid binding” (354, 12.96%) and “RNA binding” (212, 12.96%) were the largest and second-largest classes, respectively.

To compare non-GM lines with GM lines, the majority of the DEGs were assigned to the category biological processes (79/166, 47.59%), followed by molecular functions (52/166, 31.33%) and cellular components (35/166, 21.08%). Within the biological processes category, “biosynthetic process” (1,465, 14.26%) and “cellular biosynthetic process” (1,351, 13.15%) were prominently represented, indicating that important biosynthetic processes occur in the analyzed tissues (Fig. 3B, Supplementary Table S7). For the cellular component category, “intracellular” (1,262, 14.41%) and “intracellular part” (1,173, 13.4%) were the two major sub-categories. The other remaining categories: organelle, intracellular organelle, macromolecular complex, and cytoplasm, accounted for 23.04% of the DEGs. The molecular functions category included “structural molecule activity” (318, 9.39%) and “nucleic-acid binding” (316, 9.33%) as the first and second-largest sub-categories, respectively. The categories of “structural constituent of ribosome”, “RNA binding”, “oxidoreductase activity”, “helicase activity” and “endonuclease activity” accounted for approximately 26.26% of the DEGs.

The GO annotations of the two types of comparisons were subsequently compared. The number of GO-term categories (199) for the non-GM lines exceeded the number for the comparison between non-GM and GM lines (166) (Supplementary Table S7). A total of 81 categories appeared in both types of comparisons. The DEGs among the non-GM lines were functionally more varied than the DEGs between the GM lines and their donor parents. These common categories included different biological functions, such as “cellular biosynthetic process”, “nuclear transport”, and “lipid transport”, which are essential for the maintenance of cellular function. Among the three categories of GO terms in non-GM line

comparisons, 85 specific GO items were present, such as “regulation of DNA replication”, “regulation of DNA metabolic process”, “DNA replication origin binding”, and “sequence-specific DNA binding”. This enrichment of DNA-related terms suggests that differences in gene-expression resulting from genetic engineering are primarily associated with changes in DNA-level regulation.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs

The KEGG pathway database is a knowledge base for the systematic analysis of gene functions in terms of networks of genes and molecules in cells and their variants specific to particular organisms. To analyze further the DEGs between GM and non-GM plants, all the DEGs were analyzed with respect to the KEGG pathway database.

Among the 7,491 DEGs in the non-GM line comparisons, 126 (28.64%) that significantly matched to sequences in the database were assigned to eight KEGG pathways. Among these eight pathways, protein processing in endoplasmic reticulum contained the greatest number of DEGs, followed by the pathways ribosome and ribosome biogenesis in eukaryotes (Fig. 4, Supplementary Table S8). This pattern highlights that active metabolic processes occur in these tissues. Among the 6,836 DEGs between the non-GM and GM lines, the pathway protein processing in the endoplasmic reticulum contained the greatest number of DEGs, followed by ribosome and spliceosome pathways. Comparisons of the KEGG pathways revealed seven significant pathways, including photosynthesis, glutathione metabolism, arachidonic acid metabolism, ribosome biogenesis in eukaryotes, ribosome, RNA transport and protein processing in endoplasmic reticulum (Supplementary Table S8). Most of the DEGs between the two comparisons belonged to the same classes and did not differ greatly.

Discussion

In this study, four different tissues, cotyledon (C), germ (G), hypocotyl (H), and radicle (R) of four GMPs and three donor parents were selected to analyze the unintended effects. In previous studies, most studies that have analyzed unintended effects in GM and non-GM lines have performed sequencing and analysis on only one specific tissue or mixed tissue [18, 19]. For example, Liu et al.(2020) used the leave tissues of four GE rice lines expressing Bacillus thuringiensis (Bt) genes that developed by genetic engineering and seven rice lines developed by conventional cross-breeding to discover the unintended effects, and there were only tens of DEGs identified. While there were thousands of DEGs identified in the current study, which allows more reliable and robust analysis of potential unintended effects.

The PCA analyses of the raw datasets showed a distinct separation between samples with different genetic background, while there was no discrimination between GM and non-GM counterparts in transcriptomic levels. Several previous studies showed the same tendency, such as in wheat and barley [20, 21]. For example, a higher similarity between a GE variety and its non-GE near-isogenic line was

identified than between two common bean varieties in both leaf and grain proteomic profiles in Embrapa 5.1 common bean[8, 22]. So, the current finding together with previous results suggested that the intrinsic differences in genetic background bring much greater variation on plant transcriptome than by the introduction of foreign genes by genetic manipulation methods.

Besides the PCA analysis results, the other main result of the study was that significant differences existing among natural plant varieties, but the absence of significant differences between GM plants and their non-GM donor parents from differentially gene expression analysis. The differences in numbers of DEGs among the natural varieties were even larger than those between the GM lines and their donor parents. Besides the DEG numbers, the GO and KEGG analysis results showed that different types of DEGs exsiting because of gene operation or conventional breeding methods. For subgroup “biological process”, the biggest category for non-GM/non-GM line comparisons was “cellular metabolic process”, and the category “biosynthetic process” was the largest type for non-GM/GM comparasion. For “molecular function” subgroup, the biggest type of non-GM/non-GM comparisons was “cellular component”, and the biggest type for GM/non-GM comparisons was “structural molecular activity”; for “cellular component” subgroup, the biggest type of non-GM/non-GM comparisons was “organelle”, and the biggest type for GM/non-GM comparisons was “intracellular”. These results mean that although there was no significant difference of gene numbers for the two types of comparisons, while the gene types were not same because of the different gene introgression method. While all of the above results were only based on the transcriptomic data, the integrated application of multi-omics approaches, such as proteomics, metabonomics, could be used to evaluate the changes in the plants and their biological relevance.

Conclusion

In this study, four different tissues from seven soybean lines, including four GM lines and their three donor parents, were analyzed for DEGs. After gene expression values were calculated, two types of comparisons were performed: either among the non-GM lines or between GM lines and the non-GM natural varieties. In total, 7,941 DEGs were identified among the non-GM lines, and 8,461 DEGs in the comparison between the GM and non-GM lines. The GO and KEGG analyses showed that the categories and biological functions of DEGs among the natural varieties were more varied than those in the GM/non-GM line comparison. We conclude that intrinsic differences in genetic background result in considerably more variation in the plant transcriptome than the introduction of exogenous genes by genetic manipulation methods.

Methods

Plant materials

A total of seven soybean lines, including three lines obtained by conventional breeding and four lines developed by genetic engineering (GE) that express *EPSPS* genes, were used in this study (Fig. 5). The

transgenic lines MON87725, MON87708 and MON87701×MON89788 (M0188), which express *cp4 EPSPS* genes, are derived from the donor parent A3525. The other transgenic line FG72, which expresses the *2m EPSPS* gene, is derived from the donor parent JACK. The third non-GM line was Zhonghuang13, which is a widely planted variety in China. All of the materials were harvested and stored at −80℃.

Sample collection, RNA isolation and RNA-seq library preparation

Total RNA from four different tissues, including cotyledon (C), germ (G), hypocotyl (H), and radicle (R), was extracted from the soybean plants with TRIzol Reagent (Invitrogen, 15596-026) according to the manufacturer's instructions. Each RNA-seq library was constructed from the pooled RNA from 10 samples. A total of 3 μg of RNA per sample was used as input material for the RNA sample preparations. The RNA-seq library was generated using NEBNext Ultra RNA Library Prep Kits for Illumina (NEB, USA). Following the instructions provided by Illumina, mRNA was purified from the pooled total RNA using polyT oligo-attached magnetic beads (Novogene, China). Fragmentation buffer was added to disrupt the mRNA into short fragments. Reverse transcriptase and random primers were used to synthesize the first-strand cDNA from the cleaved mRNA fragments. Second-strand cDNA was synthesized using buffer, dNTPs, RNase H, and DNA polymerase I. The double-stranded cDNA was purified using the QIAquick PCR Extraction Kit (QIAGEN, Hilden, Germany) and washed with EB buffer for end repair and single nucleotide A (adenine) addition. Finally, sequencing adaptors were ligated to the fragments. The resulting fragments were purified by AMPure XP beads and enriched by PCR to construct a library for deep sequencing. The sequence data was submitted to NCBI database and the number code is PRJNA668923.

RNA-seq library sequencing and mapping

The transcriptome library was sequenced using the Illumina HiSeq 2000 system. After the raw data were generated and the data-processing steps were completed, the clean reads were then mapped to the soybean reference sequences using RSEM software [23]. Mismatches of no more than two bases were allowed in the alignments. The read count for each gene was obtained from the mapping results. The normalized data were fed to SIMCA 14.1 software (Umetrics, Umea, Sweden) for principal component analysis (PCA).

Differentially expressed gene identification

Gene-expression levels were calculated based on the numbers of reads mapped to the reference sequence using the FPKM [24] method. The differentially expressed genes (DEGs) were then screened by comparing gene-expression levels. Differential expression analysis of each comparison was performed using the DESeq R package (1.10.1) using the method described by Anders [25]. DESeq provides statistical routines for determining differential expression in digital gene-expression data using a model

based on the negative binomial distribution. The resulting *P* values were adjusted using Benjamini and Hochberg's approach to control the false discovery rate. In this study, unigenes with an adjusted $P < 0.05$ identified by DESeq were considered to be differentially expressed.

Functional annotation of DEGs

For functional annotation, the assembled unigenes that putatively encode proteins were searched against the nr (http://www.ncbi.nlm.nih.gov/), SWISS-PROT (http://www.expasy.ch/sprot/), KEGG (http://www.genome.jp/kegg/) and COG (http://www.ncbi.nlm.nih.gov/cog/) databases using the BLASTX algorithm. A typical cut-off value of E-value<1e−5 was used. The Blast2GO program [26] was used to assign GO annotations to the genes with nr annotations, according to the component function, biological process and cellular component ontologies. After obtaining GO annotations for all genes, WEGO software [27] was used to assign GO functional classifications to all the genes and to understand the distribution of gene functions for the species on the macro level.

Abbreviations

GM genetic modified

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

DEG differentially expressed gene

Declarations

Acknowledgments

The authors appreciate Dr. Qingyu Wu (Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences) for suggestions and revisions of the paper.

Author contributions

M. Dong and C. Liu performed the experiments. Y. Long, W. Xu, W. Liu, and R. Chen performed the data analysis and prepared the figures. Y. Long and L. Li prepared the manuscript, X. Pei, R. Chen and W. Jin revised the manuscript. All authors read and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was financially supported by The National Major Special Project for the Development of Transgenic Organisms (2016ZX08012-003, 2016ZX08011-001) and Central Public-interest Scientific Institution Basal Research Fund (1610392020001).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Raman R: The impact of Genetically Modified (GM) crops in modern agriculture: A review. GM crops & food 2017, 8(4):195-208.
2. ISAA: Biotech Crops Continue to Help Meet the Challenges of Increased Population and Climate Change. 2018.
3. Klumper W, Qaim M: A meta-analysis of the impacts of genetically modified crops. PloS one 2014, 9(11):e111629.
4. Kok EJ, Pedersen J, Onori R, Sowa S, Schauzu M, De Schrijver A, Teeri TH: Plants with stacked genetically modified events: to assess or not to assess? Trends in Biotechnology 2014, 32(2):70-73.
5. Kovalic D, Garnaat C, Guo L, Yan YP, Groat J, Silvanovich A, Ralston L, Huang MY, Tian Q, Christian A et al: The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology. Plant Genome-Us 2012, 5(3):149-163.
6. Yang L, Wang C, Holst-Jensen A, Morisset D, Lin Y, Zhang D: Characterization of GM events by insert knowledge adapted re-sequencing approaches. Scientific reports 2013, 3:2839.
7. Chu Y, Corey DR: RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic acid therapeutics 2012, 22(4):271-274.
8. Balsamo GM, Valentim-Neto PA, Mello CS, Arisi AC: Comparative Proteomic Analysis of Two Varieties of Genetically Modified (GM) Embrapa 5.1 Common Bean (Phaseolus vulgaris L.) and Their Non-GM Counterparts. J Agric Food Chem 2015, 63(48):10569-10577.

9. Oms-Oliu G, Odriozola-Serrano I, Martin-Belloso O: Metabolomics for assessing safety and quality of plant-derived food. Food Res Int 2013, 54(1):1172-1183.
10. Vilperte V, Agapito-Tenfen SZ, Wikmark OG, Nodari RO: Levels of DNA methylation and transcript accumulation in leaves of transgenic maize varieties. Environ Sci Eur 2016, 28.
11. Haynes E, Jimenez E, Pardo MA, Helyar SJ: The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. Food Control 2019, 101:134-143.
12. Wei W, Liang DW, Bian XH, Shen M, Xiao JH, Zhang WK, Ma B, Lin Q, Lv J, Chen X et al: GmWRKY54 improves drought tolerance through activating genes in abscisic acid and Ca(2+) signaling pathways in transgenic soybean. Plant J 2019, 100(2):384-398.
13. Wang D, Liu YX, Yu Q, Zhao SP, Zhao JY, Ru JN, Cao XY, Fang ZW: Functional Analysis of the Soybean GmCDPK3 Gene Responding to Drought and Salt Stresses. 2019, 20(23).
14. Cheng H, Jin HX, Gai JY, Yu DY: [Transgenic technology and soybean quality improvement]. Yi chuan = Hereditas 2011, 33(5):431-436.
15. Lin J, Mazarei M, Zhao N, Hatcher CN, Wuddineh WA, Rudis M, Tschaplinski TJ, Pantalone VR, Arelli PR, Hewezi T et al: Transgenic soybean overexpressing GmSAMT1 exhibits resistance to multiple-HG types of soybean cyst nematode Heterodera glycines. Plant Biotechnol J 2016, 14(11):2100-2109.
16. Yang X, Niu L, Zhang W, He H, Yang J, Xing G, Guo D, Zhao Q, Zhong X, Li H et al: Increased multiple virus resistance in transgenic soybean overexpressing the double-strand RNA-specific ribonuclease gene PAC1. Transgenic research 2019, 28(1):129-140.
17. Liu WX, Xu WT, Li L, Dong M, Wan YS, He XY, Huang KL, Jin WJ: iTRAQ-based quantitative tissue proteomic analysis of differentially expressed proteins (DEPs) in non-transgenic and transgenic soybean seeds. Scientific reports 2018, 8.
18. Liu Q, Yang X, Tzin V, Peng Y, Romeis J, Li Y: Plant breeding involving genetic engineering does not result in unacceptable unintended effects in rice relative to conventional cross-breeding. 2020.
19. Fu W, Wang C, Xu W, Zhu P, Lu Y, Wei S, Wu X, Wu Y, Zhao Y, Zhu S: Unintended effects of transgenic rice revealed by transcriptome and metabolism. GM crops & food 2019, 10(1):20-34.
20. Ioset JR, Urbaniak B, Ndjoko-Ioset K, Wirth J, Martin F, Gruissem W, Hostettmann K, Sautter C: Flavonoid profiling among wild type and related GM wheat varieties. Plant molecular biology 2007, 65(5):645-654.
21. Kogel KH, Voll LM, Schafer P, Jansen C, Wu YC, Langen G, Imani J, Hofmann J, Schmiedl A, Sonnewald S et al: Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. Proceedings of the National Academy of Sciences of the United States of America 2010, 107(14):6198-6203.
22. Valentim-Neto PA, Rossi GB, Anacleto KB, de Mello CS, Balsamo GM, Arisi ACM: Leaf proteome comparison of two GM common bean varieties and their non-GM counterparts by principal component analysis. J Sci Food Agr 2016, 96(3):927-932.
23. Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. Bmc Bioinformatics 2011, 12.

24. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010, 28(5):511-515.
25. Anders S, Huber W: Differential expression analysis for sequence count data. Genome biology 2010, 11(10):R106.
26. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005, 21(18):3674-3676.
27. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L et al: WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 2006, 34(Web Server issue):W293-297.

Figures

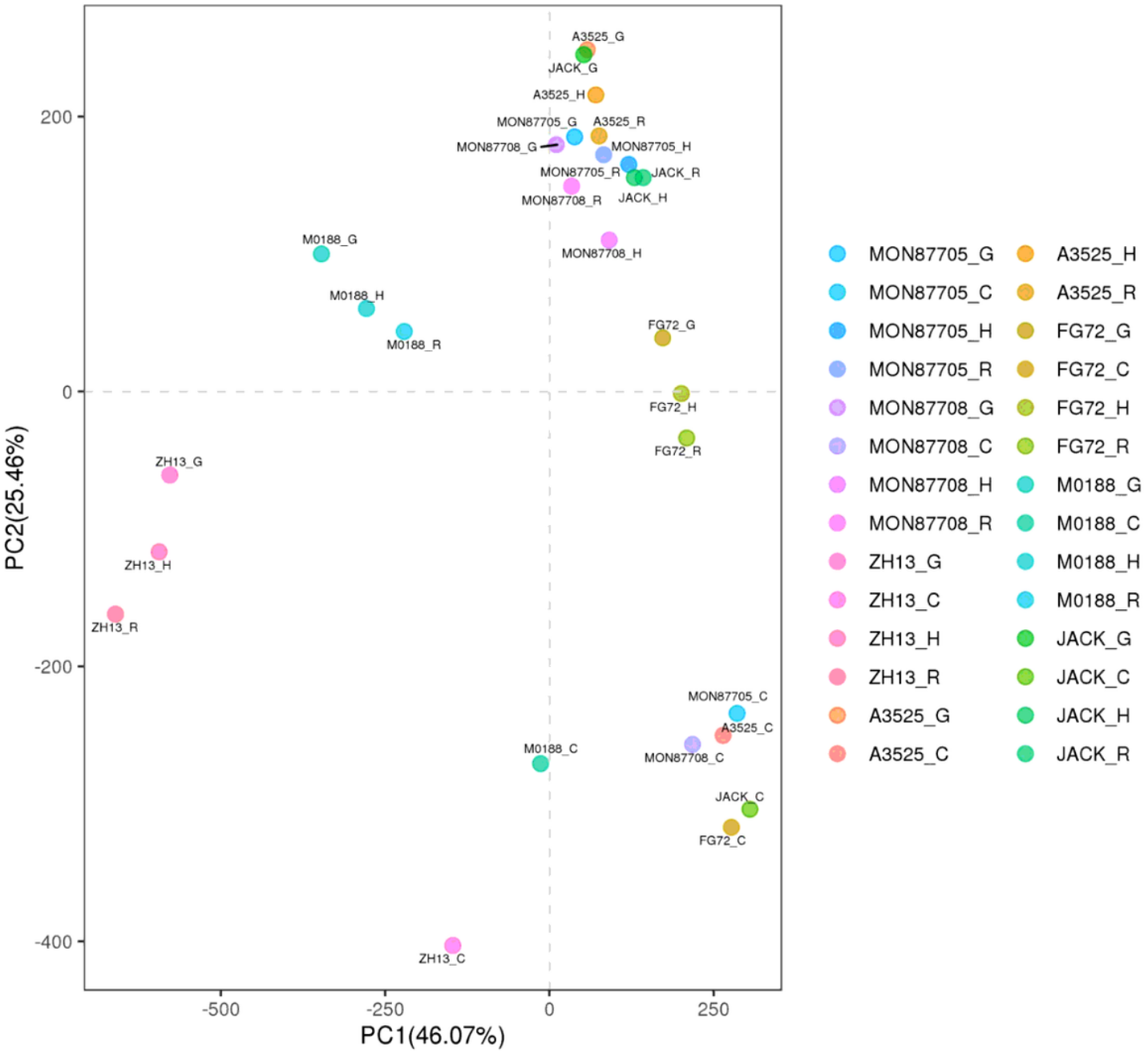

Figure 1

Principal component analyses (PCA) of gene-expression levels in the four tissues of seven soybean lines.

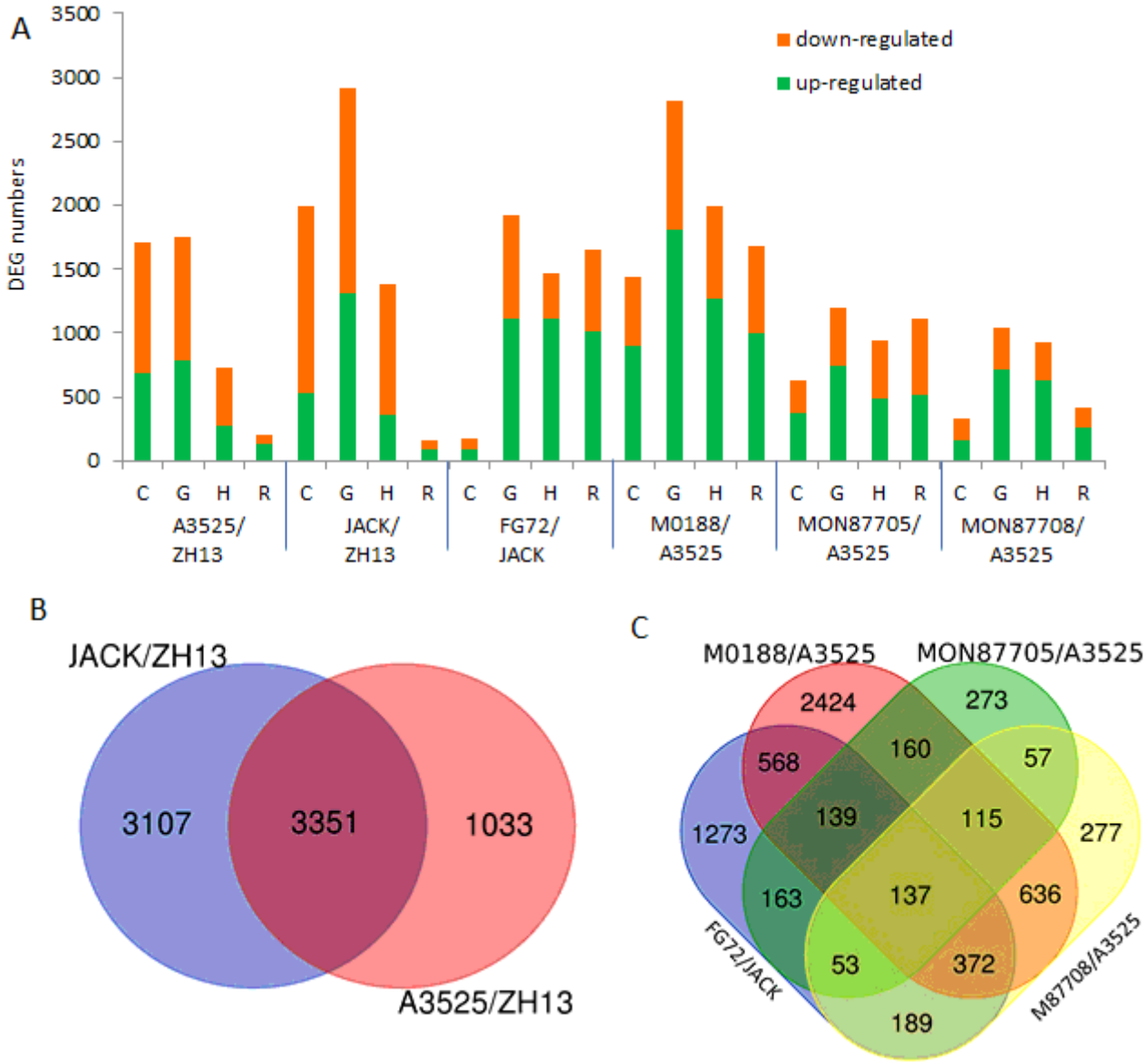

Figure 2

Differentially expressed genes for the two types of comparisons among GM and non-GM lines. (A) Numbers of DEGs for the six pairs of comparisons. For each comparison, the green bar represents the upregulated genes, and the orange bar represents the downregulated genes. C: cotyledon, G: germ, H: hypocotyl, R: radicle. (B) Venn diagram for the three non-GM line comparisons. (C) Venn diagram for the four GM lines compared with their donor parents.

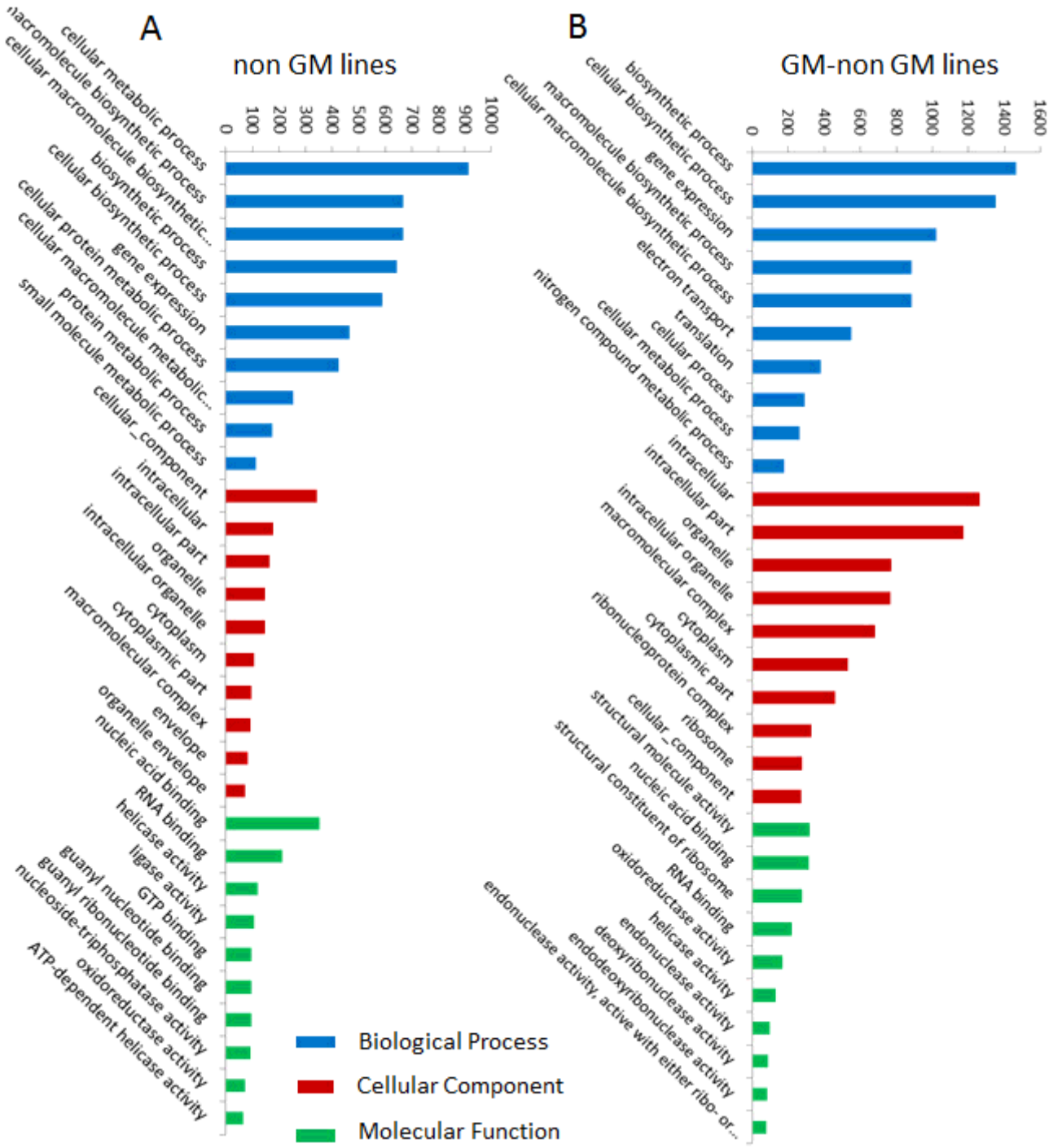

Figure 3

GO analysis results for DEGs in the two types of comparisons among GM and non-GM lines. (A) GO analysis tree for the three non-GM line comparisons. (C) GO analysis for the four GM lines compared with their donor parents.

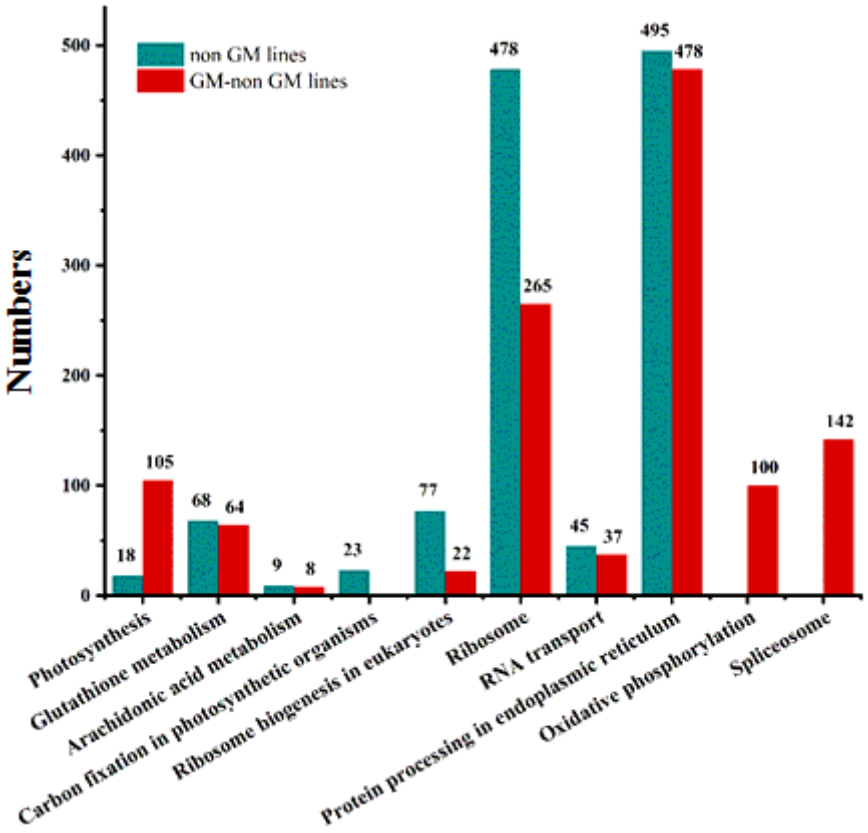

Figure 4

KEGG analysis results for DEGs in the two types of comparisons among the GM and non-GM lines. (A) KEGG pathway analysis of the three non-GM line comparisons. (C) KEGG pathway analysis of the four GM lines compared with their donor parents.

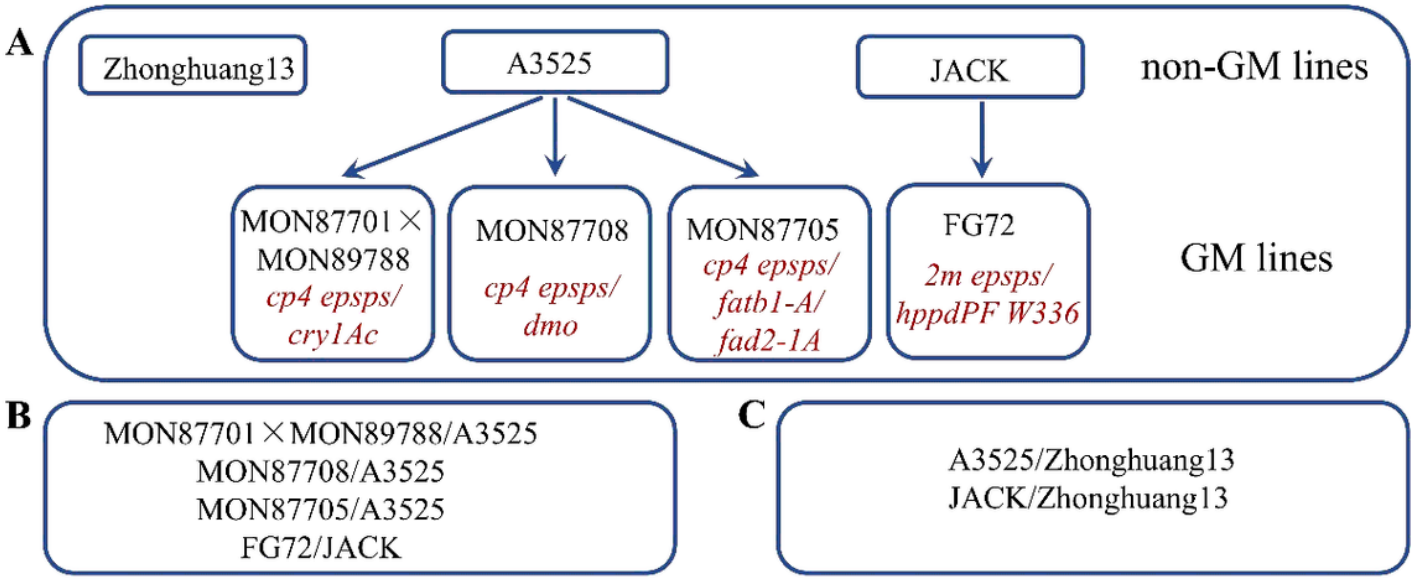

Figure 5

Genetic relationships among the studied soybean lines and design of the grouping comparisons. (a) Genetic relationships among the studied soybean lines. (b) Experimental design of pairwise comparisons of gene expression between non-GM lines and GM lines. (C) Comparisons among non-GM parental soybean lines.

Supplementary Files

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

- SupplementaryTablesS1S8.xlsx