



Comparative transcriptome profiling of different tissues from beta-carotene-enhanced transgenic soybean and its non-transgenic counterpart

Yang Qin¹ · Hee-Jong Woo¹ · Kong-Sik Shin¹ · Myung-Ho Lim¹ · Seong-Kon Lee¹

Received: 7 August 2019 / Accepted: 1 November 2019 / Published online: 9 November 2019
© The Author(s) 2019

Abstract

Beta-carotene-enhanced transgenic soybeans, harboring genes encoding phytoene synthase and carotene desaturase under the control of a seed-specific promoter, were developed to alleviate vitamin A deficiency in populations, the diet of which was deficient in this vitamin. However, metabolic engineering of carotenoid biosynthetic pathways often has unintended effects, leading to major metabolic changes in plants that harbor endogenous beta-carotene biosynthesis pathways. In the present study, we performed transcriptome profiling analysis using RNA-seq to investigate the changes in the transcriptome and some unintended pleiotropic effects on the leaves, stems, roots, and seeds of beta-carotene-enhanced transgenic soybean lines, and compared them to those of their non-transgenic counterpart donor variety Kwangan. We observed that transgenic soybeans showed significant changes in secondary metabolic biosynthesis in leaves and down-regulated galactose metabolism in roots. Differentially expressed genes in the transgenic group, which were significantly up-regulated, included those encoding glycine-aspartic acid-serine-leucine-motif esterase/lipase, known as cutin synthase and cutinase. These results suggested enhanced beta-carotene biosynthesis may affect related enzymes to carbohydrate metabolism and fatty acid metabolism. Hence, we speculated that upregulation of cutin polymerization resulted in thickened seed coat and delayed seed germination of transgenic soybeans. Furthermore, downregulation of raffinose family oligosaccharide biosynthesis may cause redundancy of myo-inositol, a substrate of phytin formation. This could lead to phytic globoids accumulation in transgenic soybean seeds. The present information would be important for transgenic plant development via carotenoid metabolic engineering, with focus on beta-carotene over-production.

Key message

Gene expression changes related to secondary metabolite biosynthesis, cutin metabolism, and raffinose-family oligosaccharide biosynthesis in leaf, stem, and root, and phytin globoid accumulation in seeds, occurred in beta-carotene-enhanced transgenic soybeans.

Keywords Soybean · Beta-carotene-enhanced · Transgenic · Transcriptome profiles

Communicated by Sergio J. Ochatt.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11240-019-01731-2>) contains supplementary material, which is available to authorized users.

✉ Seong-Kon Lee
goryeong@korea.kr
Myung-Ho Lim
mlim312@korea.kr

¹ Biosafety Division, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju 54874, Republic of Korea

Introduction

Carotenoids are antioxidants particularly important for human health, but cannot be synthesized by humans and other animals (Ye et al. 2000; Beyer et al. 2010). Therefore, humans and animals depend on plant-derived products to meet their carotenoid requirements for physiological functions. Beta-carotene, a carotenoid, is a precursor of vitamin A, an essential nutrient for humans. Its deficiency leads to severe clinical symptoms associated with night blindness and increases the susceptibility of pre-school children to

infectious diseases (Paine et al. 2005; Gayen et al. 2016). Considerable efforts have been made toward metabolic engineering of the plant carotenoid biosynthetic pathway of staple crops such as rice, tomato, potato, maize, and soybean to overcome vitamin A deficiency via food intake (Ye et al. 2000; Paine et al. 2005; Römers et al. 2000; Diretto et al. 2007; Zhu et al. 2008; Schmidt et al. 2014). However, crops created via genetic manipulation of single or multiple genes to construct carotenoid biosynthesis pathways may have unintended effects such as competition for particular precursors, which leads to significant changes in major metabolic pathways or closely related pathways (Sandmann 2001). Fray et al. (1995) studied transgenic tomato with constitutive phytoene synthase expression and observed dwarfism, chlorosis, and decrease in gibberellin and abscisic acid synthesis. Seed-specific overexpression of phytoene synthase in *Arabidopsis*, reported by Lindgren et al. (2003), resulted in delayed seed germination, increased levels of chlorophyll and abscisic acid, and increased production of carotenoids such as beta-carotene, lutein, and violaxanthin. Transgenic soybean overexpressing a seed-specific bacterial phytoene synthase showed changes in seed oil composition and protein content with elevated beta-carotene levels, although significant differences that can account for the altered phenotypes were not observed using proteomic and metabolomic analysis (Schmidt et al. 2014). High carotenoid-containing transgenic golden rice, obtained by introducing genes encoding phytoene synthase and phytoene desaturase to manipulate endosperm-specific carotenoid pathways, was subjected to proteomics and metabolite analyses (Gayen et al. 2016). Results revealed significant up-regulation of carbohydrate metabolism-related enzymes such as pullulanase, UDP-glucose pyrophosphorylase, and glucose-1-phosphate adenylyltransferase in rice seeds and enhanced pyruvate biosynthesis, but decreased levels of galactonic acid and gentiobiose. As mentioned in previous studies, accumulation of beta-carotenes in seeds often leads to changes in oil composition, cytoplasm structure, and protein and metabolite content. From another point of view, safety assessment of transgenic crops generally focuses on molecular characterization, evaluation of nutritional value, possible toxicity, and allergy-triggering components, and analysis of metabolites directly related to the introduction of the T-DNA (OECD2012). Therefore, a combination of transcriptome, proteome, and metabolome analyses will contribute to our understanding of the pleiotropic effects and potential unintended consequences caused by genetic manipulation of carotenoid biosynthesis in transgenic crops.

Soybean (*Glycine max*) is a staple crop and an important source of oil, protein, and secondary metabolites (Kim et al. 2012). In previous studies, beta-carotene-enhanced soybean lines were created by insertions of a herbicide resistant *bar* gene as a selection marker under the control

of the constitutive nopaline synthase (*nos*) promoter, and a T-DNA bicistronic system harboring two biosynthetic genes encoding the *Capsicum* phytoene synthase (*psy*) and *Pantoea* carotene desaturase (*crtI*) under the control of the seed-specific β -conglycinin promoter, into the conventional Korean soybean Kwangan variety (Kim et al. 2012). After self-pollination and reproduction, three transgenic lines were selected and two of them (7-1-1 and 10-19-1) showed agronomic traits similar to their donor variety Kwangan (Qin et al. 2015). Molecular characterization and next-generation sequencing (NGS) analysis had indicated the intergenic insertion of only one copy of the T-DNA cassette into transgenic lines 7-1-1 and 10-19-1, but six copies in the transgenic line 9-1-2. The T-DNA was inserted between 10,873,131–10,872,988 bp on chromosome 14 and between 23,918,809–23,918,840 bp on chromosome 13 of transgenic lines 7-1-1 and 10-19-1, respectively (Qin et al. 2017b). In addition, three transgenic lines 7-1-1, 10-19-1, and 9-1-2 were confirmed to respectively produce 184.71 μ g, 170.47 μ g, and 213.58 μ g of beta-carotene in seeds, whereas commercial conventional soybean seeds generally contain 0.17–0.46 μ g beta-carotene. On the other hand, nutrient analysis of beta-carotene-enhanced transgenic soybean seeds had revealed that most nutrient levels remained similar to those of the non-transgenic donor variety Kwangan; however, the levels of certain nutrients such as crude fat, carbohydrate, oleic acid, calcium, iron, and potassium differed significantly (Qin et al. 2017a). Furthermore, it is noteworthy that we observed considerably different liquid phase separation of the mixture containing TRIzol, chloroform, seed powder of the beta-carotene-enhanced soybean lines from that of their donor Kwangan during RNA isolation from seeds. After centrifugation, we had observed that the RNA-containing supernatant of the transgenic soybeans was located in the intermediate layer, with a white grease layer on the top. We speculated that the cell morphology of the transgenic soybean seeds will differ from those of the non-transgenic donor Kwangan, in addition to the enrichment of beta-carotene content. Thus, in the present study, transmission electron microscopy (TEM) was performed to compare the inner cell structure such as protein body, oil body, and protein storage vacuole between the transgenic and non-transgenic seeds. Furthermore, we observed that transgenic soybean seeds showed some extent of delayed germination. The nutritional differences and phenotypic alternations of the seed could be associated with *Agrobacterium*-mediated transformation, transgene insertion, and somaclonal variations induced by in vitro tissue culture. Similarly, soybean possesses a native carotenoid biosynthesis pathway. Beta-carotene-enhanced transgenic soybeans were developed by inserting *crtI* in the place of genes encoding four endogenous enzymes required for beta-carotene biosynthesis, including phytoene desaturase (*PDS*), zeta-carotene isomerase (*ZISO*), zeta-carotene

desaturase (*ZDS*), and *crtI-b* lycopene beta-cyclase (*LCY-b*) (KEGG; Ha et al. 2010; Kim et al. 2012; Reviewed by Giuliano 2014). Thus, unintended effects could be triggered due to competition for common substrates such as geranyl–geranyl pyrophosphate (GGPP) and phytoene, and over-production of beta-carotene via secondary metabolism. Therefore, genome-wide comparative transcriptome analysis between beta-carotene-enhanced transgenic soybeans and their non-transgenic donor counterpart Kwangan variety was performed on four soybean tissues, namely, leaves, stems, roots, and seeds, to better understand the cause of the differences in nutrient and seed morphology and other potential unintended consequences.

Materials and methods

Plant material preparation and RNA sample collection

Soybean seeds (T_6 generations) of three beta-carotene-enhanced transgenic soybean lines 7-1-1, 10-19-1, and 9-1-2, and their non-transgenic donor variety Kwangan, were harvested from the experimental field of genetically modified organisms (GMO) of the National Institute of Agricultural Science (NAS), Rural Development Administration (RDA), in October 2016. The harvested seeds were naturally dried in the greenhouse for one month. For transmission electron microscopy (TEM) analysis, we selected soybean seeds of uniform size and smooth and intact surface from three transgenic lines and their non-transgenic donor Kwangan to observe the internal morphology of the seeds. For RNA-sequencing and real-time polymerase chain reaction (PCR) analysis, we sowed 10 seeds of two one-copy T-DNA-inserted transgenic soybean lines 7-1-1 and 10-19-1, and their non-transgenic donor Kwangan, in pots in the GMO greenhouse of NAS, RDA, in November 2016. At the first trifoliolate, third trifoliolate, and fifth trifoliolate of the vegetative stages, we collected the leaf, stem, and root samples from the transgenic and non-transgenic lines. Seed samples were collected from yellow pods of soybean lines at the reproductive stage. Tissue collections were performed on different plants and each tissue sample was collected from only one plant. The collected samples were flash-frozen in liquid nitrogen and stored at $-80\text{ }^{\circ}\text{C}$ until RNA extraction.

RNA extraction

Each soybean sample of leaves, stems, roots, and seeds was ground to a fine powder with a mortar and pestle, and then transferred into a nucleotide-free 2.0 mL spin tube. Total RNA was extracted from the leaves and stems using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) following

the manufacturer's protocols. For extracting total RNA from the roots and seeds, pretreatment with phenol–chloroform–isoamyl alcohol (pH 5.6) and extraction buffer (50 mM Tri-Cl (pH 9.0), 1% SDS) was performed to remove the lipids and proteins maximally, followed by the TRIzol reagent protocol. The extracted RNA samples were treated with the TURBO DNA-Free DNase kit (Ambion, Austin, TX, USA) to remove DNA, and RNA purification was performed using one volume of chloroform. Next, 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of 100% ethanol were used for RNA precipitation. The total RNA concentration was measured at 260 and 280 nm using a Nanodrop ND 1000-spectrophotometer (Thermo Fisher Scientific, Wilmington, NC, USA), and samples with wavelength ratios within the range of 1.9–2.1 were used. Aliquots of RNAs (20 μL) were stored at $-80\text{ }^{\circ}\text{C}$ for Illumina sequencing and real-time PCR analysis.

Illumina sequencing and RNA sequencing analysis

The prepared total RNA samples from four different tissues of the transgenic soybean lines 7-1-1 and 10-19-1, and their non-transgenic donor Kwangan, were used for cDNA library construction using the TruSeq RNA sample preparation kit (Illumina, San Diego, CA, USA) following the manufacturer's recommended protocols. cDNA fragments ranging from 400 to 500 bp were selected using agarose gel purification to generate libraries for cluster generation. Next, each sample was sequenced on a HiSeq 2500 system (Illumina, San Diego, CA, USA). Quality control analysis of the resulting fastq sequencing files was performed using FastQC (v0.11.2) (Babraham Bioinformatics, Cambridgeshire, UK), followed by adapter trimming and quality trimming using Trimmomatic (v0.32; Bolger et al. 2014). The read production and average coverages of the four soybean tissues are listed on Online Resource Table 1.

RNA sequencing was performed using the CLC Genomic Workbench v9.0 (Qiagen, USA). To create soybean reference files, four track files of 'Genome', 'CDS', 'mRNA' and 'Gene' were produced using genome mapping (Glycine_max_v2.0_genomic.fna.gz) and transcript annotation references (Glycine_max_v2.0_genomic.gff.gz) from the NCBI database. RNA sequencing analysis was performed for each soybean sample using the function of 'genome annotated with genes and transcripts', with the following parameters: mismatch/insertion/deletion cost = 2/3/3, length fraction = 0.8, similarity fraction = 0.8, strand specific = both, and maximum number of hits for a read = 10. The sequence reads assigned to the soybean reference genome for four tissues are shown in Online Resource Table 2.

Differential expression analysis and gene set enrichment (GSE) analysis

Using the tool ‘transcriptome analysis’ of the CLC genomic workbench, differential gene expression analysis of soybean transcriptomes was performed on four soybean tissues each of transgenic soybeans 7-1-1 and 10-19-1, which were then compared to those of their non-transgenic donor Kwangan. Similarly, using a function of ‘Set up experiment’, all samples were divided into the transgenic group and non-transgenic group. The transgenic group consisted of four tissues each of the transgenic soybean lines 7-1-1 and 10-19-1, whereas the non-transgenic group consisted of four tissues from Kwangan. The differentially expressed genes (DEGs) were identified by comparing the transgenic group with the non-transgenic group. The outputs of DEGs were filtered with an absolute value of \log_2 fold-change > 2 , and false discovery rate (FDR) P-value < 0.005 . After filtering, functional enrichment analysis, including gene enrichment analysis, kyoto encyclopedia of genes and genomes (KEGG) pathway analysis, and protein functional annotations, were performed to identify the DEGs significantly enriched in gene ontology (GO) terms and metabolic pathways at a Bonferroni corrected P-value ≤ 0.05 , compared to the whole transcriptome background. GO functional enrichment and KEGG pathway analysis were performed using a gene functional annotation tool (<https://david.ncicfcrf.gov/tools.jsp>).

Real-time PCR

Total RNA (5 μg) from each sample was used to synthesize cDNA using the RevertAid first strand cDNA synthesis kit (Thermo Fisher Scientific) following the manufacturer’s protocol. Gene expression was analyzed using the diluted cDNA, a Stratagene MX3005P qPCR system (Agilent, Santa Clara, CA, USA), and amfiSure qGreen qPCR master mix (GenDEPOT, Barker, TX, USA). The comparative $\Delta\Delta\text{CT}$ method was used for relative quantitation of genes (Schmittgen et al. 2008). To evaluate the transcript levels of the transgenes and the DEGs, low-copy soybean housekeeping genes encoding lectin 1 (*Le1*, K00821.1) for soybean seeds, and elongation factor 1-alpha (*Ef-1 α* , Glyma.19g052400) for soybean leaves, stems, and roots, and the high-copy housekeeping gene encoding the 60S ribosomal protein L24 (*60S*, Glyma17g05270) for all four tissues were separately used as internal controls for normalizing the real-time PCR data. Three duplicates were performed for analyzing the expression of each target gene. The primer sequences of the transgenes and DEGs for RNA verifications are shown in Online Resource Table 3.

TEM

TEM material preparations were performed as described by Schmidt and Herman (2008) as follows. The central part of the soybean seed was cut into approximately 1 mm³ cubes and fixed in 4% (v/v) formaldehyde and 2% (v/v) glutaraldehyde in 0.05 M phosphate buffer (pH 7.4). Next, the sample was post-fixed with 2% (v/v) osmium tetroxide in acetone and embedded in Epon plastic (Electron Microscopy Sciences, Hatfield, PA, USA). A Balzer’s high-pressure freezing apparatus (HPM 010, Balzers, Liechtenstein) was used for specimen cryofixation and freeze-substitution. Thin sections were stained with 2% uranyl acetate in 70% ethanol and lead citrate, and observed on a JEOL JEM-1200EXII transmission electron microscope (JEOL UK, Hertfordshire, UK) at 100 kV.

Results

Illumina sequencing and RNA sequencing analysis

Illumina sequencing generated 11 to 18 million, 13 to 48 million, 30 to 37 million, and 38 to 39 million reads, respectively, from seed, leaf, stem, and root libraries of the beta-carotene-enhanced transgenic soybean lines 7-1-1 and 10-19-1, and their non-transgenic donor variety Kwangan. Average sequencing coverage of the reference genome (*G. max*) ranged from 1.52 \times to 2.44 \times for seeds, 1.75 \times to 6.72 \times for leaves, 4.18 \times to 5.00 \times for stems, and 5.04 \times to 5.28 \times for roots (Online Resource Table 1). RNA sequencing analysis aligned 24.15 to 30.99% of the mapped reads in seeds, 26.02 to 53.92% in leaves, 68.37 to 75.23% in stems, and 72.44 to 73.49% in roots to the genome and transcriptome mapping references (Online Resource Table 2).

Transgene expression analysis of transgenic soybeans

The RNA read coverage maps of the transgenes *bar*, *psy*, and *crtI* in different tissues of the two transgenic soybeans and their non-transgenic donor Kwangan are shown in Fig. 1. For the herbicide resistance gene *bar*, a large number of sequence reads were observed on read coverage maps of all four tissues in two transgenic soybeans, but no significant accumulation was observed on the corresponding region in their non-transgenic donor variety Kwangan. Similar expression patterns of *bar* in four tissues of two transgenic soybeans were also confirmed using real-time PCR analysis. In addition, RNA-seq analysis indicated that *crtI* under the control of a seed-specific promoter was expressed only in the seeds as shown in read coverage maps, but not in other tissues, of the two transgenic soybeans. Real-time PCR

Table 1 Enriched differentially expressed genes (DEGs) assigned to the kyoto encyclopedia of genes and genomes (KEGG) pathway in the leaves and roots of beta-carotene-enhanced transgenic soybeans

Enriched DEGs	Annotations	Pathways (tissues)	Log2 fold change	FDR-p value
DXR2	1-Deoxy-D-xylulose 5-phosphate reductoisomerase, chloroplastic	gmx01110: Biosynthesis of secondary metabolites (Leaves)	-5.57	2.10E-05
LOC100797583	Alanine-glyoxylate aminotransferase 2 homolog 3, mitochondrial		-5.60	3.06E-06
LOC100783858	Alcohol dehydrogenase 1		-3.49	3.09E-03
LOC100037474	Aureusidin synthase		-3.54	4.01E-03
LOC100802778	Beta-carotene 3-hydroxylase, chloroplastic-like		-3.36	4.87E-03
OAS-TL6	Cysteine synthase		-5.40	1.12E-06
LOC100789922	Dehydrololichyl diphosphate synthase 2-like		-3.84	2.69E-06
LOC100808546	Gibberellin 3-beta-dioxygenase 1-like		-3.04	4.69E-03
LOC100776695	Probable glycerol-3-phosphate acyltransferase 3		-3.77	9.28E-05
LOC100788868	Probable NAD(P)H dehydrogenase (quinone) FQR1-like 3		-2.27	4.62E-03
LOC100783538	Probable solanesyl-diphosphate synthase 3, chloroplastic		-3.28	1.43E-03
LOC100805842	Probable solanesyl-diphosphate synthase 3, chloroplastic		-2.88	1.35E-03
LOC100776947	3-Ketoacyl-CoA synthase 12-like		4.54	2.30E-03
LOC100803579	3-Ketoacyl-CoA synthase 19-like		4.29	6.52E-04
LOC100777403	3-Ketoacyl-CoA synthase 21-like		3.47	2.00E-05
LOC100789515	Cytochrome P450 90B1		2.64	1.31E-05
LOC100815256	Dehydrololichyl diphosphate synthase 2-like		7.17	1.17E-06
LOC100794772	GDP-mannose 3,5-epimerase 2-like		4.96	1.64E-04
LOC100815822	Plastidial pyruvate kinase 2-like		3.45	9.27E-04
LOC100037473	Polyphenol oxidase A1, chloroplastic		4.51	5.23E-07
LOC100801369	Polyphenol oxidase A1, chloroplastic-like		6.31	3.09E-06
LOC100801754	Polyphenol oxidase I, chloroplastic-like		5.98	2.82E-05
LOC100800090	Primary amine oxidase-like		3.23	4.20E-03
LOC100777014	Probable glycerol-3-phosphate acyltransferase 3		5.11	3.07E-08
LOC100781133	Probable glycerol-3-phosphate acyltransferase 3		4.65	1.14E-08
PDH	Proline dehydrogenase		3.28	1.20E-03
LOC100805835	Protein ECERIFERUM 1-like		2.93	2.92E-04
LOC100807448	Cytochrome P450 84A1-like	gmx01110: Biosynthesis of secondary metabolites; gmx00940: Phenylpropanoid biosynthesis (Leaves)	-3.36	2.14E-05
HIUHase	Hydroxyisourate hydrolase		-4.58	2.00E-05
LOC100791275	Peroxidase 55		-3.40	5.43E-04
LOC100783115	Peroxidase A2-like		-4.98	4.35E-06
LOC100810355	Peroxidase A2-like		-5.74	6.99E-05
LOC100779814	Spermidine hydroxycinnamoyl transferase-like		-5.05	4.13E-06
LOC100815268	Beta-glucosidase BoGH3B-like		3.56	2.39E-04
LOC100775837	Peroxidase 12-like		3.18	5.60E-04
LOC100777390	Peroxidase 12-like		2.46	3.70E-03
LOC100804829	Peroxidase 15-like		3.00	5.33E-04
LOC100819684	Probable mannitol dehydrogenase-like		2.94	5.74E-04
LOC100817721	Aldehyde dehydrogenase family 2 member C4-like		-3.91	4.24E-03

Table 1 (continued)

Enriched DEGs	Annotations	Pathways (tissues)	Log ₂ fold change	FDR-p value
INR2	Inducible nitrate reductase 2	gmx00910: Nitrogen metabolism (Roots)	3.03	2.76E–07
LOC100807710	High affinity nitrate transporter 2.4		4.00	2.91E–12
NIR	Nitrate reductase		2.68	1.15E–12
NRT2	NRT2 protein		2.76	2.45E–05
LOC100813471	Nitrate reductase [NADH] 2		2.47	6.13E–05
LOC100782969	High affinity nitrate transporter 2.4-like		2.16	3.01E–05
LOC100775519	Ferredoxin–nitrite reductase, chloroplastic		2.49	2.00E–05
LOC100804007	Galactinol synthase 2	gmx00052: Galactose metabolism (Roots)	–2.40	1.79E–05
LOC100800249	Galactinol synthase 2		–3.22	8.68E–08
LOC100792129	Galactinol synthase 2		–4.13	1.36E–03
LOC100790775	Galactinol-sucrose galactosyltransferase		–2.33	3.85E–05
LOC100815922	Stachyose synthase		–3.26	1.40E–03

analysis revealed that the expression of *crtI* in the leaves, stems, and seeds was consistent with the read coverage maps generated using RNA-seq. However, relatively low expression levels of *crtI* were detected in transgenic soybean roots using real-time PCR (Online Resource Fig. 1). This is a controversial issue as *crtI* should be expressed only in transgenic soybean seeds under the control of seed-specific promoters. For *psy* expression levels in different tissues, we observed many discontinuous read sequences clustering in the leaves and stems of transgenic soybeans and non-transgenic Kwanan. However, full sequence reads of *psy* were only mapped in seeds of transgenic soybeans, indicating seed-specific expression of the inserted *psy*. Soybean has seven endogenous *psy* genes, which respectively share 66.42–73.20% homology with the *Capsicum psy* gene. This resulted in low efficiency of the *Capsicum psy* gene-specific primers for real-time PCR analysis. Under these circumstances, RNA sequencing analysis is an easy alternative approach for evaluation of transgene expression.

DEG identification and GSE analysis in four soybean tissues

To comprehensively understand the changes in the entire transcriptome of beta-carotene-enhanced transgenic soybeans caused by the insertion of a foreign phytoene synthase gene, DEGs were analyzed in four different tissues compared to those in the non-transgenic donor variety Kwanan. A heat map sorted by hierarchical clustering of the top 500 highly expressed genes of the four different soybean tissues revealed similar gene expression patterns in seeds, stems, and roots between transgenic soybeans and their non-transgenic donor Kwanan (Fig. 2a). However, the transgenic soybean leaves were grouped into classes that were different from those of the non-transgenic Kwanan, indicating vast changes in the leaf transcriptome levels of transgenic

soybeans. In total, we detected 795 DEGs in leaves, 84 in stems, 208 in roots, and 13 in seeds from transgenic soybeans based on a threshold of log₂ fold-change > 2 with $P < 0.005$ (Fig. 2b). GSE analysis revealed that 624 DEGs in leaves, 148 in roots, and 47 in stems were enriched in GO categories; however, no enriched DEGs were observed in the seeds.

The enriched DEGs of leaves were classified into 13 GO terms in the biological process category, with the highest enrichment scores in the GO term microtubule-based movement (GO: 0007018) in 26 DEGs. In the cellular component category, 11 GO terms showed significant enrichment, with the most representative term being nucleus (GO: 0005634) in 84 DEGs. In total, 16 enriched GO terms were assigned to the molecular function category, with the most significant term being heme binding (GO: 0020037) in 25 DEGs. The two enriched GO terms that were classified in the KEGG pathway were biosynthesis of secondary metabolites (gmx01110) and phenylpropanoid biosynthesis (gmx00940) (Fig. 3a). In total, 38 enriched DEGs in leaves, which included 20 up-regulated and 18 down-regulated genes, were involved in the biosynthesis of secondary metabolites. Of the 12 DEGs, five up-regulated and seven down-regulated genes were also involved in phenylpropanoid biosynthesis (Table 1).

For DEGs detected in stems, three GO terms in the biological process category, two in the cellular component category, and two in the molecular function category were enriched with the term cell wall organization (GO: 0071555) in seven DEGs, cell wall (GO: 0005618) in nine DEGs, and hydrolase activity/hydrolyzing O-glycosyl compounds (GO: 0004553) in seven DEGs (Fig. 3b). Finally, 21 GO categories were enriched in root DEGs, with eight biological process GO terms, two cellular component GO terms, nine molecular function GO terms, and two terms in the KEGG pathway (Fig. 3c). The GO terms extracellular region

(GO: 0005576) and plasma membrane (GO: 0005886) of the cellular component category, involved in 22 and 20 DEGs, respectively, showed the most significant enrichment scores. In the molecular function category, the GO term heme binding (GO: 0020037) showed the highest enrichment score in 20 DEGs. In the biological process category, the most significantly enriched GO term was response to oxidative stress (GO: 0006979), which was associated with 10 DEGs. Two significantly enriched KEGG pathways, nitrogen metabolism (gmx00910) and galactose metabolism (gmx00052), were associated with seven up-regulated and five down-regulated DEGs, respectively, as shown in Table 1.

Differential gene expression in the transgenic group

To further dissect transcriptomic changes, we analyzed differential gene expression between the transgenic group and non-transgenic group using four tissues combinations. The analysis revealed 99 DEGs consisting of 67 up-regulated and 32 down-regulated genes, which were differentially expressed in the transgenic soybean group (Fig. 2c). GSE analysis revealed 40 enriched DEGs that were classified into two GO terms of molecular function and cellular component, each with only one functional category included (Table 2). In the molecular function GO category, eight up-regulated DEGs belonged to the functional category of hydrolase activity, acting on ester bonds (GO: 0016788) and annotated as glycine-aspartic acid-serine-leucine motif (GDSL) esterase/lipase. In the cellular component GO category,

nine DEGs belonged to the functional category extracellular region (GO: 0005576) and all were up-regulated. At the protein level, all 40 DEGs were closely related to signaling (UP_KEYWORDS), with 28 up-regulated and 12 down-regulated DEGs (Table 3). Among them, seven down-regulated DEGs comprising four fasciclin-like arabinogalactan protein (FLA) 11 and three FLA 12 were entirely categorized into the FAS1 domain (IRP000782), whereas six up-regulated DEGs were categorized into GDSL lipase (IRP001087).

To validate DEGs detected via RNA sequencing analysis, 13 DEGs, consisting of five down-regulated genes involved in raffinose family oligosaccharide biosynthesis and eight up-regulated genes related to GDSL esterase/lipase, were analyzed using real-time PCR. Compared to the non-transgenic Kwangan, the expression levels of three DEGs, which were annotated as galactinol synthase 2 (GOLS2), were significantly down-regulated in the roots of two transgenic soybean lines. Furthermore, the expression of two DEGs, which were respectively annotated as galactinol-sucrose galactosyltransferase (RFS) and stachyose synthase (STS), were significantly repressed in the stems of two transgenic soybeans (Fig. 4a). In addition, real-time PCR analysis indicated that six of eight GDSL esterase/lipase DEGs were significantly up-regulated in the stems, leaves, and seeds of two transgenic soybeans; however, two DEGs (LOC102669928 and LOC100820358) did not show significant changes in expression in transgenic soybean tissues. One of the DEGs (LOC100780781), which was annotated as GDSL esterase/lipase At5g33370,

Table 2 Gene ontology (GO) of enriched differentially expressed genes (DEGs) identified from four tissue combinations of the transgenic group compared to those of the non-transgenic group

GO terms	Enriched DEGs	Annotations	Log2 fold change	FDR-p value
Molecular function (MF): Hydrolase activity, acting on ester bonds (GO: 0016788)	LOC102669928	GDSL esterase/lipase At1g71691-like	3.51	1.61×10^{-3}
	LOC100800835	GDSL esterase/lipase At4g28780	2.62	2.43×10^{-7}
	LOC100791595	GDSL esterase/lipase At4g28780-like	2.24	2.74×10^{-6}
	LOC100780781	GDSL esterase/lipase At5g33370	2.89	3.67×10^{-5}
	LOC100820358	GDSL esterase/lipase At5g33370-like	5.77	8.49×10^{-5}
	LOC100812614	GDSL esterase/lipase At5g45950-like	2.98	1.65×10^{-5}
	LOC100819691	GDSL esterase/lipase At5g45950-like	2.41	5.39×10^{-4}
	LOC100819292	GDSL esterase/lipase LTL1	4.94	7.80×10^{-8}
	Cellular component (CC): Extracellular region (GO: 0005576)	LOC102666289	L-ascorbate oxidase-like	4.38
PR1A		PR1a precursor	9.04	8.68×10^{-5}
LOC100788189		Germin-like protein subfamily 1 member 7-like	6.09	4.02×10^{-5}
LOC106794095		Pathogenesis-related protein 1-like	7.37	1.34×10^{-3}
LOC100775970		Peroxidase 3-like	4.57	6.30×10^{-4}
LOC100792351		Peroxidase 5-like	2.77	5.37×10^{-4}
LOC100812309		Peroxidase 54-like	2.53	5.61×10^{-5}
LOC100776347		RuBisCO-associated protein-like	2.34	1.51×10^{-3}
LOC100527249		Uncharacterized protein	2.71	1.14×10^{-3}

Table 3 Protein functional classifications for enriched differentially expressed genes (DEGs) identified from four tissue combinations of the transgenic group compared to those of the non-transgenic group

Enriched DEGs	Log ₂ fold change	FDR-p value	Gene annotations	Protein category	
LOC100806257	-2.98	2.92×10^{-6}	Laccase-4-like	UP_KEYWORDS: Signal	
LOC102667854	-3.56	1.24×10^{-3}	LRR receptor-like serine/threonine-protein kinase FLS2		
LOC100783861	-3.88	3.31×10^{-3}	Probable receptor-like protein kinase		
LOC100789695	-2.33	1.43×10^{-3}	Probable serine/threonine-protein kinase		
LOC100784385	-3.54	1.39×10^{-6}	Galactose oxidase-like		
BURP2	4.42	7.41×10^{-11}	BURP domain protein 2		
LOC100801057	3.11	9.32×10^{-4}	Glycine-rich cell wall structural protein-like		
LOC102666289	4.38	6.30×10^{-5}	L-ascorbate oxidase-like		
LOC100814809	4.04	1.65×10^{-10}	Metalloendoproteinase 5-MMP-like		
LOC100775970	4.57	6.30×10^{-4}	Peroxidase 3-like		
LOC100812309	2.53	5.61×10^{-5}	Peroxidase 54-like		
LOC100792351	2.77	5.37×10^{-4}	Peroxidase 5-like		
LOC100788189	6.09	4.02×10^{-5}	Germin-like protein subfamily 1 member 7-like		
LOC106794095	7.37	1.34×10^{-3}	Pathogenesis-related protein 1-like		
PR1A	9.04	8.68×10^{-5}	PR1a precursor		
LOC100786014	3.56	1.42×10^{-3}	Proline-rich protein 4-like		
LOC100818122	4.30	3.38×10^{-4}	Protein HOTHEAD-like		
LOC100778987	3.09	3.21×10^{-3}	Thaumatococcus-like protein 1b		
LOC100789557	2.84	5.55×10^{-6}	Zinc transporter 8-like		
LOC100305549	2.02	2.76×10^{-3}	Uncharacterized protein		
LOC100527618	6.43	6.98×10^{-4}	Uncharacterized protein		
LOC100527742	2.11	1.08×10^{-3}	Uncharacterized protein		
LOC100798749	5.69	2.20×10^{-4}	Uncharacterized protein		
LOC100799805	5.13	1.11×10^{-3}	Uncharacterized protein		
LOC100805566	2.77	1.11×10^{-3}	Uncharacterized protein		
LOC100810866	4.50	1.93×10^{-7}	Uncharacterized protein		
LOC100527249	2.71	1.11×10^{-3}	Uncharacterized protein		
LOC100790242	-2.93	2.27×10^{-3}	Fasciclin-like arabinogalactan protein 11	IPR000782: FAS1 domain	
LOC100793220	-2.76	1.43×10^{-3}	Fasciclin-like arabinogalactan protein 11		
LOC100812914	-2.53	1.87×10^{-3}	Fasciclin-like arabinogalactan protein 11		
LOC106794109	-2.32	3.09×10^{-3}	Fasciclin-like arabinogalactan protein 11		
LOC100783215	-2.46	1.51×10^{-3}	Fasciclin-like arabinogalactan protein 12		
LOC100793759	-2.58	4.95×10^{-3}	Fasciclin-like arabinogalactan protein 12		
LOC100795331	-2.82	2.47×10^{-3}	Fasciclin-like arabinogalactan protein 12		
LOC102669928	3.51	1.61×10^{-3}	GDSL esterase/lipase		IPR001087: Lipase, GDSL
LOC100800835	2.62	2.43×10^{-7}	GDSL esterase/lipase		
LOC100791595	2.24	2.74×10^{-6}	GDSL esterase/lipase		
LOC100780781	2.89	3.69×10^{-5}	GDSL esterase/lipase		
LOC100812614	2.98	1.65×10^{-5}	GDSL esterase/lipase		
LOC100819691	2.41	5.39×10^{-4}	GDSL esterase/lipase		

showed 4.52–6.23 log₂ fold up-regulation in the stems and 1.72–4.20 log₂ fold in the leaves of two transgenic soybeans (Fig. 4b). Thus, 11 of the 13 DEGs detected using RNA sequencing transcriptome analysis were confirmed to show significant changes in expression using real-time PCR analysis.

Internal morphology of seeds

The cellular and intracellular structures of soybean seeds of transgenic lines 7-1-1, 9-1-2, and 10-19-1, and their non-transgenic donor variety Kwangan, were observed at $\times 1600$ and $\times 6300$ magnification. Similar cellular morphologies and number

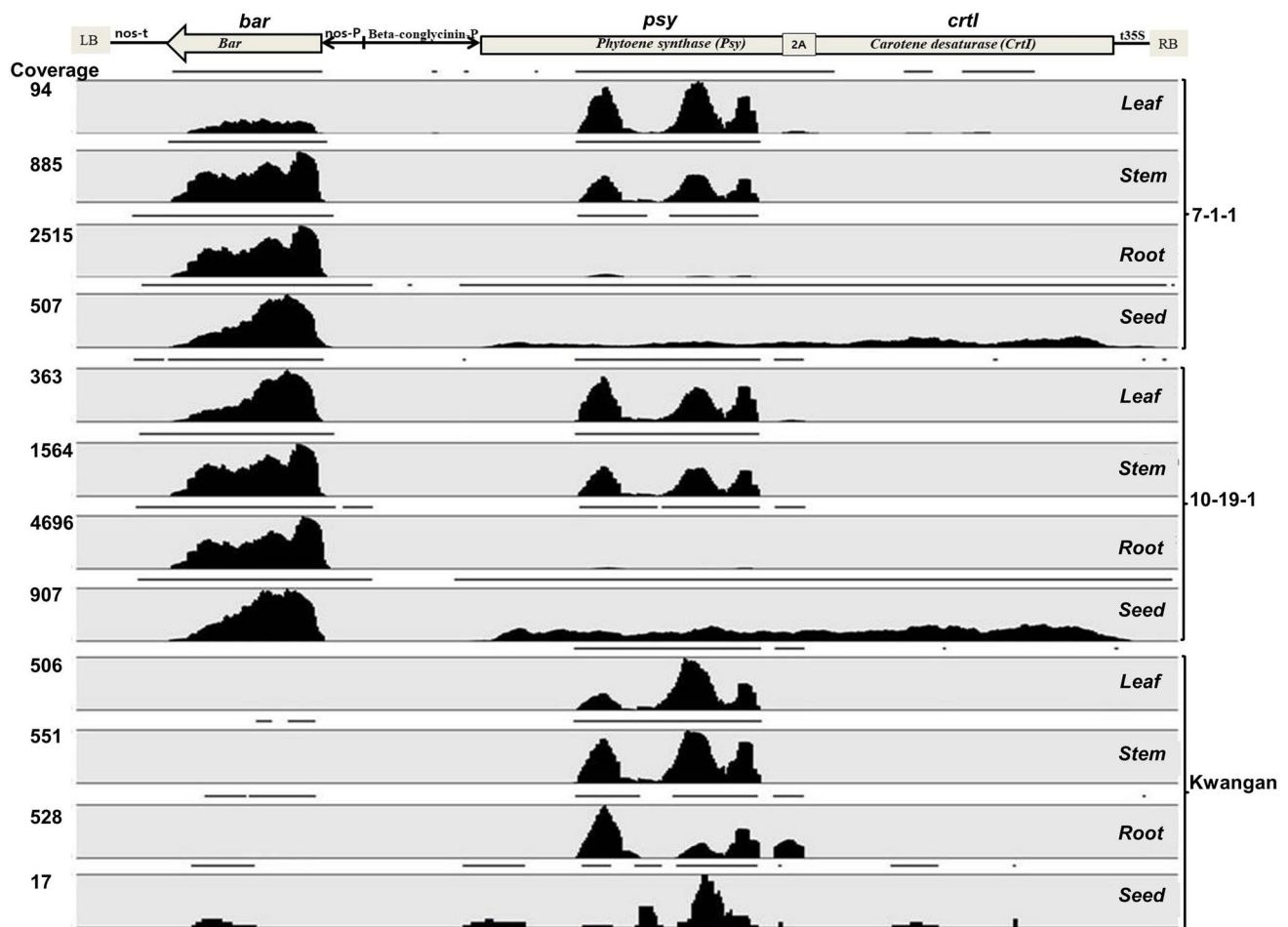


Fig. 1 Read coverage maps of the transcript levels of transgenes using RNA-sequencing analysis

and size of protein storage vacuole (PSV) were observed in the four soybean samples (Fig. 5a). However, we observed increase in phytin globoid (PG) deposition in protein bodies in the beta-carotene-enhanced transgenic soybeans, with the highest quantity in the 7-1-1 line. No PG was observed in the TEM images of seeds from the non-transgenic variety Kwangan. In addition, compared to the non-transgenic variety Kwangan, the transgenic soybeans exhibited diversity in the size, distribution, number, and appearance of oil bodies (OB) (Fig. 5b). The transgenic soybean line 7-1-1 showed OBs with high density and small size, whereas 10-19-1 showed relatively larger OBs and PSVs with irregular surroundings. However, the three transgenic soybean lines presented OBs that almost occupied the entire cytoplasm (Fig. 5b).

Discussion

To overcome vitamin A deficiency in infants and pregnant women in Africa and Southeast Asia via food intake, many efforts had been done to conduct beta-carotene production

by manipulating enzymes related to carotenoid biosynthesis in staple crops including of rice, tomatoes, maize and soybeans (Ye et al. 2000; Paine et al. 2005; Römers et al. 2000; Diretto et al. 2007; Zhu et al. 2008; Schmidt et al. 2014). However, some studies had reported unintended effects caused by seed-specific carotenoid biosynthesis influenced core metabolic processes such as carbohydrate metabolism and fatty acid metabolism (Decourcelle et al. 2015; Gayen et al. 2016). In the present study, transcriptome profiles in beta-carotene enhanced transgenic soybean leaf, stem, root and seed were comparatively analyzed to provide an insight into the transcriptome alternations resulting from over-expression of *psy* and *crtI* gene cassette. Theoretically, the inserted *psy* and *crtI* genes under the control of seed-specific promoter beta-conglycinin should be expressed only in developing soybean seeds. However, real-time PCR was performed in this study, the results of which indicated that *crtI* was expressed not only in seeds, but also in root tissues with low expression levels of the two transgenic soybean lines (Online Resource Fig. 1). Even the *crtI* expression was not detected in transgenic soybean roots by RNA-seq

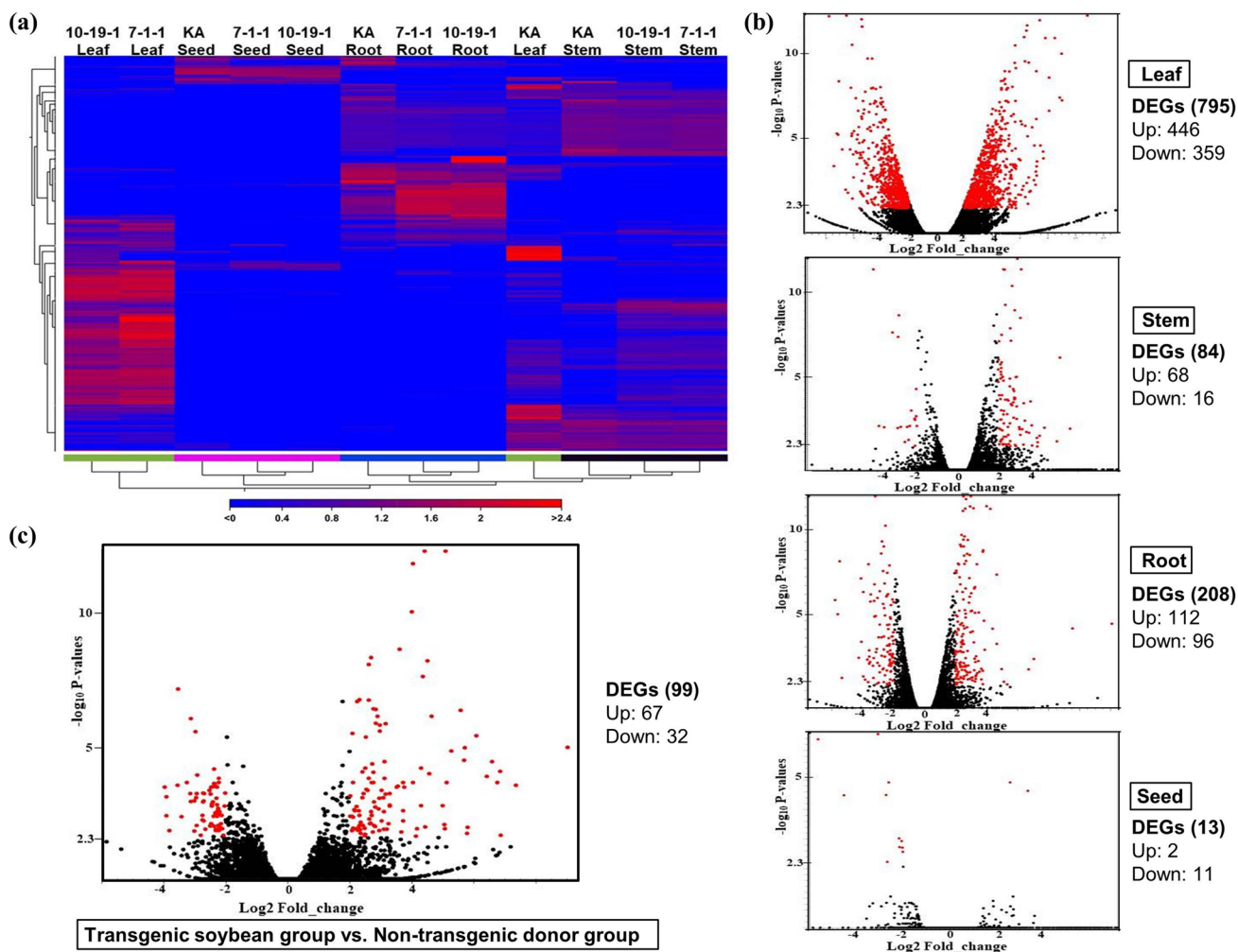


Fig. 2 Differentially expressed genes (DEGs) detected in beta-carotene-enhanced transgenic soybeans after comparison with their non-transgenic counterpart donor Kwangan (KA). **a** Hierarchical clustering analysis of the top 500 highly expressed genes in different tissues.

b Volcano plots of DEGs detected in leaf, stem, root, and seed of transgenic soybean lines. **c** Differential gene expression in the transgenic soybean group compared to those of the non-transgenic donor group using four soybean tissue combinations

analysis, might be because low read depth of RNA-sequencing was difficult to detect low level gene expressions. Several studies have reported seed-specific promoter directed gene expression in non-seed tissues (Baumlein et al. 1991; Zakharov et al. 2004). Baumlein et al. (1991) reported that the *usp*-promoter-controlled *uidA* was not only observed in developing seeds but also in the root tips of *Nicotiana tabacum* and *Arabidopsis thaliana* seedlings. Zakharov et al. (2004) re-investigated some representatives of the 7S and 11S storage globulin gene promoter families and observed that several storage globulin and non-storage protein gene promoters are not only active in the cotyledons and seed coat of developing seeds, but also in the pollen. Beta-conglycinin is a subunit of 7S globulin and the promoter mainly directs gene expressions related to 7S storage protein. We believed that the beta-conglycinin promoter directs low levels of gene expression in the root, root tips, or other root tissues in

the present study. In addition, seed samples were collected in the mature stage with yellow pods. Comparing to seed developing stage, genes related to seed maturation, storage protein and lipid biosynthesis are mainly expressed in seed maturity stage. There was no enriched DEGs detection from transgenic soybean seeds, which indicated that beta-carotene enhanced transgenic soybeans had the identical transcriptome profiles of maturity seeds to their donor variety Kwangan in the present study.

Previous studies have reported that transgenic plants are regenerated via tissue culture manipulations, which can elicit somaclonal variation due to mutations, transposable elements, and epigenetic alterations. In addition, T-DNA insertion can trigger some unintended effects such as position effects and chromosome translocations (Jiang et al. 2011; Sabot et al. 2011; Endo et al. 2015; Anderson et al. 2016; Ko et al. 2018). By using next-generation sequencing

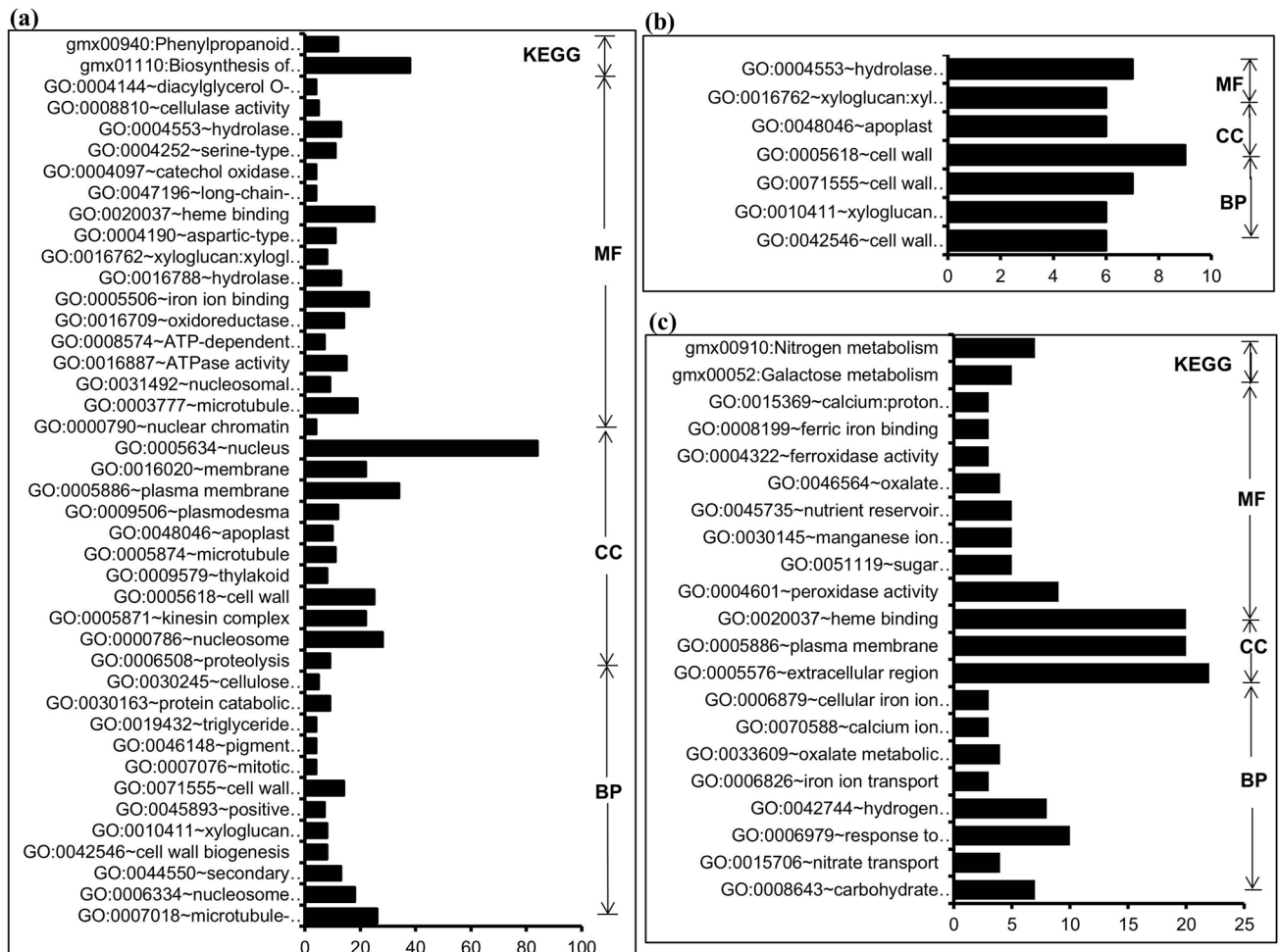


Fig. 3 Gene ontology (GO) analysis for enriched differentially expressed genes (DEGs) detected in the leaf (a), stem (b), and root (c) of beta-carotene-enhanced transgenic soybeans. KEGG, Kyoto

Encyclopedia of Genes and Genomes; *MF* molecular function, *CC* cellular component, *BP* biological process

in *Arabidopsis* and rice, genome-wide single nucleotide changes resulting from tissue culture have been observed. These studies suggested that tissue culture might increase the single nucleotide mutation rate and may activate transposons (Jiang et al. 2011; Sabot et al. 2011; Endo et al. 2015). A study was carried out to investigate effects of transgene insertion and somaclonal variation of transgenic soybeans derived from *Agrobacterium*-mediated transformation and fast neutron mutants (Anderson et al. 2016). They found that the genetic transformation process infrequently results in novel genetic variation, but single nucleotide substitution rates were similar between fast neutron and transformed plants. Transcriptomic variability caused by plant transformation in transgenic potato were studied and revealed not only transcription factors but also specific transcription factor binding sites were also enriched in promoter region of differentially expressed gene in transgenic lines (Ko et al. 2018). In the present study, we did not prepare the negative

control of transgenic soybean with empty vector transformation due to the incomprehensive considerations. Therefore, we considered that many enriched DEGs detected from leaves, stems and roots had a few of possibilities to be induced by somaclonal variations or position effects by transgene insertion. To reduce the possible effects induced by somaclonal variations, comparative transcriptome profiles of four soybean tissues of the transgenic and non-transgenic groups were analyzed individually and in combination. Subsequently, gene ontology and KEGG analysis of all the detected DEGs were performed to determine the related function categories and biosynthetic metabolic pathways. As a result, a total of 624, 47 and 148 of enriched DEGs were detected in beta-carotene enhanced transgenic soybean leaves, stems, and roots, respectively.

In soybean leaf transcriptomes, 38 enriched DEGs involved in the biosynthesis pathway of secondary metabolites were identified from beta-carotene enhanced transgenic

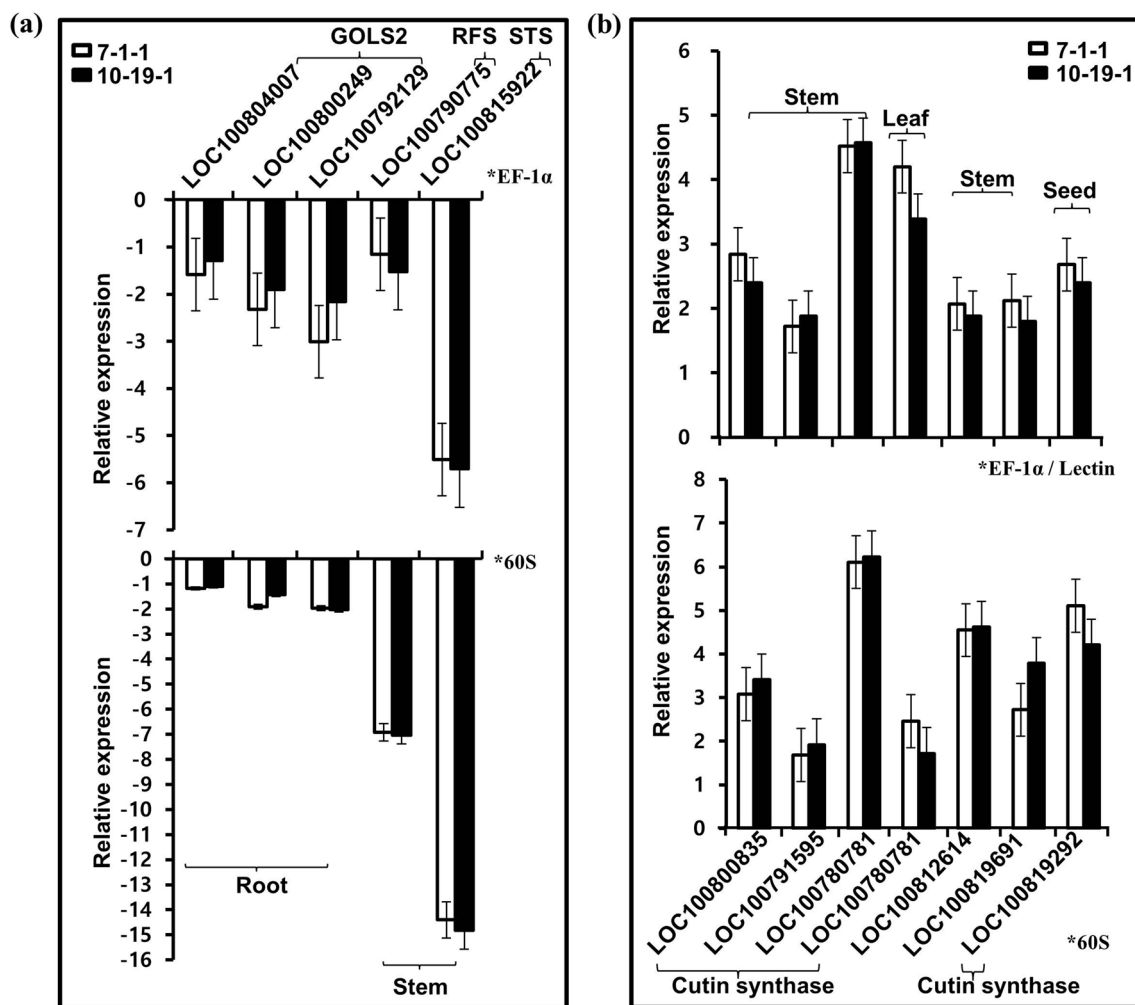


Fig. 4 Differentially expressed genes in beta-carotene-enhanced transgenic soybeans were analyzed using real-time PCR. **a** Down-regulation of five genes related to raffinose family oligosaccharide

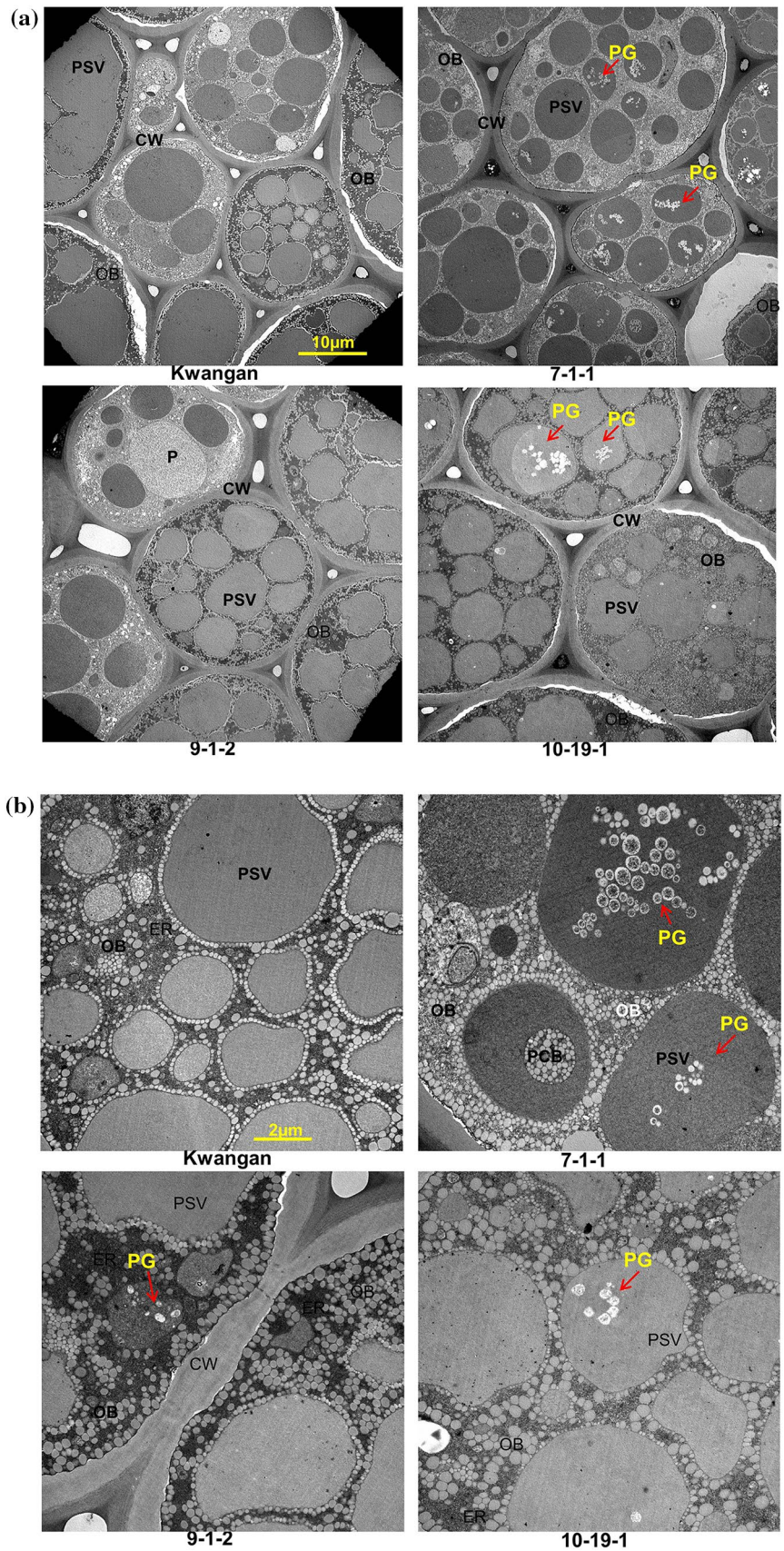
biosynthesis in transgenic soybean roots and stems. **b** Up-regulation of six genes related to glycine-aspartic acid-serine-leucine-motif (GDSL) esterase/lipase in transgenic soybean leaves, stems, and seeds

soybeans. Among these DEGs, genes encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*DXR2*) and beta-carotene 3-hydroxylase, which are located upstream of the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway and downstream of the carotenoid biosynthesis pathway, respectively, were significantly down-regulated (Table 1). In soybean (*G. max.*), geranyl-geranyl diphosphate (GGPP) is synthesized via the MEP pathway and is used as a substrate for producing phytoene. The desaturation and isomerization reactions were catalyzed by *PDS*, *ZISO*, *ZDS*, and *LCY-b* in the carotenoid biosynthetic pathway following beta-carotene production (KEGG; reviewed by Giuliano 2014). Overexpression of the inserted *psy* might trigger competition among common substrates and/or precursors so as to repress the expression of endogenous *psy* genes and enzymes involved in the upstream process. In addition, beta-carotene 3-hydroxylase is a critical enzyme required for the biosynthesis of

zeaxanthin. Diretto et al. (2007) reported that silencing of the beta-carotene hydroxylases *CHY1* and *CHY2* in potato tuber decreased the level of zeaxanthin, the intermediate product of beta-carotene hydroxylation, and increased beta-carotene content by up to 14-fold. Results from the present study indicated that beta-carotene overproduction possibly triggered the down-regulation of beta-carotene hydroxylase and inhibited zeaxanthin production.

Furthermore, we identified 40 enriched DEGs in the beta-carotene-enhanced transgenic soybean group with four tissue combinations by comparing to those of non-transgenic group. Among them, four up-regulated DEGs, which were validated using qRT-PCR, were functionally categorized into cutin synthase-encoding genes, including *LOC100800835* (GDSL esterase/lipase *At4g28780*), *LOC100791595* (GDSL esterase/lipase *At4g28780-like*), *LOC100780781* (GDSL esterase/lipase *At5g33370*), and

Fig. 5 Comparison of the seed cytoplasmic morphology of beta-carotene-enhanced transgenic soybean lines with their non-transgenic donor Kwangan using transmission electron microscopy (TEM). **a** TEM at $\times 1600$ magnification. **b** TEM at $\times 6300$ magnification. Red arrows indicate phytin globoid. *P* protein body, *PSV* protein storage vacuole, *PG* phytin globoid, *OB* oil body, *PCB* protein carbohydrate body, *ER* endoplasmic reticulum, *CW* cell wall. (Color figure online)



LOC100819292 (GDSL esterase/lipase *LTL1*). Several studies have shown these to encode GDSL family proteins and to localize to the developing cuticle, affecting cutin deposition (Girard et al. 2012; Yeats et al. 2012; Bakan and Marion 2017; Hong et al. 2017). Two other up-regulated DEGs, *LOC100812614* and *LOC100819691*, were annotated as GDSL esterase/lipase *At5g45950-like*, which confirmed the increase in the expression of beta-carotene-enhanced transgenic soybean stems. They were reported to be related to cuticle metabolism by Panikashvili et al. (2010). A study on transcriptome profiling of an orange-pericarp pummelo mutant revealed that some biochemical and molecular alterations on carbon metabolism, starch and sucrose metabolism, biosynthesis of amino acids were related to the high beta-carotene accumulation (Guo et al. 2015). They identified top ten most DEGs involved in the regulation of carotenoid biosynthesis. One of DEGs encoding GDSL esterase/lipase was significantly increased in the mutant lines. The GDSL esterase/lipase was involved in fatty acid metabolism. The biosynthesis of fatty acid and carotenoids both required the common precursors from pyruvate. Schmidt et al. (2014) studied nutrient compositions, transcriptome, proteomic and metabolomic analysis for transgenic soya bean plants over-expressing a seed-sepcific bacterial phytoene synthase gene. The transcript comparative analysis revealed down-regulated gene expressions of fatty acid desaturase 2 (FAD2), FAD3b and FAD3-2b and up-regulated gene expression of FAD3a in enhanced beta-carotene seeds, resulting in an increase of oleic acid and a decrease of linoleic acid content. Three transgenic maize lines engineered for enhanced carotenoid synthesis with *psy* and *crtI* gene integrations were studied to assess whether endosperm-specific carotenoid biosynthesis influenced core metabolic processes by using combined transcript, proteome and metabolite analysis (Decourcelle et al. 2015). Results revealed changes in carbohydrate metabolism and increase of terpenoids and fatty acids in embryo and endosperm of maize seeds. These results provided some insights that beta-carotene accumulation by phytoene synthase over-expression could affect gene expressions in fatty acid metabolism. In the present study, we considered that over-expression of *psy* and *crtI* in carotene biosynthesis pathway could result in a redundancy of the common precursor pyruvate, consequently promote up-regulations of GDSL esterase/lipase genes.

During the selection and reproduction processes of the beta-carotene-enhanced transgenic soybean lines, we also observed relatively lower and delayed seed germination rates than those of the non-transgenic donor Kwangan. Lindgren et al. (2003) studied transgenic *Arabidopsis* in which the phytoene synthase gene was introduced under the control of a seed-specific promoter; results indicated that over-production of carotenoids leads to increased levels of abscisic acid (ABA), which correlates positively with delayed seed

germination. Previously, we have measured the ABA levels in beta-carotene-enhanced transgenic soybean seeds, but did not observe any significant difference from that of the non-transgenic donor Kwangan (Qin et al. 2016). Results of the seed germination test for non-transgenic Kwangan and three beta-carotene-enhanced transgenic soybeans revealed 52% and 0–14% germination rates, respectively, after 44 h of water absorption (Online Resource Fig. 2). After 48 h, the average radicle lengths of germinated seeds ranged from 0.6 ± 0.2 cm to 1.6 ± 0.4 cm in the transgenic soybean lines, compared to an average of 2.1 ± 0.2 cm in Kwangan. Seed water imbibition was also measured for the three beta-carotene-enhanced transgenic soybean lines and their donor Kwangan. After 2 h of water uptake, the weights of the transgenic soybeans increased from 113.9 to 124.2%, compared to that of Kwangan, which showed 149.6% weight increase (Online Resource Fig. 3). After water uptake for 24 h, the transgenic soybeans showed peak in weight increase ranging from 53.3 to 104.6% and the substantial weight increased from 23.6 to 29% after 48 h of water take. There were still 5–15 hard seeds in three transgenic soybean lines at 48 h of water uptake. The weight increase of Kwangan stopped at 48 h of water uptake. These results indicated that beta-carotene-enhanced transgenic soybeans had low and delayed seed germination. One of the reasons might be that the thickening of the seed coat was caused by up-regulated cutin polymerization. In addition, the seed internal structures of non-transgenic Kwangan and three transgenic soybeans showed abnormal OBs in transgenic soybean seeds. Several studies indicate that irregular OB size (large or small) is related to the ratio of oleosin to triglyceride (TAG) oils, which affect seed dormancy and oil content (Siloto et al. 2006; Schmidt and Herman 2008; Shimada and Hara-Nishimura 2010). Therefore, we suggested that the anomalous OBs disrupt cell structure, which may be partially responsible for its delayed seed germination.

A proteomics study on transgenic carotenoid-enriched golden rice lines revealed the main alteration to be in carbohydrate metabolism pathways, such as up-regulation of pullulanase, UDP-glucose pyrophosphorylase, and glucose-1-phosphate adenylyltransferase in transgenic rice seeds (Gayen et al. 2016). They also performed metabolite profiling and demonstrated increase in glyceric acid, fructo-furanose, and galactose content and decrease in galactonic acid and gentiobiose content in transgenic golden rice seeds. By using a combined analysis of transcript, proteomes and metabolites analysis on transgenic maize lines with enhanced carotenoid biosynthesis by *psy* and *crtI* insertion revealed significant compound increases of sucrose and sorbitol and decreases of glucose, fructose and xylose. The relevant enzymes of carbohydrate metabolism such as glucose 6-phosphate isomerase and fructose-bisphosphate aldolase, providing the precursor

for carotenoid biosynthesis, were increased in transgenic lines (Decourcelle et al. 2015). In the present study, we revealed significant down-regulation of galactose metabolism in the root transcriptome of beta-carotene-enhanced soybeans. The related enzymes were encoded by three *GOLS2* genes, one *STS* gene, and one *RFS* gene. *GOLS2* initiates the biosynthesis of raffinose family oligosaccharides by catalyzing the conversion of UDP-galactose and myo-inositol to galactinol, whereas *RFS* and *STS* catalyze the conversion of galactinol and raffinose to myo-inositol and stachyose (Gangl and Tenhaken 2016). UDP-galactose is made from UDP-glucose by nucleotide-donor transfer to galactose 1-phosphate or by epimerization of the glucose moiety. Carbohydrate metabolism provides the key precursors not only for UDP-glucose formation but also carotenoids biosynthesis (Decourcelle et al. 2015). We speculated enhanced carotenoid biosynthesis could trigger precursor competitions leads to decreasing in UDP-glucose formation. Therefore, down-regulation of *GOLS2* may affect galactinol biosynthesis, owing to the redundancy of myo-inositol as a substrate. In addition, myo-inositol and glucose-6-phosphate participate in the biosynthesis of phytic acid (myo-inositol-1,2,3,4,5,6-hexakisphosphate; InsP_6). InsP_6 , in the form of phytate salts, bind important mineral cations such as iron, potassium and calcium, and is located within PSVs, known as PG (Panzeri et al. 2011). The internal morphology of the seeds of beta-carotene-enhanced transgenic soybeans suggested a different degree of PG deposition in seeds, although no PG was observed in the non-transgenic Kwangan seeds. We speculated that the myo-inositol redundancy in beta-carotene-enhanced transgenic soybeans increased the InsP_6 levels, resulting in PG accumulation. A previous study on nutritional analysis of beta-carotene-enhanced soybean seeds indirectly confirmed that the content of calcium, potassium, and iron were significantly higher in transgenic soybeans than in non-transgenic donor Kwangan (Qin et al. 2017a).

In conclusion, the above analysis indicated that DEGs related to the GDSL esterase/lipase family and galactose metabolism were highly connected with insertions and overexpression of *psy* and *crtl* genes, but were not triggered by position effects or somaclonal variations. Therefore, genetic manipulations of the carotenoid biosynthetic pathway may affect the transcriptomes of soybean leaves and roots, resulting in some unintended effects in soybean seeds, such as low germination rates and delayed germination, and accumulation of indigestible phytin globoid. Herein, we suggested that metabolic engineering of plant carotenoid biosynthesis, focusing on beta-carotene overproduction, should be carefully performed on plants such as legumes, which possess a native beta-carotene biosynthetic pathway.

Acknowledgements This work was supported by a grant (PJ01090201 and PJ01432206) from the National Institute of Agricultural Sciences (Rural Development Administration, Republic of Korea).

Author contributions YQ, HJW, and SKL conceived the experiments; YQ performed the experiment, data analysis, and writing; MHL and SKL reviewed and edited the manuscript; KSS and MHL supported the experimental devices; MHL and SKL were involved in project administration; all authors reviewed the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson JE, Michno JM, Kono TJY, Stec AO, Campbell BW, Curtin SJ, Stupar RM (2016) Genomic variation and DNA repair associated with soybean transgenesis: a comparison to cultivars and mutagenized plants. *BMC Biotechnol* 16:41. <https://doi.org/10.1186/s12896-016-0271-z>
- Bakan B, Marion D (2017) Assembly of the cutin polyester: from cells to extracellular cell walls. *Plants* 6(4):57. <https://doi.org/10.3390/plants6040057>
- Baumlein H, Boerjan W, Nagy I, Bassuner R, Van Montague M, Inze D WU (1991) A novel seed protein gene from *Vicia faba* is developmentally regulated in transgenic tobacco and *Arabidopsis* plants. *Mol Gen Genet* 225:459–467
- Beyer P (2010) Golden rice and ‘golden’ crops for human nutrition. *New Biotechnol* 27:478–481. <https://doi.org/10.1016/j.nbt.2010.05.010>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Decourcelle M, Perez-Fons L, Baulande S, Steiger S, Couvelard L, Hem S, Zhu CF, Capell T, Christou P, Fraser P, Sandmann G (2015) Combined transcript, proteome, and metabolite analysis of transgenic maize seeds engineered for enhanced carotenoid synthesis reveals pleiotropic effects in core metabolism. *J Exp Bot* 66:3141–3150
- Diretto G, Al-Babili S, Tavazza R, Papacchioli V, Beyer P, Giuliano G (2007) Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. *PLoS ONE* 4:e350. <https://doi.org/10.1371/journal.pone.0000350>
- Endo M, Kumagai M, Motoyama R, Sasaki-Yamagata H, Mori-Hosokawa S, Hamada M, Kanamori H, Nagamura Y, Katayose Y, Itoh T, Toki S (2015) Whole-genome analysis of herbicide-tolerant mutant rice generated by *Agrobacterium*-mediated gene targeting. *Plant Cell Physiol* 56(1):116–125. <https://doi.org/10.1093/pcp/pcu153>
- Fray RG, Wallace A, Fraser PD, Valero D, Hedden P, Bramley PM, Grierson D (1995) Constitutive expression of a fruit phytoene synthase gene in transgenic tomatoes causes dwarfism by redirecting metabolites from the gibberellin pathway. *Plant J* 8:693–701. <https://doi.org/10.1046/j.1365-3113X.1995.08050693.x>

- Gangl R, Tenhaken R (2016) Raffinose family oligosaccharides act as galactose stores in seeds and are required for rapid germination of *Arabidopsis* in the dark. *Front Plant Sci* 7:1115. <https://doi.org/10.3389/fpls.2016.01115>
- Gayen D, Ghosh S, Paul S, Sarkar SN, Datta SK, Datta K (2016) Metabolic regulation of carotenoid-enriched golden rice line. *Front Plant Sci* 7:1622
- Girard AL, Mounet F, Lemaire-Chamley M, Gaillard C, Elmorjani K, Vivancos J, Runavot JL, Quemener B, Petit J, Germain V, Rothan C, Marion D, Bakan B (2012) Tomato GDSL1 is required for cutin deposition in the fruit cuticle. *Plant Cell* 24:3119–3134. <https://doi.org/10.1105/tpc.112.101055>
- Giuliano G (2014) Plant carotenoids: genomics meets multi-gene engineering. *Curr Opin Plant Biol* 19:111–117. <https://doi.org/10.1016/j.pbi.2014.05.006>
- Guo F, Yu HW, Xu Q, Deng XX (2015) Transcriptomic analysis of differentially expressed genes in an *orange-pericarp* mutant and wild type in pummelo (*Citrus grandis*). *BMC Plant Biol* 15:44. <https://doi.org/10.1186/s12870-015-0435-3>
- Ha SH, Liang YS, Jung H, Ahn MJ, Suh SC, Kweon SJ, Kim DH, Kim YM, Kim JK (2010) Application of two bicistronic systems involving 2A and IRES sequences to the biosynthesis of carotenoids in rice endosperm. *Plant Biotechnol J* 8:928–938. <https://doi.org/10.1111/j.1467-7652.2010.00543.x>
- Hong L, Brown J, Segerson NA, Rose JKC, Roeder AHK (2017) Cutin synthase 2 maintains progressively developing cuticular ridges in *Arabidopsis* sepals. *Mol Plant* 10:560–574. <https://doi.org/10.1016/j.molp.2017.01.002>
- Jiang C, Mithani A, Gan X, Belfield EJ, Klingler JP, Zhu JK, Ragoussis J, Mott R, Harberd NP (2011) Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Curr Biol* 21(16):1385–1390
- Kim MJ, Kim JK, Kim HJ, Pak JH, Lee JH, Kim DH, Choi HK, Jung HW, Lee JD, Chung YS, Ha SH (2012) Genetic modification of the soybean to enhance the beta-carotene content through seed-specific expression. *PLoS ONE* 7:e48287. <https://doi.org/10.1371/journal.pone.0048287>
- Ko DK, Nadakuduti SS, Douches DS, Buell CR (2018) Transcriptome profiling of transgenic potato plants provides insights into variability caused by plant transformation. *PLoS ONE* 13:e0206055. <https://doi.org/10.1371/journal.pone.0206055>
- Lindgren LO, Stålberg KG, Höglund AS, (2003) Seed-specific over-expression of an endogenous *Arabidopsis* phytoene synthase gene results in delayed germination and increased levels of carotenoids, chlorophyll, and abscisic acid. *Plant Physiol* 132:779–785
- Organization for Economic Cooperation and Development (OECD) Environment, Health and Safety Publications (2012) Revised consensus document on compositional considerations for new varieties of [*Glycine max* (L.) Merr.]: key food and feed nutrients, antinutrients, toxicants and allergens. Series on the safety of novel foods and feeds. No. 25. Environment Directorate. Organisation for economic co-operation and development, Paris, France.
- Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of golden rice through increased pro-vitamin A content. *Nat Biotechnol* 23:482–487
- Panikashvilia D, Shia JX, Bocobzaa S, Frankeb R, Schreiberb L, Aharoni A (2010) Cutin metabolism in reproductive organs and suberin in roots. *Mol Plant* 3(3):563–575. <https://doi.org/10.1093/mp/ssp103>
- Panzeri D, Cassani E, Doria E, Tagliabue G, Forti L, Campion B, Bollini R, Brearley CA, Pilu R, Nielsen E, Sparvoli F (2011) A defective ABC transporter of the MRP family, responsible for the bean *lpa1* mutation, affects the regulation of the phytic acid pathway, reduces seed myo-inositol and alters ABA sensitivity. *New Phytol* 191(1):70–83. <https://doi.org/10.1111/j.1469-8137.2011.03666.x>
- Qin Y, Kweon SJ, Chung YS, Ha SH, Shin KS, Lim MH, Kwon TR, Cho HS, Park SK, Woo HJ (2015) Selection of beta-carotene enhanced transgenic soybean containing single-copy transgene and analysis of integration sites. *Korean J Breed Sci* 47:111–117. <https://doi.org/10.9787/KJBS.2015.47.2.111>
- Qin Y, Shin KS, Seo JK, Lim MH, Chung YS, Cho HS, Lee SK, Woo HJ (2016) Assessment of the response of beta-carotene enhanced transgenic soybeans to soybean mosaic virus. *Plant Breed Biotechnol* 4:158–169. <https://doi.org/10.9787/PBB.2016.4.2.158>
- Qin Y, Park SY, Oh SW, Lim MH, Shin KS, Cho HS, Lee SK, Woo HJ (2017) Nutritional composition analysis for beta-carotene-enhanced transgenic soybeans (*Glycine max* L.) *Appl Biol Chem* 60(3), 299–309.
- Qin Y, Woo HJ, Shin KS, Lim MH, Cho HS, Lee SK (2017) Flanking sequence and copy-number analysis of transformation events by integrating next-generation sequencing technology with southern blot hybridization. *Plant Breed Biotechnol* 5:269–281. <https://doi.org/10.9787/PBB.2017.5.4.269>
- Römer S, Fraser PD, Kiano JW, Shipton CA, Misawa N, Schuch W, Bramley PM (2000) Elevation of the provitamin A content of transgenic tomato plants. *Nat Biotechnol* 18:666–669
- Sabot F, Picault N, El-Baidouri M, Llauro C, Chaparro C, Piegu B, Roulin A, Guiderdoni E, Delabastide M, McCombie R, Panaud O (2011) Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant J* 66(2):241–246. <https://doi.org/10.1111/j.1365-313X.2011.04492.x>
- Sandmann G (2001) Genetic manipulation of carotenoid biosynthesis: strategies, problems and achievements. *Trends Plant Sci* 6:14–17
- Schmidt MA, Herman EM (2008) Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes. *Mol Plant* 1:910–924. <https://doi.org/10.1093/mp/ssn049>
- Schmidt MA, Parrott WA, Hildebrand DF, Berg HR, Cooksey A, Pendarvis K, He YH, McCarthy F, Herman EM (2014) Transgenic soya bean seeds accumulating β -carotene exhibit the collateral enhancements of oleate and protein content traits. *Plant Biotechnol J*. <https://doi.org/10.1111/pbi.12286>
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3:1101–1108
- Shimada TL, Hara-Nishimura I (2010) Oil-body-membrane proteins and their physiological functions in plants. *Biol Pharm Bull* 33:360–363
- Siloto RM, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, Moloney MM (2006) The accumulation of oleosins determines the size of seed oil bodies in *Arabidopsis*. *Plant Cell* 18:1961–1974
- Ye X, Al-Babili S, Klott A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (β -carotene) biosynthetic pathway. *Science* 287:303–305
- Yeats TH, Martin LBB, Viart HMF, Isaacson T, He YH, Zhao LX, Matas AJ, Buda GJ, Domozych DS, Clausen MH, Rose JKC (2012) The identification of cutin synthase: formation of the plant polyester cutin. *Nat Chem Biol* 8(7):609–611. <https://doi.org/10.1038/NCHEMBIO.960>
- Zakharov A, Giersberg M, Hosein F, Melzer M, Müntz K, Saalbach, (2004) Seed-specific promoters direct gene expression in non-seed tissue. *J Exp Bot* 55(402):1463–1471
- Zhu C, Naqvi S, Breitenbach J, Sandmann G, Christou P, Capell T (2008) Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA* 105:18232–18237. <https://doi.org/10.1073/pnas.0809737105>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.