PLANT SCIENCE TODAY
Vol x(x): xx–xx
https://doi.org/10.14719/pst.12065

eISSN 2348-1900

RESEARCH ARTICLE

Differentially expressed gene profiles of soybean seeds with contrasting seed coat colour

Vidya C J[1], Nethra N[1*], Satya S[1], Karthik N[2], Nagesha S N[3], Parashivamurthy[1] & Vasudevan S N[1]

[1]Department of Seed Science and Technology, University of Agricultural Sciences, Gandhi Krishi Vignana Kendra, Bengaluru 560 065, Karnataka, India
[2]Centre for Advanced Studies in Botany, University of Madras, Guindy Campus, Chennai 600 025, Tamil Nadu, India
[3]Department of Bio-technology, University of Agricultural Sciences, Gandhi Krishi Vignana Kendra, Bengaluru 560 065, Karnataka, India

*Correspondence email - nethraharsha@gmail.com

Received: 21 October 2025; Accepted: 11 March 2026; Available online: Version 1.0: 27 April 2026

Cite this article: Vidya CJ, Nethra N, Satya S, Karthik N, Nagesha SN, Parashivamurthy, Vasudevan SN. Differentially expressed gene profiles of soybean seeds with contrasting seed coat colour. Plant Science Today (Early Access). https://doi.org/10.14719/pst.12065

Abstract

Soybeans are a vital source of plant-based protein and are widely utilised for various agricultural and industrial purposes. One notable characteristic of soybeans is the diversity in their seed coat colour, which ranges from yellow, black, brown and green to bicoloured variants. This study aimed to identify key genes associated with seed quality by analysing transcriptome profiles of differentially expressed genes in seeds, naked seeds and seed coats of black and white soybean genotypes. The analysis revealed that several upregulated genes are involved in hormone signalling pathways and metabolic processes, such as lysine, starch, sucrose, protein and galactose metabolism. These genes also participate in biosynthetic pathways for ethylene, lipids, brassinosteroids, lignin and sulfur-containing amino acids. Such molecular activities are likely linked to the enhanced seed quality observed in black-coated cultivars, which exhibit greater longevity, improved resistance to ageing, moisture and physical stress. Furthermore, the identification of key transcription factors provided deeper insight into the regulatory mechanisms underlying these traits. This research offers a comprehensive understanding of the genomic and metabolic pathways that influence seed quality in soybeans and lays a foundation for future gene-silencing studies to further explore the biological significance of black seed coat characteristics.

Keywords: lignin; lysine; protein; seed coat colour; soyabean; sucrose; transcriptome analysis

Introduction

The success of the plant kingdom largely comes from the evolution and diversification of angiosperms. Angiosperms, with around 300000 species of flowering plants, form the largest and most diverse group in the plant kingdom (1, 2). A key factor in their success is the characteristics of their seed coats. The seed coat regulates gas exchange between the embryo and the environment and protects the embryo from mechanical damage, pests and diseases (3, 4). A more compact seed coat offers better protection, highlighting the importance of its physical structure and chemical composition for seed longevity (5). The seed coat, or testa, is the outer layer of a mature seed, covering the embryo and nutritive tissue (4). During seed development, the outer integument forms various layers that become the testa, while the inner integument usually disappears (6, 7). The colour of the soybean seed coat is an important agronomic trait that determines seed quality and is also an evolutionary trait (5). Differences in the structure and composition of these layers vary by species or variety. The chemical composition of the seed coat also varies between genotypes, affecting its permeability (8). The presence of tannins and lignins in cuticle and macrosclereids of the seed coat is a major contributing factor to seed coat strength, hardness and impermeability, which showed the association of seed coat colour with water absorption and lower imbibition of dark seed coat than light ones (9, 10). Pigments in the seed coat resulting from the production of phenolic compounds (e.g. tannins) are mostly associated with antioxidant content and defence activity against pathogens (11, 12). The proanthocyanidins accumulation and oxidation in the seed coat (mainly in the endothelial layer indirectly help in the desiccation of seeds and thus prolong longevity (13, 14). Thus, the seed coats' chemical composition also has a critical role in deciding seed longevity. The formation of black or brown-colored seeds is primarily due to the accumulation of metabolic products such as flavonoids and anthocyanins within the epidermal layer of the seed coat (15). Environmental conditions that influence seed longevity are relative humidity, moisture content, oxygen pressure and temperature of storage. The key player in regulating all these environmental conditions is the seed coat, which acts as an interface to seed-environment relationships. Seed coats are found to impart both physical and chemical resistance to seeds by preventing mechanical injury and free radical damage (8).

Despite the variety of soybean cultivars available, variability persists within the species across several traits, notably seed coat colour, which ranges from yellow, green, black and brown to bicolour. The development of the soybean seed coat involves various biochemical pathways, including those for pigment biosynthesis, cell wall formation and structural integrity (5). The natural products causing differences in seed coat and hilum colour include flavonoids and anthocyanins (16, 17). Flavonoids are aromatic molecules derived from phenylalanine and acetyl-CoA

through the fatty acid pathway. The presence of two anthocyanidin glycosides, cyanidin-3-monoglucoside and delphinidin-3-monoglucoside, is primarily responsible for the black seed coats in soybeans (18, 19). It was reported that when the expression of both the ANR1 (proanthocyanidin reductase 1) and ANR2 (proanthocyanidin reductase 2) genes was inhibited, the seed coat of soybeans showed a distinctive red-brown colour (20).

A unique class of secondary metabolites found in plants, isoflavones are primarily generated by soybeans and other members of the Papilionoideae family via the phenylpropanoid pathway. Dietary isoflavones such as genistein, glycitein and daidzein are mostly found in soybean seeds (21). The enzyme phenylalanine ammonia lyase (PAL) converts l-phenylalanine, the starting point for the production of isoflavones, into cinnamic acid. Then, 4-hydroxylase (C4H) and 4-coumarate CoA ligase (4CL) transform cinamic acid into p-coumaryol CoA. A crucial enzyme in the biosynthesis of isoflavones, chalcone synthase (CHS), changes p-coumaryl-CoA into naringenin chalcone (17). CHS7 and CHS8 are seed-specific in soybeans and catalyse the conversion of p-coumaryl-CoA into naringenin chalcone (22). Chalcone isomerase (CHI) and chalcone reductase (CHR) are also crucial enzymes for isoflavone synthesis (23). Isoflavone synthase, or 2-hydroxyisoflavanone synthase, is the key enzyme that distinguishes between plants that produce isoflavones and those that do not. Microsomal cytochrome P_{450} monooxygenase, or isoflavone synthase, is responsible for the following hydroxylation of the generated C-2 radical after flavanones migrate through a 2, 3 aryl ring migration to their corresponding flavones (24, 25). Two IFS genes (IFS1 and IFS2) with a 14 amino acid difference have been found in the soybean genome (25-27). The naringenin and liquiritigenin flavonones are converted to their equivalent isoflavones by both isoforms of IFS (25–28). Different soybean cultivars accumulate isoflavones differently, which can be attributed to the combination of genetic and environmental variables whose regulation is still unknown. Similar to other plant secondary metabolic pathways, the isoflavone biosynthesis pathway is divided into multiple branches that utilise common substrates (17). This intricacy poses difficulties for isoflavone biosynthesis metabolic engineering. Different genes control each pathway branch and stage, which affects how different substances build up over time (19).

The colouration of soybean seed coats is primarily controlled by five classic genetic loci: I, R, T, W1 and O (29, 30). The synthesis of seed coat pigments, which largely determines seed coat colour, is regulated by three independent loci: I, R and T (31, 32). The R and T loci dictate the type of anthocyanin and proanthocyanidin synthesised, which results in specific seed coat colours: black (R, T), imperfect black (R, t), brown (r, T) and buff (r, t). The T locus encodes a flavonoid 3′-hydroxylase (F3′H) responsible for synthesising cyanidin-based anthocyanins and proanthocyanidins (16, 33, 34). Unlike the R and T loci, the I locus (inhibitor) with its four alleles (I, ii, ik, i) determines the distribution of pigments in the seed coats' epidermal layer. The I allele inhibits pigment production, resulting in yellow seeds, while the i allele allows full pigmentation, leading to completely colored seeds (10, 35-39). The dominance hierarchy among these alleles is I > ii > ik > i. Inhibition of seed coat pigmentation by the I locus, especially the I and ii alleles, occurs through RNA silencing of chalcone synthase (CHS) genes (39, 40).

Seed coat colour can indicate traits with commercial value, providing insights for genetic improvement or product development. Black-seeded soybeans exhibit superior physiological quality and storage longevity compared to white-seeded ones. The complex trait of soybean seed coat pigmentation involves multiple interacting loci, with mechanisms such as post-transcriptional gene silencing further regulating this trait. Despite current knowledge, the full range of beneficial molecules associated with black-seeded soybeans requires extensive exploration, presenting opportunities for crop improvement and value-added product development. Therefore, a comprehensive transcriptome analysis was carried out to identify differentially expressed genes in seeds, naked seeds and seed coats of black and white-seeded soybean genotypes. By examining genes involved in metabolic pathways, hormone signalling, pigment biosynthesis and transcriptional regulation, the study seeks to uncover key genetic determinants linked to seed coat composition, structural integrity and longevity. Furthermore, this investigation intends to provide insights into regulatory networks governing seed quality traits and to establish a molecular foundation for future functional genomics and gene-silencing studies targeting the improvement of soybean seed quality and storage performance.

Materials and Methods

Seed growth and development

To explore the differences between the transcriptomes of naked seeds and seed coats with contrasting tegument colouration, two cultivars were selected: the cultivar EC993950 representing the genotype of black seed coat colouration and the cultivar JS- 335 representing the white seed coat colour genotype. These seed varieties were chosen for their contrasting pigmentation colourations available in Karnataka, India. The matured stored seeds were imbibed in distilled water and only the germinated seeds were further taken for the RNA isolation.

RNA isolation

A Trizol reagent was used to isolate the RNA. The samples of whole seed, naked seed (seed pulp without the seed coat) and seed coat made up about 100 mg of preserved (-80 °C) tissue. The tissue was then immediately immersed in liquid nitrogen and pulverised to a fine consistency. After adding 1 mL of Trizol reagent to a 2 mL Eppendorf tube containing tissue powder, the tube was allowed to stand at room temperature for five minutes. After pipetting it well, it was placed on ice for fifteen minutes. Liquid phase was transferred to a new Eppendorf tube after centrifuging at 10000 rpm for five minutes at 4 °C. After adding 200 µL of cooled chloroform and vortexing for 30 sec, the mixture was centrifuged at 13000 rpm for 15 min at 4 °C. Aqueous phase was pipetted into a fresh Eppendorf tube, 200 µL of chloroform was added and the tube was centrifuged for 15 min at 4°C at 13000 rpm. Following pipetting of the aqueous phase into a 1.5 mL Eppendorf tube, 500 µL of ice-cold isopropanol was added and the mixture was incubated at -20°C for an entire night. The pellet was centrifuged for 15 min at 4°C at 13000 rpm and after that, it was air dried for 30 min on ice and twice cleaned with 70 % ethanol. In the end, the pellet was dissolved in 50 µL of nuclease-free water. Using an RNA HS test kit (Thermofisher #Q32851) and a Qubit 4.0 fluorometer (Thermofisher #Q33238), the extracted RNAs' quality was confirmed in accordance with the manufacturer's instructions. On a 1 % agarose gel, ethidium bromide staining was used to evaluate the RNAs' integrity. An additional assessment of the samples'

purity and abundance was conducted with a Thermo Scientific, Wilmington, DE, Nano DropTM1000 spectrophotometer.

RNA sequencing and quantification

The library preparation for RNA-Seq was carried out at Molsys Scientific, Bangalore, India, according to Illumina HiSeq sequencer RNA Library Prep. The samples for RNA sequencing, 0.1 - 1 µg high-quality total RNA, were used for TruSeq standard total RNA-Illumina # 2002059. The library was carefully checked before the sequencing using insert size and Qubit concentrations. The concentration of the insert size was checked in a bioanalyser for its size. The concentration of qubits is determined internally to calculate the data output for the Novaseq6000 sequencer, according to the machine manufacturer and internal standardisation. We normalised libraries for a concentration of 300 µmol and the concentrations varied further based on the requirement of data output from each sample when mixed with other barcoded samples. The Illumina NOVASEQ 6000 sequencer was employed to generate 150 bp paired-end (PE) data.

Quality assessment of the raw fastq reads of the sample was performed. The raw fastq reads were preprocessed using FastQC (FastQC v.0.11.9) (41). The processed reads were aligned to the Silva database using Bowtie2 (Bowtie2 v2.4.5) to filter rRNA reads. The non-rRNA reads are aligned to the STAR-indexed *G. max* (soybean) (GCA_000004515) genome (STAR aligner v 2.7.9a) using parameters:- outSAMtype BAM Sorted By Coordina - outSAMunmapped Within -quantMode TranscriptomeSAM outFilterScoreMinOverLread 0.33 -outFilterMatchNminOverLread 0.33 -outSAMattributes Standard). The rRNA and tRNA features were removed from the GTF file of *G. max* (42). The alignment file (sorted BAM) from individual samples was quantified using feature Counts, based on the filtered GTF file, to obtain gene counts (43). The 'regularised log' transformation in DESeq2 (transcript counts for *G. max* were used as inputs to DESeq2) was used for principal component and clustering analysis. Normalisation and differential expression analysis were performed using DESeq2 (Fig.1).

Gene annotation

The geneIDs obtained from DESeq2 were submitted to bioDBnet for annotation, specifying taxon: id as '000004515' (*Glycine max*). Nucleotide sequences were then retrieved for geneIDs lacking annotations via bioDBnet. These sequences were subsequently subjected to the KAAS KEGG Automatic Annotation Server, which assigned KO IDs (KEGG Orthology Identifiers) based on the best hits for each gene. These KO IDs were further processed through GAEV (Gene Annotation Easy Viewer) to compile information such as gene names, KO (KEGG Orthology) numbers, functional definitions of orthologs and functional pathways associated with the queried genes. In cases where gene sequences remained unannotated by KAAS, their protein sequences were extracted and annotated using Blast KOALA. Lastly, for Gene Ontology (GO) annotation, complete protein sequences corresponding to the gene IDs were retrieved and analysed with PANNZER2 (organism: *Glycine max*) (Fig.1).

The DESeq2 result files were filtered based on "Un-adjusted *p*-value <= 0.05 and Log Fold Change ∓ 2". The volcano plots were generated using Enhanced Volcano and the MA plots were plotted using the ggmaplot function of the ggpubr R package (44, 45). The significant genes were subjected to gene ontology (GO) and KEGG pathway enrichment analysis using the ShinyGO software based on the *G. max* genome as a model, with a significance level of 0.05, with false discovery rate (FDR) as an adjustment method. FPKM calculations removed variability in gene length and overall read distribution in terms of gene expression levels. With the statistically significant criterion of $p < 0.05$, differentially expressed genes were identified by a percentage of log2 fold change≥1,i.e., the change in gene expression between the treated and control groups was greater than 2-fold. The overview of the workflow of the methodology from raw data to gene annotation is represented in Fig. 1. All the QC reports, DESEQ2 cut off, functional data and analysis reports are available in the open-source platform https://osf.io/p6us5/overview?view_only=2fb4f6c3c9ac4620be3edad252113598.

Statistical analysis

All data sets had three biological replicates. Differentially expressed genes were defined as genes with a false discovery rate (FDR) < 0.05 and a fold change > two-fold or greater. An adjusted *p-value* < 0.05 was considered significant when identifying enriched GO terms and an adjusted *p-value* < 0.05 was considered indicative of significantly enriched KEGG pathways.

To know the differences between the seed coats with contrasting tegument colouration, two cultivars were selected, the cultivar EC993950 representing the genotype of black seed coat colouration and the cultivar JS- 335 representing white seed coat colour. The paired-end library type was used for an overview of twelve samples using reference-based whole transcriptome analysis (Table 1). Four comparisons were made with the help of four samples, that is, seed coat- white, seed coat- black, naked seed- white, naked seed- black, each sample with three replications, for a total of twelve samples, which is represented in Table 2. Four comparisons include: comparison 1: seed coat black vs seed coat white of soybean, comparison 2: naked seed black vs naked seed white of soybean, comparison 3: naked seed white vs seed coat white of soybean and comparison 4: naked seed black vs seed coat black of soybean (Table 2).

Due to the low alignment rate, S_4, which is 0.02 % with 2964 counts and includes high unmapped reads of 99.95 % with

Table 1. Sample overview

Library type	Number of samples	Approach used
Paired-end	12	Reference-based whole transcriptome analysis

Table 2. Grouping information

Comparison 1: seed coat black vs seed coat white of soybean		**Comparison 2:** naked seed black vs naked seed white of soybean		**Comparison 3:** naked seed white vs seed coat white of soybean		**Comparison 4:** naked seed black vs the seed coat black of soybean	
Samples	**Group**	**Samples**	**Group**	**Samples**	**Group**	**Samples**	**Group**
S_1	Seed coat- white	S_4	Naked seed- white	S_1	Seed coat- white	S_7	Seed coat- black
S_2	Seed coat- white	S_5	Naked seed- white	S_2	Seed coat- white	S_8	Seed coat- black
S_3	Seed coat- white	S_6	Naked seed- white	S_3	Seed coat- white	S_9	Seed coat- black
S_7	Seed coat- black	S_{10}	Naked seed- black	S_4	Naked seed- white	S_{10}	Naked seed- black
S_8	Seed coat- black	S_{11}	Naked seed- black	S_5	Naked seed- white	S_{11}	Naked seed- black
S_9	Seed coat- black	S_{12}	Naked seed- black	S_6	Naked seed- white	S_{12}	Naked seed- black

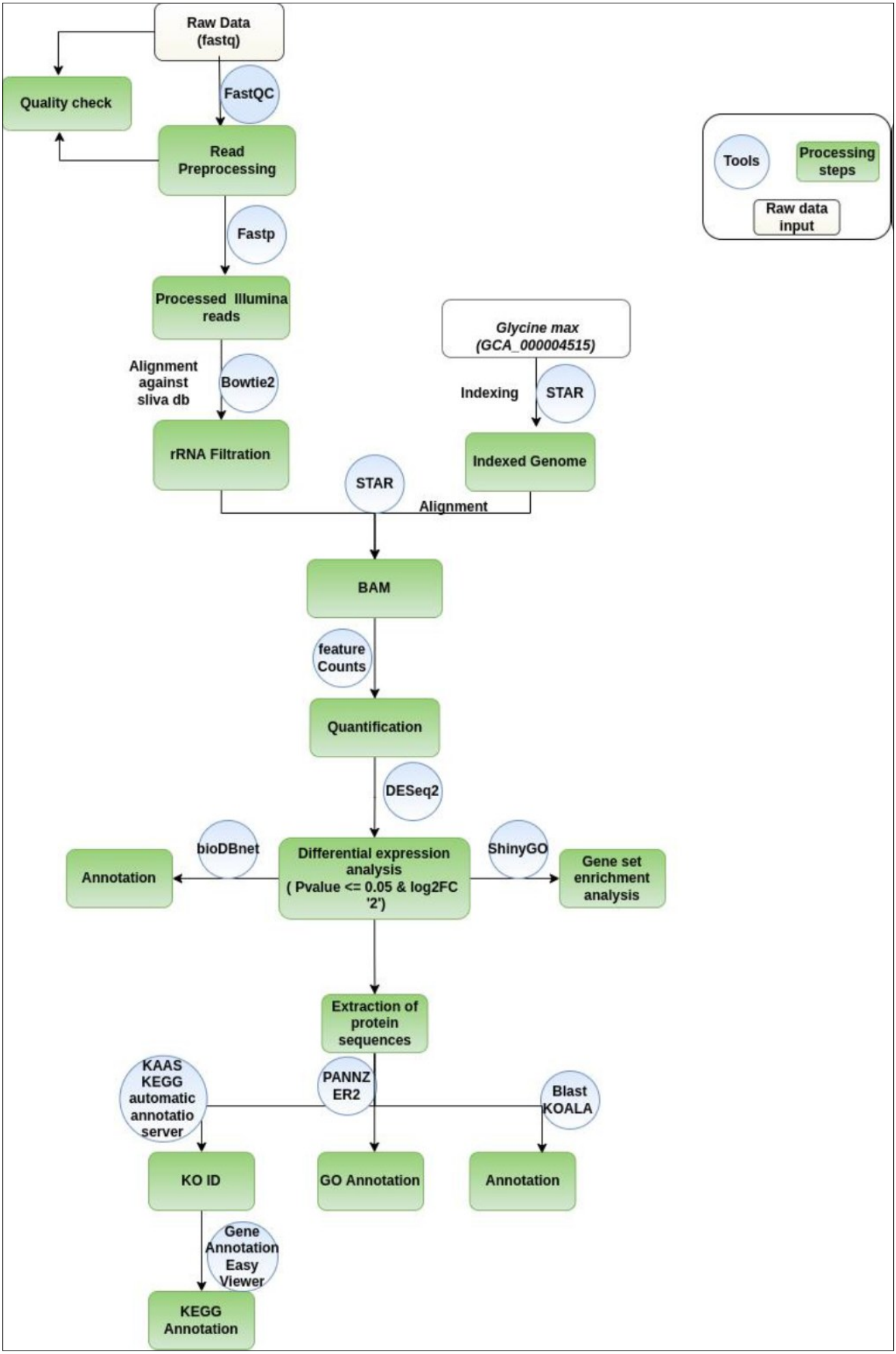

Fig. 1. Workflow of methodology.

Table 3. Alignment statistics of processed samples against the SILVA database v138 for rRNA filtration using Bowtie2 and rRNA filtered reads against the reference genome (*Glycine max*) using the STAR aligner

Sr. No	Sample Name	Aligned concordantly exactly 1 time		Uniquely mapped reads number		Number of reads unmapped: too short	
		Count	%	Count	%	Count	%
1	S_1	701671	4.95	1604897	26.41	4060328	66.82
2	S_2	858240	4.12	2481185	27.5	6030707	66.85
3	S_3	1535925	4.82	4088324	26.87	9780424	64.27
4	S_4	2964	0.02	4358	0.03	14627725	99.95
5	S_5	463858	2.48	2930143	23.42	8624795	68.94
6	S_6	576767	3.07	2663988	24.64	7432304	68.74
7	S_7	237516	2.67	474437	8.3	5122033	89.58
8	S_8	975815	4.66	2191450	25.42	5829744	67.63
9	S_9	1088519	4.55	2866458	24.29	7611891	64.5
10	S_{10}	1096862	4.53	2860014	25.44	7555887	67.21
11	S_{11}	1470907	4.94	3210424	25.9	8308577	67.03
12	S_{12}	711195	2.72	4222368	24.38	11816673	68.24

14627725 counts (Table 3), S_4 was not included in the differential expression analysis.

Note: "too short" means that the best alignments STAR found were too short to pass the filters. This is controlled by our filter score min over L read, which was set to 0.5 which means that half of the total read length (sum of mates) should be mapped.

Results and Discussion

Comparison 1 (seed coat black vs seed coat white)

Comparative transcriptomic profiling of seed coat black vs seed coat white

258 genes were found to be differentially expressed in the black seed coats of the cultivar EC993950 when compared to the white seed coat cultivar (JS-335). In comparison to JS-335, EC993950 showed 206 upregulated genes and 52 downregulated genes out of 258 total genes (Table 4). Principal component analysis (PCA) revealed a significant degree of variation between the black and white seed coats. The explained dispersion percentage for the PCA model was 72 %, as displayed in Fig. 2. This model uses regularised log count data and a plot that illustrates variance both within and across groups to explain data variability. The two main components that account for the largest percentage of variation, 46 % for PC1 and 26 % for PC2, are displayed on the horizontal and vertical axes. The distribution of gene expression between the groups, seed coat black and seed coat white, is displayed by the MA-plot. The log of the mean of the samples' normalised expression counts is represented by the X axis, while the Y axis displays the Log2 fold change. Based on the *p*-value (<0.05), red dots indicate genes that are up-regulated (> +2) and blue dots indicate genes that are down-regulated (< -2). A grey dot indicates non-significant genes with a p-value greater than 0.05. It was evident from Fig. 3 that more genes were elevated in the black seed coat than in the white seed coat. The heatmap of regularised log-transformed data displays the profile of differentially expressed transcripts, highlighting the top 50 genes with the most variance across samples. In a similar vein, the volcano plot revealed the expression of numerous DEGs, of which 19 were up-regulated and 7 were down-regulated (Fig. 4). This suggests that the contrasted seed coat colour is influenced by the expression of particular genes.

Characterisation and functional annotation of DEGs in the black seed coat and white seed coat

DEGs in the black and white seed coats were used to map the GO database using the ShinyGO 0.76 toolkit, with a false discovery rate (FDR of 0.05) set as the cutoff. This allowed us to look at the significantly enriched genes in comparison to the genomic background. The results showed that 34 significant GO keywords were identified between the two gene ontologies in the up-regulated DEGs of a black seed coat in comparison to the white seed coat. L-lysine catabolic process to acetyl-CoA, L-lysine metabolic process, lysine catabolic process, lysine biosynthetic process via aminoadipic acid, photosynthetic electron transport in photosystem ii, response to chitin, response to water, response to acid chemical, protein complex oligomerisation, response to heat, response to temperature stimulus, response to salt stress, response to inorganic substance, protein folding, response to oxygen-containing compound, cellular carbohydrate metabolic process, response to abiotic stimulus were the main classes that contributed to the biological process (Fig. 5.).

Electron transporter, transferring electrons within the cyclic electron transport, misfolded protein binding, protein folding chaperone, protein self-association, oxidoreductase activity, acting on single donors with incorporation of molecules, unfolded protein binding, hydrolase activity, hydrolysing O-glycosyl compounds was the primary class under the molecular-function category. which were displayed in Fig. 6. The major molecular networks connected to every gene that was found to be considerably changed seem to be connected to important mediators that differ in seed coat colour (Fig. 5-6).

Functional regulatory network analysis (KEGG pathway enrichment) of black seed coat and white seed coat

The distinct metabolic processes involved in the development of the soybean seed coat were elucidated by searching the KEGG pathway database for DEGs that were specifically identified in the black and white seed coats. Alpha-linolenic acid metabolism, ubiquitin-mediated proteolysis, starch and sucrose metabolism, protein processing in the endoplasmic reticulum and biosynthesis

Table 4. Up and down-regulated gene count based on *p* value < 0.05 and Log2FoldChange ± 2 using DESeq2

Comparisons	*p* value <=0.05 and log2FC ∓ '2'	
	Up	Down
Comparison 1 (seed coat black vs seed coat white)	206	52
Comparison 2 (naked seed black vs naked seed white)	59	85
Comparison 3 (naked seed white vs seed coat white)	396	234
Comparison 4 (naked seed black vs seed coat black)	302	608

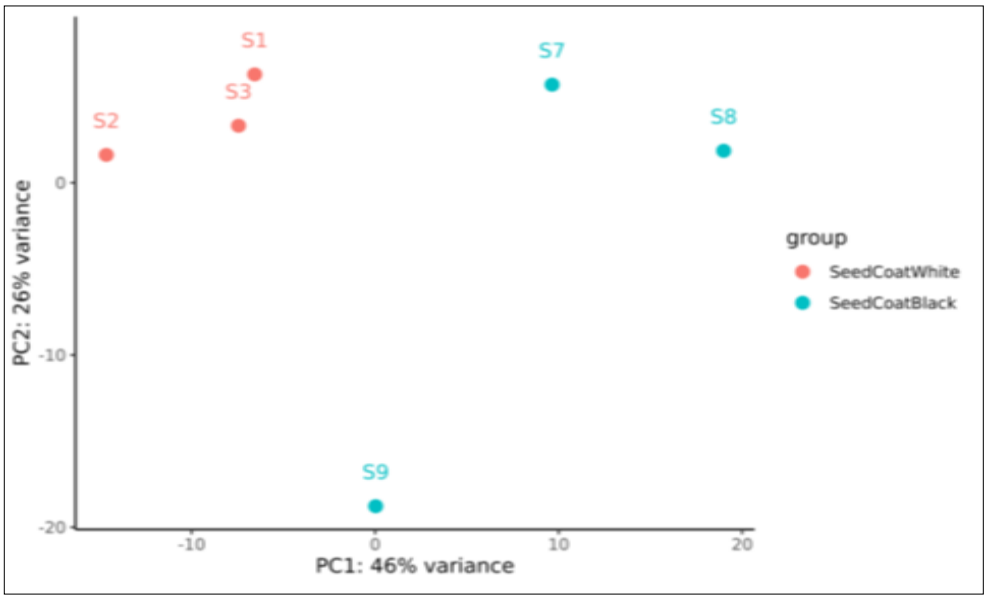

Fig. 2. Principal component analysis (PCA) plot of comparison 1.

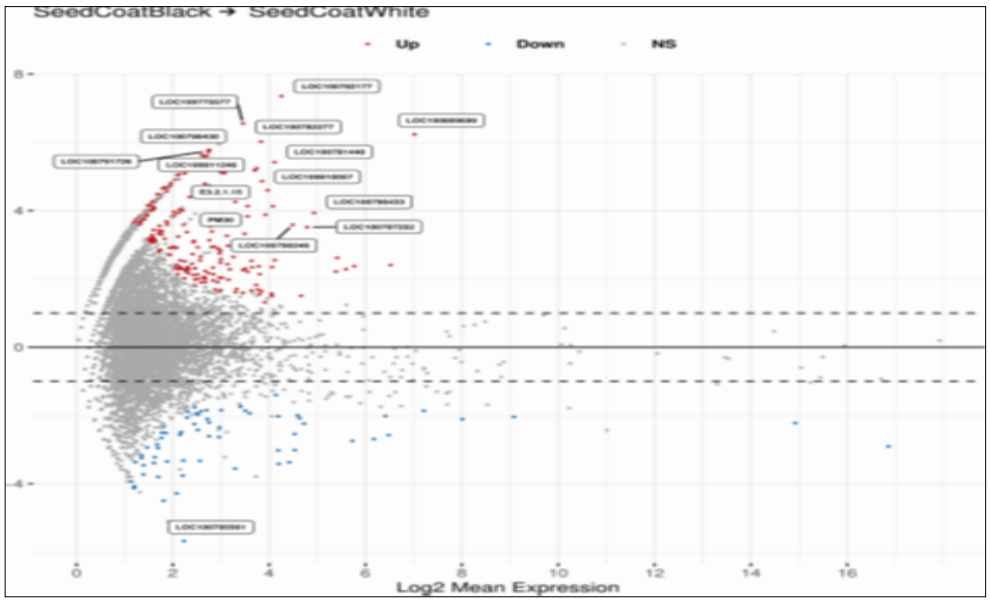

Fig. 3. MA plot of comparison 1.

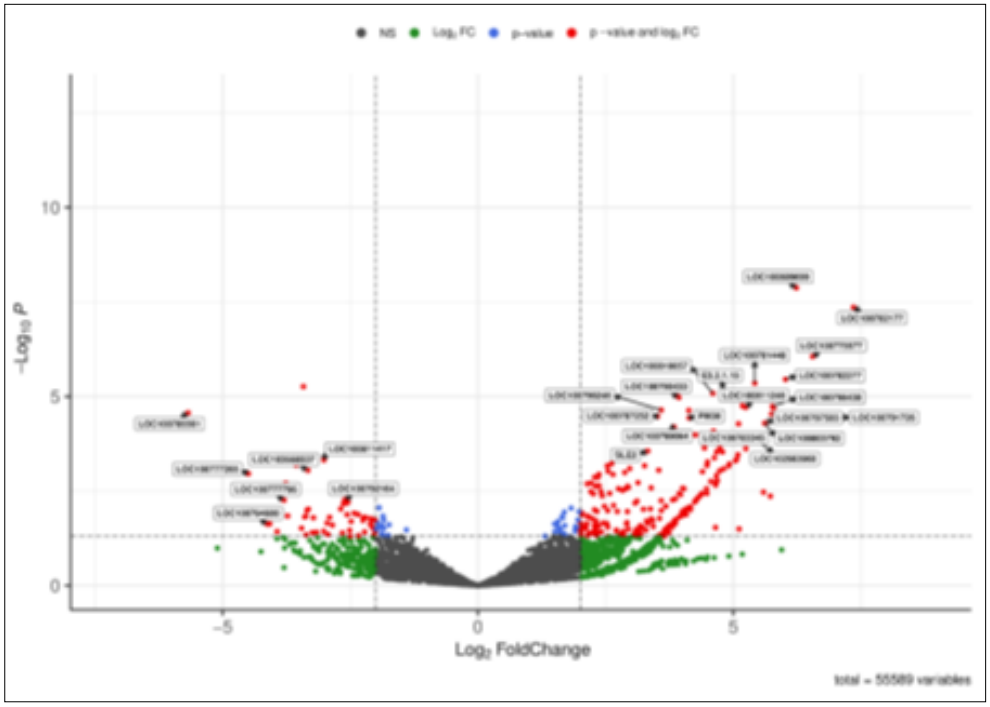

Fig. 4. Volcano plot of comparison 1.

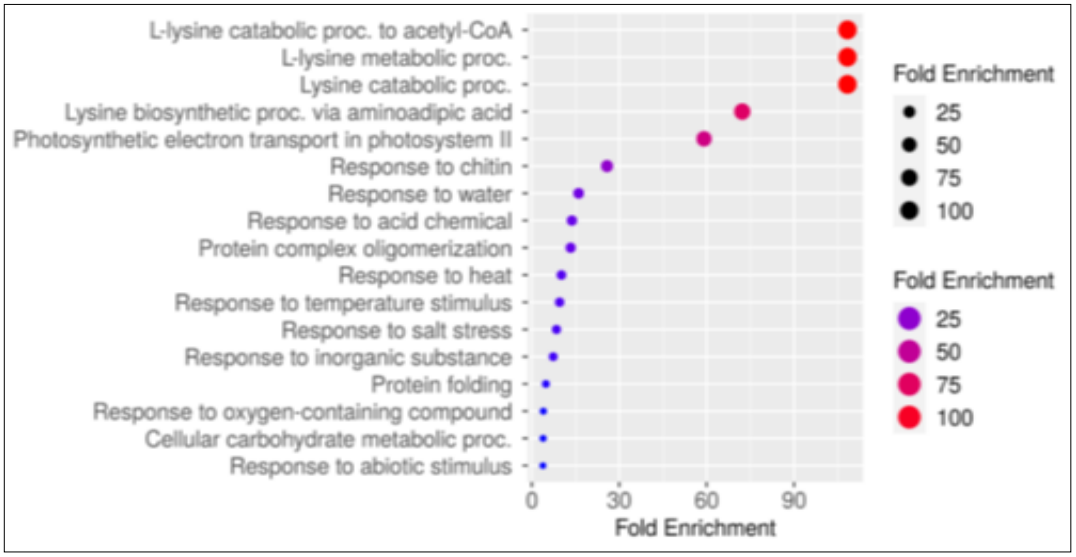

Fig. 5. Functional enrichment of the top-most significant (FDR<0.05) categories of biological processes (GO terms) for comparison 1.

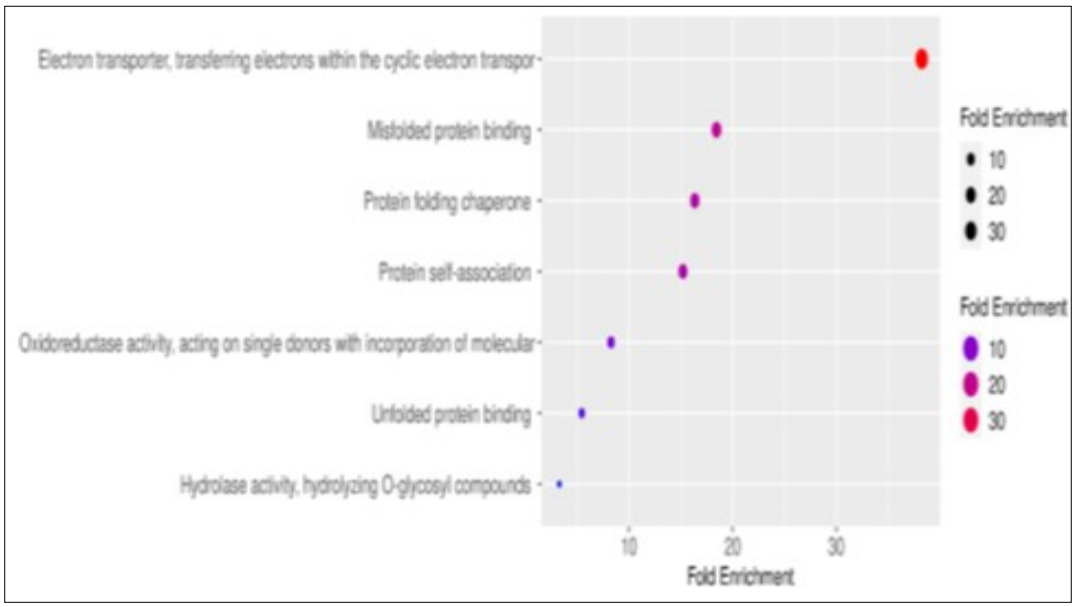

Fig. 6. Functional enrichment of the top-most significant (FDR<0.05) categories of molecular function (GO terms) for comparison 1.

of secondary metabolites are among the major mechanisms enhanced in the black seed coat (Fig. 7).

Identification of transcription factors in black seed coat and white seed coat

Furthermore, the DEGs encode the transcription factor (TF). Several transcription factors were discovered to control the colour of the soybean seed coat. These include the following. LOC100782177, LOC100775577, LOC100791726, LOC100798430, LOC100811246, PM30 (Seed maturation protein), LOC100790246, LOC100787252, LOC100790433, LOC100810057, LOC100781448, LOC100782277, LOC100809698, LOC100809699, LOC100787252, SLE2 (Protein SLE2, Transcriptional regulator WhiB), LOC100783345, LOC100803792, LOC102663960 and LOC100782177 were among the TFs that were upregulated in the black seed coat. LOC100785561, LOC100811417, LOC100792164, LOC102668537, LOC100777796, LOC100794600 and LOC100777265 were the downregulated transcription factors in the black seed coat.

Overview of metabolism

The elevated expression of lysine pathway genes may also contribute to improved seed vigour and longevity. Lysine-rich storage proteins are known to support seed germination efficiency, stress tolerance and structural stability of storage reserves. Furthermore, lysine metabolism is closely linked with energy metabolism and nitrogen assimilation, which are essential for proper seed development and maturation. Therefore, the upregulation of lysine-related genes identified in the differential expression analysis provides a molecular basis for the superior seed quality, enhanced nutritional composition and improved stress resilience observed in black seed coat soybean cultivars (46).

Soybeans can become a more valuable and complete source of protein by increasing their lysine level, which will also improve their nutritional profile. The black seed coat exhibits a higher level of functional enrichment in the L-lysine catabolic pathway to acetyl-CoA, L-lysine metabolic pathway, L-lysine catabolic pathway and L-lysine synthesis pathway via aminoadipic acid genes. Beyond its nutritional importance, lysine is essential for enzyme function (47, 48). It is a component of many enzymes' structures and activities, including those of hydrolases, oxidoreductases, metabolism of alpha-linolenic acid (α-Linolenic acid metabolism pathways in soybean signal transduction from KEGG with query genes highlighted in red, which is represented in Fig. 8. Ubiquitin-mediated proteolysis and genes involved in the metabolism of starch and sucrose that are crucial for the growth and development of seeds (49). Lysine degradation and starch and sucrose metabolism pathways in soybean signal transduction from

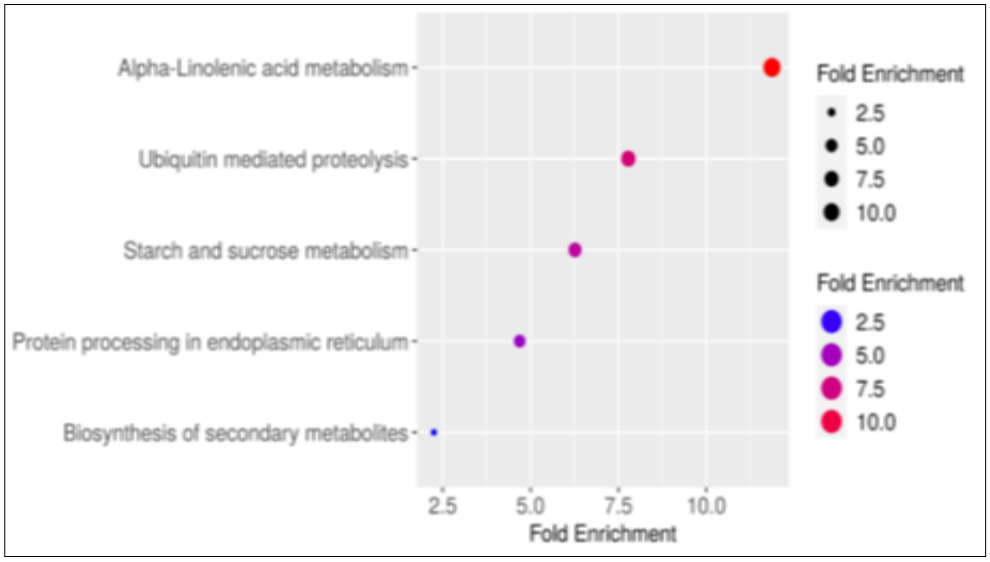

Fig. 7. Functional enrichment of the top-most significant (FDR<0.05) categories of KEGG terms for comparison 1.

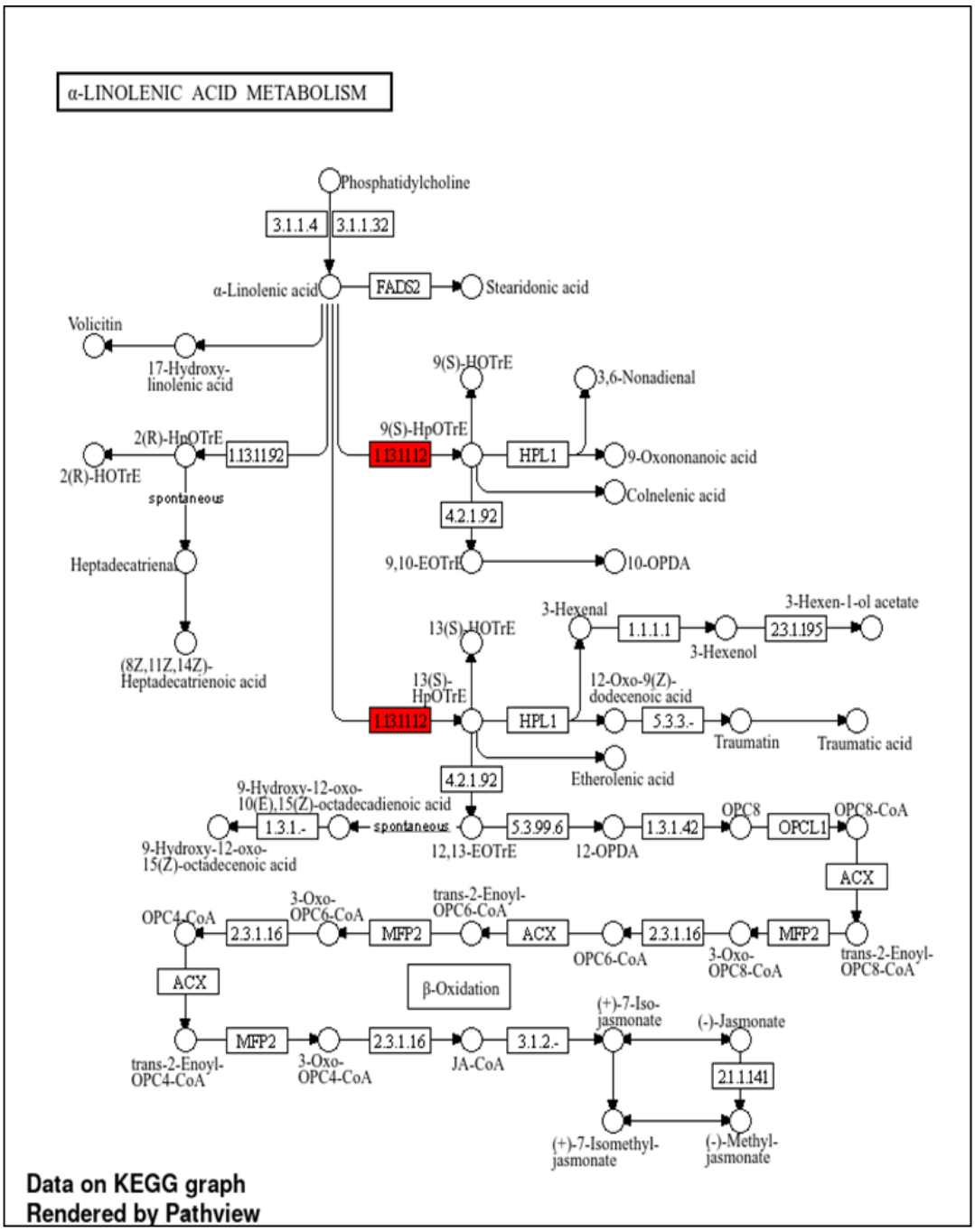

Fig. 8. α-Linolenic acid metabolism pathways in soybean signal transduction from KEGG with query genes highlighted in red.

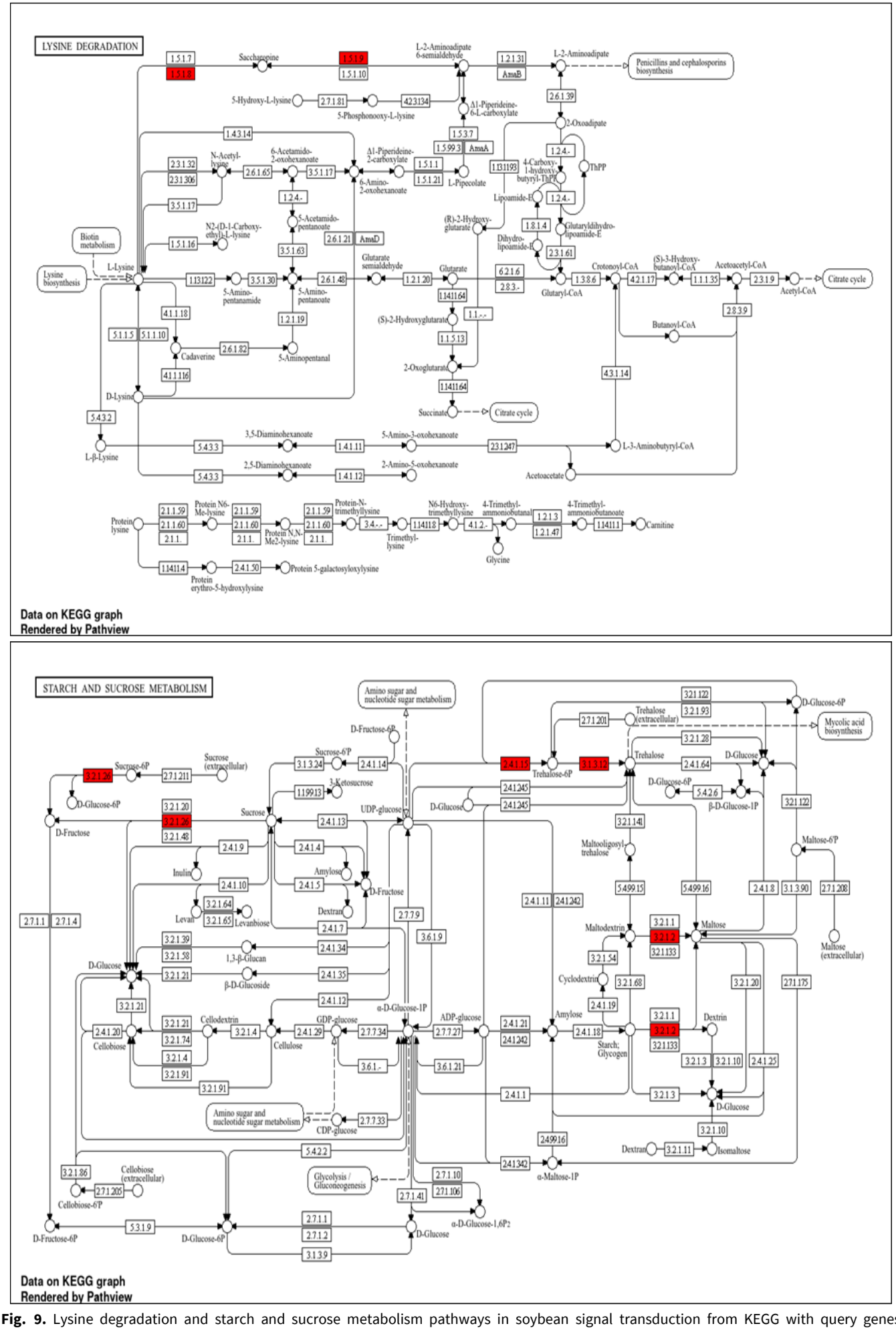

Fig. 9. Lysine degradation and starch and sucrose metabolism pathways in soybean signal transduction from KEGG with query genes highlighted in red.

KEGG, with query genes highlighted in red, are represented in Fig. 9.

It also functions as a precursor to numerous proteins, which is supported by the discovery that functional enrichment of cellular metabolic processes involving carbohydrates provides necessary energy and structural elements. Appropriate protein construction and function are guaranteed by protein complex oligomerisation and self-association (50, 51). By stopping misfolding and supporting proper protein structure, chaperones help with protein folding. Protein mistakes can be managed and corrected with the aid of unfolded and misfolded protein binding mechanisms (52, 53). Protein processing in the endoplasmic reticulum guarantees that proteins are appropriately modified, promoting efficient seed growth and germination (54).

Additionally, lysine has a role in the plants' stress response systems, which help the plant tolerate environmental stressors like disease, drought and pest infestations (47). Greater fold enrichment of genes in the black seed coat correlates with resistance. By inducing immunological responses, chitin-responsive genes aid in the defence against fungal infections (55). Osmotic equilibrium is regulated by genes that react to salt and water stress, preventing toxicity or dehydration. Stress tolerance and pH balance are regulated in response to acidic substances (56). Genes that respond to heat and temperature shield seeds from harm and control growth in temperature swings. Ion homeostasis and nutrient absorption are controlled by inorganic substance response genes. Oxygen-consuming compound response genes deal with stress caused by oxidation (57).

Black seed coats also have greater levels of secondary metabolite biosynthesis enrichment. These substances, which have antioxidant qualities and contribute to broad protection, are what give rich black colouring; they include anthocyanins, flavonoids and tannins. They also strengthen the resistance of the seed against pests and diseases (58, 59). Black soybeans are an excellent functional food because of these metabolites, which also have anti-inflammatory and anti-cancer characteristics.

Comparison 2 (naked seed black vs naked seed white)

Comparative transcriptomic profiling of seed coat black vs seed coat white

Comparing the naked white seed cultivar JS-335 to the naked black seed of cultivar EC993950, 144 genes were found to express differently. Comparing EC993950 to JS-335, 59 genes were upregulated and 85 genes were downregulated among the 144 genes (Table 4). There are significant differences between the black and white bare seeds, according to the principal component analysis (PCA). As seen in Fig. 10. which uses regularised log count data and a plot to show variation both within and across groups to explain data variability, the PCA models' percentage of explained dispersion was 80 %. The largest percentage of variation (46 % for PC1 and 34 % for PC2) may be explained by the two main components that are displayed on the horizontal and vertical axes.

The gene expression distribution between the groups of naked seed black and naked seed white is displayed using an MA-plot. The X axis displays the log of the samples' normalised expression count mean, while the Y axis displays the log of the fold change. Based on the *p*-value (<0.05), red dots represent up-regulated genes (> +2) and blue dots represent down-regulated genes (< -2). When the p-value is greater than 0.05, the grey dot represents the non-significant genes. It was evident that, in comparison to naked white seeds, more genes were elevated in naked black seeds (Fig. 11). The differentially expressed transcripts profile in the heatmap of regularised log transformed values showing the top 50 genes with the highest variance across samples. Similarly, the volcano plot showed that many DEGs were expressed, among which 9 DEGs were up-regulated and 11 were down-regulated, which depicts that the gene expression plays a role in the naked soybean seed of distinct seed coat colour (Fig. 12).

Characterisation and functional annotation of DEGs in the naked seed black and naked seed white

The significantly enriched genes compared with the genomic background were examined using DEGs in the black naked seed and white naked seed by mapping the GO database using the ShinyGO 0.76 toolkit and setting a false discovery rate (FDR = 0.05) as the cutoff. The results showed that in the two gene ontologies of the up-regulated DEGs of a black naked seed, a total of 21 significant GO keywords were identified when compared to the white naked seed.

Protein complex oligomerisation, response to hydrogen peroxide, response to reactive oxygen species, heat, response to salt stress, response to inorganic substance, response to temperature stimulus, neg. reg. of peptidase activity, protein folding, reg. of proteolysis, response to oxidative stress, fatty acid metabolic process, response to oxygen-containing compound, response to abiotic stimulus, protein-containing complex assembly, protein-containing complex subunit organisation, cellular component assembly, amide biosynthetic process are the main classes that contribute to the biological process (Fig. 13). The main class in the molecular-function category was protein self-association, unfolded protein binding (Fig. 14). Significantly differentially expressed genes were associated with major molecular networks involving key regulatory mediators in naked soybean seeds of contrasting seed coat colours.

Functional regulatory network analysis (KEGG pathway enrichment) of naked seed black and naked seed white

To shed light on the distinct metabolic processes involved in the establishment of the soybean seed coat, DEGs that were specifically found in the black and white naked seeds were then compared to the KEGG pathway database. Protein processing in the endoplasmic reticulum is one of the main pathways enhanced in the black naked seed (Fig. 15).

Identification of transcription factors in naked seed black and naked seed white

The DEGs encoding the transcription factor (TF) were also examined. Many transcription factors, such as the following, were discovered to control the colour of the seed coat on bare soybean seeds. The TFs that were upregulated in black naked seeds included BG7SNA2, LOC100526893, HSP17.5NAE, LOC100779800, LOC100500475, GY1/GY2, LOC100775665, LOC100816661 and LOC100790057. The LOC IDs represent the predicted gene loci annotated in the NCBI *G. max* genome database. These LOC IDs are yet to be assigned official gene symbols, but are functionally annotated based on sequence similarity. Many of these LOC genes correspond to predicted transcriptional regulators and stress-responsive proteins in *G. max*. For instance, LOC100526893 encodes a putative MYB-like transcription factor associated with the regulation of secondary

metabolism and stress responses. LOC100779800 and LOC100500475 are predicted regulatory proteins involved in transcriptional control and signal transduction pathways linked to seed development and metabolic activity. LOC100775665 and LOC100816661 are annotated as hypothetical or uncharacterised transcription-related proteins but are predicted to participate in gene expression regulation and stress adaptation. LOC100790057 is associated with regulatory protein activity and may contribute to metabolic and developmental processes during seed maturation. Heat shock protein gene HSP17.5NAE plays a crucial role in stress tolerance, protein stabilisation and seed longevity, which may contribute to the enhanced durability observed in black seed coats. In contrast, several transcription factors were downregulated in black naked seeds, including NAC22, NAC domain-containing proteins, LOC100793362, LOC100807861, LOC100806180, LOC100790683, LOC100788162, LOC100780168, LOC100812502, LOC100807214, LOC100794867, LOC100775287, LOC100775904 and LOC1007954. NAC transcription factors are well known for their roles in senescence, stress signalling and secondary wall biosynthesis. Downregulation of NAC22 and related NAC domain proteins suggests reduced activation of senescence-associated pathways and altered lignin or secondary metabolite regulation. Several LOC genes, including LOC100793362 and LOC100807861, are predicted transcriptional regulators involved in stress signalling and developmental processes, while LOC100806180 and LOC100790683 are associated with regulatory protein binding and transcriptional modulation. Other LOC genes represent uncharacterised transcription-related proteins that may participate in seed coat development, metabolic regulation and environmental stress responses. Overall, the differential regulation of these transcription factors indicates complex regulatory networks governing seed coat pigmentation, metabolic activity, stress tolerance and seed quality traits in black soybean genotypes.

Overview of metabolism

Seed degradation is largely caused by hydrogen peroxide and reactive oxygen species (ROS). These substances harm DNA, proteins and lipids within cells by oxidatively stressing them (60). The damage reduces the viability and vigour of seeds, which compromises seedling development and germination rates. The buildup of hydrogen peroxide and reactive oxygen species (ROS) during storage speeds up ageing processes and exacerbates the deterioration of seed quality (61). In black naked seed, there was a higher enrichment of genes responding to hydrogen peroxide and reactive oxygen species (ROS). To reduce this oxidative damage and preserve seed health, effective antioxidant defence systems are necessary. Thus, controlling ROS levels is essential to maintaining the longevity of seeds and guaranteeing successful germination (57, 62).

The naked black seed has higher expression of genes involved in the production of fatty acids. The malonyl-CoA/ methylmalonyl-CoA synthetase enzyme is responsible for catalysing the reaction that results in the malonyl-CoA product and the malonate substrate. One of the primary precursors for elongating and generating fatty acids is this substance (63). It also plays a crucial role in the synthesis of other substances such as phytoalexins, flavonoids and anthocyanins. The CHS enzyme, which condenses three molecules of malonyl-CoA with one molecule of p-coumaroyl-CoA, is the first enzyme in the route leading to the synthesis of flavonoids and isoflavonoids. Malonyl-CoA is one of the precursors required for this enzyme. This process produces the other flavonoids found in soybean seeds (59).

Since black seeds have more genes that are enriched in reactions to heat, salt stress, inorganic chemicals, temperature stimuli, oxidative stress, oxygen-containing compounds and other abiotic stimuli, they are more resistant to unfavourable conditions, which is important for maintaining seed viability. Heat and temperature responses shield cellular membranes and proteins from harm (64). Toxic effects are avoided by regulating ion homeostasis and responding to inorganic substances under salt stress. Cellular components are shielded from harm by oxidative stress reactions, which include detoxification of reactive oxygen species (65-67). Metabolism is regulated by reactions to substances that contain oxygen. When these adaptive responses work together, seeds may withstand a variety of environmental stressors, leading to effective maturation, dormancy and eventual germination in a variety of settings (31, 68).

Higher expression in black seed was seen for genes associated with the assembly of cellular components. A process called amide biosynthesis aids in the production of essential amino acids, which are needed for the creation of proteins. Premature protein degradation is prevented by the negative control of peptidase activity (69). Forming functional protein complexes required for different biological functions requires component organisation and protein-containing complex assembly. The correct three-dimensional structure of newly generated proteins is ensured by protein folding. By regulating the mechanisms involved in protein degradation, proteolysis regulation preserves protein stability (64, 70, 71). Furthermore, the process of protein processing within the endoplasmic reticulum guarantees that proteins are accurately altered and ready for their functions in seed development and growth, promoting effective germination and the establishment of seedlings.

Comparison 3 (naked seed white vs seed coat white)

Comparative transcriptomic profiling of naked seed white vs seed coat white

A total 630 genes were found to be differently expressed in the white naked seed of the cultivar JS-335 as compared to the white seed coat cultivar. In comparison to the JS-335 seed coat, 234 genes were downregulated and 396 genes were upregulated among 630 genes in the white naked seed (Table 4). The white seed coat and white naked seed differed greatly, as the PCA revealed. The regularised log count data and plot illustrating variance within and across groups are used to explain the variability in the data where the PCA models' percentage of explained dispersion was 91 %, as demonstrated in Fig. 16. The two principal components that account for the largest percentage of variation (81 % for PC1 and 10 % for PC2) are displayed on the horizontal and vertical axes. The distribution of gene expression between the white bare seed and white seed coat groups is displayed by the MA-plot. The log of the mean of the samples' normalised expression counts is represented by the X axis, while the Y axis displays the Log2 fold change. Based on the *p*-value (<0.05), red dots indicate genes that are up-regulated (> +2) and blue dots indicate genes that are down-regulated (< -2). A grey dot indicates non-significant genes with a *p*-value greater than 0.05. A higher number of genes were elevated in the white bare seed as

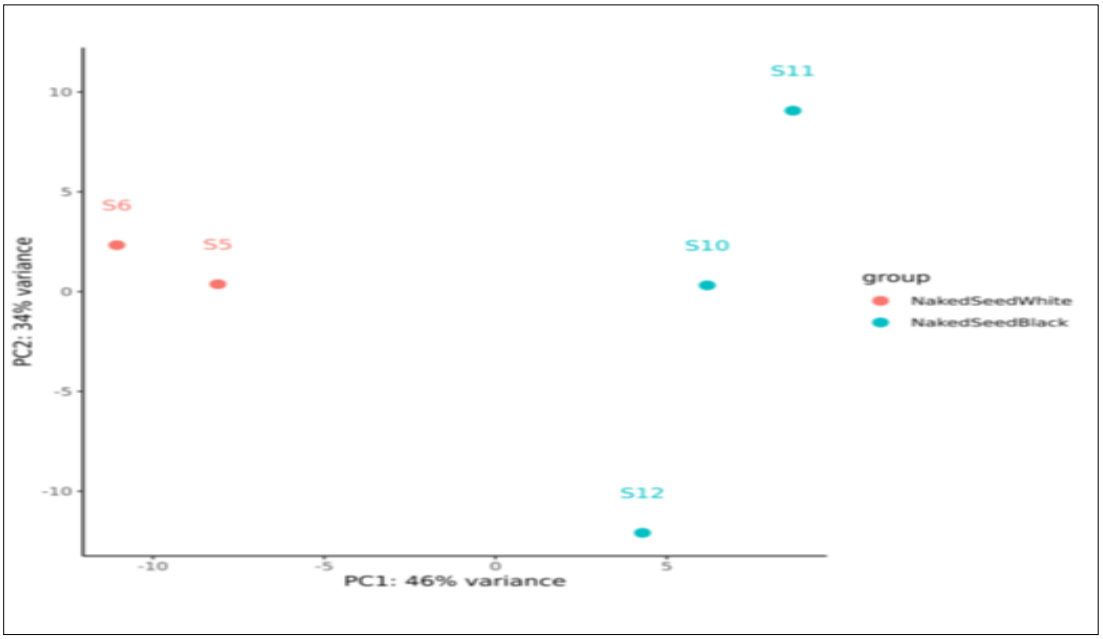

Fig. 10. Principal component analysis (PCA) plot of comparison 2.

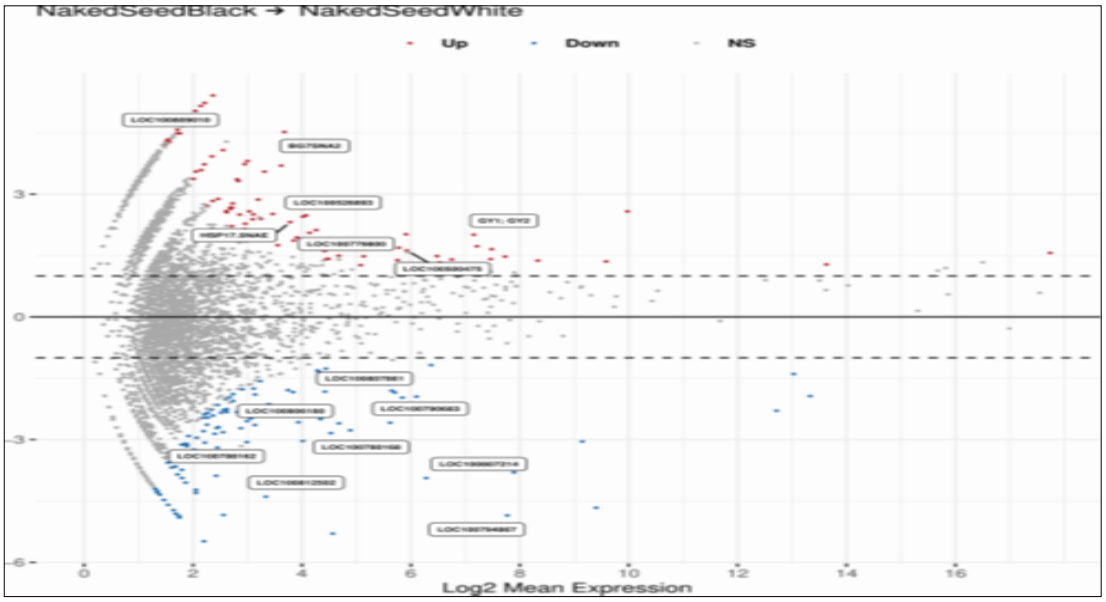

Fig. 11. MA plot of comparison 2.

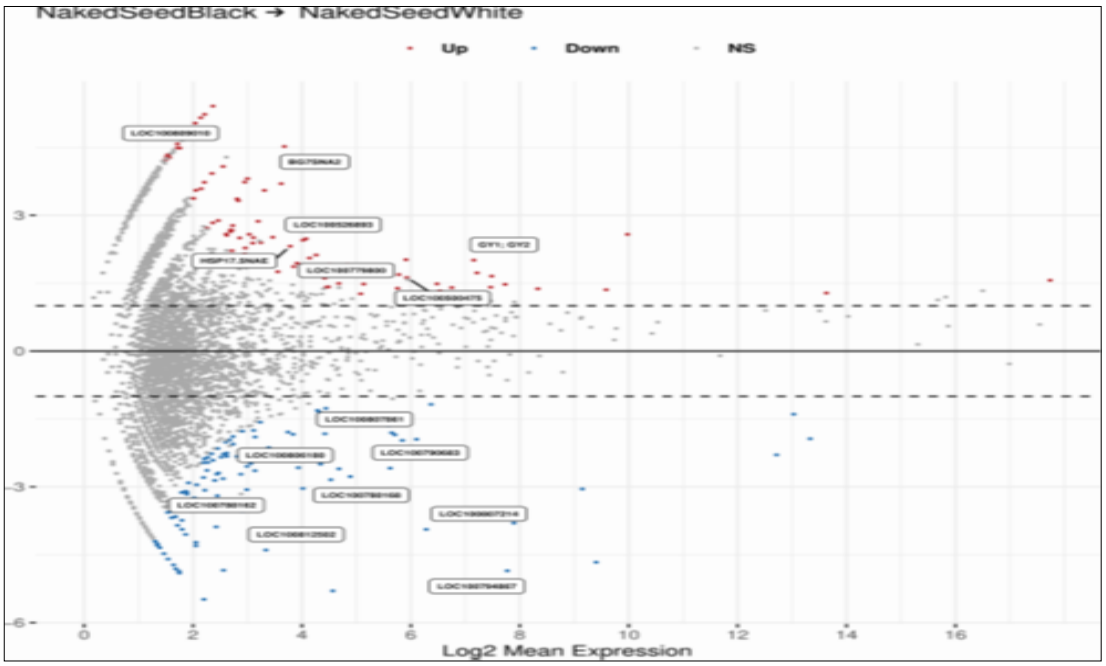

Fig. 12. Volcano plot of comparison 2.

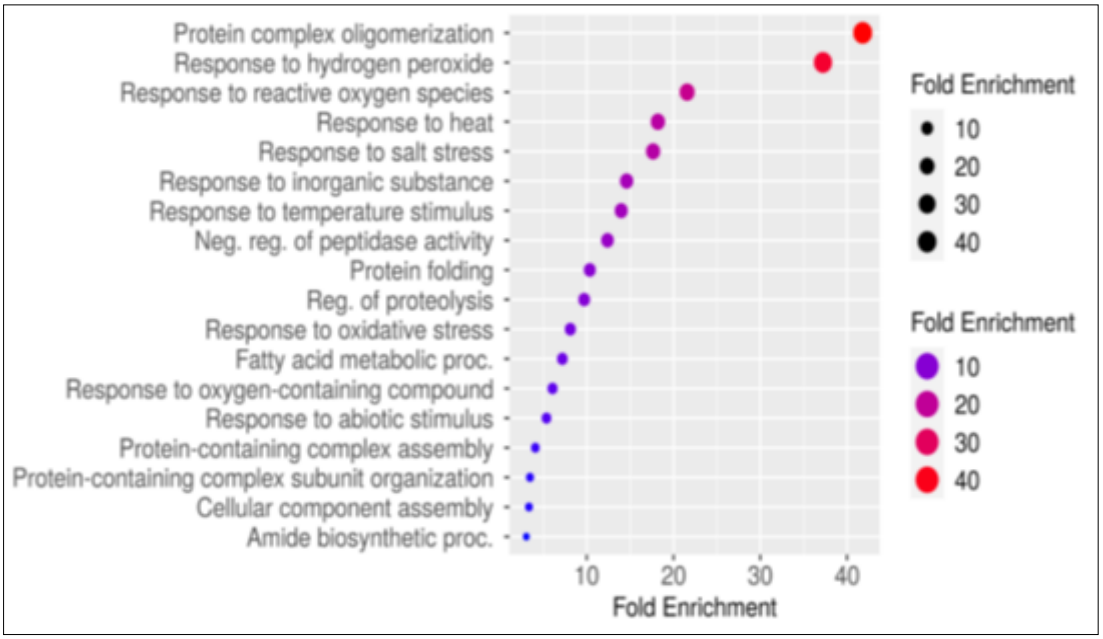

Fig. 13. Functional enrichment of the top-most significant (FDR<0.05) categories of biological processes (GO terms) for comparison 2.

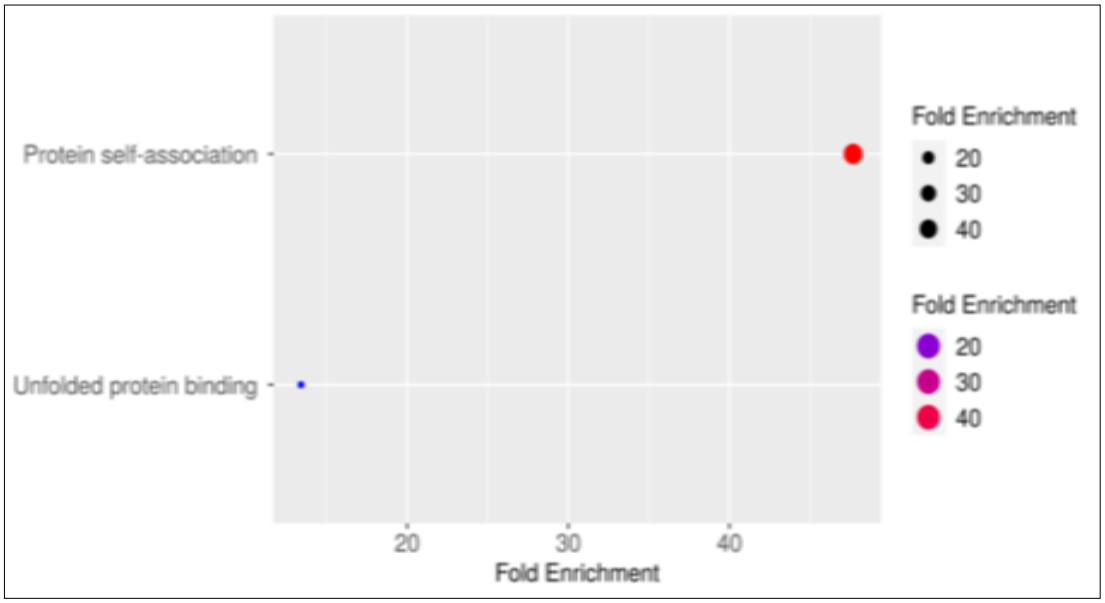

Fig. 14. Functional enrichment of the top-most significant (FDR<0.05) categories of molecular function (GO terms) for comparison 2.

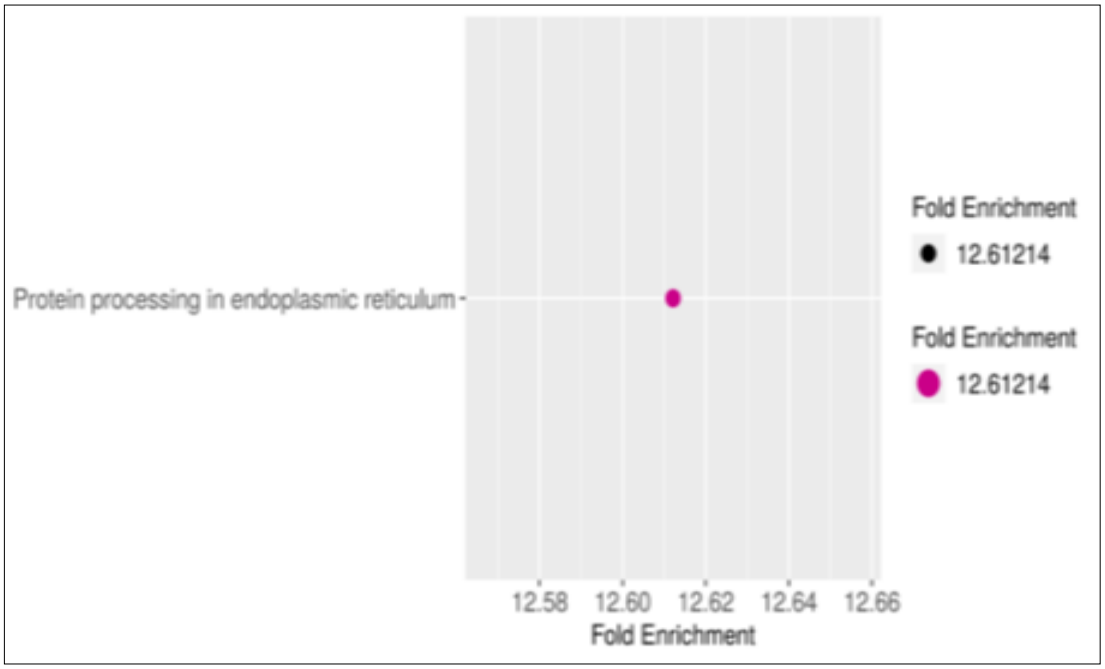

Fig. 15. Functional enrichment of the top-most significant (FDR<0.05) categories of KEGG terms for comparison 2.

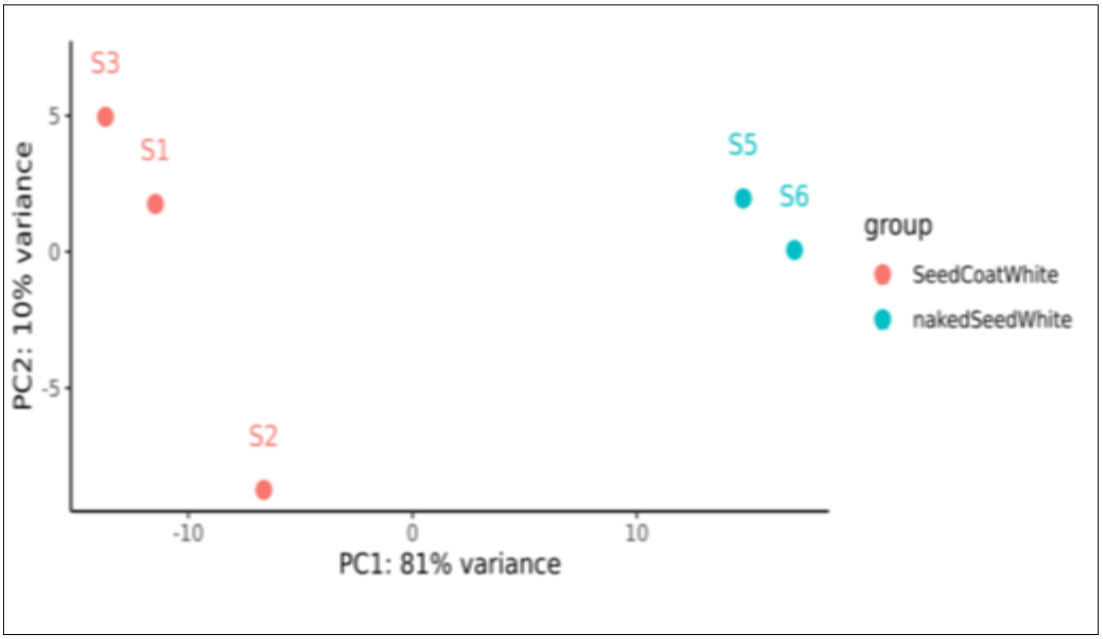

Fig. 16. Principal component analysis (PCA) plot of comparison 3.

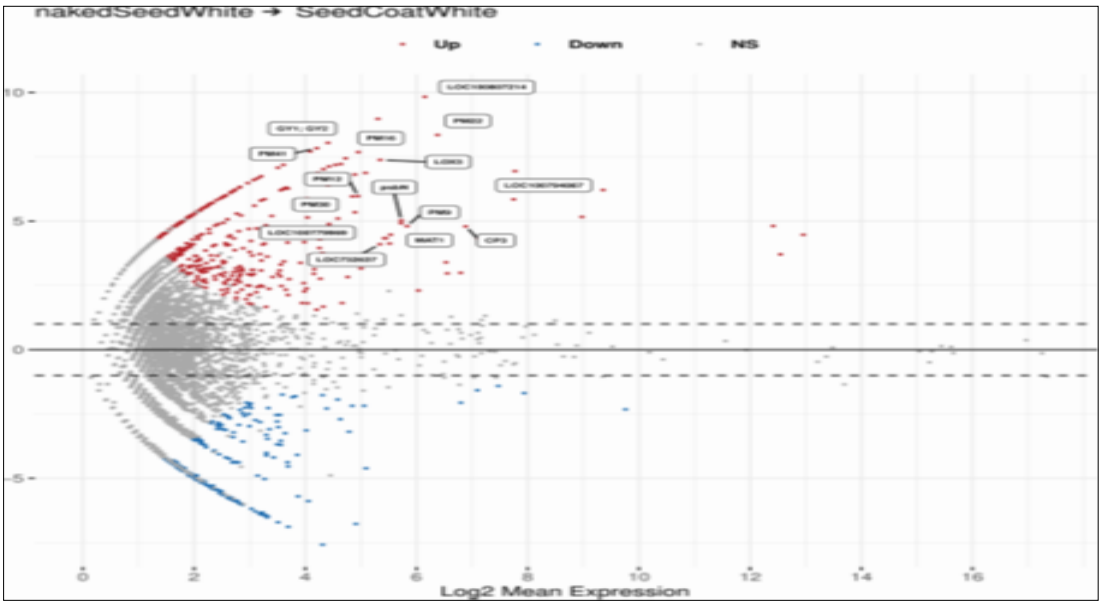

Fig. 17. MA plot of comparison 3.

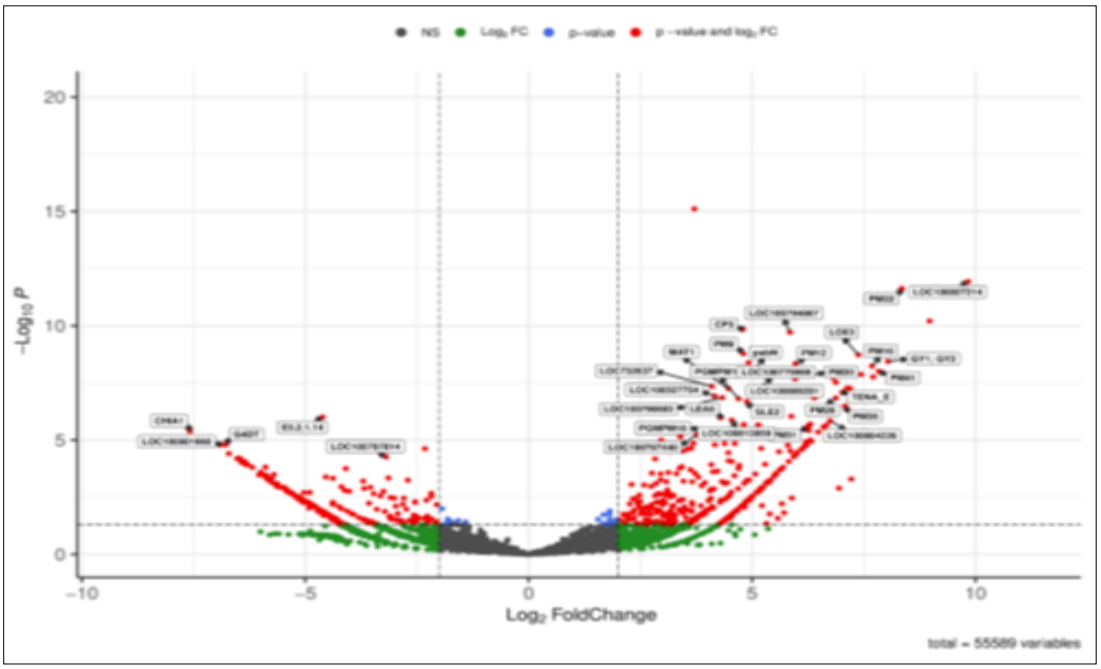

Fig. 18. Volcano plot of comparison 3.

opposed to the white seed cover (Fig. 17).

The differentially expressed transcripts profile in the heatmap of regularised log-transformed values, shows the top 50 genes with the highest variance across samples. Similarly, the volcano plot showed that many DEGs were expressed, among which 29 DEGs were up-regulated and 5 were down-regulated (Fig. 18), which depicts that the gene expression plays a role in the white naked seed and white seed coat colour.

Characterisation and functional annotation of DEGs in the white naked seed and white seed coat

DEGs in the white naked seed and white seed coat were used to map the GO database using the ShinyGO 0.76 toolkit, with a false discovery rate (FDR values 0.05) set as the cutoff. This allowed us to look at the significantly enriched genes in comparison to the genomic background. The results showed that in the up-regulated DEGs of a white naked seed, 63 significant GO keywords were identified between the two gene ontologies in comparison to the white seed coat. Among the primary classes involved in the biological process were "seed oilbody biogenesis, lipid storage, response to cold, response to water, fruit development, reaction to reactive oxygen species, reaction to hydrogen peroxide and protein complex oligomerisation reactions to stimuli involving temperature, inorganic substances, abscisic acid, alcohol and compounds containing oxygen, between others amide biosynthesis, cellular amide metabolic process, response to an abiotic stimulus, translation, peptide biosynthesis and peptide metabolic process (Fig. 19).

The principal classes within the cellular-function category were lipid droplet, actin filament, monolayer-surrounded lipid storage body, storage vacuole, protein storage vacuole, cytosolic ribosome, ribosome, cytosolic large ribosomal subunit, ribosome, large ribosomal subunit, cytosolic ribosome, small ribosomal subunit, the organelle envelope, envelope, ribonucleoprotein complex, chloroplast thylakoid, plastid thylakoid and mitochondrial inner membrane are all seen in Fig. 20. The three main classes under the molecular-function category were "oxidoreductase activity, cytochrome-c oxidase activity and inositol 3-alpha-galactosyltransferase activity, working on a heme group of donors, Protein self-association, binding to quaternary ammonium groups, binding to chitin, binding to galactosyltransferase, binding to abscisic acid, oxidoreduction-driven active transmembrane transporter activity, binding to monocarboxylic acid, binding to translation initiation factor, binding to translation factor, binding to RNA, binding to nucleic acid, binding to translation regulator and binding to unfolded protein, primary active transmembrane transporter activity, structural molecular activity and ribosome structural component, as shown in Fig. 21. Key mediators in naked white soybean seed and white seed coat appear to be connected to the top molecular networks linked to each of the significantly changed genes.

Functional regulatory network analysis (KEGG pathway enrichment) of white naked seed and white seed coat

To identify the distinct metabolic processes involved in the development of the white naked seed and white seed coat, specifically discovered DEGs in these two tissues were then compared to the KEGG pathway database. The primary processes that are enhanced in the white naked seed are represented in Fig. 22, which includes protein processing in the endoplasmic reticulum, ribosome and galactose metabolism pathways in soybean signal transduction from KEGG, with query genes highlighted in red, which is represented in Fig. 23.

Identification of transcription factors in white naked seed and white seed coat

Furthermore, the DEGs encode the TF. Many transcription factors, such as the ones listed below, were discovered to control soybean seeds. In white naked seeds, upregulated TF were LOC100807214, PM22 (peripheral myelin protein 22), PM16 (lysozyme, probable outer membrane protein pmp16, DNA-directed RNA polymerase), GY1;GY2, LOX3 (lipoxygenase 3, chloroplastic, seed linoleate 9S-lipoxygenase-3), PM41 (seed maturation protein PM41), PM12 (cytochrome c oxidase subunit 1), PM30 (isocitrate lyase, seed maturation protein PM30), LOC100779969, LOC732637, psbN (photosystem biogenesis factor 1), PM9 (ribonucleoside-diphosphate reductase, terminase, large subunit), MAT1 (CDK-activating kinase assembly factor MAT1, Phenolic glucoside malonyltransferase 1), CP3 (cytochrome P450 3A11), LOC100794867, TENA_E (probable bifunctional TENA-E protein), PM35 (seed maturation protein PM21), LOC100804226, PM28 (seed maturation protein PM28), LOC100809281, SLE2 (protein SLE2, transcriptional

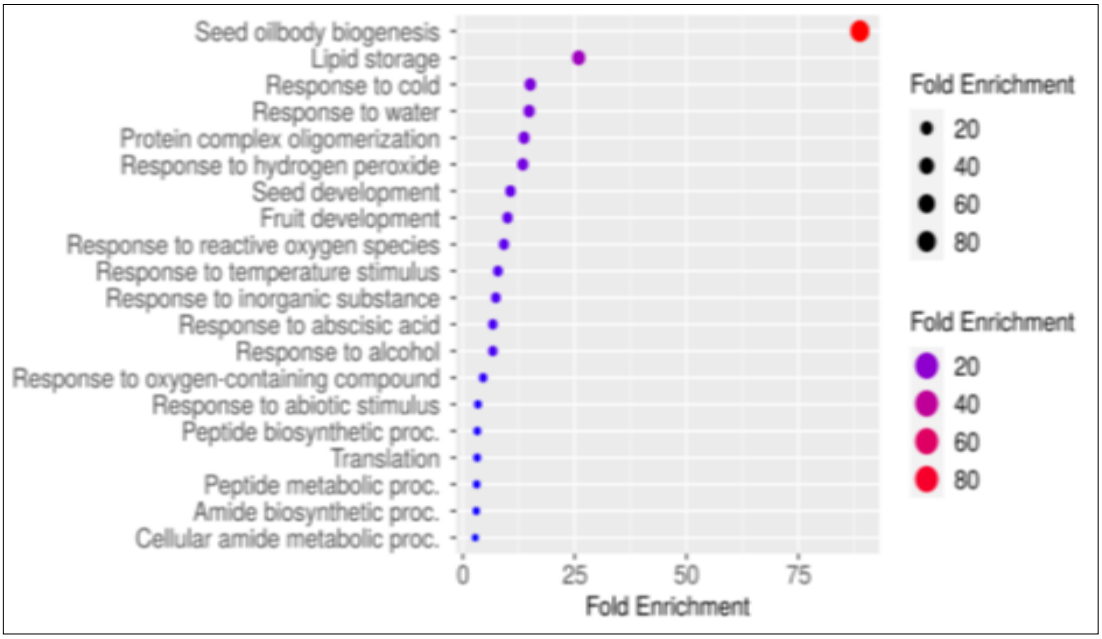

Fig. 19. Functional enrichment of the top-most significant (FDR<0.05) categories of biological processes (GO terms) for comparison 3.

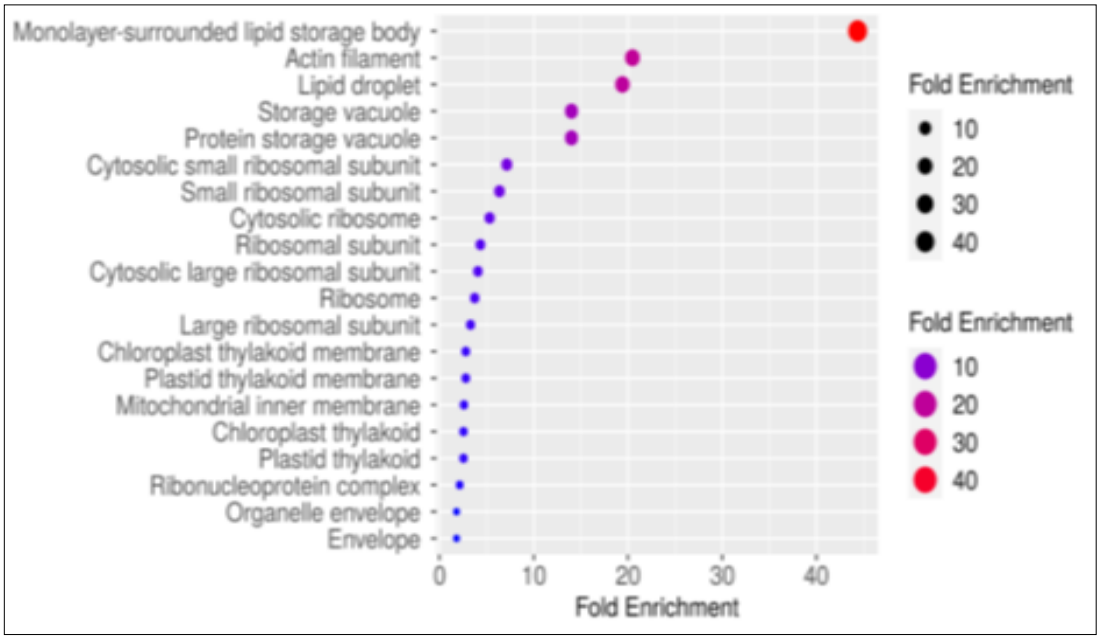

Fig. 20. Functional enrichment of the most significant (FDR<0.05) categories of cellular component (GO terms) for comparison 3.

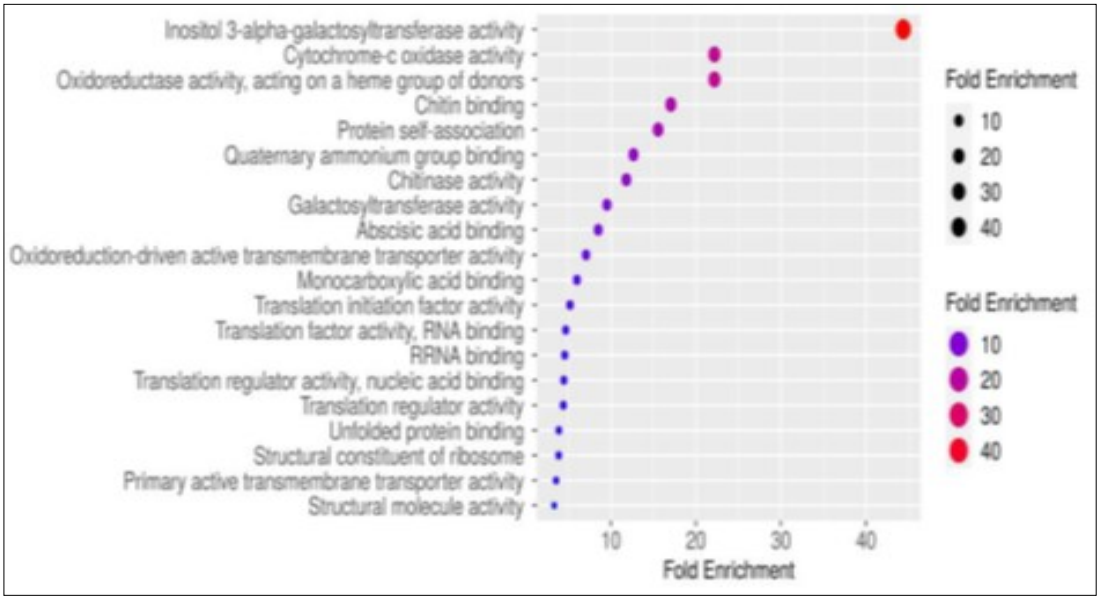

Fig. 21. Functional enrichment of top-most significant (FDR<0.05) categories of molecular function (GO terms) for comparison 3.

Fig. 22. Functional enrichment of top most significant(FDR<0.05) categories of KEGG terms for comparison 3.

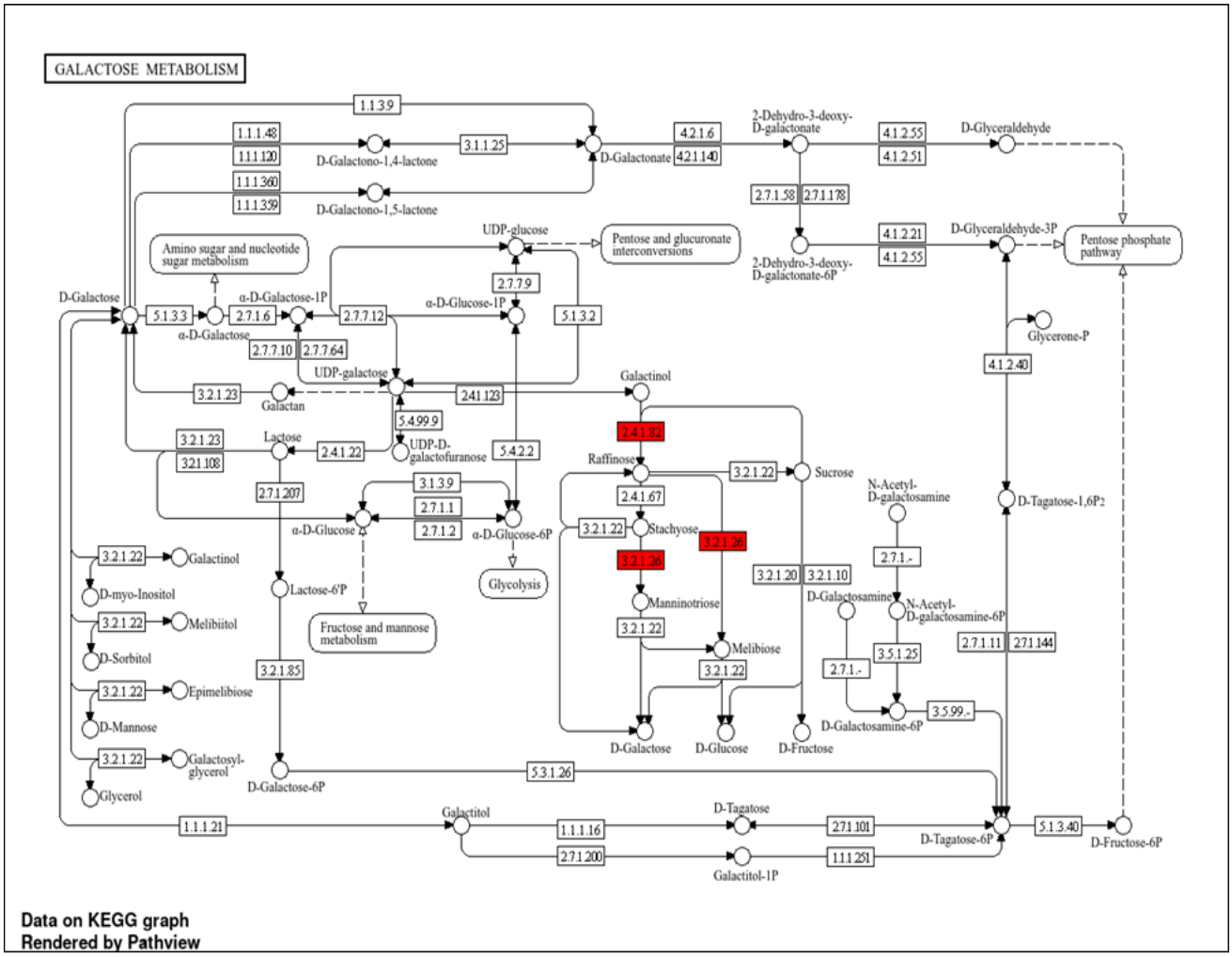

Fig. 23. Galactose metabolism pathways in soybean signal transduction from KEGG with query genes highlighted in red.

regulator WhiB), PM31 (seed maturation protein PM31), LOC100813859, LEA5 (protein SENESCENCE-ASSOCIATED GENE 21, mitochondrial, protein SLE2), PGMPM18 (35 kDa seed maturation protein), LOC100787440, LOC100790683, LOC100527754, LOC732637. CHIA1 (chitinase), LOC100801060, G4DT (glycinol 4-dimethylallyltransferase), E3.2.1.14 (E3 ubiquitin-protein ligase ORTHRUS 2) and LOC100787814 were the downregulated transcription factors in white naked seed.

Overview of metabolism

Lipid storage and seed oilbody biogenesis are essential for seed growth and germination. Triacylglycerols are stored in oilbodies, which offer a rich energy supply that is essential for seedling growth after germination. Compared to the white seed coat, white bare seed had a greater enrichment of genes linked to lipid storage and seed oilbody synthesis. Stable oilbodies are formed during proper biogenesis, shielding the stored lipids from oxidation and destruction. When photosynthesis is still in its early phases, this energy store powers metabolic functions and seedling development. Therefore, seed survival, successful germination and the development of robust, healthy seedlings depend on efficient lipid storage and utilisation. Due to their ability to cause oxidative stress, which damages cellular constituents like lipids, proteins and DNA, ROS and hydrogen peroxide are key contributors to seed toxicity. Weakening seedling growth and reduced germination rates are the outcomes of this oxidative damage, which also reduces seed viability and vigour. ROS and hydrogen peroxide accumulation during storage accelerates ageing and lowers seed quality even more. Compared to white seed coat, white naked seeds had greater levels of gene enrichment linked to responses to ROS and hydrogen peroxide. For seeds to remain healthy and resist oxidative damage, strong antioxidant defence systems are essential.

Multiple cellular structures are important during seed formation. Lipid storage bodies, sometimes called lipid droplets, are monolayer-surrounded structures that hold lipids rich in energy for germination. These lipid droplets are transported together with other organelles by actin filaments, which are a component of the cytoskeleton. The proteins required for the growth of seedlings are accumulated by storage vacuoles, especially protein storage vacuoles. For converting mRNA into proteins, cytosolic ribosomes, which are made up of the cytosolic small and large ribosomal subunits, are required. Photosynthesis produces energy for the post-germination development of seedlings and this process is carried out by the plastid and chloroplast thylakoid membranes. To produce ATP, which provides energy for cellular functions, the inner membrane of the mitochondria must function. Protein synthesis is made effective by ribonucleoprotein complexes, such as ribosomes. For chloroplasts and mitochondria in particular, the organelle envelope preserves the interior conditions required for their proper operation.

In order to build functional cellular machinery, cellular component assembly and protein-containing complex formation are essential. Amino acids that are necessary for protein synthesis are produced via amide biosynthesis mechanisms. Protein

stability is achieved by inhibiting unintended degradation through the modulation of proteolysis and negative regulation of peptidase activity. The endoplasmic reticulum performs folding and processing of proteins to guarantee that they take on their proper functional conformations. The correct construction, function and control of proteins, all necessary for seed viability and germination, are guaranteed by these mechanisms working together. White naked seed had an enrichment of genes linked to their function.

It was discovered that white seeds without a coat had more genes linked to resistance to unfavourable conditions. Responses to heat, salt stress, inorganic materials, temperature stimuli, oxidative stress, oxygen-containing chemicals and abiotic stimuli are essential for the survival and healthy growth of seeds during development. In order to maintain cellular stability and enzyme function, seeds evolve defences against heat and temperature changes. In order to preserve osmotic balance, they regulate ions and compartmentalise inorganic compounds in response to salt stress. Antioxidant enzymes prevent oxidative stress and shield cellular constituents from harm. To avoid causing harm to cells, reactions to substances containing oxygen, including ROS, are strictly controlled.

Different enzymatic and binding activities are crucial for molecular functional enrichment in seed development. The production and modification of cell walls are facilitated by the activities of galactosyltransferase and isomaltose 3-alpha-galactosyltransferase. Energy production during respiration is facilitated by the actions of oxidoreductase and cytochrome c oxidase. Pathogen protection involves both chitin binding and chitinase activity. Quaternary ammonium group binding and protein self-association contribute to the development and stability of complexes. The binding of abscisic acid controls stress reactions. RNA and rRNA binding, translation initiation and regulator functions all work together to promote effective protein synthesis. Transmembrane transporter functions preserve cellular homeostasis and nutrient transport, both of which are necessary for seed viability and germination. Unfolded protein binding and structural components of ribosomes are also critical for appropriate protein folding and assembly.

Comparison 4 (black naked seed vs black seed coat)

Comparative transcriptomic profiling of black naked seed vs black seed coat

910 genes that were found to be differently expressed in the black naked seed of the cultivar EC993950 were found to be different from those in the black seed coat cultivar. There were 302 upregulated genes and 608 downregulated genes out of 910 genes in the black naked seed as compared to the black seed coat (Table 4). There are significant differences between the black naked seed and the black seed coat, as shown by the PCA. In Fig. 24, which uses regularised log count data and a plot to illustrate variation both within and across groups to explain data variability, the PCA models' percentage of explained dispersion was 83 %. The largest percentage of variation (67 % for PC1 and 16 % for PC2) may be explained by the two main components that are displayed on the horizontal and vertical axes.

Between the groups of black naked seed and black seed coat, the MA-plot displays the distribution of gene expression. The average of the samples' normalised expression counts is represented by the log on the X axis, while the Log2 fold change is displayed on the Y axis. Based on the *p*-value (<0.05), genes that are up-regulated (> +2) and down-regulated (< -2) are shown by red and blue dots, respectively. The non-significant genes with *p*-values greater than 0.05 are represented by a grey dot. As seen in Fig. 25, it was evident that the black bare seed had more upregulated genes than the black seed coat. A heatmap of regularised log-transformed values displaying the top 50 genes with the highest variance across samples displays the differentially expressed transcripts profile. Similarly, the volcano plot revealed the expression of numerous DEGs, of which 20 were up-regulated and 19 were down-regulated (Fig. 26). This indicates that the expression of certain genes contributes to the formation of the black seed coat and black bare seed.

Characterisation and functional annotation of DEGs in the black naked seed and black seed coat

To investigate the highly enriched genes in comparison to the genomic background, DEGs in the black naked seed and black seed coat were employed to map the GO database, utilising the ShinyGO 0.76 toolkit, with a false discovery rate (FDR values 0.05) set as the cutoff. A total 54 significant GO keywords were identified between the two gene ontologies in the up-regulated DEGs of a naked black seed, as compared to the black seed coat, according to the data.

The primary classes that contributed to the biological process were "seed oilbody biogenesis, response to freezing, lipid storage, protein complex oligomerisation, response to hydrogen peroxide, response to water, response to acid chemical, response to ROS, seed development, response to temperature stimulus, fruit development, response to heat, neg. reg. of peptidase activity, response to salt stress, response to inorganic substance, neg. reg. of hydrolase activity, lipid localisation, response to oxygen containing compound, response to abiotic stimulus (Fig. 27).

The principal classes within the cellular-function category were aleurone grain, storage vacuole, protein storage vacuole, lipid droplet, 1,3-beta-D-glucan synthase complex, monolayer surrounded lipid storage body (Fig. 28). In the molecular-function category, the primary class were " NAD+ ADP-ribosyltransferase activity, serine-type endopeptidase inhibitor activity, protein self-association 1,3-beta-d-glucan synthase activity, endopeptidase inhibitor activity, peptidase inhibitor activity, endopeptidase regulator activity, peptidase regulator activity p-type ion transporter activity, nutrient reservoir activity, atpase-coupled cation transmembrane transporter activity, unfolded protein binding, carbon-carbon lyase activity, calmodulin binding, atpase-coupled transmembrane transporter activity, glucosyltransferase activity, primary active transmembrane transporter activity, ubiquitin-protein transferase activity, ubiquitin-like protein transferase activity, which were shown in Fig. 29. Key mediators in soybean black naked seed and black seed coat are connected to the top molecular networks linked to each of the significantly changed genes.

Functional regulatory network analysis (KEGG pathway enrichment) of black naked seed and black seed coat

The KEGG pathway database was then examined for specifically identified DEGs in the black naked seed and black seed coat to elucidate the distinct biochemical processes involved in the formation of soybean seeds in the black naked seed and black seed coat. Linoleic acid metabolism pathways in soybean signal transduction from KEGG with query genes highlighted in red, as shown in Fig. 31, alpha-linolenic acid metabolism, protein

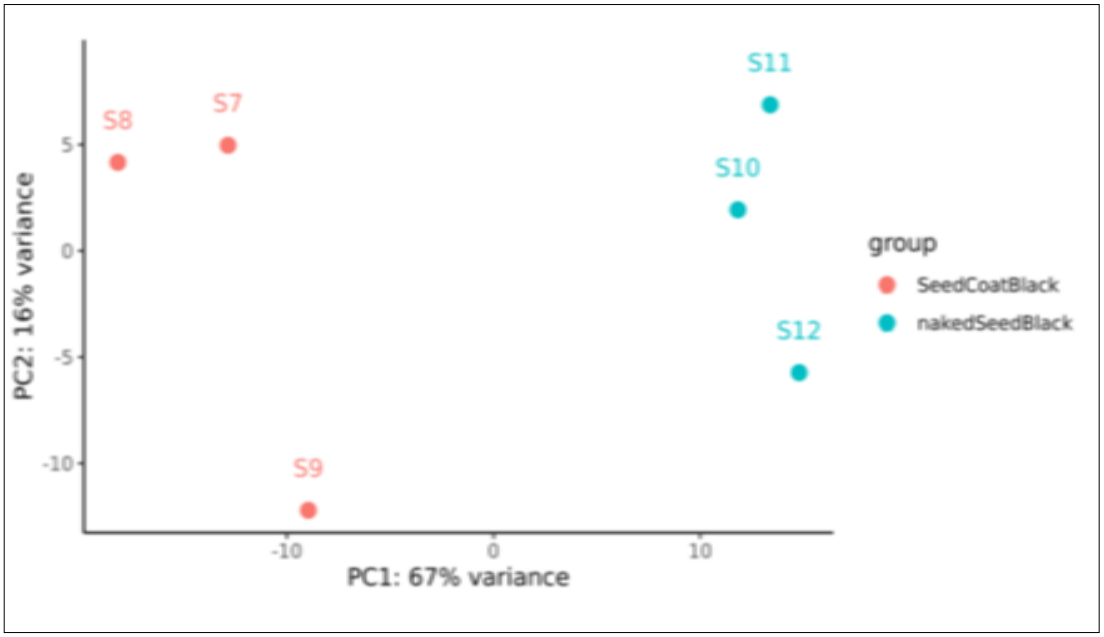

Fig. 24. Principal component analysis (PCA) plot of comparison 4.

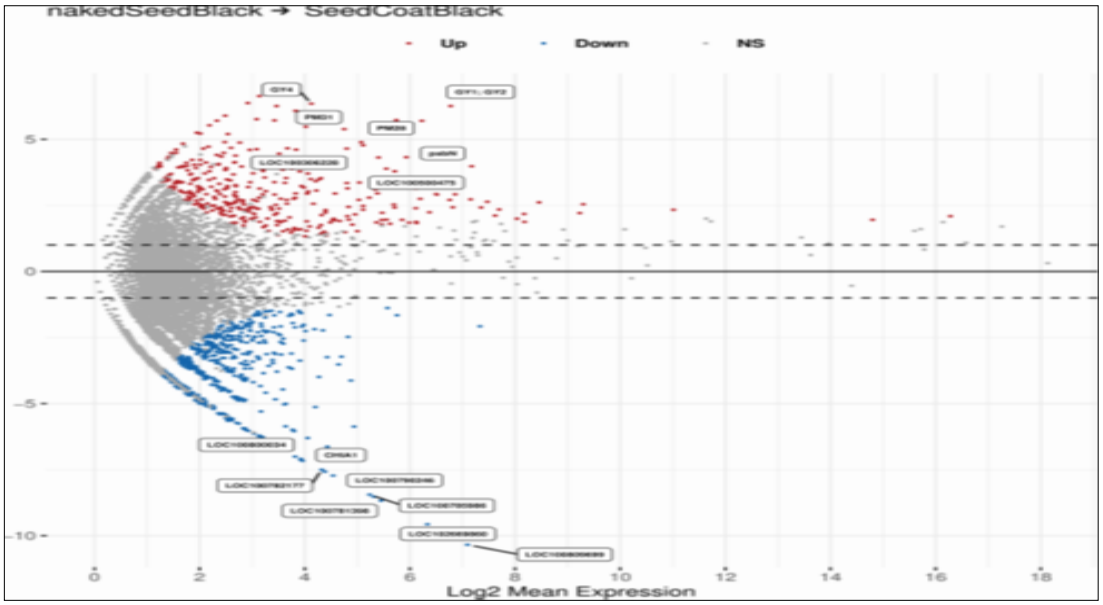

Fig. 25. MA plot of comparison 4.

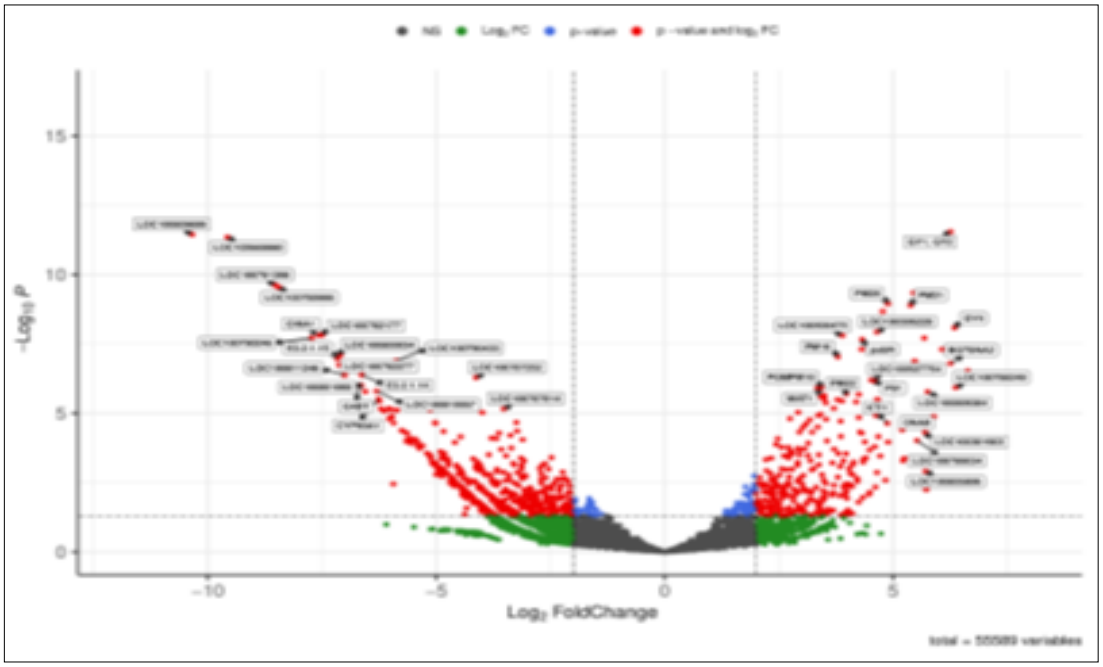

Fig. 26. Volcano plot of comparison 4.

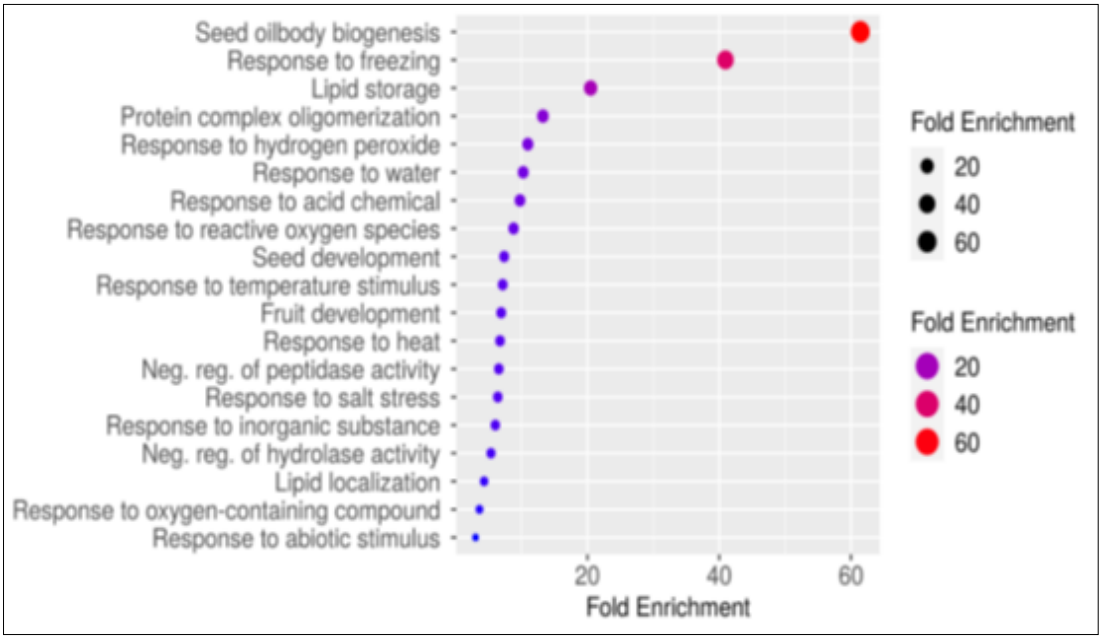

Fig. 27. Functional enrichment of top-most significant (FDR<0.05) categories of biological processes (GO terms) for comparison 4.

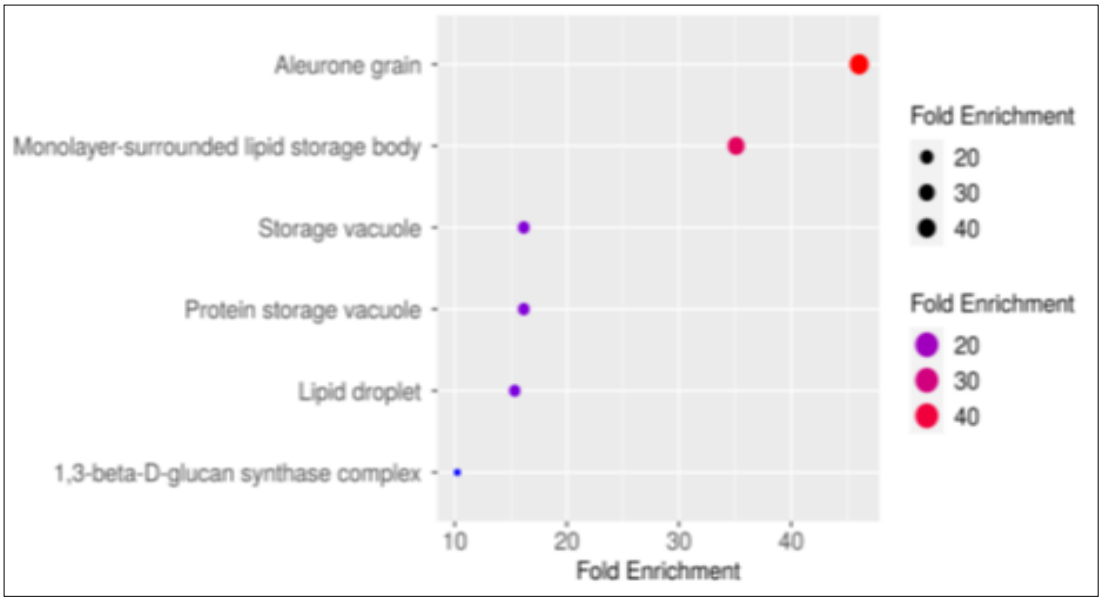

Fig. 28. Functional enrichment of top-most significant (FDR<0.05) categories of cellular component (GO terms) for comparison 4.

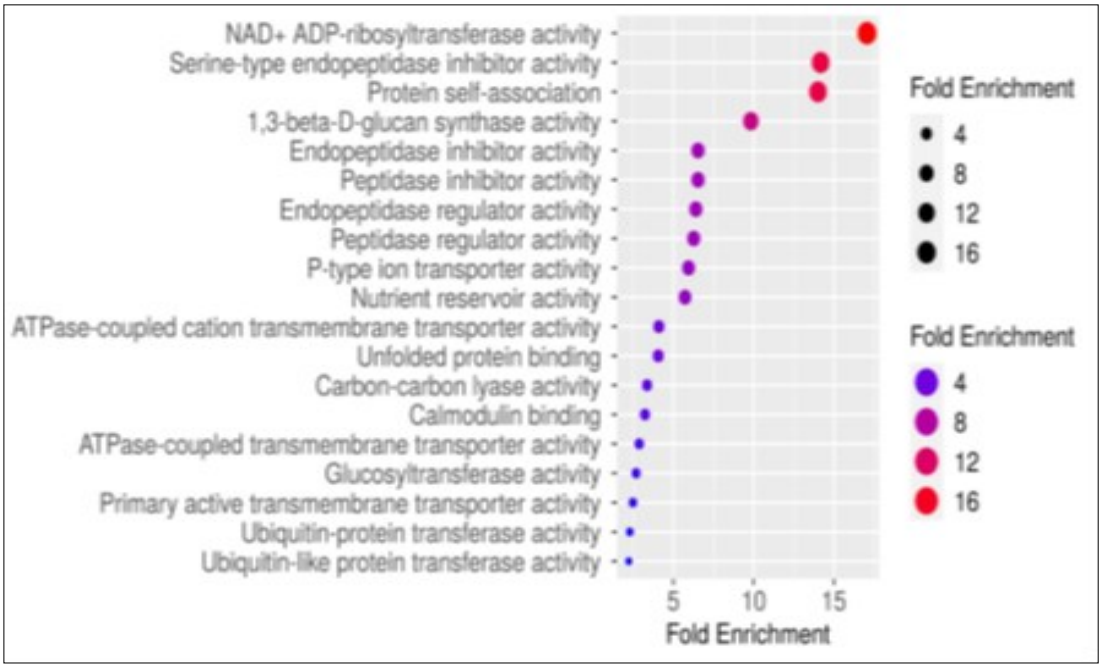

Fig. 29. Functional enrichment of top-most significant (FDR<0.05) categories of molecular function (GO terms) for comparison 4.

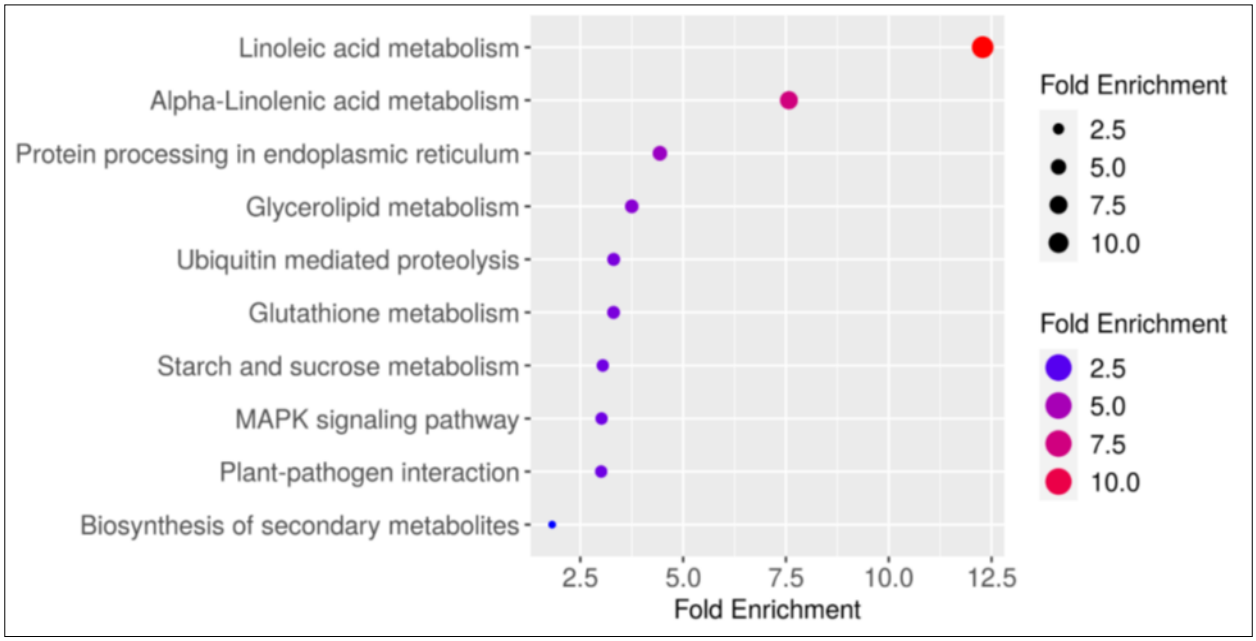

Fig. 30. Functional enrichment of top most significant(FDR<0.05) categories of KEGG terms for comparison 4.

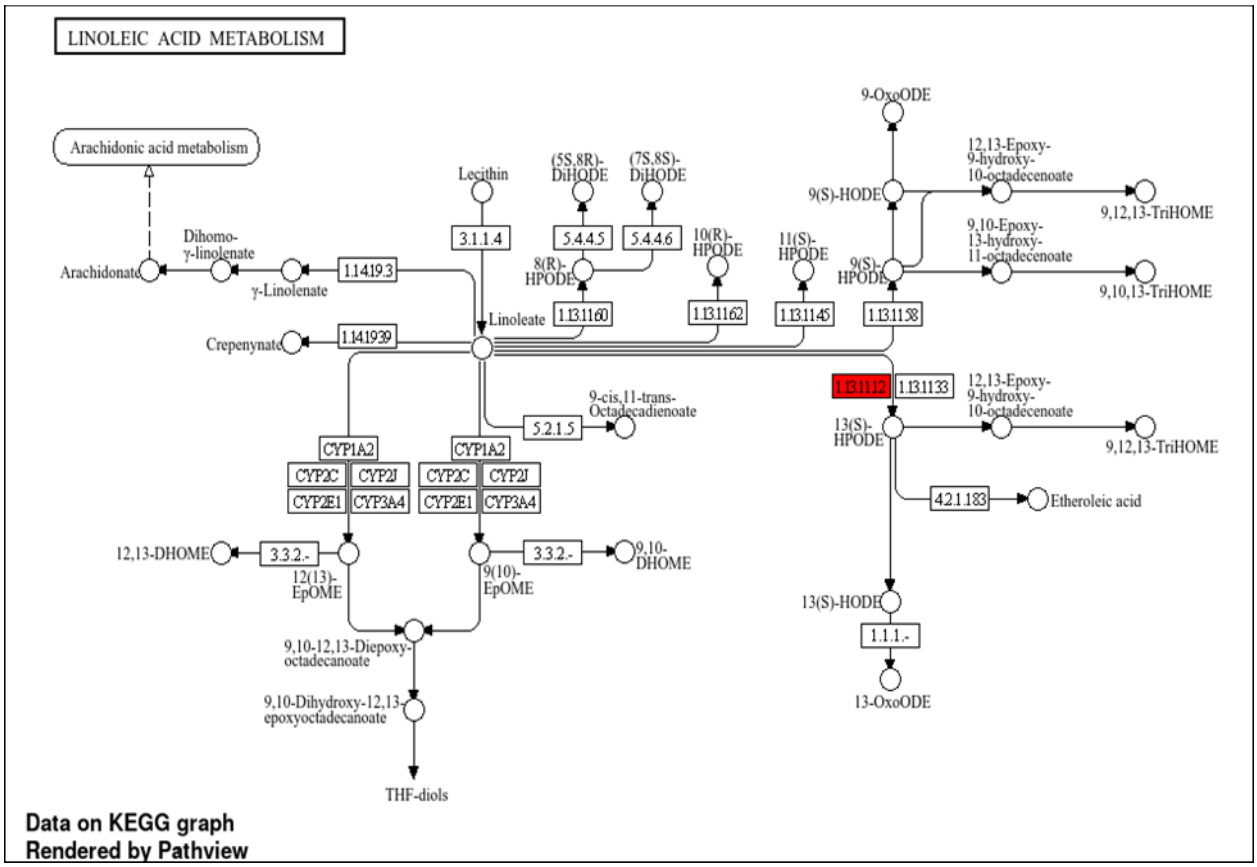

Fig. 31. Linoleic acid metabolism pathways in soybean signal transduction from KEGG with query genes highlighted in red.

processing in endoplasmic reticulum, glycerolipid metabolism, ubiquitin-mediated proteolysis, glutathione metabolism, starch and sucrose metabolism, MAPK signalling pathway, plant-pathogen interaction, biosynthesis of secondary metabolites, are among the major mechanisms elevated in the black naked seed coat (Fig. 30).

Identification of transcription factors in black naked seed and black seed coat

Furthermore, the DEGs encode the TF. The following transcription factors were discovered to be involved in controlling the colour of the black seed coat and black naked seed. GY4 (glycinin G4), PM31 (seed maturation protein PM31), PM28 (seed maturation protein PM28) and psbN (photosystem biogenesis factor 1) were the upregulated transcription factors in black naked seeds. LOC100306228, LOC100500475, GY1;GY2, PM16 (phenolic glucoside malonyltransferase 1), PM22 (peripheral myelin protein 22), LOC100527754, P91 (P24 oleosin isoform B), KTI1 (kunitz-type trypsin inhibitor KTI1), LOC100788249, LOC100809384, DNAII (2',3'-cyclic-nucleotide 3'-phosphodiesterase) LOC100301903, LOC100799034, LOC100805806.

Downregulated TF in black naked seed were LOC100800034, CHIA1 (chitinase), LOC100782177, LOC100790246, LOC102669860, LOC100809699, LOC100809699, LOC102669860, LOC100781398, LOC100785986, LOC100790246, E3.2.1.15 (E3 ubiquitin-protein ligase brl1), LOC100790433, LOC100811246,

LOC100782277, LOC100787252, LOC100801060, E3.2.1.14 (E3 ubiquitin-protein ligase listerin), G4DT (Glycinol 4-dimethylallyltransferase), CYP93A1 (3,9-dihydroxypterocarpan 6A-monooxygenase), LOC100810057, LOC00787814.

Comparative assessment of seed quality and storability in black and white seed coat soybean genotypes

Values represent mean observations of seed quality parameters, including germination percentage, seedling vigour indices I and II, electrical conductivity and size of hilar opening in black and white seed coat soybean seeds under fresh and accelerated ageing conditions. Electrical conductivity is expressed as $\mu S\ cm^{-1}\ g^{-1}$ and indicates membrane integrity, while hilar opening size is presented as mean ± standard error (µm).

The comparative evaluation of seed quality parameters revealed that black-seeded soybean genotypes exhibited superior storability and stress tolerance compared to white seed coat seeds (Table 5). Although fresh white seeds showed slightly higher germination (96 %) than black seeds (90 %), black seeds maintained better performance after accelerated ageing, with higher germination (75 %) and significantly greater seedling vigour indices I and II than white seeds (65 %). Lower electrical conductivity values in black seeds under both fresh and aged conditions indicated better membrane integrity and reduced solute leakage, reflecting enhanced physiological quality. Furthermore, black seeds possessed a smaller hilar opening size compared to white seeds, both in fresh and aged conditions, which likely contributed to reduced moisture ingress and slower deterioration during storage. Overall, the results demonstrate that black-seeded soybean seeds have improved longevity, membrane stability and vigour retention under ageing stress, making them more suitable for long-term storage and better seed quality maintenance than white seed coat counterparts.

Overview of metabolism

Many biological systems make sure that seeds develop with resistance and healthy growth. Lipid storage and seed oilbody biogenesis supply vital energy reserves. Protein complex oligomerisation guarantees the assembly of functional proteins. Seeds are able to tolerate environmental stressors such as freezing, heat, water, salt stress, inorganic chemicals and temperature stimulation. Detoxifying toxic substances is part of the reaction to ROS and hydrogen peroxide. Unwanted protein degradation is avoided via the negative control of peptidase and hydrolase activity. Within the seed, appropriate lipid distribution is ensured by lipid localisation. Cellular stability is preserved by reactions to substances that are acidic and those that contain oxygen. Together, these mechanisms facilitate fruit development and seed germination, guaranteeing successful growth and germination. Black naked seeds have a higher functional enrichment of these genes than the black seed coat.

Minerals and proteins required for germination are stored by aleurone grains during seed formation. Lipid storage structures and lipid droplets encircled by a monolayer retain lipids that are high in energy and essential for the seedlings' energy requirements. Proteins necessary for growth are accumulated and shielded by storage vacuoles, which include protein storage vacuoles. The 1,3-beta-D-glucan synthase complex is engaged in the synthesis of beta-glucans, which contribute to the structural integrity and strength of cell walls. The activity of NAD+ ADP-ribosyltransferase is involved in signalling and DNA repair. Protein self-association promotes complex formation, while peptidase inhibitor, regulator and serine-type endopeptidase activities stop protein breakdown. Components of the cell wall are synthesised by the 1,3-beta-D-glucan synthase complex. Ion homeostasis is preserved via P-type ion and ATPase-coupled cation transmembrane transporter activities. The operation of the nutrient reservoir guarantees a supply of necessary nutrients. Calmodulin binding controls calcium signalling, while unfolded protein binding facilitates protein folding. Carbohydrates are changed by glucosyltransferase activity and protein breakdown is controlled by ubiquitin-protein transferase activity.

Conclusion

This study analysed transcriptional profiles of black and white soybean seed coat genotypes and revealed significant differences in gene expression associated with seed quality. Principal component analysis demonstrated clear segregation between the two genotypes, indicating distinct expression patterns. Notably, black seed coat genotypes exhibited upregulation of genes involved in lysine biosynthesis and related metabolic pathways. The enhanced expression of these genes may contribute to improved protein quality, increased stress tolerance and greater seed longevity observed in black-coated soybean varieties. Key pathways included hormonal regulation and metabolism of lysine, starch, sucrose, protein and galactose, which collectively contributed to the superior seed quality observed in black-coated soybean genotypes. The claim is also supported by lab experiments on seed quality parameters. The differential expression of transcription factors provided insights into the regulatory networks underlying these traits. These findings offer a genomic foundation for improving seed quality, with future studies potentially using gene silencing to validate gene functions and enhance seed longevity. Future work will involve validating key candidate genes involved in pigment biosynthesis, hormone signalling and metabolic pathways using qRT-PCR to strengthen the reliability of the identified differentially expressed genes.

Acknowledgements

The authors would like to acknowledge the TRuCapSols, Bethlehem,

Table 5. Comparative evaluation of seed quality parameters in black and white seed coat soybean seeds under fresh and accelerated ageing conditions

Parameters	Black seed coat		White seed coat	
	Fresh seeds	Accelerated-aged seeds	Fresh seeds	Accelerated-aged seeds
Germination (%)	90	75	96	65
Seedling vigour index I	3424	2321	3345	1798
Seedling vigour index II	5268	3252	4892	2285
Electrical conductivity ($\mu S\ cm^{-1}\ g^{-1}$)	11.57	42.62	15.11	52.32
Size of hilar opening (µm)	1.4 ± 0.04	8.44 ± 0.29	3.9 ± 0.74	13.21 ± 0.55

the USA, for the funding of the expenses of this research under the project Ab/Ac 5371, Understanding the biochemical composition of seed coat in different crops to develop a biodegradable membrane for encapsulation. The authors also thank the University Grant Commission (UGC), India, for providing the NFSC fellowship .

Authors' contributions

VCJ contributed to conceptualisation, methodology, writing review, original draft and analysis. NN contributed to conceptualisation, funding acquisition, project administration, resources and editing. SS contributed to conceptualisation, methodology and editing. KN contributed to methodology, supervision and editing. NSN contributed to conceptualisation, methodology and editing. P contributed to conceptualisation and editing. VSN conceptualisation, methodology and editing. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

References

1. Crane PR, Friis EM, Pedersen KR. The origin and early diversification of angiosperms. Nature. 1995;374:27–33. https://doi.org/10.1038/374027a0
2. Soltis P, Soltis D. The origin and diversification of angiosperms. Am J Bot. 2004;91:1614–26. https://doi.org/10.3732/ajb.91.10.1614
3. Radchuk V, Borisjuk L. Physical, metabolic and developmental functions of the seed coat. Front Plant Sci. 2014;5:510. https://doi.org/10.3389/fpls.2014.00510
4. Souza FH, Marcos Filho J. The seed coat as a modulator of seed environment relationships in Fabaceae. Braz J Bot. 2001;24:365–75. https://doi.org/10.1590/s0100-84042001000400002
5. Qiu HM, Chen L, Hou YL, Wang XF, Chen J, Ma XP, et al. Research progress on the genetic regulatory mechanism of seed colour in soybean (*Glycine max*). Acta Agron Sin. 2021;47(12):2299–313. https://doi.org/10.3724/SP.J.1006.2021.14022
6. Esau K. Anatomy of seed plants. 2nd ed. New York: John Wiley; 1977.
7. Miller SS, Bowman LA, Gijzen M, Miki BLA. Early development of the seed coat of soybean. Ann Bot. 1999;84:297–304. https://doi.org/10.1006/anbo.1999.0915
8. Wang YN, Qi GX, Zhao HK, Yuan CP, Liu XD, Li YQ, et al. Genetic diversity of soybean landraces with different seed coat colours. Mol Plant Breed. 2020;18:1–18. https://doi.org/10.13271/j.mpb.019.007984
9. McKee GW, Peiffer RA, Mohsenin NN. Seedcoat structure in Coronilla varia and its relations to hard seed. Agron J. 1977;69:53–8.
10. Eckardt NA. Tissue specific siRNAs that silence CHS genes in soybean. Plant Cell. 2009;21(10):2983–4. https://doi.org/10.1105/tpc.109.072421
11. Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, et al. Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol. 2006;57:405–30. https://doi.org/10.1146/annurev.arplant.57.032905.105252
12. Smykal P, Vanessa V, Blair MW, Aleš S, Richard DT. The role of the testa during development and in establishment of dormancy of the legume seed. Front Plant Sci. 2014;5:351. https://doi.org/10.3389/fpls.2014.00351
13. Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I. TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerisation of flavonoids in *Arabidopsis* seed coat. Plant Cell. 2005;17:2966–80. https://doi.org/10.1105/tpc.105.035154
14. Rajjou L, Debeaujon I. Seed longevity: survival and maintenance of high germination ability of dry seeds. C R Biol. 2008;331(10):796–805.
15. Sun W, Meng X, Liang L, Jiang W, Huang Y, He J, et al. Molecular and biochemical analysis of chalcone synthase from freesia hybrid in flavonoid biosynthetic pathway. PLoS One. 2015;10(3):e0119054. https://doi.org/10.1371/journal.pone.0119054
16. Zabala G, Vodkin L. Cloning of pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of encoded flavonoid 3′ hydroxylase. Genetics. 2003;163:295–309. https://doi.org/10.1093/genetics/163.1.295
17. Song J, Guo Y, Yu LJ, Qiu LJ. Progress in genes related to seed coat colour in soybean. Yi Chuan. 2012;34(6):687–94. https://doi.org/10.3724/sp.j.1005.2012.00687
18. Yoshikura K, Hamaguchi Y. Anthocyanins of black soybean. Eiyo To Shokuryo. 1969;22(6):367–70. https://doi.org/10.4327/jsnfs1949.22.367
19. Winkel-Shirley B. Flavonoid biosynthesis: a colourful model for genetics, biochemistry, cell biology and biotechnology. Plant Physiol. 2001;126(2):485–93. https://doi.org/10.1104/pp.126.2.485
20. Kovinich N, Saleem A, Rintoul TL, Brown DCW, Arnason JT, Miki B. Colouring genetically modified soybean grains with anthocyanins by suppression of the proanthocyanidin genes ANR1 and ANR2. Transgenic Res. 2012;21(4):757–71. https://doi.org/10.1007/s11248-011-9566-y
21. Wang H, Murphy PA. Isoflavone content in commercial soybean foods. J Agric Food Chem. 1994;42:1666–73. https://doi.org/10.1021/jf00044a016
22. Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M. Transcriptome analysis reveals a critical role of CHS7 and CHS8 genes for isoflavone synthesis in soybean seeds. Plant Physiol. 2007;143:326–38. https://doi.org/10.1104/pp.106.086306
23. Ralston L, Subramanian S, Matsuno M, Yu O. Partial reconstruction of flavonoid and isoflavone biosynthesis in yeast using soybean type I and type II chalcone isomerases. Plant Physiol. 2005;137:1375–88. https://doi.org/10.1104/pp.104.054502
24. Akashi T, Aoki T, Ayabe S. Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. Plant Physiol. 1999;121(3):821–8. https://doi.org/10.1104/pp.121.3.821
25. Jung W, Yu O, Lau SC, O'Keefe DP, Odell J, Fader G. Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat Biotechnol. 2000;18:208–12. https://doi.org/10.1038/72671
26. Kuchlan MK, Kuchlan P, Onkar M, Ramesh A, Husain SM. Influence of seed coat compactness around cotyledons, protein and mineral composition on mechanical strength of soybean (*Glycine max* (L.) Merrill) seed coat. Legume Res. 2018;41:246–52.
27. Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, et al. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol. 2011;11:5. https://doi.org/10.1186/1471-2229-11-5
28. Dhaubhadel S, McGarvey BD, Williams R, Gijzen M. Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol. 2003;53:733–43. https://doi.org/10.1023/B:PLAN.0000023666.30358.ae
29. Palmer RG, Pfeiffer TW, Buss GR, Kilen TC. Qualitative genetics. In: Boerma HR, Specht JE, editors. Soybeans: improvement,

production and uses. 3rd ed. Madison (WI): ASA, CSSA, SSSA; 2004. p. 137–214.

30. Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, et al. Genetic analysis and characterisation of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet. 2010;121(3):417–31. https://doi.org/10.1007/s00122-010-1320-y

31. Song J, Liu Z, Hong H, Ma Y, Tian L, Li X, et al. Identification and validation of loci governing seed coat color by combining association mapping and bulk segregation analysis in soybean. PLoS One. 2016;11:e0159064. https://doi.org/10.1371/journal.pone.0159064

32. Bernard RL, Weiss MG. Qualitative genetics. In: Caldwell BE, editor. Soybean: improvement, production and uses. Madison (WI): American Society of Agronomy; 1973. p. 117–54.

33. Nagamatsu A, Masuta C, Senda M, Matsuura H, Kasai A, Hong JS, et al. Functional analysis of soybean genes involved in flavonoid biosynthesis by virus induced gene silencing. Plant Biotechnol J. 2007;5:778–90. https://doi.org/10.1111/j.1467-7652.2007.00288.x

34. Toda K, Yang D, Yamanaka N, Watanabe S, Harada K, Takahashi R. A single base deletion in soybean flavonoid 3′ hydroxylase gene is associated with grey pubescence color. Plant Mol Biol. 2002;50:187–96. https://doi.org/10.1023/A:1016087221334

35. Clough SJ, Tuteja JH, Li M, Marek LF, Shoemaker RC, Vodkin LO. Features of a 103 kb gene rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the I locus. Genome. 2004;47(5):819–31. https://doi.org/10.1139/g04-049

36. Senda M. Patterning of virus infected *Glycine max* seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. Plant Cell. 2004;16(4):807–18. https://doi.org/10.1105/tpc.019885

37. Tuteja JH. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell. 2004;16(4):819–35. https://doi.org/10.1105/tpc.021352.

38. Kasai A, Kasai K, Yumoto S, Senda M. Structural features of GmIRCHS, candidate of the I gene inhibiting seed coat pigmentation in soybean: implications for inducing endogenous RNA silencing of chalcone synthase genes. Plant Mol Biol. 2007;64 (4):467–79. https://doi.org/10.1007/s11103-007-9169-4.

39. Tuteja JH, Zabala G, Varala K, Hudson M, Vodkin LO. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in *Glycine max* seed coats. Plant Cell. 2009;21:3063–77.

40. Senda M, Kasai A, Yumoto S, Akada S, Ishikawa R, Harada T, et al. Sequence divergence at chalcone synthase gene in pigmented seed coat soybean mutants of the inhibitor locus. Genes Genet Syst. 2002;77(5):341–50. https://doi.org/10.1266/ggs.77.341.

41. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–8. https://doi.org/10.1093/bioinformatics/btw354

42. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29 (1):15–21. https://doi.org/10.1093/bioinformatics/bts635

43. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30. https://doi.org/10.1093/bioinformatics/btt656

44. Blighe K, Rana S, Lewis M. Enhanced Volcano: publication-ready volcano plots with enhanced colouring and labeling. R package. 2019;Version 1.0.

45. Kassambara A, Kassambara MA. ggpubr: 'ggplot2' based publication-ready plots. R package. 2020;Version 0.1.

46. Singer WM, Zhang B, Mian MAR, Huang H. Soybean amino acids in health, genetics and evaluation. In: Soybean: human consumption and animal feed. 2019.

47. Aguirre M, Kiegle E, Leo G, Ezquer I. Carbohydrate reserves and seed development: an overview. Plant Reprod. 2018;31(3):263–90. https://doi.org/10.1007/s00497-018-0336-3

48. Tayade R, Kulkarni KP, Jo H, Song JT, Lee JD. Insight into the prospects for the improvement of seed starch in legumes: a review. Front Plant Sci. 2019;10:1213. https://doi.org/10.3389/fpls.2019.01213

49. Krishnan HB, Jez JM. The promise and limits for enhancing sulfur containing amino acid content of soybean seed. Plant Sci. 2018;272:14–21. https://doi.org/10.1016/j.plantsci.2018.03.030

50. Ma Y, Ma W, Hu D, Zhang X, Yuan W, He X, et al. QTL mapping for protein and sulfur-containing amino acid contents using a high-density bin map in soybean (*Glycine max* L. Merr.). J Agric Food Chem. 2019;67:12313–21. https://doi.org/10.1021/acs.jafc.9b04497

51. Ding Y, Zhou X, Zuo L, Wang H, Yu D. Identification and functional characterization of the sulfate transporter gene GmSULTR1;2b in soybean. BMC Genomics. 2016;17:1–19. https://doi.org/10.1186/s12864-016-2705-3

52. Bai Z, Qi T, Liu Y, Wu Z, Ma L, Liu W, et al. Alteration of S-adenosylhomocysteine levels affects lignin biosynthesis in switchgrass. Plant Biotechnol J. 2018;16:2016–26. https://doi.org/10.1111/pbi.12935

53. Malle S, Eskandari M, Morrison M, Belzile F. Genome wide association identifies several QTLs controlling cysteine and methionine content in soybean seed including some promising candidate genes. Sci Rep. 2020;10:1–14. https://doi.org/10.1038/s41598-020-78907-w

54. Zhang X, Wang Y, Yan Y, Peng H, Long Y, Zhang Y, et al. Transcriptome sequencing analysis of maize embryonic callus during early redifferentiation. BMC Genomics. 2019;20:1–22. https://doi.org/10.1186/s12864-019-5506-7

55. Miyakawa T, Hatano KI, Miyauchi Y, Suwa YI, Sawano Y, Tanokura M. A secreted protein with plant specific cysteine rich motif functions as a mannose binding lectin that exhibits antifungal activity. Plant Physiol. 2014;166:766–78. https://doi.org/10.1104/pp.114.242636

56. Belkhadir Y, Yang L, Hetzel J, Dangl JL, Chory J. The growth–defense pivot: crisis management in plants mediated by LRR-RK surface receptors. Trends Biochem Sci. 2014;39:447–56. https://doi.org/10.1016/j.tibs.2014.06.006

57. Ishibashi Y, Koda Y, Zheng SH, Yuasa T, Iwaya-Inoue M. Regulation of soybean seed germination through ethylene production in response to reactive oxygen species. Ann Bot. 2013;111:95–102. https://doi.org/10.1093/aob/mcs240

58. Ciabotti S, Silva ACBB, Juhasz ACP, Mendonça CD, Tavano OL, Mandarino JMG, et al. Chemical composition, protein profile and isoflavones content in soybean genotypes with different seed coat colors. Int Food Res J. 2016.

59. Chandra S, Talukdar A, Taak Y, Yadav RR, Saini M, Sipani NS. Seed longevity studies in wild type, cultivated and inter specific recombinant inbred lines (RILs) of soybean (*Glycine max* (L.) Merr.). Genet Resour Crop Evol. 2022;69:399–409. https://doi.org/10.1007/s10722-021-01240-2

60. Ohnishi T, Godza B, Watanabe B, Fujioka S, Hategan L, Ide K, et al. CYP90A1/CPD, a brassinosteroid biosynthetic cytochrome P450 of *Arabidopsis*, catalyzes C-3 oxidation. J Biol Chem. 2012;287:31551–60. https://doi.org/10.1074/jbc.M112.392720

61. Ahammed GJ, Gantait S, Mitra M, Yang Y, Li X. Role of ethylene crosstalk in seed germination and early seedling development: a review. Plant Physiol Biochem. 2020;151:124–31. https://doi.org/10.1016/j.plaphy.2020.03.016

62. Yin W, Dong N, Niu M, Zhang X, Li L, Liu J, et al. Brassinosteroid regulated plant growth and development and gene expression in soybean. Crop J. 2019;7:411–8. https://doi.org/10.1016/j.cj.2018.10.003

63. Chen H, Kim HU, Weng H, Browse J. Malonyl-CoA synthetase,

encoded by ACYL ACTIVATING ENZYME13, is essential for growth and development of Arabidopsis. Plant Cell. 2011;23:2247–62. https://doi.org/10.1105/tpc.111.086140

64. Griebel T, Zeier J. A role for β-sitosterol to stigmasterol conversion in plant pathogen interactions. Plant J. 2010;63:254–68. https://doi.org/10.1111/j.1365-313X.2010.04235.x

65. Huth C, Mertz-Henning LM, Lopes SJ, Tabaldi LA, Rossato LV, Krzyzanowski FC, et al. Susceptibility to weathering damage and oxidative stress on soybean seeds with different lignin contents in the seed coat. J Seed Sci. 2016;38:296–304. https://doi.org/10.1590/2317-1545v38n4162115

66. Bellaloui N, Mengistu A, Fisher DK, Abel CA. Soybean seed composition constituents as affected by drought and Phomopsis in Phomopsis susceptible and resistant genotypes. J Crop Improv. 2012;26:428–53. https://doi.org/10.1080/15427528.2011.651774

67. Kuchlan MK, Dadlani M, Samuel DVK. Seed coat properties and longevity of soybean seeds. J New Seeds. 2010;11:239–49. https://doi.org/10.1080/1522886X.2010.497960

68. Ariyoshi Y, Itoyama H, Nakagawa ACS, Ario N, Kondo Y, Tomita Y, et al. Regulation of brassinosteroid on pod growth through cell hypertrophy in soybean (*Glycine max* (L.) Merr.). Plant Growth Regul. 2016;80(3):391–5. https://doi.org/10.1007/s10725-016-0176-9

69. Cabianca A, Müller L, Pawlowski K, Dahlin P. Changes in the plant β-sitosterol/stigmasterol ratio caused by the plant parasitic nematode Meloidogyne incognita. Plants. 2021;10:292. https://doi.org/10.3390/plants10020292

70. Wang M, Xu X, Zhang X, Sun S, Wu C, Hou W, et al. Functional analysis of GmCPDs and investigation of their roles in flowering. PLoS One. 2015;10:e0118476. https://doi.org/10.1371/journal.pone.0118476

71. Ertekin M, Kirdar E. Effects of seed coat colour on seed characteristics of honeylocust (*Gleditsia triacanthos*). Afr J Agric Res. 2010;5(17):2434–8.

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.