Deciphering the transcriptional regulatory networks that control size, color, and oil content in Brassica rapa seeds

Yue Niu

Southwest University

Limin Wu

InnoTech Alberta

Yanhua Li

Chongqing Academy of Agricultural Sciences

Hualei Huang

Chongqing Academy of Agricultural Sciences

Mingchao Qian

Southwest University

Wei Sun

Southwest University

Hong Zhu

Southwest University

Yuanfang Xu

Southwest University

Yonghai Fan

Southwest University

Umer Mahmood

Southwest University

Benbo Xu

Yangtze University

Kai Zhang

Southwest University

Cunmin Qu

Southwest University

Jiana Li

Southwest University

Kun Lu

Southwest University

Research

Keywords: Brassica rapa, oil content, regulatory network, seed color, seed size

Abstract

Background Brassica rapa is an important oilseed and vegetable crop species and is the A subgenome donor of two important oilseed Brassica crops, Brassica napus and Brassica juncea. Although seed size (SZ), seed color (SC), and oil content (OC) substantially affect seed yield and quality, the mechanisms regulating these traits in Brassica crops remain unclear.

Results We collected seeds from a pair of B. rapa accessions with significantly different SZ, SC, and OC at seven seed developmental stages (every 7 days from 7 to 49 days after pollination), and identified 24,835 differentially expressed genes (DEGs) from seven pairwise comparisons between accessions at each developmental stage. K-means clustering identified a group of cell cycle-related genes closely connected to variation in SZ of B. rapa. A weighted correlation analysis using the WGCNA package in R revealed two important co-expression modules comprising genes whose expression was positively correlated with SZ increase and negatively correlated with seed yellowness, respectively. Upregulated expression of cell cycle-related genes in one module was important for the G2/M cell cycle transition, and the transcription factor Bra.A05TSO1 seemed to positively stimulate the expression of two CYCB1;2 genes to promote seed development. In the second module, a conserved complex regulated by the transcription factor TT8 appear to determine SC through downregulation of TT8 and its target genes TT3, TT18, and ANR. Further, upregulation of genes involved in triacylglycerol biosynthesis and storage in the seed oil body may increase OC. We further validated the accuracy of the transcriptome data by quantitative real-time PCR of 15 DEGs. Finally, we used our results to construct detailed models to clarify the regulatory mechanisms underlying variations in SZ, SC, and OC in B. rapa.

Conclusions This study based on transcriptome comparison provides insight into the regulatory mechanisms underlying the variations of SZ, SC, and OC in plants. The findings hold great promise for improving seed yield, quality and OC through genetic engineering of critical genes in future molecular breeding.

Background

Seeds are the most important harvested organ in many crop plants, storing large amounts of essential substances used by humans, and contain genetic materials that could be potentially be used to improve crops. In angiosperms, seed development is initiated by a double fertilization event in which one of the two spermatids unites with an egg cell, while the other fertilizes the diploid central cell to form a triploid endosperm [1]. In monocotyledons and some dicotyledonous species, the endosperm occupies most of the mature seed, but in most dicotyledons, the embryo consumes a large amount of endosperm and occupies the bulk of the mature seed [2].

Seed size (SZ) is an important factor in crop yield, not only directly affecting the yield of seed crops, but also positively influencing seed germination and seedling growth in all species [2]. To date, several pathways regulating this significant agronomic trait have been identified and characterized. In the HAIKU

(IKU) pathway, a classical SZ pathway controlling endosperm development, the triple mutant iku1/iku2/mini3 of Arabidopsis thaliana (Arabidopsis) shows reduced SZ due to precocious cell formation of the endosperm [3, 4]. Furthermore, the MINI3 pathway directly regulates CYTOKININ OXIDASE 2 (CKX2), inhibiting the activity of functional cytokinin (CK) to maintain endosperm growth [5]. SHORT HYPOCOTYL UNDER BLUE 1 (SHB1) is recruited by the promoters of IKU2 and MINI3 to promote their transcription. CKX2 is also regulated epigenetically [5, 6]. In the ubiquitin-proteasome pathway, the ubiquitin receptor DA1 acts synergistically with the E3 ubiquitin ligases DA2 or BIG BROTHER (BB)/ENHANCER OF DA1 (EOD1) to degrade various substrates and thereby reduce SZ by restricting cell proliferation [7]. In addition, DA1 functions synergistically with unknown E3 ubiquitin ligases to ubiquitinate UBIQUITIN-SPECIFIC PROTEASE 15 (UBP15), encoded by SUPPRESSOR OF DA1 (SOD2), resulting in small seeds and organs [6]. Overexpression of two Ara transcription factors (TFs), TRANSPARENT TESTA GLABRA2 (TTG2), and KLU/CYP78A5, results in larger seeds due to positive regulation of cell expansion and cell proliferation in Arabidopsis [8]. In rice (Oryza sativa), the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (OsSPL13) TF promotes cell expansion and cell elongation to augment grain length and weight [9]. By contrast, APETALA2 (AP2) TFs reduce SZ through negative control of cell expansion [10]. In addition, plant hormones such as brassinosteroid (BR) and CK also play key roles in SZ variation. Overexpression of genes involved in BR biosynthesis leads to bigger seeds in Arabidopsis and rice, whereas mutants of BR biosynthesis genes show reduced SZ [11]. In Arabidopsis, the simultaneous silencing of the CK response factors AHK2, AHK3, and CRE1 produces larger but fewer seeds [12]. In Brassica crops, however, although several pathways associated with SZ variation have been identified, the molecular mechanism underlying this trait, in terms of cell cycle and cell division pathways, remains largely unknown.

Yellow seed is a desirable trait of Brassica species due to the higher oil and protein content, but lower fiber content compared with dark seed [13]. Oxidized procyanidins are the key pigments in seed color (SC) formation in plants, and they are secondary products in the flavonoid and phenylpropane biosynthesis pathways [14]. Numerous TRANSPARENT TESTA (TT) genes, which participate in various stages of flavonoid synthesis, have been identified in A. thaliana [15, 16]. A complex structure, composed of MYB, basic helix−loop−helix (bHLH) TFs, and WD40 proteins, has been identified as a crucial factor in determining SC by regulating the flavonoid biosynthesis pathway [17]. Although several homologous TT genes have been identified in Brassica crops, the regulatory mechanism of SC formation is still not particularly clear [18, 19], retarding the progress of yellow-seeded breeding of Brassica crops. In addition, Brassica species have been widely cultivated as oilseed crop, such as Brassica napus, Brassica juncea, and Brassica rapa. Triacylglycerols (TAG), as the main form of seed oil, not only play an important role in edible oil, but also can be used as a biofuel or biodiesel for industrial production [20]. Since 1961, when the TAG biosynthesis pathway was first proposed by Kennedy, the synthesis of glycerol has become a hallmark of lipid biochemistry. In A. thaliana, the TAG biosynthetic, metabolic, and degradation pathways have been studied comprehensively [21], and the genes encoding most of the enzymes involved in TAG metabolism have been and cloned characterized [22]. However, the molecular mechanism underlying

fatty acid (FA) accumulation in seeds of Brassica crops remains poorly understood, resulting in a slow improvement of breeding for high oil content (OC) in Brassica crops.

B. rapa (AA, 2n = 20) is an important oil and vegetable Brassica crop that is widely cultivated worldwide [23]. It is also the diploid parental species of two important oilseed Brassica crops, B. napus (AACC, 2n = 38) and B. juncea (AABB, 2n = 36) [24]. Therefore, unravelling the regulatory mechanisms governing seed development and SZ in B. rapa would lay a solid foundation for breeding high-yield Brassica crops and related plants. Hence, we employed a pair of B. rapa accessions with significant differences in SZ, SC, and OC to analyze transcriptomic variations of seeds at seven seed developmental stages. These results provide insight not only into the regulatory mechanism influencing SZ by regulating cell cycle progression, but also into the molecular mechanisms of SC formation and lipid synthesis, which are regulated by flavonoid and oil body synthesis pathways, respectively.

Results

Phenotypic variations between two B. rapa accessions

The two B. rapa accessions were remarkably different in both SC and SZ after harvest (Fig. 1a,b; Additional file 1: Table S1). The SC of B. rapa accession SWUK4 was yellow, whereas that of accession SWUK3 was black (Fig. 1a). The final SZ of SWUK4 seed (mean ± standard error of the longest dimension, 2.42 ± 0.05 mm) was significantly higher than that of SWUK3 (1.23 ± 0.01 mm), representing an approximately two-fold difference. Moreover, the fresh seeds of SWUK4 were visibly larger than those of SWUK3 at all seven sampling stages (Fig. 1b).

The speed of seed development is an important factor affecting the seed-filling process in B. rapa. The SZ of accession SWUK3 showed the greatest increase at 21−28 days after pollination (DAP) (0.75 ± 0.004 mm), followed by 7−14 DAP (0.4 ± 0.006 mm), whereas that of SWUK4 varied most at 21−28 DAP (0.94 ± 0.009 mm), followed by 28−35 DAP (0.63 ± 0.002 mm) (Fig. 1c), suggesting that the critical stage for SZ formation was 21−28 DAP in both B. rapa accessions, and the following stage, 28−35 DAP, was also quite important for the larger-seed accession.

The silique length (SL) and thousand-seed weight (TSW) of the two B. rapa accessions were also significantly different. The SL of SWUK4 (40.90 ± 1.14 mm) was 1.6-fold longer than that of SWUK3 (25.00 ± 0.36 mm) (Fig. 1d), and the TSW of SWUK4 (7.06 ± 0.14 g) was approximate 2.9-fold greater than that of SWUK3 (2.42 ± 0.05 g) (Fig. 1e). Likewise, near-infrared reflection spectroscopy (NIRS) assays indicated that seed OC was significantly higher in SWUK4 than in SWUK3 (Fig. 1f). Gas chromatography (GC) measurements also identified differences in the abundances of the FA components (palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid) between the two accessions (Additional file 1: Table S1).

Transcriptome sequencing analysis

To generate transcriptional networks and identify key regulatory genes regulating SZ, SC, and OC in B. rapa, we collected 42 seed samples from two distinct accessions at seven seed developmental stages. Transcriptome sequencing generated a total of 320 Gb of raw data, including 1,076.28 million reads, averaging 7.62 G and 25.63 million reads per sample (Additional file 1: Table S2). Filtering of low-quality and contaminant reads left 1,034 million clean reads, with an average of 24.64 million reads per sample. The G + C content of all samples varied from 44.39−51.51%, and the range of Q30 was from 91.12−94.28%. After mapping, three samples (one sampled from SWUK3 at 14 DAP, the rest sampled from SWUK4 at 14 and 21 DAP) were removed because they had unique mapping rates of < 70%. This resulted in an average unique mapping rate of 87.27% (Additional file 1: Table S2), indicating that the quality of our transcriptome sequencing reads was good enough for further regulatory network construction and identification of DEGs.

We eliminated one sample from SWUK4 at 49 DAP from further analysis because the correlation coefficient calculated for two of its biological replicates was only ~ 0.62, suggesting that this sample may have been contaminated. The remaining 38 samples all showed high reproducibility ($r^2 > 0.9$) among biological replicates at each seed sampling stage (Additional file 2: Fig. S1). The correlations among the samples harvested at different seed developmental stages were much lower than those among biological replicates. Furthermore, the samples could be divided into two distinct groups in both B. rapa accessions —samples collected at 7, 14, and 21 DAP grouped together, whereas those from the other four time points formed a second group—suggesting that there might be a clear transition event at 21−28 DAP during seed development in B. rapa (Additional file 2: Fig. S1b,d).

Identification of DEGs

Comparing the two B. rapa accessions at the same seven seed developmental stages revealed a total of 24,835 unique DEGs (Fig. 2, Additional file 1: Table S3). The number of DEGs rose over the course of seed development, suggesting that the variation in seed transcriptome between the two accessions gradually increased from seed germination to maturation (Fig. 2). The highest number of DEGs, 14,725, was observed at 49 DAP, and the number of downregulated DEGs was higher than that of upregulated DEGs at six stages, with the exception being 49 DAP (Fig. 2).

An analysis of TFs, transcription repressors (TRs), and kinases using the online program iTAK identified a total of 3,053 TFs, 602 TRs, and 1,500 kinase genes in B. rapa (Additional file 1: Table S4). Among the DEGs, we detected 1,867 unique differentially expressed TFs, indicating that the transcription of TFs (P = 2.48E-12) was regulated to a greater degree during seed development than was that of other genes, based on a chi-square test of the ratio between differentially expressed TFs and DEGs versus that between all TFs and all B. rapa genes. Among the TFs in B. rapa, the AP2/ERF-ERF, MYB, and bHLH families seem to play relatively important roles in seed development, as 169, 136, and 130 members of these families, respectively, were differentially expressed between the two accessions. Similar to TFs, the expression of 889 unique kinase genes and 355 unique TRs was also readily reprogrammed during seed development, but the significance levels (P = 6.33E-5 for kinase genes and P = 0.04 for TRs) determined using chi-square tests were lower than those for TFs.

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genome (KEGG) enrichment analyses of DEGs

Enrichment analysis of DEGs identified several interesting GO and KEGG terms associated with genes that were differentially expressed during B. rapa seed development (Additional file 1: Table S5,6). The term "photosynthesis" (GO:0015979) was identified in upregulated DEGs at all of the seed developmental stages except for at 28 DAP. Upregulated genes were also enriched in the photosynthesis-related GO terms "electron transport chain" (GO:0022900) and "oxidation reduction" (GO:0055114) at five of the seven stages, suggesting that the synthesis of photosynthetic products in the seeds also might be an important factor affecting seed development and determining SZ in B. rapa. The GO terms "generation of precursor metabolites and energy" (GO:0006091), "ribosome biogenesis" (GO:0042254), and "carbohydrate catabolic process" (GO:0016052) were overrepresented for upregulated genes identified in at least four stages, implying that energy supply might be another factor controlling SZ in B. rapa.

GO enrichment analysis revealed that the genes upregulated at 35, 42, and 49 DAP were significantly enriched in the term "fatty acid biosynthetic process" (GO:0006633), and also showed enrichment of FA and lipid biosynthesis/metabolism terms by KEGG enrichment analysis. Given that OC differed significantly between the two B. rapa accessions, we deduce that transcriptional variation of these genes may play a crucial role in FA accumulation in B. rapa. Another group of upregulated genes may result in larger seeds through mechanisms involving the cell cycle regulation and CK signaling pathway, as they showed enrichment in "response to cytokinin stimulus" (GO:0009735), "cell cycle" (GO:0007049), "cell division" (GO:0051301), "M phase" (GO:0000279), "M phase of mitotic cell cycle" (GO:0000087), and "regulation of cyclin-dependent protein kinase activity" (GO:0000079) at the middle to late stages of seed development.

Among the downregulated genes, important functions related to SC could be identified (Additional file 1: Table S5,6). The GO and KEGG enrichment analyses showed that "phenylpropanoid biosynthetic process" (GO:0009699) and "flavonoid biosynthesis process" (GO:0009813) were enriched at all stages of seed development according to both methods, except at 49 DAP in the GO enrichment analysis. These results indicated that a reduction of pigment biosynthesis and accumulation may be the cause for the yellow seeds of B. rapa accession SWUK4 (Fig. 1b).

Identification of expression patterns in DEGs by K-means clustering

To identify DEGs with similar expression patterns during seed development, we performed two independent K-means clustering tests for 24,836 DEGs in two accessions (Fig. 3a,b), and generated 12 optimal clusters for each accession. We identified several clusters with similar expression patterns between the two accessions, such as cluster 12 of SWUK3 and cluster 5 of SWUK4, both of which showed gene transcription levels that increased continuously with seed development (Fig. 3a,b). GO and KEGG enrichment analyses revealed that these genes were enriched in lipid storage, seed oil body

biogenesis, and related pathways, suggesting that oil body formation in B. rapa occurs mainly in the later stages of seed development.

A comparison of gene clusters in the two accessions revealed an interesting result in cluster 2 of SWUK3 and cluster 4 in SWUK4. Although the two clusters have different expression patterns, they were significantly enriched in similar functions involved in regulating the SZ, including cell cycle, cell division, and mitosis (Fig. 3c,d). In Arabidopsis, a total of 59 genes have been identified as core genes regulating the cell cycle and cell division. Based on a BLASTP analysis, we identified 101 homologs of these genes in B. rapa (Additional file 1: Table S7). Among these, the majority of positive regulatory genes showed similar patterns upon K-means clustering: they were highly transcribed at 7 and 28 DAP in SWUK3, but at 21−35 DAP in SWUK4 (Fig. 4e). Notably, the expression patterns were consistent with the respective phenotypic variations of SZ in the two accessions (Fig. 1c). Therefore, genes in the two K-means clusters might be key regulators controlling SZ, and merit further investigation.

Construction of co-expression networks

In crops, complex traits are generally regulated by several transcriptional networks. To identify the co-expression networks associated with our target traits, we used the R WGCNA software based on the fragments per kilobase per million mapped (FPKM) expression matrix and phenotypic data of four traits: developmental stage, SZ, SZ increase, and seed yellowness. The sample clustering and correlation coefficients revealed strong repeatability among the biological replicates, and no outliers needed to be removed (Additional file 2: Fig. S2a). The results were also in agreement with those from the calculation of correlation coefficients, indicating that the seed samples from 7−21 DAP were grouped together, whereas the samples from the remaining stages formed a second group, again suggesting that seed transcriptome reprogramming is closely related to seed developmental stage, but weakly affected by accession.

In the WGCNA pipeline, pickSoftThreshold calculation revealed that the optimal soft threshold was 18, where the fitting curve approached 0.9 (Additional file 2: Fig. S2b). Then, we used the automatic blockwiseModules network construction approach to identify co-expression modules (Additional file 2: Fig. S2c). This allows the visualization of modules by color scheme, showing genes that are highly correlated in the same color, and genes that are weakly correlated in different colors (Additional file 2: Fig. S2d). The module construction process demonstrated that our functional color modules were clearly divided. After merging of modules with similar expression pattern, this process produced a total of 15 color modules, each composed of genes with similar expression patterns over time (Additional file 2: Fig. S3a).

As shown in Additional file 2: Fig. S3a, the abovementioned four traits (developmental stage, SZ, SZ increase, and seed yellowness) showed significant correlation with different modules. After enrichment analysis of GO for each color module (Additional file 1: Table S8), the genes of the module defined as MEgreen, which had the highest correlation with SZ increase ($r^2 = 0.73$, P = 2E-7), were significantly

enriched in "cell cycle" (GO: 0007049), "cell division" (GO: 0051301), "regulation of cell cycle" (GO: 0051726), and "DNA replication" (GO: 006260) (Additional file 2: Fig. S3b). This indicated that the genes related to cell cycle participated in the positive regulation of seed growth rate. The functions of genes in the MEsalmon color module, which showed a significant negative correlation with seed coat transparency (r^2 = − 0.69, P = 1E-6), were enriched in the terms "flavonoid metabolic process" (GO: 0009812), "flavonoid biosynthetic process" (GO: 0009813), "phenylpropanoid metabolic process" (GO: 0009698), and "phenylpropanoid biosynthetic process" (GO: 0009699) (Additional file 2: Fig. S3c). This revealed that flavonoids negatively regulate yellow seed coat, such as that found in SWUK4. Furthermore, the MEbrown module, which is closely correlated with SZ (r^2 = 0.7, P = 1E-6), was enriched in "photosynthesis" (GO: 015979), "photosynthesis, light harvesting" (0009765), "lipid biosynthetic process" (GO: 0008610), and "fatty acid biosynthetic process" (GO: 0006633) (Additional file 2: Fig. S3d). This demonstrated that lipid synthesis occurs in the late stages of seed development in B. rapa, and requires photosynthesis to provide necessary materials and energy.

Co-expression modules regulating SZ and SC

Co-expression network construction indicated that a SZ increase was most highly correlated with the MEgreen module (r^2 = 0.73, P = 2E-7) (Additional file 2: Fig. S3a). The expression profiles of a large number of genes were highly correlated with both the module eigengene (average expression profile of module genes) and SZ in the MEgreen module (Fig. 4a). GO enrichment analysis indicated that the MEgreen module genes were significantly enriched in "cell cycle" and "cell division" (Additional file 2: Fig. S3b). A heatmap based on the expression levels of the MEgreen module genes revealed that the module eigengene of the MEgreen module was similar to the average expression profiles of cluster 3 in SWUK3 and cluster 4 in SWUK4 from the K-means clustering results (Fig. 3a,b, Fig. 4b), suggesting that the SZ of B. rapa was regulated mainly by genes related to the cell cycle and cell division.

To identify hub genes in our modules of interest, we first assessed gene connectivity (K_{within}) on the basis of the absolute value of Pearson's correlations. We then considered genes with the top 30% K_{within} in each module as hub genes of those modules. In the 800 genes of the MEgreen module, the K_{within} values ranged from 29.01 to 285.97. Based on the iTAK results, a total of 21 TF, 17 TR, and 36 kinase genes were identified in this module (Additional file 2: Fig. S4a). Two TF genes with high K_{within} values, Bra.A05TSO1 (BraA05g024430, K_{within} = 217.73) and Bra.A09GRAS (BraA09g015380, K_{within} = 201.97), were selected as hub genes of the MEgreen module. Their Arabidopsis orthologs are AtTSO1 (AT3G22780) and AtSCL28 (AT1G63100, a TF of the GRAS family), respectively. In the primary Bra.A05TSO1 network, Bra.A05TSO1 was directly co-expressed with 147 genes, including 2 TF (belonging to the E2F and B3 families, respectively), 7 TR, and 11 kinase genes (Fig. 4c). The primary network of Bra.A09GRAS contained 248 co-expressed genes, including 5 TF (2 B3, 1 GARP and 1 TUB family), 9 TR, and 15 kinase genes (Additional file 2: Fig. S4b).

Another module of interest was MEsalmon, which showed a significant negative correlation with seed yellowness (r^2 = − 0.69, P = 1E-6) (Additional file 2: Fig. S3a). The genes in this module were enriched in

the GO terms "flavonoid biosynthetic process," "regulation of flavonoid biosynthetic process," and "phenylpropanoid biosynthetic process" (Additional file 2: Fig. S3c). The gene expression levels in this module were clearly lower in SWUK4 than in SWUK3 at most of the sampling stages (Fig. 5a). These results suggest that the formation of yellow seeds in SWUK4 is most likely due to weak expression or silencing of flavonoid biosynthesis pathway genes. The co-expression network of the MMsalmon module comprised 91 genes, including 5 TF, 1 TR, and 2 kinase genes (Fig. 5b). The K_{within} values of these genes varied from 6.17 to 37.62. Two well-known phenylpropanoid biosynthetic pathway TF genes, Bra.A05TT2 (BraA05g008220, R2R3-MYB, K_{within} = 30.23) and Bra.A09TT8 (BraA09g028560, bHLH, K_{within} = 27.54), were identified as hub genes underlying SC formation (Fig. 5c, d).

Role of flavonoid pathway genes in SC formation

Because the abovementioned analysis suggested that the differences in SC and OC between the two B. rapa accessions might be caused by flavonoid pathway and FA biosynthesis genes, we performed a further expression comparison analysis of these two types of genes. All expression values were subjected to log_2(FPKM + 1) transformation and Z-score normalization.

For pathways involved in the formation of compounds contributing to SC (Fig. 6), there were 43 homologous genes in B. rapa (Additional file 1: Table S9). At the beginning of the pathway, the generation of p-cinnamoyl-CoA requires the three enzymes PAL, CH4, and 4CL, encoded by 14 genes in B. rapa (Additional file 1: Table S9). All of these genes showed significantly differential expression between the two accessions at one or more stages of seed development. Their expression patterns were similar, with expression increasing over time and being higher in the black-seeded accession SWUK3 at almost all stages. In the first step of flavonoid biosynthesis, the formation of p-cinnamoyl-CoA is catalyzed by CHS to produce the chalcone compound naringenin, which is then converted to its isomer naringenin in a reaction catalyzed by CHI. Next, F3H and F3'H catalyze reactions converting naringenin to dihydrokaempferol and then eriodictyol. Of the 12 genes encoding these four enzymes in B. rapa, only two (one CHI and one F3H gene) were not differentially expressed between the two accessions.

The downstream portion of flavonoid biosynthesis is divided into two branches. In one branch, dihydrokaempferol and eriodictyol are catalyzed by FLS to generate dihydrokaempferol and dihydroquercetin. Of the five genes encoding this enzyme, only one gene was not differentially transcribed between accessions. In the other branch of the pathway, dihydrokaempferol and dihydroquercetin are catalyzed to become (−)-epiafazelechin and (−)-epicatechin by the enzymes DFR, LDOX, and BAN, which are encoded by five B. rapa genes. Notably, four of these genes were silenced and the fifth showed dramatically decreased transcription in the yellow-seeded accession SWUK4 at all seed developmental stages, indicating that the mutation of epicatechin branch genes may block pigment accumulation in B. rapa seeds.

Role of lipid biosynthesis and accumulation genes in OC variation

In plants, lipid biosynthesis and accumulation in seeds can be divided into four steps (Fig. 7). (1) Pyruvate and other substances form C16-C18 FAs in the plastid. (2) FAs are transported into the cytoplasm, where they undergo elongation and desaturation of their carbon chains. (3) Various FAs and glycerol are catalyzed to synthesize triacylglycerols (TAG) and store them in the seed oil body, after which some are (4) degraded by products of the GDSL-type Seed Fatty Acid Reducer (SFAR) genes.

In the first step, genes shared similar expression patterns in both accessions, being more highly expressed at 28–49 DAP than at 7–21 DAP (Fig. 7). However, most genes encoding proteins involved in this step were generally more highly expressed in the higher-OC accession SWUK4 at each developmental stage, and could be found in cluster 12 of SWUK3 and cluster 5 of SWUK4 in K-means clustering, indicating that oil synthesis in B. rapa seed is mainly initiated at a later stage of seed development. In the second step, the four genes encoding the relevant catalytic enzymes (3-Ketoacyl-ACP Synthase II (KAS II), Stearoylacyl Carrier Protein Desaturase (SAD), Fatty Acid Export 1 (FAX1), and Long Chain Acyl-CoA Synthetase 9 (LACS9)) were more highly expressed in SWUK4 (Fig. 7), suggesting that SWUK4 seeds have a larger acyl-CoA pool than SWUK3 seeds. Differential expression of genes encoding Lysophosphatidylcholine Acyltransferase (LPCAT) and FA Desaturase 2/3 (FAD2/3) between the two accessions may cause the significant variation in FA components, including oleic acid, linoleic acid, and linolenic acid, in mature seeds detected by GC analysis (Additional file 1: Table S1). In the third step, TAG is formed in the endoplasmic reticulum catalyzed by four enzymes, glycerol-3-phosphate acyltransferase 9 (GPAT9), 1-acylglycerol-3-phosphate acyltransferase (LPAAT), phosphatidic acid phosphatase (PAP), and diacylglycerol acyltransferase (DGAT) [25]. The majority of genes encoding these four enzymes were more highly expressed in SWUK4 than in SWUK3 at stages 28–49 DAP, implying that greater TAG biosynthesis and storage in seed oil body might be the cause for the higher OC in SWUK4. In the fourth step, SFAR genes negatively regulate FA storage and seed oil body size, leading to FA degradation. B. rapa has 12 SFAR genes, but only two of them showed significant upregulation in the higher-OC accession SWUK4 at late sampling stages, suggesting that this accession does not adopt the strategy of negatively regulating SFAR activity to increase OC. However, whether the OC of accession SWUK4 could be increased by reducing FA degradation remains to be established.

qRT-PCR validation

The relative expression levels of the 15 key DEGs at the seven seed developmental stages were analyzed by qRT-PCR to assess the accuracy of the RNA-seq results. The relative expression patterns of DEGs tested were positively correlated with the fold change variations obtained from the RNA-seq results (Fig. 8). The correlation coefficients between qRT-PCR and RNA-seq for the 15 DEGs ranged from 0.75 to 0.92. Further comparison showed that all the lower correlation coefficients between the two approaches belonged to three flavonoid pathway genes, Bra.A01TT18, Bra.A01ANR, and Bra.A09TT3. Due to the silencing of the three genes in the yellow-seeded B. rapa accession, their FPKM values were zero, but had to be set to 0.001 for fold change calculation, leading to lower correlation coefficients. However, the remaining 12 genes associated with FA metabolism and cell cycle progression all showed better

correlation than the flavonoid pathway genes, demonstrating the reliability and accuracy of our RNA-seq results.

Discussion

Cell cycle-related genes may regulate SZ in B. rapa

B. rapa is a model plant for Brassica species. Understanding the molecular mechanism regulating B. rapa SZ can provide essential reference information for other Brassica crops. In this study, we investigated these mechanisms in two B. rapa accessions with significant differences in seed development progression and final SZ. WGCNA co-expression network analysis showed that one expression module, i.e., the MEgreen module, had the highest correlation with SZ increase (Additional file 2: Fig. S3a), and the genes in the module were significantly enriched in cell cycle-related functions. Using the K-means clustering method, we obtained two clusters with different expression profiles but similar functions in the cell cycle from two B. rapa accessions. A comparison between the WGCNA and K-means clustering results indicated that both methods identified as critical genes those associated with the cell cycle, cell division, and mitotic cell cycle, and that these were significantly correlated with traits of interest, suggesting that transcriptional regulators of cell cycle-related genes might be potential targets for improvement of seed yield in B. rapa and related Brassica crops.

Cell cycle regulation plays a crucial role in the growth and development of plant organs, affecting both cell size and cell number. We identified 101 cell cycle core genes in B. rapa, most of which were more highly expressed in the large-seeded accession SWUK4 than in the small-seeded accession SWUK3 (Fig. 3e). The cyclin-dependent kinases (CDKs) play a central role in cell cycle regulation, and their activity is also controlled by regulatory subunits, such as cyclins and E2F TFs [26].

Previous studies have shown that SZ is regulated by cell cycle-related genes. A genome-wide association study of Arabidopsis SZ variation identified a B1-type CDK gene, AtCYCB1;4, and revealed that AtCYCB1;4 overexpression plants produced larger seeds than wild-type plants [27]. Arabidopsis AtCYCB1;2 is responsible for stimulating the G_2/M transition and is critical for cell division cycle progression during seed development [28, 29]. Brassica rapa has four CYCB1 genes (one CYCB1;1, two CYCB1;2, and one CYCB1;3 genes), of which three, Bra.A02CYCB1;2, Bra.A03CYCB1;2, and Bra.A01CYCB1;3, showed higher expression levels in the large-seeded accession SWUK4 at seed developmental stages from 28−49 DAP, the fast-growing stages of seeds in B. rapa.

In addition, the expression patterns of the two CYCB1;2 genes were positively correlated with SZ increase, suggesting that CYCB1;2 might be a key regulatory target for SZ trait improvement in Brassica crops. Eight other CDK genes (including four CYCB, three CYCA, and one CDKB genes) were also differentially expressed between the two B. rapa accessions. Although they were not included in the co-expression networks, the relationship of their expression to that of genes in these networks is also worthy of further investigation given their importance in cell cycle regulation.

The conserved DREAM complex may control seed size as an upstream coordinator

In the past 20 years, an evolutionarily conserved complex known as the DREAM complex has been identified as a master coordinator of cell cycle-dependent gene expression in animals [30]. DREAM consists of a core complex of five members, LIN9, LIN37, LIN52, LIN 53 (RBBP4), and LIN54, called MuvB (multivulva class B proteins) or MIP (MYB-interacting proteins) in Caenorhabditis elegans and mammals [30]. Perturbations in DREAM complex regulation trigger a shift of the balance from quiescence towards proliferation and result in increased mitotic gene expression that is frequently observed in human cancers.

The conserved function of the DREAM complex has also been identified in plants. In Arabidopsis, the homolog of LIN54, AtTSO1, coordinates with MYB3R1 to activate or inhibit cell cycle-related genes, and regulates the development of the root apical meristem and shoot apical meristem [31]. LIN54 is a core subunit of the DREAM complex. Downregulating LIN54 expression disrupts the DREAM complex, thereby reducing the expression of G_2/M cell cycle genes [32].

In this study, we constructed a co-expression network with the hub TF Bra.A05TSO1, and identified the regulatory relationship between Bra.A05TSO1 and its 23 downstream target genes using PlantRegMap (Fig. 4c). Bra.A05TSO1 could bind cis-regulatory elements in the promoters of two cell cycle core genes, Bra.A02CYCB1;2 and Bra.A03CYCB1;2, to stimulate their expression (Fig. 4c), suggesting that the regulatory module TSO1-CYCB1;2 may contribute to the G_2/M transition so as to control the cell cycle progression and affect the final SZ in plants, reminiscent of the similar mechanism adopted by LIN54 in animals (Fig. 4d). Hence, it is likely that the function of the DREAM complex in cell cycle regulation is conserved between plants and animals.

MYB binding with MuvB is a key factor in DREAM complex regulation of the cell cycle. When the cell leaves the S phase and enters the G_2 phase, MYB binds the MuvB complex to activate genes required for mitosis [30]. However, B-MYB may also cause transcriptional suppression in mammals [28]. The Arabidopsis homologs of the Myb oncoprotein are the MYB3R (R1R2R3-MYB) TFs, which contain similar domains and also regulate the expression of G_2 and M phase cell cycle genes [29]. Among the five Arabidopsis MYB3Rs, MYB3R1 not only can be redundant with MYB3R4 in transactivating the expression of specific genes at the G_2 and M phases, including that of CYCB1 [26, 33], but also can act as a repressor that inhibits the expression of cell cycle-related genes, similarly to MYB3R3 and MYB3R5 [30].

In our study, we identified 11 members of the MYB3R gene family. Though we did not identify MYB3R genes in the MEgreen module, we did identify Bra.A05TSO1, Bra.A02CYCB1;2, and Bra.A01MYB3R1 (BraA01g005340) in cluster 6 of SWUK4 in the K-means clustering analysis, suggesting that the three genes have similar expression patterns in the larger-seeded accession. Hence, we propose a model of SZ regulation in B. rapa whereby the TF Bra.A05TSO1 positively regulates BraCYCB1;2 either directly or in

coordination with Bra.A01MYB3R1 (as in the DREAM complex in animals), and then promotes cell division to produce larger seeds (Fig. 4d).

TT8 might be a conserved key target for creating yellow-seeded Brassica crops

Yellow seed color is a desirable trait with great potential for improving oil and protein content and meal quality in Brassica crops. Though several yellow-seeded varieties of B. rapa, B. napus, B. juncea, and Brassica carinata have been bred, and the inheritance of SC traits in these varieties hasbeen studied, the molecular mechanism controlling the target trait remains to be unraveled. However, a similar trait regulated by the flavonoid pathway has been well characterized at the molecular level by utilizing transparent testa (tt) mutants, which carry mutations that hinder flavonoid accumulation and modify pigmentation in the Arabidopsis seed coat [34–36]

Our WGCNA identified a module, denoted MEsalmon, that is significantly negatively correlated with B. rapa SC (Additional file 2: Fig. S3a). The two TF genes Bra.A09TT8 and Bra.A08TT2 were identified as hub genes in this co-expression network (Fig. 5c, d). Previous studies had revealed that three TF regulatory genes, AtTT2, AtTT8, and TTG1 (encoding a WD40 protein), could form an MYB-bHLH-WD40 (MBW) complex that activates proanthocyanidin (PA)-specific genes (such as TT3, TT18, and ANR) during seed coat development [37]. In our results, two of the genes encoding MBW complex members, Bra.A08TT2 and Bra.A06TTG1, seem to be unimportant for SC formation in B. rapa, since their expression was either unregulated or higher in the yellow-seeded accession at one stage only. However, the product of the third MBW gene, Bra.A09TT8, may play a critical regulatory role in SC formation as a core TF in the flavonoid pathway. Our results indicated that Bra.A09TT8 expression in the yellow-seeded accession was very weak, and significantly lower than in the black-seeded accession, at all seed developmental stages except at 49 DAP. Therefore, it is likely that the accumulation of the PA precursor epicatechin might be blocked due to the silencing of Bra.A09TT8, consistent with a previous study showing that the insertion of a transposable element in BraTT8 caused a loss of function of BraTT8 in the yellow-seeded B. rapa var. yellow sarson [38]. Recent studies have also revealed that simultaneous natural mutations of two homologous TT8 genes in allotetraploid B. juncea resulted in the yellow-seeded trait [39], and a similar mutant also could be created in B. napus using CRISPR/Cas9 technology [40].

We also observed that the regulatory network of TT8 was well conserved in Arabidopsis and Brassica crops. In Arabidopsis seeds, the MBW complex AtTT8-AtTT2-AtTTG1 positively regulates PA synthesis by directly binding to the cis-regulatory elements of DFR, LODX, and BAN in the flavonoid biosynthesis pathway [41]. In the primary co-expression network of Bra.A09TT8, we identified three downstream target genes, Bra.A09TT3 (BraA09g019440), Bra.A01ANR (BraA01g029500), and Bra.A03ANR (BraA03g064730), that were almost silenced in the yellow-seeded B. rapa accession during seed development. Though two TT18 genes (Bra.A01TT18 and Bra.A03TT18) in B. rapa were not involved in the Bra.TT8-containing primary network, a positive regulatory relationship between TT8 and TT18 could be observed due to their similar weak expression in the yellow-seeded B. rapa accession. A similar network has also been identified in B. napus [40, 42, 43], Thus, we infer that the regulatory network of the

TT8-involved complex was evolutionarily conserved between Arabidopsis and Brassica crop species. Considering the divergence of the two genera, the conserved TT8 genes might be optimal target loci for manipulation in the breeding of yellow-seeded varieties of Brassica and related crops.

Conclusion

In this study, we identified 24,835 DEGs from transcriptome comparisons of seeds sampled from a pair of B. rapa accessions with different SZ, SC, and OC at seven seed developmental stages. Both K-means clustering and WGCNA identified a group of cell cycle-related genes whose expression was also positively correlated with SZ increase, and indicated that the TF Bra.A05TSO1 may positively stimulate the expression of two CYCB1;2 genes to increase the SZ through regulation of the G_2/M transition. In a module whose expression was negatively correlated with seed yellowness, a conserved TT8-involved complex may determine the SC through downregulation of the key TF gene TT8 and its targets TT3, TT18, and ANR in the flavonoid pathway. Upregulated genes involved in triacylglycerol biosynthesis and storage in the seed oil body may increase the OC of B. rapa. This study thus unravels the regulatory mechanisms underlying the variation of SZ, SC, and OC in B. rapa seeds, and may facilitate genetic engineering efforts to breed Brassica crops with improved yield and OC.

Materials And Methods

Plant materials

Seeds of two B. rapa accessions, SWUK3 (small black seeds with lower OC) and SWUK4 (large yellow seeds with higher OC), were obtained from the Chongqing Rapeseed Engineering Research Center. All seeds were sown at the beginning of October, 2018, and transferred into the field at Southwest University, Beibei, Chonqging, China (29°45' N, 106°22' E, 238.57 m), 1 month later. Each accession was planted in a plot of five rows, 10 plants per row, with 20 cm between plants within each row, and 30 cm between rows.

Phenotype measurements

Seeds were collected at seven seed developmental stages: 7, 14, 21, 28, 35, 42, and 49 DAP. Fifty seeds were harvested from the siliques on the main inflorescences at each stage to measure SZ. Based on the Biologische Bundesanstalt, Bundessortenamt, and CHemical (BBCH) industry scale [44], six representative plants from the middle of each plot were harvested at growth stage 99 (harvested product) for measurement of SZ, TSW, and SL. The OC and seed coat transparency were assayed using NIRS (DS2500) with previously established models for Brassica crops [45]. The fatty acid (FA) composition, specifically the compositions of palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid, was analyzed by gas-liquid chromatography on a Perkin Elmer Gas Chromatograph Model GC-2010 (Shimadzu, Kyoto, Japan) equipped with a fused silica capillary column DB-WAX (30 m × 0.25 mm id, 0.25 μm film thickness, J&W, Folsom, CA, USA), as reported previously [46]. For each trait, at least five independent biological replicates were measured.

Transcriptome sequencing and identification of DEGs

To compare the transcriptomic variations between the two B. rapa accessions, a total of 42 seed samples (2 accessions × 7 stages × 3 biological replicates) were collected. Total RNAs of abovementioned samples were isolated using an RNAprep Pure Plant Kit (Tiangen, Beijing, China), and sent to Novogene Corporation (Beijing, China) for library construction and transcriptome sequencing on an Illumina HiSeq 2500 platform. The raw sequencing data were deposited in the BIG Data Center (BIGD) under BioProject accession number PRJCA002339.

Low-quality reads, connectors, and barcode sequences were eliminated using Trimmomatic-0.39 [47] Then, the clean data were aligned to the B. rapa reference genome, v3.0 (http://brassicadb.org/brad/), using STAR-2.5.3 [48]. Gene expression levels were quantified as count number and FPKM using the programs featurecounts and cuffquant, respectively [49]. The correlation relationships among all the samples were examined by principal component analysis (PCA), and the correlation coefficients were determined through the R package ggfortify and SPSS15.0 [50]. DEGs were identified using the R package DEseq2, based on the criteria of false discovery rate (FDR, Benjamini–Hochberg multiple test correction) < 0.01 and absolute fold change > 2 [51].

Identification of transcription factors, and genes associated with SC and acyl-lipid metabolism

To identify and classify transcription factors (TFs), transcription repressors (TRs), and kinases, all the protein sequences of B. rapa were analyzed using the online program iTAK v1.6, (http://itak.feilab.net) [52].

Since the SC and OC differed significantly between the two B. rapa accessions, we obtained all of the genes involved in the flavonoid pathway that play critical roles in SC formation, as well as those involved in acyl-lipid metabolism, in Arabidopsis from the Arabidopsis Acyl-Lipid Metabolism database (ARALIP) (http://aralip.plantbiology.msu.edu) [53]. Subsequently, a reciprocal BLASTP with an E-value cut-off of 1E-5 was used to identify the homologous relationships between the Arabidopsis and B. rapa genes [54], and then the protein sequences were analyzed with Pfam Scan (https://www.ebi.ac.uk/Tools/pfa/pfamscan/) to further confirm the existence of corresponding functional domains.

K -means clustering of DEGs

K-means clustering is an effective approach to identify gene expression patterns for transcriptome data. To establish the expression profiles of DEGs with potential biological function in regulating our target traits in B. rapa, cluster analysis was performed by the K-means method with Pearson's correlation distance using the cluster package in R. Optimal number of clusters for K-means were estimated by the gap statistic that computed using the clusGap function in R package factoextra [55]. K-means clustering was then performed with the optimal cluster number 12 for both two B. rapa accessions. Heatmaps were

created using the expression values with $\log_2(\text{FPKM} + 1)$ and Z-score normalization and visualized using the pheatmap package in R.

Weighted gene co-expression network analysis (WGCNA)

To detect co-expression modules and key regulatory genes associated with target seed traits in B. rapa, we generated co-expression networks using the WGCNA package in R as previously described [56]. Briefly, only expressed genes with average FPMK values higher than 1 in any seeds were retained and were then $\log_2(\text{FPKM} + 1)$ transformed before further processing. The soft thresholding power was determined using pickSoftThreshold function based on the scale-free topology model fit (R^2) > 0.9. Then, the automatic blockwiseModules network construction approach was applied to obtain the highly correlated modules, with the following parameters: power, 18; TOM-type, unsigned; miniModuleSize, 50; maxBlockSize, 35,000; mergeCutHeight, 0.25. The Plant Transcriptional Regulatory Map (PlantRegMap, http://plantregmap.gao-lab.org) was used to analyze regulatory relationships between TFs and their co-expressed genes in the same network [57]. The co-expression and transcriptional regulatory networks were displayed using Cytoscape v3.5.1 [58].

GO and KEGG enrichment analyses

All the B. rapa genes were annotated with BLASTP against the Arabidopsis proteome (TAIR10) with an E-value cut-off of 1E-5 (Altschul et al., 1997). GO enrichment analysis was performed using the BiNGO plug-in in Cytoscape v3.5.1 [58]. Significantly overrepresented GO terms were identified with the threshold of FDR < 0.05. The online OmicShare tool (https://www.omicshare.com/tools), a free online platform, was used to perform KEGG pathway enrichment analysis. Bubble plots for GO terms and KEGG pathways were generated using the R package ggplot2 [59].

qRT-PCR validation

To confirm the accuracy of transcriptomic sequencing and identification of DEGs, the cDNA was synthesized from 1 μg of total RNA that used for transcriptomic sequencing using a PrimeScript RT Master Mix Kit (TaKaRa, Dalian, China). A total of 15 DEGs (3, 4, and 8 genes involved in flavonoid pathway, acyl-lipid metabolism, and seed development, respectively) were selected for qRT-PCR assays. All gene-specific primers were retrieved from the qPrimerDB database (https://biodb.swu.edu.cn/qprimerdb) [60]. All qRT-PCR assays were carried out following the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines. Bna.UBC21 and BnaACT7 were used as internal controls, and relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [24].

Abbreviations

AP2
APETALA2; ARALIP:Arabidopsis Acyl-Lipid Metabolism database; BB:BIG BROTHER; BIGD:BIG Data Center; BR:brassinosteroid; CK:cytokinin; CKX2:CYTOKININ OXIDASE 2; DAP:days after pollination;

DEG:differentially expressed gene; DGAT:Diacylglycerol Acyltransferase; EOD1:ENHANCER OF DA1; FA:fatty acid; FAD:FA Desaturase; FAX1:Fatty Acid Export 1; FPKM:fragments per kilobase per million mapped reads; GC:gas chromatography; GO:Gene Ontology; GPAT9:Glycerol-3-Phosphate Acyltransferase 9; GRF4:GROWTH-REGULATING FACTORS; IKU:HAIKU; KAS II:3-Ketoacyl-ACP Synthase II; KEGG:Kyoto Encyclopedia of Genes and Genomes; LACS9:Long Chain Acyl-Coa Synthetase 9; LPAAT:1-Acylglycerol-3-Phosphate Acyltransferase; LPCAT:Lysophosphatidylcholine Acyltransferase; MIQE:Minimum Information for Publication of Quantitative Real-Time PCR Experiments; NIRS:Near-Infrared Reflection Spectroscopy; OC:oil content; PAP Phosphatidic Acid Phosphatase; PCA:principal component analysis; qRT-PCR:quantitative real-time PCR; SAD:Stearoylacyl Carrier Protein Desaturase; SC:seed color; SFAR:Seed Fatty Acid Reducer; SHB1:SHORT HYPOCOTYL UNDER BLUE 1; SL:silique length; SOD2:SUPPRESSOR OF DA1; SPL13:SQUAMOSA PROMOTER BINDING PROTEIN-LIKE; SZ:seed size; TAG:triacylglycerols; TF:transcription factor; TR:transcription repressor; TSW:thousand-seed weight; TT:TRANSPARENT TESTA; UBP15:UBIQUITIN-SPECIFIC PROTEASE 15; WGCNA:Weighted Gene Co-expression Network Analysis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files.

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by grants from the National Key Research and Development Plan (2018YFD0100500 and 2016YFD0101007), the National Natural Science Foundation of China (31871653), the 111 project (B12006); the Natural Science Foundation of Chongqing, China (cstc2018jcyjA1219), and the Chongqing Agricultural Development Fund (NKY-2019AB008-2).

Authors' contributions

KL and LW conceived and designed the experiments; YN, YL and MQ performed the sampling and experiments; HH, HZ, and YX measured the phenotypic traits; WS, YF, UM, BX, and KZ contributed to data processing, analysis, interpretation, and discussion; KL and YN wrote the manuscript; CQ and JL contributed to discussion and helped with manuscript reviewing. All authors reviewed and approved the manuscript.

Acknowledgements

The authors are grateful to the reviewers and editors for their constructive review and suggestions for this paper.

References

1. Syrenne RD, Shi W, Stewart CN, Yuan JS. Omics platforms: Importance of twenty-first century genome-enabled technologies in seed developmental research for improved seed quality and crop yield. In: Agrawal G, Rakwal R. Seed development: OMICS technologies toward improvement of seed quality and crop yield. Dordrecht: Springer; 2012. p. 43–57.

2. Sundaresan V. Control of seed size in plants. Proc Natl Acad Sci USA. 2005;102: 17887–17888.

3. Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A. MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proc Natl Acad Sci USA. 2005;102:17531–17536.

4. Wang A, Garcia D, Zhang H, Feng K, Chaudhury A, Berger F, Peacock WJ, Dennis ES, Luo, M. The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. Plant J. 2010;63:670–

5. Li J, Nie X, Tan JLH, Berger F. Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. Proc Natl Acad Sci USA. 2013;110:15479–15484.

6. Li N, Li Y. Signaling pathways of seed size control in plants. Curr Opin Plant Biol. 2016;33:23–32.

7. Xia T, Li N, Dumenil J, Li J, Kamenski A, Bevan MW, Gao F, Li Y.The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. Plant Cell. 2013;25:3347–59.

8. Doughty J, Aljabri M, Scott RJ. Flavonoids and the regulation of seed size in Arabidopsis. Biochem Soc T. 2014;42(2):364–

9. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, Zhou T, Lu T, Zhu J, Shangguan, Y. OsSPL13 controls grain size in cultivated rice. Nat Genet. 2016;48:447.

10. Ohto M, Floyd SK, Fischer RL, Goldberg RB, Harada JJ. Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. Sex Plant Reprod. 2009;22:277–289.

11. Zuo J, Li J. Molecular genetic dissection of quantitative trait loci regulating rice grain size. Annu Rev Genet. 2014;48:99–118.

12. Riefler M, Novak O, Strnad M, Schmülling T. Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. Plant Cell. 2006;18:40–54.

13. Tang ZL, Li JN, Zhang XK, Chen L, Wang R. Genetic variation of yellow-seeded rapeseed lines (*Brassica napus*) from different genetic sources. Plant Breed. 1997;116:471–474.

14. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 2001;126(2):485–

15. Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec L. Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta. 2006;224(1):96–

16. Appelhagen I, Thiedig K, Nordholt N, Schmidt N, Huep G, Sagasser M, Weisshaar B. Update on *transparent testa* mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. Planta. 2014;240(5):955–970.

17. Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A. A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development. 2003;130: 4859–4869.

18. Wang Y, Xiao L, Dun X, Liu K, Du D. Characterization of the *BrTT1* gene responsible for seed coat color formation in Dahuang (*Brassica rapa L*. landrace). Mol Breed. 2017;37(11):137.

19. Zhang J, Lu Y, Yuan Y, Zhang X, Geng J, Chen Y, Cloutier S, McVetty PBE, Li G. Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in *Brassica rapa*. Plant Mol Biol. 2009; 69: 553–563.

20. Kalscheuer R, Steinbüchel A. A novel bifunctional wax ester synthase/acyl-CoA: diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* J Biol Chem. 2003;278.

21. Sébastien B, Lepiniec L. Regulation of *de novo* fatty acid synthesis in maturing oilseeds of Arabidopsis. Plant Physiol Biochem. 2009;47(6):448–455.

22. Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, Franke RB, Graham IA, Katayama K, Kelly AA, Larson T, Markham JE, Miquel M, Molina I, Nishida I, Rowland O, Samuels L, Schmid KM, Wada H, Welti R, Xu C, Zallot R, Ohlrogge J. Acyl-lipid metabolism. Arabidopsis 2013;11:e0161.

23. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z, *Brassica rapa* Genome Sequencing

Project C. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 2010;43(10):1035–1039.

24. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, Zhang C, Chen Z, Xiao Z, Jian H, Cheng F, Zhang K, Du H, Cheng X, Qu C, Qian W, Liu L, Wang R, Zou Q, Ying J, Xu X, Mei J, Liang Y, Chai YR, Tang Z, Wan H, Ni Y, He Y, Lin N, Fan Y, Sun W, Li NN, Zhou G, Zheng H, Wang X, Paterson AH, Li J. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nat. Commun. 2019;10: 1154.

25. Chapman KD, Ohlrogge JB. Compartmentation of triacylglycerol accumulation in plants. J Biol Chem. 2012;287: 2288–2294.

26. De Veylder L, Joubès J, Inzé D. Plant cell cycle transitions. Curr Opin Plant Biol. 2003;6: 536–543.

27. Ren D, Wang X, Yang M, Yang L, He G, Deng XW. A new regulator of seed size control in Arabidopsis identified by a genome-wide association study. New Phytol. 2019;222:895–906.

28. Schnittger A, Schöbinger U, Stierhof Y-D, Hülskamp M. Ectopic B-type cyclin expression induces mitotic cycles in endoreduplicating Arabidopsis Curr Biol. 2002;12:415–420.

29. Weingartner M, Pelayo HR, Binarova P, Zwerger K, Melikant B, de la Torre C, Heberle-Bors E, Bögre L. A plant cyclin B2 is degraded early in mitosis and its ectopic expression shortens G2-phase and alleviates the DNA-damage checkpoint. J Cell Sci. 2003;116:487–498.

30. Sadasivam S, DeCaprio JA. The DREAM complex: master coordinator of cell cycle-dependent gene expression. Nat Rev Cancer. 2013;13:585–595.

31. Wang W, Sijacic P, Xu P, Lian H, Liu Z. Arabidopsis TSO1 and MYB3R1 form a regulatory module to coordinate cell proliferation with differentiation in shoot and root. Proc Natl Acad Sci USA. 2018;115:E3045–E3054.

32. Schmit, F, Cremer S, Gaubatz S. LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdc2 promoter in a sequence-specific manner. FEBS J. 2009;276:5703–5716.

33. Fischer M, DeCaprio JA. Does *Arabidopsis thaliana* DREAM of cell cycle control? EMBO J. 2015;34:1987–1989.

34. Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM. Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. Plant J. 1995;8:659–671.

35. Peer WA, Brown DE, Tague BW, Muday GK, Taiz L, Murphy AS. Flavonoid accumulation patterns of *transparent testa* mutants of Arabidopsis. Plant Physiol. 2001;126:536–548.

36. Zhang B, Schrader A. TRANSPARENT TESTA GLABRA 1-dependent regulation of flavonoid biosynthesis. Plants. 2017;6:65.

37. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015;20:176–185.

38. Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J, Yi B, Ma C, Shen J, Tu J. A large insertion in bHLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. PLoS ONE. 2012;7(9):e44145.

39. Padmaja LK, Agarwal P, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK. Natural mutations in two homoeologous TT8 genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). Theor Appl Genet. 2014;127:339–347.

40. Zhai Y, Yu K, Cai S, Hu L, Amoo O, Xu L, Yang Y, Ma B, Jiao Y, Zhang C. Targeted mutagenesis of BnTT8 homologs controls yellow seed coat development for effective oil production in *Brassica napus* Plant Biotechnol J. 2019; doi:10.1111/pbi.13281

41. Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, Lepiniec L, Dubos, C. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB–bHLH–WDR complexes and their targets in Arabidopsis seed. New Phytol. 2014;202:132–144.

42. Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J. Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. Front Plant Sci. 2017;8:1674.

43. Qu C, Fu F, Lu K, Zhang K, Wang R, Xu X, Wang M, Lu J, Wan H, Zhanglin T, Li J. Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. J Exp Bot. 2013;64:2885–2898.

44. Lancashire PD, Bleiholder H, Boom TVD, Langelüddeke P, Stauss R, Weber E, Witzenberger A. A uniform decimal code for growth stages of crops and weeds. Ann Appl Biol. 1991;119:561–

45. Liu L, Qu C, Wittkop B, Yi B, Xiao Y, He Y, Snowdon RJ, Li J. A high-density snp map for accurate mapping of seed fibre QTL in *Brassica napus* PLoS ONE. 2013;8:1–9.

46. Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, Xu X, Liang Y, Li S, Wang R, Li J. Genome-wide association mapping and Identification of candidate genes for fatty acid composition in *Brassica napus* using SNP markers. BMC Genom. 2017;18:232.

47. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120.

48. Dobin, A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

49. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–578.

50. Tang Y, Horikoshi M, Li W. ggfortify: unified interface to visualize statistical results of popular r packages. R J. 2016;8:474–485.

51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B. 1995;527: 289–300.

52. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei Z. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol Plant. 2016;9:1667–1670.

53. McGlew K, Shaw V, Zhang M, Kim RJ, Yang W, Shorrosh B, Suh MC, Ohlrogge J. An annotated database of Arabidopsis mutants of acyl lipid metabolism. Plant Cell Rep. 2015;34:519–532.

54. Altschul SF, Madden, TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–3402.

55. Kassambara A, Mundt F. Factoextra: extract and visualize the results of multivariate data analyses. R Packag. Version. 2017;1(4):2017.

56. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

57. Tian F, Yang DC, Meng YQ, Jin J, Gao G. PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res. 2020;48:D1104–D1113.

58. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics. 2011;27:431–432.

59. Ginestet C. ggplot2: elegant graphics for data analysis. J R Stat Soc Ser A Stat. 2011;174:245–246.

60. Lu K, Li T, He J, Chang W, Zhang R, Liu M, Yu M, Fan Y, Sun W, Ma J, Qu C, Liu L, Li N, Liang Y, Wang R, Tang Z, Qian W, Xu X, Lei B, Zhang K, Li J. qPrimerDB: a thermodynamics-based gene-specific qPCR primer database for 147 organisms. Nucleic Acids Res. 2018;46:D1229–D1236.

Figures
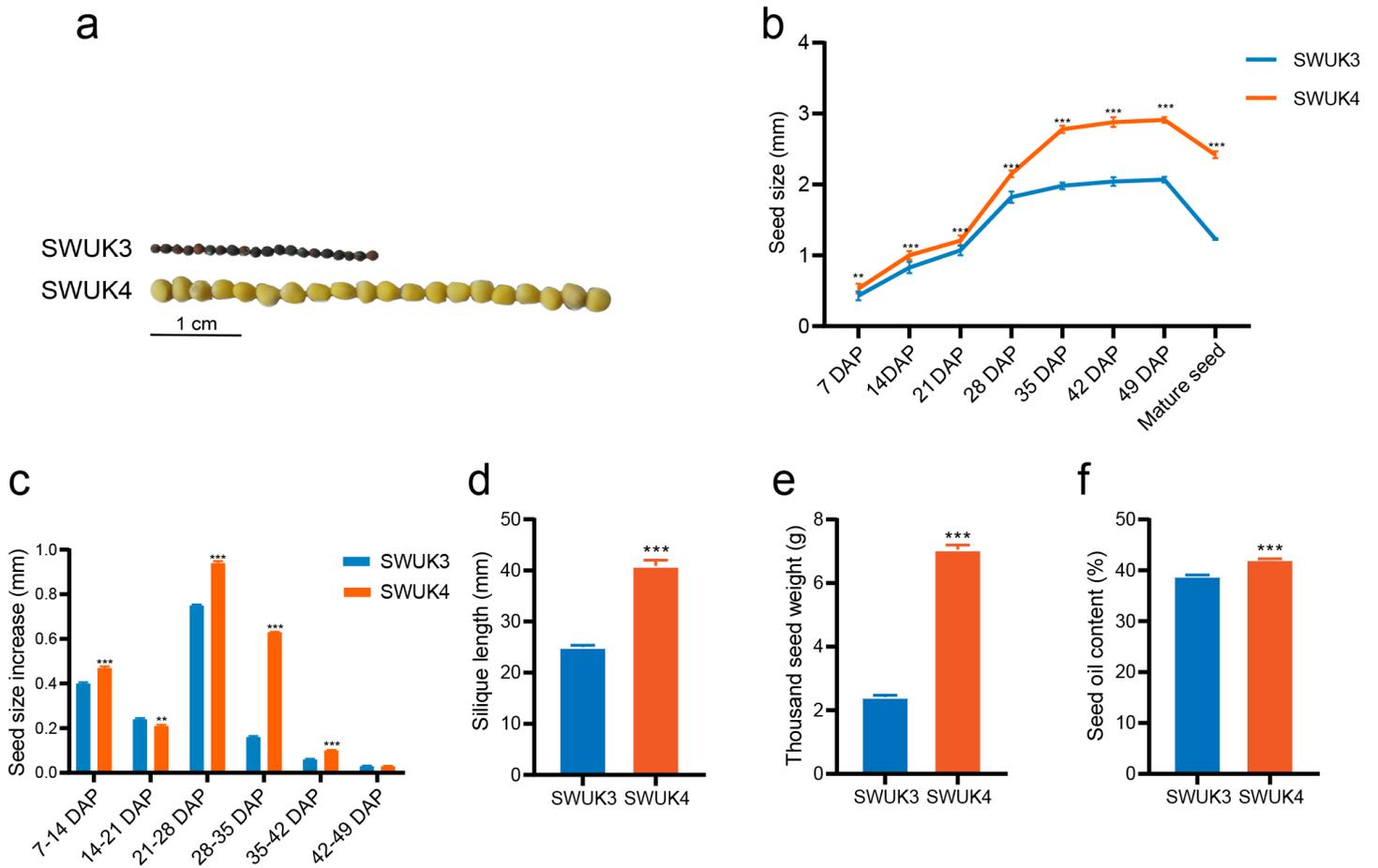
Figure 1

Phenotypic comparison between two B. rapa accessions. (a) Comparison of SZ and SC in SWUK3 and SWUK4. (b) Comparison of SZ of seeds sampled at seven seed developmental stages and dry mature seeds. (c) Comparison of SZ increase between two adjacent sampling stages. (d) Comparison of SL. (e) Comparison of TSW. (f) Comparison of seed OC. Values are the means ± standard error (SE) of three biological replicates. Student's t-test was used for statistical analysis of data from the two B. rapa accessions (**, P < 0.01; ***, P < 0.001).
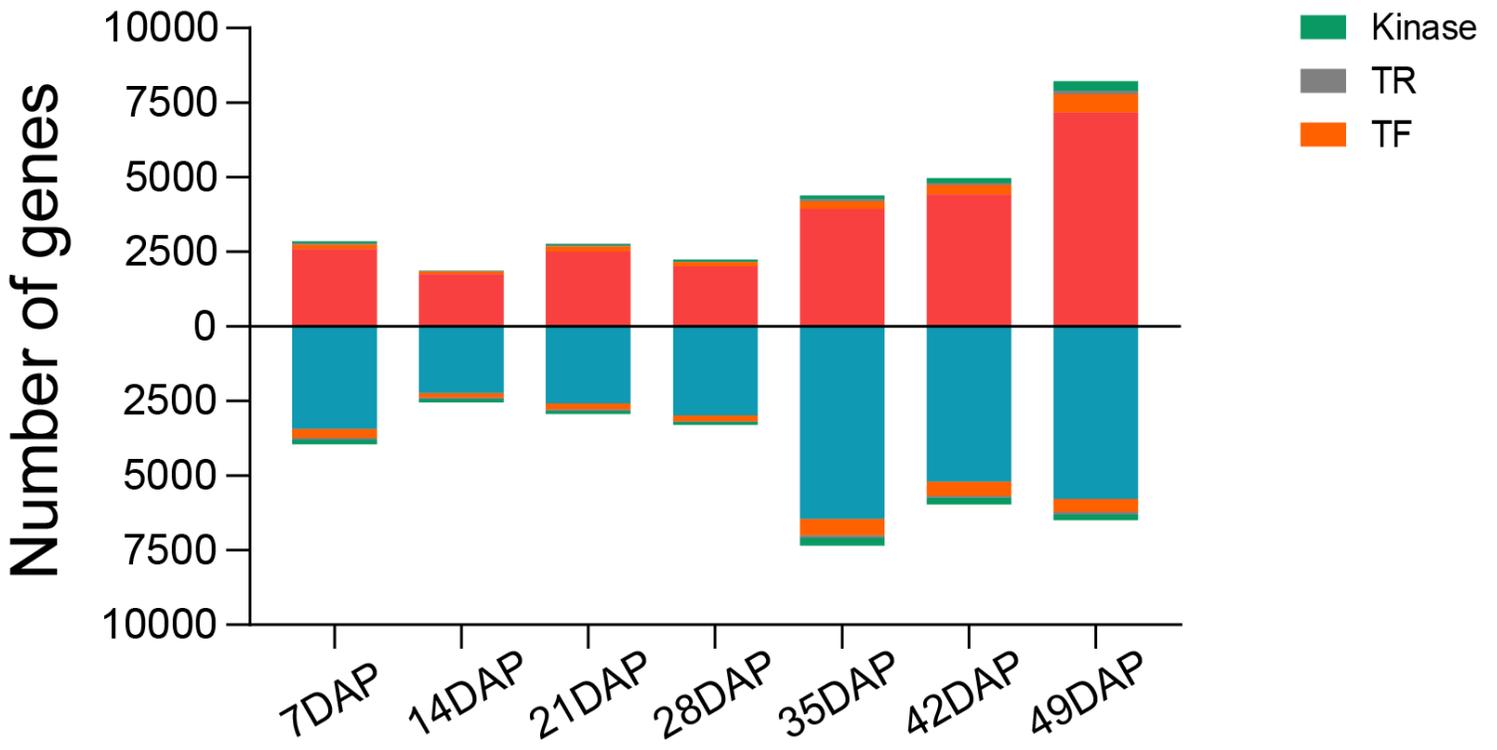
Figure 2

Number of DEGs identified at each of seven seed developmental stages. Classification of TF, TR, and kinase genes was performed using iTAK v1.6, (http://itak.feilab.net).
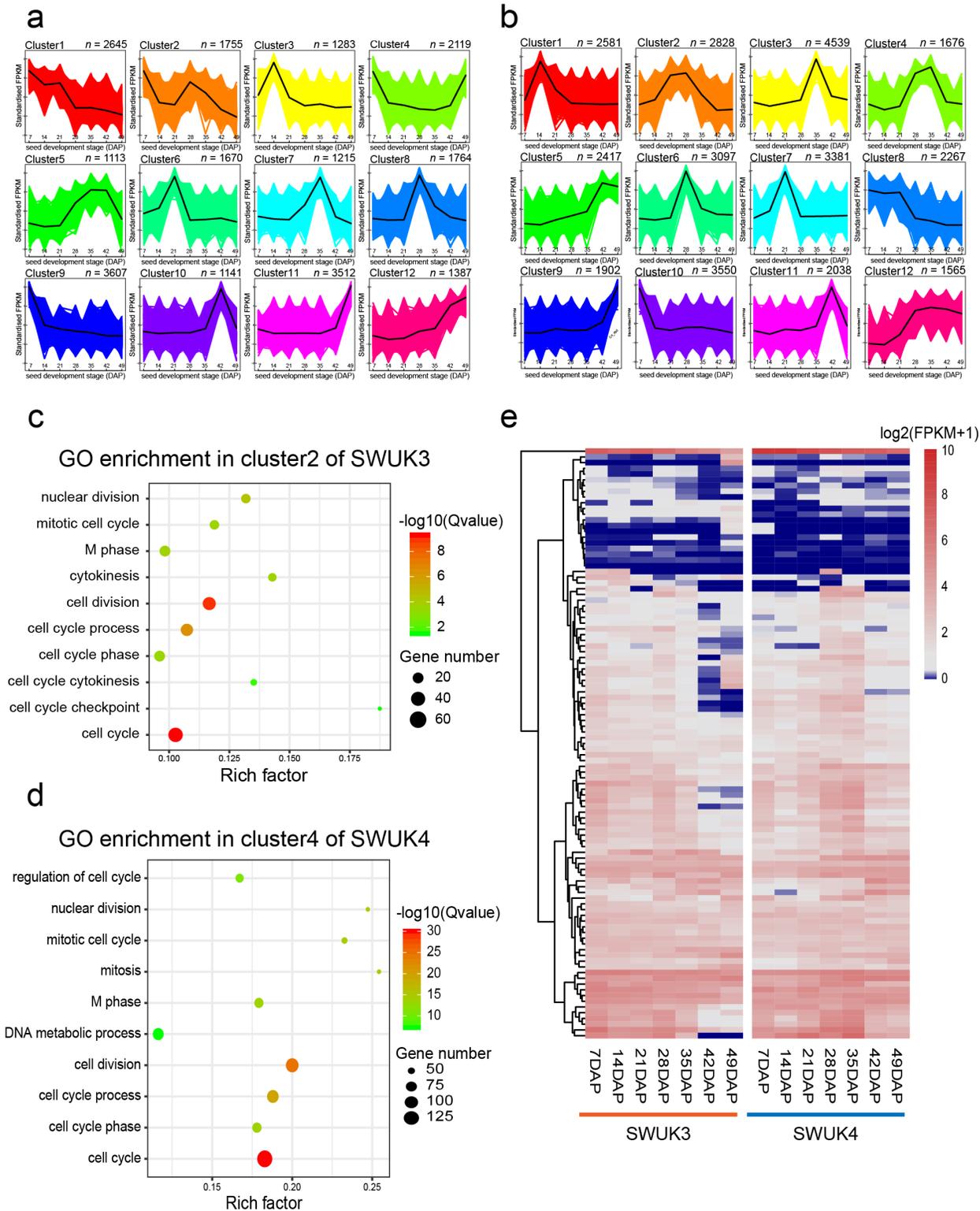
Figure 3

K-means clustering of DEGs. (a,b) Twelve clusters of (a) SWUK3 and (b) SWUK4 based on FPKM values at seven seed developmental stages. (c,d) GO enrichment analysis results for genes in (c) cluster 2 of SWUK3 and (d) cluster 4 of SWUK4. (e) Heatmap of expression of cell cycle core genes in B. rapa. Gene expression levels were transformed with the log2(FPKM+1). All genes are listed in Additional file 1: Table S7.
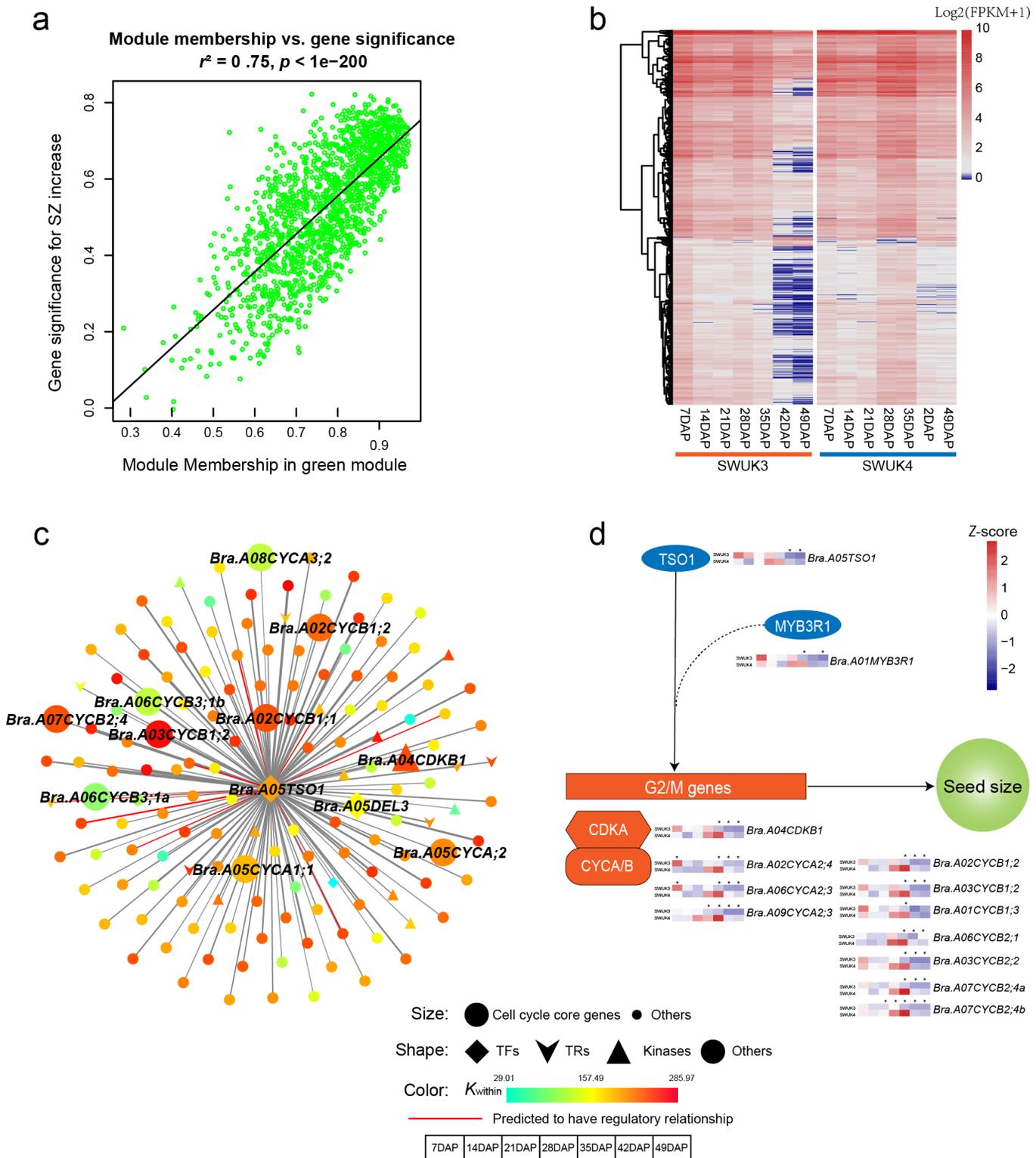
Figure 4

The WGCNA MEgreen module is significantly associated with SZ formation. (a) Scatter plot of the correlation of module membership (correlation coefficients between genes with MEgreen module) and gene significance (correlation coefficients between genes with trait of SZ increase). (b) Heatmap of MEgreen module genes in two B. rapa accessions, which was displayed based on log2(FPKM+1). (c) Primary co-expression network of Bra.A05TSO1. Square, down arrow, triangle, and disc represent TF, TR,

kinase, and other genes, respectively. The edge width represents the weight value between the two nodes: the higher the value of the weight between the nodes, the wider the edge. The regulatory relationships between the two nodes obtained from PlantRegMap are represented by red edges. (d) The regulatory model of TSO1 controlling seed size, with the heatmap of genes profiles in the two accessions. TSO1 play an important role in regulating the genes of G2/M phase, and MYB3R1 may bind to TSO1.
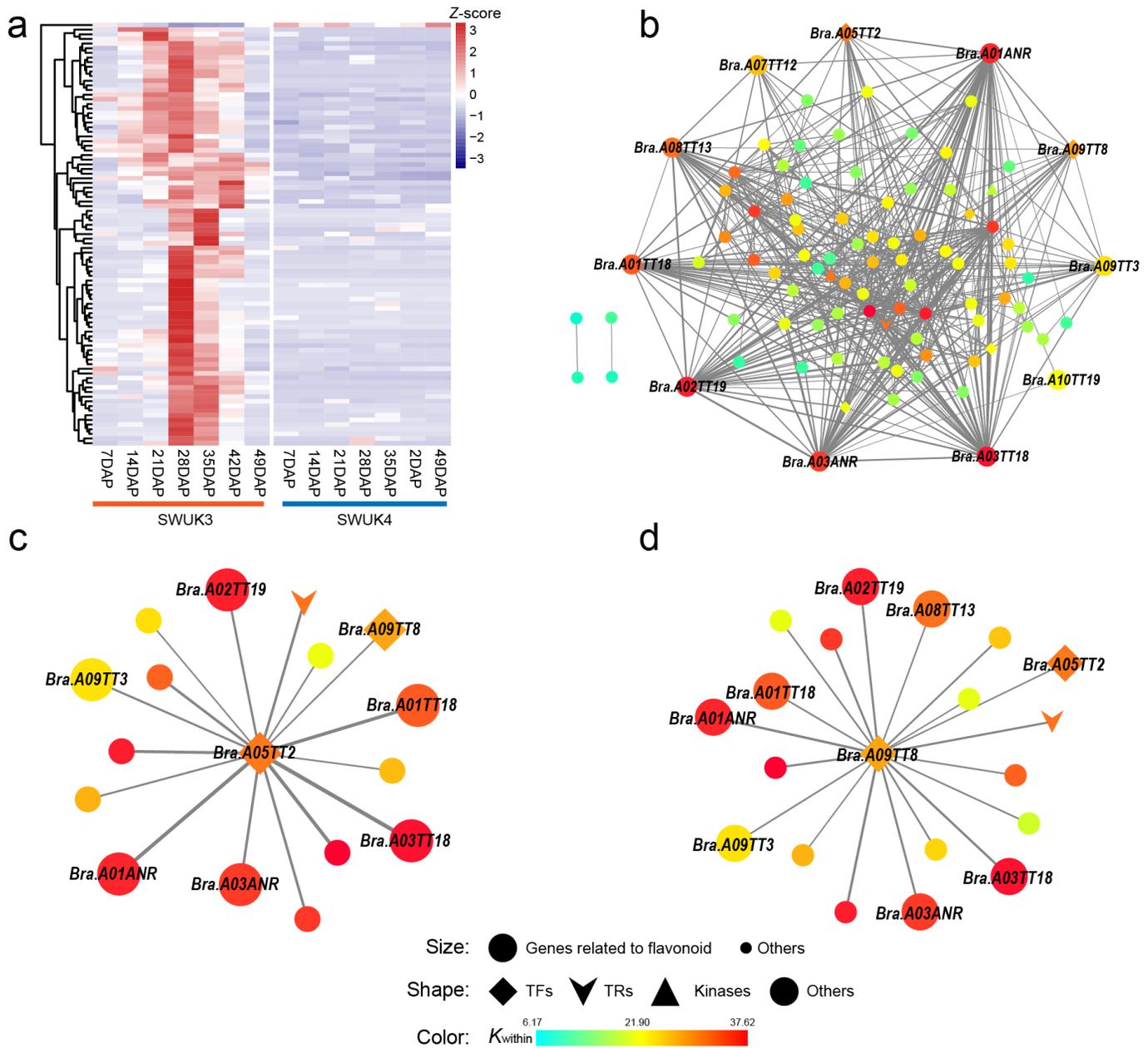
Figure 5

The WGCNA MEsalmon module is negatively related to SC formation. (a) Heatmap of MEsalmon module genes in two B. rapa accessions. Gene expression levels are normalized by Z-score. (b) The MEsalmon co-expression network. (c) Primary co-expression network of Bra.A05TT2. (d) Primary co-expression network of Bra.A09TT8. Square, down arrow, triangle, and disc represent TF, TR, kinase, and other genes, respectively. The edge width represents the weight value between the
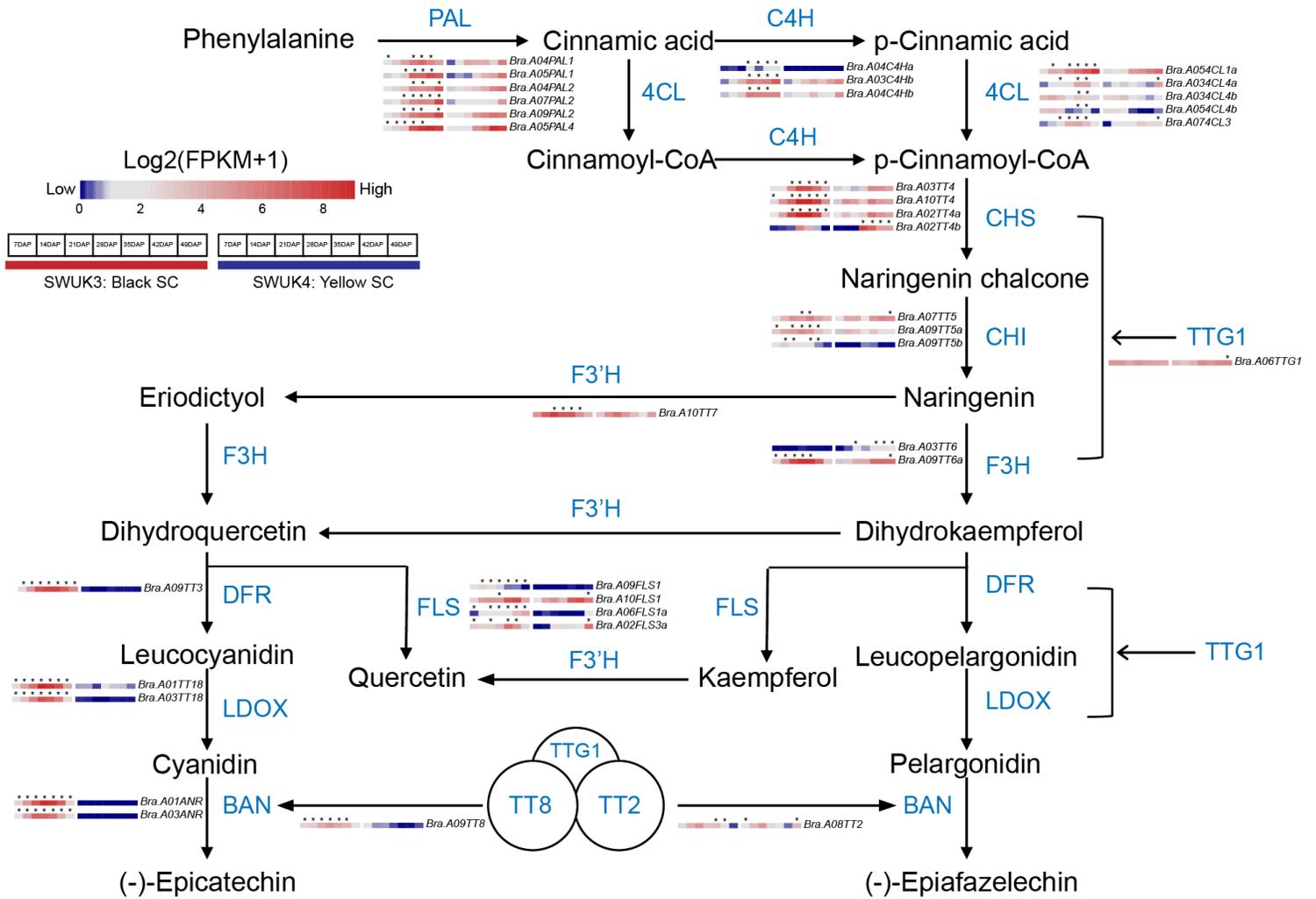
Figure 6

Flavonoid pathways, with enzymes encoded by genes involved in SC formation. Gene expression values in the heatmap were normalized by log2(FPKM+1). All of these genes are listed in Additional file 1: Table S9. Asterisks (*) above the heatmap images indicate that a given gene is significantly more highly expressed in the B. rapa accession marked with an asterisk. Protein abbreviations: PAL, phenylalanineammonialyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate coenzyme A ligase; CHS; chalcone synthase; CHI, chalcone isomerase; F3H, flavonol 3-hydroxylase; F3'H, flavonol 3ᐟ-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase; BAN, BANYULS; TTG1, TRANSPARENT TESTA GLABRA1.
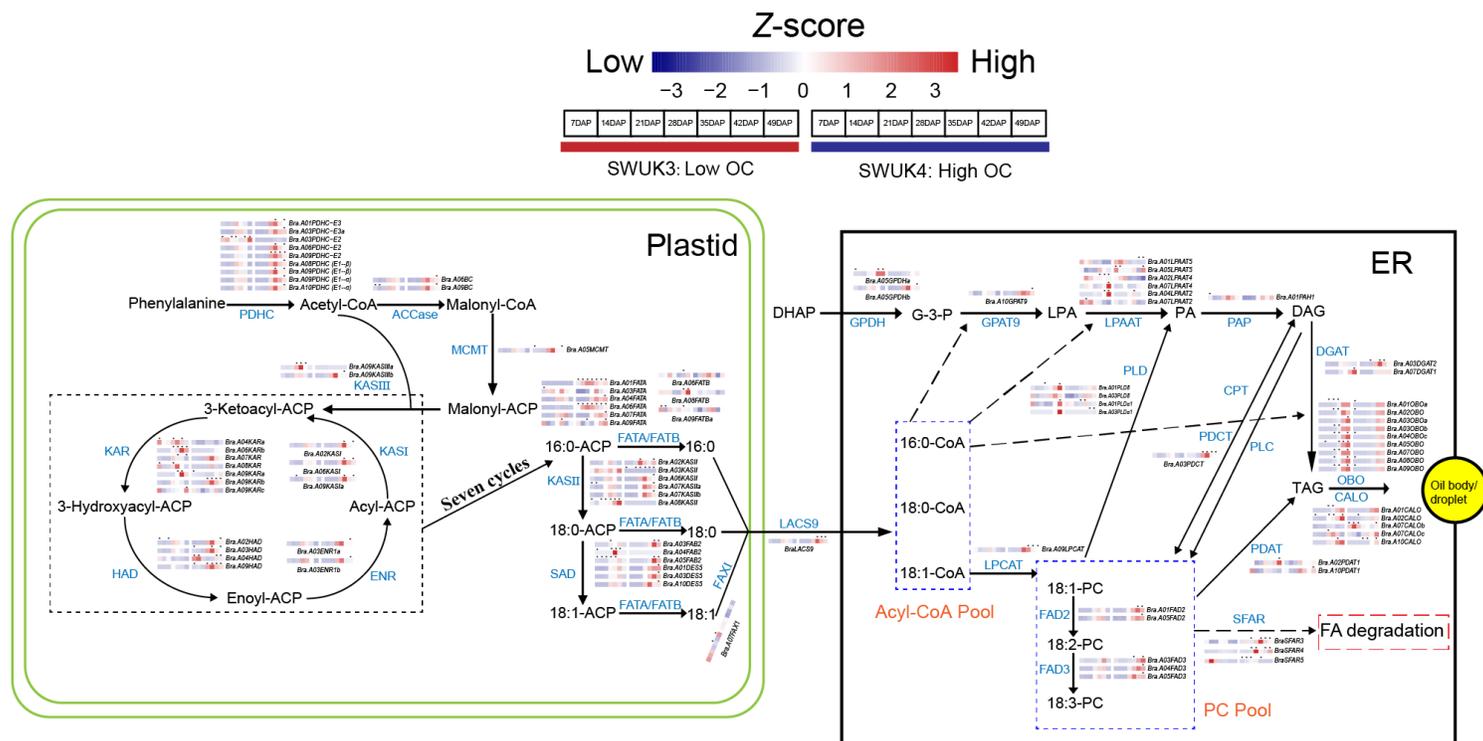
Figure 7

FA metabolism pathway genes regulating the OC. All the genes are listed in Additional file 1: Table S10. Gene expression values were normalized by Z-score. Asterisks (*) above the heatmap indicate that a given gene is expressed at significantly higher levels in the B. rapa accession marked with an asterisk. Protein abbreviations: PDHC, pyruvate dehydrogenase complex; ACCase, acetyl-CoA carboxylase; MCMT, malonyl-CoA:ACP malonyltransferase; ACP, acyl carrier protein; KAS II, 3-ketoacyl-ACP synthase II; KAR, ketoacyl-ACP reductase; HAD, hydroxyacyl-ACP dehydrase; ENR, enoyl-ACP reductase; SAD, stearoyl-acyl carrier protein desaturase; FATA/B, fatty acyl-ACP thioesterase A/B; FAX1, plastid fatty acid export 1; LACS9, long-chain acyl-CoA synthetase 9; GPDH, glycerol-3-phosphate dehydrogenase; GPAT9, glycerol-3-phosphate acyltransferase 9; LPAAT, lysophosphatidic acid acyltransferase; PAP, phosphatidic acid phosphatase, DGAT, diacylglycerol acyltransferase; LPCAT, lysophosphatidylcholine acyltransferase; FAD2/3, fatty acid desaturase 2/3; PLD, phospholipase D; PDAT, phospholipid:diacylglycerol acyltransferase; PDCT, phosphatidylcholine:diacylglycerol cholinephosphotransferase; CPT, CDP-choline: diacylglycerol cholinephosphotransferase, PLC, phospholipase C; OBO, oil body oleosin; CALO, caleosin; SFAR, seed fatty acid reducer.
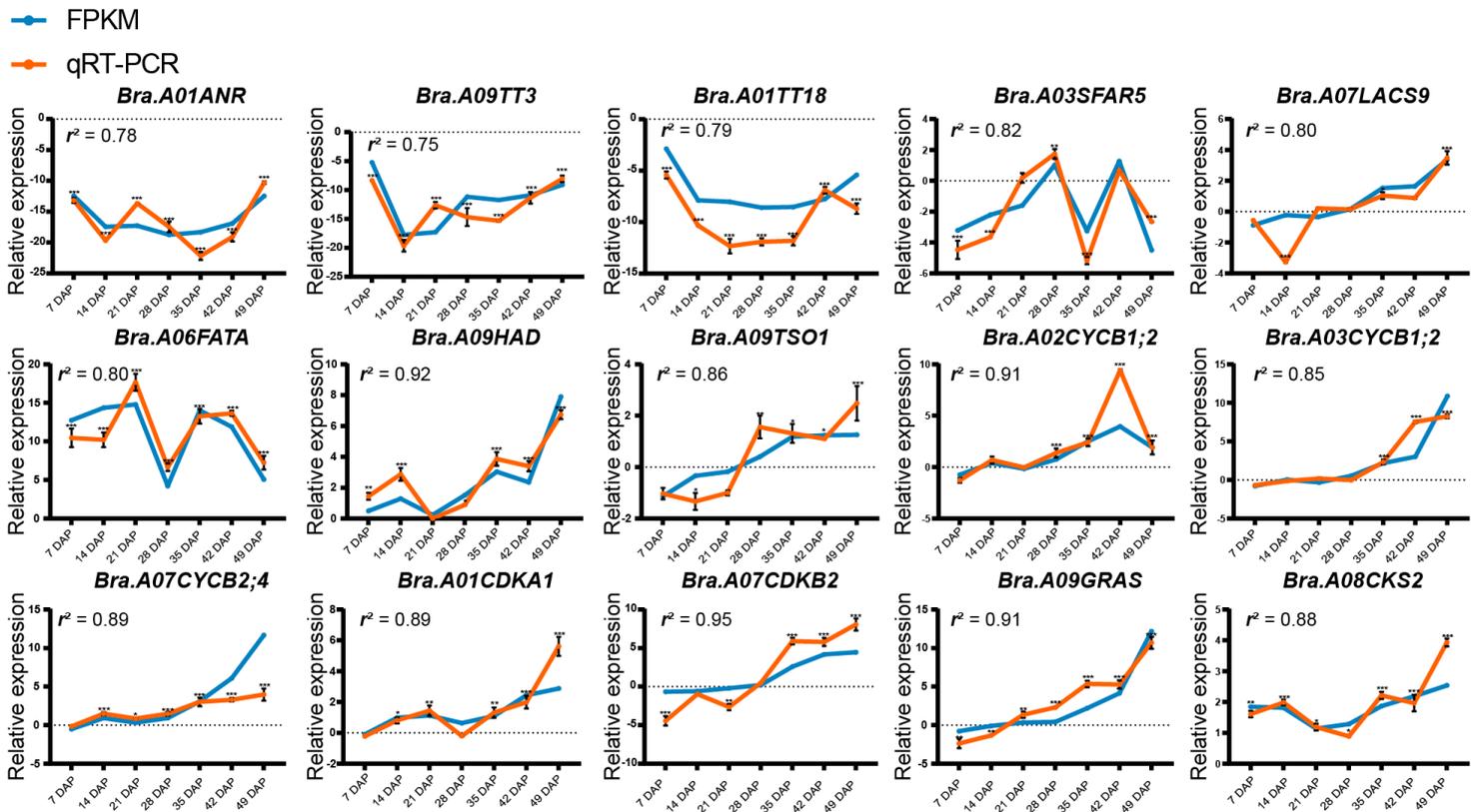
Figure 8

qRT-PCR validation of DEGs. Fifteen genes were selected for validation, including three encoding proteins involved in SC formation (Bra.A01ANR, Bra.A09TT3, and Bra.A01TT18) and four related to lipid biosynthesis and degradation (Bra.A03SFAR5, Bra.A07LACS9, Bra.A06FATA, and Bra.A09HAD); the remaining genes are related to the cell cycle. The blue lines represent RNA-seq results and the orange lines represent qRT-PCR results. Student's t-test was used for statistical analysis of data from the two B. rapa accessions (*, P < 0.05; **, P < 0.01; ***, P < 0.001).