Gene expression and DNA methylation altering lead to the high oil content in wild allotetraploid peanut (*A. monticola*)

Nian Liu
National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University
Manish Pandey
Bei Wu
Li Huang
Huaiyong Luo
Yuning Chen
Xiaojing Zhou
Weigang Chen
Dongxin Huai
Yong Lei
Boshou Liao
Rajeev Varshney
Huifang Jiang (peanutlab@oilcrops.cn)
Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430062,

Article

Keywords: peanut, oil biosynthesis, expression profiling, DNA methylation, functional genomics

Posted Date: January 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1222296/v1

Abstract

The wild allotetraploid peanut *Arachis monticola* contains higher oil content than cultivated allotetraploid *Arachis hypogaea*. To investigate its molecular mechanism controlling oil accumulation, we performed comparative transcriptomics from developing seeds between three *Arachis monticola* and five *Arachis hypogaea* varieties. The analysis not only showed species-specific grouping based on transcriptional profiles but also detected two gene clusters with divergent expression patterns enriched in lipid metabolism. Further, the differential expression gene analysis also indicated expression alteration in wild peanut leading to enhanced activity of oil biogenesis and limiting the rate of lipid degradation. We also constructed a regulatory network of lipid metabolic DEGs with co-expressed transcription factors. In addition, bisulfite sequencing was conducted to characterize the variation of DNA methylation between wild allotetraploid (245, WH 10025) and cultivated allotetraploid (Z16, Zhh 7720) genotypes. Genome-wide DNA methylation was found antagonistically correlated with gene expression during seed development. The results indicated that CG and CHG contexts methylation may negatively regulate specific lipid metabolic genes and transcription factors to subtly affect the difference of oil accumulation. Our work provided the first glimpse on the regulatory mechanism of gene expression altering for oil accumulation in wild peanut and gene resources for future breeding applications.

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is one of important oil crops, which is widely grown in more than 100 countries. The annual production is 48.76 Mt, with 29.60 Mha of global planting area for cultivated peanut (FAOSTAT, 2019, https://www.fao.org/faostat/en/#data/). It was domesticated from a wild relative, *A. monticola*, which harbors high oil content as well as resistance to several biotic stresses (Bertioli et al. 2011; Huang et al. 2012; Moretzsohn et al. 2013; Yin et al. 2020). Since cultivated peanut is one of the major sources for edible oil in the world, enhancing oil content is the second most vital objective after yield in peanut breeding. There are examples in several crops that novel genes from wild counterparts were successfully introgressed into cultivated accessions for crop improvement (Hufford et al. 2012; Sang and Ge 2013; Qi et al. 2014; Tian et al. 2019). The wild counterpart (*A. monticola*) being not having any reproductive barrier with domesticated peanut (*A. hypogaea*) could provide favorable alleles for high-oil accumulation. Developing better understanding on the molecular mechanism and genomic control of high lipid accumulation in wild peanut would facilitate significant increase in oil content in newly developed peanut varieties.

Lipids can be classified into fatty acids, galactolipids, phospholipids, sphingolipid, and acylglycerol (Manan et al. 2017). Tri-acylglycerol (TAG) is the major form of lipid in seed oil which provides calories and essential nutrients to human body. In model plant *Arabidopsis*, significant progress has been made in understanding lipid biosynthesis, transport, and degradation. Many genes encoding enzymes involved in lipid metabolism have been characterized, and several transcription families such as B3, NFY-B, AP2/EREBP, and bZIP, have been reported to regulate these structural genes to control seed oil accumulation (Beisson and F. 2003; Li-Beisson 2013; Manan et al. 2017). Only couple of studies on

expression variation underlying oil accumulation of peanut seed are available that is too only for domesticated peanut, *A. hypogaea* (Wang et al. 2018; Zhang et al. 2021). In case of wild tetraploid, *A. monticola*, a study very recently has been reported on the alteration of genes expression by structural variations affecting pod size (Yin et al. 2020) and not the oil accumulation. Ample literature generated in past provided convincing evidence related to alteration of gene transcription contributing to phenotypic variations between domesticated crops and their wild counterparts (Koenig et al. 2013; Ichihashi et al. 2014; Yoo and Wendel 2014; Lu et al. 2016). Last decade has witnessed monumental progress in terms of genomic resources including reference genomes, gene expression atlas and genotyping assays in peanut to accelerate the genomics and breeding applications in peanut(Pandey et al. 2020). Availability of reference genomes for wild and cultivated tetraploid provides unique opportunity to explore comparative structural and functional genomics to generate more information for further expanding our understanding the genomic and regulatory mechanism of enhanced oil accumulation in *A. monticola*.

In addition to gene transcription, recent studies have also revealed association between difference at DNA methylation level with phenotypic variation in crops (Shen et al. 2018; Xu et al. 2019). In plants, DNA methylation occurs in the three sequence contexts of cytosine, namely CG, CHG, and CHH (H represents A, T or C) (Zhang et al. 2006; Zhang et al. 2018). In heterochromatin or euchromatic chromosome arms, DNA methylation plays a role in the control of transposon silencing and chromosome interaction (Zhang et al. 2006; Cokus et al. 2008). DNA methylation in gene promoters or bodies usually represses transcription, but might increase transcription in some cases (Zhang et al. 2006; Wang et al. 2015; Bewick and Schmitz 2017; Lang et al. 2017; Huang et al. 2019). The DNA methylation has been widely reported to play important roles in vegetable growth, fruit ripening, seed development, and response to biotic and abiotic stress (Zhong et al. 2013; Baubec et al. 2014; Yong-Villalobos et al. 2015; Hewezi et al. 2017; Narsai et al. 2017; Huang et al. 2019; Rajkumar et al. 2020). However, the role of DNA methylation on the regulation of protein-coding genes involved in oil accumulation is largely unknown including in peanut. Genome-wide methylome would expand our view of regulatory mechanism of gene transcription in peanut.

In the present study, we conducted comparative transcriptome profiling of developing seeds from multiple accessions of wild and cultivated peanuts using RNA-seq. Two developmental stages in seed were studied, R5 and R8, representing stages of lipid initial accumulation and lipid rapid increase, respectively. In addition, bisulfite-seq was performed to investigate the variation of DNA methylation between wild accession (245, WH 10025) and cultivated accession (Z16, Zhh 7720). The objective of this study is to reveal a regulatory mechanism of gene expression altering for enhanced oil accumulation in peanut.

Results

Evaluation of RNA-seq data

For analyzing phenotypic variation in seed oil accumulation between wild tetraploid and cultivated peanuts, we measured oil content in three accessions of A. *monticola* and five accessions of *A.*

hypogaea. The oil content of mature seed in three wild accessions, namely 171 (WH 4335), 172 (WH 4334), and 245 (WH 10025), ranged from 58.1% to 59.8%, and that of five cultivated peanuts, namely 003 (Zhh 0225), 145 (Zhh 0888), 492 (Zhh 0003), 502 (Zhh 0602), and Z16 (Zhh 7720), ranged from 48.3% to 52.0% (Figure 1A). Multiple comparison analyses indicated that wild tetraploid species has significantly higher oil content than cultivated peanuts. To explore the reason for oil accumulation difference during peanut domestication, developing seeds at two stages (R5 and R8) were collected to generate RNA-seq datasets for both the species (Figure 1B). Sixteen samples with three biological replications were used to construct RNA-seq libraries followed by generation of 44.1 to 45.4 million clean reads per library (Table S1). The union set of genomes of diploid ancestors *A. ipaensis* and *A. duranensis* (Bertioli et al. 2016) were used as a reference in this study. An average of 83.72% and 83.59% of clean reads were mapped on the reference genome for wild and cultivated peanuts, respectively (Table S1). This similar mapping rate indicated that the reference can be used to quantify the gene expression level in both wild and cultivated species. The number of genes among different libraries ranged from 32148 to 45226 genes with expression of total 56236 genes among 16 samples (Table S1 and S2). Hierarchical clustering analysis was performed to cluster the samples according to developmental stages (Figure 1C). In stage R8 group, the samples could be further divided into wild (BW) or cultivated (BC) subgroups. However, the samples in R5 group could not be clearly distinguished between wild and cultivated subgroups. The first two PCA components (Figure 1D) of transcriptional profiles explained 37.8% (PC1) and 13.2% (PC2) of sample-to-sample variance, respectively. Similar to hierarchical clustering analysis, the PCA analysis also grouped samples according to developmental stages (PC1) prior to species (PC2).

Differentiation in gene expression between wild and cultivated peanuts

To explore divergence of gene expression pattern between two species, transcriptional profiles of 16 samples were used to perform K-mean clustering analysis. The expressed genes (36850) were grouped into 9 clusters designated as C1-C9 (Figure 2A). Five clusters (C1-C5) showed a decreasing tendency from R5 to R8 stage in both peanut species. Conversely, C6 and C8 clusters showed an increasing trend across accessions from R5 to R8 stage. For C7 cluster, wild accessions (171, 172, and 245) showed a decreasing trend, while cultivated accessions showed an increasing pattern or no obvious change during seed developmental stages. In C2 cluster, gene expression level was higher in wild accessions at R5 stage. In C8 cluster, gene expression level was in general higher in cultivated accessions at both stages. GO category analysis was performed to identify over-represented GO terms of the 9 clusters (Figure 2B and Table S3). A total of 30, 14, and 22 GO terms were enriched in biological process, cellular component, and molecular function, respectively. Each cluster had differently enriched GO terms, and the number of over-represented GO terms ranged from 3 (C9) to 35 (C2). The GO term of lipid metabolism process was found over-represented in clusters C1, C2, and C7 (P-value < 0.05). Interestingly, gene expression level was different between wild and cultivated peanuts in C2 and C7 clusters, indicating that transcriptional profile of lipid metabolism may be divergent between the two species.

To further dissect the expression difference of lipid metabolism, a comparative transcriptomic analysis was performed to identify differentially expressed genes (DEGs) between wild and cultivated peanuts. There were 5647 and 3184 DEGs at R5 (SW vs SC) and R8 (BW vs BC) stages, respectively (Figure S1). A total of 1578 DEGs were detected at both R5 and R8 stages. The number of up-regulated and down-regulated DEGs was overall equivalent at both R5 and R8 stages (Figure S1). Heatmaps of DEGs involved in fatty acid synthesis, fatty acid elongation, TAG synthesis, phospholipid synthesis, sphingolipid synthesis, galactolipid and sulfolipid synthesis, wax synthesis, and β-oxidation were profiled to represent a transcriptional change in lipid metabolism (Figure 3). According to expression pattern of DEGs involved in lipid metabolism, samples in both heatmaps (R5 and R8 stages) could be divided into two groups (wild and cultivated). At R5 stage, most lipid metabolism-related pathways, such as fatty acid synthesis, fatty acid elongation, TAG synthesis, phospholipid synthesis, sphingolipid synthesis, galactolipid, and sulfolipid synthesis were dramatically up-regulated in wild species (Figure 3A). Most down-regulated DEGs at R5 stage were mainly distributed on β-oxidation and wax synthesis pathways. Compared with R5 stage, the portion of up-regulated DEGs in wild species was generally lower at R8 stage (Figure 3B). However, DEGs in fatty acid synthesis, TAG synthesis, phospholipid synthesis, sphingolipid synthesis were mainly up-regulated in wild peanut accessions at R8 stage. Conversely, all the DEGs in β-oxidation were down-regulated at R8 stage. There were 16 lipid metabolic DEGs simultaneously identified at both R5 and R8 stages belonging to fatty acid synthesis (5), fatty acid elongation (3), TAG synthesis (2), sphingolipid synthesis (4), and β-oxidation (3) pathways.

Construction of co-expression network

To further explore genes with high connectivity to lipid metabolic-related DEGs, weighted gene co-expression network analysis (Langfelder and Horvath 2008) was performed to construct a topological overlap matrix of expression similarity between genes. Genes co-expressed with lipid metabolic-related DEGs (expression similarity ≥ 0.1) were selected to perform GO enrichment analysis. They were enriched in 14 GO terms for molecular function category, including transcription regulator activity, DNA-binding transcription factor activity, and DNA binding (Figure 4A). It was suggested that transcription factors (TFs) may play a role in the co-expression network. Many TFs were differentially expressed between wild cultivated peanuts (Figure S3). In specific TF families, such as AP2, B3, and bZIP, the family members were predominately upregulated in wild cultivated peanuts at both R5 and R8 stages. 391 transcription factors were belonging to 47 families that were co-expressed with DEGs involved in lipid synthesis pathways, such as fatty acid synthesis, fatty acid elongation, TAG synthesis, Phospholipid synthesis, Sphingolipid synthesis, Galactolipid and Sulfolipid synthesis (Figure 4B). According to interaction number, the top10 notes in the co-expressed transcription factors (Table S6) included AP2 (*Araip.E0UEG*), B3 (*Aradu.07I6M*), B3 (*Araip.S9XVH*), bZIP (*Aradu.898PR*), bZIP (*Araip.0GM4I*), bZIP (*Araip.R3LNH*), C2H2 (*Araip.X9IXZ*), G2-like (*Araip.8YC75*), MYB-related (*Araip.L6TM2*), NF-YC (*Aradu.3GN04*), and Trihelix (*Aradu.RW5KN*). Expression ratio (wild/cultivated) of the transcription factors ranged from 1.4 to 2.5 at R5 stage with mean value of 1.8 while the expression ratio ranged from 1.1 to 2.4 at R8 stage with mean

value of 1.4 (Figure 4C). In general, the abundance of the co-expressed transcription factors was higher in wild peanuts than cultivated peanuts, especially at R5 stage.

Influence of DNA methylation on gene expression

The available literature in multiple crops proved that DNA methylation could help to regulate gene expression (Huang et al., 2019; Rajkumar et al., 2020; Wang et al., 2015; Wang et al., 2016; Xing et al., 2015). For revealing the relationship between DNA methylation and gene expression in peanut, we performed bisulfite sequencing of the genomic DNA isolated from R5 and R8 stages of seeds in wild peanut (245) and cultivated peanut (Z16). Four samples (S245, SZ16, B245, BZ16) representing seeds at R5 (S) and R8 (B) stages for 245 and Z16, were sequenced with three biological replicates. About 480 M clean reads were generated for each sample (Table S7) and approximately 76% of the clean reads were uniquely mapped to the reference genomes covering >87% of the genomic cytosine positions. Each methylome had >16-fold average depth per strand. Methyltcytosines were identified in CG, CHG, and CHH contexts across samples. Compared with average methylation of CHH context (11.2-22.3%), the level was much higher in CG (81.7%-86.0%) and CHG (73.2%-78.2%) contexts (Figure S3A). The fraction of methylcytosine (mC) was 20.1%-25.2% in CG, 25.6-31.8% in CHG, and 43.0-54.4% in CHH. Overall methylation levels in four samples were similar (29.3-34.7%) (Figure 5A and Figure S3A). The distribution of mC showed much lower methylation in the terminal chromosomes in contrast to much higher gene expression in the terminal chromosomes (Figure 5A). Genome-wide correlation coefficient between overall DNA methylation and gene expression in the four samples was ~-0.56 ($p < 2.2$ e-16), indicating significantly antagonistical correlation (Figure 5B).

To further characterize epigenetic variations between 245 (wild) and Z16 (cultivated), differentially methylation regions of CG, CHG, and CHH contexts were detected at R5 and R8 stages (Table S8). Chromosome-wide view of DMR distribution indicated CG-DMR being most likely to be enriched in gene-rich regions (Figure 6A and 6B). Approximately 40% of CG-DMRs were located in the genic region (gene body+2k upstream), the portions were deduced to ~15% and ~10% for CHG-DMRs and CHH-DMRs, respectively (Figure 6C). Conjoint analysis of the expression profile and DMRs was indicated that CG-DMRs and CHG-DMRs were significantly enriched in the genic region (gene body+2k-upstream) of DEGs between 245 and Z16 (Figure 6D). It is hinted that CG- and CHG-DMRs, not CHH-DMR, may play a role in the change of gene expression. Among DEG-overlapped CG-DMRs, the portion of hyper-DMRs in S245-upregulated DEGs was 50.3% and 51.9% at R5 and R8 stages, respectively, and that in S245-downregulated DEGs could increase to 70.5% and 71.1% at R5 and R8 stages, respectively (Figure 6E). Similarly, the portion of hyper-DMRs was much higher in S245-downregulated DEGs (83.7% and 82.4% at R5 and R8, respectively) than in S245-upregulated DEGs (42.7% and 12.5% at R5 and R8, respectively) for DEG-overlapped CHG-DMRs. The results suggested that dynamic change of methylated CG (mCG) and methylated CHG (mCHG) on a genic scale correlated negatively with the difference in gene abundance between 245 and Z16.

Role of DNA methylation in divergence of lipid metabolism

The CG-DMRs between 245 and Z16 were overlapped with 8639 and 2147 DEGs at R5 and R8 stages, respectively (Figure 6F). For CHG-DMRs, there were 4024 and 1357 CHG-DMR associated DEGs at R5 and R8 stages, respectively (Figure 6G). In general, more than 50% and 25% of DEGs between 245 and Z16 may be associated with CG-DMRs and CHG-DMRs, respectively. GO enrichment analysis showed that CG-DMR associated DEGs were enriched in carbohydrate metabolic process and response to endogenous stimulus at both stages (Figure 6H). The CHG-DMR associated DEGs were enriched in response to chemical and response to endogenous stimulus at both stages (Figure 6I). Interestingly, lipid metabolism was one of enriched terms at R8 stage for CG-DMR associated DEGs.

There were 41 and 19 lipid metabolic DEGs that showed a negative correlation with DMRs in CG and CHG context between 245 and Z16 at R5 stage, respectively (Figure 7A). Most genes involved in fatty acid synthesis, fatty acid elongation, phospholipid synthesis, sphingolipid synthesis, galactolipid and sulfolipid synthesis, and TAG synthesis pathways, exhibited hypomethylation with higher expression in 245 at R5 stage. In addition, 33 and 20 TFs known to involve in regulation of lipid metabolism, were found to be negatively correlated with CG- and CHG- DMRs. Among the CG- or CHG- DMR associated TFs, 15 members had been identified to co-express with lipid metabolic DEGs (Figure 7A and Table S6). Most TFs involved in regulation of lipid metabolism, especially the co-expressed TFs, showed hypomethylation with higher abundance in 245 at R5 stages. At R8 stage, 6 and 6 lipid metabolic DEGs were found to be negatively correlated with CG- and CHG- DMRs between 245 and Z16 at R8 stage, respectively (Figure 7B). Genes involved in β-Oxidation showed hypermethylation with lower expression in 245 at R8 stage. Reversely, genes involved in fatty acid elongation, phospholipid synthesis, and TAG synthesis pathways showed hypermethylation with lower abundance in 245. Meanwhile, 17 TFs known to regulate lipid metabolism, exhibited a negative correlation with DMR in CG or CHG context at R8 stage. Among them, three members were identified to co-express with lipid metabolic DEGs (Figure 7B and Table S6). Two DEGs showed hypermethylation with lower abundance, while another one showed hypomethylation with higher expression in 245 at R8 stage. Two examples exhibited a relationship between DMRs and lipid metabolic DEGs. The one gene (*Araip.H6S1B*) encoding acyl transporter protein was down-regulated in 245 at both stages. Hyper-DMRs were observed at 5’ UTR of the gene (Figure 7B). Another one (*Aradu.IPH1G*) encoding B3 transcript factor, was up-regulated in 245 at R5. Hypo-DMRs were located at upstream of the gene (Figure 7C).

Discussion

Cultivated peanut (*A. hypogaea*) was domesticated from the wild tetraploid *A. monticola*, originated because of hybridization between two diploid *A. duranensis* and *A. ipaensis* (Bertioli et al. 2011; Moretzsohn et al. 2013; Yin et al. 2020). Compared to cultivated species, wild peanut possess high genetic diversity as well as several superior features including high oil content in seed (Huang et al.

2012). Understanding lipid accumulation in *A. monticola* at the level of regulation of gene expression would contribute to high-oil improvement in peanut breeding. In this study, we used the RNA-seq approach to analyze transcriptome divergence between wild and cultivated peanut at two seed developing stages (R5 and R8), and investigated the molecular mechanism underlying difference in seed oil content. More than 70% of the reference genes (56236) were identified and profiled in at least one sample. According to the result of PCA for expression pattern, developmental stage prior to species was the first principal component to distinguish the samples. However, expression profiles at R8 stage could be obviously divided into two species in the hierarchical clustering diagram, indicating that gene expression profiles in developing seed (R8 stage) have been altered during peanut domestication. Available literature suggest that gene expression divergence is essential to drive phenotypic variation during domestication (Lin et al. 2012; Koenig et al. 2013; Lu et al. 2016). Interestingly, we detected two divergent gene clusters (C2 and C7) in peanuts enriched in lipid metabolism (Figure 2). Our results suggest that there may be the existence of divergence at gene expression level between wild and cultivated peanut contributing to the difference of lipid accumulation.

The lipid metabolism-related DEGs were presumed to play key roles in oil accumulation. Based on the expression pattern of lipid metabolic DEGs, samples could be clearly clustered into two groups (wild and cultivated) (Figure 3). Most DEGs in lipid biosynthesis pathways were upregulated in wild species, especially at lipid initial accumulation stage (R5) (Figure 3). Meanwhile, *β*-oxidation related DEGs, which are involved in degradation of lipid, were downregulated in wild peanut from lipid initiation to lipid rapid increase stages (R5 and R8). The DEG analysis revealed a robust activity of oil biosynthesis with a constraint of lipid degradation in developing seed of wild peanuts. It is worth noting that 16 lipid metabolic DEGs between wild and cultivated peanut were repeatedly identified at both R5 and R8 stages. These consistently divergent genes during seed development may contribute to a part of the difference in oil accumulation between wild and cultivated species. For example, *Aradu.CP1HR* encoded a biotin carboxyl carrier protein, which is a subunit of acetyl-CoA carboxylase (ACC). ACC catalyzes malonyl-CoA and bicarbonate to yield malonyl-CoA, which is the first committed step in fatty acid synthesis(Li-Beisson 2013). The expression level of *Aradu.CP1HR* was upregulated in wild species. In contrast, the expression abundance of another gene (*Araip.H6S1B*) was lower in wild species. It was an acyl transporter protein involved in the import of *β*-oxidation substrate (fatty acids) to peroxisome to break down fatty acids(Li-Beisson 2013). In addition, many transcription factors (TFs) known in the regulatory circuitry during seed development (Agarwal et al. 2011; Li and Li 2016), were identified to co-express with lipid metabolic DEGs (Figure 4). The TF families, such as AP2, B3, b-ZIP, and C3H, whose members showed different abundance between wild and cultivated peanuts (Figure 4 and S2), have been well known in the regulation of oil production (Pouvreau et al. 2011; Song et al. 2013; Kim et al. 2015; Manan et al. 2017; Lu et al. 2021). Since a set of genes with a similar expression pattern was likely to take part in the same biological process, we deduced that coexpressed network of TFs and lipid metabolic DEGs may construct transcriptional modules affecting differential oil accumulation between wild and cultivated peanuts. Altogether, TFs may coordinate expression abundance of lipid metabolic genes to promote oil biosynthesis in wild species. It might explain higher oil content in seed of wild peanuts.

DNA methylation is a conserved epigenetic marker that regulate gene expression. There are examples of natural epialleles in several crops showing varied DNA methylation affecting multiple biological processes (Manning et al. 2006; Zhang et al. 2015; Song et al. 2017). Here, we sought to investigate the role of DNA methylation on genes involved in lipid accumulation. The genome-wide DNA methylome and trancriptome for developing seeds of 245 (wild species) and Z16 (cultivated species) were conjointly profiled (Figure 5). A significantly negative correlation on chromosome-scale between DNA methylation and gene expression was observed, indicating that DNA methylation generally inhibits gene transcription in developing peant seeds (R5 and R8). Meanwhile, differentially methylated regions between 245 and Z16 were displayed on 20 reference chromosomes (Figure 6A and 6B), providing a first glimpse of the epigenetic changes in seed development during peanut domestication. DMRs in CG and CHG, but not CHH contexts, were positively correlated with gene density on chromosomes and tended to enrich in genic regions (gene body+2k-upstream) of DEGs (Figure 6D). Previously studies have shown that DNA methylation occurring in the promoter or within the transcribed gene body would regulate gene transcription (Zhang et al. 2006; Wang et al. 2015; Bewick and Schmitz 2017; Lang et al. 2017; Huang et al. 2019). Therefore, the DMRs, especially in CG and CHG contexts, would play a role in the differential expression of genes in the present study. There were 50% and 25% of DEGs between 245 and Z16 were associated with CG-DMRs and CHG-DMRs, respectively. GO terms of CG- and CHG-DMRs associated DEGs were enriched in several biological processes, including lipid metabolism. There was an obvious trend that most differentially expressed enzymes and co-expressed TFs involved in lipid production were hypomethylated in wild peanut (245) at R5 stage. In contrast, DEGs involved in β-Oxidation were found hypermethylated in 245 at R8 stage. Cyl transporter protein (*Araip.H6S1B*) and B3 transcription factor (*Aradu.IPH1G*) were examples to demonstrate that the influence of DNA methylation on expression of lipid-related genes varied between wild (245) and cultivated peanut (Z16). Altogether, the methylome and transcriptome data depicted a possible regulatory network in which DNA methylation and TFs co-regulate the expression of lipid metabolism genes in peanut seed (Figure 8). In wild peanut, comprehensive alteration of genes transcription by TFs and DNA methylation would promote oil production and constrain oil degradation simultaneously, which finally contribute to higher oil accumulation.

In summary, transcriptomic and methylomic comparisons revealed gene expression and DNA methylation variations in seed development between wild and cultivated peanuts. In wild peanut seed, transcription factors and DNA methylation may coordinately regulate specific lipid metabolic genes to active oil biosynthesis and simultaneously constrain lipid degradation. Thus, gene expression change would contribute to increasing oil content in wild peanut. Our study reveals a regulation mechanism of oil accumulation in seed and provides gene resources for oil improvement in cultivated peanut.

Methods

Plant material

Three accessions, 171 (WH 4335), 172 (WH 4334), and 245 (WH 10025), were collected to represent wild peanuts (*A. monticola*). Five accessions, 003 (Zhh 0225), 145 (Zhh 0888), 492 (Zhh 0003), 502 (Zhh

0602), and Z16 (Zhh 7720), were used to represent cultivated peanuts (*A. hypogaea*). 003, 145, 492, 502 were belonging to four mainly agronomic types (*vulgaris*, *hypogaea*, *fastigiata*, and *hirsuta*) in cultivated peanut, respectively. Z16 is a modern cultivar with mixed parentage. Wild and cultivated peanuts were planted in the experimental nursery of OCRI-CAAS, Wuhan. Nursery management followed standard agricultural practices. Three biological replications were grown for each accession. Developing seeds from eight accessions were collected at previously characterized stages (Pattee et al. 1974): lipid initial accumulation stage (R5) and lipid rapid increase stage (R8). The collected samples were immediately frozen in liquid nitrogen and stored at -70 ℃ for RNA and DNA isolation. The oil content (%) of mature seeds from each accession was measured as described previously (Liu et al. 2020).

Processing of RNA-seq Data

Total RNA was extracted using an RNAprep pure plant kit (Tiangen, Being, China) according to the manufacturer's instructions. Forty-eight libraries from the developing seeds of wild and cultivated peanuts were constructed and sequenced using a Hiseq XTen platform in Beijing Genomics Institute (BGI, https://www.genomics.cn/). After obtaining raw data, we used software SOAPnuke to perform quality filtering and read trimming (https://github.com/BGI-flexlab/SOAPnuke). The clean RNA-seq reads were mapped to the reference genome using HISAT2 software (Gao et al. 2016). The reference genome consisted of two diploid ancestors' genomes of *A. ipaensis* (V14167) and *A. duranensis* (V14167) (Bertioli et al. 2016). The RSEM package was used to calculate and normalize gene expression level as fragments per kilobase of transcript per million mapped reads (FPKM) (Koenig et al. 2013). Differential gene expression (fold change ≥ 2 and adjusted *P*-value ≤0 .001) was identified using DEGseq package (Lin et al. 2012).

Gene clustering, functional annotation, and Weighted gene co-expression network analysis

K-means clustering was used to visualize genes exhibiting a similar expression pattern, and it was performed on normalized FPKM values using MeV software (Agarwal et al. 2011). The distance metric for K-means clustering was set as Euclidean distance. Annotation analyses were performed by BLASTing public protein databases, including Nr (http://www.ncbi.nlm.nih.gov), GO (http://www.geneontology.org), and KEGG (http://www.genome.jp/kegg). AgriGO and TBtools were used to identify and show over-represented Gene Ontology (GO) terms (Zhang et al. 2015; Lang et al. 2017). The R package (WGCNA) was used to build weighted gene coexpression networks (Langfelder and Horvath 2008). Network construction was performed using the block wise Modules function with default parameters. The topological overlap matrix (TOM) was calculated to measure the strength of a coexpression relationship, i.e. connectivity between any two genes with respect to all other genes in the network.

Whole-genome bisulfite sequencing

Genome DNA from each sample was fragmented to a mean size of 250 bp through Bioruptor (Diagenode, Belgium). Adapters were ligated to the fragment DNA and treated with sodium bisulfite using EZDNA Methylation-Goldkit (ZYMO, USA). Sequencing was performed using Illumina HiSeq platform. Three biological replications were sequenced for each stage of developing seed in both wild and cultivated peanuts.

Read alignment and methylcytosine identification

The raw reads from each library were processed to remove low-quality reads, adaptor sequences, and contamination. The clean reads were aligned to the reference genome (Bertioli et al. 2016), and only the reads mapped at unique position were retained. The binomial test was performed for each cytosine base to identified true methylcytosine. Cytosine sites with *p*-value < 0.0001 were defined as methylated cytosine sites. Methylation level at each methylcytosine site was determined by percentage of reads giving methylation call to all the reads aligned at the same site.

Identification of differentially methylated regions (DMRs)

Putative DMRs were identified using windows that contained at least 5 CG (CHG or CHH) sites with different methylation level (cut-off value > 0.1 between two compared samples) and *Fisher*-test *p*-value <= 0.05. In addition, two nearby DMRs would be considered interdependent and joined into one continuous DMR if the genomic region from the start of an upstream DMR to the end of a downstream DMR also had differences methylation levels (cut-off value > 0.1 between two compared samples) with a *p*-value <= 0.05. Otherwise, the two DMRs were viewed as independent. When the genic region (2kb upstream or body) was overlapped with DMR, the gene was defined as a DMR-associated gene. The fold enrichment of DEGs in DMR-associated genes was calculated using the formula, and *P*-value significance was generated using the hypergeometric test.

Data availability

RNA sequence data (accession no. PRJNA781013) and Bisulfite sequence data (accession no. PRJNA782686) have been deposited in Sequence Read Archive at the National Center for Biotechnology Information.

References

1. Agarwal P, Kapoor S, Tyagi AK (2011) Transcription factors regulating the progression of monocot and dicot seed development. BioEssays 33: 189–202
2. Baubec T, Finke A, Mittelsten Scheid O, Pecinka A (2014) Meristem-specific expression of epigenetic regulators safeguards transposon silencing in *Arabidopsis*. EMBO Rep 15: 446–452
3. Beisson, F. (2003) *Arabidopsis* genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. **Plant Physiol** 132: 681-697
4. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araújo ACG, Kozik A, Do Kim K, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimarães PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SCM, Xun X, Jackson SA, Michelmore R, Ozias-Akins P (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. **Nat Genet** 48: 438-446
5. Bertioli DJ, Seijo G, Freitas FO, Valls JFM, Leal-Bertioli SCM, Moretzsohn MC (2011) An overview of peanut and its wild relatives. Plant Genetic Resources 9: 134–149
6. Bewick AJ, Schmitz RJ (2017) Gene body DNA methylation in plants. Curr Opin Plant Biol 36: 103–110
7. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature 452: 215–219
8. Gao X, Chen J, Dai X, Zhang D, Zhao Y (2016) An effective strategy for reliably isolating heritable and Cas9-free *Arabidopsis* mutants generated by CRISPR/Cas9-mediated genome editing. Plant Physiol 171: 1794
9. Hewezi T, Lane T, Piya S, Rambani A, Rice JH, Staton M (2017) Cyst nematode parasitism induces dynamic changes in the root epigenome. **Plant Physiol** 174: 405-420
10. Huang H, Liu RE, Niu QF, Tang K, Zhang B, Zhang H, Chen KS, Zhu JK, Lang ZB (2019) Global increase in DNA methylation during orange fruit development and ripening. Proc Natl Acad Sci USA 116: 1430–1436
11. Huang L, Jiang HF, Ren XP, Chen YN, Xiao YJ, Zhao XY, Tang M, Huang JQ, Upadhyaya HD, Liao BS (2012) Abundant microsatellite diversity and oil content in wild *Arachis* species. **PLoS ONE** 7:
12. Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. **Nat Genet** 44: 808-811
13. Ichihashi Y, Aguilarmartínez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, Peng J, Maloof JN, Sinha NR (2014) Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. Proc Natl Acad Sci USA 111: 2616–2621

14. Kim HU, Lee KR, Jung SJ, Shin HA, Go YS, Suh MC, Kim JB (2015) Senescence-inducible *LEC2* enhances triacylglycerol accumulation in leaves without negatively affecting plant growth. Plant Biotechnol J 13: 1346–1359
15. Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeples M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR, Maloof JN (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci USA 110: E2655-2662
16. Lang Z, Wang Y, Tang K, Tang D, Datsenka T, Cheng J, Zhang Y, Handa AK, Zhu JK (2017) Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. Proc Natl Acad Sci USA 114: E4511-E4519
17. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9
18. Li-Beisson Y (2013) Acyl-lipid metabolism. **Arabidopsis Book** 11
19. Li N, Li Y (2016) Signaling pathways of seed size control in plants. Curr Opin Plant Biol 33: 23–32
20. Lin ZW, Li XR, Shannon LM, Yeh CT, Wang ML, Bai GH, Peng Z, Li JR, Trick HN, Clemente TE, Doebley J, Schnable PS, Tuinstra MR, Tesso TT, White F, Yu JM (2012) Parallel domestication of the Shattering1 genes in cereals. Nat Genet 44: 720–724
21. Liu N, Guo J, Zhou X, Wu B, Huang L, Luo H, Chen Y, Chen W, Lei Y, Huang Y, Liao B, Jiang H (2020) High-resolution mapping of a major and consensus quantitative trait locus for oil content to a ~ 0.8-Mb region on chromosome A08 in peanut (*Arachis hypogaea* L.). Theor Appl Genet 133: 37–49
22. Lu L, Wei W, Li QT, Bian XH, Lu X, Hu Y, Cheng T, Wang ZY, Jin M, Tao JJ, Yin CC, He SJ, Man WQ, Li W, Lai YC, Zhang WK, Chen SY, Zhang JS (2021) A transcriptional regulatory module controls lipid accumulation in soybean. **New Phytol**
23. Lu X, Li QT, Xiong Q, Li W, Bi YD, Lai YC, Liu XL, Man WQ, Zhang WK, Ma B, Chen SY, Zhang JS (2016) The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. Plant J 86: 530–544
24. Manan S, Chen B, She G, Wan X, Zhao J (2017) Transport and transcriptional regulation of oil production in plants. Crit Rev Biotechnol 37: 641–655
25. Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet 38: 948–952
26. Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SC, Valls JF, Bertioli DJ (2013) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111: 113–126
27. Narsai R, Secco D, Schultz MD, Ecker JR, Lister R, Whelan J (2017) Dynamic and rapid changes in the transcriptome and epigenome during germination and in developing rice (*Oryza sativa*) coleoptiles under anoxia and re-oxygenation. Plant J 89: 805–824

28. Pandey MK, Pandey AK, Kumar R, Nwosu CV, Guo B, Wright GC, Bhat RS, Chen X, Bera SK, Yuan M, Jiang H, Faye I, Radhakrishnan T, Wang X, Liang X, Liao B, Zhang X, Varshney RK, Zhuang W (2020) Translational genomics for achieving higher genetic gains in groundnut. Theor Appl Genet 133: 1679–1702
29. Pattee HE, Johns EB, Singleton JA, Sanders TH (1974) Composition changes of peanut fruit parts during maturation. Peanut Sci 1: 57–62
30. Pouvreau B, Baud S, Vernoud V, Morin V, Py C, Gendrot G, Pichon JP, Rouster J, Paul W, Rogowsky PM (2011) Duplicate maize wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis. **Plant Physiol** 156: 674-686
31. Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen Yim A, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang TH, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu X, Lam HM (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. **Nat Commun** 5: 4340
32. Rajkumar MS, Gupta K, Khemka NK, Garg R, Jain M (2020) DNA methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. Communications Biology 3: 340
33. Sang T, Ge S (2013) Understanding rice domestication and implications for cultivar improvement. Curr Opin Plant Biol 16: 139–146
34. Shen Y, Zhang J, Liu Y, Liu S, Liu Z, Duan Z, Wang Z, Zhu B, Guo YL, Tian Z (2018) DNA methylation footprints during soybean domestication and improvement. Genome Biol 19: 128
35. Song QX, Li QT, Liu YF, Zhang FX, Ma B, Zhang WK, Man WQ, Du WG, Wang GD, Chen SY, Zhang JS (2013) Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants. J Exp Bot 64: 4329–4341
36. Song Q, Zhang T, Stelly DM, Chen ZJ (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol 18: 99
37. Tian J, Wang C, Xia J, Wu L, Xu G, Wu W, Li D, Qin W, Han X, Chen Q, Jin W, Tian F (2019) Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. Science 365: 658–664
38. Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I (2015) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. Proc Natl Acad Sci USA 112: 13729–13734
39. Wang X, Xu P, Yin L, Ren Y, Li S, Shi Y, Alcock TD, Xiong Q, Qian W, Chi X, Pandey MK, Varshney RK, Yuan M (2018) Genomic and transcriptomic analysis identified gene clusters and candidate genes for oil content in peanut (*Arachis hypogaea* L.). Plant Mol Biol Report 36(3): 518–529
40. Xu J, Chen G, Hermanson PJ, Xu Q, Sun C, Chen W, Kan Q, Li M, Crisp PA, Yan J, Li L, Springer NM, Li Q (2019) Population-level analysis reveals the widespread occurrence and phenotypic consequence of DNA methylation variation not tagged by genetic variation in maize. **Genome Biol** 20: 243.

41. Yin D, Ji C, Song Q, Zhang W, Zhang X, Zhao K, Chen CY, Wang C, He G, Liang Z, Ma X, Li Z, Tang Y, Wang Y, Li K, Ning L, Zhang H, Li X, Yu H, Lei Y, Wang M, Ma L, Zheng H, Zhang Y, Zhang J, Hu W, Chen ZJ (2020) Comparison of arachis monticola with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of Peanut. Adv Sci 7: 1901672
42. Yong-Villalobos L, Gonzalez-Morales SI, Wrobel K, Gutierrez-Alanis D, Cervantes-Perez SA, Hayano-Kanashiro C, Oropeza-Aburto A, Cruz-Ramirez A, Martinez O, Herrera-Estrella L (2015) Methylome analysis reveals an important role for epigenetic changes in the regulation of the *Arabidopsis* response to phosphate starvation. Proc Natl Acad Sci USA 112: E7293-E7302
43. Yoo MJ, Wendel JF (2014) Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. PLoS Genet 10: e1004073
44. Zhang H, Wang ML, Dang P, Jiang T, Zhao SZ, Lamb M, Chen C (2021) Identification of potential QTLs and genes associated with seed composition traits in peanut (*Arachis hypogaea* L.) using GWAS and RNA-Seq analysis. **Gene** 769:
45. Zhang HM, Lang ZB, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19: 489–506
46. Zhang X, Sun J, Cao X, Song X (2015) Epigenetic mutation of RAV6 affects leaf angle and seed size in Rice. **Plant Physiol** 169: 2118-2128
47. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker Joseph R (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. Cell 126: 1189–1201
48. Zhong S, Fei Z, Chen YR, Zheng Y (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol 31: 154–159

Declarations

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 31801403, No. 31871666, and No. 31761143005), the National Peanut Industry Technology System Construction (CARS-13), National Crop Germplasm Resources Center (NCGRC-2021-036), National Program for Crop Germplasm Protection of China (19210163), Central Public-interest Scientific Institution Basal Research Fund (No.1610172019003), and the Science and technology innovation Project of Chinese Academy of Agricultural Sciences.

Author contribution

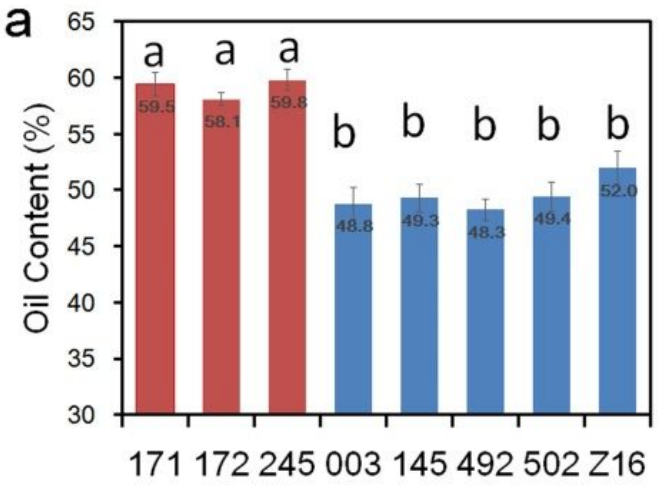
N.L., M.K.P., R.K.V., and H.J. conceived, designed, and supervised the experiments. L.H., Y.C., and X.Z. managed materials planted in the experimental nursery. N.L, B.W., W.C., and D.H. conducted sampling and

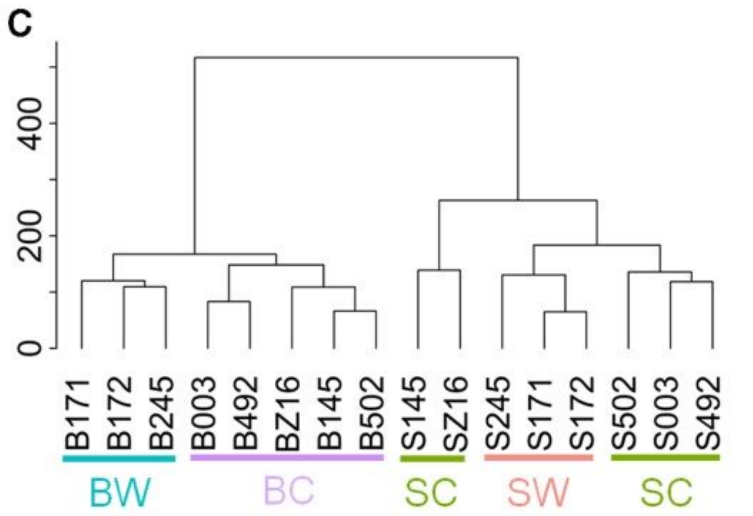
phenotyping. N.L. and B.W. performed DNA and RNA extraction. N.L. and M.K.P. performed bioinformatic analysis and interpreted the results. N.L. prepared the first draft and N.L., M.K.P., Y.L., B.L., R.K.V. and H.J. contributed to the final editing of manuscript. All authors read and approved the final manuscript.

Competing interests

No potential conflict of interest was reported by the authors

Figures





b

Sample	Group	Species	Tissue	Stage
S171	SW	*A. monticola*	Developing kernel	R5
S172	SW	*A. monticola*	Developing kernel	R5
S245	SW	*A. monticola*	Developing kernel	R5
S003	SC	*A. hypogaea*	Developing kernel	R5
S145	SC	*A. hypogaea*	Developing kernel	R5
S492	SC	*A. hypogaea*	Developing kernel	R5
S502	SC	*A. hypogaea*	Developing kernel	R5
SZ16	SC	*A. hypogaea*	Developing kernel	R5
B171	BW	*A. monticola*	Developing kernel	R8
B172	BW	*A. monticola*	Developing kernel	R8
B245	BW	*A. monticola*	Developing kernel	R8
B003	BC	*A. hypogaea*	Developing kernel	R8
B145	BC	*A. hypogaea*	Developing kernel	R8
B492	BC	*A. hypogaea*	Developing kernel	R8
B502	BC	*A. hypogaea*	Developing kernel	R8
BZ16	BC	*A. monticola*	Developing kernel	R8

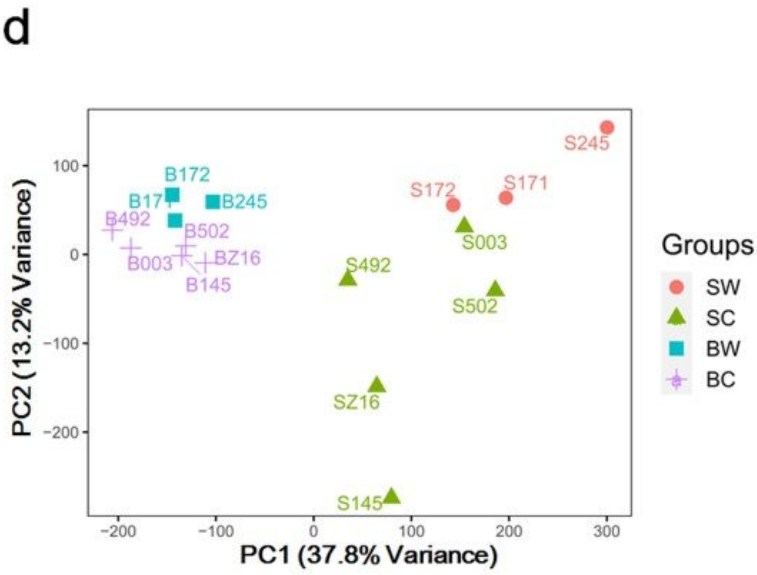


Figure 1

Analysis of wild and cultivated peanuts at two seed developmental stages. (a) Seed oil content of three wild and five cultivated accessions. (b) Illustration of 16 samples for RNA-seq. (c) Hierarchical clustering between 16 samples. (d) Principal component analysis (PCA) of gene expression profiles in 16 samples.

SW and SC indicated seeds of wild and cultivated peanuts at R5 stage, respectively. BW and BC denoted seeds of wild and cultivated peanuts at R8 stage, respectively.

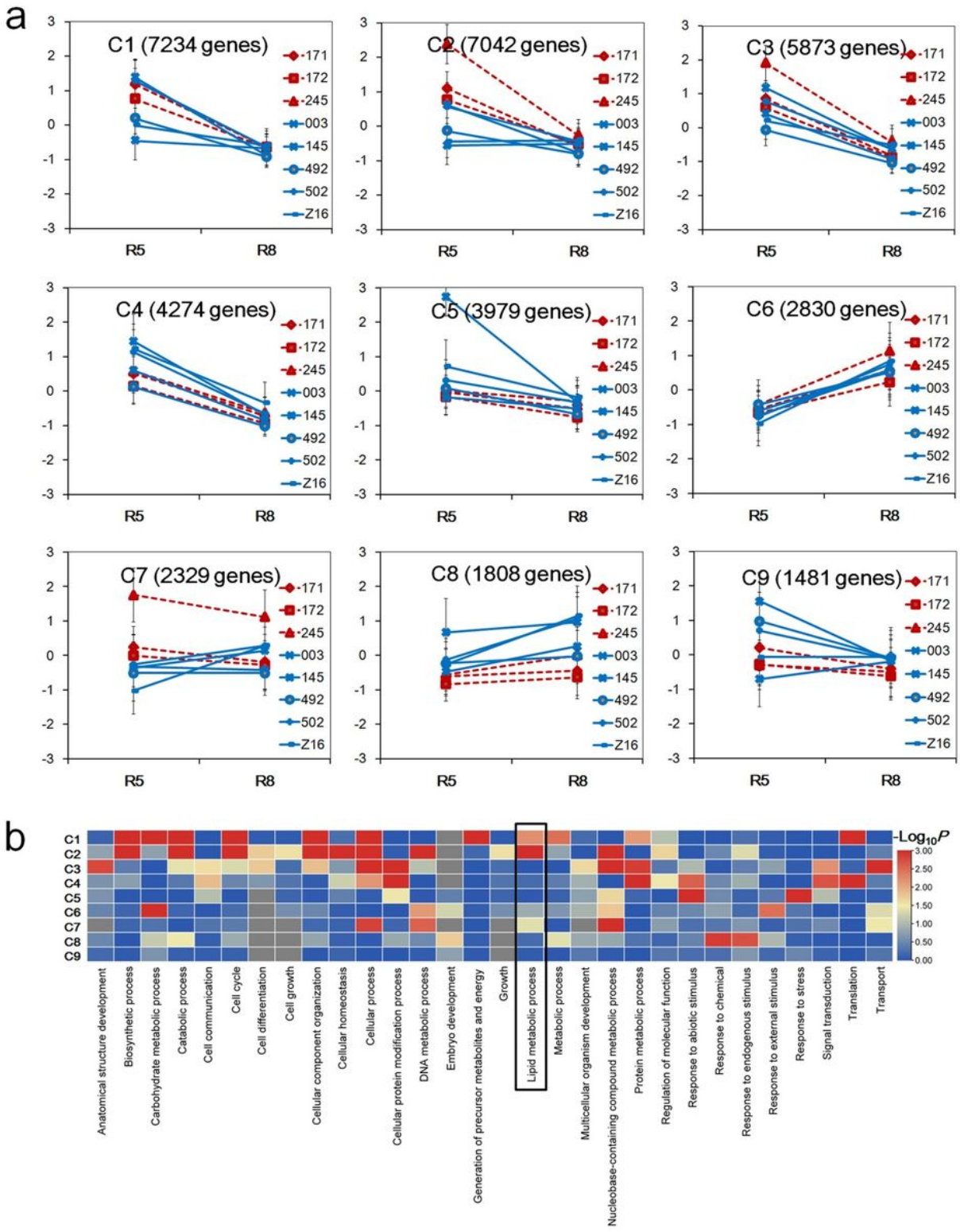


Figure 2

Gene expression pattern between wild and cultivated peanuts. (a) Identification of gene clusters in 16 samples. Nine gene clusters (C1-C9) were identified using K-means clustering. (b) Heatmap of enrichment

of biological process category among nine clusters. The color was indicated –log10 (P-Value). The lipid metabolism process was boxed in the heatmap.

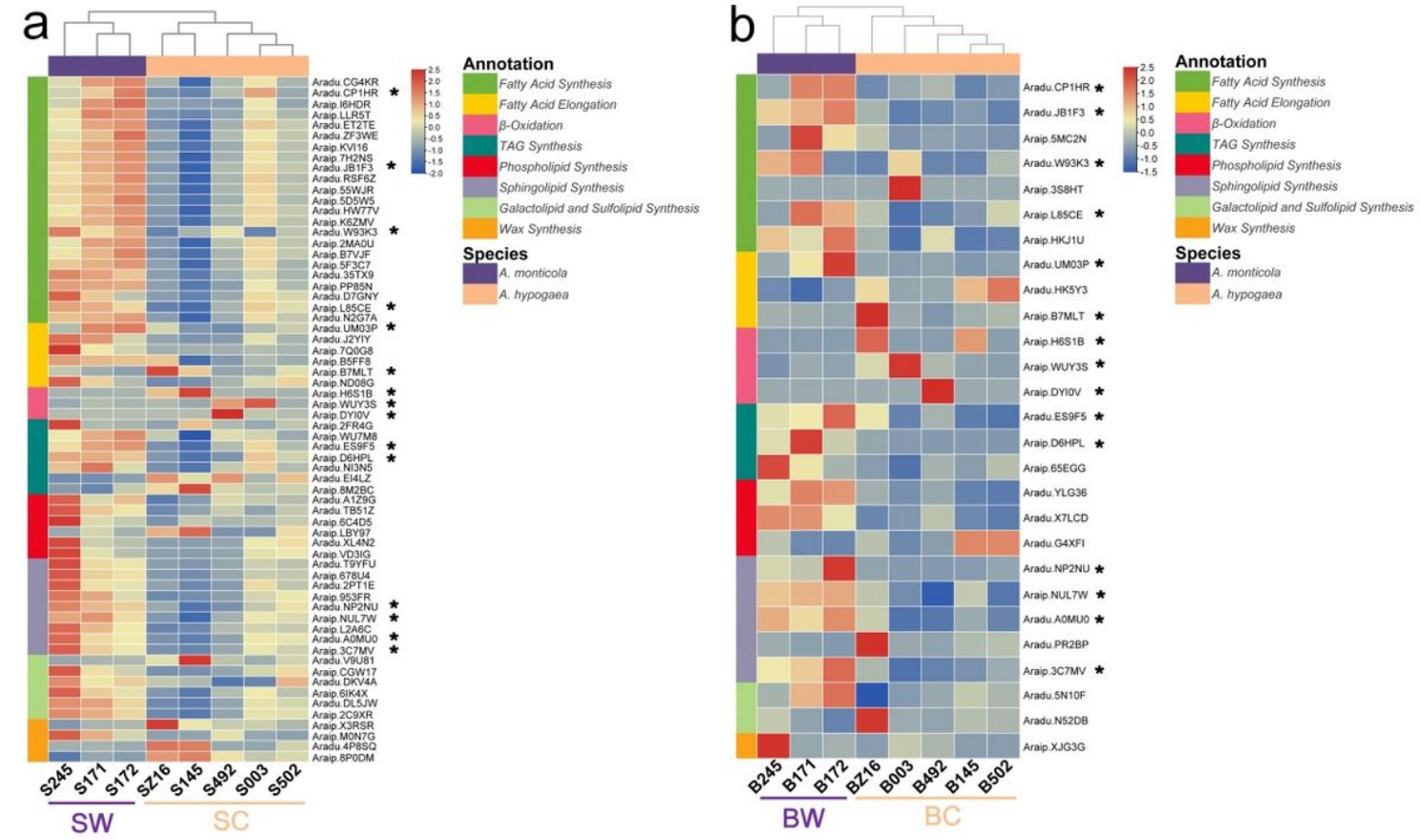


Figure 3

Heatmaps of lipid metabolism-related DEGs between wild and cultivated peanuts. (a and b) represented R5 and R8 stages, respectively. * denoted the differential expression genes identified at both stages. SW and SC indicated seeds of wild and cultivated peanuts at R5 stage, respectively. BW and BC denoted seeds of wild and cultivated peanuts at R8 stage, respectively.

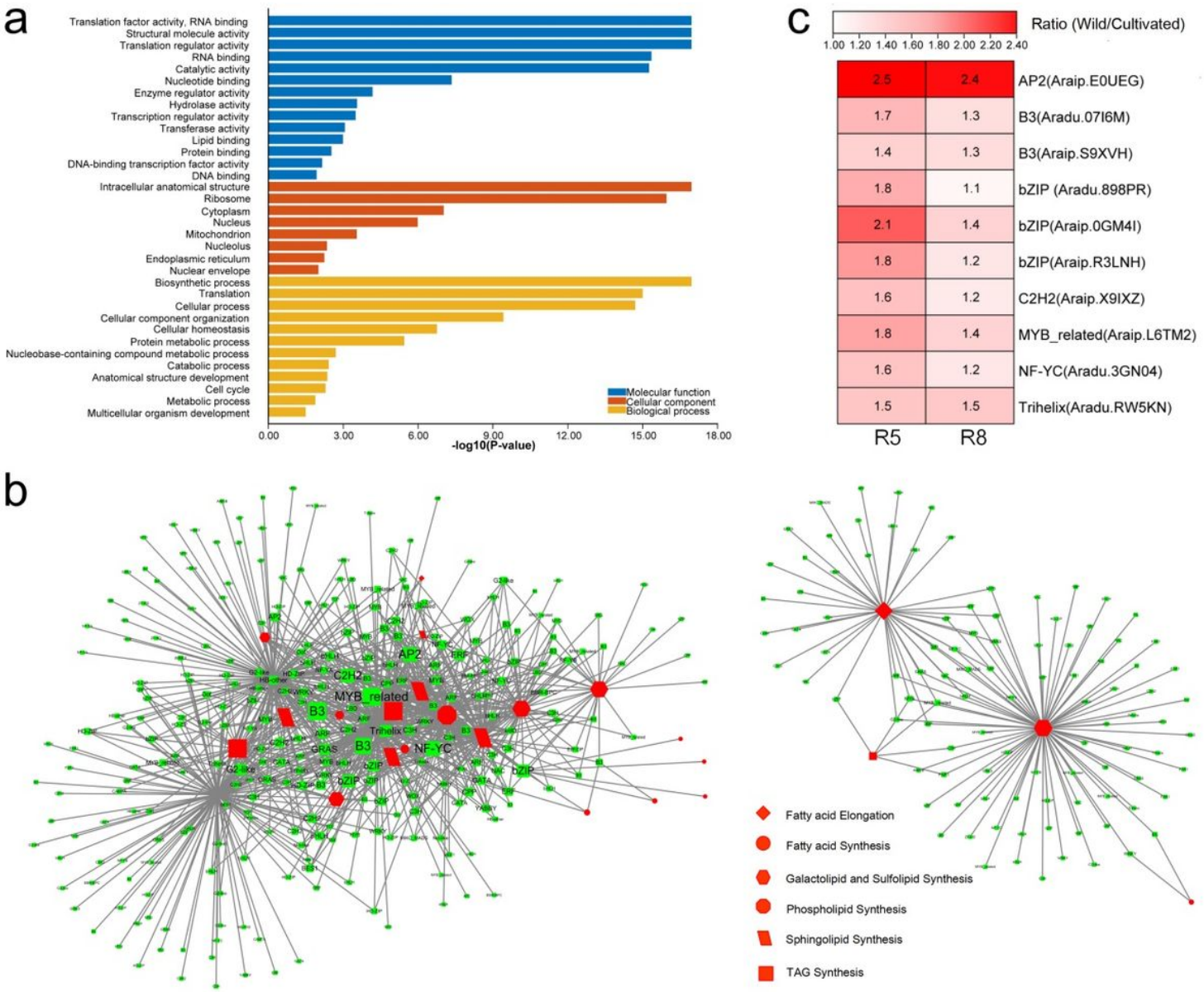


Figure 4

Identification of co-expression genes with DEGs involved in lipid metabolism. (a) GO enrichment analysis of co-expression genes. GO terms with p-value < 0.05 were listed. (b) Network of co-expression TFs with lipid metabolic DEGs. (c) Differential expression of top10 co-expression TFs between wild and cultivated peanuts.

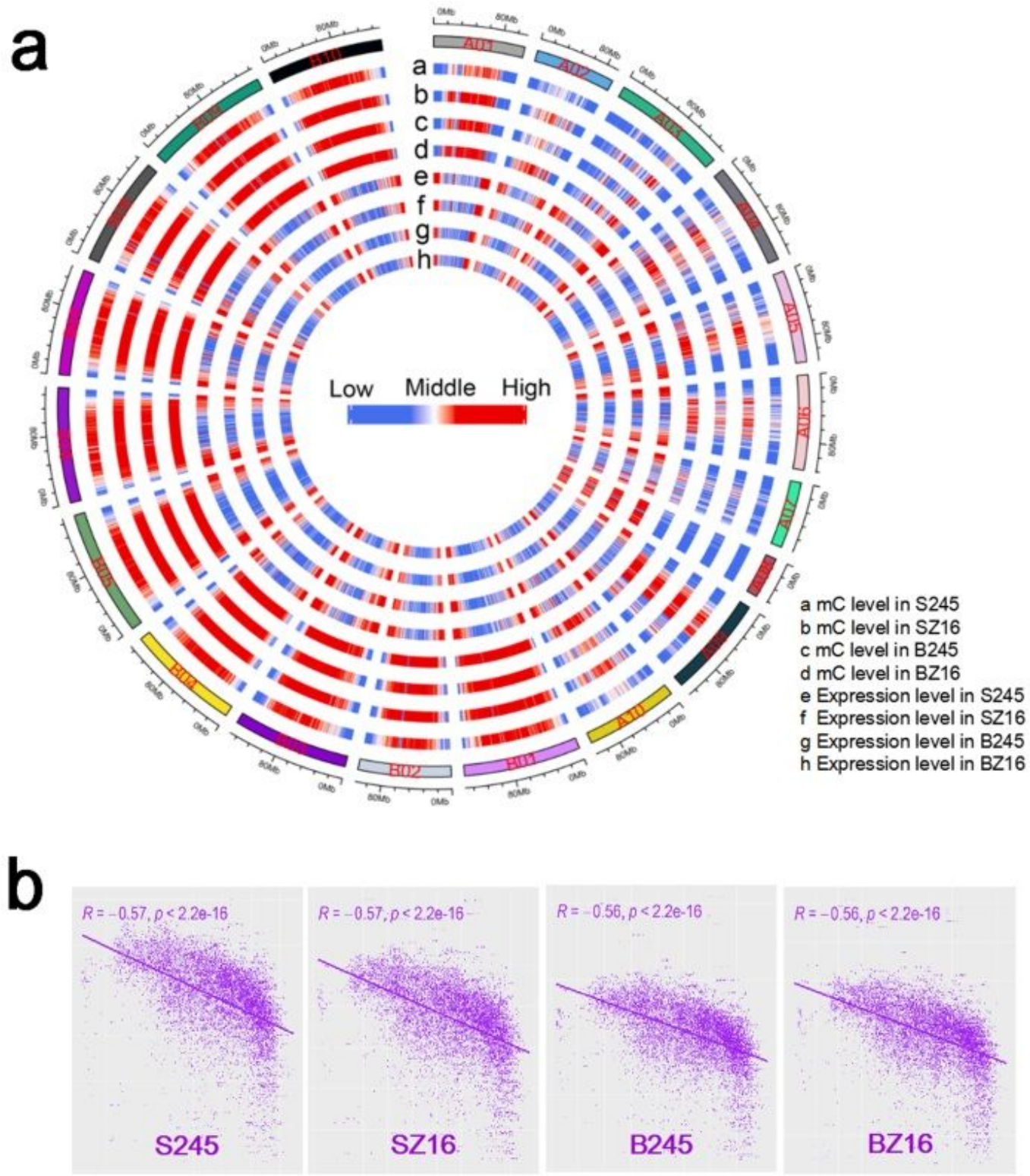


Figure 5

Genome-wide correlation between DNA methylation and gene expression. (a) A circle plot showed the overall DNA methylation level and gene expression level (log10 fpkm) in the reference chromosomes. The data for each chromosome was analyzed in 1 Mb windows sliding 200 kb. (b) Analysis of correlation coefficient between DNA methylation and gene expression. S245 and SZ16 indicated seeds of 245 and Z16 at R5 stage, respectively. B245 and BZ16 indicated seeds of 245 and Z16 at R8 stage, respectively.

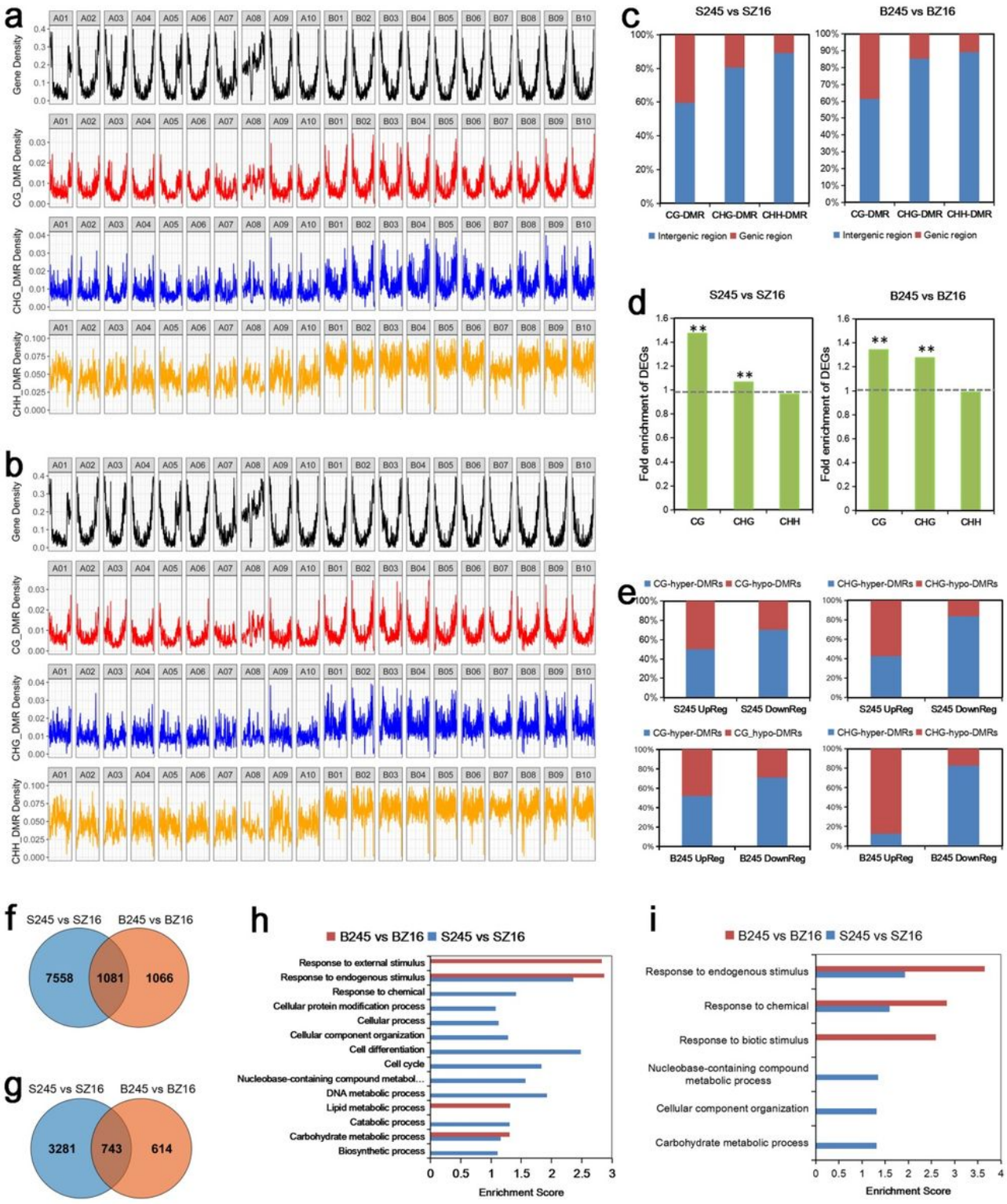
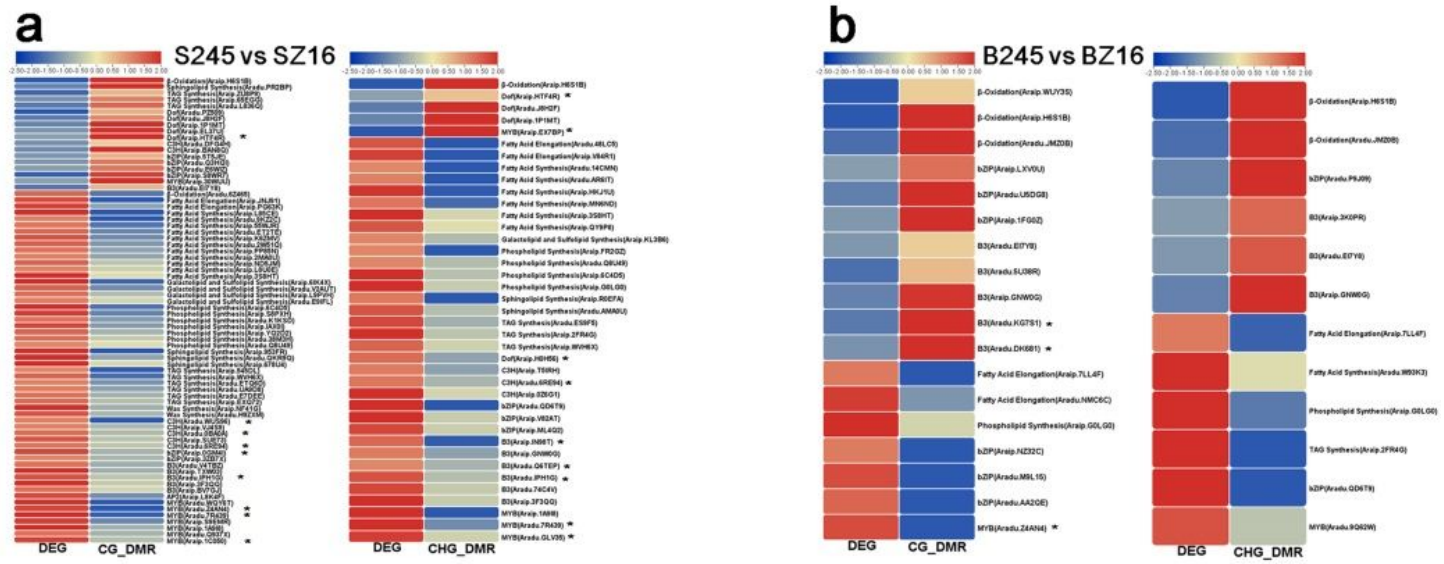
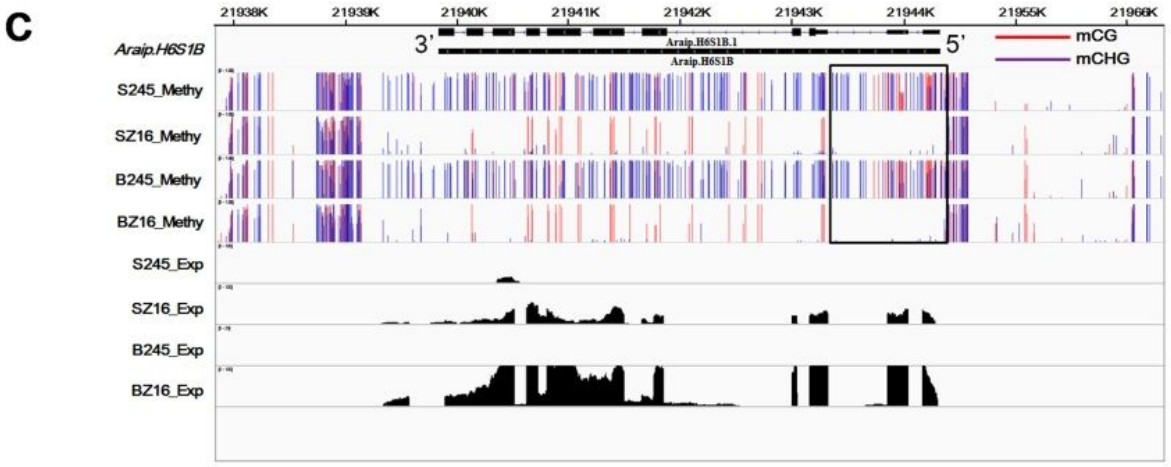


Figure 6

Influence of DMR on differential gene expression. (a and b) Genome-wide distribution of protein-coding genes and DMRs (CG, CHG, and CHH) between wild and cultivated seeds at R5 stage (a) and R8 stage (b). (c) Percentage of DMRs in the genic region and intergenic region. (d) Fold-enrichment of DEGs in those overlapping with DMRs in gene body with 2 kb upstream. The hypergeometric test was used to infer statistical significance (**: P < 0.01). (e) The ratio of (hyper/hypo) DMRs in CG or CHG contexts

overlapped with DEGs (up/down) between seeds of 245 and Z16 at R5 and R8 stages. (f) Venn diagram of CG-DMR associated DEGs at R5 and R8 stages. (g) Venn diagram of CHG-DMR associated DEGs between 245 and Z16 at R5 and R8 stages. (h) GO enrichment analysis of CG-DMR associated DEGs between 245 and Z16 at R5 and R8 stages. (i) GO enrichment analysis of CHG-DMR associated DEGs between 245 and Z16 at R5 and R8 stages. S245 and SZ16 denoted seeds of 245 and Z16 at R5 stage, respectively. B245 and BZ16 denoted seeds of 245 and Z16 at R8 stage, respectively.





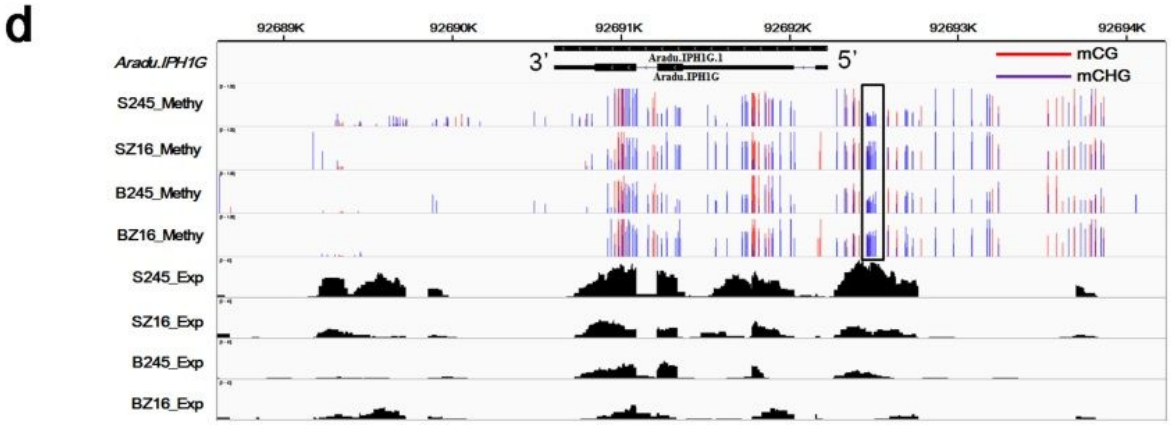


Figure 7

CG- or CHG-DMRs and their association with the expression of lipid metabolism-related genes. (a and b) Heatmaps showing CG- and CHG- DMRs between 245 and Z16 associated with differential expression at R5 (a) and R8 (b) stages for sets of enzymes and TFs involved in lipid metabolism, respectively. Scales represent differential expression and differential methylation in log2 fold-change. (c) A visualization showing the methylation level and expression level of acyl transporter (*Araip.H6S1B*) in seeds of 245 and Z16 at R5 and R8 stages. (d) A visualization showing the methylation level and expression level of B3 transcription factor (*Aradu.IPH1G*) in seeds of 245 and Z16 at R5 and R8 stages. S245 and SZ16 represented seeds of 245 and Z16 at R5 stage, respectively. B245 and BZ16 denoted seeds of 245 and Z16 at R8 stage, respectively.

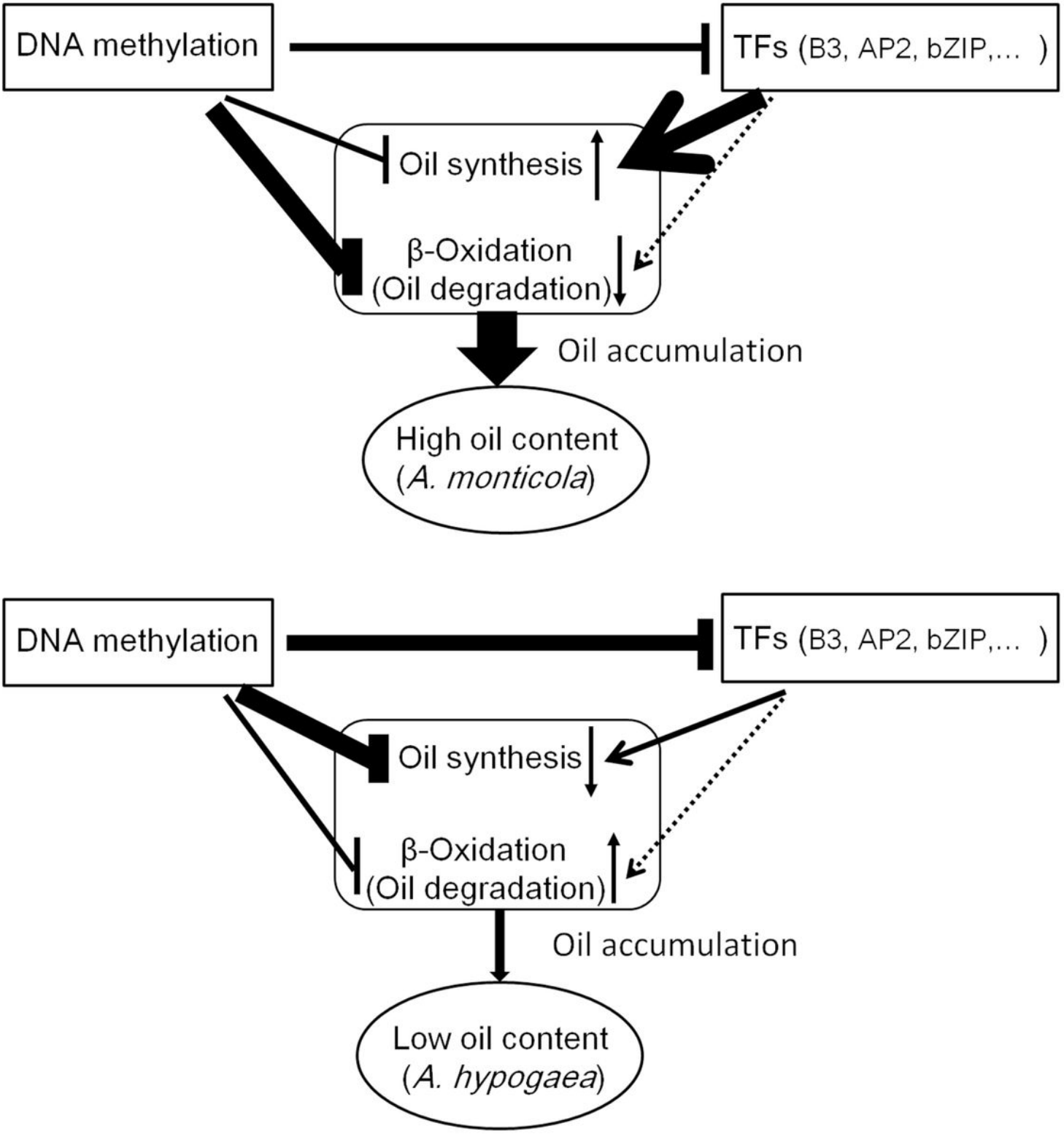


Figure 8

A regulatory model of lipid accumulation in peanut seed. Compared to *A. hypogaea*, *A. monticola* activates transcription of genes invovled in oil synthesis and depressed transcription of genes invovled in β-oxidatation via DNA methylation and TFs in seed. Thus, DNA methylation and TFs coordinately tuning expression of genes contributes to high oil accumulation in *A. monticola*.

Supplementary Files

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

- SupplementalTableS1.xlsx
- SupplementalTableS2.xlsx
- SupplementalTableS3.xlsx
- SupplementalTableS4.xlsx
- SupplementalTableS5.xlsx
- SupplementalTableS6.xlsx
- SupplementalTableS7.xlsx
- SupplementalTableS8.xlsx
- SupplementalTableS9.xlsx
- supplementalfigure.pdf
- AdditionalInformation.docx