the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) **103,** 1103–1124 doi: 10.1111/tpj.14788

Three-dimensional genetic networks among seed oil-related traits, metabolites and genes reveal the genetic foundations of oil synthesis in soybean

Jin-Yang Liu[1,2], Pei Li[2], Ya-Wen Zhang[2], Jian-Fang Zuo[2], Guo Li[2], Xu Han[2], Jim M. Dunwell[3] and Yuan-Ming Zhang[1,2,*]

[1]*State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China,*

[2]*Crop Information Center, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China, and*

[3]*School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, UK*

Received 11 December 2019; accepted 21 April 2020; published online 28 April 2020.
*For correspondence (e-mails soyzhang@njau.edu.cn; soyzhang@mail.hzau.edu.cn).

SUMMARY

Although the biochemical and genetic basis of lipid metabolism is clear in Arabidopsis, there is limited information concerning the relevant genes in *Glycine max* (soybean). To address this issue, we constructed three-dimensional genetic networks using six seed oil-related traits, 52 lipid metabolism-related metabolites and 54 294 SNPs in 286 soybean accessions in total. As a result, 284 and 279 candidate genes were found to be significantly associated with seed oil-related traits and metabolites by phenotypic and metabolic genome-wide association studies and multi-omics analyses, respectively. Using minimax concave penalty (MCP) and smoothly clipped absolute deviation (SCAD) analyses, six seed oil-related traits were found to be significantly related to 31 metabolites. Among the above candidate genes, 36 genes were found to be associated with oil synthesis (27 genes), amino acid synthesis (four genes) and the tricarboxylic acid (TCA) cycle (five genes), and four genes (*GmFATB1a*, *GmPDAT*, *GmPLDα1* and *GmDAGAT1*) are already known to be related to oil synthesis. Using this information, 133 three-dimensional genetic networks were constructed, 24 of which are known, e.g. pyruvate–*GmPDAT*–*GmFATA2*–oil content. Using these networks, *GmPDAT*, *GmAGT* and *GmACP4* reveal the genetic relationships between pyruvate and the three major nutrients, and *GmPDAT*, *GmZF351* and *GmPgs1* reveal the genetic relationships between amino acids and seed oil content. In addition, *GmCds1*, along with average temperature in July and the rainfall from June to September, influence seed oil content across years. This study provides a new approach for the construction of three-dimensional genetic networks and reveals new information for soybean seed oil improvement and the identification of gene function.

Keywords: seed oil-related traits, lipid-related metabolites, metabolome-based genome-wide association studies, three-dimensional genetic networks, soybean.

INTRODUCTION

Scientists have focused on the genetic basis of seed oil-related traits in *Glycine max* (soybean) for a long time, with the purpose of improving seed oil content and quality in this crop (Fang *et al.*, 2017). The significant negative correlation between seed oil and protein contents (Chaudhary *et al.*, 2015; Patil *et al.*, 2017) has resulted in very slow progress in improving soybean quality by means of conventional breeding, however (Charron *et al.*, 2005). Recently, metabolites, which act as a bridge between a trait phenotype and its genes, have been shown to determine crop nutritional traits like seed oil content and composition via a wide range of intermediate compounds, such as fatty acids, phospholipids and carbohydrates (Wen *et al.*, 2015; Chen *et al.*, 2016). Although many genes have been found to be associated with seed oil-related traits and lipid synthesis, these studies have usually involved phenotypic genome-wide association studies (GWAS) and linkage analysis (Hwang *et al.*, 2014; Meng *et al.*, 2016; Fang *et al.*, 2017; Van and McHale, 2017; Leamy *et al.*, 2017; Zuo *et al.*, 2019; Zhang *et al.*, 2019c). Therefore, modern crop breeding necessitates the construction of three-dimensional genetic networks among seed oil-related traits, genes and oil biosynthesis metabolites.

To date, many genes have been reported to be involved in seed oil biosynthesis in Arabidopsis. For example: *GPAT* (Li *et al.*, 2007), *PDHC* (Shen *et al.*, 2006), *ACCase* (Roesler *et al.*, 1994), *KASI* (Xiong *et al.*, 2017), and *FATB* and *FATA2* (Bonaventure *et al.*, 2003; Moreno-Pérez *et al.*, 2012) were found to be involved in the synthesis of short-chain fatty acids; *DGAT* and *PDAT* (Jako *et al.*, 2001; Zhang *et al.*, 2009; Fan *et al.*, 2013; Pan *et al.*, 2013) were found to be involved in triacylglycerol (TAG) biosynthesis; *LACS* (Lü *et al.*, 2009; Katavic *et al.*, 2014) was found to be involved in the synthesis of very long-chain fatty acids; *PLP2/PLA2A* (La Camera *et al.*, 2009; Yang *et al.*, 2012), *Pgs1* or *PGP1* (Tanoue *et al.*, 2014), *Cds1* (Zhou *et al.*, 2013), *LPEAT2* (Jasieniecka-Gazarkiewicz *et al.*, 2017) and *TIM/PDTPI* (López-Castillo *et al.*, 2016) were found to be involved in lipid synthesis; and *OLE1* (oleosin) was found to be involved in the storage of lipid droplets (Siloto *et al.*, 2006; Shimada and Hara-Nishimura, 2010). Although a hundred genes relating to lipid synthesis have been reported to participate in the process of carbohydrate metabolism (Zhang *et al.*, 2018b), few genes have been reported to be related to the tricarboxylic acid (TCA) cycle and amino acid synthesis (Wen *et al.*, 2015; Zhang *et al.*, 2018a,b). In Arabidopsis, *SDH1* (Huang *et al.*, 2013), *ACO1* (Park *et al.*, 2018), *MDH* (Selinski and Scheibe, 2019), *FUM1* (Zubimendi *et al.*, 2018), *IDH-V* (Lemaitre and Hodges, 2006) and *2-OGDH* (Araújo *et al.*, 2014) were reported to participate in reactions of the TCA cycle; *AGT* (Zhang *et al.*, 2013), *P5C1* (Giberti *et al.*, 2014), *MTO* (Goto and Naito, 2002), *HMT2* (Ranocha *et al.*, 2000) and *AtBCAT* (Diebold *et al.*, 2002) were reported to participate in amino acid metabolism.

In soybean, some transcription factors and genes encoding other functional proteins have been reported to be responsible for seed oil biosynthesis. The transcription factors *GmDof4*, *GmDof11* (Wang *et al.*, 2007), *GmbZIP123* (Song *et al.*, 2013), *GmLEC1a/GmLEC1b* (Zhang *et al.*, 2017a), *GmWRI1a* (Chen *et al.*, 2017), *GmMYB73* (Liu *et al.*, 2014), *GmDREBL* (Zhang *et al.*, 2016b), *GmNFYA* (Lu *et al.*, 2016), *GmLEC2* (Manan *et al.*, 2017) and *GmZF351* (Li *et al.*, 2017) were found to participate in the regulation of lipid accumulation. The functional genes *GmDGAT1* or *GmDAGAT1* (Lardizabal *et al.*, 2008; Chen *et al.*, 2016) and *GmOLE1* (desaturase) (Zhang *et al.*, 2019a) were reported to play a key role in plant diacylglycerol/triacylglycerol (DAG/TAG) biosynthesis, and *GmPLD* (phospholipase D) and *GmLPAT* (lysophosphatidyl acyltransferase) (Zhao *et al.*, 2012; Zhao, 2013) were found to regulate lipid synthesis. Rarely have oil-synthesis genes been reported to be related to the TCA cycle or amino acid synthesis in soybean, however.

As we all know, metabolites have a significant influence on signal transmission, material synthesis and decomposition, and other differentiation processes in each cell (Chen *et al.*, 2014, 2016; Wen *et al.*, 2015). Recently, using metabolome-based GWAS (mGWAS) and metabolome profiling analysis, some genes have been identified to be associated with primary or secondary metabolites, which are responsible for complex traits (Chen *et al.*, 2016; Wu *et al.*, 2018). In Arabidopsis, for example, *OMT1* encoding 5-hydroxyferulic acid *O*-methyltransferase was found to regulate 5-hydroxyferulic acid glucoside (Wu *et al.*, 2018), which influences the synthesis of lignins and sinapoyl esters (Tohge *et al.*, 2005). In *Oryza sativa* (rice), *Os07g32060*, encoding flavone 5-*O*-glucosyltransferase, was found to regulate 5-*O*-glucoside, which influences the synthesis of flavonoids (Chen *et al.*, 2014), *Os12g27220* and *Os12g27254*, encoding spermidine hydroxycinnamoyl transferases, were found to regulate *N*-hydroxycinnamoyl spermidines, which influences phenolamide biosynthesis (Dong *et al.*, 2015), and *Os02g57760*, encoding nicotinic acid *N*-methyltransferase, was found to regulate trigonelline, which influences grain width (Chen *et al.*, 2016). At present, the studies on soybean mGWAS are relatively limited.

As described above, genetic relationships are derived mainly from seed oil-related traits and genes or metabolites and genes. In modern breeding strategies, it is crucial to construct three-dimensional genetic networks among seed oil-related traits, metabolites and genes. To address this issue, six seed oil-related traits, 52 lipid-related metabolites and 54 294 SNP markers in a total of 286 soybean accessions were used to conduct single- and multi-locus GWAS (Zhou *et al.*, 2015; Zhou *et al.*, 2015; Wang *et al.*, 2016; Tamba *et al.*, 2017; Zhang *et al.*, 2017b; Wen *et al.*, 2018; Ren *et al.*, 2018) for seed oil-related traits and metabolites, and genetic relationships between seed oil-related traits and metabolites were also established by the minimax concave penalty (MCP) (Zhang, 2010) and smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001) analyses. Candidate genes for seed oil-related traits and metabolites were predicted by bioinformatics, comparative genomics and transcriptomics. Using the above results, 133 three-dimensional genetic networks were constructed in this study. Using these networks, some new genetic relationships were revealed, e.g. the relationships between pyruvate and three major nutrients, and between amino acids and seed oil content. In addition, we also discuss the reasons for different seed oil contents across different years. Thus, this study provides a new approach for constructing three-dimensional genetic networks, which reveal some new genetic relationships among seed oil content, some metabolites (three major nutrients, malic acid and amino acids) and genes. These relationships are useful for soybean quality improvement and the identification of gene functions.

RESULTS

Distribution of six seed oil-related traits and 52 metabolites in soybean

The seed oil-related traits investigated in this study are seed oil content and its five oil constituents, including stearic acid, palmitic acid, oleic acid, linoleic acid and linolenic acid. These traits were measured from 286 soybean accessions between 2014 and 2016. The averages plus standard deviations across the 3 years for the above six traits were 17.92 ± 2.16, 3.54 ± 0.46, 11.65 ± 1.21, 24.79 ± 4.53, 52.29 ± 3.63 and 7.73 ± 1.58 (%), respectively, and their average coefficients of variation (CV) across the three years were 12.03, 10.33, 12.92, 18.24, 6.95 and 20.40 (%), respectively (Table S1). Clearly, these traits have large variation and are typical quantitative traits. Although the trends for the five seed oil constituents over the 3 years are almost the same (Figure 1a–e), the seed oil content in 2016 (16.67 ± 1.92%) was significantly lower than those recorded in 2014 (19.06 ± 2.18%) and 2015 (18.03 ± 2.37%) ($P < 0.001$).

A total of 52 lipid-related metabolites in the pathways of the TCA cycle, amino acid metabolism, oil synthesis and soybean isoflavone synthesis were measured from 214 soybean accessions in 2015. These metabolites are classified into organic acids, soybean isoflavone, phosphatidyl ethanolamines (PEs), phosphatidyl cholines (PCs), phosphatidyl inositols (PIs) and amino acids. Organic acids measured in this study included pyruvic acid, succinic acid, fumaric acid, malic acid, palmitic acid, stearic acid, oleic acid, linoleic acid and linolenic acid. Their phenotypic values were in the ranges of 175.87–50 980.18, 1.35–515.01, 1.25–440.91, 18.61–5280.87, 0.90–342.63, 0.50–105.69, 0.15–112.67, 21.71–774.08 and 8.5–102.43 μg g^{-1}, respectively, and their CVs were 181.85, 123.82, 113.08, 82.37, 79.92, 75.57, 126.59, 90.47 and 45.02%, respectively. Soybean isoflavones measured in this study included daidzein, daidzin, genistein, genistin and glycitin. Their phenotypic values were in the ranges of 0.23–163.78, 0.50–314.13, 0.22–87.65, 7.78–1611.42 and 0.002–238.69 μg g^{-1}, respectively, and their CVs were 107.06, 110.34, 104.93, 74.56 and 109.61%, respectively. The phenotypic values for PEs (six), PIs (six) and PCs (six), with 18 molecular species (for detail information, see Experimental Procedures) were in the ranges of 3.02–2160.52, 0.00–30 568.93 and 0.00–2830.26 μg g^{-1}, respectively, and their CVs were 91.88, 124.53 and 96.34%, respectively. A total of 20 amino acids were measured: their phenotypic values ranged from 0.04 to 1864.51 μg g^{-1}, and their CVs ranged from 41.89 to 236.48%. Detailed information for all 52 metabolites

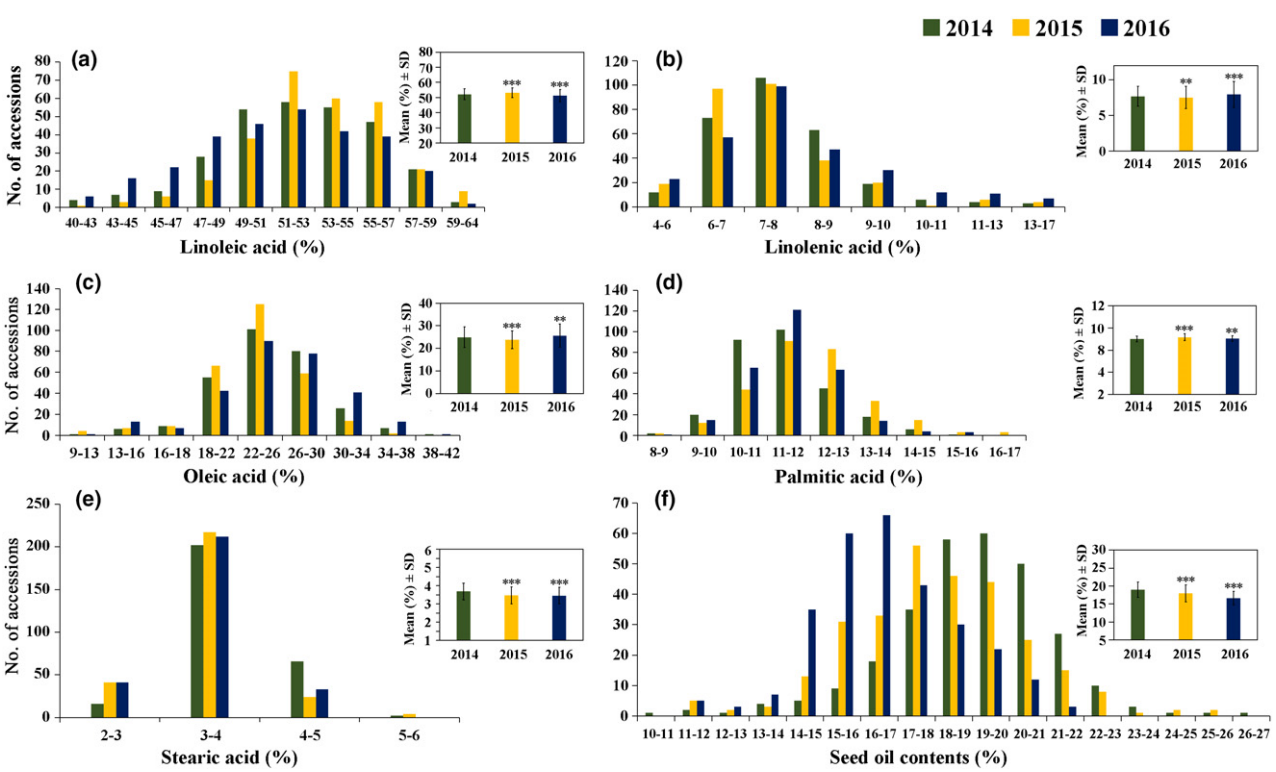

Figure 1. Frequent distributions for seed oil content (f) and constituents (a–e) in 286 soybean accessions. The results in 2014, 2015 and 2016 are indicated by green, yellow and navy-blue bars, respectively. Data are means ± standard deviations. Levels of significance by paired Student's *t*-test ($n = 286$): $*P = 0.05$, $**P = 0.01$ and $***P = 0.001$.

is given in Table S2. Clearly, these metabolites have large variations.

Genome-wide association studies for seed oil-related traits in soybean

Detection of main-effect quantitative trait nucleotides (QTNs) for oil-related traits. With 286 soybean accessions, six seed oil-related traits measured from 2014 to 2016, along with 54 294 SNPs, were used to conduct phenotypic GWAS using GEMMA, mrMLM, FASTmrEMMA, ISIS EM-BLASSO, pLARmEB and pKWmEB. As a result, 334 significant QTNs were identified (Figure S1 and Table S3). These QTNs were mainly distributed on chromosomes 5, 6, 7, 8, 9, 13, 17, 18 and 19 (≥16 QTNs for each chromosome), and had 5.51% average proportion of total phenotypic variation explained by each QTN, with 56, 46, 50, 68, 75 and 39 QTNs, respectively, for palmitic, stearic, oleic, linoleic and linolenic acids, and seed oil content. Thirty-five QTNs were detected in at least two environments, whereas 299 QTNs were identified in only one environment. A total of 77 significant QTNs for the above six oil-related traits were detected in at least two environments or two GWAS methods (Table S4). Among these common QTNs there were 11, 17, 12, 18, 7 and 12 QTNs, respectively, for linolenic, linoleic, stearic, oleic and palmitic acids, and seed oil content. Based on previous studies (https://www.soybase.org/GWAS/), there are many QTNs on chromosome 5 and almost no QTNs on chromosome 13. In this study, five significant QTNs were positioned within 38.0–41.0 Mb at the distal end of chromosome 5 and eight QTNs were positioned on chromosome 13.

Detection of QTN–environment interactions for oil-related traits. The above data sets in GWAS were also used to detect QTN–environment interactions (QEs) using the quantitative trait interaction (G × E) module in PLINK (Purcell *et al.*, 2007) (http://zzz.bwh.harvard.edu/plink/anal.shtml#qtgxe). As a result, 5, 1 and 3 significant QEs were found to be associated with linolenic acid, palmitic acid and stearic acid, respectively (Table S5). For example, the locus Chr18-4720420 was significantly associated with linolenic acid (P = 6.53E–04).

Detection of QTN–QTN interactions for oil-related traits. The above data sets in GWAS were again used to detect QTN–QTN interactions (QQs) using PEPIS (http://bioinfo.noble.org/PolyGenic_QTL/) (Zhang *et al.*, 2016c). As a result, 2, 2, 3, 1, 1 and 1 significant QQs were found to be associated with linoleic acid, seed oil content, palmitic acid, oleic acid, stearic acid and linolenic acid, respectively (Table S6). For example, the epistasis between locus Chr13-20532852 bp and locus Chr13-20704034 bp was found to be significantly responsible for linolenic acid (likelihood-ratio test, LRT = 24.37).

Candidate genes for seed oil-related traits. In order to determine candidate genes for seed oil-related traits, we adopted the following analyses. First, we found all genes between the 100-kb upstream and downstream regions for each of the 334 significant QTNs. Using the soybean metabolic pathway database, KEGG annotation (https://soycyc.soybase.org/), and the soybean genome annotation database and gene ontology terms (https://soybase.org/genomeannotation/), all of the above genes were then used to mine the potential candidate genes or their Arabidopsis homologous genes, which were annotated for fatty acid biosynthesis, phospholipid biosynthesis, phospholipid binding, phosphorylation and dephosphorylation, triacylglycerol biosynthesis, oxidoreductase activity, electron carrier activity and TCA cycle pathways. As a result, 284 genes were found to be associated with the above metabolic pathways.

Among the above 284 genes, 22 were found to be related to lipid metabolism pathways, including 14 lipid biosynthesis-related genes, four amino acid biosynthesis-related genes and four TCA cycle-related genes. In oil biosynthesis-related genes, *GmPDAT*, *GmDAGAT1*, *GmFATB1a*, *GmKASI*, *GmPgs1*, *GmACC*, *GmFATA2*, *GmCds1*, *GmWRI1b*, *GmNFYA*, *GmDof11*, *GmCYP78A10*, *Glyma.18g038400* and *GmBS1* were found to be associated, respectively, with linolenic acid (limit of detection, LOD = 4.15–4.20) and pyruvate (P = 1.44E–05) (Liu, 2020), linolenic acid (P = 8.28E–09 to 1.58E–06) (Chen *et al.*, 2016), stearic acid (LOD = 2.61–5.13) (Murad *et al.*, 2014), palmitic acid (LOD = 3.09) (Xiong *et al.*, 2017), linoleic acid (LOD = 4.86) (Tanoue *et al.*, 2014), oil content (LOD = 3.11–5.31) (Roesler *et al.*, 1994), oil content (LOD = 3.21) (Moreno-Pérez *et al.*, 2012), linolenic acid (P = 1.56E–09) (Zhou *et al.*, 2013), palmitic acid (LOD = 3.59) (Chen *et al.*, 2017), oleic acid (P = 3.82E–06) (Lu *et al.*, 2016), linolenic acid (LOD = 3.95) (Wang *et al.*, 2007), linolenic acid (LOD = 2.88) (Wang *et al.*, 2015), palmitic acid (LOD = 3.37–3.76) and palmitic acid (LOD = 5.25) (Ge *et al.*, 2016). Among these genes, *GmWRI1b*, *GmNFYA* and *GmDof11* have no annotations of biochemical metabolic processes; *GmPDAT*, *GmDAGAT1, GmFATB1a, GmPgs1* and *GmFATA2* were differentially expressed between wild and domesticated soybeans (Figure 2b; Table 1). In amino acid biosynthesis-related genes, *GmAGT*, *GmBCAT*, *GmHMT2* and *GmP5C1* were found to be associated, respectively, with palmitic acid (LOD = 3.39) (Zhang *et al.*, 2013), palmitic acid (LOD = 4.70) (Diebold *et al.*, 2002), oleic acid (P = 2.49E–09) (Ranocha *et al.*, 2000) and linoleic acid (LOD = 3.84) (Giberti *et al.*, 2014). In TCA cycle-related genes, *GmACO1* (*Glyma.01g162800*), *GmFUM1* (*Glyma.02g015700*), *GmSDH1* (*Glyma.01g175600*) and *GmMDH1* (*Glyma.13g104800*) were found to be associated, respectively, with oleic acid (P = 4.34E–06) (Park *et al.*, 2018), linolenic acid (P = 1.25E–06) (Zubimendi *et al.*, 2018), linoleic acid (LOD = 3.29–3.68)

(Huang *et al.*, 2013) and linolenic acid, P = 2.24E–07) (Selinski and Scheibe, 2019) (Figure 2a; Table 2).

Genome-wide association studies for acyl lipid-related metabolites in soybean

Genome-wide association studies for acyl lipid-related metabolites. In 214 soybean accessions, 52 acyl lipid-related metabolites measured in 2015, along with 54 294 SNPs, were used to conduct metabolic GWAS using GEMMA, mrMLM, FASTmrEMMA, ISIS EM-BLASSO, pLARmEB and pKWmEB. As a result, 1001 mQTNs were detected to be associated with the 52 acyl lipid metabolites (Figure S2; Table S7). These QTNs were distributed mainly on chromosomes 5, 7, 8, 13–18 and 20 (≥50 mQTNs for each chromosome), had an average proportion of total phenotypic variation explained by each mQTN of 6.63%, and 230, 115, 66, 111, 96 and 383 SNPs were identified to be significantly associated with nine organic acids, five soybean isoflavones, six PEs, six PIs, six PCs and 20 amino acids in soybean, respectively (Figure S2). Forty-eight mQTNs were detected by at least two approaches (Table S8). In addition, there were some large-effect mQTNs, e.g. mQTNs Chr4-3969004, Chr5-2665256, Chr8-17117978 and Chr18-62242431 were found by ISIS EM-BLASSO to be associated with glutamic acid (r^2 = 21.15%), PI (34:3) (r^2 = 9.31%), malate (r^2 = 4.97%) and isoleucine (r^2 = 6.75%), respectively, and mQTN Chr20-45754357 was found by mrMLM to be associated with pyruvate (r^2 = 6.18%).

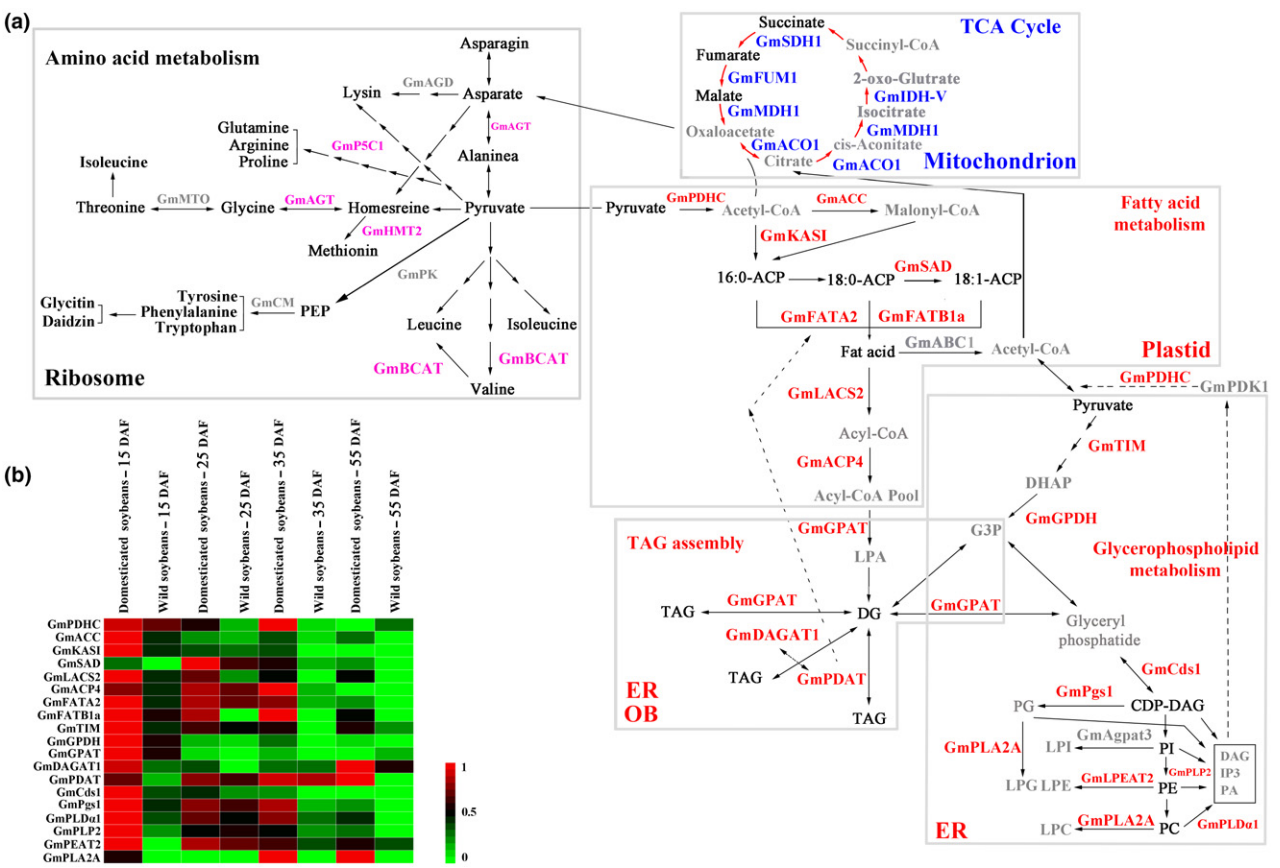

Figure 2. (a) The primary metabolic networks in soybean and (b) the expression profiling of 19 key seed oil-related genes identified in this study. Genes with red, pink and blue colors are in the pathways of oil biosynthesis, amino acid biosynthesis and TCA cycle, respectively. The metabolites and genes in gray were not identified in this study. Abbreviations: *ABC1*, activity of bc1 complex homolog 1; *ACC*, acetyl coenzyme-A carboxylase; *ACO1*, acyl-CoA oxidase 1; *ACP4*, acyl carrier protein (ACP)-4; *AGD*, diaminopimelate aminotransferase; *BCAT*, branched-chain amino acid transaminase; *AGT*, alanine glyoxylate aminotransferase; *Agpat3*, acylglycerophosphate acyltransferase; *CDS1*, CDP-diacylglycerol synthase 1; *CM*, chorismate mutase; *DAGAT1*, diacylglycerol acyltransferase enzymes 1; *FATA*, fatty acid thioesterase A; *FATB*, fatty acid thioesterase B; *LACS*, long-chain fatty acyl-CoA synthetase; *FUM1*, fumonisin synthase gene 1; *GPAT*, glycerol-3-phosphate acyltransferase; *GPDH*, glycerol phosphate dehydrogenase; *HMT2*, homocysteine *S*-methyltransferase 2; *IDH-V* isocitrate dehydrogenase V; *KASI*, β-Ketoacyl-ACP synthase I; *LPEAT2*, lyso-PE acyltransferase 2; *MDH*, malate dehydrogenase; *MTO*, mitochondrial tRNA modification gene; *P5C1*, pyrroline-carboxylic acid synthase 1; *PDAT1*, phospholipid diacylglycerol acyltransferase 1; *PDHC*, pyruvate dehydrogenase complex; *PDK1*, pyruvate dehydrogenase kinase 1; *Pgs1*, phosphatidylglycerolphosphate synthase; *PLA2A*, phospholipase A2; *PK*, pyruvate kinase *PLDα1*, phospholipase D gene 1; *PLP2*, proteolipid protein 2; *SAD*, sinapyl alcohol dehydrogenase; *SDH1*, succinate dehydrogenase 1; *TIM*, translocases inner mitochondrial membrane. DAF: days after flowering. Domesticated soybeans include four accessions with high seed oil content; wild soybeans include two accessions with low seed oil content.

Table 1 Twenty-two key candidate genes derived from genome-wide association studies for seed oil-related traits

	Genome-wide association studies				Comparative genomics					
Trait	Chr.	Position	LOD score or *P*	Method, year[a]	Candidate genes		Arabidopsis homologs	Functional annotation	*P*[b]	References
Oil content	18	42 441 603	1.47E−05	6, 2014	*Glyma18g36130*	*GmFATA2*	*AT4G13050*	Acyl-ACP thioesterase	0.050*	Moreno-Pérez *et al.* (2012)
	18	58 420 889	3.11–5.31	1, 2014; 3, 2014 & 2015	*Glyma18g50020*	*GmACC*	*AT5G15530.1*	Fatty-acid biosynthetic process	0.121	Klaus *et al.* (2004)
Linolenic acid	2	1 549 143	1.67E−08	6, 2015	*Glyma02g01920*	*GmFUM1*	*AT2G47510.1*	Fumarase 1	0.083	Zubimendi *et al.* (2018)
	5	247 186	2.88	2, 2014	*Glyma05g00220*	*GmCYP78A10*	*AT1G74110*	Control of seed size in soybean	0.086	Wang *et al.* (2015)
	13	20 274 945	2.14E−6	6, 2014	*Glyma.13g104800*	*GmMDH1*	*AT2G22780.1*	Peroxisomal NAD-malate dehydrogenase 1	0.070	Selinski and Scheibe (2019)
	13	20 532 852	8.28E−09–1.58E−06	6, 2014 & 2015	*Glyma13g16560*	*GmDAGAT1*	*AT2G19450.1*	Diacylglycerol acyltransferase 1	0.013*	Chen *et al.* (2016)
	13	20 704 034	3.17E−06	6, 2014	*Glyma13g16790*	*GmPDAT*	*AT2G19450.1*	Diacylglycerol acyltransferase 1	0.016*	Liu (2020)
	13	40 977 541	3.95	2, 2016	*Glyma.13g40420*	*GmDof11*	*AT2G28510*	Increase the content of total fatty acids and lipids	0.180	Wang *et al.* (2007)
	18	4 720 420	1.56E−09	6, 2014	*Glyma.18g055100*	*GmCds1*	*AT2G45150.3*	Phosphatidylglycerol biosynthesis I	0.170	Zhou *et al.* (2013)
	18	62 146 771	4.86	4, 2015	Glyma18g54020	GmPgs1	AT2G39290.1	Phosphatidylglycerolphosphate synthase 1	0.022*	Tanoue *et al.* (2014)
Linoleic acid	1	51 429 468	3.29–3.68	4 & 5, 2016	*Glyma05g33940*	*GmSDH1*	*AT5G66760.1*	Succinate dehydrogenase 1	0.055	Huang *et al.* (2013)
	3	36 244 172	3.84	5, 2015	*Glyma03g28476*	*GmP5C1*	*AT5G14800*	1-Pyrroline-5-carboxylate reductase	0.002*	Giberti *et al.* (2014)
Oleic acid	1	49 157 127	7.08E−06	6, 2014	*Glyma01g36750*	*GmACO1*	*AT4G35830.1*	Aconitase 1	0.031*	Park *et al.* (2018)
	2	50 913 342	3.82E−06	6, 2014	*Glyma02g47380*	*GmNFYA*	*AT3G20910.1*	Nuclear factor Y, subunit A	0.057	Lu *et al.* (2016)
	3	39 102 918	1.45E−08	6, 2014	*Glyma03g31281*	*GmHMT2*	*AT3G63250.1*	Homocysteine methyltransferase 2	0.176	Ranocha *et al.* (2000)
Stearic acid	20	36 599 310	4.94–5.38	1 & 3, 2014; 2 & 4, 2015	*Glyma05g08060*	*GmFATB1a*	*AT1G08510.1*	Fatty acyl-ACP thioesterases B	0.041*	Murad *et al.* (2014)
Palmitic acid	4	4 161 316	4.70	1, 2014	*Glyma04g05190*	*GmBCAT*	*AT5G28680.1*	Serine/threonine protein kinase	0.322	Diebold *et al.* (2002)
	8	6 430 244	3.71	1, 2016	*Glyma08g08910*	*GmKASI*	*AT5G46290.1*	β-Ketoacyl-acyl carrier protein synthase I	0.234	Xiong *et al.* (2017)
	8	16 829 990	3.59	4, 2015	*Glyma08g24420*	*GmWRI1b*	*AT3G54320.1*	Regulate the synthesis of fatty acids and triacylglycerols	0.098	Chen *et al.* (2017)
	8	41 399 047	3.39	4, 2014	*Glyma.08g302600*	*GmAGT*	*AT2G13360.1*	Glycine biosynthesis III		Zhang *et al.* (2002)
	10	46 681 643	5.25	1, 2016	*Glyma10g38970*	*GmBS1*	*AT4G14720.1*	Seed size-related gene	0.106	Ge *et al.* (2016)
	18	3 091 833	3.37–3.76	2 & 4, 2015	*Glyma.18g038400*	*Glyma.18g038400*	*AT3G55470.2*	Phospholipid-binding protein		

[a]Methods ISIS EM-BLASSO, mrMLM, FASTmrEMMA, pLARmEB, pKWmEB and GEMMA are indicated by the numerals 1–6, respectively.
[b]The *P* values were calculated using the paired Student's *t*-test from the average RPKM values at four stages between cultivated (high seed oil, $n_1 = 4$) and wild (low seed oil, $n_2 = 2$) soybeans, with significance indicated by an asterisk (*0.05 level).

Candidate genes associated with metabolites. The methodologies of determining the candidate genes for acyl lipid-related metabolites were the same as those for the above seed oil-related traits. First, we found all the genes between the 100-kb upstream and downstream regions for each of the significant mQTNs. Using the soybean metabolic pathway database, KEGG annotation (https://soycyc.soybase.org/), and soybean genome annotation database and gene ontology terms (https://soybase.org/genomeannotation/), all of the above genes were then used to mine the potential candidate genes or their Arabidopsis homologous genes, which were annotated for fatty acid biosynthesis, fatty acid activation, phospholipid biosynthesis, flavonoid biosynthesis, amino acid transporters, brassinosteroid biosynthesis, glycolysis, triacylglycerol biosynthesis, cellulose biosynthesis, jasmonic acid biosynthesis and TCA cycle pathways. As a result, 279 genes were found to be associated with the above metabolic pathways.

Among the above 279 genes, 20 were found to be related to lipid metabolism pathways, including 17 oil biosynthesis-related genes, one amino acid biosynthesis-related gene, two TCA cycle-related genes and one lipid-related gene, reported in previous studies. Among these lipid metabolism-related genes, six were the same as those for seed oil-related traits, including *GmPDAT*, *GmCds1*, *GmACO1*, *GmAGT*, *GmBS1* and *GmPgs1*.

In oil biosynthesis-related genes, *GmPDAT*, *GmLPEAT2* (*Glyma.03g019200*), *GmPDHC* (*Glyma.20g115500*), *GmLACS2* (*Glyma.11g122500*), *GmACP4* (*Glyma.20g230100*), *GmGPDH* (*Glyma.19g136100*), *GmPLDα1* (*Glyma.08g211700*), *GmPLP2* (*Glyma.05g049500*), *GmCds1* (*Glyma.18g055100*), *GmTIM* (*Glyma.13g146200*), *GmGPAT* (*Glyma.07g069700*), *GmPgs1* (*Glyma.18g302100*), *GmPLA2A* (*Glyma.14g081200*), *GmSAD* (*Glyma.14g121400*), *GmZF351* (*Glyma.06g290100*), *GmBS1* (*Glyma10g38970*) and *Glyma.08g323100* were found to be associated, respectively, with pyruvate (P = 1.44E–05) (Liu, 2020), PI (34:3) (P = 7.12E–10) (Jasieniecka-Gazarkiewicz *et al.*, 2017), phenylalanine (LOD = 4.05) (Zhang *et al.*, 2016a), linolenic acid (P = 2.63E–07) (Lü *et al.*, 2009; Katavic *et al.*, 2014), pyruvate (LOD = 14.68) (Feng *et al.*, 2018), daidzin (LOD = 4.71) (Shen *et al.*, 2006), malate (LOD = 3.11) (Zhao *et al.*, 2012; Zhang *et al.*, 2019b), PI (34:3) (LOD = 4.26) (La Camera *et al.*, 2009), aspartic acid (LOD = 5.65) (Zhou *et al.*, 2013), glycytin (LOD = 3.41) (López-Castillo *et al.*, 2016), serine (LOD = 3.55) (Li *et al.*, 2007), isoleucine (LOD = 6.75) (Tanoue *et al.*, 2014), PE (34:1) (LOD = 3.92) (Yang *et al.*, 2012), stearic acid (LOD = 5.42) (Lindqvist *et al.*, 1996), phenylalanine (LOD = 3.96) (Li *et al.*, 2017), oleic acid (LOD = 3.26) (Ge *et al.*, 2016) and fumaric acid (LOD = 4.56). Note that gene *GmZF351* has no annotation for a biochemical metabolic process, and eight genes (*GmPDAT*, *GmLPEAT2*, *GmSAD*, *GmLACS2*, *GmPLDα1*, *GmPLP2*, *GmTIM* and *GmZF351*) were differentially expressed between wild and cultivated soybean (Figure 2b; Table 2). In genes related to amino acid biosynthesis, *GmAGT* (*Glyma.08g302600*) was found to be associated with palmitic acid (LOD = 3.39) (Zhang *et al.*, 2013). In TCA cycle-related genes, *GmIDH-V* (*Glyma.13g144900*) and *GmACO1* (*Glyma.01g162800*) were found to be associated, respectively, with γ-aminobutyric acid (LOD = 2.78) (Lemaitre and Hodges, 2006) and glycytin (P = 2.63E–07) (Park *et al.*, 2018) (Figure 3b and Table 3).

Genetic relationships between seed oil-related traits and lipid metabolism-related metabolites in soybean

The MCP and SCAD algorithms were used to conduct multiple regression analysis of each seed oil-related trait on 52 acyl lipid-related metabolites, and the Student's *t*-test was further used to determine the acyl lipid-related metabolites that were significantly associated with each oil-related trait. To reduce experimental error, the average of each seed oil-related trait in each accession across 3 years was used to conduct the above analysis. As a result, seed oil content, linoleic acid, linolenic acid, oleic acid and palmitic acid were found to be significantly associated, respectively, with 7, 5, 7, 2, and 10 lipid metabolism-related metabolites (Figure 3a; Table 3). Seed oil content had significant partial regression with genistein (0.526, P = 0.002), PC (36:2) (0.679, P = 1.09E–06), glutamic acid (0.243, P = 0.038), daidzin (−0.842, P = 2.36E–06), PC (36:4) (−0.659, P = 4.75E–06), PC (36:5) (−0.316, P = 0.030) and aspartic acid (−0.172, P = 0.034); linoleic acid had significant partial regression with fumarate (0.486, P = 0.050), PC (36:5) (0.564, P = 4.84E–05), daidzin (−0.911, P = 0.003), PI (36:1) (−1.162, P = 0.009) and stearic acid (−0.324, P = 0.017); linolenic acid had significant partial regression with glycitin (0.664, P = 0.008), PI (34:1) (1.367, P = 4.19E–05), linolenic acid (metabolite) (−0.324, P = 0.017), stearic acid (metabolite) (−0.633, P = 0.014), pyruvate (−0.026, P = 0.050), fumarate (−0.662, P = 0.017) and PI (34:2) (−1.420, P = 0.045); oleic acid had significantly partial regression with daidzin (0.0732, P = 3.11E–4) and isoleucine (−0.022, P = 0.041); palmitic acid had significant partial regression with daidzin (0.086, P = 0.047), fumaric acid (0.220, P = 1.09E–4), PC (36:2) (0.739, P = 8.95E–4), PE (36:5) (0.383, P = 1.24E–4), PI (34:1) (0.294, P = 0.0387), tryptophan (0.142, P = 0.004), aspartate (0.148, P = 0.032), glutamic acid (−0.143, P = 0.042), PC (34:2) (−1.020, P = 0.002) and PI (36:2) (−0.162, P = 0.005) (Table 1). No significant partial regression of stearic acid on acyl lipid metabolites was identified.

Protein–protein interaction (PPI) analysis

The above 36 genes for seed oil-related traits and lipid-related metabolites were used to identify the PPIs using STRING (https://string-db.org/cgi/input.pl). As a result, the predicted values for 16 pairs of PPIs were larger than the medium confidence value of 0.40 (Table S9), indicating the

Table 2 Twenty key candidate genes derived from genome-wide association studies for acyl lipid-related metabolites

	Genome-wide association studies				Comparative genomics					
Trait	Chr.	Position	LOD or *P*	Method[a]	Candidate genes		Arabidopsis homologs	Functional annotation	*P*[b]	References
Pyruvate	8	41 48 8353	4.21	5	*Glyma.08g302600*	*GmAGT*	*AT2G13360.1*	Glycine biosynthesis III	NA	Zhang *et al.* (2002)
	13	20 743 520	1.44E−05	6	*Glyma13g16790*	*GmPDAT*	*AT2G19450.1*	Diacylglycerol acyltransferase 1	0.016*	Liu (2020)
PE (36:3)	1	49 466 364	5.68	4	*Glyma01g36750*	*GmACO1*	*AT4G35830.1*	Aconitase 1	0.031*	Park *et al.* (2018)
Oleic acid	10	46 505 619	3.26	1	*Glyma10g38970*	*GmBS1*	*AT4G14720.1*	Seed size related gene	0.106	Ge *et al.* (2016)
PI (34:3)	3	1 966 012	7.12E−10	6	*Glyma03g02171*	*GmLPEAT2*	*AT2G45670.1*	Predicted phosphate acyltransferase,	0.00*	Jasieniecka-Gazarkiewicz *et al.* (2017)
	5	2 665 256	4.26	1	*Glyma05g03510*	*GmPLP2*	*AT1G12640.1*	Phosphatidylcholine acyl editing	0.050*	La Camera *et al.* (2009)
Phenylalanine	20	34 798 928	4.05	2	*Glyma20g24830*	*GmPDHC*	*AT3G25860.1*	Acetyl-CoA biosynthetic process from pyruvate	0.170	Zhang *et al.*, 2016a; Shen *et al.* (2006)
Stearic acid	14	35 956 260	5.42	4	*Glyma14g27990*	*GmSAD*	*AT1G43800.1*	Plant stearoyl-acyl-carrier protein desaturase family protein	0.032*	Du *et al.* (2016)
Linolenic acid	11	9 480 133	2.63E−07	6	*Glyma11g13050*	*GmLACS2*	*AT1G49430.1*	Long-chain acyl-CoA synthetase 2	0.043*	Lü *et al.* (2009), Katavic *et al.* (2014)
Daidzein	15	7 627 221	4.33	1	*Glyma15g10520*	*GmACP4*	*AT4G25050.1*	Acyl carrier protein 4	0.090	Feng *et al.* (2018)
Daidzin	19	35 006 105	4.71	1	*Glyma19g31730*	*GmGPDH*	*AT3G26720.1*	Glycerol-3-phosphate dehydrogenase	0.231	Shen *et al.* (2006)
Malate	8	17 117 978	3.11	1	*Glyma.08g211700*	*GmPLDα1*	*AT3G15730.1*	Phospholipase D alpha 1	0.011*	Zhao (2013)
Glycytin	13	24 389 546	3.41	1	*Glyma13g20930*	*GmTIM*	*AT2G21170.1*	Triose phosphate isomerase	0.031*	López-Castillo *et al.* (2016)
Aspartic acid	18	4 792 076	5.65	1	*Glyma.18g055100*	*GmCds1*	*AT2G45150.3*	Cytidinediphosphate diacylglycerol synthase	0.170	Zhou *et al.* (2013)
Serine	7	6 389 701	3.55	5	*Glyma07g07580*	*GmGPAT*	*AT4G00400.1*	Triacylglycerol biosynthesis	0.381	Li *et al.* (2007)
Isoleucine	18	62 242 431	3.30	1	*Glyma18g54020*	*GmPgs1*	*AT2G39290.1*	Phosphatidylglycerolphosphate synthase 1	0.022*	Tanoue *et al.* (2014)
Phenylalanine	6	47 437 352	3.96	1	*Glyma06g44440*	*GmZF351*	*AT1G03790.1*	Zinc-finger protein	0.011*	Li *et al.* (2017)
PE (34:1)	14	6 990 732	3.92	5	*Glyma14g08920*	*GmPLA2A*	*AT2G26560.1*	PHOSPHOLIPASE A 2A	0.045*	Yang *et al.* (2012)
γ-aminobutyric acid	13	24 115 317	2.78	4	*Glyma13g20790*	*GmIDH-V*	*AT5G03290.1*	Isocitrate dehydrogenase V	0.097	Lemaitre and Hodges (2006)
Fumaric acid	8	43 127 956	4.56	5	*Glyma.08g323100*	*Glyma.08g323100*	*AT5G55380.1*	Long-chain-alcohol *O*-fatty-acyltransferase	0.316	

[a]Methods ISIS EM-BLASSO, mrMLM, FASTmrEMMA, pLARmEB, pKWmEB and GEMMA are indicated by the numerals 1–6, respectively.
[b]The *P* values were calculated using the paired Student's *t*-test from the average RPKM values at four stages between cultivated (high seed oil, $n_1 = 4$) and wild (low seed oil, $n_2 = 2$) soybeans, with significance indicated by an asterisk (*0.05 level).

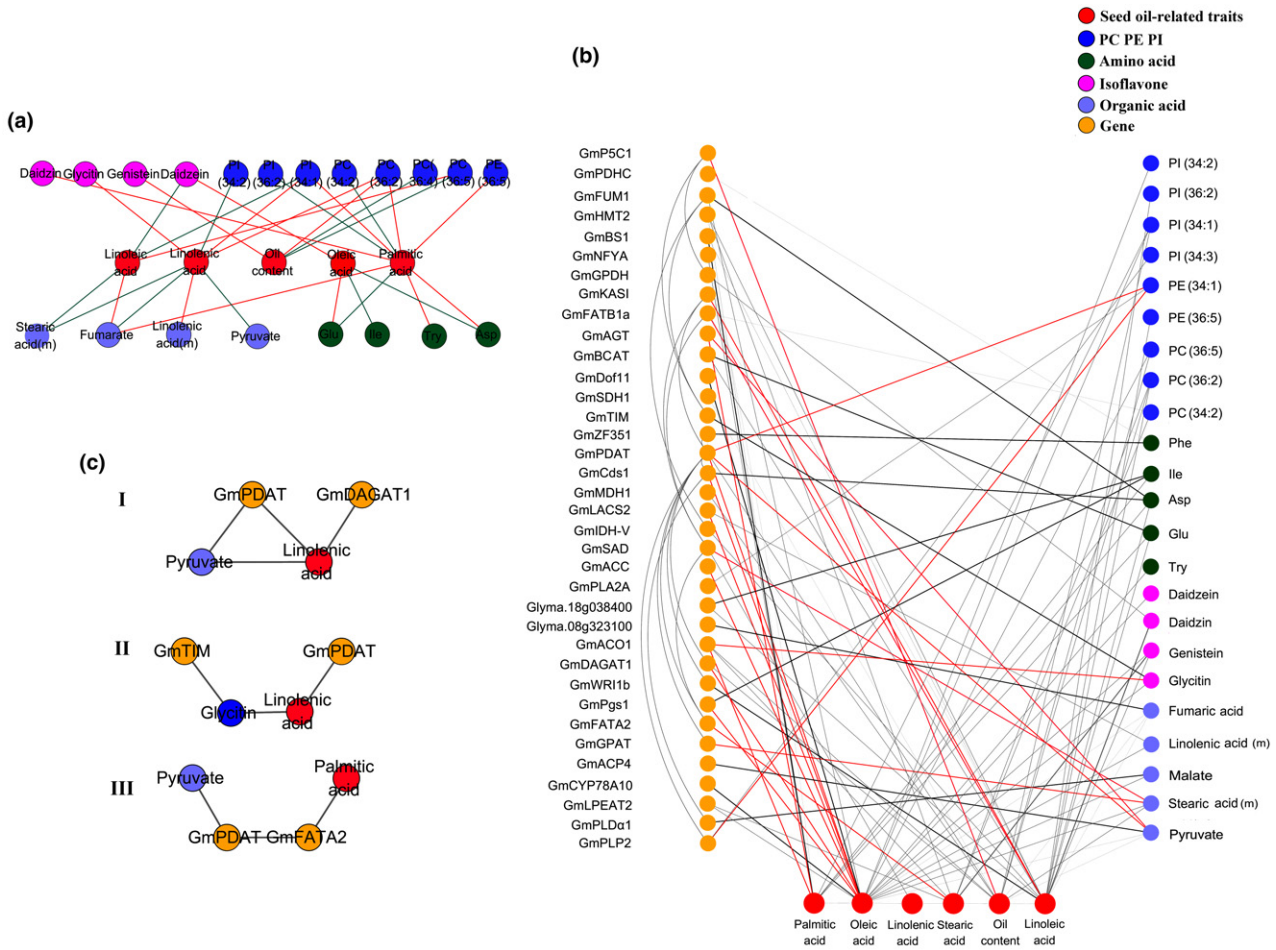

Figure 3. (a) The significant associations of soybean seed oil-related traits with metabolites. (b and c) Three-dimensional genetic networks among seed oil-related traits, metabolites and candidate genes. The red and green lines represent significantly positive and negative correlations between seed oil-related trait and metabolite, respectively. In three-dimensional genetic networks, the nodes for oil-related traits and genes are indicated in red and yellow, respectively, and the other nodes are indicated by blue (PC, PE and PI), green (amino acids), pink (isoflavone) and gray (organic acids); the edges are indicated by the relationship among seed oil-related traits, metabolites and candidate genes; bold red and black lines represent known and newly identified subnetworks, respectively. I: the first group of subnetworks, in which the candidates are commonly and significantly associated with oil-related traits and metabolites. II: the second group of subnetworks, in which oil-related traits are significantly related to metabolites. III: the third group of subnetworks, in which one interacting gene is related to oil-related traits and another interacting gene is related to metabolites.

existence of significant PPIs: e.g. Glyma13g16790.1 (GmPDAT) and Glyma18g36130.3 (GmFATA2) (0.69), GmCds1 (Glyma18g06190.1) and Glyma13g16790.1 (GmPDAT) (0.43), Glyma06g44440.1 (GmZF351) and Glyma13g16790.1 (GmPDAT) (0.43), Glyma08g22600.1 (GmPLDα1) and Glyma18g06190.1 (GmCds1) (0.69), Glyma05g03510.1 (GmPLP2) and Glyma13g16790.1 (GmPDAT) (0.57), Glyma13g16790.1 (GmPDAT) and Glyma08g08910.1 (GmKASI) (0.69), Glyma13g16560.1 (GmDAGAT1) and Glyma13g16790.1 (GmPDAT) (0.75), Glyma13g20790.1 (GmIDH-V) and Glyma02g01920.1 (GmFUM1) (0.92), and Glyma14g27990.1 (GmSAD) and Glyma20g25833.1 (GmFATB1a) (0.90). Clearly, the above two PPIs between GmDAGAT1 and GmPDAT (Liu, 2020) and between GmPDAT and GmFATA2 (Figure 4) were confirmed *in vivo* using luciferase complementation image assay. In addition, the interactions between GmIDH-V and GmFUM1 and between GmDAGAT1 and GmPDAT were reported, respectively, in Zhang *et al.* (2017c) and Liu (2020), and the PPI between GmDAGAT1 and GmPDAT was further validated by the interaction between two loci: Chr13-20532852 and Chr13-20704079 bp (Table S6).

Construction of three-dimensional genetic networks from six soybean seed oil-related traits, 23 lipid-related metabolites, and 36 candidate genes in the pathways of fatty acids, amino acid synthesis and TCA cycle

First, primary metabolic networks in soybean were constructed. Making use of gene homogeneity, 28 genes with functional annotations in the above 36 candidate genes were incorporated into primary metabolic networks in *Arabidopsis thaliana* (Li-Beisson *et al.*, 2013; Wen *et al.*, 2015;

Table 3 Significant associations between seed oil-related traits and metabolites in soybean

Seed oil-related traits	Metabolite	Partial regression coefficient	*t*-test	*F*-test	Seed oil-related traits	Metabolite	Partial regression coefficient	*t*-test	*F*-test
Linolenic acid	Glycitin	0.664	0.008**	4.61E–07***	Palmitic acid	Daidzin	0.086	0.047*	2.59E–15***
	Pyruvate	−0.026	0.050*			Fumaric acid	0.220	1.09E–4***	
	Fumaric acid	−0.662	0.017*			PC (34:2)	−1.020	0.002**	
	PI (34:1)	1.367	4.19E–05***			PC (36:2)	0.739	8.95E–4***	
	PI (34:2)	−1.420	0.045*			PE (36:5)	0.383	1.24E–4***	
	Linolenic acid (m)	0.444	0.045*			PI (34:1)	0.294	0.0387*	
	Stearic acid (m)	−0.633	0.014*			PI (36:2)	−0.162	0.005**	
Oil content	Daidzin	−0.842	2.36E–06***	3.62E–10***		Asparagine	0.148	0.032*	
	Genistein	0.526	0.002**			Glutamic acid	−0.143	0.042*	
	PC (36:2)	0.679	1.09E–06***			Tryptophan	−0.142	0.004**	
	PC(36:4)	−0.659	4.75E–06***		Linoleic acid	Daidzin	−0.911	0.003**	3.11E–05***
	PC (36:5)	−0.316	0.030*			Fumarate	0.486	0.050*	
	Asparagine	−0.172	0.034*			PC (36:5)	0.564	4.84E–05***	
	Glutamic acid	0.243	0.038*			PI (36:1)	−1.162	0.009**	
Oleic acid	Daidzin	0.073	3.11E–4***	1.13E–4***		Stearic acid (m)	−0.324	0.017*	
	Isoleucine	−0.022	0.041*						

Significance levels: *0.05, **0.01 and ***0.001.

Zhang *et al.*, 2016a). In the networks, there were 19 oil biosynthesis-related genes, four amino acid biosynthesis-related genes, five TCA cycle-related genes, six seed oil-related traits and 43 metabolites (Figure 2a). Among the 19 oil biosynthesis-related genes, 12 were differentially expressed between four cultivated and two wild soybeans (Figure 2b).

The above primary metabolic networks in soybean and all the above genetic information in this study were used to construct three-dimensional genetic networks. In these networks, six oil-related traits, 23 lipid-related metabolites and the above 36 candidate genes were used to construct 133 genetic subnetworks, which belong to one of the three types listed below.

The first group included 33 subnetworks, in which each linked gene was identified frequently by phenotypic and metabolic GWAS. In the isoleucine–*GmPgs1*–linolenic acid–*GmPDAT* subnetwork, *GmPgs1* was identified to be associated commonly with isoleucine (metabolite) and linolenic acid (trait). In the pyruvate–*GmPDAT*–linolenic acid–*GmCds1*, PE (34:1)–*GmPDAT*–linolenic acid–*GmDAGAT1* and PE (34:1)–*GmPDAT*–linolenic acid–*GmCds1* subnetworks, *GmPDAT* was identified to be associated commonly with linolenic acid (trait) and two metabolites [PE (34:1) and pyruvate]. In the pyruvate–*GmAGT*–palmitic acid–*GmKASI* subnetwork, *GmAGT* was identified to be associated with pyruvate (metabolite) and palmitic acid (trait). Among all of the 33 subnetworks, five were already known and the others were newly identified (Figure 3d; Table S10). To validate these results, five high-oil and five low-oil accessions were used to conduct hypothesis testing for each node (gene, metabolite or trait) in the above subnetworks. As a result, five, seven, 14 and seven subnetworks were found to have one, two, three and four significant nodes, respectively, although the accessions used in trait and metabolite analyses differed little from those in the gene expression analysis (Table S11).

The second group included 84 subnetworks, which were derived from the significant association of oil-related traits with metabolites (Tables 1 and S10). In the *GmPDAT*–pyruvate–linolenic acid–*GmDAGAT1* subnetwork, pyruvate was significantly associated with linolenic acid ($P < 0.050$). In the *GmLACS2*–linolenic acid (metabolite)–linolenic acid–*GmDof11* subnetwork, linolenic acid (metabolite) was significantly associated with linolenic acid ($P = 0.045$). In the *GmTIM*–glycitin–linolenic acid–*GmPDAT*/*GmDAGAT1* subnetwork, glycitin was significantly associated with linolenic acid ($P = 0.008$) (Table 1). Among all of these subnetworks, 13 were already known

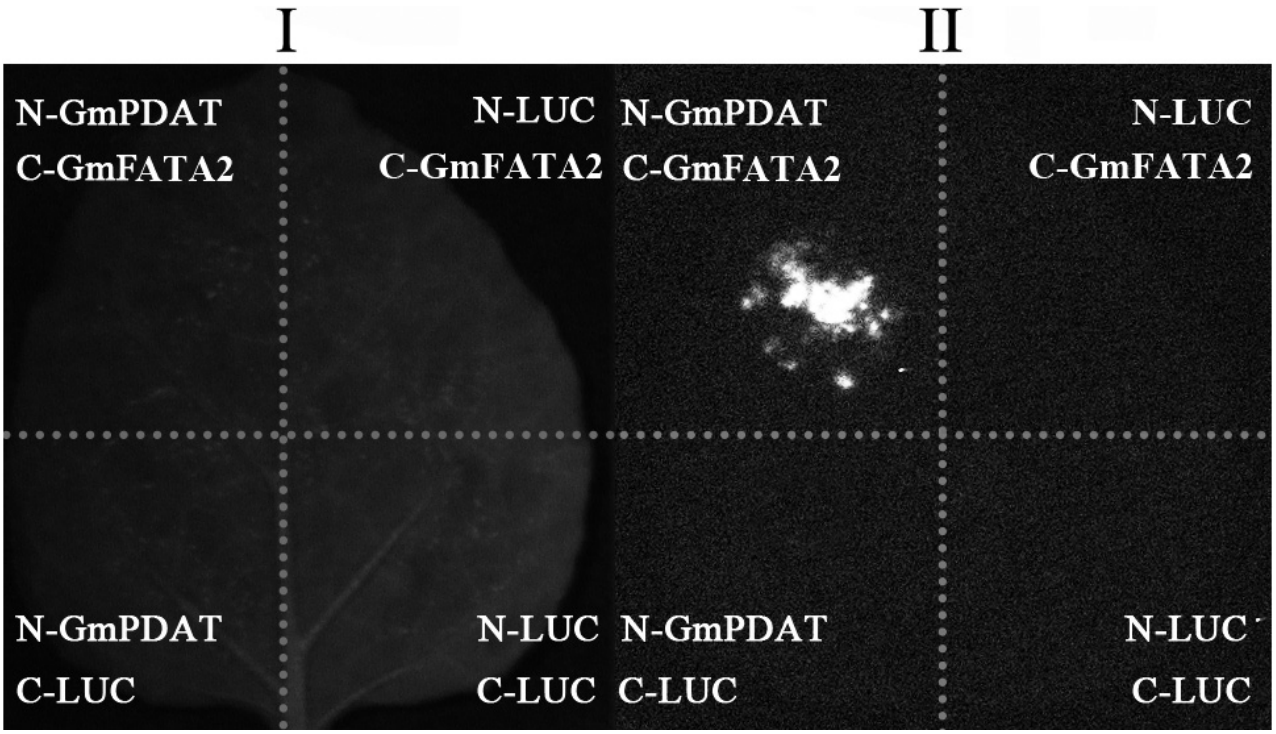

Figure 4. Luciferase complementation image assay of the interaction of GmPDAT with GmFATA2 in *Agrobacterium*-infiltrated *Nicotiana benthamiana* leaves under dark illumination. I and II represent bright and dark fields, under the same treatment. The image shows the interaction between GmPDAT and GmFATA2 in *N. benthamiana* leaves, with the LUC images of *N. benthamiana* leaves co-infiltrated with the *Agrobacterium* strains containing N-GmPDAT and C-GmFATA2 (experimental group, top left corner), N-LUC and C-GmFATA2 (control, top right corner), N-GmPDAT and C-LUC (control, bottom left corner), and N-LUC and C-LUC (control, bottom right corner). LUC fluorescence was detected 48–60 h after infiltration by confocal microscopy. The experiment was repeated three times with similar results.

and the others were newly identified (Figure 3b; Table S10). Similarly, 15, 35, 31 and three subnetworks were found to have one, two, three and four significant nodes, respectively (Table S11).

The third group included 16 subnetworks, which were derived from the interactions between the genes for oil-related traits and/or metabolites (Figure 3d; Table S10). In the pyruvate–*GmPDAT*–*GmFATA2*–oil content and pyruvate–*GmPDAT*–*GmKASI*–palmitic acid subnetworks, the statistic scores for PPIs between *GmPDAT* and *GmFATA2* and between *GmPDAT* and *GmKASI* were 0.69 and 0.69, respectively. Moreover, luciferase complementation image (LCI) assays validated the protein interaction between GmPDAT and GmFATA2 (Figure 4). In the phenylalanine–*GmZF351*–*GmPDAT*–linolenic acid subnetwork, the statistic score for PPI between GmPDAT and GmZF351 was 0.43. In the pyruvate–*GmPDAT*–*GmCds1*–linolenic acid subnetwork, the statistic score for PPI between GmPDAT and GmCds1 was 0.43, whereas *GmPDAT* was significantly associated with linolenic acid and pyruvate. Among all of these subnetworks, six were known and the others were newly identified. In the same way, nine, one and six subnetworks were found to have two, three and four significant nodes, respectively (Table S11).

DISCUSSION

One-dimensional genetic networks among genes (Lin and Lai, 2017) or metabolites (Sauvage *et al.*, 2014), and two-dimensional genetic networks between traits and genes (Wang *et al.*, 2007) and between metabolites and genes (Wen *et al.*, 2015; Chen *et al.*, 2016), have frequently been reported in previous studies. Recently, Shi *et al.* (2020) reported one two-dimensional network between metabolites and traits in Triticum aestivum (wheat). As we know, metabolites act as a bridge between traits and genes (Fiehn, 2002). Thus, it is very important and necessary to construct three-dimensional genetic networks among traits, metabolites and genes. In these networks, 36 candidate genes were obtained from pGWAS and mGWAS, 23 metabolites were significantly associated with five oil-related traits, and all the genetic information was used to construct 133 three-dimensional genetic subnetworks. This study is novel in three aspects. First, to the best of our knowledge this study reports the first three-dimensional genetic networks in soybean. Among these subnetworks, 60 were found to be partly validated in previous molecular biology studies (Table 4), 21 were found to be involved in known KEGG metabolic pathways (https://www.kegg.jp/kegg/pathway.

Table 4 Sixty genetic subnetworks that were partly validated by previous molecular biology studies

Subnetworks constructed in this study				Evidence from previous molecular biology studies	Subnetworks constructed in this study				Evidence from previous molecular biology studies
Group	No.	Subnetwork	Known[a]		Group	No.	Subnetwork	Known[a]	
I	3	Aspartic acid—*GmCds1*–Linolenic acid–*GmDAGAT1*	New	*GmCds1*–Linolenic acid (Zhou *et al.*, 2013); Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)	II	34	*Glyma.08g323100*–Fumaric acid–Linolenic acid–*GmPDAT*	New	Linolenic acid–*GmPDAT* (Liu, 2020)
I	4	Aspartic acid–*GmCds1*–Linolenic acid–*GmDof11*	New	*GmCds1*–Linolenic acid (Zhou *et al.*, 2013); Linolenic acid–*GmDof11* (Wang *et al.*, 2007)	II	35	*Glyma.08g323100*–Fumaric acid–Linolenic acid–*GmDAGAT1*	New	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)
I	7	Aspartic acid–*GmCds1*–Linolenic acid–*GmPgs1*	New	*GmCds1*–Linolenic acid (Zhou *et al.*, 2013); Linolenic acid–*GmPgs1* (Tanoue *et al.*, 2014)	II	39	*Glyma.08g323100*–Fumaric acid–Linolenic acid–*GmDof11*	New	Linolenic acid–*GmDof11* (Wang *et al.*, 2007)
I	11	Isoleucine–*GmPgs1*–Linolenic acid–*GmPDAT*	New	*GmPgs1*–Linolenic acid–Gm*Pgs1* (Tanoue *et al.*, 2014); Linolenic acid–*GmPDAT* (Liu, 2020)	II	41	*GmLACS2*–Linolenic acid (m)–Linolenic acid–*GmPDAT*	Known	Linolenic acid–*GmPDAT* (Liu, 2020)
I	12	Isoleucine–*GmPgs1*–Linolenic acid–*GmDof11*	New	*GmPgs1*–Linolenic acid (Tanoue *et al.*, 2014); Linolenic acid–*GmDof11* (Wang *et al.*, 2007)	II	42	*GmLACS2*–Linolenic acid (m)–Linolenic acid–*GmDAGAT1*	Known	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)
I	18	PE (36:3)–*GmACO1*–Oleic acid–*GmNFYA*	New	Oleic acid–*GmNFYA* (Lu *et al.*, 2016)	II	46	*GmLACS2*–Linolenic acid (m)–Linolenic acid–*GmDof11*	New	Linolenic acid–*GmDof11* (Wang *et al.*, 2007)
I	19	PE (34:1) –*GmPDAT*–Linolenic acid–*GmDAGAT1*	Known	*GmPDAT*–Linolenic acid (Liu, 2020); Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)	II	48	*GmSAD*–Stearic acid (m)–Linolenic acid–*GmPDAT*	Known	Linolenic acid–*GmPDAT* (Liu, 2020)
I	20	PE (34:1) –*GmPDAT*–Linolenic acid–*GmPDAT*	Known	*GmPDAT*–Linolenic acid (Liu, 2020); Linolenic acid–*GmPDAT* (Liu, 2020)	II	49	*GmSAD*–Stearic acid (m)–Linolenic acid–*GmDAGAT1*	Known	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)
I	22	Pyruvate–*GmAGT*–Palmitic acid–*GmBS1*	New	Palmitic acid–*GmBS1* (Ge *et al.*, 2016)	II	53	*GmSAD*–Stearic acid (m)–Linolenic acid–*GmDof11*	New	Linolenic acid–*GmDof11* (Wang *et al.*, 2007)
I	24	Pyruvate–*GmPDAT*–Linolenic acid–*GmCds1*	Known	*GmPDAT*–Linolenic acid (Liu, 2020); Linolenic acid–*GmCds1* (Zhou *et al.*, 2013)	II	56	*GmGPDH*–Daidzin–Oil content–*GmFATA2*	New	Oil content–*GmFATA2* (Moreno-Pérez *et al.*, 2012)
I	26	Pyruvate–*GmPDAT*–Linolenic acid–*GmDAGAT1*	Known	*GmPDAT*–Linolenic acid (Liu, 2020); Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)	II	58	*GmCds1*–Asparagine–Oil content–*GmFATA2*	New	Oil content–*GmFATA2* (Moreno-Pérez *et al.*, 2012)

(continued)

Table 4. (continued)

Subnetworks constructed in this study				Evidence from previous molecular biology studies	Subnetworks constructed in this study				Evidence from previous molecular biology studies
Group	No.	Subnetwork	Known[a]		Group	No.	Subnetwork	Known[a]	
I	27	Pyruvate–*GmPDAT*–Linolenic acid–*GmDof11*	New	*GmPDAT*–Linolenic acid (Liu, 2020); Linolenic acid–*GmDof11* (Wang *et al.*, 2007)	II	61	*GmGPDH*–Daidzin–Palmitic acid–*GmBS1*	New	Palmitic acid–*GmBS1* (Ge *et al.*, 2016)
I	30	Pyruvate–*GmPDAT*–Linolenic acid–*GmPgs1*	New	*GmPDAT*–Linolenic acid (Liu, 2020); Linolenic acid–*GmPgs1* (Tanoue *et al.*, 2014)	II	62	*GmGPDH*–Daidzin–Palmitic acid–*GmWRI1b*	New	Palmitic acid–*GmWRI1b* (Chen *et al.*, 2017)
I	31	Pyruvate–*GmAGT*–Palmitic acid–*GmWRI1b*	New	*GmPDAT*–Linolenic acid (Liu, 2020); Palmitic acid–*GmWRI1b* (Chen *et al.*, 2017)	II	67	*Glyma.08g323100*–Fumaric acid–Palmitic acid–*GmBS1*	New	Palmitic acid–*GmBS1* (Ge *et al.*, 2016)
II	1	*GmGPDH*–Daidzin–Linoleic acid–*GmPgs1*	New	Linoleic acid–*GmPgs1* (Tanoue *et al.*, 2014)	II	68	*Glyma.08g323100*–Fumaric acid–Palmitic acid–*GmWRI1b*	New	Palmitic acid–*GmWRI1b* (Chen *et al.*, 2017)
II	4	*GmGPDH*–Daidzin–Linoleic acid–*GmPDAT*	New	Linoleic acid–*GmPDAT* (Liu, 2020)	II	73	*GmCds1*–Asparagine–Palmitic acid–*GmBS1*	New	Palmitic acid–*GmBS1* (Ge *et al.*, 2016)
II	5	*Glyma.08g323100*–Fumarate–Linoleic acid–*GmPgs1*	New	Linoleic acid–*GmPgs1* (Tanoue *et al.*, 2014)	II	74	*GmCds1*–Asparagine–Palmitic acid–*GmWRI1b*	New	Palmitic acid–*GmWRI1b* (Chen *et al.*, 2017)
II	8	*Glyma.08g323100*–Fumarate–Linoleic acid–*GmPDAT*	New	Linoleic acid–*GmPDAT* (Liu, 2020)	II	79	*GmGPDH*–Daidzin–Oleic acid–*GmNFYA*	New	Oleic acid–*GmNFYA* (Lu *et al.*, 2016)
II	9	*GmSAD*–Stearic acid (m)–Linoleic acid–*GmPgs1*	Known	Linoleic acid–*GmPgs1* (Tanoue *et al.*, 2014)	II	80	*GmACP4*–Pyruvate–Linolenic acid–*GmDAGAT1*	New	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)
II	12	*GmSAD*–Stearic acid (m)–Linoleic acid–*GmPDAT*	Known	Linoleic acid–*GmPDAT* (Liu, 2020)	II	82	*GmACP4*–Pyruvate-Linolenic acid–*GmPDAT*	New	Linolenic acid–*GmPDAT* (Liu, 2020)
II	13	*GmTIM*–Glycitin–Linolenic acid –*GmPDAT*	New	Linolenic acid–*GmPDAT* (Liu, 2020)	II	83	*GmACP4*–Pyruvate–Linoleic acid–*GmPgs1*	New	Linoleic acid–*GmPgs1* (Tanoue *et al.*, 2014)
II	14	*GmTIM*–Glycitin–Linolenic acid –*GmDAGAT1*	New	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)	II	81	*GmACP4*–Pyruvate–Linolenic acid–*GmLACS2*	New	Linolenic acid–*GmLACS2* (Katavic *et al.*, 2014)
II	18	*GmTIM*–Glycitin–Linolenic acid –*GmDof11*	New	Linolenic acid–*GmDof11* (Wang *et al.*, 2007)	III	1	Stearic acid (m)–*GmSAD*–*GmFATA2*–Oil content	Known	*GmFATA2*–Oil content (Moreno-Pérez *et al.*, 2012)
II	20	*GmPDAT*–Pyruvate–Linolenic acid–*GmPDAT*	Known	Linolenic acid–*GmPDAT* (Liu, 2020)	III	2	Stearic acid (m)–*GmSAD*–*GmFATB1a*–Palmitic acid	Known	*GmFATB1a*–Palmitic acid (Chen *et al.*, 2017)
II	21	*GmPDAT*–Pyruvate–Linolenic acid–*GmDAGAT1*	Known	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)	III	8	Pyruvate–*GmPDAT*–*GmWRI1b*–Palmitic acid	New	*GmWRI1b*–Palmitic acid (Chen *et al.*, 2017)
II	22	*GmPDAT*–Pyruvate–Linolenic acid–*GmCds1*	Known	Linolenic acid–*GmCds1* (Zhou *et al.*, 2013)	III	9	Pyruvate–*GmPDAT*–*GmDAGAT1*–Linolenic acid	Known	*GmDAGAT1*–Linolenic acid (Chen *et al.*, 2016)
II	25	*GmPDAT*–Pyruvate–Linolenic acid–*GmDof11*	New	Linolenic acid–*GmDof11* (Wang *et al.*, 2007)	III	10	Phenylalanine–*GmZF351*–*GmPDAT*–Linolenic acid	New	*GmPDAT*–Linolenic acid (Liu, 2020)
II	27	*GmAGT*–Pyruvate–Linolenic acid–*GmPDAT*	New	Linolenic acid–*GmPDAT* (Liu, 2020)	III	12	Pyruvate–*GmPDAT*–*GmFATA2*–Oil content	Known	*GmFATA2*–Oil content (Moreno-Pérez *et al.*, 2012)

(continued)

The Plant Journal, (2020), **103**, 1103–1124

Table 4. (continued)

Subnetworks constructed in this study					Subnetworks constructed in this study				
Group	No.	Subnetwork	Known[a]	Evidence from previous molecular biology studies	Group	No.	Subnetwork	Known[a]	Evidence from previous molecular biology studies
II	28	*GmAGT*–Pyruvate–Linolenic acid–*GmDAGAT1*	New	Linolenic acid–*GmDAGAT1* (Chen *et al.*, 2016)	III	13	Pyruvate–*GmCds1*–*GmPDAT*–Linolenic acid	New	*GmPDAT*–Linolenic acid (Liu, 2020)
II	32	*GmAGT*–Pyruvate–Linolenic acid–*GmDof11*	New	Linolenic acid–*GmDof11* (Wang *et al.*, 2007)	III	15	PI (34:3) –*GmPLP2*–*GmPDAT*–Linolenic acid	Known	*GmPDAT*–Linolenic acid (Liu, 2020)

[a]'Known' subnetworks could be found at the KEGG PATHWAY website (https://www.kegg.jp/kegg/pathway.html) and 'new' subnetworks were constructed in this study.

html) (Table S10) and 112 were newly identified in this study. Then, a series of GWAS approaches were used and all of the significant QTNs across various environments or approaches were used to mine candidate genes in this study. This is because the combination of several GWAS approaches has been recommended in a series of studies to improve the power in QTN detection (Chang *et al.*, 2018; He *et al.*, 2019; Xu *et al.*, 2018; Li *et al.*, 2019; Zhang *et al.*, 2019d), and in practice some true genes for the traits of interest are found to be linked with the QTNs detected by only one GWAS method or in one environment (Zhang *et al.*, 2019e). Finally, quite constructive, reasonable and interesting issues in these subnetworks have been discussed in this study. The results provide the theoretical basis for both the functional identification of seed oil-related genes and quality improvement in soybean breeding.

Using the three-dimensional genetic networks, we may mine some candidate genes to uncover some genetic relationships: for example, pyruvate and the three major nutrients; and amino acids and seed oil content. In this discussion we will focus on these relationships (Figure 5; Table 4).

GmPDAT, GmAGT and *GmACP4* reveal the genetic relationships between pyruvate and three major nutrients

Nutrients mainly include amino acids, fatty acids and carbohydrates. In amino acid metabolism, the absence of pyruvate affected the synthesis of amino acids (Orsi and Leese, 2004; Feng *et al.*, 2018), and *AGT* participated in the metabolism of aspartic acid in *A. thaliana* (Zhang *et al.*, 2013). In this study, *GmAGT* was found to be associated commonly with pyruvate (metabolite) and palmitic acid (trait) in the pyruvate–*GmAGT*–palmitic acid–*GmBS1/GmWRI1b* subnetwork (Table 5), indicating the genetic relationship of *GmAGT* with both pyruvate and palmitic acid.

Pyruvate and adenosine triphosphate (ATP) are the basic molecules in the synthesis of acetyl-CoA, and acetyl-CoA is the main precursor in fatty acid synthesis (Weiss *et al.*, 1974). Meanwhile, *ACP* acts as a carbon carrier for fatty acid synthesis, and *GmPDAT* and *GmDAGAT1* have been reported to be related to oil synthesis (Lardizabal *et al.*, 2008; Chen *et al.*, 2016; Liu, 2020). In this study, pyruvate was found to be significantly associated with linolenic acid ($P = 0.050$) (Table 1) and both *GmPDAT* and *GmACP4* in the *GmACP4*–pyruvate–linolenic acid–*GmDAGAT1* subnetwork (Table 5). We deduce that pyruvate may regulate the synthesis of fatty acids through the action of *GmACP4*, *GmPDAT* and *GmDAGAT1*.

In addition, pyruvate is an important product of glycolysis (Chen *et al.*, 2019). Based on the above information, therefore, *GmPDAT*, *GmAGT* and *GmACP4* may be key genes in the genetic relationships between pyruvate and three major nutrients.

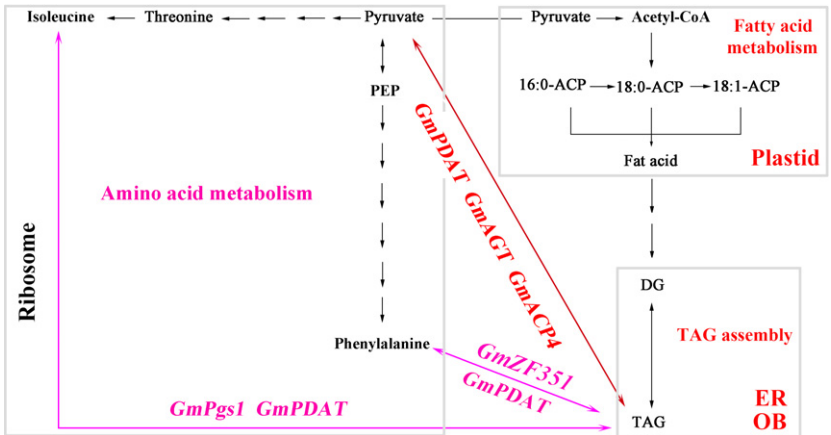

Figure 5. The genetic relationships between pyruvate and three major nutrients, and between amino acids and seed oil content are dissected by *GmPDAT*, *GmAGT* and *GmACP4* (red), and *GmPLDα* and *GmCds1* (pink), respectively, in the three-dimensional genetic networks. The genes are set in italics.

GmPDAT, *GmZF351* and *GmPgs1* reveal the genetic relationship between amino acids and seed oil content

Although seed oil content in soybean is negatively correlated with seed protein content, knowledge about the molecular mechanism of the negative correlation is limited (Chaudhary *et al.*, 2015; Patil *et al.*, 2017). Warrington *et al.* (2015) and Patil *et al.* (2017) revealed a significant correlation of crude protein with amino acids, especially threonine. Note that threonine was the upstream mediator of isoleucine (Guo *et al.*, 2015). If the isoleucine content changed, the threonine content would be influenced, followed by protein and oil contents. In this study, *GmZF351* was found to interact with *GmPDAT* in the detection of PPIs, *GmZF351* and *GmPDAT* were found to be associated with phenylalanine and linolenic acid (Table 4), respectively, and *GmZF351* was reported to increase TAG content in soybean seed (Li *et al.*, 2017). In addition, *GmPgs1* was found to be significantly associated with isoleucine and linolenic acid in this study (Table 5), whereas *Pgs1* participated in the biosynthesis of phosphatidylglycerol (Tanoue *et al.*, 2014). Thus, *GmPDAT*, *GmZF351* and *GmPgs1* may be key genes in amino acid and oil synthesis, which may reveal the genetic relationship between amino acids and seed oil synthesis.

GmCds1, along with average temperature and rainfall, reveals interannual variation of seed oil content in soybean

A paired Student's *t*-test showed that all six oil-related traits in 286 soybean accessions were significantly higher in 2015 and 2016 than in 2014 ($P < 1E{-}04$; Figure 1; Table S12). We discuss the reasons here.

From the genetic perspective, several types of evidence were obtained. In this study, *GmPDAT* was found to be significantly associated with both pyruvate and linolenic acid, *GmCds1* was found to be significantly associated with linolenic acid, and the interaction between locus Chr18-4720420 and the environment was found to be significantly associated with linolenic acid. Around Chr18-4720420, *GmCds1* was mined and annotated with phosphatidylglycerol biosynthesis in the soybean metabolic pathway database. Zhou *et al.* (2013) showed that *CDS* can influence the biosynthesis of phosphatidylglycerol in Arabidopsis. Meanwhile, *GmCds1* had significantly higher expression in cultivated soybeans than in wild soybeans (Figure 2b). More importantly, soybean seeds in the plants with overexpression and interference of *GmPDAT* showed significant changes in linolenic acid and linoleic acid, as compared with the controls (Liu, 2020). As we know, CDP-DAG synthasePAP (CDS) and phosphatidate phosphatase (PAP), along with phosphatidic acid as the substrate, can form CDP-DAG and DAG, respectively (Nakamura, 2017). In extreme environments, *GmCds1* may thus affect the synthesis of DAG, which may reduce the synthesis of TAG, with the assistance of *GmPDAT*, possibly resulting in a decrease in seed oil-related traits.

In addition, we conducted two analyses for environmental factors. First, we conducted correlation analysis between seed oil-related traits and the average temperature from June to September in 2011, 2012 and 2014–2016. As a result, average temperatures in early and all of July were found to have significant correlation with linoleic acid ($r = 0.907$, $P = 0.007$; $r = 0.831$, $P = 0.020$), respectively (Table S13). Then, we calculated the rainfall from June to September and found that the rainfall in 2015 and 2016 was 1.57 and 1.42 times larger than in 2014 (Table S14), whereas seed oil content decreased by 5.4 and 12.5% in 2015 and 2016, respectively, as compared with that in 2014.

Table 5 The significance of differences in traits (t), metabolites (m) and gene expression levels for six subnetworks between high-oil and low-oil soybean accessions

	Node 1			Node 2			Node 3			Node 4			Reference
Subnetwork	High	Low	*P*	High	Low	*P*	High	Low	*P*	High	Low	*P*	
1		Pyruvate (m)			*GmAGT*[b]			Palmitic acid (t)			*GmBS1*		Zhang *et al.* (2002), Ge *et al.* (2016)
	1339.57 ± 891.57[a]	437.61 ± 62.53	0.043*	2.19 ± 0.81	0.83 ± 0.40	0.104	10.69 ± 0.69	11.43 ± 0.54	0.049*	19.54 ± 1.71	10.71 ± 1.72	0.018*	
2		Pyruvate (m)			*GmPDAT*			Linolenic acid (t)			*GmDAGAT1*		Liu (2020), Chen *et al.* (2016)
	1339.57 ± 891.57	437.61 ± 62.53	0.043*	5.68 ± 0.63	1.52 ± 0.54	0.005**	7.51 ± 0.06	12.34 ± 0.58	0.000**	11.54 ± 2.09	1.16 ± 0.47	0.007**	
3		Isoleucine (m)			*GmPgs1*			Linolenic acid (t)			*GmPDAT*		Tanoue *et al.* (2014), Liu (2020)
	83.86 ± 43.86	31.61 ± 18.38	0.027*	7.5 ± 1.51	3.33 ± 0.08	0.035*	7.51 ± 0.06	12.34 ± 0.58	0.000**	5.68 ± 0.63	1.52 ± 0.54	0.005**	
4		Pyruvate (m)			*GmAGT*[b]			Palmitic acid (t)			*GmWRI1b*		Zhang *et al.* (2002), Chen *et al.* (2017)
	1339.57 ± 891.57	437.61 ± 62.53	0.043*	2.19 ± 0.81	0.83 ± 0.4	0.104	10.69 ± 0.69	11.43 ± 0.54	0.049*	16.67 ± 2.76	9.23 ± 1.15	0.036*	
5		Pyruvate (m)			*GmACP4*[c]			Linolenic acid (t)			*GmDAGAT1*		Feng *et al.* (2018), Chen *et al.* (2016)
	1339.57 ± 891.57	437.61 ± 62.53	0.043*	3.17 ± 1.08	0.92 ± 0.92	0.099	7.51 ± 0.06	12.34 ± 0.58	0.000**	11.54 ± 2.09	1.16 ± 0.47	0.007**	
6		Phenylalanine (m)			*GmZF351*			Linolenic acid (t)			*GmPDAT*		Li *et al.* (2017), Liu (2020)
	116.61 ± 43.74	75.16 ± 14.15	0.050*	64.71 ± 16.19	14.64 ± 7.29	0.025*	7.51 ± 0.06	12.34 ± 0.58	0.000**	5.68 ± 0.63	1.52 ± 0.54	0.005**	

[a]Average ± standard deviation. The trait phenotype for each accession was the average across 3 years (2014–2016). The *t* values for the traits (t) and metabolites (m) were calculated between five high-oil and five low-oil accessions, whereas the *t* values for gene expression levels were calculated between four high-oil and two low-oil accessions.
[b]*GmAGT* was found to have significant differences in expression ($P = 0.004$) between four high-oil accessions and one low-oil accession (no. 265) at 15, 25 and 35 DAF, respectively.
[c]*GmACP4* was found to have significant differences in expression ($P = 0.033$) between four high-oil accessions and one low-oil accession (no. 272) at 15, 25 and 35 DAF, respectively; significance levels, *0.05 and *0.01.

Therefore, *GmCds1* and *GmPDAT*, along with the average temperature in July and rainfall, may influence the change of seed oil-related traits across years.

EXPERIMENTAL PROCEDURES

Association populations for phenotypic and metabolic GWAS

As described by Zhou *et al.* (2015), the 286 soybean accessions were randomly selected from six geographic regions in China using a stratified random sampling method, and included 14 wild, 153 landrace and 119 bred accessions. All the accessions were planted in three-row plots in a completely randomized design at the Jiangpu Experimental Station of Nanjing Agricultural University (Nanjing, 31°14′N, 118°22′E) in 2014, 2015 and 2016. The plots were 1.5 m wide and 2.0 m long. Seeds for each accession in 2014–2016 were harvested from the middle row in three-row plots and used to measure seed oil content, palmitic acid, stearic acid, oleic acid, linoleic acid and linolenic acid at the State Key Laboratory of Crop Genetics and Germplasm Enhancement of Nanjing Agricultural University. Among the 286 accessions in 2015, 214 were selected at 55 days after flowering (DAF) and used to measure acyl lipid-related metabolites at Beijing Pufeng Technology Co., Ltd. (Table S15; https://masspeaks.biomart.cn/). The mixture with at least three pods each from different plants for each accession was stored at −80°C before extraction and then extracted for metabolite profiling.

Measurement for six oil-related traits in 286 soybean accessions

Approximately 10 g of seeds was collected from five plants per accession. Based on the method of Baydar and Akkurt (2001), five fatty acids (stearic, palmitic, oleic, linoleic and linolenic acids) (Fang *et al.*, 2017; Zuo *et al.*, 2019; Zhang *et al.*, 2019b) were measured for each accession by gas chromatography with a flame ionization detector and a Permabond free fatty acid phase (FFAP) stainless steel column (50 m × 0.2 mm × 0.33 μm; ThermoFisher Scientific, https://www.thermofisher.com) at Nanjing Agricultural University in 2014, 2015 and 2016. After drying at 70°C for 3 h, approximately 2 g of mature and well-rounded seeds were milled to a fine powder with an electric grinder. Solid fractions were filtered using a 0.20-mm sieve, then 0.03 g of soybean powder was placed in a 2-ml tube with 0.5 ml of 2 mg ml^{-1} heptadecanoic acid (used as an internal standard) and 1 ml *N*-hexane, and then shaken for 30 sec and placed at 20–30°C for 5 h. A 750-μl portion of the hexane layer was transferred to a new 2-ml tube with 0.5 ml of 0.4 M KOH-methanol, then shaken for 2 min and placed at room temperature for 2 h. The hexane layer was transferred to a new 2-ml tube, centrifuged for 5 min at 6000 rpm (4500 ***g***), with 500 μl of supernatant kept for further GC analysis. A 1-μl portion of the prepared sample was injected into the Trace GC system (ThermoFisher Scientific), which was equipped with a DB-23 column (60 m × 0.25 mm × 0.25 μm; Agilent, https://www.agilent.com) at a split ratio of 1:20. The oven was programmed as follows: 150°C for 1 min, ramped to 200°C at 4°C min^{-1}, ramped to 220°C at 3°C min^{-1} and finally ramped to 250°C at 25°C min^{-1}, then held for 5 min with 1.1 ml min^{-1} helium as the carrier gas (Lisec *et al.*, 2006; Marques *et al.*, 2006). Using methyl heptadecanoate (C17) as an internal standard, oil content was calculated by the method described by Zhou *et al.* (2016).

Measurement for 52 acyl lipid-related metabolites using LC–MS

A liquid chromatography–mass spectrometry system was used for the relative quantification of widely targeted metabolites in pods harvested at 55 DAF. The beans were crushed using a mixer mill (MM 200; Retsch, https://www.retsch.com) by MIX-3000 (Hangzhou Miou Instrument, Co., Ltd., http://miulab.com), 100 mg dried powder was weighed and extracted overnight at 4°C with 1.0 ml pure methanol acetonitrile water (1:1). The sample was centrifuges at 14 000 ***g*** and 4°C for 15 min, before 1 μl of the prepared sample was injected into the LC-20AD system (Shimadzu, https://www.shimadzu.com). Separation was performed in a C18 column (150 × 2.1 mm, 3.5 μm) using solvent A (water containing 0.01% heptafluorobutyric acid and 0.1% formic acid) and solvent B (acetonitrile containing 0.01% heptafluorobutyric acid and 0.1% formic acid) as mobile phases, with a column temperature of 50°C. The following MS conditions were used: gas temperature, 325°C; drying gas, 11 l/min; nebulizer, 40 psig; fragmentor, 120 V; and skimmer, 65 V. The instrument was set to acquire over the *m/z* range of 40–1200 with an acquisition rate of 1.2 spectra sec^{-1} (Nygren *et al.*, 2011). Quantification of metabolites was carried out using the standard curve method (Nygren *et al.*, 2011; Thiele *et al.*, 2012; Wen *et al.*, 2015).

The 52 acyl lipid-related metabolites measured in this study included nine organic acids [pyruvic, succinic, fumaric, malic, palmitic (metabolite, m), stearic (m), oleic (m), linoleic and linolenic acids (m)], five soybean isoflavone (daidzein, daidzin, genistein, genistin and glycitin), six PEs [PE (34:1) (16:0/18:1), PE (34:2) (16:1/18:1), PE (36:2) (18:1/18:1), PE (36:3) (18:2/18:1), PE (36:4) (16:0/20:4) and PE (36:5) (16:1/20:4)], six PCs [PC (34:1) (16:0/18:1), PC (34:2) (16:0/18:2), PC (36:2) (18:0/18:2), PC (36:3) (18:1/18:2), PC (36:4) (18:1/18:3) and PC (36:5) (20:4/16:1)], six PIs [PI (34:1) (16:0/18:1), PI (34:2) (16:0/18:2), PI (34:3) (16:1/18:2), PI (36:2) (18:0/18:2), PI (36:3) (18:0/18:3) and PI (36:4) (16:0/20:4)] and 20 amino acids (alanine, arginine, γ-aminobutyric acid, phenylalanine, glycine, glutamic acid, glutamine, methionine, lysine, tyrosine, leucine, proline, tryptophan, serine, threonine, aspartic acid, asparagine, isoleucine, valine and histidine). The number of biological replicates for each accession was two.

GWAS for oil-related traits and acyl lipid-related metabolites

The pre-processing procedures for phenotypic and metabolic GWAS were as follows. Only SNPs with minor allele frequencies (MAF) of ≥0.05 and missing rates of <0.1 in the mapping populations were used in the GWAS, the lines with more than 90% missing for trait phenotypes or metabolites were filtered out and the metabolites with more than 50% missing in 214 lines were excluded (Liaw and Wiener, 2002). The population structure was calculated using the Bayesian clustering program FASTSTRUCTURE (Raj *et al.*, 2014). Six oil-related traits in 286 accessions and 52 acyl lipid-related metabolites in 214 accessions, along with the above SNP information, were used to conduct phenotypic and metabolic GWAS using GEMMA (Zhou and Stephens, 2012), mrMLM (Wang *et al.*, 2016), ISIS EM-BLASSO (Tamba *et al.*, 2017), pLARmEB (Zhang *et al.*, 2017b), FASTmrEMMA (Wen *et al.*, 2018) and pKWmEB (Ren *et al.*, 2018) methods. The K matrix was calculated in the above GEMMA and mrMLM programs. The threshold for significant QTN in phenotypic and metabolic GWAS was set at $P \leq 1/54\ 294 = 1.84E{-}05$ for GEMMA and LOD ≥ 2.5 for the others (Xu *et al.*, 2018; Zhang *et al.*, 2019d). All the mQTNs were obtained from each biological replicate.

The interactions between QTNs and environment (QEs) were detected using the quantitative trait interaction (G × E) module in PLINK 1.9 (http://zzz.bwh.harvard.edu/plink/anal.shtml#qtgxe) (Purcell *et al.*, 2007), and the critical *P*-value for significant QEs was set at 0.001.

The QTN–QTN interactions (QQs) were detected using the online software PEPIS (Zhang *et al.*, 2016c) (http://bioinfo.noble.org//PolyGenic_QTL//Home.gy), and the critical *P* value for significant QQs was set at a LRT of ≥13.815. The protein–protein interactions for candidate genes in phenotypic and metabolic GWAS were detected using the online tools STRING (https://string-db.org//) (Jensen *et al.*, 2009).

Genetic association analysis between oil-related traits and metabolites

minimax concave penalty (MCP; Zhang *et al.*, 2006), SCAD (Fan and Li, 2001) and the Student's *t*-test were used to construct the genetic relationships between six oil-related traits and 52 acyl lipid-related metabolites. To reduce experimental error, the average of each seed oil-related trait in each accession across 2014–2016 was used to conduct the above analysis. Statistical significance was calculated using the *F*-test for the total regression of each oil-related trait on several metabolites and the Student's *t*-test for the regression of each oil-related trait on each metabolite. *, ** and *** indicated significant probability levels of 0.05, 0.01 and 0.001, respectively.

Candidate gene identification

Candidate genes for each oil-related trait and metabolite were mined in two steps. First, all genes between the 100-kb upstream and downstream regions for each of the significant QTNs or mQTNs were mined. Then, we downloaded the soybean metabolic pathway database, KEGG annotation (https://soycyc.soybase.org/) and soybean genome annotation database and gene ontology terms (https://soybase.org/genomeannotation/), and identified the genes or their Arabidopsis homologous genes, which were annotated with fatty acid biosynthesis, fatty acid activation, phosphatidylglycerol biosynthesis, flavonoid biosynthesis, amino acid transporters, brassinosteroid biosynthesis I, glycolysis, triacylglycerol biosynthesis, cellulose biosynthesis, jasmonic acid biosynthesis and TCA cycle.

Differentially expressed gene based on RNA-sequenced data

Four domesticated soybeans (accession nos 101, 236, 257 and 276) with high seed oil content (20.9, 22.3, 17.2, and 17.8%, respectively) and two wild soybeans (accession nos 265 and 272) with low seed oil content (11.9 and 12.5%, respectively) were selected for RNA-seq analysis. Seeds were collected at five seed development stages (15, 25, 35, 45, and 55 DAF) for RNA extraction in 2014. Total RNA was extracted using TRIzol reagent (Invitrogen, now ThermoFisher Scientific, https://www.thermofisher.com) according to the manufacturer's instructions. The RNA was analyzed in an Illumina Hiseq 2500 Sequencer. Sequence reads were aligned using sequence alignment map (SAM) format (Li *et al.*, 2009). The raw reads were cleaned by removing reads with adapters and those of low quality. Clean reads were mapped to reference sequences using SOAPALIGNER/SOAP2 (http://soap.genomics.org.cn/soapdenovo.html). Mismatches of no more than two bases were allowed in the alignment. The gene expression level was calculated by using the reads per kb per million reads (RPKM) method (Mortazavi *et al.*, 2008).

Construction and visualization of three-dimensional genetic networks among oil-related traits, metabolites and candidate genes

In the three-dimensional genetic networks, oil-related traits, metabolites and candidate genes were the nodes of the networks, and the genetic relationships between oil-related traits and candidate genes, between metabolites and candidate genes, between oil-related traits and metabolites, and between candidate genes were the edges of the networks. The genetic relationships between oil-related traits and candidate genes were derived from phenotypic GWAS, the relationships between metabolites and candidate genes were derived from metabolic GWAS, the relationships between oil-related traits and metabolites were derived from the MCP, SCAD and Student's *t*-test analyses, and the relationships between candidate genes were derived from the detection of both QQs and PPIs. Three-dimensional genetic networks with the above nodes, edges and interactions were constructed by the open-source software CYTOSCAPE (Saito *et al.*, 2012).

Hypothesis tests for the differences of traits, metabolites and gene expressional levels in subnetworks between five high-oil and five low-oil soybean accessions

Five high-oil (accession nos 95, 146, 159, 183, and 215, with an average oil content of 18.85 ± 0.81%, SE) and five low-oil (accession nos. 214, 260, 261, 270, and 271, with an average oil content of 13.83 ± 1.69%) soybean accessions were selected to conduct hypothesis tests for the differences of traits and metabolites in the constructed subnetworks, whereas four high-oil (accession nos. 101, 236, 257 and 276) and two low-oil (accession nos. 265 and 272) soybean accessions were selected to conduct hypothesis tests for the expression level differences of genes in the constructed subnetworks. Trait phenotypes for each accession were averages across 3 years (2004–2006), metabolites in pods harvested at 55 DAF were measured by LC-MS in 2015 and the expression levels of genes at 15 DAF were measured by the RPKM values based on RNA-sequenced data. The Student's *t*-test was adopted for hypothesis testing.

Cloning and generation of plant LUC vectors

Soybean (*Glycine max* Willimas 82) and *Nicotiana benthamiana* plants were grown under 16 h light/8 h dark at 25°C for 30–60 days. Soybean total RNA was isolated using TRIzol reagent (Invitrogen), the first-strand cDNA was then synthesized using Moloney murine leukemia virus (M-MLV) reverse transcriptase (Promega, https://www.promega.com). PCR-amplified DNA fragments were cloned into the N-LUC (LUC-luciferase) and C-LUC vector (Chen *et al.*, 2008; Zhang *et al.*, 2018a,b). The full-length coding sequence (CDS) of *GmPDAT* and *GmFATA2* were cloned into the *Bam*HI and *Sal*I sites of JW-771-N, as well as *Kpn*I and *Sal*I sites of JW-772-C, to produce N-gene and C-gene recombination vectors for the LCI assays (Krenek *et al.*, 2015). Primers are listed in Table S16.

Detection of interactions *in vivo*

As described by Zhang *et al.*, (2018a,b), the recombinant plasmids like N-*GmPDAT*+C-*GmFATA2*, N-*GmPDAT*+C-LUC, N-LUC+C-*GmGmFATA2* or N-LUC+C-LUC were transfected into *Agrobacterium tumefaciens* (GV3101). After growing for 48 h under 16 h light/8 h dark, the leaf abaxial epidermis was daubed with 1 mM luciferin (E1602; Promega), and the resulting luciferase signals were captured by the Tanon-5200 image system (Tanon, https://

www.biotanon.com). These experiments were repeated three times to get similar results.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (31871242 and 31571268), Huazhong Agricultural University Scientific & Technological Self-Innovation Foundation (2014RC020) and the State Key Laboratory of Cotton Biology Open Fund (CB2019B01).

AUTHOR CONTRIBUTIONS

YMZ conceived of the project and its components. JYL, PL, YWZ, JFZ, GL, XH and YMZ performed field experiments, bioinformatics analysis and real data analysis. JYL and JFZ performed the experimental LCI assays. YMZ, JYL and JMD wrote and revised the article. All authors reviewed the article for publication.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Supporting information is available from the Wiley Online Library or from the author.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Chromosomal distribution of oil-related trait QTNs for linoleic acid (blue), oleic acid (red), palmitic acid (green), stearic acid (pink), linolenic acid (navy blue) and seed oil content (black) on the soybean genome positions (*x*-axis, cM).

Figure S2. Chromosomal distribution of metabolic QTNs for amino acids (grey), daidzin group (green), organic acid (blue), fatty acid (orange), and PC, PE and PI (pink) on the soybean genome (*x*-axis, cM). Abbreviations: m1, alanine; m2, arginine; m3, γ-aminobutyric acid; m4, phenylalanine; m5, glycine; m6, glutamic acid; m7, glutamine; m8, methionine; m9, lysine; m10, tyrosine; m11, leucine; m12, proline; m13, tryptophan; m14, serine; m15, threonine; m16, aspartic acid; m17, asparagine; m18, isoleucine; m19, valine; m20, histidine; m21, daidzin; m22, daidzein; m23, glycitin; m24, genistein; m25, genistin; m26, pyruvate; m27, succinic acid; m28, malic acid; m29, fumaric acid; m30, linoleic acid; m31, stearic acid; m32, linolenic acid; m33, oleic acid; m34, palmitic acid; m35, PC (34:1); m36, PC (34:2); m37, PC (36:2); m38, PC (36:3); m39, PC (36:4); m40, PC (36:5); m41, PE (34:1); m42, PE (34:2); m43, PE (36:2); m44, PE (36:3); m45, PE (36:4); m46, PE (36:5); m47, PI (34:1); m48, PI (34:2); m49, PI (34:3); m50, PI (36:2); m51, PI (36:3); m52, PI (36:4).

Table S1. Phenotypic characteristics for seed oil-related traits in 286 soybean accessions.

Table S2. Phenotypic characteristics for metabolites (μg g^{-1}) in 214 soybean accessions.

Table S3. Candidate genes in genome-wide association studies for seed oil-related traits.

Table S4. Seventy-seven QTNs of seed oil-related traits detected commonly in two years or by at **least** two methods.

Table S5. Nine QTN–environment interactions for seed oil-related traits in soybean.

Table S6. Ten QTN–QTN interactions for seed oil-related traits in soybean.

Table S7. Candidate genes in genome-wide association studies for 52 metabolites.

Table S8. Forty-eight metabolic QTNs detected by at least two GWAS approaches.

Table S9. Sixteen pairs of significant PPIs between 36 candidate genes derived from phenotypic and metabolic GWAS.

Table S10. One hundred and thirty-three genetic subnetworks among oil-related traits, metabolites and candidate genes.

Table S11. The significances for the differences of traits (t), metabolites (m) and gene expression levels in 133 subnetworks between high-oil and low-oil soybean accessions.

Table S12. Paired Student's *t*-tests and their *P* values for seed oil-related traits between 2014 and others.

Table S13. Correlation analysis between seed oil-related traits and average temperature at the seed developmental stages.

Table S14. Rainfall and annual averages (in mm) for 2014–2016.

Table S15. Two hundred and fourteen accessions used to measure acyl lipid-related metabolites at 55 days after flowering in 2015.

Table S16. Primers used in Luciferase complementation image assays.

REFERENCES

Araújo, W.L., Martins, A.O., Fernie, A.R. and Tohge, T. (2014) 2-Oxoglutarate: linking TCA cycle function with amino acid, glucosinolate, flavonoid, alkaloid, and gibberellin biosynthesis. *Front. Plant. Sci.* **5**, 552.

Baydar, N.G. and Akkurt, M. (2001) Oil content and oil quality properties of some grape seeds. *Turk. J. Agric. For.* **25**, 163–168.

Bonaventure, G., Salas, J.J., Pollard, M.R. and Ohlrogge, J.B. (2003) Disruption of the *FATB* gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell*, **15**, 1020–1033.

Chang, F., Guo, C., Sun, F., Zhang, J., Wang, Z., Kong, J., He, Q., Sharmin, R.A. and Zhao, T. (2018) Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. *Front. Plant Sci.* **9**, 1184.

Charron, C.S., Allen, F.L., Johnson, R.D., Pantalone, V.R. and Sams, C.E. (2005) Correlations of oil and protein with isoflavone concentration in soybean [*Glycine max* (L.) Merr.]. *J. Agric. Food Chem.* **53**, 7128–7135.

Chaudhary, J., Patil, G.B., Sonah, H., Deshmukh, R.K., Vuong, T.D., Valliyodan, B. and Nguyen, H.T. (2015) Expanding omics resources for improvement of soybean seed composition traits. *Front. Plant Sci.* **6**, 1021.

Chen, B., Wang, J., Zhang, G., Liu, J., Manan, S., Hu, H. and Zhao, J. (2016) Two types of soybean diacylglycerol acyltransferases are differentially involved in triacylglycerol biosynthesis and response to environmental stresses and hormones. *Sci. Rep.* **6**, 28541.

Chen, L., Chen, X.W., Huang, X., Song, B.L., Wang, Y. and Wang, Y.G. (2019) Regulation of glucose and lipid metabolism in health and disease. *Sci. China Life Sci.* **62**(11), 1420–1458.

Chen, H., Zou, Y., Shang, Y., Lin, H., Wang, Y., Cai, R., Tang, X. and Zhou, J.M. (2008) Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiol.* **146**, 368–376.

Chen, L., Zheng, Y., Dong, Z., Meng, F., Sun, X., Fan, X., Zhang, Y., Wang, M. and Wang, S. (2017) Soybean (*Glycine max*) WRINKLED1 transcription factor, *GmWRI1a*, positively regulates seed oil accumulation. *Mol. Genet. Genomics*, **293**, 401–415.

Chen, W., Gao, Y., Xie, W. *et al.* (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714–721.

Chen, W., Wang, W., Peng, M. *et al.* (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* **7**, 12767.

Diebold, R., Schuster, J., Daschner, K. and Binder, S. (2002) The branched-chain amino acid transaminase gene family in *Arabidopsis* encodes plastid and mitochondrial proteins. *Plant Physiol.* **129**, 540–550.

Dong, X., Gao, Y., Chen, W., Wang, W., Gong, L., Liu, X. and Luo, J. (2015) Spatio-temporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Mol. Plant*, **8**, 111–121.

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348–1360.

Fan, J., Yan, C. and Xu, C. (2013) Phospholipid: diacylglycerol acyltransferase-mediated triacylglycerol biosynthesis is crucial for protection against fatty acid-induced cell death in growing tissues of *Arabidopsis*. *Plant J.* **76**, 930–942.

Fang, C., Ma, Y., Wu, S. *et al.* (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**, 161.

Feng, Y., Zhang, Y., Wang, Y., Liu, J., Liu, Y., Cao, X. and Xue, S. (2018) Tuning of acyl-ACP thioesterase activity directed for tailored fatty acid synthesis. *Appl. Microbiol. Biotechnol.* **102**, 3173–3182.

Fiehn, O. (2002) Metabolomics — The link between genotypes and phenotypes. *Functional Genomics*, **48**, 155–171.

Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K. and Chen, R. (2016) Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proc. Natl Acad. Sci. USA*, **113**, 12414–12419.

Giberti, S., Funck, D. and Forlani, G. (2014) Δ1-Pyrroline-5-carboxylate reductase from *Arabidopsis thaliana*: stimulation or inhibition by chloride ions and feedback regulation by proline depend on whether NADPH or NADH acts as co-substrate. *New Phytol.* **202**, 911–920.

Goto, D.B. and Naito, S. (2002) *AtMRD1* and *AtMRU1*, two novel genes with altered mRNA levels in the methionine over-accumulating mto1-1 mutant of *Arabidopsis thaliana*. *Plant Cell Physiol.* **43**, 923–931.

Guo, Y., Xu, J., Han, M. and Zhang, W. (2015) Generation of mutant threonine dehydratase and its effects on isoleucine synthesis in *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* **31**, 1369–1377.

He, L., Xiao, J., Rashid, K.Y., Yao, Z., Li, P., Jia, G., Wang, X., Cloutier, S. and You, F.M. (2019) Genome-wide association studies for pasmo resistance in Flax (*Linum usitatissimum* L.). *Front. Plant Sci.* **9**, 1982.

Hwang, E.Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J. and Cregan, P.B. (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* **15**, 1.

Huang, S., Taylor, N.L., Ströher, E., Fenske, R. and Millar, A.H. (2013) Succinate dehydrogenase assembly factor 2 is needed for assembly and activity of mitochondrial complex ii and for normal root elongation in *Arabidopsis*. *Plant J.* **733**, 429–441.

Jako, C., Kumar, A., Wei, Y., Zou, J., Barton, D.L., Giblin, E.M., Covello, P.S. and Taylor, D.C. (2001) Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.* **126**, 861–874.

Jasieniecka-Gazarkiewicz, K., Lager, I., Carlsson, A.S., Gutbrod, K., Peisker, H., Dörmann, P., Stymne, S. and Banaś, A. (2017) Acyl-CoA: Lysophosphatidylethanolamine acyltransferase activity regulates growth of Arabidopsis. *Plant Physiol.* **174**(2), 986–998.

Jensen, L.J., Kuhn, M., Stark, M. *et al.* (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, 412–416.

Katavic, V., Shi, L., Yu, Y., Zhao, L., Haughn, G.W. and Kunst, L. (2014) Investigation of the contribution of oil biosynthetic enzymes to seed oil content in brassica napus and *Arabidopsis thaliana*. *Can. J. Plant Sci.* **94**, 1109–1112.

Klaus, D., Ohlrogge, J.B., Neuhaus, H.E. and Dormann, P. (2004) Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. *Planta*, **219**, 389–396.

Krenek, P., Samajova, O., Luptovciak, I., Doskocilova, A., Komis, G. and Samaj, J. (2015) Transient plant transformation mediated by Agrobacterium tumefaciens: Principles, methods and applications. *Biotechnol Adv.* **33**, 1024–1042.

La Camera, S., Balagué, C., Göbel, C., Geoffroy, P., Legrand, M., Feussner, I., Roby, D. and Heitz, T. (2009) The Arabidopsis patatin-like protein 2 (PLP2) plays an essential role in cell death execution and differentially affects biosynthesis of oxylipins and resistance to pathogens. *Mol. Plant Microbe Interact.* **22**, 469–481.

Lardizabal, K., Effertz, R., Levering, C., Mai, J., Pedroso, M.C., Jury, T., Aasen, E., Gruys, K. and Bennett, K. (2008) Expression of *Umbelopsis ramanniana DGAT2A* in seed increases oil in soybean. *Plant Physiol.* **148**, 89–96.

Leamy, L.J., Zhang, H., Li, C., Chen, C.Y. and Song, B.H. (2017) A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). *BMC Genom.* **18**(1), 18.

Lemaitre, T. and Hodges, M. (2006) Expression analysis of *Arabidopsis thaliana* NAD- dependent isocitrate dehydrogenase genes shows the presence of a functional subunit that is mainly expressed in the pollen and absent from vegetative organs. *Plant Cell Physiol.* **47**, 634–643.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, K., Wang, D., Gong, L., Lyu, Y., Guo, H., Chen, W., Jin, C., Liu, X., Fang, C. and Luo, J. (2019) Comparative analysis of metabolome of rice seeds at three developmental stages using a recombinant inbred line population. *Plant J.* **100**, 908–922.

Li, Q.T., Lu, X., Song, Q.X. *et al.* (2017) Selection for a zinc-finger protein contributes to seed oil increase during soybean domestication. *Plant Physiol.* **173**, 2208–2224.

Li, Y., Beisson, F., Koo, A.J., Molina, I., Pollard, M. and Ohlrogge, J. (2007) Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. *Proc. Natl Acad. Sci. USA*, **104**, 18339–18344.

Liaw, A. and Wiener, M. (2002) Classification and regression by randomForest. *R. News*, **2**, 18–22.

Li-Beisson, Y., Shorrosh, B., Beisson, F. *et al.* (2013) Acyl-Lipid metabolism. *Arabidopsis Book*, **11**, e0161.

Lin, J.S. and Lai, E.M. (2017) Protein-Protein Interactions, Co-Immunoprecipitation. *Methods Mol. Biol.* **1615**, 211–219.

Lindqvist, Y., Huang, W., Schneider, G. and Shanklin, J. (1996) Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *EMBO J.* **15**, 4081–4092.

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. and Fernie, A.R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **10**, 387–396.

Liu, J.Y. (2020) Genetic network of seed oil-related traits and functional analysis of gene *GmPDAT* in soybean (*Glycine max*) [D]. Nanjing Agricultural University.

Liu, Y.F., Li, Q.T., Lu, X. *et al.* (2014) Soybean *GmMYB73* promotes lipid accumulation in transgenic plants. *BMC Plant Biol.* **14**, 73–89.

López-Castillo, L.M., Jiménez-Sandoval, P., Baruch-Torres, N., Trasviña-Arenas, C.H., Díaz-Quezada, C., Lara-González, S., Winkler, R. and Brieba, L.G. (2016) Structural basis for redox regulation of cytoplasmic and chloroplastic triosephosphate isomerases from *Arabidopsis thaliana*. *Front. Plant Sci.* **7**, 1817.

Lü, S., Song, T., Kosma, D.K., Parsons, E.P., Rowland, O. and Jenks, M.A. (2009) *Arabidopsis CER8* encodes LONG-CHAIN ACYL-COA SYNTHETASE 1 (*LACS1*) that has overlapping functions with *LACS2* in plant wax and cutin synthesis. *Plant J.* **59**, 553–564.

Lu, X., Li, Q.T., Xiong, Q. *et al.* (2016) The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* **86**, 530–544.

Manan, S., Ahmad, M.Z., Zhang, G., Chen, B., Haq, B.U., Yang, J. and Zhao, J. (2017) Soybean *LEC2* regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development. *Front. Plant Sci.* **8**, 1604.

Marques, M.A.S., Pereira, H.M.G. and de Aquino Neto, F.R. (2006) Improvements in steroid screening in doping control with special emphasis to GC-MS analytical conditions and method validation. *J. Braz. Chem. Soc.*, **17**, 382–392.

Meng, S., He, J., Zhao, T., Xing, G., Li, Y., Yang, S., Lu, J., Wang, Y. and Gai, J. (2016) Detecting the QTL-allele system of seed isoflavone content in Chinese soybean landrace population for optimal cross design and gene system exploration. *Theor. Appl. Genet.* **129**, 1557–1576.

Moreno-Pérez, A.J., Venegas-Calerón, M., Vaistij, F.E., Salas, J.J., Larson, T.R., Garcés, R., Graham, I.A. and Martínez-Force, E. (2012) Reduced expression of *FatA* thioesterases in *Arabidopsis* affects the oil content and fatty acid composition of the seeds. *Planta*, **235**, 629–639.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.

Murad, A.M., Vianna, G.R., Machado, A.M., da Cunha, N.B., Coelho, C.M., Lacerda, V.A., Coelho, M.C. and Rech, E.L. (2014) Mass spectrometry characterisation of fatty acids from metabolically engineered soybean seeds. *Anal. Bioanal. Chem.* **406**, 2873–2883.

Nakamura, Y. (2017) Plant phospholipid diversity: emerging functions in metabolism and protein-lipid interactions. *Trends Plant Sci.* **22**, 1027–1040.

Nygren, H., Seppänen-Laakso, T., Castillo, S., Hyötyläinen, T. and Orešič, M. (2011) Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics for studies of body fluids and tissues. *Methods Mol. Biol.* **708**, 247–257.

Orsi, N.M. and Leese, H.J. (2004) Ammonium exposure and pyruvate affect the amino acid metabolism of bovine blastocysts *in vitro*. *Reproduction*, **127**, 131–140.

Pan, X., Siloto, R.M., Wickramarathna, A.D., Mietkiewska, E. and Weselake, R.J. (2013) Identification of a pair of phospholipid: diacylglycerol acyltransferases from developing flax (*Linum usitatissimum* L.) seed catalyzing the selective production of trilinolenin. *J. Biol. Chem.* **288**, 24173–24188.

Park, C.H., Roh, J., Youn, J.H., Son, S.H., Park, J.H., Kim, S.Y., Kim, T.W. and Kim, S.K. (2018) *Arabidopsis* ACC oxidase 1 coordinated by multiple signals mediates ethylene biosynthesis and is involved in root development. *Mol. Cells*, **41**, 923–932.

Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., Shannon, G.J., Carter, T.C. and Nguyen, H.T. (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.* **130**, 1975–1997.

Purcell, S., Neale, B., Todd-Brown, K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.

Raj, A., Stephens, M. and Pritchard, J.K. (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573–589.

Ranocha, P., Bourgis, F., Ziemak, M.J., Rhodes, D., Gage, D.A. and Hanson, A.D. (2000) Characterization and functional expression of cDNAs encoding methionine-sensitive and - insensitive homocysteine S-methyltransferases from *Arabidopsis*. *J. Biol. Chem.* **275**, 15962–15968.

Ren, W.L., Wen, Y.J., Dunwell, J.M. and Zhang, Y.M. (2018) pKWmEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity*, **120**, 208–218.

Roesler, K.R., Shorrosh, B.S. and Ohlrogge, J.B. (1994) Structure and expression of an *Arabidopsis* acetyl-coenzyme A carboxylase gene. *Plant Physiol.* **105**, 611–617.

Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D. and Ideker, T. (2012) A travel guide to Cytoscape plugins. *Nat. Methods*, **9**, 1069–1076.

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P.T., Nikoloski, Z., Fernie, A.R. and Causse, M. (2014) Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol.* **165**, 1120–1132.

Selinski, J. and Scheibe, R. (2019) Malate valves: old shuttles with new perspectives. *Plant Biol.* **21**, 21–30.

Shen, W., Wei, Y., Dauk, M., Tan, Y., Taylor, D.C., Selvaraj, G. and Zou, J. (2006) Involvement of a glycerol-3-phosphate dehydrogenase in modulating the NADH/NAD+ ratio provides evidence of a mitochondrial glycerol-3-phosphate shuttle in *Arabidopsis*. *Plant Cell*, **18**, 422–441.

Shi, T.T., Zhu, A.T., Jia, J.Q. *et al.* (2020) Metabolomics analysis and metabolite-agronomic trait associations using kernels of wheat (*Triticum aestivum*) recombinant inbred lines. *Plant J.* https://doi.org/10.1111/tpj.14727.

Shimada, T.L. and Hara-Nishimura, I. (2010) Oil-body-membrane proteins and their physiological functions in plants. *Biol. Pharm. Bull.* **33**, 360–363.

Siloto, R.M., Findlay, K., Lopez-Villalobos, A., Yeung, E.C., Nykiforuk, C.L. and Moloney, M.M. (2006) The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. *Plant Cell*, **18**, 1961–1974.

Song, Q.X., Li, Q.T., Liu, Y.F. *et al.* (2013) Soybean *GmbZIP123* gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants. *J. Exp. Bot.* **64**, 4329–4341.

Tamba, C.L., Ni, Y.L. and Zhang, Y.M. (2017) Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* **13**, e1005357.

Tanoue, R., Kobayashi, M., Katayama, K., Nagata, N. and Wada, H. (2014) Phosphatidylglycerol biosynthesis is required for the development of embryos and normal membrane structures of chloroplasts and mitochondria in *Arabidopsis*. *FEBS Lett.* **588**, 1680–1685.

Thiele, B., Stein, N., Oldiges, M. and Hofmann, D. (2012) Direct analysis of underivatized amino acids in plant extracts by LC-MS-MS. *Methods Mol. Biol.* **828**, 317–328.

Tohge, T., Nishiyama, Y., Hirai, M.Y. *et al.* (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* **42**, 218–235.

Van, K. and McHale, L.K. (2017) Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] seed. *Int. J. Mol. Sci.* **18**(6), e1180.

Wang, H.W., Zhang, B., Hao, Y.J., Huang, J., Tian, A.G., Liao, Y., Zhang, J.S. and Chen, S.Y. (2007) The soybean Dof-type transcription factor genes, *GmDof4* and *GmDof11*, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. *Plant J.* **52**, 716–729.

Wang, S.B., Feng, J.Y., Ren, W.L., Huang, B., Zhou, L., Wen, Y.J., Zhang, J., Dunwell, J.M., Xu, S. and Zhang, Y.M. (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6**, 19444.

Wang, X., Li, Y., Zhang, H., Sun, G., Zhang, W. and Qiu, L. (2015) Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* **42**, 489–496.

Warrington, C., Abdel-Haleem, H., Hyten, D., Cregan, P., Orf, J., Killam, A., Bajjalieh, N., Li, Z. and Boerma, H. (2015) QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. *Theor. Appl. Genet.* **128**, 839–850.

Weiss, L., Löffler, G. and Wieland, O.H. (1974) Regulation by insulin of adipose tissue pyruvate dehydrogenase. A mechanism controlling fatty acid synthesis from carbohydrates. *Hoppe Seylers Z. Physiol. Chem.* **355**, 363–377.

Wen, W., Li, K., Alseekh, S. *et al.* (2015) Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell*, **27**, 1839–1856.

Wen, Y.J., Zhang, H., Ni, Y.L., Huang, B., Zhang, J., Feng, J.Y., Wang, S.B., Dunwell, J.M., Zhang, Y.M. and Wu, R. (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* **19**, 700–712.

Wu, S., Tohge, T., Cuadros-Inostroza, Á. *et al.* (2018) Mapping the *Arabidopsis* metabolic landscape by untargeted metabolomics at different environmental conditions. *Mol. Plant*, **11**, 118–134.

Xiong, W., Wei, Q., Wu, P., Zhang, S., Li, J., Chen, Y., Li, M., Jiang, H. and Wu, G. (2017) Molecular cloning and characterization of two β-ketoacyl-acyl carrier protein synthase I genes from *Jatropha curcas* L. *Plant Physiol.* **214**, 152–160.

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., Xu, S., Yang, Z. and Xu, C. (2018) Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant Sci.* **9**, 1311.

Yang, W.Y., Zheng, Y., Bahn, S.C. *et al.* (2012) The patatin-containing phospholipase A *pPLAIIα* modulates oxylipin formation and water loss in *Arabidopsis thaliana*. *Mol. Plant*, **5**, 452–460.

Zhang, C.H. (2010) Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **38**, 894–942.

Zhang, D., Zhang, H., Hu, Z. *et al.* (2019a) Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication. *PLoS Genet.* **15**, e1008267.

Zhang, D., Zhao, M., Li, S., Sun, L., Wang, W., Cai, C., Dierking, E.C. and Ma, J. (2017a) Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the Palaeopolyploid soybean. *Plant J.* **90**, 1120–1133.

Zhang, G., Bahn, S.C., Wang, G., Zhang, Y., Chen, B., Zhang, Y., Wang, X. and Zhao, J. (2019b) PLDα1-knockdown soybean seeds display higher

unsaturated glycerolipid contents and seed vigor in high temperature and humidity environments. *Biotechnol. Biofuels*, **12**, 9.

Zhang, J., Feng, J.Y., Ni, Y.L., Wen, Y.J., Niu, Y., Tamba, C.L., Yue, C., Song, Q. and Zhang, Y.M. (2017b) pLARmEB, integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity*, **118**, 517–524.

Zhang, L., Liu, J.Y., Gu, H. *et al.* (2018a) *Bradyrhizobium diazoefficiens* USDA 110-*Glycine max* interactome provides candidate proteins associated with symbiosis. *J. Proteome Res.* **17**, 3061–3074.

Zhang, L., Wang, S.B., Li, Q.G., Song, J., Hao, Y.Q., Zhou, L., Zheng, H.Q., Dunwell, J.M. and Zhang, Y.M. (2016a) An integrated bioinformatics analysis reveals divergent evolutionary pattern of oil biosynthesis in high- and low-oil plants. *PLoS ONE*, **11**, e0154882.

Zhang, M., Fan, J., Taylor, D.C. and Ohlrogge, J.B. (2009) *DGAT1* and *PDAT1* acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell*, **21**, 3885–3901.

Zhang, Q., Lee, J., Pandurangan, S., Clarke, M., Pajak, A. and Marsolais, F. (2013) Characterization of *Arabidopsis* serine: glyoxylate aminotransferase, AGT1, as an asparagine aminotransferase. *Phytochemistry*, **85**, 30–35.

Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., Hou, W., Wu, C., Han, T. and Sun, S. (2019c) A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* **20**, E5915.

Zhang, Y., Beard, K., Swart, C. *et al.* (2017c) Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nat. Commun.* **8**, 15212.

Zhang, Y.M., Jia, Z. and Dunwell, J.M. (2019d) Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* **10**, 100.

Zhang, Y.-M., Jia, Z. and Dunwell, J.M. eds. (2019e) *The Applications of New Multi-locus GWAS Methodologies in the Genetic Dissection of Complex Traits*. Lausanne: Frontiers Media.

Zhang, Y.Q., Lu, X., Zhao, F.Y., Li, Q.T., Niu, S.L., Wei, W., Zhang, W.K., Ma, B., Chen, S.Y. and Zhang, J.S. (2016b) Soybean *GmDREBL* increases lipid content in seeds of transgenic *Arabidopsis*. *Sci. Rep.* **6**, 34307.

Zhang, Z., Dunwell, J.M. and Zhang, Y.M. (2018b) An integrated omics analysis reveals molecular mechanisms that are associated with differences in seed oil content between *Glycine max* and *Brassica napus*. *BMC Plant Biol.* **18**, 328.

Zhang, W., Dai, X., Wang, Q., Xu, S. and Zhao, P.X. (2016c) PEPIS: A pipeline for estimating epistatic effects in quantitative trait locus mapping and genome-wide association studies. *PLoS Comput. Biol.* **12**, e1004925.

Zhao, J., Zhou, D., Zhang, Q. and Zhang, W. (2012) Genomic analysis of phospholipase D family and characterization of *GmPLDαs* in soybean (*Glycine max*). *J. Plant Res.* **125**, 569–578.

Zhao, J.Z. (2013) Phospholipase gene *GmPLD* and lipid synthase genes *GmDGAT* and *GmLPAT* play important role in regulating *Arabidopsis* seed oil content and growth [D]. Nanjing Agricultural University.

Zhou, L., Luo, L., Zuo, J.F. *et al.* (2016) Identification and validation of candidate genes associated with domesticated and improved traits in soybean. *Plant Genome*, **9**, https://doi.org/10.3835/plantgenome2015.09.0090.

Zhou, L., Wang, S.B., Jian, J. *et al.* (2015) Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci. Rep.* **5**, 9350.

Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed model analysis for association studies. *Nat. Genet.* **44**, 821–824.

Zhou, Y., Peisker, H., Weth, A., Baumgartner, W., Dörmann, P. and Frentzen, M. (2013) Extraplastidial cytidinediphosphate diacylglycerol synthase activity is required for vegetative development in *Arabidopsis thaliana*. *Plant J.* **75**, 867–879.

Zubimendi, J.P., Martinatto, A., Valacco, M.P., Moreno, S., Andreo, C.S., Drincovich, M.F. and Tronconi, M.A. (2018) The complex allosteric and redox regulation of the fumarate hydratase and malate dehydratase reactions of *Arabidopsis thaliana* Fumarase 1 and 2 gives clues for understanding the massive accumulation of fumarate. *FEBS J.* **285**, 2205–2224.

Zuo, J.F., Niu, Y., Cheng, P., Feng, J.Y., Han, S.F., Zhang, Y.H., Shu, G., Wang, Y. and Zhang, Y.M. (2019) Effect of marker segregation distortion on high density linkage map construction and QTL mapping in soybean (*Glycine max* L.). *Heredity*, **123**, 579–592.