ORIGINAL ARTICLE

Soybean (*Glycine max*) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation

Liang Chen[1] · Yuhong Zheng[2] · Zhimin Dong[1] · Fanfan Meng[1] · Xingmiao Sun[1] · Xuhong Fan[1] · Yunfeng Zhang[1] · Mingliang Wang[1] · Shuming Wang[1]

Abstract

Soybean is the world's most important leguminous crop producing high-quality protein and oil. Elevating oil accumulation in soybean seed is always many researchers' goal. WRINKLED1 (WRI1) encodes a transcription factor of the APETALA2/ethylene responsive element-binding protein (AP2/EREBP) family that plays important roles during plant seed oil accumulation. In this study, we isolated and characterized three distinct orthologues of *WRI1* in soybean (*Glycine max*) that display different organ-specific expression patterns, among which *GmWRI1a* was highly expressed in maturing soybean seed. Electrophoretic mobility shift assays and yeast one-hybrid experiments demonstrated that the GmWRI1a protein was capable of binding to AW-box, a conserved sequence in the proximal upstream regions of many genes involved in various steps of oil biosynthesis. Transgenic soybean seeds overexpressing *GmWRI1a* under the control of the seed-specific napin promoter showed the increased total oil and fatty acid content and the changed fatty acid composition. Furthermore, basing on the activated expressions in transgenic soybean seeds and existence of AW-box element in the promoter regions, direct downstream genes of GmWRI1a were identified, and their products were responsible for fatty acid production, elongation, desaturation and export from plastid. We conclude that GmWRI1a transcription factor can positively regulate oil accumulation in soybean seed by a complex gene expression network related to fatty acid biosynthesis.

Keywords Soybean (*Glycine max*) · Seed oil · WRI1 · AP2/EREBP transcription factor · Fatty acid biosynthesis · Transcriptional machinery

Abbreviations

AP2/EREBP	APETALA2/ethylene responsive element-binding protein
CaMV	Cauliflower mosaic virus
DAF	Days after flowering
EMSA	Electrophoretic mobility shift assay
FA	Fatty acid
ORF	Open reading frame
qRT-PCR	Quantitative real-time PCR
TAG	Triacylglycerols
TF	Transcription factor
WRI	WRINKLED

✉ Zhimin Dong
  dongzhimin2005@126.com

✉ Shuming Wang
  shumingw@263.net

[1] Soybean Research Institute, Jilin Academy of Agricultural Sciences/National Engineering Research Center for Soybean, Changchun 130033, People's Republic of China

[2] Key Laboratory of Soybean Biology in Chinese Ministry of Education, Northeast Agricultural University, Harbin 150030, People's Republic of China

Introduction

Soybean [*Glycine max* (L.) Merrill], one of the most important oilseed crops in the world, accounting for 27% of world vegetable oil production (Chen et al. 2012). World production of soybean oil in the 2016/17 growing season was estimated to be 53.62 million metric tons, an increase of nearly 48% over the past 10 years (USDA data). To meet the growing global demand for soybean oil production in the future, detailed understanding of oil biosynthesis in the

model plant *Arabidopsis thaliana*, combined with the tools of molecular biology and biotechnology, has opened the door for elevating oil accumulation in soybean seed (Clemente and Cahoon 2009; Dussert et al. 2013; Rachael and Tang 2006). The oil accumulation in plant seed is dependent on the processes of fatty acid (FA) biosynthesis, triacylglycerols (TAG) assembly and carbon partitioning between oil and other cellular metabolites during seed-filling stage (Chapman and Ohlrogge 2012). Metabolic engineering to modify (positively and negatively) the expression of specific genes encoding rate-limiting enzymes related to FA biosynthesis (ACCase), TAG assembly (GPAT, LPAAT and DGAT) or carbon partitioning (PDHK and Glu6PDH) is a potential strategy to elevate oil accumulation in plant seed (Baud and Lepiniec 2009; Liu et al. 2012; Sanjaya et al. 2011; Weselake et al. 2009). Most attempts at metabolic engineering have focused on modifying the expression of a single gene and regulating single enzyme-catalyzed reaction (Chen and Smith 2012; Nölke et al. 2006; Shen et al. 2010). However, recent researches suggested that the manipulation of metabolic pathways should be studied at the whole metabolic process rather than at the single enzyme-catalyzed reaction (Courchesne et al. 2009; Zhao et al. 2012). Transcription factors (TFs), as key regulators of metabolic pathways, can simultaneously regulate the expression of multiple genes in the whole cell context (Grotewold 2008; van Erp et al. 2014). Thus, modifying the expression of TF genes in a complex network involving genetic programs, hormonal receptors and metabolic signals represents another strategy for elevating seed oil accumulation (Cagliari et al. 2011; Iwase et al. 2009; Liu et al. 2012). Examples of key TFs regulating oil biosynthesis include the LEAFY COTYLEDON genes (LEC1 and LEC2), FUSCA3 (FUS3), ABSCISIC ACID INSENTIVE3 (ABI3) and WRINKLED1 (WRI1).

As a member of APETALA2/ethylene responsive element-binding protein (AP2/EREBP) transcription factor family, WRI1 is of particular interest with regard to the regulation of seed oil biosynthesis. Arabidopsis *wri1* mutant produced incompletely filled seeds with a dramatic decline (80%) in seed oil content, and lower transcript levels for several steps of FA biosynthesis and TAG assembly. And, overexpression of the *WRI1* homologs in Brassica, oil palm and maize significantly increased seed oil content (Laibach et al. 2015; Liu et al. 2010; Ma et al. 2013; Tajima et al. 2013). Up to now, there were no research reports about WRI1 transcription factor in soybean. In addition, several problems remain to be figured out, for instance, whether the same downstream genes are activated among different plants. To further understand soybean oil biosynthetic pathway and elevate oil accumulation, it is necessary to identify and characterize WRI1 transcription factor in soybean.

The objectives of this study were: (1) to isolate the *GmWRI1* genes in soybean and determine their organ-specific/development-dependent dynamic expression profiles; (2) to characterize in detail the properties of the predicted GmWRI1 protein, including AW box-binding affinity and transactivation activity; (3) to understand the potential function of GmWRI1a related to seed oil content and FA composition in transgenic soybean; (4) to identify direct downstream genes of GmWRI1 and unravel transcriptional machinery of GmWRI1 in soybean.

Materials and methods

Plant materials, RNA extraction and cDNA synthesis

The soybean cultivar Williams82 and transgenic Williams82 plants overexpressing *GmWRI1a* were grown in a greenhouse with a 15-h light period (25 ± 2 °C) and a 9-h dark period (23 ± 2 °C), with relative humidity near 50%. At the specified stages, the leaves, stems, roots, flowers and maturing seeds used for RNA extraction were frozen in liquid nitrogen immediately after harvested, and then stored at -80 °C prior to extraction. The organs were ground to a fine powder using liquid nitrogen and a micro pestle in a 2-ml microcentrifuge tube. Total RNA was extracted using TRIzol reagent (Invitrogen) and then purified using the RNase-Free DNase I (TaKaRa) according to the manufacturers' protocols. The purified RNA (5 μg) per sample was then used as a template to synthesize the first-strand cDNA using the PrimeScript® II 1st Strand cDNA Synthesis Kit (TaKaRa).

Isolation and sequence analysis of three *GmWRI1* homologous genes

The first-strand cDNA from the leaves, stems, roots, flowers and maturing seeds of the soybean cultivar Williams82 at 30 DAF was used as the template for isolating the full-length cDNA sequences of *GmWRI1* genes in soybean. One soybean EST (NCBI accession number AW568355), whose encoding amino acid sequence shares homology to the AP2/EREBP domain of Arabidopsis *WRI1* (At3g54320), was used as a starting sequence to amplify the partial sequences of putative *GmWRI1* genes. From the resulting partial sequences, primers specific to the genes were designed. Following the manufacturer's protocol for the rapid amplification of cDNA end (RACE), a nested PCR strategy was used to amplify 3′ and 5′ end sequences of the genes using 3′-Full RACE Core Set Ver.2.0 (TaKaRa) and 5′-Full RACE Kit (TaKaRa). Then the full-length cDNA sequences of the resulting *GmWRI1* genes were obtained by analyzing the overlaid sequences. From the maturing seed, stem and flower of soybean, the cDNA sequences of three *GmWRI1* homologous genes were isolated,

which were named *GmWRI1a* (GLYMA_15G221600), *GmWRI1b* (GLYMA_08G227700) and *GmWRI1c* (GLYMA_15G34770), respectively.

The analyses of the nucleotide and amino acid sequences were performed using DNAstar and DNA-MAN software. Using the NCBI website (http://www.ncbi.nlm.gov/blast), a BLAST search was run using *GmWRI1* sequences to identify other WRI1-type genes sharing double AP2/EREBP domains. Sequence of WRI1 and WRI1-like members found in other species was collected and used to construct a phylogenetic tree. The sequence alignment was conducted using ClustalW software. A reconstructed phylogenetic tree was built by the neighbor-joining program of MegAlign software.

Expression profiles of three *GmWRI1* homologous genes in different organs and developmental stages

Total RNA was isolated from the leaves, stems, roots, flowers and maturing seeds at 8, 16, 24, 32, 40, 48, 56 and 64 DAF, and purified following the protocol outlined above. The concentration of RNA was quantified by UV absorbance at 260 nm. Approximately 10 µg of purified RNA samples were resolved by electrophoresis in a denaturing $1 \times$ MOPS 1.2% (w/v) agarose gel containing 2% formaldehyde. The gels were run overnight at 3–4 V/cm in RNase free gel boxes. After the RNA samples were well separated, the gels were rinsed for 2×15 min in $20 \times$ SSC. Subsequently, RNA samples were blotted overnight onto positively charged nylon membranes (Roche) by capillary transfer with $20 \times$ SSC. The membrane was rinsed briefly twice in $2 \times$ SSC. RNA samples were fixed to the membrane by baking at 120 °C for 30 min. The specific 5′-terminal fragment of *GmWRI1a* (210 bp), *GmWRI1b* (135 bp) and *GmWRI1c* (150 bp) and a fragment of the soybean *18S rDNA* (400 bp), labeled by digoxygenin-11-UTP in *in vitro* transcription reaction, were used by RNA probes using the SP6/T7 RNA DIG-labeling Kit (Roche) according to the manufacturer's instructions. The blotted membranes were hybridized overnight with DIG labeled RNA probes (100 ng/ml) at 68 °C in DIG Easy Hyb buffer (Roche). After hybridization, the membranes were subsequently washed under high-stringency conditions as follows: twice for 5 min in $2 \times$ SSC, 0.1%SDS at ambient temperature under constant agitation, and twice for 15 min in $0.1 \times$ SSC, 0.1%SDS at 68 °C under constant agitation. After hybridization and stringency washes, the membranes were followed by immunological detection of DIG labeling using CDP-STAR in the DIG Northern Starter Detection Kit (Roche) according to the manufacturer's protocol. Blots were exposed to X-ray film for 15–25 min at ambient temperature.

Transactivation activity of GmWRI1a via yeast one-hybrid assay

The yeast one-hybrid assay was carried out as described before by Chen et al. (2008) and Zhang et al. (2007). The reporter plasmids were constructed by inserting the sequences of the AW-box (5′-CCGCCTTCGTAAGTTCCGCCGA-3′) and the mutant AW-box (5′-CCGCtTcCaTAAGTTCaaCCGA-3′) into upstream of the minimal promoter of the yeast iso-1-cytochrome C gene and the reporter gene *lacZ* existed in the pLacZi vector. These reporter plasmids were then digested with *Nco* I and integrated into the genome of the yeast strain YM4271 at the *URA3* locus. The effector plasmid was constructed by subcloning the full-length ORF of *GmWRI1a* between the promoter of the alcohol dehydrogenase gene and the terminator of the alcohol dehydrogenase gene existed in the pGAD424 vector, which carries the *LEU2* gene for selection in Leu-auxotrophic yeast strains. The effector plasmid was transformed into the competence cells of yeast strain YM4271 carrying the reporter plasmids using the AcLi/SSDNA/PEG method (Gietz and Woods 2002). Transformants were selected on a medium lacking uracil and leucine (SD/-Ura/-Leu), and subsequently the activity of β-galactosidase (encoded by the *lacZ* reporter gene) was performed using a filter-lift assay to investigate transactivation and interaction of GmWRI1a with AW-box.

AW box-binding affinity of GmWRI1a via electrophoretic mobility shift assay

The ORF sequence of *GmWRI1a* was subcloned into the pET-32-Ek/LIC vector. After sequencing confirmation, the $(His)_6$-GmWRI1a clone was transformed into *E. coli* BL21(DE3). The bacteria were grown in LB medium, supplemented with the ampicillin (100 mg/ml), at 37 °C until $A_{600} = 06$–0.7. Expression of the recombinant protein was induced with 1 mM IPTG. After incubating for 3 h, cells were harvested by centrifugation ($2500 \times g$, 30 min, 4 °C). Cell pellets were washed with PBS buffer, centrifuged and resuspended in 20 ml buffer containing 5 mM imidazole, 0.5 M NaCl and 20 mM NaH_2PO_4 (pH 7.4) for the pellet derived from 500 ml cell culture. The cells were lysed by one cycle of freezing at − 80 °C and thawing at 37 °C. DNA was sheared by a brief sonication and inclusion bodies were recovered by centrifugation ($2000 \times g$, 20 min, 4 °C). The pellet was washed twice with the buffer containing 5 mM imidazole, 0.5 M NaCl, 20 mM NaH_2PO_4 and 0.5% Triton™ X-100 (pH 7.4), resuspended in the buffer containing 8 M urea, 0.5 M NaCl and 20 mM NaH_2PO_4 (pH 8.0) and sonicated. After cell lysis and fractionation, soluble and recombinant GmWRI1a protein was purified using HiTrap Chelating HP columns (GE healthcare) according to the manufacturers' protocols.

The electrophoretic mobility shift assay (EMSA) was conducted according to the modified protocols of Maeo et al. (2009) and Baud et al. (2009). Probe sequences covering 500-bp sequences upstream from the ATG codon of GmWRI1a target genes are given in Supplementary Table S1. The DNA probes were labeled with α-^{32}P-dATP by Klenow DNA polymerase I and purified on the column of DNA Fragment Purification Kit Ver.2.0 (TaKaRa). The DNA-binding assays were prepared as follows: 2 µg of GmWRI1a recombinant protein were incubated with 3 ng of [^{32}P] labeled probes in banding buffer [10 mM Tris–HCl, pH 7.5, 50 mM NaCl, 2 mM Mg$_2$Cl, 0.5 mM EDTA, 5% glycerol, 1 mM DTT, 0.5 µg poly (dI-dC)] in a total volume of 15 µl. The mixture was incubated at ambient temperature for 30 min, and fractionated at 4 °C by 8% polyacrylamide gel electrophoresis in $0.5 \times$ Tris-borate-EDTA buffer followed by autoradiography.

Generation of transgenic soybean lines overexpressing *GmWRI1a*

The ORF of the *GmWRI1a* gene was subcloned and inserted between the *B. napus* napin promoter, identified to be seed specific in transgenic plant (Jako et al. 2001; Stalberg et al. 1996; Vigeolas et al. 2007), and the Agrobacterium nos terminator of a binary vector pSE, which contains the *bar* selectable marker gene inserted between the CaMV 35S promoter and the soybean vegetative storage protein terminator. The resulting plasmid was introduced into the *Agrobacterium tumefaciens* strain EHA101 by the freeze–thaw method and used for *A. tumefaciens*-mediated transformation of soybean cotyledonary node following the procedure described by Paz et al. (2004).

After herbicide resistance screening, the presence of the *GmWRI1a* gene in the positive T$_4$ transgenic soybean lines was monitored by southern blot analysis. Briefly, genomic DNA was extracted from maturing soybean seeds at 40 DAF using Universal Genomic DNA Extraction Kit Ver.3.0 (TaKaRa). 20 µg of genomic DNAs from three transgenic lines were digested with the restriction enzyme *Dra* I or *Eco*R V, and the digested DNA was separated on a 0.8% (w/v) agarose gel and blotted onto positively charged nylon membranes (Roche). The specific sequence of napin promoter and *GmWRI1a* as the probe was DIG-labeled, hybridized with the blotted membranes and detected using DIG High Prime DNA Labeling and Detection Starter Kit II (Roche) according to the manufacturer's instructions. In addition, the expression of the introduced *GmWRI1a* gene was confirmed by northern blot analysis as mentioned above.

Measurement of oil and fatty acids

The content of total oil in soybean seed was determined with approximately 10 g of seed using a NMR (Nuclear Magnetic Resonance) spectroscopy (Bruker) according to the manufacturer's instructions. The content of total oil was expressed on a% dry weight basis (Hwang et al. 2014). Samples were analyzed in triplicate.

The FA composition in soybean seed was measured by a gas chromatography (GC) assay using the model GC-14C gas chromatograph (Shimadzu). Briefly, 0.5 g of soybean seed powder was weighed accurately. Following addition of 5 ml of 0.4 M KOH–methanol solution for transforming fatty acids into methyl ester forms, the samples were mixed uniformly. After 30 min standing, 0.5 ml of 10% acetic acid and 3 ml of heptane were added, shaken and then stood for 2 min. The supernatant was used for GC analysis. The gas chromatograph was equipped with a FFAP elastic quartz capillary vessel column. The operating parameters were: column temperature (210 °C), injection and detector temperatures (250 °C), air and hydrogen flow rates (400 ml/min), nitrogen pressure (11,620 kPa), split ratio (1:50), sample volume (1 µl), and total analysis time (7 min). After the run, the content of each FA species was detected by the normalization method of peak area using N3000 GC data workstation.

Affymetrix GeneChip array and qRT-PCR analysis

Total RNA samples of napin:*GmWRI1a* transgenic plants and wild-type plants were used to the Affymetrix soybean GeneChip array, which contains 37,593 probe sets representing 35,611 soybean transcripts. The GeneChip were hybridized and scanned following standard Affymetrix procedures. Each assay was replicated three times with independent biological samples for statistical analysis for a total of six chips. The data were analyzed using the Expressionist software Pro version 3.1 (Genedata) as described previously (Bolton et al. 2008). Briefly, the raw probe-level hybridization data were loaded into the Expressionist Refiner module for assuring high data quality, and then the probe set values were condensed using the MAS 5.0 Algorithm (Affymetrix). The MAS 5.0 signal data were natural-log transformed and normalized. To quantify an expression value for each transcript, normalized raw probe set expression values were extracted by the Expressionist Analyst module. To evaluate differential expression of the same gene between transgenic soybean and wild-type, expression values were subsequently subjected to a Student's *t* test, which resulted in a *p* value for each gene. These *p* values were converted to *q* values using the method of Storey and Tibshirani (2003) to approximate the false discovery rate (FDR). The genes with statistically significantly differences were filtered out by a fold change threshold ≥ 2.0

using a stringent cut off at FDR ≤ 5%. The differentially expressed genes were annotated using the Affymetrix Gene-Chip Soybean Genome Array annotation page developed as part of SoyBase and the Soybean Breeder's Toolbox (http://soybase.org/AffyChip/).

The GeneChip array data were confirmed by qRT-PCR analysis, which was carried out using SYBR® Premix Ex Taq™ II (TaKaRa) and analyzed by the Applied Biosystems 7900 Real-Time PCR system. The primers of each selected gene were designed using primer3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cg). The soybean *18S rRNA* gene was used as an internal reference for relative quantification analyses. The relative expression of each selected gene was evaluated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001) by comparing the data with the internal reference gene. To assess the reproducibility of the data analysis, the experiments were repeated three times.

Results

Identification and sequence analysis of three homologous *GmWRI1* genes

Three full-length cDNA sequences of the homologous *WRI1* genes of soybean, designated *GmWRI1a*, *GmWRI1b* and *GmWRI1c*, were obtained, respectively, from the maturing seed, stem and flower using RT-PCR and RACE methods. The open reading frame (ORF) of *GmWRI1a*, *GmWRI1b* and *GmWRI1c* was consisted of 1230, 1116 and 993 nucleotides, respectively. The deduced amino acid sequences of these gene products showed significant sequence similarity in the conserved domain found in AP2/EREBP transcription factor family (Fig. 1a). As shown in Fig. 1b, they all include double conserved AP2/EREBP domains that specifically bind to the *cis*-acting element within the promoters of downstream
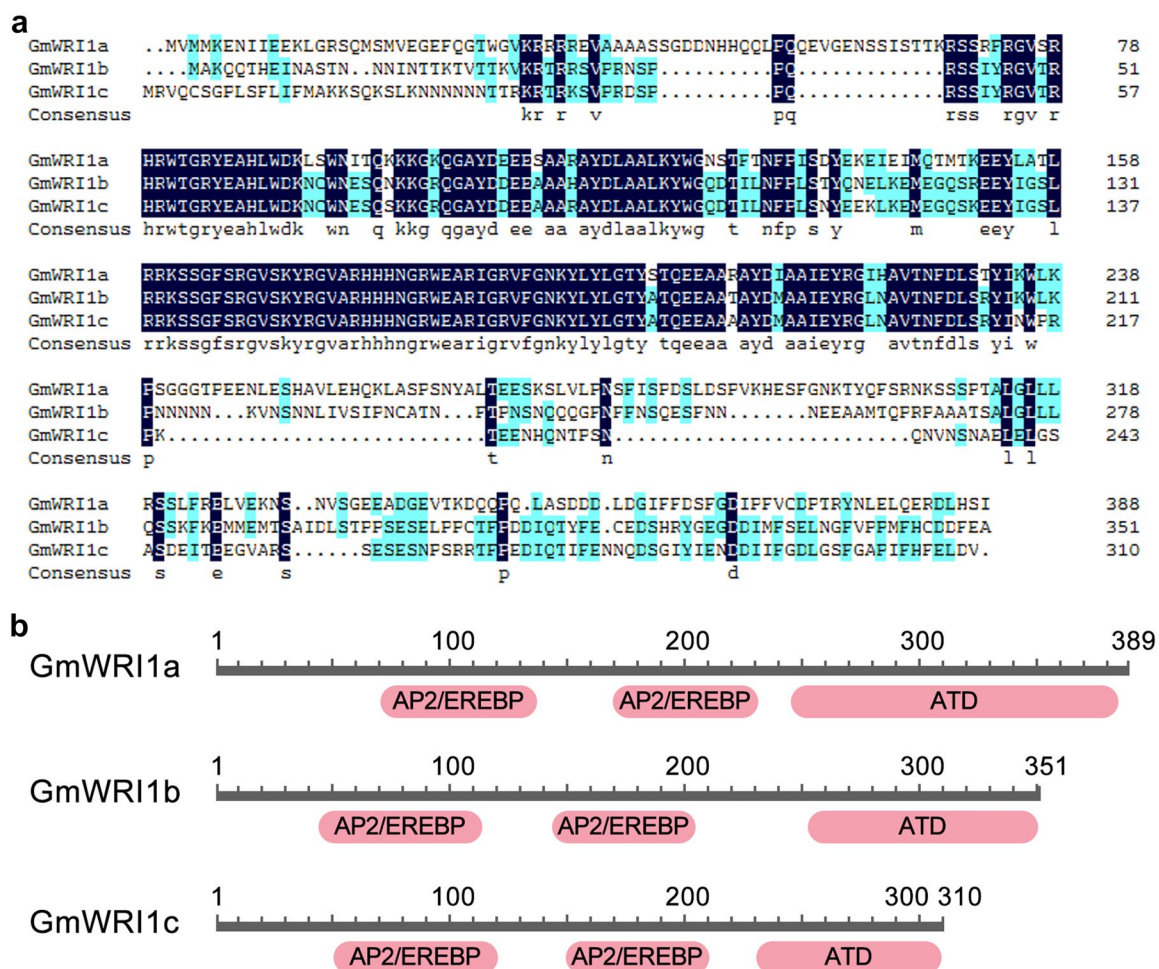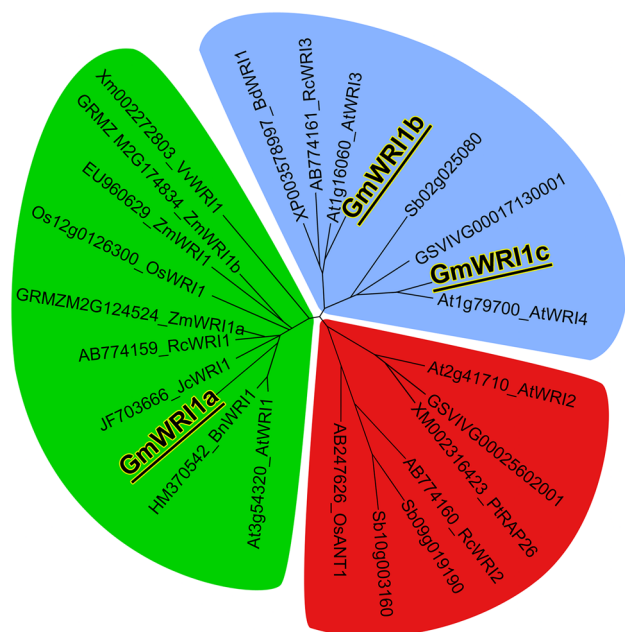

Fig. 1 Sequence analysis of GmWRI1a, GmWRI1b and GmWRI1c. **a** Protein alignment of GmWRI1a, GmWRI1b and GmWRI1c. Identical residues were shaded black, and similar residues were shaded gray. **b** The conserved domains of GmWRI1a, GmWRI1b and GmWRI1c. Numbers indicate the amino acid positions along the protein. Double AP2/EREBP DNA binding domains and single acidic transactivation domain (ATD) were predicted in all three proteins

functional genes, and a region rich in acidic amino acids act as a putative transactivation domain.
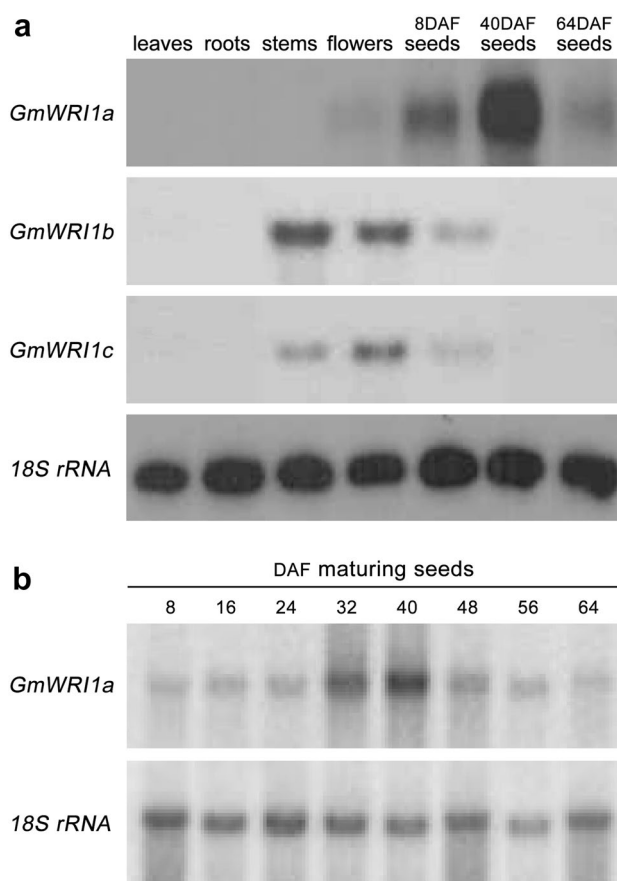
To clarify the phylogenetic relationship between soybean *WRI1* genes and other *WRI1* or *WRI1*-like genes in different plant species, a phylogenetic tree was constructed using the whole amino acid sequences (Fig. 2). The results demonstrated that GmWRI1a fell into the same clade as AtWRI1 and ZmWRI1, and GmWRI1b and GmWRI1c were present in the clades defined by AtWRI3 and AtWRI4. Sequence comparison of the complete amino acids showed that GmWRI1a, GmWRI1b and GmWRI1c had 49.47, 57.14 and 62.79% similarity with the Arabidopsis AtWRI1, AtWRI3 and AtWRI4, respectively. It is worth noting that *GmWRI1a* was isolated specifically from the maturing seed, similar to *AtWRI1* and *ZmWRI1*, and *GmWRI1b* and *GmWRI1c* were isolated from the stems and flowers, similar to *AtWRI3*/*AtWRI4*. These results demonstrated that three soybean *WRI1* genes encode novel members of AP2/EREBP transcription factor family.

Organ-specific and development-dependent dynamic expression profiles of three homologous *GmWRI1* genes

To gain further insights into the organ-specific expression profiles of *GmWRI1a*, *GmWRI1b* and *GmWRI1c*, the organ distribution of three homologous genes was investigated by northern blot analysis. The results in Fig. 3a showed that the transcript of *GmWRI1a* was detected specifically in the maturing seeds. The transcript abundances of *GmWRI1b* and *GmWRI1c* were enriched in stems and flowers. Since the highest transcript abundance of *GmWRI1a* was found in soybean seed, the development-dependent dynamic expression profile of *GmWRI1a* was further investigated

Fig. 2 Phylogenetic tree of GmWRI1a, GmWRI1b and GmWRI1c with other WRI1 or WRI1-like proteins. Accession numbers for the WRI1 or WRI1-like proteins used here were listed: Arabidopsis (At3g54320_AtWRI1, At2g41710_AtWRI2, At1g16060_AtWRI3 and At1g79700_AtWRI4); rice (Os12g0126300_OsWRI1 and AB247626_OsANT1); maize (GRMZM2G124524_ZmWRI1a, GRMZ M2G174834_ZmWRI1b and EU960209_ZmWRI1); grape (XM002272803_VvWRI1); black cottonwood (XM002316423_ PtRAP 26); castor bean (AB774159_RcWRI1, AB774160_ RcWRI2 and AB774161_RcWRI3); Jatropha curcas (JF703666_JcWRI1); rape (HM370542_BnWRI1); sorghum (Sb10g003160, Sb09g019190 and Sb02g025080); grapevine (GSVIVG0025602001 and GSVIVG0001713001); stiff brome (XP003578997_BdWRI1). Full-length deduced amino acid sequences were used for analysis

Fig. 3 Expression profiles of *GmWRI1a, GmWRI1b* and *GmWRI1c* in various organs of soybean plants. **a** Northern blot analysis of *GmWRI1a, GmWRI1b* and *GmWRI1c* in the leaves, stems, roots, flowers and seeds at 8, 40 and 64 DAF. 10 µg of total RNA was loaded for each lane. The blots were hybridized with DIG-labeled probe from the specific 5′-terminal fragment of *GmWRI1a, GmWRI1b* and *GmWRI1c*, respectively. *18S rRNA* was used as a control. **b** Northern blot analysis of *GmWRI1a* in the maturing seeds at 8, 16, 24, 32, 40, 48, 56 and 64 DAF. 10 µg of total RNA was loaded for each lane. The blot was hybridized with DIG-labeled probe from the specific 5′-terminal fragment of *GmWRI1a*. *18S rRNA* was used as a control

during seed-filling stages ranging from 8 to 64 days after flowering (DAF). As shown in Fig. 3b, GmWRI1a showed markedly activated expression within early seed-filling stages, reached a maximum at 40 DAF, and then gradually decreased throughout the seed maturation. Taken the above results together, it can be concluded that three homologous *GmWRI1* genes may play distinct roles during various growth and development stages, and GmWRI1a transcription factor probably function during soybean seed formation period. In the following researches, GmWRI1a was selected as the research focus.

## Transactivation and interaction of GmWRI1a with AW-box

The transactivation activity of GmWRI1a was investigated using the yeast one-hybrid system. The *cis*-DNA elements (wild-type AW-box and mutant AW-box) were synthesized, and each was inserted into the reporter plasmid pLacZi, upstream of the reporter gene *lacZ*. Each pLacZi-AW/pLacZi-mAW was digested and integrated into the genome of the yeast strain YM4271 at the *URA3* locus (Fig. 4a). The pGAD424-GmWRI1a effector plasmid (Fig. 4a) was constructed by subcloning the whole coding sequence to

the pGAD424 vector, and transfected into the cells of yeast strain YM4271 carrying the pLacZi-AW/pLacZi-mAW reporter plasmids. Growth of the transfected yeast cells on the SD/-Ura/-Leu medium indicated that the transformants harbored both the pGAD424-GmWRI1a effector plasmid and the pLacZi-AW/pLacZi-mAW reporter plasmid (Fig. 4b, left panel). The results of the β-galactosidase (encoded by *lacZ*) activity assay by lifted-filter (Fig. 4b, right panel) showed that the activation of the *lacZ* reporter gene and binding of the GmWRI1a protein to wild-type AW-box has occurred, but the yeast cells harboring GmWRI1a and the mutated AW-box exhibited no β-galactosidase activity. These results proved that GmWRI1a as an AP2/EREBP transcription activator can specifically interact with the AW-box *cis*-acting element and control the expression of downstream genes in soybean.

Generation and molecular characterization of transgenic soybean lines overexpressing *GmWRI1a* gene

To investigate the function of GmWRI1a, transgenic soybean plants overexpressing *GmWRI1a* gene were generated by *Agrobacterium*-mediated soybean transformation using


Fig. 4 Analysis of DNA-binding affinity and transactivation activity of GmWRI1a. **a** Scheme of the reporter and the effector constructs. P-ADH1, the promoter of the alcohol dehydrogenase gene; T-ADH1, the terminator of the alcohol dehydrogenase gene; MP-CYC1, the minimal promoter of the yeast iso-1-cytochrome C gene. **b** Comparison of transactivation activity and interaction of GmWRI1a with AW-box/mutated AW-box in yeast one-hybrid experiments. The yeast strains with integrated wild-type or mutated AW box- pLacZi construct (pLacZi-AW/pLacZi-mAW) were transformed with the GmWRI1a effector (pGAD424-GmWRI1a). Transformants were grown on a medium lacking uracil and leucine (SD/-Ura/-Leu plate). β-galactosidase colony-lift filters assays were performed on the transformants (lifted-filter). *SD* synthetic drop-out medium, *Ura* uracil, *Leu* leucine
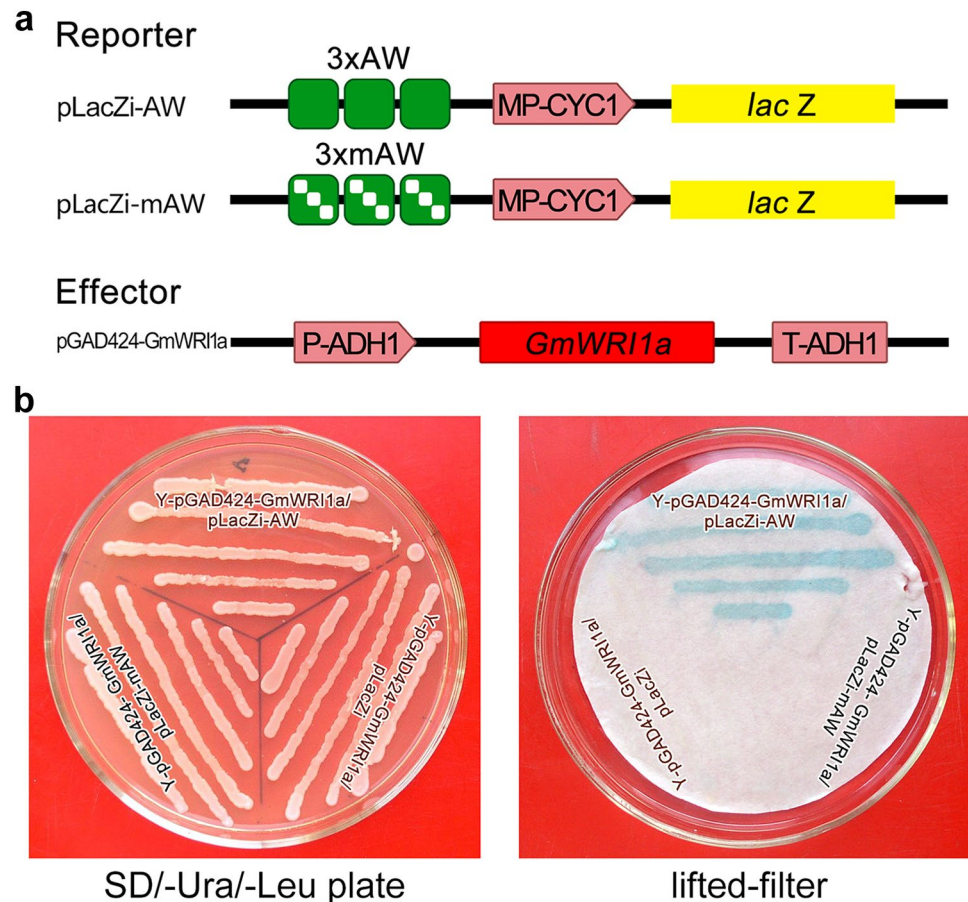
the cotyledonary node explant. The T-DNA region of the transformation vector contained the *GmWRI1a* gene and the *bar* selectable marker gene (Fig. 5a). The *GmWRI1a*-positive transgenic plants in T_0 generation were selected by Basta spray and confirmed by PCR analysis. The T_4 progenies derived from four Basta-resistant, PCR-positive and randomly selected transgenic soybean lines T19, T22, T25 and T28 were used to detect the stable integration and expression of *GmWRI1a* transgene. The result of southern hybridization showed that the transgenic lines showed the presence of three and fewer hybridization bands, whereas no hybridization band was detected in genomic DNA of wild-type plants (Fig. 5b). The hybridization bands in the progenies of four transgenic soybean lines have different patterns, indicating that the *GmWRI1a* gene was integrated into different sites of genome and these four lines were independently derived.

Expression of the *GmWRI1a* gene in four individual transgenic soybean lines was then determined by northern blotting analysis. Two independent trials were conducted to determine the expression level of *GmWI1a* transcripts.

Within the first trial, mRNAs from the leaves of the napin:*GmWRI1a* transgenic plants and wild-type plants were used. The results showed that none of *GmWRI1a* transcript was detected in the leaves of independently derived transgenic lines (Fig. 5c), because *GmWRI1a* was expressed specifically in transgenic soybean seed under the control of the seed-specific napin promoter. And none of *GmWRI1a* transcript was detected in the leaves of wild-type plants (Fig. 5c), because *GmWRI1a* was not specifically expressed in the leaf but in the maturing seed. Within the second trial, mRNAs from the 40 DAF maturing seeds of the napin:*GmWRI1a* transgenic plants and wild-type plants were used. The results showed that the transcript levels of *GmWRI1a* in the transgenic soybean lines increased dramatically, compared with that in wild-type plants (Fig. 5c). GmWRI1a transcript was detected in the seeds of wild-type plants, because of the specificity of the expression of *GmWRI1a* in the maturing seed. All the above results verified that the stable inheritance and overexpression of *GmWRI1a* in the seeds of transgenic soybean lines.
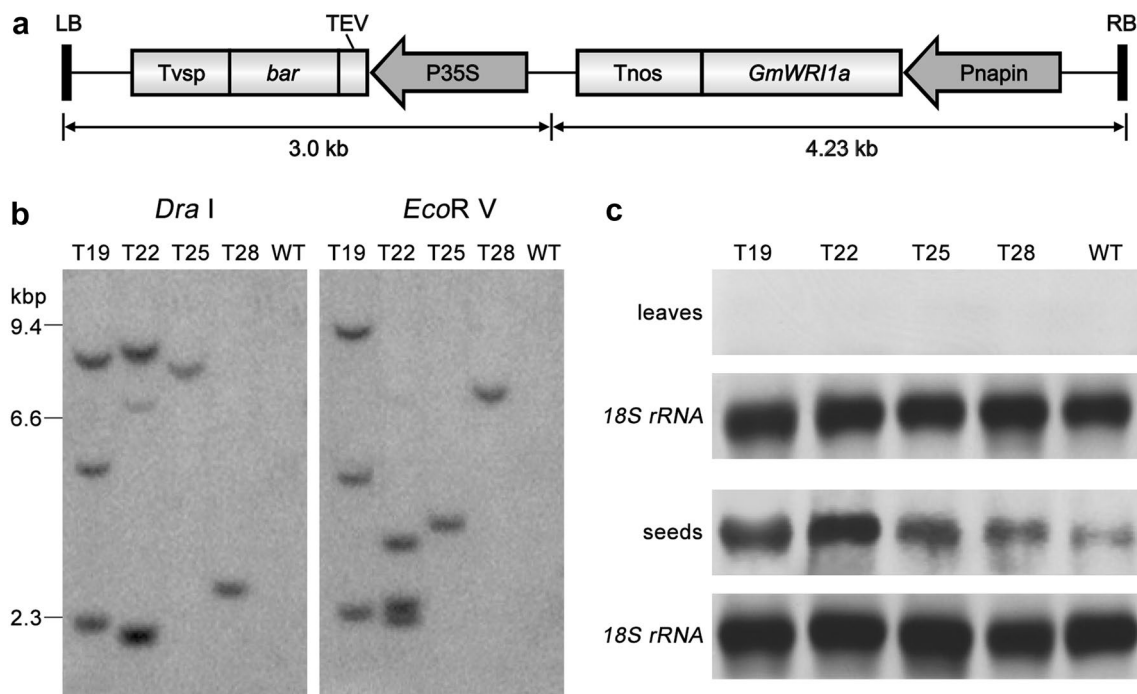

Fig. 5 Analysis of integration and overexpression of *GmWRI1a* in transgenic soybean lines. **a** T-DNA region of the binary pSE vector harboring the *GmWRIa* and bar genes. LB/RB, left/right T-DNA border sequences; bar, coding region of the phosphinothricin acetyl-transferase gene; *GmWRIa*, coding region of the *GmWRIa* gene; Pnapin, seed-specific napin promoter; Tnos, *Agrobacterium* nos terminator; P35S, CaMV 35S promoter; TEV, tobacco etch virus translational enhancer; Tvsp, soybean storage protein terminator. **b** Southern blot analysis of *GmWRI1a* in transgenic soybean lines. The genomic DNA of four T_4 transgenic soybean lines (T19, T22, T25 and T28) and wild-type (WT) plants were digested by *Dra* I and *Eco*R V, electrophoresed, blotted and hybridized with a DIG-labeled probe of the specific sequence of napin promoter and *GmWRI1a*. **c** Northern blot analysis of *GmWRI1a* in transgenic soybean lines. Total RNA from the leaves and the seeds at 40 DAF of four T_4 positive soybean lines (T19, T22, T25 and T28) and wild-type (WT) plants was loaded for each lane. The blot was hybridized with a DIG-labeled probe of the specific 5′-terminal fragment of *GmWRI1a*. 18S rRNA was used as a control

Overexpression of *GmWRI1a* increased the content of total oil and total fatty acids in soybean seed

The content of total oil and total fatty acids was measured and compared in the seeds of napin:*GmWRI1a* transgenic plants and wild-type plants. Figure 6a showed that the *GmWRI1a* transgenic lines had significantly ($P < 0.05$) higher seed oil content than wild-type plants. To further verify whether the increase of total fatty acids was due to
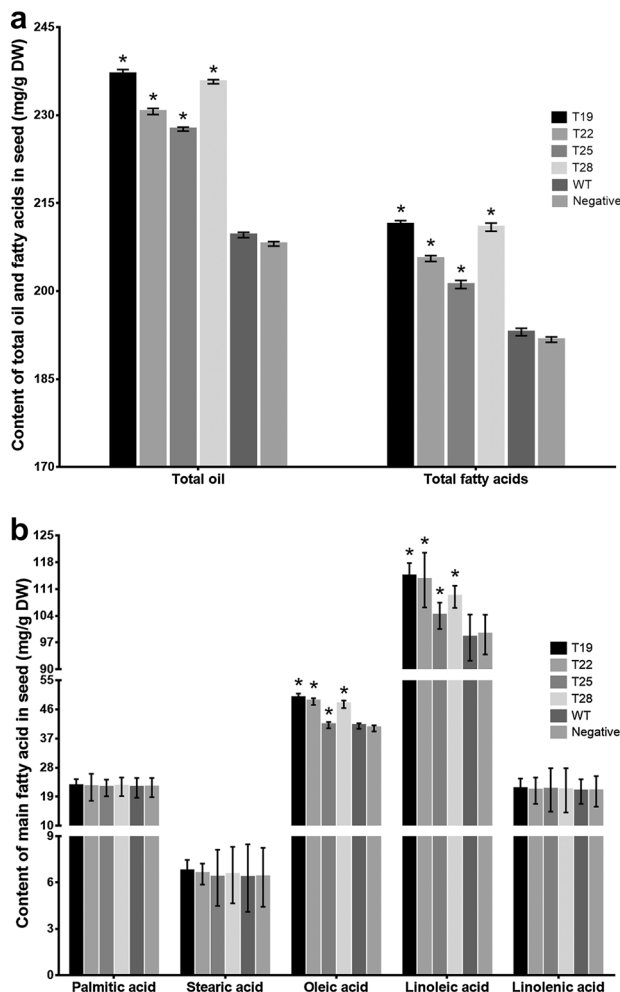
Fig. 6 Chemical composition in the seeds of transgenic soybean lines overexpressing *GmWRI1a* gene. **a** Content of total oil and total fatty acids in the seeds of WT, null transgenic and four T₄ positive *GmWRI1a*-overexpressing transgenic (T19, T22, T25 and T28) plants. The data represent the means ± SD of three replicate experiments and were analyzed by Student's *t* test ($n = 5$) and the values are in dry weight (DW) for seeds. Asterisks indicate significant differences compared with the wild-type at $P < 0.05$. **b** Content of main fatty acids in the seeds of WT, null transgenic (Negative) and four T₄ positive *GmWRI1a*-overexpressing transgenic (T19, T22, T25 and T28) plants. The data represent the means ± SD of three replicate experiments and were analyzed by Student's *t* test ($n = 5$) and the values are in dry weight (DW) for seeds. Asterisks indicate significant differences compared with the wild-type at $P < 0.05$

the increase of one or several specific FA composition, we compared major fatty acid composition in the seeds of the *GmWRI1a* transgenic lines with wild-type plants and null transgenic plants, which were segregated from heterozygous transgenic plants. As shown in Fig. 6b, two fatty acids, oleic acid (C18:1) and linoleic acid (C18:2), showed significant increases in the seeds of transgenic *GmWRI1a* lines. There were no significant differences in the other three fatty acid compositions, palmitic acid (C16:0), stearic acid (C18:0), and alpha linolenic acid (C18:3), between *GmWRI1a* transgenic lines and wild-type plants. These results suggested that overexpression of *GmWRI1a* increased the content of total oil and total fatty acids in the seeds of the *GmWRI1a* transgenic lines, and this increase was due to the increased levels of specific FA composition. As a transcription factor, GmWRI1a can positively regulate oil accumulation and change FA composition in soybean seed.

The growth of transgenic plants was compared with the growth of WT and null transgenic plants, which were segregated from heterozygous transgenic plants. Overexpression of *GmWRI1a* caused late flowering and dwarf phenotype. As showed in Table 1, the transgenic line flowered at approximately 56 DAE, but WT flowered at approximately 49 DAE. The transgenic plants were significantly shorter than WT plants and the average internode length was notably shorter in the *GmWRI1a* transgenic plants. However, the number of internodes was similar. It can be deduced that the dwarf phenotype was caused by the shortened internodes. In addition, *GmWRI1a* transgenic plants have a significantly higher seed weight per plant and per 100 seeds than WT (Table 1). Other yield parameters, such as pod number and seed number, were similar between WT and *GmWRI1a* transgenic plants (Table 1). However, the significant differences were observed between WT and *GmWRI1a* transgenic plants in the seed size traits, including 10 seed length and width (Table 1). It can be deduced that the increased seed weight of *GmWRI1a* transgenic plants was caused by large seed size.

Identification of direct downstream genes of GmWRI1a

We searched for downstream genes in transgenic plants overexpressing *GmWRI1a* using the Affymetrix Soybean GeneChip array. The total RNAs from the 40 DAF seeds of napin:*GmWRI1a* transgenic plants and wild-type plants were used for the preparation of cDNA probes. cDNA probes were then mixed and hybridized with the Affymetrix GeneChip. To assess the reproducibility of the GeneChip analysis, we repeated the experiment three times. Using high stringency parameters ($\log_R$ ratios greater than 0.3), analysis with the Affymetrix GeneChip array revealed an increase in the expression level of 39 transcripts only during soybean seed development

Table 1 Agronomic performance of WT, null transgenic and *GmWRI1a*-overexpressing transgenic (T19, T22, T25 and T28) plants in field

Genotype	WT	Negative	T19	T22	T25	T28
Flowering time (DAE)[d]	49 ± 0.5^b	50 ± 0.6^b	58 ± 0.9^a	54 ± 0.8^a	57 ± 1.5^a	56 ± 1.4^a
Plant height (cm)	92.9 ± 6.1^a	95.1 ± 3.4^a	77.2 ± 5.9^b	83.7 ± 5.1^b	81.4 ± 2.8^b	79.3 ± 2.9^b
Average internode length (cm)	6.4 ± 1.1^a	6.6 ± 1.3^a	4.4 ± 0.9^b	4.8 ± 1.1^b	4.7 ± 0.9^b	5.3 ± 1.3^b
Internode number/plant	14.5 ± 1.0^a	14.5 ± 0.8^a	14.2 ± 0.9^a	14.9 ± 1.1^a	14.2 ± 0.6^a	14.8 ± 1.2^a
Seed weight (g)/plant	43.8 ± 6.7^b	41.2 ± 7.2^b	53.7 ± 7.0^a	58.5 ± 7.2^a	54.0 ± 7.6^a	57.7 ± 11.8^a
Seed weight (g)/100 seeds	13.9 ± 1.5^b	14.4 ± 0.8^b	15.9 ± 0.8^a	15.6 ± 1.1^a	16.6 ± 1.3^a	16.1 ± 1.6^a
Pod number/plant	181 ± 35^a	$175 \pm 28^{a,b}$	185 ± 19^a	$172 \pm 21^{a,b}$	$174 \pm 54^{a,b}$	$179 \pm 55^{a,b}$
Seed number/plant	434 ± 90^a	422 ± 136^a	412 ± 81^a	420 ± 86^a	421 ± 131^a	435 ± 129^a
10 seed length (cm)	7.17 ± 0.08^c	7.20 ± 0.08^c	7.83 ± 0.09^b	7.84 ± 0.08^b	8.12 ± 0.07^a	7.96 ± 0.11^b
10 seed width (cm)	6.16 ± 0.07^c	6.28 ± 0.05^c	7.01 ± 0.10^b	7.09 ± 0.13^b	7.39 ± 0.06^a	7.11 ± 0.09^b

[a,b,c]Means $\pm$ SD in each column with the same letter are not significantly different from each other (LSD test, $P < 0.05$)

[d]Flowering time from emergence to the appearance of the first flower; DAE days after emergence

in napin:*GmWRI1a* transgenic plants as compared with the wild-type plants. 39 transcripts were further analyzed by qRT-PCR analysis using the same total RNA samples that had been used for the initial Affymetrix GeneChip array analysis. Analysis with qRT-PCR revealed that 28 of 39 transcripts showed increased level in transgenic seeds as compared with wild-type plants (data not shown). Therefore, the 28 transcripts were confirmed as candidate GmWRI1a downstream genes and were mapped to Arabidopsis locus IDs that provide functional annotation data. According to the matched Arabidopsis locus IDs and sequence similarities, Gene Ontology (GO) of functional categorization indicated that the 28 transcripts encoded nine enzymes. As shown in Table 2, the nine enzymes are ketoacyl-ACP synthase III coded by one transcript (Gma.6041.1.S1_at), ketoacyl-ACP synthase I coded by two near isogenic transcripts (Gma.248.1.S1_s_at and Gma.1967.1.S1_at), ketoacyl-ACP reductase coded by one transcript (GmaAffx.39337.1.S1_at), hydroxyacyl-ACP dehydrase coded by one transcript (Gma.9081.1.S1_at), enoyl-ACP reductase coded by four near isogenic transcripts (Gma.4516.1.S1_at, GmaAffx.47845.1.S1_at, Gma.7816.1.A1_at and GmaAffx.55127.1.A1_at), ketoacyl-ACP synthase II coded by three near isogenic transcripts (Gma.5093.1.S1_at, Gma.5093.2.S1_a_at and GmaAffx.87834.1.S1_at), stearoyl-ACP desaturase coded by three near isogenic transcripts (Gma.2857.1.S1_at, GmaAffx.43163.1.S1_at and Gma.6256.1.S1_s_at), acyl-ACP thioesterase B coded by seven near isogenic transcripts (Gma.4079.1.S1_at, GmaAffx.142.1.A1_at, GmaAffx.33180.1.S1_at, GmaAffx.48202.1.S1_at, GmaAffx.62625.1.S1_at, GmaAffx.7719.1.S1_at and GmaAffx.78262.1.S1_at) and long-chain acyl-CoA synthetase coded by six near isogenic transcripts (Gma.7613.1.S1_a_at, Gma.7613.1.S1_at,

Gma.7613.2.S1_at, GmaAffx.15196.1.S1_at, GmaAffx.36701.2.S1_at and GmaAffx.84626.1.S1_at). The above 28 transcripts represented nine candidate downstream genes.
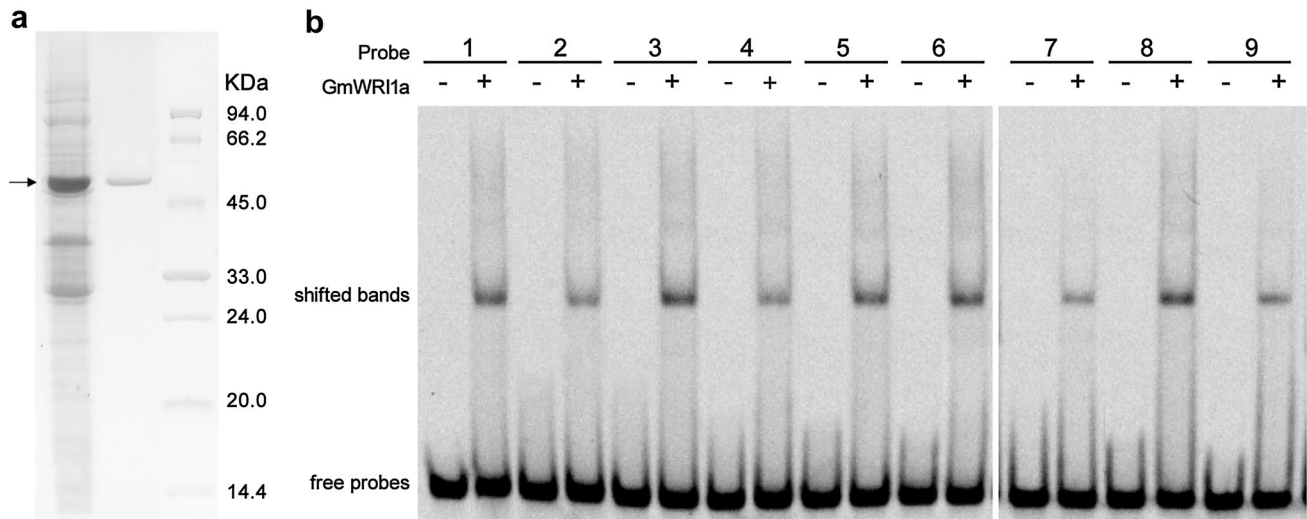
The direct downstream genes of GmWRI1a were selected from the identified candidate downstream genes based on binding affinity of their proximal upstream regions to GmWRI1a. The electrophoretic mobility shift assay (EMSA) was performed to examine whether GmWRI1a protein was able to bind to the proximal upstream regions of the nine candidate genes. The recombinant histidine-tagged GmWRI1a protein was expressed as soluble protein in *E.coli* BL21 (Fig. 7a, lane 1). $(His)_6$-GmWRI1a (46.2 kDa) was purified directly from bacterial lysates (Fig. 7a, lane 2). Addition of GmWRI1a protein to 500-bp upstream promoter fragments of the nine candidate downstream genes resulted in the formation of shifted bands (Fig. 7b). The interaction between GmWRI1a and the promoter of each downstream gene was tested by yeast one-hybrid assay. GmWRI1a can specifically bind the 500-bp sequences upstream from the ATG codon of nine candidate downstream genes and effectively activate the transcription of the *lacZ* reporter gene in the transient assay system (Supplementary Fig. S1). In addition, to examine the specificity of the binding sequence of GmWRI1a, mutant AW-box of KAS $III_{(-262/-237)}$ (ketoacyl-ACP synthase III) was prepared (Supplementary Fig. S2). As showed in Supplementary Fig. S2, mutations in the conserved nucleotides of the AW-box, $[C^{-256}\text{-}T^{-254}\text{-}G^{-252}]$ and $[C^{-244}G^{-243}]$ caused reduced GmWRI1a binding, but mutations between or outside the conserved motifs showed no effect on GmWRI1a binding. These above results indicated that the nine genes may be direct downstream genes of GmWRI1a.

Table 2 Annotation, expression level and predicted subcellular localization of target genes of GmWRI1a

No.	ID of oligonucleotide on the array[a]	Annotation	Best Glyma1.1 match	Ratio OE/wild-type[b,c] (array)	Ratio OE/wild-type[b,c] (qRT-PCR)	Predicted subcellular localization	Arabidopsis ortholog
1	Gma.6041.1.S1_at	Ketoacyl-ACP synthase III	Glyma18g44350.1	2.34**	4.77**	Plastid	AT1G62640
2	Gma.248.1.S1_s_at	Ketoacyl-ACP synthase I	Glyma08g08910.1	2.25**	4.55**	Plastid	AT5G46290
	Gma.1967.1.S1_at	Ketoacyl-ACP synthase I	No BLASTN hit	2.16*	4.13**	Plastid	AT5G46290
3	GmaAffx.39337.1.S1_at	Ketoacyl-ACP reductase	Glyma16g34820.1	2.41**	5.03**	Plastid	AT5G04050
4	Gma.9081.1.S1_at	Hydroxyacyl-ACP dehydrase	Glyma15g05800.1	2.03*	3.11*	Plastid	AT5G10160
5	Gma.4516.1.S1_at	Enoyl-ACP reductase	Glyma11g10770.4	1.82*	2.13*	Plastid	AT2G05990
	GmaAffx.47845.1.S1_at	Enoyl-ACP reductase	Glyma08g45990.1	1.71*	1.75*	Plastid	AT2G05990
	Gma.7816.1.A1_at	Enoyl-ACP reductase	Ambiguous Hit	1.97*	2.97*	Plastid	AT2G05990
	GmaAffx.55127.1.A1_at	Enoyl-ACP reductase	No BLASTN hit	1.71*	1.75*	Plastid	AT2G05990
6	Gma.5093.1.S1_at	Ketoacyl-ACP synthase II	Glyma17g05200.4	1.99*	3.08*	Plastid	AT1G74960
	Gma.5093.2.S1_a_at	Ketoacyl-ACP synthase II	Glyma13g17290.3	1.92*	2.39*	Plastid	AT1G74960
	GmaAffx.87834.1.S1_at	Ketoacyl-ACP synthase II	Glyma13g17290.3	1.96*	2.77*	Plastid	AT1G74960
7	Gma.2857.1.S1_at	Stearoyl-ACP desaturase	Glyma14g27990.1	2.08*	3.74**	Plastid	AT1G43800
	GmaAffx.43163.1.S1_at	Stearoyl-ACP desaturase	Glyma07g32850.1	2.31**	4.67**	Plastid	AT2G43710
	Gma.6256.1.S1_s_at	Stearoyl-ACP desaturase	No BLASTN hit	2.19*	4.43**	Plastid	AT2G43710
8	Gma.4079.1.S1_at	Acyl-ACP thioesterase B	Glyma05g08060.4	1.79*	1.97*	Plastid	AT1G08510
	GmaAffx.142.1.A1_at	Acyl-ACP thioesterase B	Glyma04g21910.1	2.06*	3.73**	Plastid	AT1G08510
	GmaAffx.33180.1.S1_at	Acyl-ACP thioesterase B	Glyma04g37420.2	1.76*	1.86*	Plastid	AT1G08510
	GmaAffx.48202.1.S1_at	Acyl-ACP thioesterase B	Glyma04g21910.1	2.11*	4.11**	Plastid	AT1G08510
	GmaAffx.62625.1.S1_at	Acyl-ACP thioesterase B	Glyma05g08060.4	1.97*	2.97*	Plastid	AT1G08510
	GmaAffx.7719.1.S1_at	Acyl-ACP thioesterase B	Glyma06g23560.1	2.04*	3.34*	Plastid	AT1G08510
	GmaAffx.78262.1.S1_at	Acyl-ACP thioesterase B	Glyma06g23560.1	2.09*	3.94**	Plastid	AT1G08510
9	Gma.7613.1.S1_a_at	Long-chain acyl-CoA synthetase	Glyma20g28200.1	2.27**	4.62**	Peroxisome	AT5G27600
	Gma.7613.1.S1_at	Long-chain acyl-CoA synthetase	Glyma20g28200.1	2.24**	4.52**	Peroxisome	AT5G27600
	Gma.7613.2.S1_at	Long-chain acyl-CoA synthetase	Glyma10g39540.1	2.18*	4.35**	Peroxisome	AT3G05970
	GmaAffx.15196.1.S1_at	Long-chain acyl-CoA synthetase	Glyma07g20860.1	2.36**	5.01**	None	AT1G49430
	GmaAffx.36701.2.S1_at	Long-chain acyl-CoA synthetase	Glyma12g05140.1	2.09*	3.94**	None	AT1G49430
	GmaAffx.84626.1.S1_at	Long-chain acyl-CoA synthetase	Glyma20g28200.1	2.34**	4.77**	Peroxisome	AT5G27600

[a]ID of the corresponding oligonucleotide deposited on Affymetrix Soybean Genome Array

[b]Student's *t* test value: **$P < 0.01$, *$P < 0.05$

Fig. 7 Binding of GmWRI1a to the proximal upstream regions of direct downstream genes. **a** Bacterial expression and purification of the histidine-tagged GmWRI1a, (His)[6]-GmWRI1a. Purified protein and soluble extracts of *E. coli* treated with IPTG were resolved by 12% SDS-PAGE and stained with Coomassie Brilliant Blue. Arrows mark the anticipated protein species. The size (KDa) of the protein markers was shown on the right side. **b** Electrophoretic mobility shift assays of GmWRI1a with probes covering 500-bp sequences upstream from the ATG codon of the gene encoding ketoacyl-ACP synthase III (Probe 1), ketoacyl-ACP synthase I (Probe 2), ketoacyl-ACP reductase (Probe 3), hydroxyacyl-ACP dehydratase (Probe 4), enoyl-ACP reductase (Probe 5), ketoacyl-ACP Synthase II (Probe 6), stearoyl-ACP desaturase (Probe 7), acyl-ACP thioesterases (Probe 8), or long-chain acyl-CoA synthetase (Probe 9). Positions of the shifted bands and the free probes were indicated

Discussion

Soybean [*Glycine max* (L.) Merrill] is a major leguminous seed crop providing an important source of oil, and ranks first in oil production among the major oil seed crops (Yu et al. 2014). Elevating oil accumulation in soybean seed has always been an important goal of breeders and metabolic engineers for decades (Bates et al. 2013). In recent years, with a detailed understanding of the metabolic pathways in the model plant *Arabidopsis thaliana*, it has been known that plant seed oil accumulation is regulated by a complex network involved in the biosynthetic pathway of TAG, plastid FAs, endomembrane lipids and the storage process (Eskandari et al. 2013; La Russa et al. 2012; Xia et al. 2014). Some important transcription factors regulating oil biosynthesis and seed maturation, such as WRI1, LEC2, LEC1, FUS3 and ABI3, have been reported to be involved in a complex network to regulate oil accumulation. Altering expression levels of TF genes can provide an efficient alternative for elevating seed oil accumulation to single-enzyme approaches (Chai et al. 2010; Ibáñez-Salazar et al. 2014).

Previously, several soybean transcriptional regulators have been identified promoting seed oil synthesis (Li et al. 2017; Manan et al. 2017). Overexpression of *GmDOF4* and *GmDOF11* increased lipid content in seeds of transgenic Arabidopsis plants via the direct activation of lipid biosynthesis genes and the repression of storage protein genes (Wang et al. 2007). Overexpression of *GmDOF4* in *Chlorella ellipsoidea* also enhanced lipid contents (Zhang et al. 2014). *GmMYB73* elevated the total lipid contents in the leaves and seeds of transgenic Lotus and in transgenic hairy roots of soybean plants by accelerating the conversion of phosphatidylcholine to TAG (Liu et al. 2014). *GmbZIP123* elevated lipid contents in seeds of transgenic Arabidopsis plants by activating Suc transporter genes and cell wall invertase genes for sugar translocation and sugar breakdown (Song et al. 2013). Overexpression of *GmLEC1* increased lipid content in seeds of transgenic Arabidopsis plants via interacting with other TFs to regulate distinct gene sets at different stages of seed development (Pelletier et al. 2017). Overexpression of *GmDREBL* increased lipid content in the seeds of transgenic Arabidopsis plants via participating in the regulation of fatty acid accumulation and controlling the expression of WRI1 and its downstream genes (Zhang et al. 2016). Therefore, identification of TFs involved in soybean oil biosynthesis is necessary not only for understanding oil biosynthetic pathway in

more detail, but also for applying them to improve soybean oil, both quantitatively and qualitatively (Iwase et al. 2009).

In this study, three distinct orthologues of *WRI1* transcription factor gene, *GmWRI1a, GmWRI1b* and *GmWRI1c*, were isolated in soybean. Sequence alignment indicated that the three WRI1 proteins belong to the AP2/EREBP transcription factor family and contained double conserved AP2/EREBP domains. The three soybean *WRI1* genes were classified into three groups based on their organ-specific expression patterns, among which *GmWRI1a, GmWRI1b* and *GmWRI1c* were preferentially expressed in maturing seed, stem and flower, respectively. The differences in expression pattern of three soybean *WRI1* genes likely reflect an ongoing specialization on an evolutionary time scale. It can be speculated that *GmWRI1b* and *GmWRI1c* are evolving to have extensive biological functions at vegetative and reproductive periods prior to soybean seed formation while *GmWRI1a* mainly devotes to FA biosynthesis at soybean seed formation stage (Pouvreau et al. 2011). In addition, we found that the transcript of *GmWRI1a* showed peak level on 40 DAF. This time point approximately coincides with the period when soybean seed had the most accumulation of oil during seed formation stage (Eskandari et al. 2013). It can be inferred that *GmWRI1a* may play an important role in regulating oil accumulation in soybean seed.

We further studied the effect of *GmWRI1a* on seed oil accumulation in transgenic soybean plants. Considering that overexpression of TF genes caused severe growth retardation of transgenic plants (Chen et al. 2008), we used the seed-specific napin promoter instead of the constitutive CaMV 35S promoter and negative effects on the growth of napin:*GmWRI1a* transgenic soybean plants were evidently minimized. The *GmWRI1a* transgenic soybean lines exhibit reduced height, reduced internode length, extended flowering time, increased seed size and increased seed weight compared with the wild type. More importantly, transgenic soybean seeds showed a significant increase not only in the content of total oil but also in the content of total fatty acids, compared to the non-transgenic controls. In addition, the overexpression of *GmWRI1a* changed FA composition and ratio. The transgenic seeds accumulated more oleic acid (C18:1) and linoleic acid (C18:2) than non-transgenic control. It can be inferred that the transcriptional machinery of GmWRI1a may be composed of a set of direct downstream genes involved in FA biosynthesis.

It is important to identify the direct downstream genes of GmWRI1a precisely to understand the WRI1 transcription factor-regulated signal pathway and molecular mechanisms involved in soybean oil accumulation. We identified nine candidate downstream genes of GmWRI1a using the Affymetrix GeneChip array and qRT-PCR analysis. The direct downstream genes for GmWRI1a satisfied the following

three criteria: (1) involved in FA biosynthesis; (2) existence of the AW-box in the promoter region; and (3) binding of their proximal upstream regions to GmWRI1a. We investigated the functions of the identified genes, and estimated how overexpression of *GmWRI1a* increased the content of total oil and total fatty acids and changed fatty acid composition. The identified candidate downstream genes of GmWRI1a were classified into two groups. The first group includes enzymes that probably function in catalyse reactions of FA production (Bates et al. 2013; Cagliari et al. 2011), such as ketoacyl-ACP synthase III, ketoacyl-ACP synthase I, ketoacyl-ACP reductase, hydroxyacyl-ACP dehydrase and enoyl-ACP reductase. These gene products probably function to increase the content of total oil and total fatty acids in soybean seed. The second group contains enzymes involved in further regulation of signal transduction and gene expression that probably function in catalyse reactions of FA elongation, desaturation and export from plastid (Bates et al. 2013; Cagliari et al. 2011), such as ketoacyl-ACP synthase II, stearoyl-ACP desaturase, acyl-ACP thioesterase B and long-chain acyl-CoA synthetase. These gene products probably function to change fatty acid composition in soybean seed. The sequence analysis confirmed the existence of the AW-box in their promoter regions. All the nine genes contained a conserved sequence [CnTnG](n)7[CG] (where n is any nucleotide and [CnTnG] and [CG] were separated by 7 bp) as AW-box *cis*-element in their -51 to -450 promoter regions. EMSA and yeast one-hybrid assay revealed that GmWRI1a protein was able to bind to AW-box within the promoters of the nine genes. It can be concluded that all the nine genes are direct downstream genes of GmWRI1a transcription factor. And more importantly, the wide range of functions of the direct downstream genes indicated that GmWRI1a transcription factor regulated a complex gene expression network for controlling the flux through FA biosynthesis.

In addition, it is necessary to demonstrate whether GmWRI1a activates the same downstream genes as AtWRI1 in Arabidopsis and ZmWRI1 in maize (Pouvreau et al. 2011). We found that common downstream genes between soybean, Arabidopsis and maize include ketoacyl-ACP synthase III and acyl-ACP thioesterase B; common downstream genes between soybean and Arabidopsis include hydroxyacyl-ACP dehydrase and enoyl-ACP reductase; common downstream genes between soybean and maize include long-chain acyl-CoA synthetase; novel downstream genes detected only in soybean include ketoacyl-ACP synthase I, ketoacyl-ACP reductase, ketoacyl-ACP Synthase II, stearoyl-ACP desaturase. And previous studies had shown that AtWRI1 in Arabidopsis can mainly activate genes coding for enzymes in late glycolysis (Baud and Lepiniec 2009). In this study, we also found that the direct downstream genes of GmWRI1a primarily involved in late steps of FA

biosynthesis. The exact differences and similarities of WRI1 transcriptional machinery among various plants still need to be further studied in the future.

In summary, considering the growing global demand for soybean oil, the identification of TFs regulating oil accumulation in soybean seed is a research area of great interest. We isolated and characterized soybean WRI1 transcription factor genes in this study. Among the three homologous WRI1 genes exhibiting different organ-specific expression patterns, only *GmWRI1a* was found to be expressed in maturing soybean seed. Overexpression of *GmWRI1a* can significantly increase the content of total oil and total fatty acids, and change FA composition and ratio in soybean seed. As an AW box-binding AP2/EREBP transcription activator, GmWRI1a can regulate a complex gene expression network for controlling the flux through FA production, elongation, desaturation and export from plastid. To our knowledge, this is the first report on the identification and characterization in detail of soybean WRI1 transcription factor gene. This study will contribute to a better understanding of WRI1 transcriptional machinery and oil biosynthetic pathway in soybean, and WRI1 transcription factor can be applied to improve soybean oil, both quantitatively and qualitatively.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Bates PD, Stymne S, Ohlrogge J (2013) Biochemical pathways in seed oil synthesis. Curr Opin Plant Biol 16:358–364

Baud S, Lepiniec L (2009) Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. Plant Physiol Bioch 47:448–455

Baud S, Wuillème S, To A, Rochat C, Lepiniec L (2009) Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in Arabidopsis. Plant J 60:933–947

Bolton MD, Kolmer JA, Xu WW, Garvin DF (2008) Lr34-mediated leaf rust resistance in wheat: transcript profiling reveals a high energetic demand supported by transient recruitment of multiple metabolic pathways. Mol Plant Microbe In 21:1515–1527

Cagliari A, Margis R, dos Santos MF, Turchetto-Zolet AC, Loss G, Margis-Pinheiro M (2011) Biosynthesis of Triacylglycerols (TAGs) in plants and algae. Int J Plant Biol 2:10

Chai G, Bai Z, Wei F, King GJ, Wang C, Shi L, Dong C, Chen H, Liu S (2010) Brassica GLABRA2 genes: analysis of function related to seed oil content and development of functional markers. Theor Appl Genet 120:1597–1610

Chapman KD, Ohlrogge JB (2012) Compartment of triacylglycerol accumulation in plants. J Biol Chem 287:2288–2294

Chen JE, Smith AG (2012) A look at diacylglycerol acyltransferases (DGATs) in algae. J Biotechnol 162:28–39

Chen L, Zhang ZY, Liang HX, Liu HX, Du LP, Xu HJ, Xin ZY (2008) Overexpression of TiERF1 enhances resistance to sharp eyespot in transgenic wheat. J Exp Bot 59:4195–4204

Chen H, Wang FW, Dong YY, Wang N, Sun YP, Li XY, Liu L, Fan XD, Yin HL, Jing YY, Zhang XY, Li YL, Chen G, Li HY (2012) Sequence mining and transcript profiling to explore differentially expressed genes associated with lipid biosynthesis during soybean seed development. BMC Plant Biol 12:1

Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiol 151:1030–1040

Courchesne NMD, Parisien A, Wang B, Lan CQ (2009) Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. J Biotechnol 141:31–41

Dussert S, Guerin C, Andersson M, Joët T, Tranbarger TJ, Pizot M, Sarah G, Omore A, Durand-Gasselin T, Morcillo F (2013) Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiol 162:1337–1358

Eskandari M, Cober ER, Rajcan I (2013) Using the candidate gene approach for detecting genes underlying seed oil concentration and yield in soybean. Theor Appl Genet 126:1839–1850

Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Method Enzymol 350:87–96

Grotewold E (2008) Transcription factors for predictive plant metabolic engineering: are we there yet? Curr Opin Plant Biol 19:138–144

Hwang EY, Song Q, Jia G, Specht JE, Hyten DL, Costa J, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genom 15:1

Ibáñez-Salazar A, Rosales-Mendoza S, Rocha-Uribe A, Ramírez-Alonso JI, Lara-Hernández I, Hernández-Torres A, Paz-Maldonado LMT, Silva-Ramírez AS, Bañuelos-Hernández B, Martínez-Salgado JL, Soria-Guerra RE (2014) Over-expression of Dof-type transcription factor increases lipid production in *Chlamydomonas reinhardtii*. J Biotechnol 184:27–38

Iwase A, Matsui K, Ohme-Takagi M (2009) Manipulation of plant metabolic pathways by transcription factors. Plant Biotechnol 26:29–38

Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM, Covello PS, Taylor DC (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

La Russa M, Bogen C, Uhmeyer A, Doebbe A, Filippone E, Kruse O, Mussgnug JH (2012) Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. J Biotechnol 162:13–20

Laibach N, Post J, Twyman RM, Gronover CS, Prüfer D (2015) The characteristics and potential applications of structural lipid droplet proteins in plants. J Biotechnol 201:15–27

Li QT, Lu X, Song QX, Chen HW, Wei W, Tao JJ, Bian XH, Shen M, Ma B, Zhang WK, Bi YD, Li W, Lai YC, Lam SM, Shui GH, Chen SY, Zhang JS (2017) Selection for a zinc-finger protein contributes to seed oil increase during soybean domestication. Plant Physiol 173:2208–2224

Liu J, Hua W, Zhan GM, Wei F, Wang XF, Liu GH, Wang HZ (2010) Increasing seed mass and oil content in transgenic Arabidopsis by the overexpression of wri1-like gene from *Brassica napus*. Plant Physiol Bioch 48:9–15

Liu J, Hua W, Yang HL, Zhan GM, Li RJ, Deng LB, Wang XF, Liu GH, Wang HZ (2012) The BnGRF2 gene (GRF2-like gene from Brassica napus) enhances seed oil production through regulating cell number and plant photosynthesis. J Exp Bot 63:3727–3740

Liu YF, Li QT, Lu X, Song QX, Lam SM, Zhang WK, Ma B, Lin Q, Man WQ, Du WG, Shui GH, Chen SY (2014) Soybean GmMYB73 promotes lipid accumulation in transgenic plants. BMC Plant Biol 14:73

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408

Ma W, Kong Q, Arondel V, Kilaru A, Bates PD, Thrower NA, Benning C, Ohlrogge JB (2013) WRINKLED1, a ubiquitous regulator in oil accumulating tissues from Arabidopsis embryos to oil palm mesocarp. PLoS One 8:e68887

Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai T, Tsukagoshi H, Ishiguro S, Nakamura K (2009) An AP2-type transcription factor, WRINKLED1, of Arabidopsis thaliana binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J 60:476–487

Manan S, Chen B, She G, Wan X, Zhao J (2017) Transport and transcriptional regulation of oil production in plants. Crit Rev Biotechnol 37:641–655

Nölke G, Fischer R, Schillberg S (2006) Antibody-based metabolic engineering in plants. J Biotechnol 124:271–283

Paz MM, Shou HX, Guo ZB, Zhang ZY, Banerjee AK, Wang K (2004) Assessment of conditions affecting Agrobacterium-mediated soybean transformation using the cotyledonary node explant. Euphytica 136:167–179

Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliari A, Hashimoto M, Munoz MD, Fischer RL, Goldberg RB, Harada JJ (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. Proc Natl Acad Sci USA 114:E6710–E6719

Pouvreau B, Baud S, Vernoud V, Morin V, Py C, Gendrot G, Pichon JP, Rouster J, Paul W, Rogowsky PM (2011) Duplicate maize wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis. Plant Physiol 156:674–686

Rachael S, Tang JH (2006) Modification of oil using conventional and transgenic approaches. Crop Sci 46:1225–1236

Sanjaya Durrett TP, Weise SE, Benning C (2011) Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. Plant Biotechnol J 9:874–883

Shen B, Allen WB, Zheng PZ, Li CJ, Glassman K, Ranch J, Nubel D, Tarczynski MC (2010) Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant Physiol 153:980–987

Song QX, Li QT, Liu YF, Zhang FX, Ma B, Zhang WK, Man WQ, Du WG, Wang GD, Chen SY, Zhang JS (2013) Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic Arabidopsis plants. J Exp Bot 64:4329–4341

Stalberg K, Ellerstom M, Ezcurra I, Ablov S, Rask L (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. Planta 199:515–519

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci 100:9440–9445

Tajima D, Kaneko A, Sakamoto M, Ito Y, Hue NT, Miyazaki M, Ishibashi Y, Yuasa T, Iwaya-Inoue M (2013) Wrinkled1 (WRI1) homologs, AP2-type transcription factors involving master regulation of seed storage oil synthesis in castor bean (*Ricinus communis* L.). Am J Plant Sci 4:333–339

van Erp H, Kelly AA, Menard G, Eastmond PJ (2014) Multigene engineering of triacylglycerol metabolism boosts seed oil content in Arabidopsis. Plant Physiol 165:30–36

Vigeolas H, Waldeck P, Zank T, Geigenberger P (2007) Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol J 5:431–441

Wang HW, Zhang B, Hao YJ, Huang J, Tian AG, Liao Y, Zhang JS, Chen SY (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

Weselake RJ, Taylor DC, Rahman HM, Shah S, Laroche A, McVetty PBE, Harwood JL (2009) Increasing the flow of carbon into seed oil. Biotechnol Adv 27:866–878

Xia W, Mason AS, Xiao Y, Liu Z, Yang Y, Lei X, Wu X, Ma Z, Peng M (2014) Analysis of multiple transcriptomes of the African oil palm (*Elaeis guineensis*) to identify reference genes for RT-qPCR. J Biotechnol 184:63–73

Yu J, Zhang ZH, Wei JG, Ling Y, Xu WY, Su Z (2014) SFGD: a comprehensive platform for mining functional information from soybean transcriptome data and its use in identifying acyl-lipid metabolism pathways. BMC Genom 15:1

Zhang ZY, Yao WL, Dong N, Liang HX, Liu HX, Huang RF (2007) A novel ERF transcription activator in wheat and its induction kinetics after pathogen and hormone treatments. J Exp Bot 58:2993–3003

Zhang J, Hao Q, Bai L, Xu J, Yin W, Song L, Xu L, Guo X, Fan C, Chen Y, Ruan J, Hao S, Li Y, Wang RR-C, Hu Z (2014) Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of Chlorella ellipsoidea. Biotechnol Biofuels 7:128

Zhang YQ, Lu X, Zhao FY, Li QT, Niu SL, Wei W, Zhang WK, Ma B, Chen SY, Zhang JS (2016) Soybean GmDREBL increases lipid content in seeds of transgenic Arabidopsis. Sci Rep 6:34307

Zhao JY, Huang JX, Chen F, Xu F, Ni XY, Xu HM, Wang YL, Jiang CC, Wang H, Xu AX, Huang RZ, Li DR, Meng JL (2012) Molecular mapping of *Arabidopsis thaliana* lipid-related orthologous genes in *Brassica napus*. Theor Appl Genet 124:407–421