

Integrating GWAS, QTL Mapping, and RNA-seq to Identify Candidate Genes for Kernel Weight and Oil Traits in Maize (*Zea mays* ssp. *mays*)

Lijun Zhang,^{||} Yuhan Song,^{||} Meiyan Chen, Chunyu Hu, Wanru Wang, Xunbo Zhou, Xiaoqin Hao,*
Xingtang Zhang,* and Xiping Yang*



Cite This: *J. Agric. Food Chem.* 2026, 74, 11285–11295



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

ABSTRACT: The evaluation of the kernel oil content and fatty acid composition in maize is crucial for understanding maize nutrition. This study utilized an association population of 278 inbred lines and a F_2 population of 229 individuals to identify candidate genes for 8 kernel weight and oil traits. By employing two genome-wide association study (GWAS) models, the K model and Q+K model, key SNP loci and candidate genes were identified. A total of 10 pleiotropic quantitative trait loci (QTLs) were discovered that collectively encompassed 82 candidate genes. Notably, the *SAD* gene catalyzed the conversion of stearyl-ACP to oleoyl-ACP, playing a crucial role in regulating the content of unsaturated fatty acids in plants. Three candidate genes were identified through three methods (GWAS, QTL mapping, and RNA-seq), with the *FALDH* gene playing a role in regulating the grain oil content. Through the integration of GWAS analysis and QTL mapping, one stable QTL related to the oil content (*qOCS*) and one stable QTL related to arachidic acid (*qAA9*) were identified. Among them, the *FAE2* gene was located in *qAA9*, which plays a crucial role in regulating the fatty acid content in plants. Additionally, the gene *TT12*, identified through a combination of GWAS and RNA-seq analyses, has been found to play a pivotal role in seed development. Overall, these findings provide a foundation for deeper insights into the genetic mechanisms underlying maize oil traits and offer valuable insights for molecular marker-assisted breeding.

KEYWORDS: GWAS analysis, QTL mapping, RNA-seq, Oil content, Maize

1. INTRODUCTION

Maize is the largest grain crop globally and is primarily used for industrial materials, animal feed, and human consumption. High-oil maize, a selectively bred variant with an oil content exceeding 8%, doubling that of ordinary maize, is valued for its multiple purposes in oil production, grain yield, and livestock feed. Enhancing oil content of maize is essential for increasing its economic worth.¹ However, oil trait in maize is intricate, influenced by multiple quantitative trait loci (QTLs).² Maize oil content demonstrates high heritability with phenotypic variations resulting from a combination of genetic factors, environmental conditions, and their interactions. Therefore, studies based on phenotyping of multiple years are necessary to pinpoint reliable genetic loci associated with maize oil content. The components of maize oil traits include the oil content (OC), palmitic acid (PA), stearic acid (SA), oleic acid (OA), linoleic acid (LA), linolenic acid (LAD), and arachidic acid (AA). A thorough investigation into genetic mechanisms controlling oil components is crucial for systematically identifying genes responsible for regulating maize oil content and selecting superior high-oil maize germplasm.

High-oil maize, a genetic resource of great value, has been developed through long-term artificial selection.^{3,4} Micro-endosperm super-high-oil maize, with the embryo as the main component (embryo weight ratio $\geq 40\%$) and a total oil content exceeding 20%, represents a novel maize germplasm

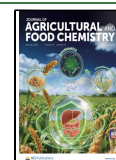
with low endosperm content. This germplasm features reduced endosperm, enlarged embryos, a high oil content, and elevated protein levels. The kernels consist solely of pericarp (including seed coat), endosperm, and embryo with minimal starch reserves. The extremely low starch content diminishes metabolic cross-talk with lipid biosynthesis, rendering this material an ideal system for dissecting embryo- and oil-specific genetic pathways.⁵ Furthermore, the whole kernels can be utilized directly for oil extraction without de-embryonation, highlighting its potential as a novel oil maize germplasm. As a unique and valuable resource, unraveling the genetic structure of oil biosynthesis in maize kernels are significant. Previous studies have shown that the synthesis and accumulation of oil in plant seeds are highly intricate, involving over 120 enzymes and more than 600 genes in the oil metabolism pathway of the model plant *Arabidopsis thaliana*.⁶ In addition, research on maize kernel oil metabolism has been limited, with only a few genes cloned that were associated with oil content and fatty acid composition in maize kernels.^{7–9} For example, Zheng et

Received: September 30, 2025

Revised: March 3, 2026

Accepted: March 3, 2026

Published: March 24, 2026



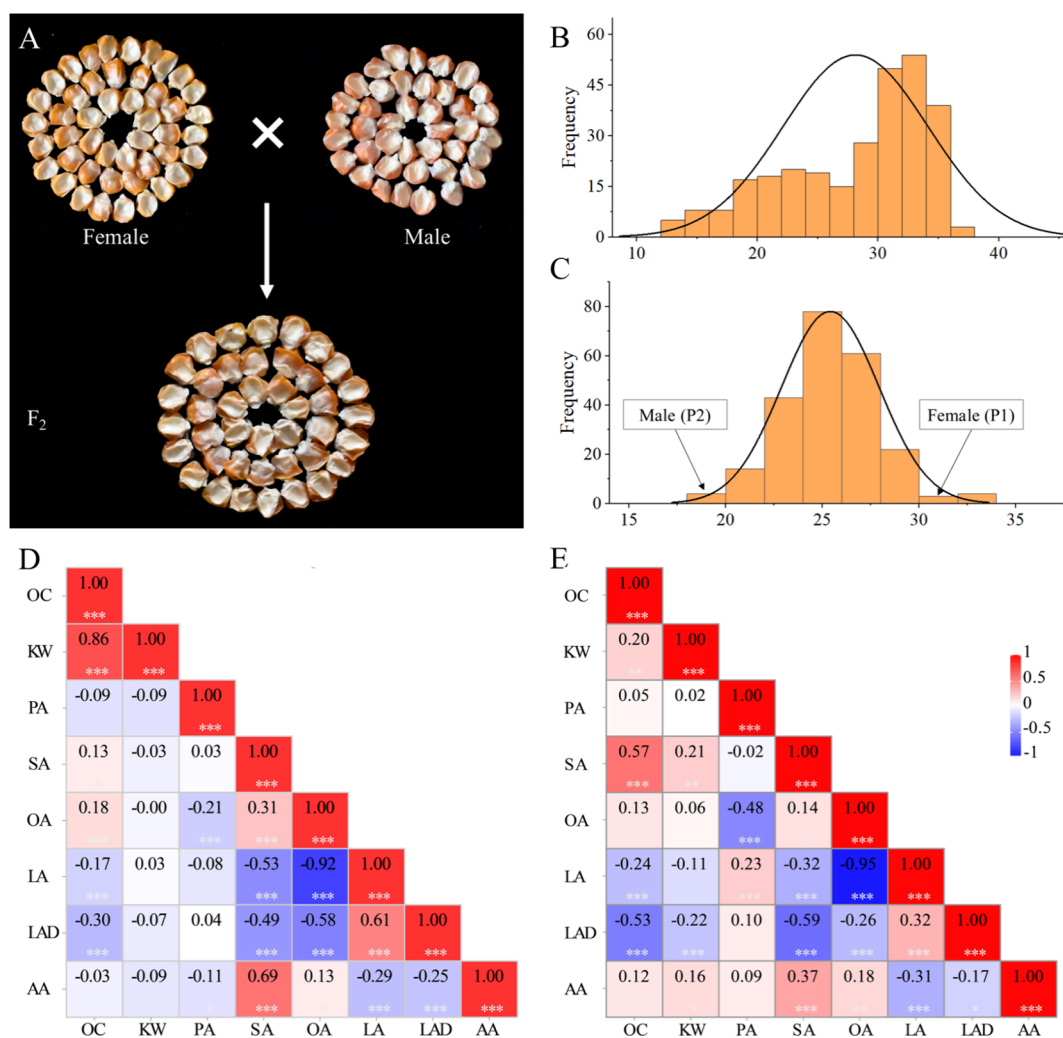


Figure 1. Correlation between the phenotypic traits and oil characteristics of maize kernels. (A) Kernel phenotypes of F₂ population; (B) oil content of inbred line population; (C) oil content of F₂ population. Correlation between the phenotypic traits and oil characteristics of (D) inbred line population and (E) F₂ population. Numbers in correlation panels D and E indicate the correlation between the two traits. ***, **, and * denote significances at the 0.1%, 1%, and 5% levels of probability, respectively.

al. demonstrated the critical role of the *DGAT1-2* gene, cloned through positional cloning, in regulating maize oil and oleic acid content.¹⁰ Overexpression of *ZmLEC1* and *ZmWRI1* can enhance maize seed oil production.⁹ *ZmSAD1* is involved in regulating fatty acid composition and the ratio of saturated to unsaturated fatty acids in maize kernels.¹¹ *ZmGE2* affects oil content by modulating the ratio of embryo to endosperm in kernels.¹² Genes *su1* and *sh2* decrease maize endosperm size by suppressing starch synthesis, thereby regulating kernel oil content.^{1,13} In conclusion, elucidating the molecular mechanisms underlying high-oil content in maize kernels is instrumental in expediting the breeding of new high-oil maize varieties.

The formation of high-oil populations and their regulatory mechanisms have always been the focus of researchers. QTL mapping analysis is an important method for elucidating the genes controlling the quantitative traits. To date, multiple molecular markers controlling maize kernel oil content have been detected on different chromosomes.¹⁴ For instance, Alrefai et al. studied the IHO (Illinois High Oil) and ILO (Illinois Low Oil) lines and identified 69 gene loci controlling maize oil content.¹⁵ Mangolin et al. constructed an F₂

population from two maize inbred lines and detected 13 QTLs by QTL mapping of maize oil content.¹⁶ Yang et al. identified 11 QTLs linked to maize oil content using an F3:4 population resulting from a cross between Zheng58 and B73.¹⁷ Genome-wide association analysis is also a powerful approach for dissecting candidate genes linked to complex traits. In recent years, GWAS has been extensively utilized in genetic studies in maize traits.^{18,19} For instance, Liu et al. identified 4 SNPs significantly associated with maize starch content, uncovering 77 candidate genes related to starch synthesis in a natural population of 263 maize inbred lines.²⁰ Similarly, Zhu et al. found 20 SNPs linked to ear traits in 292 maize inbred lines.²¹ Li et al. conducted a genome-wide association study on maize oil content with 1.03 million SNPs, identifying 74 QTLs significantly associated with maize oil content or fatty acid composition.²² In addition, by integrating GWAS and RNA-seq data, researchers can conduct a more comprehensive investigation into the relationship between genotype and phenotype, aiding in the discovery of potential genetic regulatory mechanisms and novel functional genes. For example, through GWAS and RNA-seq, *Zm00001d038250* and *Zm00001d001960* were identified as candidate genes that

influence alkali tolerance in maize seedlings.²³ Currently, the identification of candidate genes related to maize traits predominantly relies on single-omics or dual-omics approaches, with limited studies integrating three or more omics analyses. To delve deeper into the molecular mechanisms of maize oil biosynthesis, this study employs a multiomics approach that combines GWAS, QTL mapping, and RNA-seq to identify key genes involved in maize oil synthesis.

In this research, phenotypic and genotypic data from 278 inbred lines and 229 individuals from F₂ populations were utilized to achieve three main objectives: (1) Detect trait-associated SNPs via two GWAS models; (2) identify stable QTLs for oil traits using an F₂ bin map; (3) uncover DEGs via RNA-seq of high- and low-oil lines; (4) integrate GWAS, QTL mapping, and RNA-seq to prioritize candidate genes. The primary goal of this study is to enhance our comprehension of the intricate genetic makeup of oil content in high-oil maize kernels and to establish a groundwork for molecular marker-assisted breeding and the identification of key genes that regulate oil content in high-oil maize.

2. MATERIALS AND METHODS

2.1. Plant Materials and Field Experiments

An association panel was constructed using 278 microendosperm maize inbred lines from the breeding program at Guangxi University. The F₂ population consisted of 229 lines derived from two maize inbred lines (2-3-4 (female, high-oil line with elevated oleic acid content, 33.12% oil content) × 2-1-2 (male, low-oil line with reduced oleic acid content, 18.82% oil content)) (Figure 1A–C). Both the association panel and F₂ population were planted at the research field of Guangxi University in both 2021 and 2022. Field experiments were conducted using a randomized complete block design with three replications of single-row plots, and standard cultivation practices were employed for plant management.

2.2. Phenotypic Data Collection and Analysis

The association panel and F₂ population were simultaneously harvested upon reaching full maturity. Maize oil content was measured directly using an MQC23 NMR analyzer (Bruker Optics) after thorough drying.²⁴ Palmitic acid (PA), stearic acid (SA), oleic acid (OA), linoleic acid (LA), linolenic acid (LAD), and arachidic acid (AA) were quantified through gas chromatography in 2022.²⁵ For a single environment, the average of three replicates for each inbred line was calculated as the phenotypic data for descriptive statistics. The data analysis module of Excel was utilized to determine the standard deviation (SD), coefficient of variance (CV), and minimum and maximum values. Pearson correlation coefficients (r^2) were calculated for each pair of traits using the “Hmisc” and “rrplot” packages in R software.^{26,27} Additionally, phenotypic variation was assessed¹² and broad-sense heritability of OC and KW was computed.

2.3. Genotyping and Constructing a Genetic Linkage Map

Genomic DNA was extracted from the leaf tissues of maize using the CTAB method, and DNA library sequencing was performed on Huada Gene Technology Co. Ltd. (Shenzhen, China).²⁸ Raw reads were quality-controlled using FastQCv0.12.1 and then filtered with the fastp software to obtain clean reads. We used an integrated commercial suite GTXv2.2 (<http://www.gtxlab.com/product/cat>) based on the Zm-B73-REFERENCE-NAM-5.0 (<https://www.maizgdb.org/genome/>) reference genome for variant detection. Raw variants were filtered using bcftools software v1.15.1 (<http://www.htslib.org/>) with the parameters (QD < 2||FS > 60||MQ < 40||MQRankSum < -12.5||SOR > 3||ReadPosRankSum < -8), then filtered using vcftools software v0.1.16 (<https://vcftools.sourceforge.net/>) based on minor allele frequency (MAF) > 0.05 and missing rate < 0.2.

2.4. Population Structure and GWAS Analysis

We pruned the SNPs in high levels of pairwise LD using PLINK v1.9 with the parameter to perform principal component analysis (PCA) and fastStructure (<https://rajanil.github.io/fastStructure/>) analysis. The “chooseK.py” subroutine suggested the optimal number of clusters as $K = 9$. Then, the Q matrix for $K = 9$ was inputted into the mixed linear model. The kinship matrix was generated by the GEMMA software. Using GEMMA software, two linear mixed models (a K model and a $Q+K$ model) that could effectively control population structure and kinship were selected for GWAS analysis.²⁹ The Manhattan and quantile–quantile ($Q-Q$) plots were generated using the qqman and ggplot2 tools, respectively.^{30,31} The PopLDdecay software was used to visualize the local GWAS regions and linkage disequilibrium patterns. Meanwhile, the Bonferroni correction P -value ($\alpha = 0.05$) threshold was also verified.³²

2.5. QTL Linkage Mapping Using the F₂ Population

A genetic map was constructed based on the genotype data from an F₂ maize population. First, the clean SNPs were filtered according to the parental genotypes (nonhomozygous and identical genotype loci in parents were removed), yielding 1450801 SNPs. Then, the progeny genotypes were corrected using a sliding window approach³³ (window size 20kb) and merged into bin markers (<https://github.com/lileiting/BinMarkers-v2>). Based on 31733 bin markers, the genetic map was constructed using lemap3 (<https://sourceforge.net/projects/lep-map3/files/>) software (lodLimit = 30), resulting in 10 linkage groups. Subsequently, bin markers within the same linkage group that had identical genetic distances and genotypes were filtered, leaving 2916 high-quality bin markers for QTL analysis. We used the R/qtl package for QTL analysis.³⁴

2.6. RNA-seq

We conducted an RNA-seq profiling of two maize parental lines utilized for QTL mapping. The female parent 2-3-4 (P1) is a high-oil line with an elevated oleic acid content, while the male parent 2-1-2 (P2) is a low-oil line with a low oleic acid content. Specifically, we collected kernels from the middle part of maize ears 30 days after pollination. RNA-seq libraries were prepared from these maize kernel tissues using the Illumina NEBNext Ultra RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and sequenced on the BGISEQ platform (Shenzhen Huada Gene Science and Technology Service Co., Shenzhen, China). We constructed a reference genome index for the maize genome Zm-B73-REFERENCE-NAM-5.0 (<https://www.maizgdb.org/genome/>) using HISAT2-build. Cleaned data were then aligned to this reference genome using HISAT2 v2.1.0 (<https://daehwankimlab.github.io/hisat2/>). Transcript or gene expression levels were quantified based on TPM values. Differential gene expression analysis was conducted using DESeq2 (<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>), with significantly differentially expressed genes identified by a log₂ Fold Change absolute value ≥ 1 and a false discovery rate (FDR) P -value < 0.05.

2.7. Identification and Annotation of Candidate Genes

We conducted a gene ontology (GO) enrichment analysis to identify significantly enriched GO terms among the background genes. The GO terms were classified and organized by using publicly available databases. Specifically, we employed a hypergeometric test in conjunction with phytozome to determine the significantly enriched GO terms. These terms were categorized into three main classes: “biological process”, “cellular component”, and “molecular function”. Furthermore, to gain deeper insight into the biological functions of differentially expressed genes (DEGs), we performed an enrichment analysis using the KEGG pathway database. Gene annotations were obtained from the MaizeGDB and NCBI databases. Finally, heatmaps were generated using the online ChiPlot tool (<https://www.chiplot.online/>) to visualize the data.

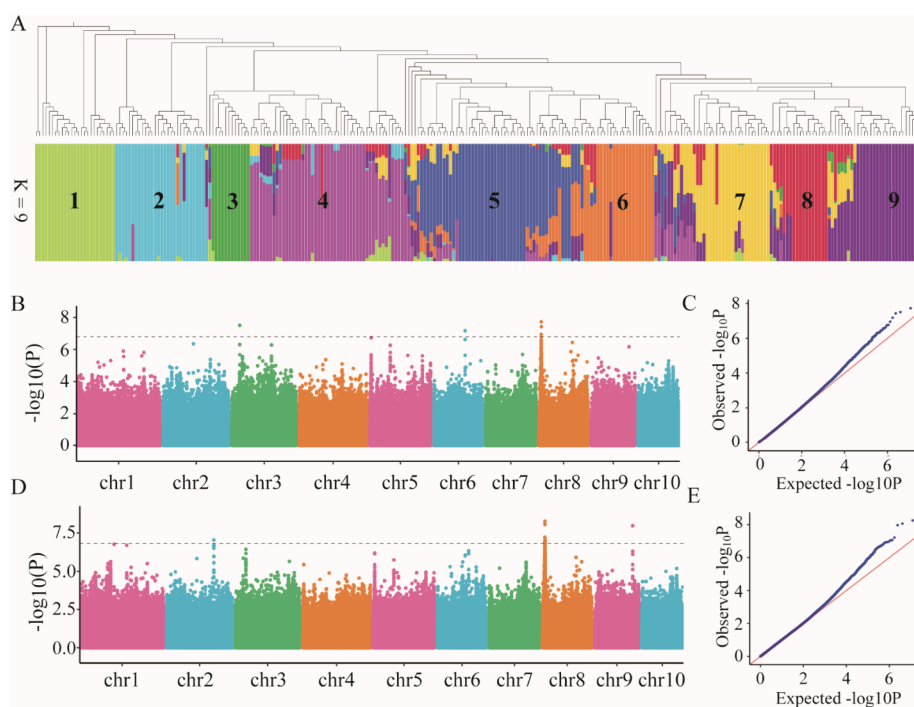


Figure 2. Manhattan and quantile–quantile (Q–Q) plots of GWAS for maize kernel oil content. (A) Population structure of the inbred line population; (B, C) Manhattan and Q–Q plots of the K model; (D, E) Manhattan and Q–Q plots of the Q+K model. The x-axis displays the chromosome label, and the y-axis displays $-\log_{10}(P\text{-value})$. The solid gray line shows significant associations between SNPs and the color dots above the threshold values indicate the significant SNPs for maize kernel oil content traits in B and D figure.

3. RESULTS

3.1. Phenotypic Analyses and Heritability Estimation of Maize Kernel Oil Traits

This study measured eight phenotypic kernel weight and oil traits in 278 maize inbreds and 229 F_2 individuals, including OC, KW, PA, SA, OA, LA, LAD, and AA. Means of three biological replicates were calculated for all phenotypes, and two years evaluations were performed for OC and KW. The results indicated significant variation in these eight traits among maize inbreds, which exhibited an approximately normal distribution alongside notable correlation (Supporting Information Figure S1A). Phenotypes of the F_2 individuals demonstrated a similar pattern (Figure S1B). The coefficients of variation (CV) were calculated for maize inbred lines and F_2 lines separately. In the inbred line population, SA exhibited the highest CV (25.99%), while OA had the lowest CV (9.29%) (Table S1). In contrast, in the F_2 population, KW had the highest CV (13.92%) in 2021 and PA had the lowest CV (5.37%) (Table S2). Overall, complex traits such as OC and KW displayed a high CV in both populations. The CV for the same traits were slightly higher in the inbred line population compared to the F_2 population, with some differences observed between different years' observation. Correlation analysis revealed that OC was positively correlated with KW and OA ($P < 0.01$) (Figure 1D,E). OA was positively correlated with SA but negatively correlated with PA, LA, and LAD ($P < 0.05$). The results suggested that the composition of the maize oil trait is complex, with interactions among different oil components. In the inbred population, the broad-sense heritability (hB^2) values for OC and KW both exceeded 80% (Table S1). In the F_2 individuals' population, OC had an hB^2 above 50%, whereas the hB^2 of KW surpassed 90% (Table S2).

These findings suggest that their phenotypic variations were mainly contributed by genetic factors.

3.2. GWAS Analysis of Maize Kernel Oil Traits

In this study, the inbred line population and F_2 population were genotyped using whole-genome resequencing technology with an average coverage depth of $15\times$ per sample. The Q20 percentage (proportion of nucleotides with a quality value greater than 20 within the read) was greater than 95%, and the Q30 percentage was greater than 90%. Consequently, 10352895 and 2840095 SNPs were identified for the inbred line population and F_2 population, respectively. The data set was filtered with missing rates < 0.2 and retained SNPs with a minor allele frequency (MAF) > 0.05 . A total of 6053624 high-quality SNPs were retained for the inbred line population for association analysis. In the inbred population, the number of SNPs per chromosome ranged from 462335 (chromosome 10) to 902294 (chromosome 1), with an average SNP number of 605362 (Figure S2A). Population structure analysis and phylogenetic analysis of the inbred line population revealed that the 278 accessions were divided into nine subgroups, each containing 25, 29, 12, 52, 52, 25, 29, 19, and 35 samples, respectively (Figure 2A).

Association analysis was conducted using the genotypic data and eight kernel weight and oil traits in the inbred population with two models, the K model and the Q+K model, respectively. A total of 338 SNPs were identified, which were either (1) detected by both K and Q+K models or (2) associated with ≥ 2 oil traits. Given the phenotypic variation explained (PVE) by SNPs identified through the two methods, the mean value was employed to represent the proportion of PVEs. Individual SNPs explained 9.41–13.88% of the variance, and all SNPs associated with a single trait together explained 33.58–88.25% of the variance. SNPs associated with maize

kernel oil content had relatively high PVEs, and a total of 16 SNPs associated with oil content were identified, which were located on chromosome 8 (Chr8; 14 SNPs), chromosome 6 (Chr6; 1 SNP), and chromosome 3 (Chr3; 1 SNPs) (Figure 2). Regions of 100 Kb surrounding the associated SNP were considered as QTLs controlling traits, and this resulted in 132 QTLs. Eight QTL loci controlling oil content were identified in phenotypes of two years, except for *qOC8.5* and *qOC9.1*, which were detected in the K model and Q+K model, respectively (Table S3).

3.3. QTLs and Candidate Genes Associated with Kernel Weight and Multiple Oil Traits

Some SNPs were associated with kernel weight and multiple oil traits, and their corresponding QTLs might have pleiotropic effects on multiple traits. This study identified 10 such QTLs (Table S5, Figure S9) and 82 candidate genes in these QTLs considering the linkage disequilibrium (LD) decay distance (Figure S7A). Among these genes, 70 were annotated and 12 were unknown (Table S5, Figure S7B). Through gene ontology (GO) functional enrichment analysis, candidate genes for OA and LA were enriched in short-chain fatty acid esterase activity, acylglycerol lipase activity, and phenylalanine-tRNA aminoacylation (Figure S8A, Table S6). Previous reports show that short-chain carboxylic esterase catalyzes the synthesis of short-chain carboxylic esters, substrates for fatty acid synthesis, and acylglycerol lipase catalyzes the hydrolysis of glycerides into glycerol and free fatty acids.³⁵ Furthermore, the candidate genes associated with multiple traits were also enriched in transaminase activity, the nucleotide–sugar metabolism process, and acyl-[acyl-carrier-protein] desaturase activity, which directly or indirectly impacted fatty acid synthesis or metabolism (Table S6). KEGG analysis showed candidate genes were mainly enriched in fatty acid biosynthesis, biosynthesis of unsaturated fatty acids, and amino sugar and nucleotide sugar metabolism (Figure S8B), and fatty acid biosynthesis was the core pathway (Figure S8C). The candidate genes might play crucial roles in fatty acid metabolism. For instance, stearoyl-acyl carrier protein desaturase 1 (*SAD*, *Zm00001eb143300*) catalyzes the synthesis of monounsaturated oleic acid or palmitoleic acid in the plastids, serving as a key enzyme controlling the ratio of saturated to unsaturated fatty acids in plant cells.^{11,36} While the *SAD* gene showed no differential expression across various tissues in maize, its expression level in the high-oil maize Zheng58 endosperm was significantly higher than that in regular maize B73 (Figure S12A,B), strengthening the reliability of our results.

3.4. QTL Mapping of Kernel Weight and Oil Traits Using the Maize F₂ Population

Using 1450801 high-quality SNPs generated in the maize F₂ population including 229 lines, a highly dense genetic map was constructed, spanning 1174.38 cM with an average marker interval of 0.258 cM. Chromosome 3 (Chr3) exhibited the highest marker number (294116 SNPs, 20.27% of the total mapped SNPs), while Chromosome 8 (Chr8) had the lowest number (48606 SNPs, 3.35% of the total), with an average marker density of 679 markers per megabase (Mb) (Figure S2B). With Bonferroni correction as the threshold, QTL mapping of kernel weight and oil traits identified seven significant QTLs. QTLs with a PVE exceeding 10% were considered as major QTLs, and all 7 QTLs were classified as major QTLs. In multiple years' phenotypes, three QTLs

controlling oil content were located at 57.13, 43.17, and 56.92 cM on Chromosome 5 with PVE of 15.65, 28.13, and 26.07%, respectively. The same QTL for oil content was consistently identified in the phenotypes of 2021, 2022, and BLUP (Figure 3A,B). Additionally, a significant QTL was identified for each

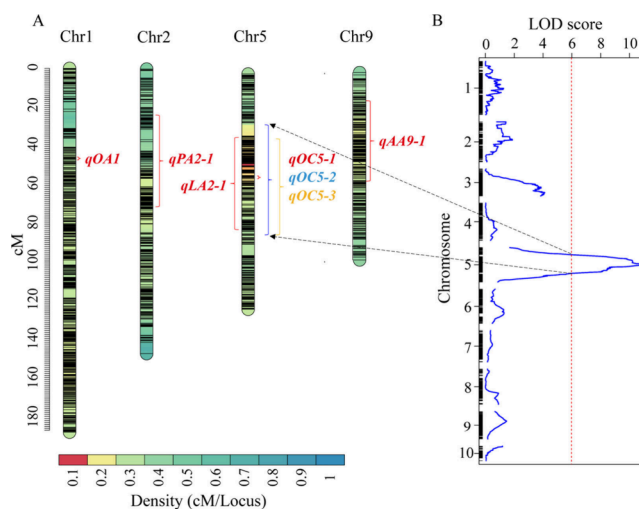


Figure 3. QTLs detected for maize kernel oil traits using the F₂ population. (A) QTL loci in maize genetic maps; (B) QTL detected for the oil content BLUP value.

of the phenotypes, PA, OA, LAD, and AA (Figure 3A). The total PVEs of related traits in multiple years' traits and the BLUP were estimated (Table S7), and PVE for the oil content in 2022 was the largest (28.13%).

3.5. Transcriptome Analysis

To delve deeper into the molecular mechanisms of oil accumulation in microendosperm maize kernels, we selected the parental lines with different oil contents (Figure 4A) and performed high-throughput transcriptomics sequencing to identify differentially expressed genes (DEGs) in kernels. In the RNA-seq analysis, a total of 77.32 Gb of data was obtained from six samples. The Q20 and Q30 values were higher than 96.36 and 89.59%, suggesting that the sequencing data met the quality requirements of subsequent analyses (Table S8). A total of 821 DEGs were identified across the two parental lines. A total of 234 genes were found to be upregulated in P1 relative to P2, while 625 genes exhibited downregulation in P1 compared to that in P2 (Figure 4B, Table S9). To further study the metabolic pathways and other pathways related to oil accumulation in microendosperm maize kernels, KEGG analysis was conducted. Upregulated genes were enriched in eight pathways for metabolism, while downregulated genes were enriched in four pathways for metabolism. The most representative pathways of upregulated genes were linoleic acid metabolism, biosynthesis of various plant secondary metabolites, sphingolipid metabolism, starch and sucrose metabolism, and biosynthesis of secondary metabolites. In addition, the pathways of downregulated genes were other glycan degradation, benzoxazinoid biosynthesis, nitrogen metabolism, and glyoxylate and dicarboxylate metabolism (Figure 4C). These results indicated that lipid metabolism, carbohydrate metabolism and global and overview maps play important roles in maize oil accumulation.

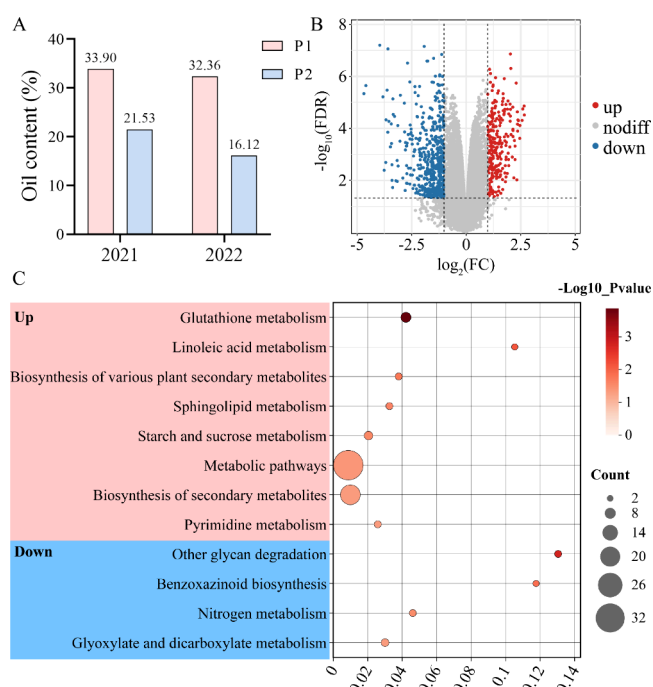


Figure 4. Transcriptome analysis and KEGG pathway enrichment analysis. (A) Oil content in 2021 and 2022; (B) volcano plot of differentially expressed genes (DEGs) between P1 and P2; (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway classification enrichment analysis of DEGs. P1, high oil content and high oleic acid varieties; P2, low oil content and low oleic acid variety.

3.6. Candidate Genes Identified by GWAS, QTL Mapping and RNA-seq

Integrating the results of GWAS, QTL mapping, and RNA-seq can enhance the accuracy of gene mining. We performed a joint analysis of candidate genes associated with the oil content identified by GWAS, QTL mapping, and RNA-seq. The results of the Venn diagram analysis revealed three candidate genes identified by all three methods (GWAS, QTL mapping, and RNA-seq). Additionally, 12 candidate genes were identified by GWAS and RNA-seq, 190 candidate genes were identified by GWAS and QTL mapping, and 171 candidate genes were identified by QTL mapping and RNA-seq (Figure S11, Table S10). Then, we further analyzed three candidate genes identified by the three methods, and the results showed that fatty aldehyde dehydrogenase (*FALDH*, *Zm00001eb248870*, chr5:194359742-194364076) played a central function in the metabolism of lipid intermediates, oxidizing fatty aldehydes into the corresponding fatty acids (Figure 5A,B). The expression level of *FALDH* was significantly higher in P1 (OC: 33.12%) than in P2 (OC: 18.82%) ($P < 0.001$) (Figure 5C). In addition, *FALDH* genes expression level was significantly higher in the high-oil Zheng58 maize endosperm compared to the normal B73 maize (Figure S12C,D). The second gene, *Zm00001eb215810*, is classified as a 7-methyl-GTP pyrophosphatase. Its expression level was significantly higher in P1 (female) than in P2 (male) ($P < 0.01$) (Figure S13A, Table S11). The third gene, *Zm00001eb221670*, encodes a protein of unknown function, which requires further investigation. The expression level of *Zm00001eb221670* was significantly higher in P1 (female) than in P2 (male) ($P < 0.05$) (Figure S13B, Table S11). Furthermore, we identified two distinct colocalized regions associated with OC and AA on

chromosomes 5 and 9 by both GWAS and QTL mapping. A total of 13 candidate genes were colocalized, among which *Zm00001eb385870* encodes fatty acid elongases (*FAE2*), participating in the conversion of fatty acid molecules with shorter chain lengths than those with longer chain lengths. Other genes' potential roles in maize oil traits needed further research (Table 1, Figures S5D and S14). Finally, integration of findings from GWAS with the RNA-seq data set allowed us to identify a total of 15 candidate genes (Figure S11). Among them, the gene *Zm00001eb146460* encodes the TRANS-PARENT TESTA 12 (TT12) protein, which is known to play a critical role in seed development. The *Zm00001eb154570* gene encodes an FCS-like zinc finger protein that participates in the regulation of plant growth and development. In addition, some genes with function not annotated are highly expressed in P2, which needs further research. In summary, by interrogation of GWAS, QTL mapping, and RNA-seq, we identified a substantial number of candidate genes, offering new gene resources for future investigation and application to improve maize oil content.

4. DISCUSSION

With the advancement of sequencing technologies, GWAS has become a powerful tool for identifying genotype–phenotype associations. The release of the maize B73 reference genome has significantly promoted maize population genetic research into the genomic era. To date, GWAS has been applied in agronomic traits (plant height, yield, hundred kernel weight, etc.), biotic stresses (maize rust, *Fusarium* ear rot, maize rootworm, etc.), and abiotic stresses (salt tolerance, drought tolerance, heat tolerance, etc.) in maize, revealing a plethora of key genes.^{42,22,43} Therefore, using GWAS to identify genes associated with maize kernel oil traits and elucidate their genetic basis will facilitate advancements in research and breeding high-oil maize varieties.

While GWAS offers insights in uncovering genetic architecture of complex agronomic traits, it also poses a high risk of false positives due to the influence of population structures.^{44,45} In crops with complex population structures such as maize, wheat, and potatoes, the K model and Q+K model within mixed linear models were effective approaches for conducting marker–phenotype association analyses.^{46,47} The K model, with faster computation time, effectively reduces false associations influenced by kinship relationships resulting from inbreeding, selection, and drift factors among maize populations.⁴⁸ In comparison, the Q+K model, including population structure and kinship relationships as covariates, decreases the occurrence of false positives. Nevertheless, in cases where complex traits were associated with population structures exhibiting extensive genetic variation, correction of the Q+K model may be excessively stringent, likely resulting in false negatives.^{49,50} Therefore, when dealing with complex phenotypes with diverse genetic backgrounds in GWAS, selecting appropriate models is crucial for identifying reliable marker–phenotype associations and genes. For example, Yu et al. assessed effectiveness of four models in association analysis, focusing on fitness and error rate control.⁵¹ They found that the Q+K model outperformed others for flowering time and ear height. However, for ear diameter, the K model exhibited slightly higher efficacy than the Q+K model, consistent with findings in other crops.⁴⁸ Current study suggested that false associations can occur using a single model. For instance, false negative associations were observed in the Q+K model for

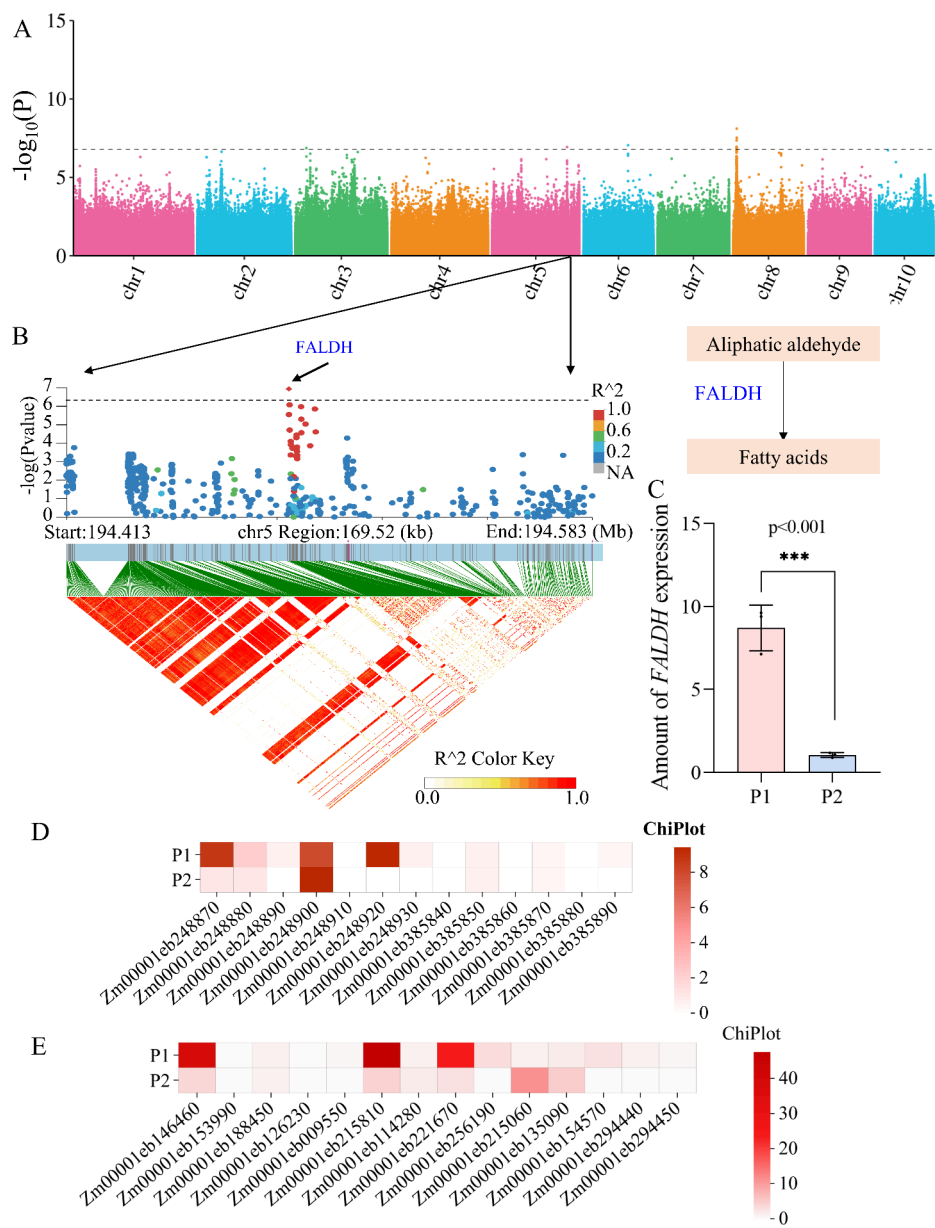


Figure 5. Colocalization of QTLs through integration of GWAS and QTL mapping. (A) Manhattan of OC in 2021; (B) chromosome localization of OC in 2021; (C) amount of *FALDH* expression; (D) expression levels of genes identified by both GWAS and QTL in RNA-seq data; (E) expression levels of genes identified by both GWAS and RNA-seq. Gene *FALDH* was colocated on chromosome 5 by GWAS and QTL mapping. P1, high oil content and high oleic acid varieties; P2, low oil content and low oleic acid variety.

Table 1. Colocalized Genes by GWAS Analysis and QTL Mapping

Trait	Gene Name	Position	Homologue	Gene Function
OC	<i>Zm00001eb248870</i>	chr5:194358242-194365576	<i>FALDH</i>	Fatty aldehyde dehydrogenase ³⁷
OC	<i>Zm00001eb248880</i>	chr5:194367754-194369423	<i>APK1A</i>	Protein kinase <i>APK1A</i> ³⁸
OC	<i>Zm00001eb248890</i>	chr5:194366403-194370940		Predicted
OC	<i>Zm00001eb248900</i>	chr5:194370961-194377904	<i>ZmThx20</i>	Plays an essential role in maize seed development ³⁹
OC	<i>Zm00001eb248910</i>	chr5:194503681-194504344		Unknown
OC	<i>Zm00001eb248920</i>	chr5:194582627-194586282		Unknown
OC	<i>Zm00001eb248930</i>	chr5:194617190-194618421	<i>ATHB-6</i>	Homeobox-leucine zipper protein <i>ATHB-6</i> ⁴⁰
AA	<i>Zm00001eb385840</i>	chr9:91854468-91856369	<i>NicC</i>	A flavin-dependent decarboxylative hydroxylase
AA	<i>Zm00001eb385850</i>	chr9:91983651-91988573	<i>FAR1</i>	Plant growth and development ⁶³
AA	<i>Zm00001eb385860</i>	chr9:91989066-91994169		SOUL family protein
AA	<i>Zm00001eb385870</i>	chr9:92186031-92192076	<i>FAE2</i>	Fatty acid elongase 2 ⁸
AA	<i>Zm00001eb385880</i>	chr9:92191081-92198233		Serine–threonine kinase receptor-associated protein ⁴¹
AA	<i>Zm00001eb385890</i>	chr9:92253387-92253893		Unknown

maize OC in 2022 and the K model for KW in 2021 (Figures S3B and S4A), while false positive associations were detected in the K model for oleic acid (Figure S5C). To reduce false associations, this study employed the two models in GWAS to identify shared SNP loci for further investigation.

Previous research indicates that depending solely on either method, like GWAS, QTL mapping, or RNA-seq, has significant drawbacks in identifying genes controlling complex traits in maize.⁵² QTL mapping based on biparental populations often shows lower precision in gene localization due to the limited number of recombinations existing in mapping populations. In addition, constructing a high genetic linkage map poses challenges as these populations only display allelic diversity between parental lines, and resolution relies heavily on the number of recombination markers.⁵³ Therefore, a limited number and uneven distribution of molecular markers have hindered efficiency and accuracy of QTL mapping, and QTLs detected typically explain only a small portion of genetic variations for target traits.⁵⁴ In contrast, GWAS using a diversity panel with accumulated recombination has a higher resolution to identify QTLs with small effects.²⁸ Therefore, combining GWAS, QTL mapping and RNA-seq can enhance the power of detection, compensating for each other's limitations and yielding more reliable QTLs and candidate genes identified.⁵⁵ For example, through a joint analysis of GWAS, QTL mapping and RNA-seq, Guo et al. dissected 7 major effect QTLs and 69 key genes associated with seed vigor in rice (*Oryza sativa* L.).⁵⁵ In another study, Zhang et al. identified the crucial gene, *Zm00001d016656*, encoding serine/threonine protein kinase using QTL mapping and GWAS.¹⁸ In our research, we identified two shared QTL loci associated with oleic or linoleic acid, containing 13 key genes involved in maize oil synthesis metabolism. Some of these genes were validated through RNA sequencing (RNA-seq), demonstrating significant differences in expression. The functions of some of these genes were validated in previous studies.^{38,40} The results highlighted the power of integration of GWAS, QTL mapping, and RNA-seq in revealing essential genes controlling maize kernel oil components.

Maize oil content is a complex quantitative trait controlled by multiple genes, with its synthesis influenced by factors like grain weight, oleic acid, and linoleic acid.⁵⁶ Significant correlations between maize oil content and grain weight, oleic acid, and linoleic acid levels in this study highlighted the impact of these traits on maize oil synthesis, consistent with previous research.⁸ Currently, numerous genes related to maize kernel oil traits were discovered and characterized by various technologies and methods. For instance, *FAD2* and *FAD3* genes were found to be involved in unsaturated fatty acid synthesis in lipid pathway.^{7,57} *ZmLEC1* is a key transcription factor in the lipid biosynthesis pathway, regulating expression of multiple key genes in the lipid metabolism pathway, thereby affecting lipid synthesis. Overexpression of *ZmWRI1* led to a significant increase of seed oil content up to 46%, with no observed impacts on seed germination or plant growth.⁹ In addition, other important genes have been mined, such as *ZmDGAT1*, *ZmGPAT*, *ACP*, and *COPII*.^{22,58,59} Despite this, these candidate genes of maize oil traits still require in-depth investigation, and many critical genes need to be discovered. In this study, *ZmSAD1* (Stearoyl-Acyl Carrier Protein Desaturase 1, *Zm00001eb143300*) was found associated with stearic acid, linoleic acid, and oleic acid content (Figure 1A). The *SAD1* gene facilitates conversion of stearoyl acyl carrier protein

(ACP) to oleoyl-ACP, an essential precursor for polyunsaturated fatty acid biosynthesis.⁶⁰ Additionally, the *SAD* enzyme is pivotal in initiating double bond formation in saturated fatty acids, thereby regulating the balance between saturated and unsaturated fatty acids.⁶¹ Previous research has demonstrated the significance of *SAD1* in converting saturated fatty acids to oleic acid in crops like maize, *Arabidopsis*, and Ginkgo.¹¹ The expression level of the *ZmSAD1* gene in P1 was higher than that in P2, however, the difference was not significant. In addition, this study identified the gene *fatty aldehyde dehydrogenase*, *FALDH*, encoding a type of aldehyde dehydrogenase ALDH to convert acetaldehyde to acetic acid, which is catalyzed by ACS to form acetyl-coenzyme A in fatty acid synthesis pathway.³⁷ Although *FALDH* has been extensively studied in animals, there is limited research on this gene in plants, especially in monocotyledonous plants. *FALDH* catalyzes the direct oxidation of saturated and α,β -unsaturated aldehydes to their corresponding fatty acids, thereby detoxifying the endosperm and replenishing the acyl-CoA pool necessary for the reassembly of triacylglycerol (TAG). This dual functionality elucidates the enzyme's markedly higher activity in the high-oil parent P1 (33.12% oil content) compared to the low-oil parent P2 (18.82% oil content). Interestingly, trihelix transcription factor *ZmThx20*, adjacent to *FALDH* in the same QTL region in this study, was found to play a significant role in regulating grain development and responding to environmental stresses.³⁹ Additionally, two key genes, *FAE2* and *FAR*, have been linked to arachidonic acid. The *FAE2* gene is a homologous allele of *GRMZM2G022558* identified in QTL mapping of oil-related traits in the KUI3/SC55 RIL population using high-density SNP markers, which are directly involved in the biosynthesis of very long chain fatty acids.⁶² *FAE2* elongates C18 fatty acids to C20 (AA), which is consistent with its localization in *qAA9* and the positive correlation observed between the *FAE2* expression and AA content. *FAR* is crucial in regulating plant growth and development processes.⁶³ Moreover, some genes identified were predicted with unknown functions, which necessitates further research for gene function. Given that RNA sequencing was conducted exclusively on grains during the rapid oil accumulation period, some genes may not have shown significant differences in expression levels. This limitation is likely due to the specific sampling period selected for this study. Future research should involve sampling kernels at 10 (seed formation stage), 20 (rapid oil accumulation phase), 40 (dough stage), and 50 (physiological maturity stage) days after pollination (DAP) to capture the dynamic expression of both early (e.g., ACCase, MCMT) and late (e.g., DGAT, GPAT) oil biosynthesis genes and to validate candidate genes through CRISPR/Cas9 knockout techniques.

In conclusion, the combination of GWAS, QTL mapping, and RNA-seq offers a robust method for dissection genes controlling main oil traits, providing valuable gene resources for research and breeding high-oil maize.

■ ASSOCIATED CONTENT

Data Availability Statement

The data that support the findings of this study have been deposited into China National GeneBank DataBase (CNCBdb) with accession number CNP0005817. The reviewer link is available at http://db.cngb.org/cnsa/variant/CNP0005817_dabf8618/reviewlink/.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jafc.5c13129>.

(Figure S1) Phenotypic population data; (Figure S2) number of SNP density; (Figures S3–S6) Manhattan and quantile–quantile plots; (Figure S7) LD decay distance and candidate gene annotation status; (Figure S8) candidate genes analysis; (Figure S9) phenotype LOD plots; (Figure S10) chained genes atlas; (Figure S11) candidate genes Venn diagram of candidate genes identified through GWAS, QTL mapping, and RNA-seq; (Figures S12 and S13) gene expressions; (Figure S14) arachidic acid localized Manhattan diagram; (Tables S1 and S2) phenotypic data; (Tables S3–S5, S7) loci and loci information; (Table S6) GO functional enrichment analysis; (Table S8) list of RNA-Seq data quality; (Table S9) differentially expressed genes; (Table S10) candidate gene; (Table S11) screening of GWAS, QTL mapping, and RNA-seq jointly (PDF)

AUTHOR INFORMATION

Corresponding Authors

Xiaoqin Hao – College of Agriculture, Guangxi University, Nanning 530005, China; Email: haoxq74@163.com

Xingtian Zhang – National Key Laboratory for Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangzhou 518120, China; orcid.org/0000-0002-5207-0882; Email: zhangxingtian@caas.cn

Xiping Yang – Guangxi Sugarcane Bio-breeding Laboratory, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530005, China; College of Agriculture, Guangxi University, Nanning 530005, China; orcid.org/0000-0002-3452-211X; Phone: (+86) 135 5810 9155; Email: xipinyang@gxu.edu.cn

Authors

Lijun Zhang – Guangxi Sugarcane Bio-breeding Laboratory, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530005, China

Yuhan Song – National Key Laboratory for Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangzhou 518120, China

Meiyan Chen – Guangxi Sugarcane Bio-breeding Laboratory, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530005, China

Chunyu Hu – Guangxi Sugarcane Bio-breeding Laboratory, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530005, China

Wanru Wang – Guangxi Sugarcane Bio-breeding Laboratory, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530005, China

Xunbo Zhou – College of Agriculture, Guangxi University, Nanning 530005, China; orcid.org/0000-0002-3191-7191

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jafc.5c13129>

Author Contributions

L.Z. and Y.S. contributed equally to this work. X.Y., X.H., and Xi.Z. planned and supervised the research. L.Z. and Y.S. analyzed the data and completed the manuscript. L.Z., M.C., C.H., and W.W. conducted phenotype investigations and participated in data analysis. X.Y., X.H., Xi.Z., and Xu.Z. contributed to the manuscript revisions. All authors have read and approved the final version of the manuscript.

Notes

Ethical Approval: The authors declare that no ethical standards have been violated during the study. The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Science and Technology Major Project of Guangxi (GK AD25069107), the National Natural Science Foundation of China (31960429), and the “One Hundred Person” Project of Guangxi Province.

REFERENCES

- (1) Luo, M. J.; Lu, B. S.; Shi, Y. X.; Zhao, Y. X.; Liu, J. L.; Zhang, C. Y.; Wang, Y. D.; Liu, H.; Shi, Y. M.; Fan, Y. L.; Xu, L.; Wang, R.; Zhao, J. Genetic basis of the oil biosynthesis in ultra-high-oil maize grains with an oil content exceeding 20. *Front. Plant Sci.* **2023**, *14*, No. 1168216.
- (2) Ndlovu, N.; Kachapur, R. M.; Beyene, Y.; Das, B.; Ogugo, V.; Makumbi, D.; Spillane, C.; McKeown, P. C.; Prasanna, B. M.; Gowda, M. Linkage mapping and genomic prediction of grain quality traits in tropical maize (*Zea mays* L.). *Front. Genet.* **2024**, *15*, No. 1353289.
- (3) Shi, M.; Sun, J.; Jiang, F.; Shaw, R. K.; Ijaz, B.; Fan, X. Mining of oil content genes in recombinant maize inbred lines with introgression from temperate and tropical germplasm. *Int. J. Mol. Sci.* **2024**, *25* (19), 10813.
- (4) Wei, M.; Fu, J.; Li, X.; Wang, Y.; Li, Y. Influence of dent corn genetic backgrounds on QTL detection for plant-height traits and their relationships in high-oil maize. *J. Appl. Genet.* **2009**, *50* (3), 225–234.
- (5) Shi, T. Q.; Wang, L. R.; Zhang, Z. X.; Sun, X. M.; Huang, H. Stresses as first-line tools for enhancing lipid and carotenoid production in microalgae. *Front. Bioeng. Biotechnol.* **2020**, *8*, 610.
- (6) Li-Beisson, Y.; Shorrosh, B.; Beisson, F.; Andersson, M. X.; Arondel, V.; Bates, P. D.; Baud, S.; Bird, D.; Debono, A.; Durrett, T. P.; Franke, R. B.; Graham, I. A.; Katayama, K.; Kelly, A. A.; Larson, T.; Markham, J. E.; Miquel, M.; Molina, I.; Nishida, I.; Rowland, O.; Samuels, L.; Schmid, K. M.; Wada, H.; Welti, R.; Xu, C.; Zallot, R.; Ohlrogge, J. Acyl-lipid metabolism. *Arabidopsis Book* **2013**, *11*, No. e0161.
- (7) Beló, A.; Zheng, P.; Luck, S.; Shen, B.; Meyer, D. J.; Li, B.; Tingey, S.; Rafalski, A. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. *Mol. Genet. Genomics* **2008**, *279* (1), 1–10.
- (8) Fang, H.; Fu, X.; Ge, H.; Zhang, A.; Shan, T.; Wang, Y.; Li, P.; Wang, B. Genetic basis of maize kernel oil-related traits revealed by high-density SNP markers in a recombinant inbred line population. *BMC Plant Biol.* **2021**, *21* (1), 344.
- (9) Shen, B.; Allen, W. B.; Zheng, P.; Li, C.; Glassman, K.; Ranch, J.; Nubel, D.; Tarczynski, M. C. Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. *Plant Physiol.* **2010**, *153* (3), 980–987.

- (10) Zheng, P.; Allen, W. B.; Roesler, K.; Williams, M. E.; Zhang, S.; Li, J.; Glassman, K.; Ranch, J.; Nubel, D.; Solawetz, W.; Bhattaramakki, D.; Llaca, V.; Deschamps, S.; Zhong, G.-Y.; Tarczynski, M. C.; Shen, B. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* **2008**, *40* (3), 367–372.
- (11) Du, H.; Huang, M.; Hu, J.; Li, J. Modification of the fatty acid composition in *Arabidopsis* and maize seeds using a *stearoyl-acyl carrier protein desaturase-1* (*ZmSAD1*) gene. *BMC Plant Biol.* **2016**, *16* (1), 137.
- (12) Zhang, P.; Allen, W. B.; Nagasawa, N.; Ching, A. S.; Heppard, E. P.; Li, H.; Hao, X.; Li, X.; Yang, X.; Yan, J.; Nagato, Y.; Sakai, H.; Shen, B.; Li, J. A transposable element insertion within *ZmGE2* gene is associated with increase in embryo to endosperm ratio in maize. *Theor. Appl. Genet.* **2012**, *125* (7), 1463–1471.
- (13) Hu, Y.; Colantonio, V.; Muller, B. S. F.; Leach, K. A.; Nanni, A.; Finegan, C.; Wang, B.; Baseggio, M.; Newton, C. J.; Juhl, E. M.; Hislop, L.; Gonzalez, J. M.; Rios, E. F.; Hannah, L. C.; Swarts, K.; Gore, M. A.; Hennen-Bierwagen, T. A.; Myers, A. M.; Settles, A. M.; Tracy, W. F.; Resende, M. F. R. Genome assembly and population genomic analysis provide insights into the evolution of modern sweet corn. *Nat. Commun.* **2021**, *12* (1), 1227.
- (14) Guo, Y.; Yang, X.; Chander, S.; Yan, J.; Zhang, J.; Song, T.; Li, J. Identification of unconditional and conditional QTL for oil, protein and starch content in maize. *Crop Journal* **2013**, *1* (1), 34–42.
- (15) Alrefai, R.; Berke, T. G.; Rocheford, T. R. Quantitative trait locus analysis of fatty acid concentrations in maize. *Genome* **1995**, *38* (5), 894–901.
- (16) Mangolin, C.; de Souza, C., Jr; Garcia, A. A. F.; Garcia, A.; Sibov, S.; De Souza, A. Mapping QTLs for kernel oil content in a tropical maize population. *Euphytica* **2004**, *137* (2), 251–259.
- (17) Yang, Z.; Li, X.; Zhang, N.; Zhang, Y. N.; Jiang, H. W.; Gao, J.; Kuai, B. K.; Ding, Y. L.; Huang, X. Q. Detection of quantitative trait loci for kernel oil and protein concentration in a B73 and Zheng58 maize cross. *Genet. Mol. Res.* **2016**, *15*, (3) DOI: 10.4238/gmr.15038951.
- (18) Zhang, X.; Ren, Z.; Luo, B.; Zhong, H.; Ma, P.; Zhang, H.; Hu, H.; Wang, Y.; Zhang, H.; Liu, D.; Wu, L.; Nie, Z.; Zhu, Y.; He, W.; Zhang, S.; Su, S.; Shen, Y.; Gao, S. Genetic architecture of maize yield traits dissected by QTL mapping and GWAS in maize. *Crop Journal* **2022**, *10* (2), 436–446.
- (19) Luo, X.; Wang, B.; Gao, S.; Zhang, F.; Terzaghi, W.; Dai, M. Genome-wide association study dissects the genetic bases of salt tolerance in maize seedlings. *J. Integr. Plant Biol.* **2019**, *61* (6), 658–674.
- (20) Liu, N.; Xue, Y.; Guo, Z.; Li, W.; Tang, J. Genome-wide association study identifies candidate genes for starch content regulation in maize kernels. *Front. Plant Sci.* **2016**, *7*, 1046.
- (21) Zhu, X. M.; Shao, X. Y.; Pei, Y. H.; Guo, X. M.; Li, J.; Song, X. Y.; Zhao, M. A. Genetic diversity and genome-wide association study of major ear quantitative traits using high-density SNPs in maize. *Front. Plant Sci.* **2018**, *9*, 966.
- (22) Li, H.; Peng, Z.; Yang, X.; Wang, W.; Fu, J.; Wang, J.; Han, Y.; Chai, Y.; Guo, T.; Yang, N.; Liu, J.; Warburton, M. L.; Cheng, Y.; Hao, X.; Zhang, P.; Zhao, J.; Liu, Y.; Wang, G.; Li, J.; Yan, J. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **2013**, *45* (1), 43–50.
- (23) Li, C.; Jia, Y.; Zhou, R.; Liu, L.; Cao, M.; Zhou, Y.; Wang, Z.; Di, H. GWAS and RNA-seq analysis uncover candidate genes associated with alkaline stress tolerance in maize (*Zea mays* L.) seedlings. *Front. Plant Sci.* **2022**, *13*, No. 963874.
- (24) Wan, X.; Liu, Q.; Dong, B.; Vibhakaran Pillai, S.; Huang, F. H.; Singh, S. P.; Zhou, X. R. Molecular and biochemical analysis of the castor caruncle reveals a set of unique genes involved in oil accumulation in non-seed tissues. *Biotechnol. Biofuels* **2019**, *12*, 158.
- (25) Furuhashi, M.; Hiramitsu, S.; Mita, T.; Omori, A.; Fuseya, T.; Ishimura, S.; Watanabe, Y.; Hoshina, K.; Matsumoto, M.; Tanaka, M.; Moniwa, N.; Yoshida, H.; Ishii, J.; Miura, T. Reduction of circulating FABP4 level by treatment with omega-3 fatty acid ethyl esters. *Lipids Health Dis* **2016**, *15*, 5.
- (26) Ruan, Y.; Yu, B.; Knox, R. E.; Zhang, W.; Singh, A. K.; Cuthbert, R.; Fobert, P.; DePauw, R.; Berraies, S.; Sharpe, A.; Fu, B. X.; Sangha, J. Conditional mapping identified quantitative trait loci for grain protein concentration expressing independently of grain yield in Canadian Durum wheat. *Front. Plant Sci.* **2021**, *12*, No. 642955.
- (27) Sandhu, K. S.; You, F. M.; Conner, R. L.; Balasubramanian, P. M.; Hou, A. Genetic analysis and QTL mapping of the seed hardness trait in a black common bean (*Phaseolus vulgaris*) recombinant inbred line (RIL) population. *Mol. Breed* **2018**, *38* (3), 34.
- (28) Zhang, L.; Wang, W.; Chen, M.; Wang, Z.; Yang, H.; Duan, Y.; Hu, C.; Tang, S.; Liu, P.; Zhou, X.; Yang, X. Deciphering the genetic basis of sugar cane (*Saccharum spontaneum* L.) root system and related traits under nitrogen stress through the integration of genome-wide association studies and RNA-seq. *J. Agric. Food Chem.* **2025**, *73* (5), 3246–3256.
- (29) Zhou, X.; Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **2012**, *44* (7), 821–824.
- (30) Turner, S. D. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv Preprint*, 2014. <https://doi.org/10.1101/005165> (2024-8-15).
- (31) Zhang, B.; Huang, Y.; Zhang, L.; Zhou, Z.; Zhou, S.; Duan, W.; Yang, C.; Gao, Y.; Li, S.; Chen, M.; Li, Y.; Yang, X.; Zhang, G.; Huang, D. Genome-wide association study unravels quantitative trait loci and genes associated with yield-related traits in sugarcane. *J. Agric. Food Chem.* **2023**, *71* (44), 16815–16826.
- (32) Zhang, C.; Dong, S. S.; Xu, J. Y.; He, W. M.; Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35* (10), 1786–1788.
- (33) Qin, M. F.; Li, L. T.; Singh, J.; Sun, M. Y.; Bai, B.; Li, S. W.; Ni, J. P.; Zhang, J. Y.; Zhang, X.; Wei, W. L. Construction of a high-density bin-map and identification of fruit quality-related quantitative trait loci and functional genes in pear. *Horticulture Research* **2022**, *9*, uhac141.
- (34) Broman, K. W.; Wu, H.; Sen, S.; Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **2003**, *19* (7), 889–890.
- (35) González-Thuillier, I.; Pellny, T. K.; Tosi, P.; Mitchell, R. A. C.; Haslam, R.; Shewry, P. R. Accumulation and deposition of triacylglycerols in the starchy endosperm of wheat grain. *J. Cereal Sci.* **2021**, *98*, No. 103167.
- (36) Kibe, M.; Nyaga, C.; Nair, S. K.; Beyene, Y.; Das, B.; M, S. L.; Bright, J. M.; Makumbi, D.; Kinyua, J.; Olsen, M. S. Combination of linkage mapping, GWAS, and GP to dissect the genetic basis of common rust resistance in tropical maize germplasm. *Int. J. Mol. Sci.* **2020**, *21* (18), 6518.
- (37) Wei, Y.; Lin, M.; Oliver, D. J.; Schnable, P. S. The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of *Arabidopsis*. *BMC Biochem* **2009**, *10*, 7.
- (38) Hirayama, T.; Oka, A. Novel protein kinase of *Arabidopsis thaliana* (APK1) that phosphorylates tyrosine, serine and threonine. *Plant Mol. Biol.* **1992**, *20* (4), 653–662.
- (39) Li, P.; Li, Z.; Xie, G.; Zhang, J. Trihelix transcription factor *ZmThx20* is required for kernel development in maize. *Int. J. Mol. Sci.* **2021**, *22* (22), 12137.
- (40) Jiao, P.; Jiang, Z.; Wei, X.; Liu, S.; Qu, J.; Guan, S.; Ma, Y. Overexpression of the homeobox-leucine zipper protein ATHB-6 improves the drought tolerance of maize (*Zea mays* L.). *Plant Sci.* **2022**, *316*, No. 111159.
- (41) Jia, H.; Li, M.; Li, W.; Liu, L.; Jian, Y.; Yang, Z.; Shen, X.; Ning, Q.; Du, Y.; Zhao, R.; Jackson, D.; Yang, X.; Zhang, Z. A serine/threonine protein kinase encoding gene *KERNEL NUMBER PER ROW6* regulates maize grain yield. *Nat. Commun.* **2020**, *11* (1), 988.
- (42) Li, C.; Song, Y.; Zhu, Y.; Cao, M.; Han, X.; Fan, J.; Lu, Z.; Xu, Y.; Zhou, Y.; Zeng, X.; Zhang, Z.; Dong, L.; Sun, D.; Wang, Z.; Di, H. GWAS analysis reveals candidate genes associated with density tolerance (ear leaf structure) in maize (*Zea mays* L.). *Journal of Integrative Agriculture* **2025**, *24* (6), 2046–2062.

- (43) Wu, X.; Feng, H.; Wu, D.; Yan, S.; Zhang, P.; Wang, W.; Zhang, J.; Ye, J.; Dai, G.; Fan, Y.; Li, W.; Song, B.; Geng, Z.; Yang, W.; Chen, G.; Qin, F.; Terzaghi, W.; Stitzer, M.; Li, L.; Xiong, L.; Yan, J.; Buckler, E.; Yang, W.; Dai, M. Using high-throughput multiple optical phenotyping to decipher the genetic architecture of maize drought tolerance. *Genome Biol.* **2021**, *22* (1), 185.
- (44) Li, C.; Huang, Y.; Huang, R.; Wu, Y.; Wang, W. The genetic architecture of amylose biosynthesis in maize kernel. *Plant Biotechnol J.* **2018**, *16* (2), 688–695.
- (45) Borrego-Benjumea, A.; Carter, A.; Zhu, M.; Tucker, J. R.; Zhou, M.; Badea, A. Genome-wide association study of waterlogging tolerance in barley (*Hordeum vulgare* L.) under controlled field conditions. *Front Plant Sci.* **2021**, *12*, No. 711654.
- (46) Duan, H.; Li, J.; Sun, Y.; Xiong, X.; Sun, L.; Li, W.; Gao, J.; Li, N.; Zhang, J.; Cui, J.; Fu, Z.; Zhang, X.; Tang, J. Candidate loci for leaf angle in maize revealed by a combination of genome-wide association study and meta-analysis. *Front Genet.* **2022**, *13*, No. 1004211.
- (47) Sharma, S. K.; MacKenzie, K.; McLean, K.; Dale, F.; Daniels, S.; Bryan, G. J. Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3 (Bethesda)* **2018**, *8* (10), 3185–3202.
- (48) Pasam, R. K.; Sharma, R.; Malosetti, M.; van Eeuwijk, F. A.; Haseneyer, G.; Kilian, B.; Graner, A. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* **2012**, *12*, 16.
- (49) Kaler, A. S.; Gillman, J. D.; Beissinger, T.; Purcell, L. C. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front Plant Sci.* **2020**, *10*, 1794.
- (50) Zhao, Z.; Zhang, H.; Fu, Z.; Chen, H.; Lin, Y.; Yan, P.; Li, W.; Xie, H.; Guo, Z.; Zhang, X.; Tang, J. Genetic-based dissection of arsenic accumulation in maize using a genome-wide association analysis method. *Plant Biotechnol. J.* **2018**, *16* (5), 1085–1093.
- (51) Yu, J.; Pressoir, G.; Briggs, W. H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J. F.; McMullen, M. D.; Gaut, B. S.; Nielsen, D. M.; Holland, J. B.; Kresovich, S.; Buckler, E. S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38* (2), 203–208.
- (52) Wang, Y.; Ran, F.; Yin, X.; Jiang, F.; Bi, Y.; Shaw, R. K.; Fan, X. Genome-wide association studies on the kernel row number in a multi-parent maize population. *Int. J. Mol. Sci.* **2024**, *25* (6), 3377.
- (53) Yang, Y.; Zhao, T.; Wang, F.; Liu, L.; Liu, B.; Zhang, K.; Qin, J.; Yang, C.; Qiao, Y. Identification of candidate genes for soybean seed coat-related traits using QTL mapping and GWAS. *Front Plant Sci.* **2023**, *14*, No. 1190503.
- (54) Hu, D.; Kan, G.; Hu, W.; Li, Y.; Hao, D.; Li, X.; Yang, H.; Yang, Z.; He, X.; Huang, F.; Yu, D. Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. *Front. Plant Sci.* **2019**, *10*, 811.
- (55) Guo, T.; Yang, J.; Li, D.; Sun, K.; Luo, L.; Xiao, W.; Wang, J.; Liu, Y.; Wang, S.; Wang, H.; Chen, Z. Integrating GWAS, QTL, mapping and RNA-seq to identify candidate genes for seed vigor in rice (*Oryza sativa* L.). *Molecular Breeding* **2019**, *39* (6), 87.
- (56) Zhong, Y.; Zhao, Y.; Wang, Y.; Niu, J.; Sun, Z.; Chen, J.; Luan, M. Transcriptome analysis and GC-MS profiling of key fatty acid biosynthesis genes in *Akebia trifoliata* (Thunb.) Koidz Seeds. *Biology (Basel)* **2022**, *11* (6), 855.
- (57) Pushkova, E. N.; Povkhova, L. V.; Dvorianinova, E. M.; Novakovskiy, R. O.; Rozhmina, T. A.; Gryzunov, A. A.; Sigova, E. A.; Zhernova, D. A.; Borkhert, E. V.; Turba, A. A.; Yablokov, A. G.; Bolsheva, N. L.; Dmitriev, A. A.; Melnikova, N. V. Expression of *FAD* and *SAD* genes in developing seeds of flax varieties under different growth conditions. *Plants (Basel)* **2024**, *13* (7), 956.
- (58) Yan, B.; Xu, X.; Gu, Y.; Zhao, Y.; Zhao, X.; He, L.; Zhao, C.; Li, Z.; Xu, J. Genome-wide characterization and expression profiling of diacylglycerol acyltransferase genes from maize. *Genome* **2018**, *61* (10), 735–743.
- (59) Zhu, T.; Wu, S.; Zhang, D.; Li, Z.; Xie, K.; An, X.; Ma, B.; Hou, Q.; Dong, Z.; Tian, Y.; Li, J.; Wan, X. Genome-wide analysis of maize GPAT gene family and cytological characterization and breeding application of ZmMs33/ZmGPAT6 gene. *Theor. Appl. Genet.* **2019**, *132* (7), 2137–2154.
- (60) Mustiga, G. M.; Morrissey, J.; Stack, J. C.; DuVal, A.; Royaert, S.; Jansen, J.; Bizzotto, C.; Villela-Dias, C.; Mei, L.; Cahoon, E. B.; Seguíne, E.; Marelli, J. P.; Motamayor, J. C. Identification of climate and genetic factors that control fat content and fatty acid composition of *Theobroma cacao* L. Beans. *Front. Plant Sci.* **2019**, *10*, 1159.
- (61) Lindqvist, Y.; Huang, W.; Schneider, G.; Shanklin, J. Crystal structure of delta9 stearoyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *Embo j* **1996**, *15* (16), 4081–4092.
- (62) Haslam, T. M.; Kunst, L. Extending the story of very-long-chain fatty acid elongation. *Plant Sci.* **2013**, *210*, 93–107.
- (63) Ma, L.; Li, G. FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in *Arabidopsis* growth and development. *Front Plant Sci.* **2018**, *9*, 692.