

Identification and expression characterization of a WRKY transcription factor affecting squalene synthesis in *Camellia oleifera*

Qinhui Du

Central South University of Forestry and Technology

Luyao Ge

Central South University of Forestry and Technology

YanLing Zeng

zengyanling110@126.com

Central South University of Forestry and Technology

Aori Li

Central South University of Forestry and Technology

Ziyan Zhu

Central South University of Forestry and Technology

Xiaofeng Tan

Central South University of Forestry and Technology

Research Article

Keywords: *Camellia oleifera*, Squalene, CoWRKY transcription factor, expression

Posted Date: April 9th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4204992/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

WRKY-like transcriptional regulators are widely involved in physiological processes such as growth and development, metabolic regulation and environmental response. In this study, we obtained six CoWRKY transcription factors by yeast one-hybrid screening library with reference to the *Camellia oleifera* genome sequence, using squalene synthase gene (CoSQS) as bait. AOS (Antibody Optimization System) analysis showed that CoWRKY15 had the highest interactions with a confidence level of 0.9026. Bioinformatics analysis showed that CoWRKY15 encodes 346 amino acid residues, was a basic hydrophilic protein, did not contain a transmembrane region, contained one WRKY conserved structural domain and one C2H2 zinc finger structural domain. and belonged to class 2 of the WRKY gene family, and had the closest genetic distance of 0.5564 to the homologous protein of *Panax quinquefolius* PqWRKY1. The results of prokaryotic expression showed that the CoWRK15 protein with a size of 38.3 kD was successfully induced by adding a final concentration of 0.5 mM ITPG for 4 h at 37°C. The results of subcellular localization showed that CoWRKY15 functioned in the nucleus. The results of CoWRKY15 promoter analysis showed that 8 out of 14 cis-elements with annotatable functions were related to the light response, indicating that the expression of CoWRKY15 was strongly affected by light. The correlation analysis of CoWRKY15 expression and squalene content in *Camellia oleifera* seed kernels treated under different light quality conditions showed a significant positive correlation.

1. Introduction

Camellia oleifera (Oil Tea), in a broad sense, refers to more than 60 shrubs of the genus *Camellia* (Theaceae) whose seed kernels produce high-quality edible oils^[1]. *Camellia* oil is not only rich in unsaturated fatty acids, but also rich in squalene and other bioactive substances^[2]. Squalene is an open-chain triterpene organic substance. Squalene is involved in tumour suppression and immune enhancement, has antibacterial and antiviral activity and can be used as a drug carrier and adjuvant for vaccines^[3]. Although some studies have shown that *C.oleifera* seed kernels are rich in squalene, the mechanism involved is not clear. Kim et al^[4]. proved that squalene synthase (SQS) regulation is the key factor affecting squalene accumulation as assessed through the SQS deletion yeast *erg9* mutant functional complementarity test^[5]. However, under the same promoter, the ability of SQS mutant to resume squalene synthesis was significantly different after the complementary expression, which indicated the importance of transcription factors.

WRKY belongs to the WRKY-GCM1 zinc finger transcription factor superfamily, which is involved in plant growth, development, and physiological responses^[6, 7]. WRKY proteins in the plant kingdom form a large family of transcription factors that can be divided into at least three groups (Group , Group , and Group) according to the characteristics of the structural domain^[8]. The members of the WRKY family Group contains two WRKYGQK domains, and the zinc fingerprint pattern is C-X4-5-C-X22-23H-X-H, such as AtWRKY33 in *Arabidopsis thaliana*^[9]. The members of the WRKY family Group contains only one WRKYGQK domain, and the zinc fingerprint pattern is the same as that of the first category, such as

GmWRYK13^[10] in soybean. The members of the WRKY family Group contains only one WRKYGQK domain, but the zinc fingerprint is the same. The structural amino acids changed to C-X7-C-X22-23-H-X-C, such as JrWRKY4^[11] of *walnuts*.

Yongzhen Sun^[12] showed that the expression levels of HMGR, FPS2, SQS1, and SQE2 were higher than those of the control group after transformation of PqWRKY1 gene from *P. quinquefolius* into *Arabidopsis thaliana*. WRKY binds to the W-box in the promoter region of key genes involved in triterpenoid synthesis in *Arabidopsis thaliana* and PqWRKY1 was characterised by structural domains typical of WRKY family class 2 members. Our group compared the changes of squalene content in *C. oleifera* seed kernels during different developmental periods, constructed the transcriptome database, and based on the analysis that the CoSQS promoter contains a W-box cis-element, and retrieved the CoWRKY1 gene in the transcriptome database, which may interoperate with CoSQS, for yeast one-hybrid validation, we preliminarily demonstrated that squalene synthesis in *C. oleifera* seed kernels is affected by the CoWRKY transcription factor. However, the CoWRKY transcription factor is a large gene family^[13], and it is not clear whether there are other members that also combine with CoSQS to affect squalene synthesis in *C. oleifera* seed kernels.

In this study, we combined *C. oleifera* genomic data, constructed *C. oleifera* seed kernel yeast libraries, used the CoSQS promoter as bait to regulate transcription factors that may interact with each other, and through AOS (Antibody Optimization System) analysis, identified the CoWRKY transcription gene CoWRKY15 with the highest interaction confidence, based on which we carried out bioinformatics analyses, prokaryotic expression and subcellular localization, and clarified the exogenous factor that most affects the expression of the CoWRKY genes through analysing their promoter sequences. By analysing its promoter sequence, the exogenous factor that most affects the expression of CoWRKY gene was clarified, and the squalene content was associated with the regulation of the expression of this CoWRKY through this exogenous factor, which further proved the role of CoWRKY transcription factor in affecting the squalene content of *C. oleifera* seed kernel, and provided scientific basis for revealing the function of WRKY transcription factor in *C. oleifera* seed kernel, and regulating the biosynthesis of squalene in plants by means of genetic engineering.

2. Materials and Methods

2.1 Plant materials

In the gene cloning experiment, the seed kernels of national certified cultivar 'HS' were collected from the base of *C. oleifera* in Dongcheng, Hunan, Wangcheng County, Changsha Province. The fluorescent quantitative test materials were *C. oleifera* fruit under different light exposure conditions (Fig. 1) .

Light not only affected the growth of *C. oleifera*, but also had an impact on the metabolism of squalene^[14, 15]. Light exposure treatments included blue light (B), white light (W), red light (R), a red–blue light combination (RB; red–blue ratio: 1:1), and sunlight (CK). LED lights were used as light sources for

treatment except for CK, including blue (450–480 nm), red (610–640 nm), and white (420–720 nm) LED lights. Spectral characteristics of the different light quality treatments were measured with a HopooColor OHSP-350SF Spectral Color Luminance Meter (HopooColor, Hangzhou, China). The light source was placed above the three year-old *C.oleifera*, and the height of the LED light source was adjusted to ensure that the photon flux density (PPFD) of each treated plant canopy is $200 \pm 10 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (Fig. 1)

2.2 Real-Time PCR Quantification of Gene Expression

The relative expression of CoWRKY15 gene was analysed by real-time PCR using the cDNA of the test material as template, and CoEF1 α gene was used as the internal reference gene, and three parallel replicates of the assay were done for each sample. The PCR primers for CoWRKY and CoEF1 α /CoEF1 α R were F 5'-CTTTCGTCGTTGTCGGCGTTTC-3', R 5'-ACAATGGCAGCGGTCGGAAG-3', F 5'-CAAAGAAGGGTGCCAAGTGA-3' and R 5'-ACCAAACAACCGACCTACGA-3', respectively. The PCR reaction system was: 2 $\times$ ChamQSYBR qPCR Master Mix 10.0 μL , forward and reverse primers (10 Pmol.L⁻¹) 0.4 μL each, 50 $\times$ ROX Reference Dye1 0.4 μL , cDNA 0.5 μL , and ultrapure water 8.3 μL . The PCR amplification conditions were as follows: 95 $^{\circ}\text{C}$ pre-denaturation for 30 sec, 95 $^{\circ}\text{C}$ denaturation for 10 sec, and 60 $^{\circ}\text{C}$ annealing for 30 sec for a total of 40 cycles; dissolution curve 95 $^{\circ}\text{C}$ denaturation for 30 sec, and dissolution curve 95 $^{\circ}\text{C}$ denaturation for 30 sec for 40 cycles.

2.3 Yeast one-hybrid screening

RNA was extracted from *Camellia oleifera* using the Trizol method using (reverse transcription-polymerase chain reaction, RT-PCR) to obtain cDNA, the reaction system are: 2 μL of 5 $\times$ PrimeScriptRT Mix, 2 μg of Total RNA, RNase Free ddH₂O to 10 μL . The reaction conditions were as follows: 25 $^{\circ}\text{C}$, 10 min; 42 $^{\circ}\text{C}$, 30 min; 85 $^{\circ}\text{C}$, 5min; 4 $^{\circ}\text{C}$, 3min.

After purification, it was homologously recombined with linear pGADT7 and transformed into *Escherichia coli* to construct the *C.oleifera* kernels cDNA library. The 1600bp upstream sequence of the *C.oleifera* squalene synthase gene (*CoSQS*) promoter was introduced into pHIS2 and transformed into Y187 strain for background screening (Fig. 2), to determine the subsequent screening concentration using 75mM 3AT. The primers used for plasmid construction were pHIS2-CoSQS F: 5'-TACGACTCACTATAGGGCGAATTCGTTATGTTAG-3' and pHIS2-CoSQS R: 5'-ATCGATTGCGAACGCGTGAGCTCTTTTCTCTCT-3'. Using Y187 yeast containing the correct pHIS2-CoSQS- promoter bait plasmid as the receptor strain to prepare the competent cell, the cDNA library plasmid was transferred into it and 100 μL was applied to SD-LTH/75 mmol/L 3-AT culture. Nutrient-based plates were incubated inversion at 30 $^{\circ}\text{C}$ for 4–6 days. The positive transformants grown on the SD-TLH + 75 mM 3AT screening plate were scraped and subjected to NGS screening and sequencing.

2.4 Prokaryotic expression

Using the *C.oleifer* cDNA of the test material as template, pET32a-CoWRKY15F 5'-GCCATGGCTGATATCGGATCCATGATGGCCGTCGAGCTCGTGATG-3' and pET32a-CoWRKY15R: 5'-

CTCGAGTGC GGCCGCAAGCTTTCAAGAAGACTCTAAGATAAG-3' as primers, the products obtained from PCR amplification were recovered and used in the ClonExpress II One Step Cloning Kit (Vazyme) to pET32a using homologous recombination. The 10 µL reaction system was: 5×CEII Buffer 2.0 µL, Exuase II 1.0 µL, recovered product (20–100 ng) 2.0 µL, linearized vector 4.0 µL, and PCR-Grade Water 1.0 µL. The reaction conditions were: 37 °C for 30 min; 4 °C for 5 min.

Single colonies of positive and transformed empty carriers were selected and cultured 3 mL liquid LB medium containing ampicillin for 12 h at 37°C. Subsequently, the culture was expanded in 200 mL with 1% inoculum and induced with 0.5 mM IPTG at OD600 = 0.8, followed by culture at 37°C for 4 h. Then, the cell pellet was obtained by centrifugation for 2 min at 8000rpm and resuspended in an appropriate volume of Laemmli buffer, denatured at 100°C for 10 min, and subjected to SDS-PAGE.

2.5 subcellular localization

Using the *C. oleifer* cDNA of the test material as template, pCAMBIA1300-CoWRKY15F: 5'-gatcaagagacaggatccgATGATGGCCGTCGAGCTCGTGATG-3' and pCAMBIA1300-CoWRKY15R: 5'-tcaccatcggtgcactagtgtCAAGAAGACTCTAAGATAAG-3' as primers, the products obtained from PCR amplification were recovered and used in the ClonExpress II One Step Cloning Kit (Vazyme) to pCAMBIA1300 using homologous recombination.

Choose *Nicotiana benthamiana* leaves that are suitable for healthy, referring the live infection method of *Agrobacterium tumefaciens* by Zeng Yanling et al^[16]. *Agrobacterium tumefaciens* containing recombinant plasmids were injected into *N. benthamiana* leaves. The leaves were cultivated for 48 hours, and observed the fluorescence signal of eGFP under a laser confocal microscope excitation wavelength of 488 nm^[17].

2.6 Squalene content analysis

Camellia oil was extracted using Soxhlet extraction method^[18]. Liquid chromatography (HPLC) was used to determine the squalene content in *Camellia* oil. The detection conditions were as follows: Agilent reverse C18 column, mobile phase (methanol: acetonitrile = 35:65), sample injection volume of 20 µL, flow rate of 1 mL/min, column temperature of 40 °C, and detection wavelength of 210 nm.

2.7 Data Analysis

Interaction prediction was performed using AOS (Antibody Optimisation System), a software developed by Pronetbio Co^[19]. The physical and chemical properties and structural characteristics of CoWRKY15 were predicted by on-line software ProtParam (<https://web.expasy.org/protparam>) and TMHMM (<https://services.healthtech.dtu.dk/services/TMHMM-2.0>). CoWRKY15 promoter cis-acting elements were predicted by PLACE (<https://www.dna.affrc.go.jp/PLACE/?action=newplace>).

The phylogenetic tree was constructed and the cluster relationship was analyzed by using offline software MEGA11^[20]. Fluorescence quantitative PCR(real-time PCR) data were analyzed by Bio-Rad CFX Manager software. The correlation was analyzed by SPSS Statistics22.0 software.

3. Results

3.1 Obtaining WRKY transcription factors based on the CoSQS promoter yeast one-hybrid screen library

As the transcription factors of endogenous yeast may cause histidine excretion leakage, so it is necessary to add 3-AT to inhibit this leakage, in amino acids Auxotrophic medium SD-LTH was supplemented with 0, 5, 10, 20, 30, 40, 50, 75, 100mM, respectively mmol/L of 3-AT, which is determined by observing yeast growth. Appropriate 3-AT inhibitory concentrations. When the concentration of 3-AT is increased to 75 mmol/L, Yeast in the test group and negative control the growth of the cells is completely inhibited, The positive control grew normally, The results indicated that pHIS2-CoSQS did not occur in yeast AH109 activation, Therefore, 75 mmol/L 3-AT can be used as the lowest inhibitory concentration yeast single bulk library screening was performed.

Based on the CoSQS promoter yeast single heterozygous screening library, a total of 1632 NGS sequences were obtained, which were related to 43 metabolic pathways such as signal transduction, environmental adaptation, fatty acid synthesis, and triterpenoid skeleton synthesis, including 379 transcription factors and 394 kinases (Fig. 3a). AOS analysis revealed 6 CoWRKY transcription factors interacting with CoSQS promoter that the CoWRKY transcription factor with the highest similarity to *Camellia sinensis* WRKY15^[21] (GenBank number: XM_028196157) had the highest confidence in interacting with the CoSQS promoter sequence, which was 0.9026. The predicted results of both indicate 14 binding sites (Fig. 3b-e).

3.2 Bioinformatics analysis of CoWRKY15

The physical and chemical properties of CoWRKY15 were analyzed using on-line and off-line bioinformatics software. The results showed that the gene encoded 346 amino acids, the theoretical molecular weight (Mw) was 38297.31 Da, and the isoelectric point (pI) was 9.60, indicating that the protein was alkaline under physiological conditions. The number of negatively charged residues (Asp + Glu) was 33, the number of positively charged residues (Asp + Lys) is 49, and the molecular formula was C₁₆₃₉H₂₆₇₄N₅₀₈O₅₁₈S₁₆. The instability index (II) is 52.84 for the typical unstable protein. The total average hydrophobicity (GRAVY) of the protein was - 0.616, which was predicted hydrophilicity.

ProtParam predicted that 9.83% of CoWRKY15 protein was α -helical structure, 11.27% was β -folded, and 78.90% was irregularly coiled with no signal peptide. CoWRKY15 was predicted to function in eukaryotic nuclei with a confidence level of 43. TMHMM predicted the absence of a transmembrane region for the CoWRKY15 protein (Fig. 4a). The hydrophilic index of CoWRKY15 protein ranged from - 3.356 to 1.700, in which hydrophobic residues accounted for 67.34% of the entire amino acid residues, which was typical of hydrophobic proteins (Fig. 4b). The CoWRKY15 protein existed at one functional site, the WRKY structural domain, located at positions 268–334 Aa, with the amino acid sequence "KMSDIPPDDYSWRKYGQKPIKGSPHPRGYKCSSVRGCPARKHVERALDDPKMLIVTYEGEHNHLSL".

Judging from the conserved amino acid sequences contained and the structural features of the zinc finger, CoWRKY15 belonged to class 2 members of the WRKY family.

3.3 Phylogenetic tree analysis

Based on previous studies^[22, 23] and the results of BLAST on <http://www.ncbi.nlm.nih.gov> website, 19 plant WRKY amino acid sequences and CoWRKY15 amino acid sequences were selected for cluster analysis. The results showed that the genetic distance ranged from 0.0063 to 1.7279. CoWRKY15 had a genetic distance of 0.5564 with PqWRKY1 of *P. quinquefolius*, thereby exhibiting a close genetic association. The genetic distance between CoWRKY15 and AtWRKY2 homologous proteins of *Arabidopsis thaliana* was 1.5746 with the farthest genetic correlation (Fig. 5a). This phenomenon indicated that the WRKY gene family altered greatly during evolution. The conserved domain of the WRKY protein, was analyzed, and the AtWRKY66 of *A. thaliana* was found to belong to type of the WRKY family. It contains only one WRKYGQK domain and has a typical zinc fingerprint of C-X7-C-X22-23-H-X-C. The AtWRKY26, AtWRKY33, and AtWRKY2 of *A. thaliana* belonged to type of the WRKY family and contain two WRKYGQK domains. The zinc fingerprinting pattern was C-X4-5-C-X22-23H-X-H. The other members belonged to type of the WRKY family with only one WRKYGQK domain (Fig. 5b). The zinc fingerprinting pattern was C-X4-5-C-X22-23H-X-H. CoWRKY15 of *C. oleifera* belonged to type of the WRKY family ascribed to the typical characteristics of type members. However, the genetic distances with AtWRKY26, AtWRKY33, and AtWRKY2 of type members were closer than those of AtWRKY40, AtWRKY18, and AtWRKY60 of type members. This phenomenon might be attributed to the fact that the first type members of the WRKY family constituted the branch of origin, that differentiated into several subclasses in order to adapt to the environment. Thus, the genetic distance between different subclasses became longer.

3.4 Analysis of CoWRKY15 expression in prokaryotes

The recombinant plasmid pET32a-CoWRKY15 and empty plasmid pET32a were transformed into *E. coli* Rosseta (DE3), respectively. The expression of CoWRKY15 in *E. coli* was analyzed by SDS-PAGE after IPTG induction. ProtParam, an online software, demonstrated that CoWRKY15 gene was expressed at about 38.2 kDa. The protein expressed as a His-tag sequence at N-terminal of the expression system also induced to express about 23.8 kDa, such that the expected expression level of the fusion protein was about 60 kDa. SDS-PAGE displayed a fusion protein band of about 60 kDa, while only the labeled protein band was induced in the empty plasmid strains, and no specific protein band was detected in the non-induced transformed recombinant plasmid strains (Fig. 6). The recombinant plasmid pET32a-CoWRKY15 was successfully induced to express the CoWRKY15 protein in *E. coli*.

3.5 CoWRKY15 subcellular localization

The fusion expression vector was transformed into tobacco leaves by the tobacco injection method, and CoWRKY15 was efficiently transiently expressed in the vicinity of the injected area after 48 h of

incubation. eGFP fluorescence signals in tobacco leaves were observed under a confocal laser microscope, and the results showed that eGFP fluorescence signals completely overlapped with autofluorescence signals of the nucleus under the excitation of 543 nm light (Fig. 7), which indicated that CoWRKY15 might be located in the nucleus, which is in agreement with the results of the bioinformatics analyses.

3.6 Analysis of Cis-Acting Elements in the CoWRKY15 Promoter

Plant Care, an online promoter analysis database, analyzed the 2000 bp sequence of the upstream non-coding region of *CoWRKY15* for cis-elements (Table 1), which showed that the sequence contained 14 types of cis-acting elements with annotated functions, including eight elements involved in the light response (AE-box, AT1-motif, Box 4, G-box, L-box, LAMP-element, Sp1, Chs-CMA2a), one cis-acting element involved in abscisic acid response (ABRE), and two methyl jasmonate-responsive elements (CGTCA-motif, TGACG-motif). These results suggested that *CoWRKY15* gene expression might be regulated by photosystems and hormone-like signaling substances.

Table 1
Prediction of cis-acting elements in the CoWRKY15 promoter.

cis-regulatory element	core sequence	Position and Strand	function
A-box	CCGTCC	9(-)/679(+)	Cis-acting regulatory element
ABRE	ACGTG	235(+)	Cis-acting elements involved in the abscisic acid responsiveness
AE-box	AGAAACAA	1146(-)	Part of a module for light response
ARE	AAACCA	1524(-)	Cis-acting regulatory element essential for the anaerobic induction
AT-rich element	ATAGAAATCAA	899(-)/1284(+)	Binding site of AT-rich DNA binding protein (ATBP-1)
AT1-motif	AATTATTTTTTATT	1796(-)	Part of a light responsive module
Box4	ATTAAT	288(+)/1807(-)/1476(-) 968(+)/1618(-)	Part of a conserved DNAmodule involved in light responsiveness
CGTCA-motif	CGTCA	1308(-)	cis-acting regulatory element involved in the MeJA-responsiveness
G-box	TACGTG	234(+)	cis-acting regulatory element involved in light responsiveness
L-box	ATCCACCTAC	707(+)	Part of a light responsive element
LAMP-element	CTTTATCA	143(-)	Part of a light responsive element
Sp1	GGGCGG	728(-)/1657(-)/774(-)	Light responsive element
TGACG-motif	TGACG	1308(+)	cis-acting regulatory element involved in the MeJA-responsiveness
Chs-CMA2a	TCACTTGA	1099(-)	Part of a light responsive element

3.7 Effect of CoWRKY15 expression on squalene content under different light qualities

The *CoWRKY15* promoter sequence contained at least eight light-responsive elements (Table 1). Significant differences in the *CoWRKY15* expression and squalene content were observed after different light quality treatments (Fig. 6), and bivariate correlation analysis of the two parameters yielded a Pearson correlation of 0.798 and a significance of 0.106.

4. Discussion

Camellia oleifera is rich in squalene, that is a pivotal product of the terpene metabolic pathway. Some studies have shown that the key enzyme gene affecting squalene synthesis is SQS^[24]. A previous study by our group found that the expression of CoSQS was regulated by the transcription factor WRKY^[25]. WRKY is a transcriptional regulator occurring widely throughout the plant kingdom. It participates in regulating the biological and abiotic stresses and physiological responses of various metabolic pathways and is closely related to plant growth, development, and senescence^[26, 27]. WRKY TFs can specifically recognize the W-box, with sequence TTGAC/T within the target genes' promoter regions^[28]. Hanting Yang^[29] also found that WRKY functions as a transcriptional regulator of plant secondary metabolite synthesis. In this study, we used bioinformatics to analyze six WRKY transcription factors that might interact with *CoSQS*, among which the CoWRKY15 transcription factor, which had the highest confidence of interaction, was not the same as the CoWRKY1 transcription factor that was obtained by our group based on transcriptome analysis in the previous stage. Multiple WRKY transcription factors that might affect terpenoid metabolism had also been found in *Panax ginseng*^[30]. The synthesis of terpenoid metabolites could be induced by herbivory, wounding, light, low temperatures, and other stress conditions^[31]. It was possible that the WRKY transcription factors that bind to the CoSQS promoter were not the same under different conditions, leading to differences in the outcome of transcriptional regulation, which ultimately exhibited differences in squalene content.

According to the number of WRKY domains and the characteristics of a zinc finger structure, WRKY could be divided into Groups I, II, and III. Group II can be further divided into five sub-categories: IIa, IIb, IIc, IId, and IIe^[32]. The deduced amino acid sequence of CoWRKY15 gene of *C.oleifera* consisted of a typical WRKY domain and C2H2zinc finger structure, categorized into the second group of WRKY members. CoWRKY15 is genetically most distant from PqWRKY1, which was the 1st transcription factor gene cloned and identified in *P.quinquefolius* that might be involved in the regulation of triterpenoid saponin synthesis, and could regulate squalene synthesis in combination with the W-box structural domain of the promoter region of the SQS gene^[33]. CoWRKY15 and PqWRKY1 were clustered together and had similar structural characteristics, further indicating that CoWRKY15 had the function of transcriptional regulation of squalene synthesis in *C.oleifera*.

UV-B treatment of one-year-old potted *C. sinensis* cv. Longjing-43 differentially altered the metabolism of terpenoids with significant effects after 8 h of treatment, demonstrating the strong potential of UV-B application for flavor improvement in tea^[34]. This study used different wavelengths light to treat reproductive period *C.oleifera*, and the results showed that different wavelengths light also had a significant impact on the content of squalene in *C.oleifera* kernels. The correlation analysis between squalene content in *C. oleifera* seeds and CoW RKY15 expression under different light quality conditions showed a positive correlation between the two, but the correlation did not reach a very significant level. This may be because CoSQS directly regulated the synthesis of squalene, while CoWRKY15 affected the squalene content in *C. oleifera* kernels by regulating the expression of CoSQS. In addition, bioinformatics software predicts that CoWRKY15 was a hydrophilic non transmembrane 38.2 kD protein located in the nucleus. This study confirmed the predicted results through prokaryotic expression and subcellular

localization methods. CoWRKY15 was localized in the nucleus consistent with a location where transcription factors played a role.

This study analyzed the effect of CoWRKY15 transcription factor on the accumulation of squalene in *C. oleifera* seeds through multiple pathways. However, whether there are other members of the CoWRKY family interacting with *CoSQS*, and what factors affect interaction selection still need to be explored.

5. Conclusion

CoWRKY transcription factors were present in *C. oleifera* seed kernel in the form of a multigene family, among which CoWRKY15 interacts with CoSQS with the highest predicted confidence. CoWRKY15 belonged to the second class of members in the WRKY family, encoding 346 aa, which was a hydrophilic protein without transmembrane structure of about 38.2 kD in size, and it functioned in the nucleus. The wavelength of light had a significant effect on the squalene content of *C. oleifera* seed kernels, and CoWRKY15 transcription factor positively regulated the squalene content of *C. oleifera* seed kernels under this condition.

Declarations

Author Contributions: Y.Z. conceived and designed the experiments; Q.D., L.G., A. L. and Z.Z. performed the experiment, data collection, lab analysis, and data analysis. Q.D and L.G. wrote the paper; Y.Z. edited the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Major Program of Natural Science Foundation of Hunan Province (2021JC0007), the Natural Science Foundation of Hunan Province (2022JJ30998), and the Fundamental Research Funds of Forestry Department of Hunan Province (XLKY202221).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

Ethics approval and consent to participate: Not applicable.

References

1. LIN P, WANG K, WANG Y, et al. The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication [J]. *Genome biology*, 2022, 23(1): 14.
2. HAO L I, XUE-ZHI F, HAI-YAN Z, et al. Variation of Physicochemical Properties and Nutritional Components of Oil-tea Camellia Seeds during Riping [J]. *Forest Research*, 2014, 27(1): 86-91.
3. TATENO M, STONE B J, SRODULSKI S J, et al. Synthetic Biology-derived triterpenes as efficacious immunomodulating adjuvants [J]. *Scientific Reports*, 2020, 10(1): 17090.

4. KIM T D, HAN J Y, HUH G H, et al. Expression and Functional Characterization of Three Squalene Synthase Genes Associated with Saponin Biosynthesis in *Panax ginseng* [J]. *Plant and Cell Physiology*, 2011, 52(1): 125-37.
5. BRAMLEY P M, ELMADFA I, KAFATOS A, et al. Vitamin E (p 913-938) [J]. *journal of the science of food & agriculture*, 2010, 80(7).
6. BAKSHI M, OELMÜLLER R. WRKY transcription factors [J]. *Plant Signaling & Behavior*, 2014.
7. FENG W, XIAO-ZHU W U, LIANG L, et al. Research advances on physiological function of WRKY transcription factor in plant stress resistance [J]. *Guihaia*, 2017.
8. RUSHTON P J, SOMSSICH I E, RINGLER P, et al. WRKY transcription factors [J]. *Trends in Plant Science*, 2010, 15(5): 247-58.
9. ABUQAMAR S, CHEN X, DHAWAN R, et al. Expression profiling and mutant analysis reveals complex regulatory networks involved in *Arabidopsis* response to *Botrytis* infection [J]. *The Plant Journal*, 2006, 48(1): 28-44.
10. ZHOU Q Y, TIAN A G, ZOU H F, et al. Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants [J]. *Plant Biotechnology Journal*, 2008, 6(5): 486-503.
11. LI X U, XIN C, HAI-RONG W, et al. Cloning and Expression Analysis of WRKY4 Gene from *Juglans regia* L [J]. *Journal of Nuclear Agricultural Sciences*, 2014, 28(7): 1188-96.
12. HARA M, FURUKAWA J, SATO A, et al. *Abiotic Stress and Role of Salicylic Acid in Plants* [J]. Springer New York, 2012.
13. MUHAMMAD, WAQAS, MUHAMMAD, et al. Genome-wide identification and expression analyses of WRKY transcription factor family members from chickpea (*Cicer arietinum* L.) reveal their role in abiotic stress-responses [J]. *Genes & Genomics*, 2019.
14. BELÉN, GONZÁLEZ, AMORÓS, et al. A model study into the effects of light and temperature on the degradation of fingerprint constituents [J]. *Science & Justice*, 2014.
15. HE C, ZENG Y, FU Y, et al. Light quality affects the proliferation of in vitro cultured plantlets of *Camellia oleifera* Huajin [J]. *PeerJ*, 2020, 8(3): e10016.
16. YAN-LING Z, DANG-QUAN Z, XIAO-FENG Z, et al. Molecular Characterization and Expression Analysis of Squalene Synthetase Gene(CoSQS) from *Camellia oleifera* [J]. *Journal of Plant Genetic Resources*, 2016.
17. HIEI Y. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA [J]. *Plant J*, 1994, 6.
18. YANG H, BI W Y, CHEN H X, et al. Contrastive Studies of Coal between Direct Ultrasonic Extraction and Soxhlet Extraction [J]. *Advanced Materials Research*, 2011, 236-238.
19. OHUE M, SHIMODA T, SUZUKI S, et al. MEGADOCK 4.0: an ultra-high-performance protein-protein docking software for heterogeneous supercomputers [J]. *Bioinformatics*, 2014, 30(22).

20. KOICHIRO T, GLEN S, SUDHIR K. MEGA11: Molecular Evolutionary Genetics Analysis Version 11 [J]. *Molecular Biology and Evolution*, (7): 7.
21. EN-HUA, HAI-BIN, ZHANG, et al. The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis [J]. , 2017, (6): 866-77.
22. REN X, CHEN Z, LIU Y, et al. ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in *Arabidopsis* [J]. *The Plant Journal*, 2010.
23. WENBO, JIANGDIQIU, YU. *Arabidopsis* WRKY2 transcription factor mediates seed germination and postgermination arrest of development by abscisic acid [J]. *Bmc Plant Biology*, 2009.
24. SUO J, TONG K, WU J, et al. Comparative transcriptome analysis reveals key genes in the regulation of squalene and β -sitosterol biosynthesis in *Torreya grandis* [J]. *Industrial Crops and Products*, 2019, 131: 182-93.
25. YANG R, YAN Y, ZENG Y, et al. Correlation between squalene synthase promoter and WRKY transcription factor in *Camellia oleifera* [J]. *Journal of Horticultural Science and Biotechnology*, 2020: 1-10.
26. LAGACÉ M, MATTON D P. Characterization of a WRKY transcription factor expressed in late torpedo-stage embryos of *Solanum chacoense* [J]. *Planta*, 2004, 219(1): 185-9.
27. XU X, CHEN C, CHEN F Z. Physical and Functional Interactions between Pathogen-Induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 Transcription Factors [J]. *Plant Cell*, 2006, 18(5): 1310-26.
28. CIOLKOWSKI I, WANKE D, BIRKENBIHL R P, et al. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function [J]. *Plant Molecular Biology*, 2008, 68(1-2): 81-92.
29. YANG H, LI H, LI Q. Biosynthetic regulatory network of flavonoid metabolites in stems and leaves of [J]. *Scientific Reports*.
30. DI P, WANG P, YAN M, et al. Genome-wide characterization and analysis of WRKY transcription factors in *Panax ginseng* [J]. *BMC Genomics*, 2021, (1).
31. WEN B, LUO Y, LIU D, et al. The R2R3-MYB transcription factor CsMYB73 negatively regulates L-Theanine biosynthesis in tea plants (*Camellia sinensis* L.) [J]. *Plant Science: An International Journal of Experimental Plant Biology*, 2020, (298-): 298.
32. LEI L, BINGYAN X, XIAOFENG D, et al. WRKY Transcription Factors and Their Roles in Plant Defense Responses [J]. *Molecular Plant Breeding*, 2005, 3: 401-8.
33. ZENG, GUO, LP, et al. Arbuscular mycorrhizal symbiosis and active ingredients of medicinal plants: current research status and perspectives [J]. *MYCORRHIZA*, 2013, 2013,23(4)(-): 253-65.
34. SHAMALA L, ZHOU H, HAN Z, et al. UV-B Induces Distinct Transcriptional Re-programing in UVR8-Signal Transduction, Flavonoid, and Terpenoids Pathways in *Camellia sinensis* [J]. *Frontiers in plant science*, 2020, 11: 234.

Figures

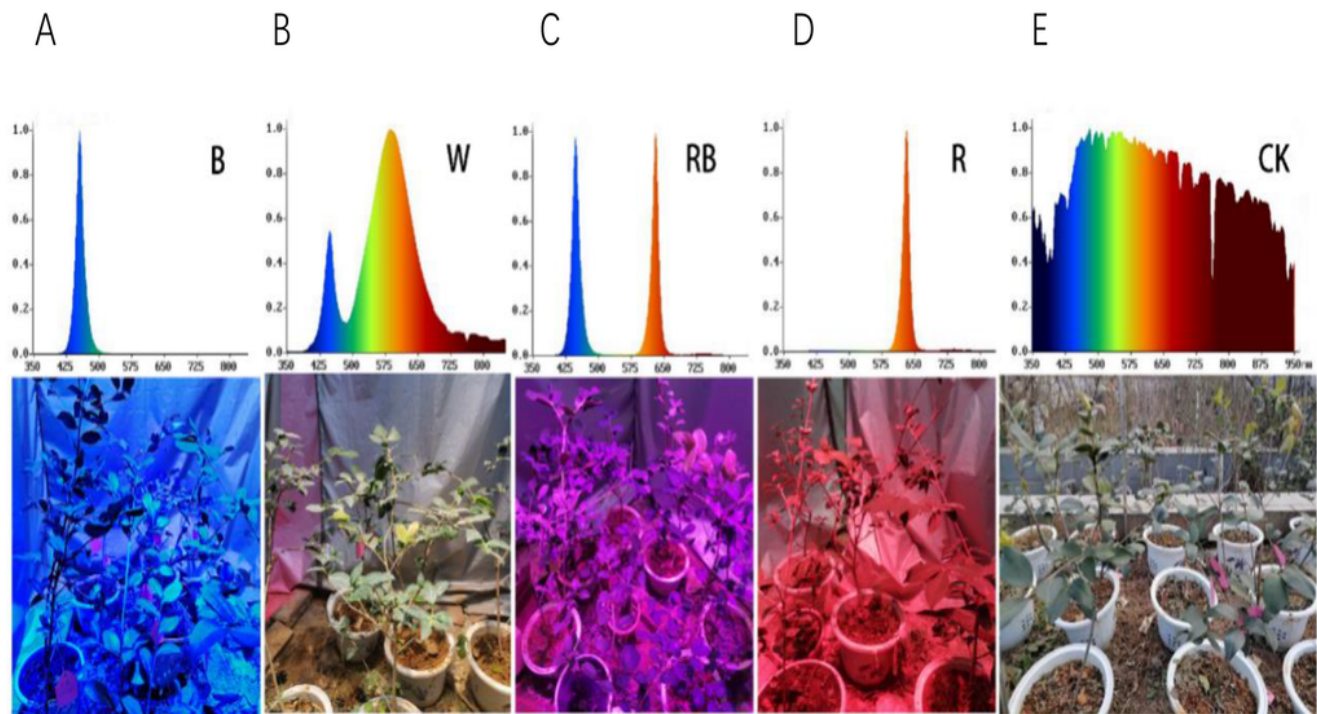


Figure 1. Light spectra of different treatments and *C. oleifera* growing under different light quality

Note : A is blue light treatment (450 ~ 480nm), B is white light treatment (420 ~ 720nm), C is red and blue light treatment (red and blue ratio : 1 : 1) ; D is red light treatment (610 ~ 640nm), E is sunlight treatment (CK).

Figure 1

See image above for figure legend

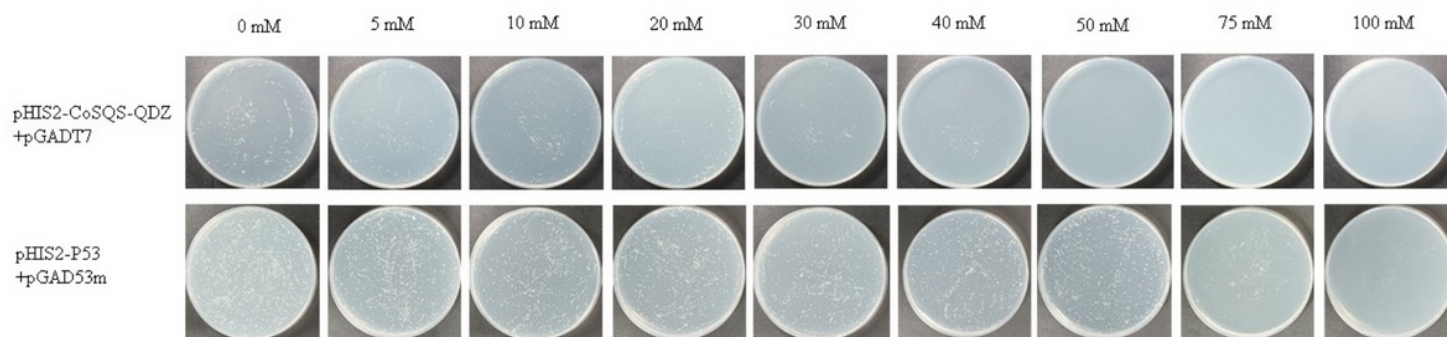


Figure 2

Background screening results

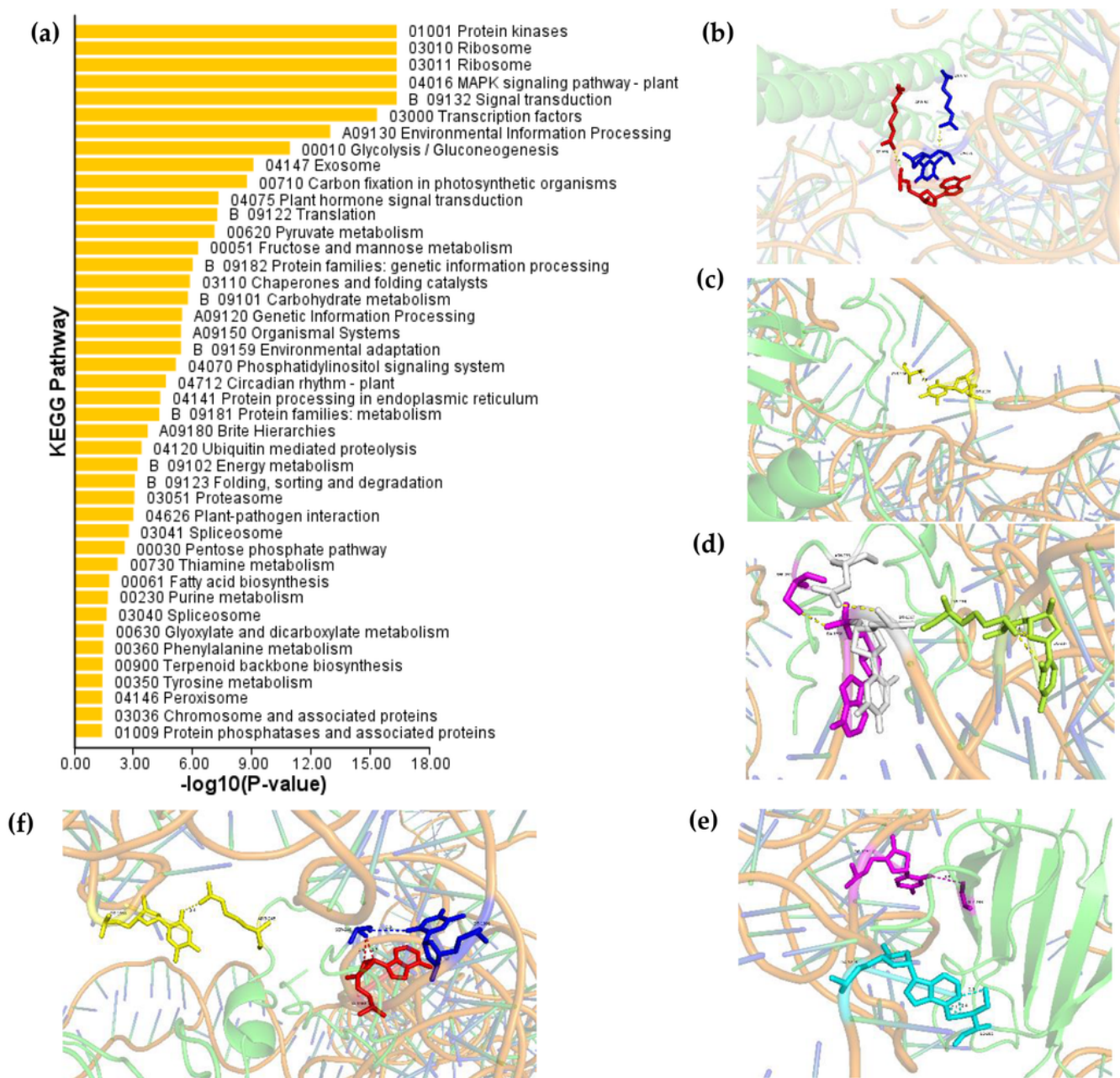


Figure 3. yeast one-hybrid screen library. (a) The metabolic pathways to which NGS belongs; (b)-(f) AOS predicted interaction sites between CoWRKY15 and CoSQS promoters.

Figure 3

See image above for figure legend

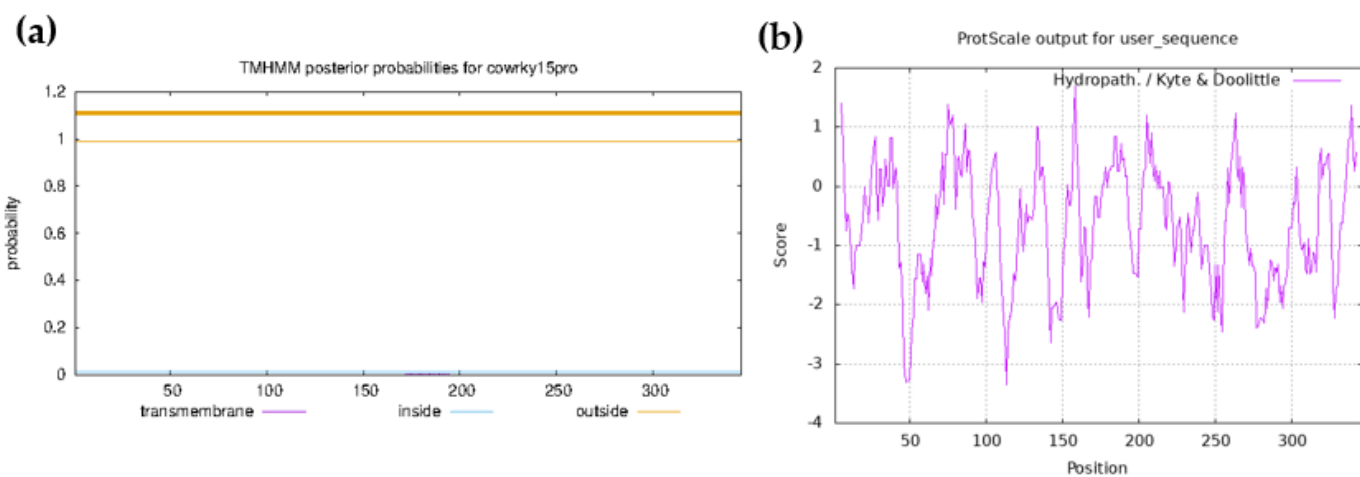


Figure 4

Transmembrane domain prediction and hydrophobicity prediction of CoWRKY15. (a) CoWRKY15 transmembrane prediction; (b) CoWRKY15 hydrophobicity prediction.

(a) (b)

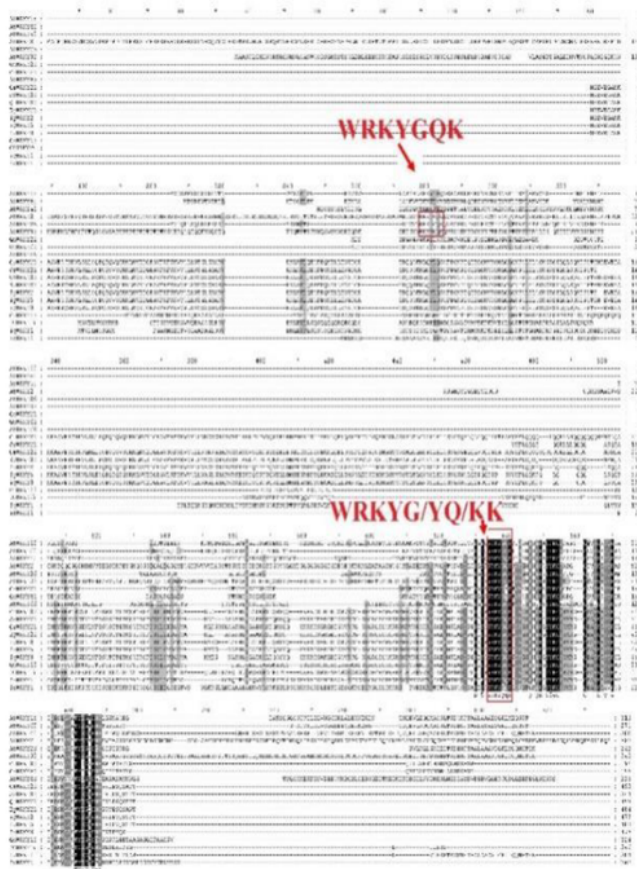
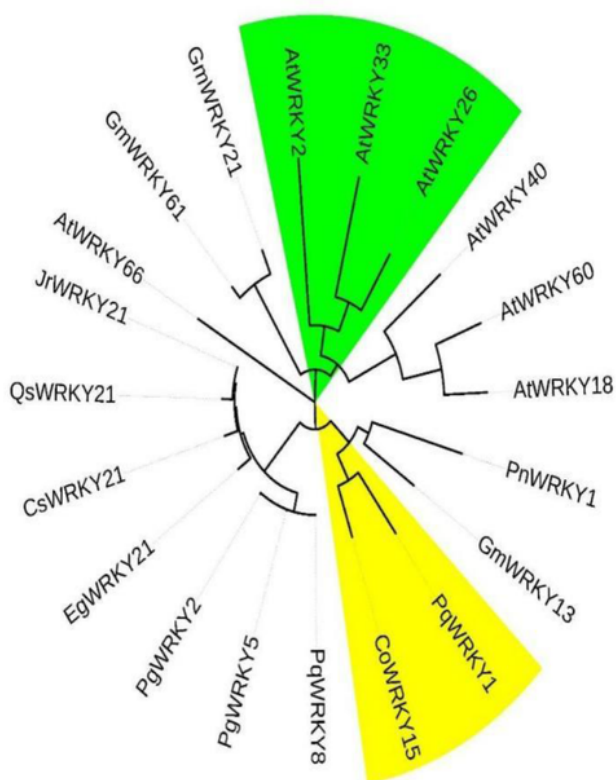


Figure 5. Cluster analysis of the deduced amino acid from CoWRKY15 CDS sequence with homologs of other plants WRKY sequences. (a) Cluster tree. The yellow label indicates the branch with the closest genetic distance to CoWRKY15. The green label indicates the type I branch; (b) Comparison of amino acid sequences. The red box represents the conservative domain .

Figure 5

See image above for figure legend

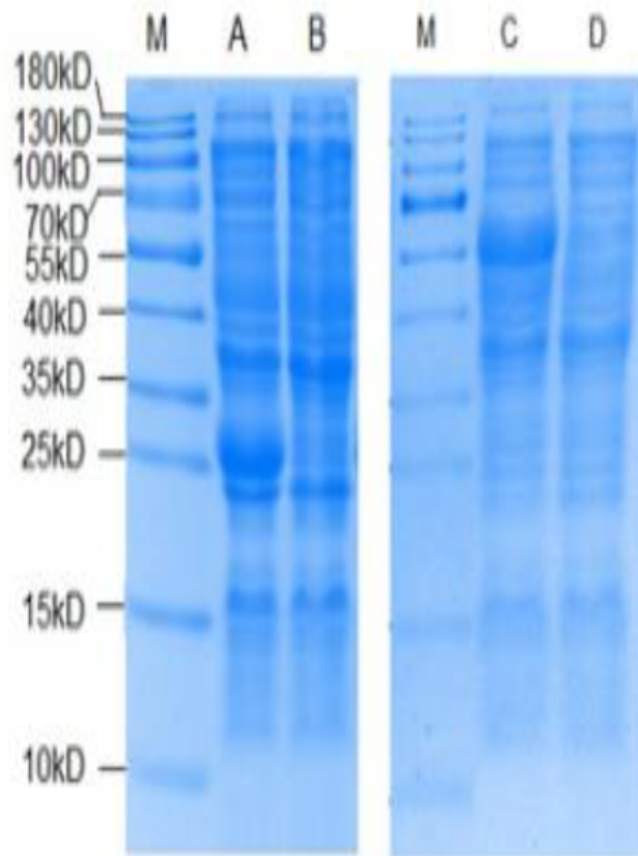


Figure 6. Analysis of induced expression products by SDS-PAGE. M: protein maker; A: pET32a empty plasmid with induced expression products; B: pET32a empty plasmid without the induced expression; C: pET32a-CoWRKY15 recombinant plasmid with induced expression products; D: pET32a-CoWRKY15 recombinant plasmid without induced expression.

Figure 6

See image above for figure legend

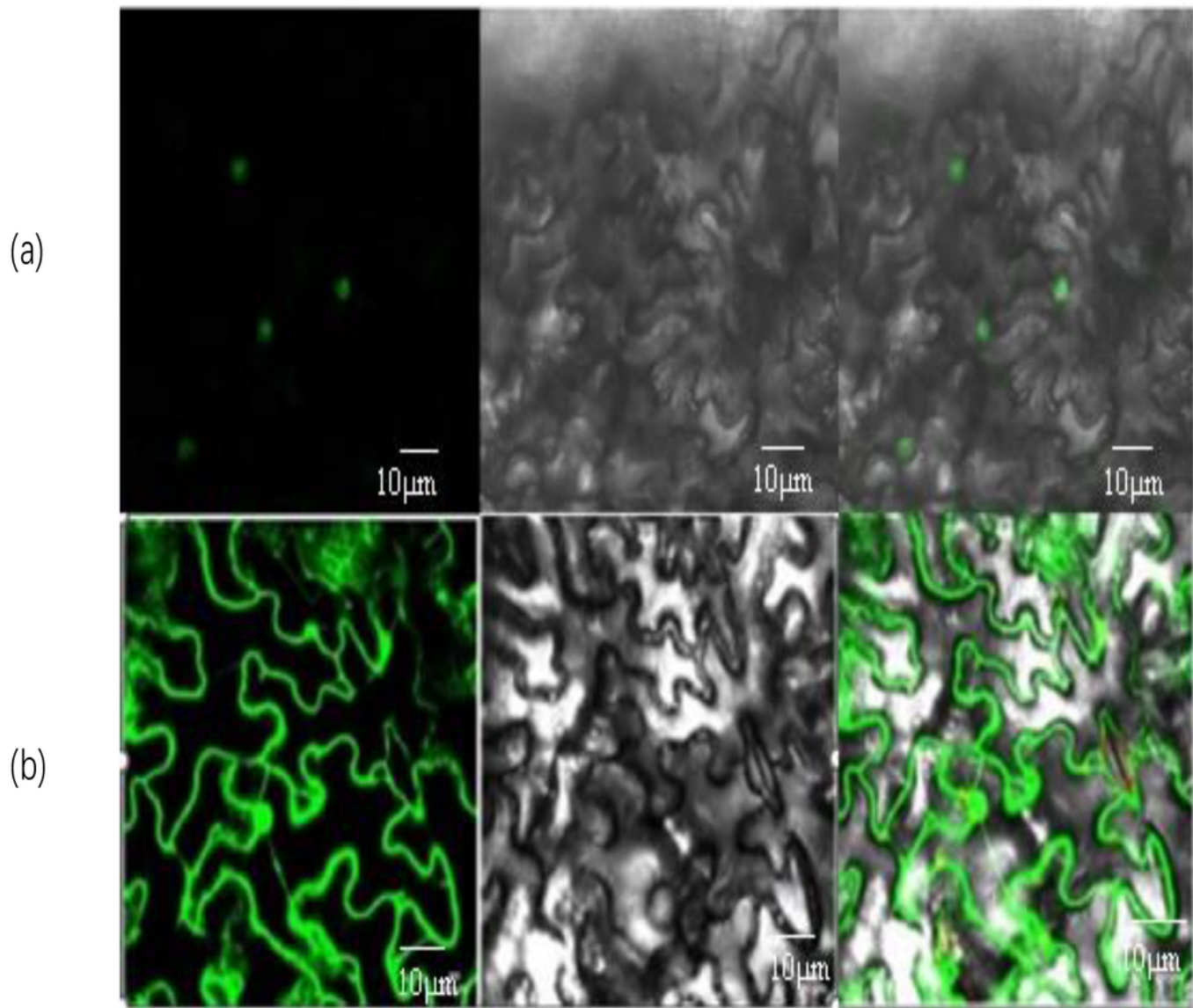


Figure 7. Subcellular localization of CoWRKY15. (a) pCAMBIA1300-CoWRKY15-GFP; (b) pCAMBIA1300-GFP

Figure 7

See image above for figure legend

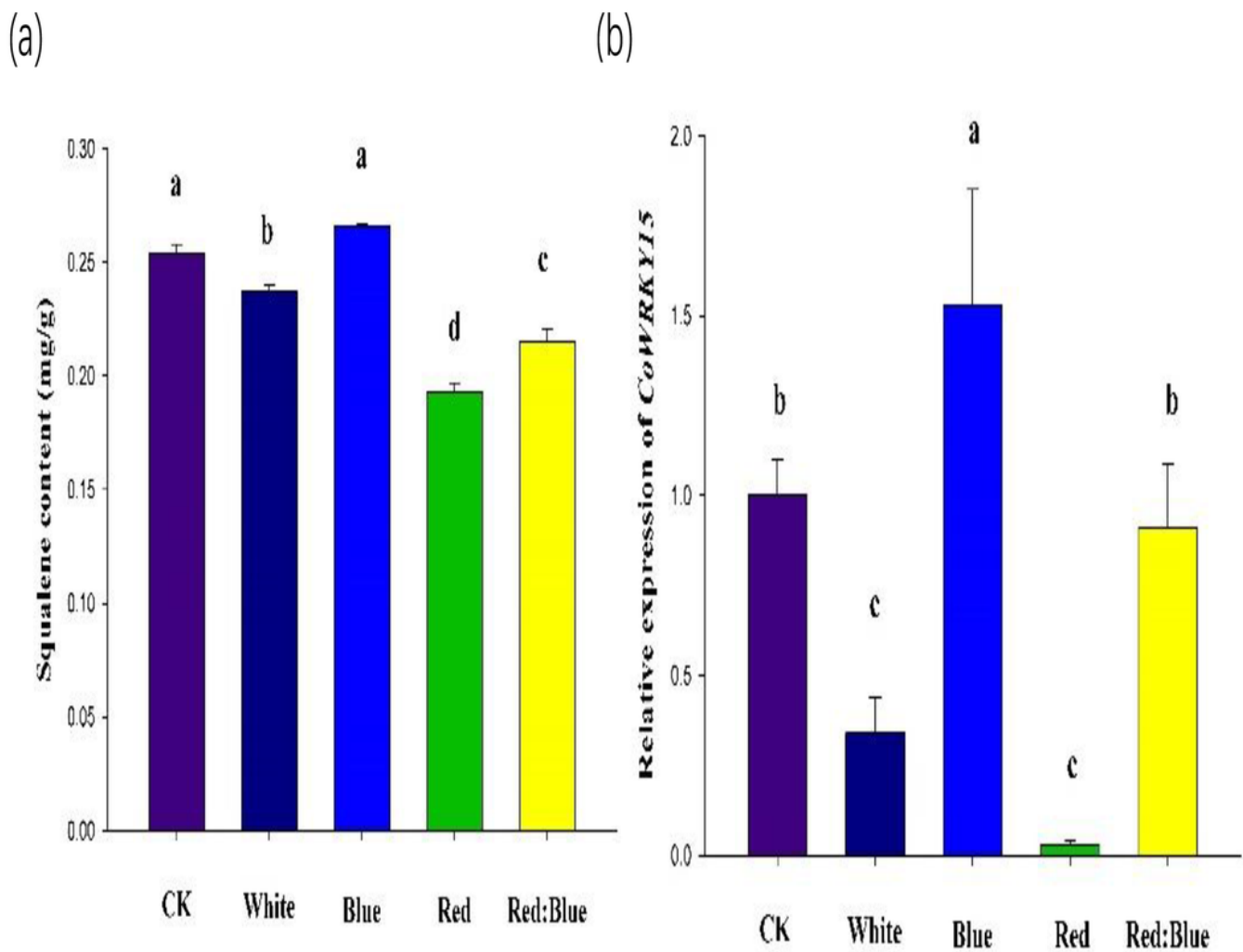


Figure 8. Squalene content and *CoWRKY15* expression under different light quality conditions. (a) Squalene content of *C. oleifera* kernel oil. (b) *CoWRKY15* relative expression in *C. oleifera* kernels. The light quality conditions were as follows: natural light: White, Blue, Red, and Red:Blue = 1:1.

Figure 8

See image above for figure legend