De novo transcriptome assembly of the cotyledon of *Camellia oleifera* for discovery of genes regulating seed germination

Wei Long (✉ ylslongwei@caf.ac.cn)

Research Institute of Subtropical Forestry

Xiaohua Yao

Research Institute of Subtropical Forestry

Kailiang Wang

Research Institute of Subtropical Forestry

Yu Sheng

Research Institute of Subtropical Forestry

Leyan Lv

Zhejiang Tongji Vocational College of Science and Technology

Research Article

Keywords: *Camellia oleifera*, Cotyledon, Seed germination, Transcriptome, Hormone

Posted Date: February 14th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1289104/v1>

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

[Read Full License](#)

Abstract

Background: *Camellia oleifera* (*C. oleifera*) is one of the most important wood oil species in the world. It was propagated by nurse seedling graft in *C.oleifera*. Since morphology of rootstock has a significant impact on grafting efficiency and seedling quality, it is necessary to have a clear understanding molecular mechanism of morphogenesis for cultivating high quality and controllable rootstocks,. However, the genomic resource for this species is relatively limited, thus hampering our understanding of the molecular mechanisms of seed germination in *C.oleifera*.

Results: In this paper, we measured the gene expression in the *C.oleifera* cotyledon in different development stages and the global gene expression profiles through transcriptome sequencing. Approximately 45.4 gigabases (GB) paired-end clean reads were obtained and assembled into 113,582 unigenes with an average length of 396 bp. 61.5% (68,217) of the unigenes were annotated in six public protein databases. We identified 11,391 differentially expressed genes (DEGs) between the different stages in the germination process. Enrichment analysis revealed that the DEGs were mainly involved in hormone signal transduction and starch sucrose metabolism pathways. The gravitropism regulator *UNE10*, the meristem regulator *STM*, *KNAT1*, *PLT2*, and root-specific transcription factor *WOX11* has higher expression level in the CAM2 stage (seed soaking), indicating the cotyledon-regulated program upon seed germination has been initiated to establish in early imbibition. Our data showed that multiple hormone-related genes in cotyledons were differentially reprogrammed during *C. oleifera* seed germination.

Conclusion: The cotyledons play key roles, both as the main nutrient provider and as one of the main instructors for seed germination and seedling growth. Together, our study will remarkably enrich the genomic resources of *Camellia* and contribute to our understanding of the molecular mechanisms of the development in the seed germination and seedling growth of *C.oleifera*. It is helpful to culture in standardizing and superior quality rootstock for *C. oleifera* breeding.

Background

The *C. oleifera*, one of the most important wood oil species in the world, is rich in unsaturated fatty acids and other nutrients, such as vitamins, camelliaside, and tea polyphenols, and has been increasingly recognized in China owing to the seed oil has similar nutritional value as olive oil [1-3]. In breeding practice, the reproduction of *C.oleifera* depends on grafting, and the rootstocks required for grafting are derived from seedlings after seed germination. The growth of rootstocks has a significant impact on the quality of grafting. Therefore, understanding the growth mechanism of rootstock seeds is essential for achieving efficient cultivation of rootstocks.

Seeds are able to sense and respond to environmental factors such as light [4], temperature [5], nutrients [6] and water [7], in order to control the precise timing of germination. "Once the embryonic growth potential exceeds the mechanical constraint of the surrounding tissues, including the cotyledon,

germination is then complete.”[8]. This suggests that these environmental factors regulate common downstream events, probably targeting and acting on plant hormone metabolism and signaling. It has been shown that environmental signals can regulate hormone metabolism in the seed, as well as seed responsiveness to hormones [9-11]. Sand storage (low temperature stratification treatment) is the most common method for ending seed dormancy. In the dormancy breaking treatment, seeds are stored under wet conditions at 4°C.

The cotyledons play an important role in supporting embryonic growth by supplying nutrients, protecting the embryo and controlling embryo growth by acting as a mechanical barrier during seed development and germination. Its structure and function in mature, dry seeds are divergent and specialized among different plant species. A subset of cotyledon tissues is composed of living cells even after seed maturation, which play a critical role in the regulation of seed germination. Cotyledons, as the nutrient source of the seed germination stage, will participate in each stage of seedling morphogenesis after germination. So far, many studies have been conducted to look at endogenous hormones [12,13], exogenous hormones, temperature and other factors during seed germination of *C.oleifera* [14,15], but the molecular mechanism of cotyledon participating in morphogenesis of seedling has been rarely investigated.

In this paper, we explored the dynamic transcript profile of the functional transition of cotyledonary tissue during seed germination and the emergence of seedling of *C.oleifera*, aiming to provide a clear understanding of the genetic elements that participate in this physiological process. This can help to achieve more accurate cultivation strategies and improve the propagation efficiency and quality of seedlings to satisfy the demand for low cost, high efficiency, and automation for seedling breeding of *C.oleifera*.

Methods

Plant material

CL18, as a variety granted number: S-SC-CO-007-2008 in *C.oleifera*, was obtained from the Research Institute of Subtropical Forestry, China Academy of Forestry. Characteristics of CL18 were described by He et al [33]. Mature seeds of CL18 variety were collected from Dongfanghong Forestry, Jinhua City, Zhejiang Province, and stored at 4 °C for 2 months. Matured seeds of CL18 variety were collected from Dongfanghong Forestry Farm, Jinhua City, Zhejiang Province, China, and stored at 4 °C for 2 months. Then, in dark conditions at 25 °C, seeds were flooded using deionized water for 2 days, followed by sand stratification treatment in a germination box. After the seeds cracked open, vernier calipers were used to measure the total length of root and stem after germination. By removing the seed coat and embryo quickly, we collected cotyledon tissues of seeds that were stored at 4 °C (CAM0), soaked in purified water (CAM2), shell-broken (CAM4S), and the cotyledon tissues of seedlings with root growth to 2 cm (CAM6S) or 10 cm (CAM26S) in the sand stratification (**Fig. 8**). The samples were immediately frozen

in liquid nitrogen and stored at -80 °C for subsequent RNA-seq or hormone determination. Each test consisted of three biological replicates.

Hormone determination

The content of hormones was determined using liquid chromatography-mass spectrometry (LC-MS) method [34, 13]. Approximately 0.5 g of cotyledon tissues were ground with liquid nitrogen and then suspended in 30 ml of 80% methanol at 4°C for 24 h in dark. The extraction mixture was centrifuged at 3 500 r /min for 10 minutes, and the supernatant was collected. The precipitation was re-suspended in 20 ml of 80% methanol at 4°C for 1 h after centrifuged. The supernatant was merged and evaporated at 40°C on the rotary evaporator until there was no methanol residual. The remaining water phase was extracted and decolorized twice with 30 ml of petroleum ether, and the ether phase was discarded. The pH of the aqueous phase was adjusted to 2.9, and 30 ml of ethyl acetate was used to extract the solution 3 times, then the ester phase was combined, and the solution was decompressed and dried at 40°C. The dry powder was dissolved in 2 ml methanol, filtered with 0.45 m microporous membranes, stored in a refrigerator at 4°C. Chromatographic column: Hypersil BDS C18 chromatographic column; The mobile phase was methanol-water (containing 0.75% glacial acetic acid, 35:65, V/V). Detection wavelength: 254 nm; Flow rate: 0.8 ml /min; Injection volume 10 ul; Column temperature: 30°C.

RNA isolation and illumina sequencing of the transcriptome

Approximately 100 mg of fresh cotyledon tissue was used for total RNA extraction, according to the manufacturer's protocol of Trizol reagent (Invitrogen, CA, USA). The total RNA quantity and purity were analysis of Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA) with RIN number >7.0.

Approximately 10 ug of total RNA was subjected to isolate poly(A) mRNA with poly-T oligoattached magnetic beads (Invitrogen). Following purification, the poly(A)- or poly(A)+ RNA fractions is fragmented into small pieces using divalent cations under elevated temperature. Then the cleaved RNA fragments were reverse-transcribed to create the final cDNA library in accordance with the protocol for the Illumina TruSeq Stranded mRNA Library Prep kit (Illumina, San Diego, USA), the average insert size for the paired-end libraries was 300 bp (±50 bp). And then we performed the paired-end sequencing on an Illumina platform (Hiseq 4000) at the LC-BIO (Hangzhou, China) following the vendor's recommended protocol.

De novo assembly and functional annotation

The raw reads were quality controled using Cutadapt and Perl scripts in-house and the sequence quality was further verified using FastQC. De novo assembly of the transcriptome was performed with Trinity 2.4.0 [35]. Trinity groups transcripts into clusters based on shared sequence content and the longest transcript in the cluster was selected as the representative gene (aka Unigene). For functional annotation, the unigenes were aligned against the non-redundant (NR) protein database

(<http://www.ncbi.nlm.nih.gov/>), Gene ontology (GO), SwissProt, Kyoto Encyclopedia of Genes and Genomes (KEGG) and eggNOG databases using DIAMOND with a threshold of E-value < 0.00001[36].

Identification of differentially expressed gene (DEG)

Salmon [37] was used to perform expression level for Unigenes by calculating TPM (Transcripts Per kilobase Million) [38]. The DEGs were specified with a threshold of $\frac{1}{2}\log_2(FC)^{\frac{1}{2}}$ at $^3 1$, p value < 0.01, using edgeR [39].

Heatmap plotting of DEGs

Heatmap was generated with Z-score normalized TPM values of the DEGs using the online OmicShare tools (<https://www.omicstudio.cn/analysis>).

GO and KEGG enrichment analysis

The GO and KEGG pathway enrichment analysis of DEGs was implemented by Goseq R packages [40] and KOBAS software [41], respectively. Both analysis was tested at a significance cutoff $FDR \leq 0.05$.

Identification of transcription factor

The TF families were identified by BLASTX against known plant TFs identified in iTAK database (<http://itak.feilab.net/cgi-bin/itak/index.cgi>, version: v1.7) with E-value threshold $\leq 10^{-5}$.

Weight gene co-expression network analysis

The input data for the WGCNA were the normalized values for each transcript. The DEGs co-expression network of was constructed using R package WGCNA(V1.6) with a soft threshold was chosen to create networks with a scale-free topology using the method described by Langfelder and Horvath [42, 43]. After the networks were built, modules of transcripts with similar expression patterns were created and eigengens for these modules were calculated. Finally, correlations between these eigengenes and the content of hormones and length of seedling were calculated (**Table S10**).

Quantitative Real-time RT-PCR (qRT-PCR) analysis

12 representative DEGs were selected to validate the RNA-Seq analysis using qRT-PCR. The Tublin α -3 (*TUB α -3*) gene and Glyceraldehyde-3-phosphate dehydrogenase (*GADPH*) was used as the reference gene [44]. The primer pairs were designed according to the selected unigene sequences using Primer 3.0 Plus (<http://primer3plus.com/cgi-bin/dev/primer3plus.cgi>). The primers were 19-21 bp in length, and had amplicon lengths of 200-260 bp (**Table S11**). Each qRT-PCR was represented by three biological and three technical replicates. Relative transcription abundance was derived using the $2^{-\Delta\Delta CT}$ method [45, 46]. All reactions were carried out in 96-well plate in QuantStudio™ Real-Time PCR Software (Thome) using the TB Green® Premix Ex Taq™ (Tli RNaseH Plus)(TaKaRa) kit. The amplification procedure is 95 °C for 30 s, followed by 40 cycles of 95 °C for 5s, 60 °C for 30 s.

Results

Illumina paired-end sequencing and de novo assembly, functional annotation

In total, we obtained 707,382,370 raw reads, from which a total of 690,806,162 clean reads were yield after quality control (**Table S1**). All clean reads were deposited at NCBI and can be accessed via SRP349608. All of high-quality reads from these fifteen libraries were mixed to do transcriptome assembly by using the Trinity software. All clean reads were de novo assembled into 284,782 transcripts with a mean length of 441bp and N50 of 1,308 bp. These further assembled into 113,582 unigenes, with a mean length of 396 bp and N50 of 934 bp (**Table S2**). Based on homologous searches, the total of 68,217 (60.06%) unigenes had significant blast by DIAMOND program hits in at least one of the six database (**Table S3**). Further analysis of homologies had the highest homology with sequences from *Vitis vinifera* (19.25%), followed by *Oryza sativa* (5.71%), *Juglans regia* (4.51%) (**Fig. S1**).

Functional categories of the unigenes

To gain insight into the functional categorization of the assembled unigenes, GO classification and KEGG pathway-based analysis was performed. A total of 38,657 unigenes were assigned to GO terms, which were classified into 50 functional groups under three principle categories, i.e., biological process, Molecular function and Cellular components (**Fig. 1A**). Under the Biological process category, the largest subgroups were “biological process”, “regulation of transcription, DNA templated”, and “transcription, DNA templated”; With the cellular components category, “nucleus”, “cytoplasm”, “plasma membrane” were the most highly represented categories. For the Molecular function category, the most abundant of genes were associated with “molecular function”, “protein binding”, and “ATP binding”.

Besides, a total of 20,652 unigenes were matched in the KEGG database and assigned to 138 pathway in five KEGG biochemical pathways, including “metabolism”, “Genetic information processing”, “Organismal systems”, “Cellular processes”, “Environmental information processing” (**Fig. 1B**). The three most represented pathways were “translation”, “carbohydrate metabolism”, and “folding, sorting and degradation”, followed by “environment adaptation” and “overviews”, whereas “membrane transport” and “replication and repair” pathways represented the smallest categories. There were relation with seed germination in pathways, including plant hormone signal transduction, starch and sucrose metabolism, biosynthesis of amino acids.

To evaluate the effectiveness of the annotation and the completeness of the transcriptome library, all the assembled unigenes were subjected to search against the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) database. Unigenes were assigned into 23 categories (**Fig. 1C**). The top three categories were ‘function unknown’, ‘postranslational modification, protein turnover, chaperones’, and ‘signal transduction mechanisms’, followed by ‘transcription’ and ‘replication, recombination and repair’; the smallest category was ‘cell motility’ and ‘nuclear structure’.

Identification and functional classification of DEGs

To find the germination-induced genes in seed, DEGs during seed germination were identified by pairwise comparisons at different time-points. In general, there was a minimum number of DEGs between the CAM2 and CAM4S time points, and higher proportions of DEGs between the CAM2 and CAM6S time-point and between the CAM6S and CAM26S time point. Thus, most of the changes in gene expression occurred after the stage of water uptake (CAM4S) (**Fig. 2A**|**Table S4**). A Venn analysis showed the distribution of common and unique DEGs of comparison between different adjacent stages, and between CAM0 and CAM4S, CAM6S, and CAM26S time-points. Compared to the 12 common DEGs between adjacent stages (**Fig. 2B**), the comparison of CAM0 between CAM4S, CAM6S, and CAM26S time-point have more common DEGs (297) (**Fig. 2C**), indicating that there were substantial alterations of gene expression among different germination stages in *C.oleifera* seeds. These phase-changes of gene expression during seed germination stages may have important functional implications in the growth of *C.oleifera* seed.

To gain insights into the functional categorization and metabolic pathways of DEGs involved in seeds germination, DEGs in cotyledons during *C.oleifera* seed germination were subjected to function enrichment analysis. Abundant GO terms like “nucleus”, “transcription, DNA-templated”, “molecular function”, “zinc ion binding”, “plasma membrane”, and “protein serine/threonine kinase activity” were enriched in comparisons of adjacent stages, and CAM0 compared to CAM2, CAM4S, CAM6S, CAM26S (**Fig. S2**). Besides, “response to light stimulus” and “thylakoid” were significantly enriched in the comparison of CAM4S vs CAM2.

In addition, the results of DEGs mapped to the KEGG database showed that a large number of genes were involved in the pathways related to “plant-pathogen interaction”, “carbon metabolism”, “endocytosis”, “ribosome”, “RNA transport”, “Protein processing in endoplasmic reticulum”, “spliceosome”, “starch and sucrose metabolism”, “phenylpropanoid biosynthesis”, and “plant hormone signal transduction” (**Fig. S3**). This suggests a crucial role of hormones in the regulation of *C.oleifera* seed germination

Identification of hormone signaling-related genes in seed germination

To gain insights into the functional roles of hormones during *C.oleifera* seed germination, we mapped the DEGs to hormone signaling and transduction pathways, and analyzed their expression in different stages (**Fig. 3**). A total of 86 genes were identified to be associated with the biosynthesis, metabolism and signaling of ten hormones, including Absciscic acid (ABA), Auxin (IAA), Ethylene (ETH), Karrikins (KARs), Brassinosteroid (BR), Cytokinin (CTK), Gibberellin acid (GAs), Strigolactone (STR), Salicylic acid (SA) and Jasmonic acid (JA). The genes associated with ABA biosynthesis and metabolism represented as the the largest group with 17 members, followed by GAs with 13 members, KARs and Auxin with 12 members, CTK with 10 members. In contrast, the genes with JA signaling was the smallest group with only 1 members. All these hormone signaling and transduction-related genes exhibited differential expression across the different stages examined. A significant proportion of DEGs related to GA, Auxin and ETH had the highly expression level in the CAM26S stage, suggesting their potential roles in seedling growth. Besides, most of the genes that respond to KARs and BR are relatively highly expressed at CAM0 and/or CAM2 stage, suggesting that they may functioned in the early stage of seed germination.

Identification of transcription factors in *C.oleifera*

Transcription factors (TFs) play critical roles in various plant development process. To provide insights into the regulatory network underlying seed germination, we examined the expression of TFs in different especially their dynamic differentially expression. Totally, 310 TFs belonging to 36 different families were found to be differentially expressed in the stages examined. The TFs displayed expression patterns across the five stages. A majority of TFs showed relatively broad expression patterns in all stages (**Fig. S4**), while some exhibited distinctive stage-specific patterns (**Fig. 4**). Interestingly, multiple members of the WRKY and ERF transcription factor family, which respond to GA signals [16], were identified to be significantly high expressed in CAM2 stage (soaking seeds). In addition, the meristem regulator *SHOOTMERISTEMLESS* (*STM*), *KNOTTED-like from Arabidopsis thaliana* (*KNAT1*), *PLETHORA 2* (*PLT2*), and *WOX11*, as the root-specific transcription factor, was also observed to be highly expressed in this stage [17, 18].

Coexpression network analysis with WGCNA

To identify the clusters of highly interconnected genes that were specific to tissues, coexpression networks are constructed on the basis of pairwise correlations between genes in their common expression trends across all sampled tissues. As observed in the dendrogram, 20 unique modules eigengenes were identified (**Fig. 5A**). Notably, 4 out of 20 coexpression modules are comprised of genes that are significantly relevant to hormones and growth during seed germination ($r \geq 0.7$, $P \leq 10^{-2}$) (**Fig. 5B**).

Here, two modules (MEyellow and MEgrey60) were listed for further analysis. The MEyellow module, including 935 genes (**Table S5**), was highly correlation with content of ABA (**Fig. 5B, 5C**). According to the GO analysis, all genes in the MEyellow module were highly enriched in molecular function, zinc ion binding, quercetin 7-O-glucosyltransferase activity, transcription factor activity, UDP-glycosyltransferase activity, sugar:proton symporter activity, etc (**Table S5**). Those enriched in the KEGG pathways were associated with plant-pathogen interaction, plant hormone signal transduction, circadian rhythm-plant, amino sugar and nucleotide sugar metabolism, etc (**Table S5**).

WGCNA can also be employed to construct gene networks, in which each node represents a gene and the connecting lines (edges) between genes represent coexpression correlations. Hub genes are those that show most connections in the network. In the MEyellow module network, there were twenty seven genes encode transcription factors and four genes encode hormone-related. Most of these transcription factor genes were highly expressed during soaking seed, and may regulate or participate in activities operating in the seed germination (**Table S6**).

The 50 most highly connected hub genes in the MEyellow module were used for analysing gene expression network. Gene expression showed that the expression level in soaking seed stages was higher than in other stages (**Fig. 6A, 6B; Table S7**). Co-expression network showed core hub genes, namely 25556_c1_g1 and 26030_c3_g4. The 25556_c1_g1 gene is of unknown function. The 26030_c3_g4 (*AT3G29970*) gene is belongs to molecular function. Other highly connected hub genes

include signal transduction *CoPH1*(25268_c3_g2), sugar:proton symporter activity *CoPMT3*(26833_c0_g1), vacuole *CoAT3G62550*(27851_c0_g1), *CoDJC66*(29838_c0_g1) and so on (**Fig. 7B**). *CoPMT3*(26833_c0_g1) is annotated to protein mannosyl transferases gene family (PMT), which is an important sugar proton symporter activity, and produce SUMO (small, ubiquitin-like modifier). PMTs have differential tissue-specific functions in phosphatidylcholine (PC) biosynthesis and plant growth. *PMT3* are the primary enzymes for phosphocholine (PCho) biosynthesis and are involved in PtdCho biosynthesis and vascular development in *Arabidopsis* seedlings [19, 20].

The 110 genes in the MEgrey60 module were highly correlated with the growth of seedling (**Fig. 5B, 5D**). These genes were highly enriched in GO terms, including molecular function, protein serine/threonine kinase activity, transferase activity, sugar:proton symporter activity and transcription factor, secondary metabolite biosynthetic process, signal transduction, transcription, transmembrane transport (**Table S8**). Besides, KEGG analysis showed that those genes were enriched in pathways associated with plant-pathogen interaction, plant hormone signal transduction, phenylpropanoid biosynthesis, cyanoamino acid metabolism, limonene and pinene degradation, brassinosteroid biosynthesis (**Table S8**). The hub genes with the highest edge number (81) is *CoVIT_13s0047g00260* (28624_c1_g1), a member of the shikimate 3-O-methyltransferase (ST) protein kinase family, is associated with protein autophosphorylation, transporter activity [*CoAT1G24430* (29788_c0_g1), *CoNPF3.1* (40839_c0_g1), *CoK4D641*(gene43291_c0_g3)], xyloglucan: xyloglucosyl transferase activity [*CoXTH7* (30876_c0_g5)], protein binding [*CoP02879*(41442_c2_g1)], peroxidase activity [*CoPOD* (42101_c0_g1)], xylem development [*CoPRX52* (29405_c1_g1)] (**Fig. 6C, D; Table S9**). *CoXTH7* (30876_c0_g5) can catalyze xyloglucan endohydrolysis (XEH) and/or endotransglycosylation (XET). It may be an essential constituent of the primary cell wall, and participates in cell wall biogenesis. *CoNPF3.1* (40839_c0_g1) was annotated to nitrate transporter1/peptide transporter (NPF) family, and it has been described as transmembrane transporter, gibberellic acid homeostasis and nitrate assimilation in *Arabidopsis thaliana* [21]. It may improve development of bud and growth of root [22].

Validation of the Expression of DEGs by qRT-PCR

To validate the reliability of expression profiling obtained by RNA-seq, 12 DEGs with different expression in cotyledon of seed were selected for qRT-PCR analysis. The genes chosen for qRT-PCR analysis included 9 TFs and 3 hormone signaling-related genes. For all these genes, the results of qRT-PCR exhibited almost similar expression patterns as compared to those in the RNA-seq data (**Fig. 7A**), confirming a high correlation between the RNA-seq and qRT-PCR data (**Fig. 7B**).

Discussion

Cotyledons play a critical role in regulating seed germination and morphogenesis

Seed germination refers to some biological processes including the radicle prominence, the seed coat and the resumption of embryo growth, accompanied with nutrient metabolism, DNA repair and transcription [23, 24]. Cotyledon or endosperm, the source of nutrient storage, plays an important role

during seed germination. In *Arabidopsis*, the transcriptomes of the endosperm and embryo were quite similar. Thus, the two tissues likely express similar genetic programs that are intrinsic to the seed [25, 26]. It can be assumed that genes expressed in cotyledon are similar to those appear in embryo, and cotyledon acts as a key partner in seed germination and subsequent growth.

In this study, we characterized the gene expression profiling in cotyledon of *C.oleifera* in different stages during seed germination. Several key pathways in seed germination were obtained by KEGG annotation analysis. These pathways are principally involved in cell wall biosynthesis, cell proliferation, primary metabolism and hormone signaling (**Fig. S3**). Similar to the process of seed germination involved the transformation of nutrient metabolism and photosynthesis in other species, a large number of DEGs were identified to be involved in various metabolic pathways, and most of them were related to the regulation of gene expression, followed by energy production and metastasis (**Fig. S3**). These reflect the drastic physiological and genetic regulation activities in the cotyledons, the nutrient storage sources. These results suggested that a certain percent of genes in cotyledons might play largely overlapped roles in governing the germination and the emergence of seedling of *C.oleifera* seed, compared with the whole gene set throughout the seed.

Candidate hormone associated with C.oleifera seed germination

In present study, many genes were highly expressed in CAM26S, including *GASA3*, *GASA9*, *GASA11* and *OS08G0141300*. Particularly, gibberellin-regulated protein that may function in hormonal controlled steps of development, such as seed germination, flowering and seed maturation [27]. Meantime, there were many SAGUR (small auxin up-regulated RNA) genes highly expressed in CAM26S—including *CoSAUR32*, *CoSAUR36*, *CoSAUR55*, *CoSAUR59* and *CoSAUR74*. The genes of the SAUR family may be play a central role in auxin-induced acid growth by regulating cell wall acidification, but can also act independently of auxin, tissue specifically regulated by various other hormone pathways and transcription factors [28]. It have revealed the functions of several SAUR genes by molecular genetic studies in *Arabidopsis*, such as *AtSARU32*—*AtSAUR36* in cell elongation [29]. Furthermore, some SAURs may be perform roles in processes other than cell elongation, such as leaf senescence or cell division. Besides, The *SAUR55*, presented in our data (**Fig. 3**), was previously reported to encode auxin response protein that plays important role in plant growth [30].

Distinct expression pattern specific to stage was observed from that most of the identified genes that relate to AUXIN, GA and ETH, were relatively highly expressed in CAM26S, compared with the genes related to KARs and BR were highly expressed at CAM0 and CAM2 stage (**Fig. 3**). This suggests a potential roles of AUXIN, GA and ETH in the stage of seedling growth. Our data showed that multiple phytohormone related genes in cotyledons were differentially reprogrammed during germination, indicating that at least in part, multiple hormone-related factors in cotyledons synergistically contribute to the seed germination and subsequent vegative growth of *C.oleifera*. The cotyledons are not only the main nutrient provider but also one of the main instructor for seed germination and seedling growth.

Candidate TFs associated with C.oleifera seed germination

TFs play important roles in various plant development processes. There were encoded more than 1500 TFs of over 5% of the genome in *Arabidopsis* [31]. It was found about 60 kinds of TFs families in higher plants [32], which they were involved in regulations of various processes. In the present study, our transcriptome profiling revealed that a subset of TFs belonging to AP2/ERF, bHLH, MYB, WRKY, C2H2, C3H, NAC, GRAS and HB-HD-ZIP families were predominantly expressed in *Camellia* seed germination (**Fig. S4**). In CAM2, the soaking seeds make the sealed seed embryo and cotyledon to exchange substance and energy with the outside world and enter into an active metabolic state, which may prompt the seeds to enter the germination stage from the dormant state. Since ABA and GA as the main plant hormones regulate seed dormancy and germination, the balance of their biosynthesis and catabolism are regulate the stability of dormancy and germination to seed. In this study, we found that *CobHLH093*(38274_c0_g1) as a transcription factor regulating gibberellin biosynthesis, and *CoAREB3*(37687_c2_g6) and *MYB3*(Co31728_c1_g5) as a transcription factor promoting seed dormancy and ABA activation , were highly expressed in the CAM2 stage (**Fig. 4**), indicating that two hormone-related genes may synergistic effect. In addition, high expression of *CoWOX11*(33796_c4_g1), *CoSPT* (32638_c0_g1), *CoKNAT1*(37948_c2_g1), *CoTINY2*(32839_c0_g2), *CoSTM* (43196_c1_g1), and *CoPLT2*(38998_c0_g2) (**Fig. 4**), which genes are regulate growth of stem or root in meristem tissue, was found at this stage. Meantime, *CoSGR5*(30478_c0_g2) and *CoUNE10*(44750_c2_g2), which are involved in root gravity, were also highly expressed in this stage. It is implied that the switch of seed germination had be triggered during imbibition.

Conclusion

In this study, a de novo assembly of transcriptome data from five stages of *C.oleifera* was performed with the main objective to provide preliminary insights into the change of cotyledon during seed germination. Enrichment analysis revealed that the DEGs were mainly involved in hormone signal transduction and starch sucrose metabolism pathways. The gravitropism regulator *UNE10*, the meristem regulator *STM*, *KNAT1*, *PLT2*, and root-specific transcription factor *WOX11* has higher expression level in the CAM2 stage (seed soaking), indicating the cotyledon-regulated program upon seed germination has been initiated to establish in early imbibition. The cotyledons play key roles, both as the main nutrient provider and as one of the main instructors for seed germination and seedling growth. Taken together, our study will remarkably enrich the genomic resources of *Camellia* and laid a foundation contribute for further understanding of the molecular mechanisms of the development in the seed germination and seedling growth of *C.oleifera*.

Declarations

ACKNOWLEDGMENTS

Data summarized in this paper have been generated through works of several authors which we would like to thank for their continuous efforts for the emergence of sesame crop.

AUTHOR CONTRINBUTIONS

WL designed the experiment, performed data processing and drafted the manuscript. KW helped in bioinformatics analysis and data interpretation. LL prepared the materials and performed the experiments. XY and KW participated in the design to the study, helped in data processing, and revision of the manuscript. LL assisted in results interpretation and manuscript preparation. XY and YS conceived the study and revised the manuscript. All authors read and approved final version of the manuscript.

Funding

This research was funded by National Key R&D Program of China (2019YFD1001602), the Fundamental Research Funds for the Central Public Welfare Scientific Research Institutes (CAFYBB2017SY016).

Availability of Supporting Data

All the sequencing data used in this research are available from the NCBI Sequence Read Archive (SRA) database with accession number SRP349608.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare no conflict of interest

Permission to use varieties

We have been authorized by the variety owner to use the CL18 in this study.

Author details

¹ Zhejiang Provincial Key laboratory of Tree breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, No.73 Daqiao Road, Fuyang District, Hangzhou 311400, China.

²Department of Hydraulic, Zhejiang Tongji Vocational College of Science and Technology, No. 418, Gengwen Road, Xiaoshan District, Hangzhou 311231, China.

Received: date; Accepted: date; Published: date

References

1. Zhuang R. China Oil Tea. Second edition. Beijing: China Forestry Press.2008.(in chinese)
2. Wang N, Zhang W, Qin M, Li S, Qiao M, Liu Z, et al. Drought tolerance conferred in soybean (*Glycine max.* L) by *GmMYB84*, a novel *R2R3-MYB* transcription factor. Plant Cell Physiol. 2017; 58(10): 1764-1776. <https://doi.org/10.1093/pcp/pcx111>
3. Zeng W, Endo Y. Lipid characteristics of *Camellia* seed oil. J oleo sci. 2019; 68(7): 649-658. <https://doi.org/10.5650/jos.ess18234>.
4. Oh E, Yamaguchi S, Hu J, Yusuke J, Jung B, Paik I, et al. *PIL5*, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in *Arabidopsis* seeds. Plant Cell. 2007; 19(4): 1192-1208. <https://doi.org/10.1105/tpc.107.050153>
5. Tsuchiya Y. Thermoinhibition uncovers a role for strigolactones in *Arabidopsis* seed germination. Plant & Cell Physiol. 2012; 53(1): 107-17. <https://doi.org/10.1093/pcp/pcr176>
6. Matakiadis T, Alboresi A, Jikumaru Y, Tatematsu K, Pichon O, Renou JP, et al. The *Arabidopsis* abscisic acid catabolic gene *CYP707A2* plays a key role in nitrate control of seed dormancy. Plant Physiol. 2008; 149(2): 949-960. <https://doi.org/10.1104/pp.108.126938>
7. Bai X, Chen J, Kong X, Todd C, Yang Y, Hu X, et al. Carbon monoxide enhances the chilling tolerance of recalcitrant *Baccaurea ramiflora* seeds via nitric oxide-mediated glutathione homeostasis. Free Radic Biol Med. 2012; 53(4): 710-720. <https://doi.org/10.1016/j.freeradbiomed.2012.05.042>
8. Finch-Savage WE, Leubner-Metzger G. Seed dormancy and the control of germination. New phytol. 2006; 171(3): 501-523. <https://doi.org/10.1111/j.1469-8137.2006.01787.x>
9. Duermeyer L, Khodapanahi E, Yan D, Krapp A, Rothstein SJ, Nambara E. Regulation of seed dormancy and germination by nitrate. Seed Sci Res. 2018; 28(3): 150-157. <https://doi.org/10.1017/S096025851800020X>
10. Yang L, Liu S, Lin R. The role of light in regulating seed dormancy and germination. J Integr Plant Biol. 2020; 62(9): 1310-1326. <https://doi.org/10.1111/jipb.13001>
11. Xu H, Lantzouni O, Bruggink T, Benjamins R, Lanfermeijer F, Denby K, et al. A molecular signal integration network underpinning *Arabidopsis* seed germination. Curr Biol. 2020; 30(19): 3703-3712. <https://doi.org/10.1016/j.cub.2020.07.012>
12. Shi J, Wang J, Zeng WJ. Changes of several ingredients in *Camellia oleifera* seeds during storage and germination. Tropicl For. 2018; 46(4): 14-17. <https://doi.org/10.3969/j.issn.1672-0938.2018.04.004>. (in chinese)
13. Chen BW, Liu H, Chen XM, Cai L, Huang J, Qin Z, et al. Hormone physiology of two *Camellia oleifera* Abel. species during seed Germination. J Shanxi Agri Sci. 2012; 40(8): 840-843. <https://doi.org/10.3969/j.issn.1002-2481.2012.08.09>. (in chinese)
14. Yuan J, Tan X, Luo J, Li W. Germination effect of *Camellia Oleifera* seed with different methods. China Seed Industry. 2009; 9: 50-51. <https://doi.org/10.3969/j.issn.1671-895X.2009.09.024>
15. Cheng JY, Zhou XH, Deng XZ, Li Y, Du YW, Huang Y. Effect of different treatment measures of seed germination rate of *Camellia oleifera*. Hubei Forestry Sci Tech. 2014; 43(6), 26-27,56. <https://doi.org/CNKI: SUN:FBLI.0.2014-06-009>. (in chinese)

16. Li W, Yong YB, Zhang Y, Lyu YM. Transcriptional regulatory network of ga floral induction pathway in la hybrid Lily. *Int J Mol Sci.* 2019; 20(11): 2694. <https://doi.org/10.3390/ijms20112694>
17. Zhou SL, Jiang W, Long F, Cheng SF, Yang WJ, Zhao Y, Zhou D.X. Rice homeodomain protein *WOX11* recruits a histone acetyltransferase complex to establish programs of cell proliferation of crown root meristem. *The Plant Cell.* 2017; 29(5): 1088-1104. <https://doi.org/10.1105/tpc.16.00908>
18. Rupp HM, Frank M, Werner T, Strnad M, Schmülling T. Increased steady state mRNA levels of the *STM* and *KNAT1* homeobox genes in cytokinin overproducing *Arabidopsis thaliana* indicate a role for cytokinins in the shoot apical meristem. *The Plant J.* 2002; 18(5): 557-563. <https://doi.org/10.1046/j.1365-313X.1999.00472.x>
19. Liu YC., Lin YC, Kanehara K, Nakamura Y. A pair of phospho-base methyltransferases important for phosphatidylcholine biosynthesis in *Arabidopsis*. *The Plant J.* 2018; 96(5): 1064-1075. <https://doi.org/10.1111/tpj.14090>
20. Liu YC, Lin YC, Kanehara K, Nakamura Y. A methyltransferase trio essential for phosphatidylcholine biosynthesis and growth. *Plant physiol.* 2019; 179(2): 433-445. <https://doi.org/10.1104/pp.18.01408>
21. Sugiura M, Georgescu MN, Takahashi M. A nitrite transporter associated with nitrite uptake by higher plant chloroplasts. *Plant Cell Physiol.* 2007; 48(7): 1022-1035. <https://doi.org/10.1093/pcp/pcm073>
22. Vittozzi Y, Nadzieja M, Rogato A, Radutoiu S, Valkov VT, Chiurazzi M. The *Lotus japonicus* *NPF3.1* is a nodule-induced gene that plays a positive role in nodule functioning. *Front Plant Sci.* 2021; 12: 688187. <https://doi.org/10.3389/fpls.2021.688187>
23. Barrôco RM, Van Poucke K, Bergervoet JH, De Veylder L, Groot SP, Inzé D, et al. The role of the cell cycle machinery in resumption of postembryonic development. *Plant Physiol.* 2005; 137(1): 127-140. <https://doi.org/10.1104/pp.104.049361>
24. Masubelele NH, Dewitte W, Menges M, Maughan S, Collins C, Huntley R, et al. D-type cyclins activate division in the root apex to promote seed germination in *Arabidopsis*. *PNAS.* 2005; 102(43): 15694-15699. <https://doi.org/10.1073/pnas.0507581102>
25. Penfield S, Gilday AD., Halliday KJ, Graham IA. DELLA-mediated cotyledon expansion breaks coat-imposed seed dormancy. *Curr Biol.* 2006; 16(23): 2366-2370. <https://doi.org/10.1016/j.cub.2006.10.057>
26. Penfield S, Li Y, Gilday AD, Graham S, Graham IA. *Arabidopsis* ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *The Plant Cell.* 2006; 18(8): 1887-1899. <https://doi.org/10.1105/tpc.106.041277>
27. Ahmad S, Lu C, Wei Y, Gao J, Jin J, Zheng C, et al. Stage specificity, the dynamic regulators and the unique orchid *Arundina graminifolia*. *Int J Mol Sci.* 2021; 22(20): 10935. <https://doi.org/10.3390/ijms222010935>
28. Stortenbeker N, & Bemer M. The SAUR gene family: the plant's toolbox for adaptation of growth and development. *J Exper Bot.* 2019; 70(1): 17-27. <https://doi.org/10.1093/jxb/ery332>

29. Stamm P, Kumar PP. Auxin and gibberellin responsive *Arabidopsis* *SMALL AUXIN UP RNA36* regulates hypocotyl elongation in the light. *Plant Cell Rep.* 2013; 32: 759-769. <https://doi.org/10.1007/s00299-013-1406-5>
30. Narsai R, Law SR, Carrie C, Xu L, Whelan J. In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in *Arabidopsis*. *Plant Physiol.* 2011; 157: 1342-1362. <https://doi.org/10.1104/pp.111.183129>
31. Dröge-Laser W, Snoek BL, Snel B, Weiste C. The *Arabidopsis* bZIP transcription factor family—an update. *Curr Opin Plant Biol.* 2018; 45: 36-49. <https://doi.org/10.1016/j.pbi.2018.05.001>
32. Jin J, Tian F, Yang D, Meng Y, Kong L, Luo J, et al. Plant TFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 2017; 45(D1): 1040-1045. <https://doi.org/10.1093/nar/gkw982>
33. He F, Yao XH. Cultivation of Oil Tea in China. Beijing: China Forestry Press. 2013. pp22. (in chinese)
34. DU W, Yuan CJ, Ding J, Li JB, Li H, Wang L. Simultaneous determination of five plant hormones in different organs of *Camellia oleifera* by high performance liquid chromatography-tandem mass spectrometry. *J Anal Sci.* 2018; 34(1):37-42. <https://doi.org/10.13526/j.issn.1006-6144.2018.01.008>. (in chinese)
35. Grabherr M. G, Haas JB, Yassour M, Levin ZJ, Thompson AD, Ido A, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011; 29(7): 644-652. <https://doi.org/10.1038/nbt.1883>
36. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat methods,* 2015; 12(1): 59-60. <https://doi.org/10.1038/nmeth.3176>
37. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017; 14(4): 417-419. <https://doi.org/10.1038/nmeth.4197>
38. Mortazavi A, Williams B, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5(7): 621-628. <https://doi.org/10.1038/nmeth.1226>
39. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26(1): 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
40. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010; 11: R14. <https://doi.org/10.1186/gb-2010-11-2-r14>
41. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 2006; 34 (suppl_2): W720-W724. <https://doi.org/10.1093/nar/gkl167>
42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9: 559. <https://doi.org/10.1186/1471-2105-9-559>

43. Botia JA, Vandrovcova J, Forabosco P, Guelfi S, D'Sa Karishma, United Kingdom brain expression consortium, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. BMC Syst Biol. 2017; 11(1): 47-62. <https://doi.org/10.1186/s12918-017-0420-6>
44. Zhou CF, Lin P, Yao XH, Chang J, Han XJ. Selection of reference genes for quantitative Real-Time PCR In six oil-tea *Camellia* based on RNA-seq. Mole Biol. 2013; 47(6): 836-851. <https://doi.org/10.1134/S0026893313060198>
45. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biol. 2002; 3(7): 1-12. <https://doi.org/10.1242/jcs.115.6.1113>
46. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome biol. 2007; 8(2): 1-14. <https://doi.org/10.1186/gb-2007-8-2-r19>

Figures

Figure 1

See image above for figure legend.

Figure 2

See image above for figure legend.

Fig. 3 Expression profile of hormone signaling-related across five different stages. The expression of genes involved in hormone signaling and transduction-related pathways was Z-score normalized and hierarchically clustered in heatmap. A color scale is shown at the top. Green color indicate lower expression, red color indicates higher expression.

Figure 3

See image above for figure legend.

Figure 4

See image above for figure legend.

Figure 5

Figure 7

See image above for figure legend.

Fig. 8 Characterization of *C.oleifera* seed, and samples for RNA-seq. (A) Schematic illustration showing the structures including the plumule, cotyledonary petiole, hypocotyl, and huge cotyledon in *C. oleifera* seed. (B) *C. oleifera* seed or seedling showing the cotyledon tissues (covered in brown seed coat) sampled for RNA-seq and hormone determination.

CAM0: matured dry seed stored at 4 °C, CAM2: imbibed seed soaked in purified water for 2 days, CAM4S: seed with the shell is broking, CAM6S: seed with the root growing to 2 cm in the sand stratification, CAM26S: seedling with the root growing to 10 cm in the sand stratification.

Figure 8

See image above for figure legend.

Supplementary Files

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

- [TableS1Sequencingstatisticsofthetranscriptomefromfivestages.xlsx](#)
- [TableS2Summaryoffunctionalannotationforassembledunigenes.xlsx](#)
- [TableS3Annotationstatisticaofunigeneinpubliclyavailabledatabases.docx](#)
- [TableS4ComparablenumbersofDEGsbetweenupregulatedanddownregulatedindifferentstages.xlsx](#)
- [TableS5ThefunctionofgenesinMEyellowmodule.xlsx](#)
- [TableS6ThefunctionoftranscriptionfactorsandhormonesinMEyellowmodule.xlsx](#)
- [TableS7Thefunctionof50mosthighlyconnectedhubgenesintheMEyellowmodule..xlsx](#)

- [TableS8ThefunctionofgenesinMEgrey60module..xlsx](#)
- [TableS9ThefunctionofhubgenesintheMEgrey60module.xlsx](#)
- [TableS10.ThecontentofhormoneinseedsofC.oleifera.docx](#)
- [TableS11PrimersequencesusedinqRTPCRanalysis.docx](#)
- [Fig.S1SpeciesdistributionoftopBLSATHitsformatchedunigenesequences.pdf](#)
- [Fig.S2HistogrampresentationofGOclassificationindifferentgenes..pdf](#)
- [Fig.S3ClassificationofdifferentgenesinKEGGpathway.pdf](#)
- [Fig.S4ThedifferentialexpressionofTFsacrossfivestages.pdf](#)