BMC Plant Biology

RESEARCH

Open Access

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

Wei Long[1], Xiaohua Yao[1*], Kailiang Wang[1], Yu Sheng[1] and Leyan Lv[2]

Abstract

Background: *Camellia oleifera* (*C.oleifera*) is one of the most important wood oil species in the world. *C.oleifera* was propagated by nurse seedling grafting. Since the morphology of rootstocks has a significant impact on grafting efficiency and seedling quality, it is necessary to understand the molecular mechanism of morphogenesis for cultivating high-quality and controllable rootstocks. However, the genomic resource for this species is relatively limited, which hinders us from fully understanding the molecular mechanisms of seed germination in *C.oleifera*.

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

Conclusion: Cotyledons play vital roles, both as the main nutrient provider and as one primary instructor for seed germination and seedling growth. Together, our study will significantly enrich the genomic resources of *Camellia* and help us understand the molecular mechanisms of the development in the seed germination and seedling growth of *C.oleifera*. It is helpful to culture standard and superior quality rootstock for *C.oleifera* breeding.

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

Background

The *C. oleifera*, one of the most important wood oil species in the world, is rich in unsaturated fatty acids and other nutrients, including vitamins, camelliaside, and tea polyphenols. It has been increasingly recognized in China because its 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

*Correspondence: yaoxh168@163.com

[1] Zhejiang Provincial Key Laboratory of Tree Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, Zhejiang, China
Full list of author information is available at the end of the article

required for grafting are derived from seedlings after seed germination. The growth of rootstocks has a significant impact on the grafting quality. Therefore, understanding the growth mechanism of rootstock seeds is essential for achieving efficient cultivation of rootstocks.

Seeds can 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℃.

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 vary in different plant species. A subset of cotyledon tissues is composed of living cells after seed maturation, which play a critical role in regulating seed germination. As the nutrient source for seed germination, Cotyledons will participate in each stage of seedling morphogenesis after germination. So far, many studies have been conducted to examine endogenous hormones [12, 13], exogenous hormones, temperature, and other factors during seed germination of *C.oleifera* [14, 15], but molecular mechanisms of cotyledons participating in the morphogenesis of seedlings have been rarely investigated.

In this paper, we explored the dynamic transcript profile of the functional transition of cotyledonary tissues during seed germination and seedling emergence of *C.oleifera* (Fig. 1), aiming to provide a clear understanding of the genetic elements that participate in this physiological process. The research can provide more accurate cultivation strategies and improve seedlings' propagation efficiency and quality, which ultimately helps achieve low-cost, high-efficiency, and automatic seedling breeding on *C.oleifera*.

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 produced after quality control (Additional file 2: Table S1). All clean reads were deposited at NCBI and can be accessed via SRP349608. All of the high-quality reads from these fifteen libraries were mixed for transcriptome assembly

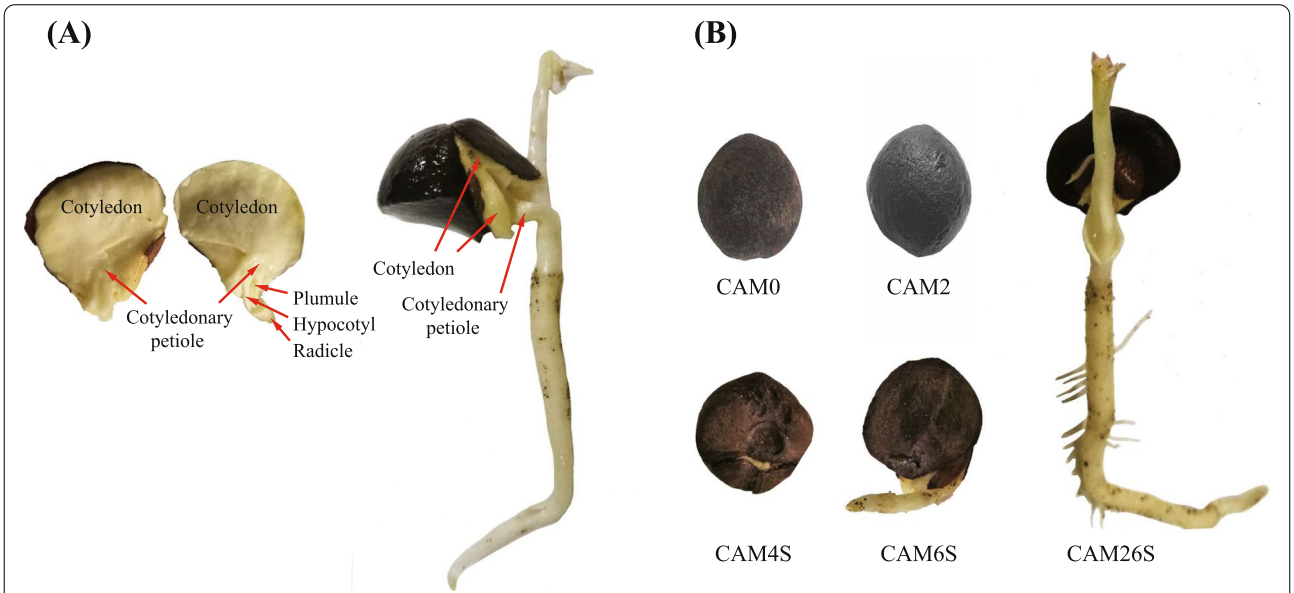

Fig. 1 Characterization of *C.oleifera* seeds, and samples for RNA-seq. **A** Schematic illustration showing the structures including the plumule, cotyledonary petiole, hypocotyl, and huge cotyledon in seed of *C.oleifera*. **B** Seed or seedling of *C.oleifera* showing the cotyledon tissues (covered in brown seed coat) sampled for RNA-seq and hormone determination. CAM0: matured dry seed stored at 4℃, 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

using the Trinity software. All clean reads were de novo assembled into 284,782 transcripts with a mean length of 441 bp and N50 of 1,308 bp. These reads would be assembled into 113,582 unigenes, with a mean length of 396 bp and N50 of 934 bp (Additional file 3: Table S2). Based on homologous searches, 68,217 (60.06%) unigenes had achieved blast by DIAMOND program hits in at least one of the six databases (Additional file 4: Table S3). Further analysis of homologies had the highest homology with sequences from *Vitis vinfera* (19.25%), followed by *Oryza sativa* (5.71%), *Juglans regia* (4.51%) (Additional file 1: Fig. S1).

Functional categories of the unigenes

GO classification and KEGG pathway-based analysis were performed to gain insight into the assembled unigenes' functional categorization.. A total of 38,657 unigenes were assigned to GO terms, which were classified into 50 functional groups under three principal categories, i.e., biological process, Molecular function, and Cellular components (Fig. 2A, Additional file 5: Table S4). In the Biological process category, the largest subgroups were "biological process", "regulation of transcription, DNA templated", and "transcription, DNA templated"; In the cellular components category, "nucleus", "cytoplasm", "plasma membrane" were the most highly represented ones. As for the Molecular function category, the most abundant 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 pathways in five KEGG biochemical pathways, including "metabolism", "Genetic information processing", "Organismal systems", "Cellular processes", and "Environmental information processing" (Fig. 2B, Additional file 5: Table S4). 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. The pathways are associated with seed germination, including "plant hormone signal transduction", "starch, and sucrose metabolism", and "biosynthesis of amino acids".

All the assembled unigenes were subjected to a search against the evolutionary genealogy of genes to evaluate the annotation's effectiveness and the transcriptome library's completeness: Non-supervised Orthologous Groups (eggNOG) database. Unigenes were assigned into 23 categories (Fig. 2C, Additional file 6: Table S5). The top three categories were "function unknown", "post-translational 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

The fifteen samples were separated into five groups in the principal component analysis (PCA) plot, and three replicates of each sample were grouped. Three seed samples at different developmental stages were closer to each other than to other samples (Additional file 1: Fig. S2). DEGs during seed germination were identified by pairwise comparisons at different time points to find germination-induced genes in seeds. The number of DEGs in CAM2 vs. CAM4S was minimum, and there were much more DEGs in CAM2 vs. CAM6S and CAM6S vs. CAM26S. Thus, most gene expression changes occurred after the water uptake stage (CAM4S) (Fig. 3A, Additional file 7: Table S6). A Venn analysis showed the distribution of common and unique DEGs between different adjacent stages and pairs of CAM0 vs. CAM2, CAM0 vs. CAM4S, CAM0 vs. CAM6S, and CAM0 vs. CAM26S. There were 12 common DEGs in adjacent stages (Fig. 3B, Additional file 8: Table S7). Comparatively speaking, there were more common DEGs (297) in pairs of CAM0 vs. CAM2, CAM0 vs. CAM4S, CAM0 vs. CAM6S, and CAM0 vs. CAM26S (Fig. 3C, Additional file 8: Table S7), indicating that the gene expression was substantially altered throughout different seed germination stages of *C.oleifera*. These phase-related gene expression changes during seed germination may have important functional implications for the growth of *C.oleifera* seed.

To gain insights into the functional categorization and metabolic pathways of DEGs involved in seed germination, 6820 of the 11,391 DEGs in cotyledons during *C.oleifera* seed germination were subjected to function annotation and function enrichment analysis. Abundant GO terms like "nucleus", "transcription, DNA-templated", "molecular function", "zinc ion binding", "plasma membrane, and "protein serine/threonine kinase activity"

(See figure on next page.)

Fig. 2 Functional categories of the unigenes. **A** Histogram presentation of GO classification. Bars represent the numbers of unigenes matched to each GO term of three categories: Biological process (blue), Cellular component (green). Molecular function (red). **B** The metabolism pathways are assigned into five categories. (a) Cellular processes. (b) Environmental information processing. (c) Genetic information processing. (d) Metabolism, and (e) Organismal systems. **C** Histogram presentation of eggNOG classification. Unigenes with significant matches in the NOG database were classified into 24 categories

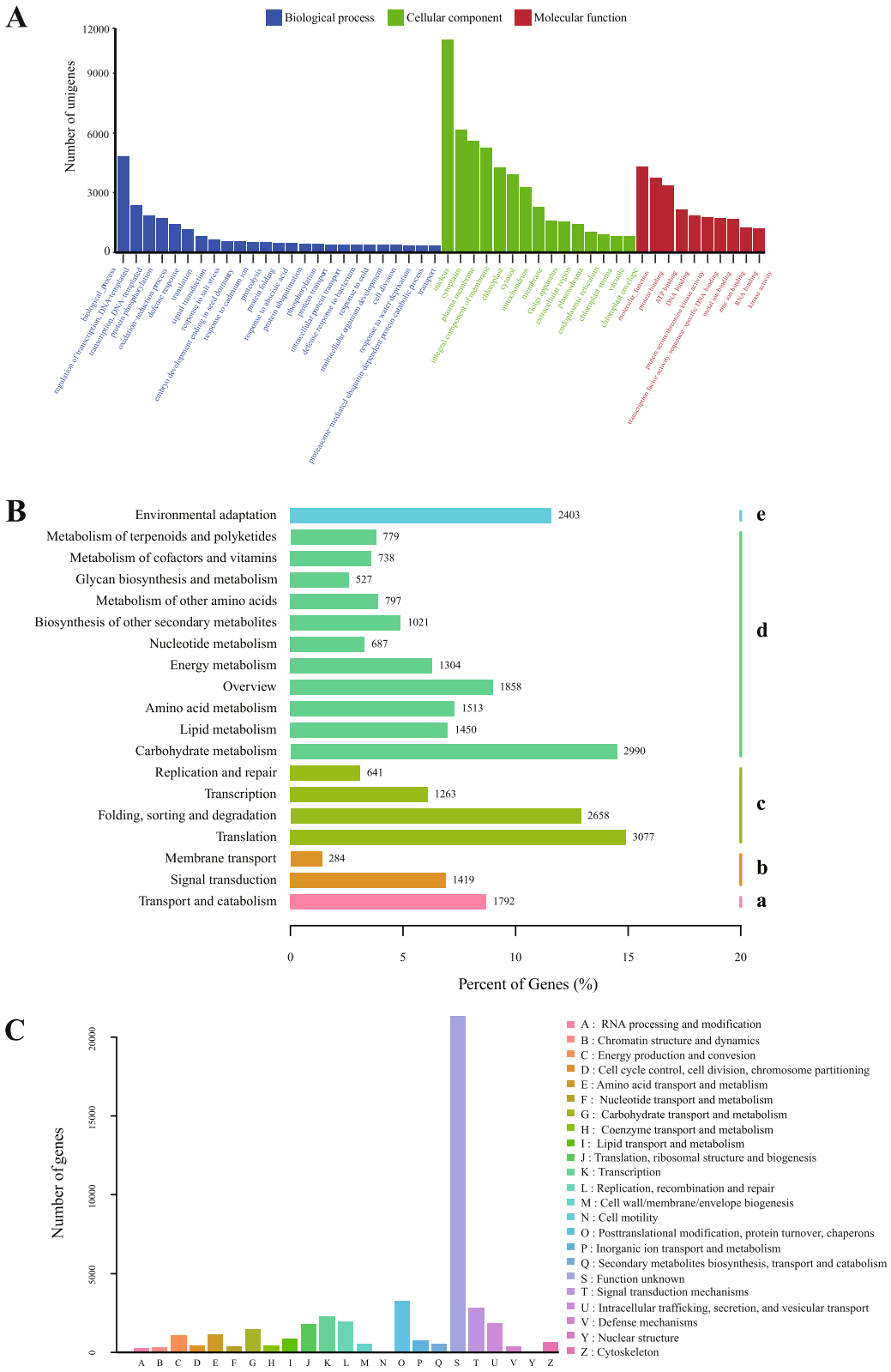

Fig. 2 (See legend on previous page.)

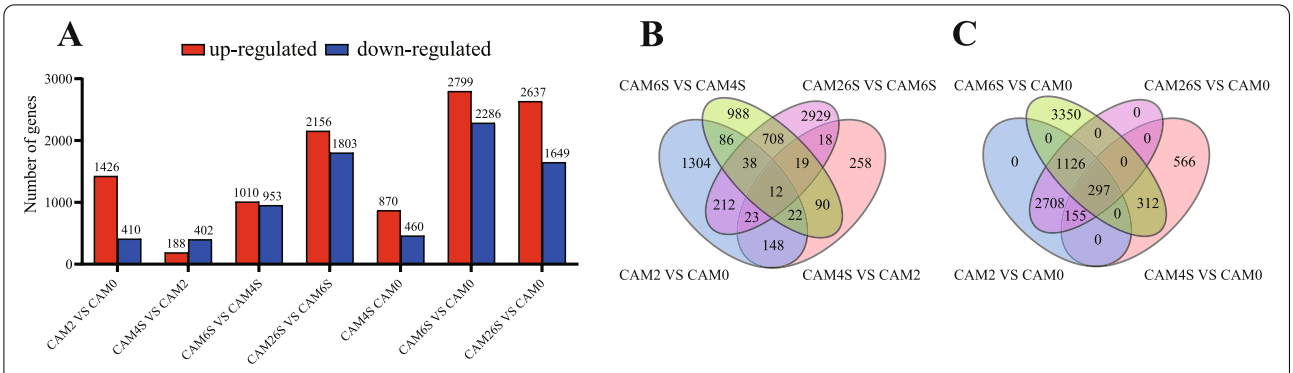

Fig. 3 Comparative analysis of tissue-enriched and differentially expressed genes. **A** Histogram shows the distribution of different genes in adjacent stages and different stages compared to CAM0. **B** Venn diagram showing the genes in tissue compared to adajacent stages. **C** Venn diagram showing the genes in tissues compared to CAM0

were enriched in adjacent stages, and pair of CAM0 vs. CAM2, CAM0 vs. CAM4S, CAM0 vs. CAM6S, and CAM0 vs. CAM26S (Fig. S3, Additional file 9: Table S8). In addition, the results of the 3810 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. S4, Additional file 9: Table S8). This suggests a crucial role of hormones regulating *C.oleifera* seed germination.

Besides, the GO terms of 12 common DEGs among adjacent stages mainly included plasma membrane, dioxygenase activity, and nucleosome. The GO terms of 297 common DEGs with pairs of CAM0 vs. CAM2, CAM0 vs. CAM4S, CAM0 vs. CAM6S, and CAM0 vs. CAM26S, mainly included "nucleosome, "molecular function", "secondary metabolic biosynthesis pathway", "signal transduction", and "vacuolar membrane". The genes involved in the KEGG pathways were related to the mRNA surveillance pathway, including "cyanoamino acid metabolism", "plant hormone signal transmission", "endocytosis", "plant pathogen interaction", "protein processing in endoplasmic reticulum", and "2-oxocarboxylic acid metabolism" (Additional file 8: Table S7).

Identification of genes associated with hormones signaling in seed germination

To gain insights into the functional roles of hormones during the seed germination *of C.oleifera*, we mapped the DEGs to hormone signaling and transduction pathways and analyzed their expression in different stages (Fig. 4). A total of 86 genes were identified to be associated with the biosynthesis, metabolism, and signaling of ten hormones, including Abscisic 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 were chosen as the largest group with 17 members, followed by GAs with 13 members, KARs and Auxin with 12 members, and CTK with 10 members. In contrast, the genes with JA signaling belonged to the smallest group with only 1 member. These hormone-signaling and transduction-related genes exhibited differential expression across the different examined stages. A significant proportion of DEGs related to GA, Auxin, and ETH had high expression levels in the CAM26S stage, indicative of 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 function in the early stage of seed germination.

Identification of transcription factors in *C.oleifera*

Transcription factors (TFs) play critical roles in various plant development processes. To provide insights into the regulatory network underlying seed germination, we examined the expression of TFs, especially their dynamic differential expression. 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 (Additional file 1: Fig. S5), while some exhibited distinctive stage-specific patterns (Fig. 5). Interestingly, multiple transcription factor members of the WRKY and ERF family, which respond to GA signals [16], were significantly

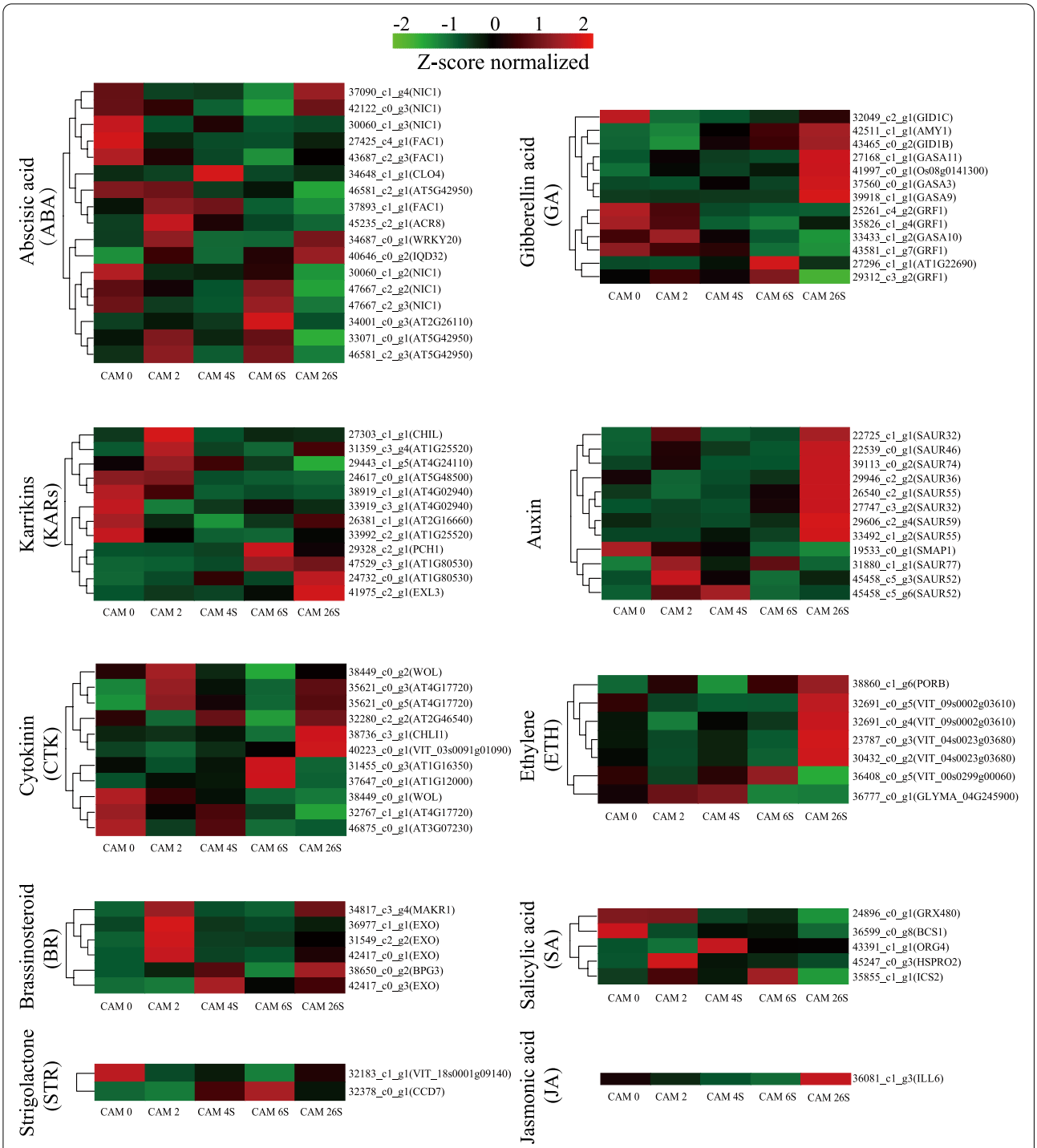

Fig. 4 Expression profiles of genes associated with hormones signaling across five different stages. The expression of genes involved in hormones signaling and transduciton-related pathways was Z-score normalized and hierarchically clustered in heatmap. Z-scores were calculated from the log2 fold change values. Z-score = (log2 fold change of TPM values—mean of log2 fold changes of all TPM values of the samples) / standard deviation of log2 fold changes of all TPM value of the samples. A color scale is showed at the top. Green color indicates lower expression, and red color indicates higher expression

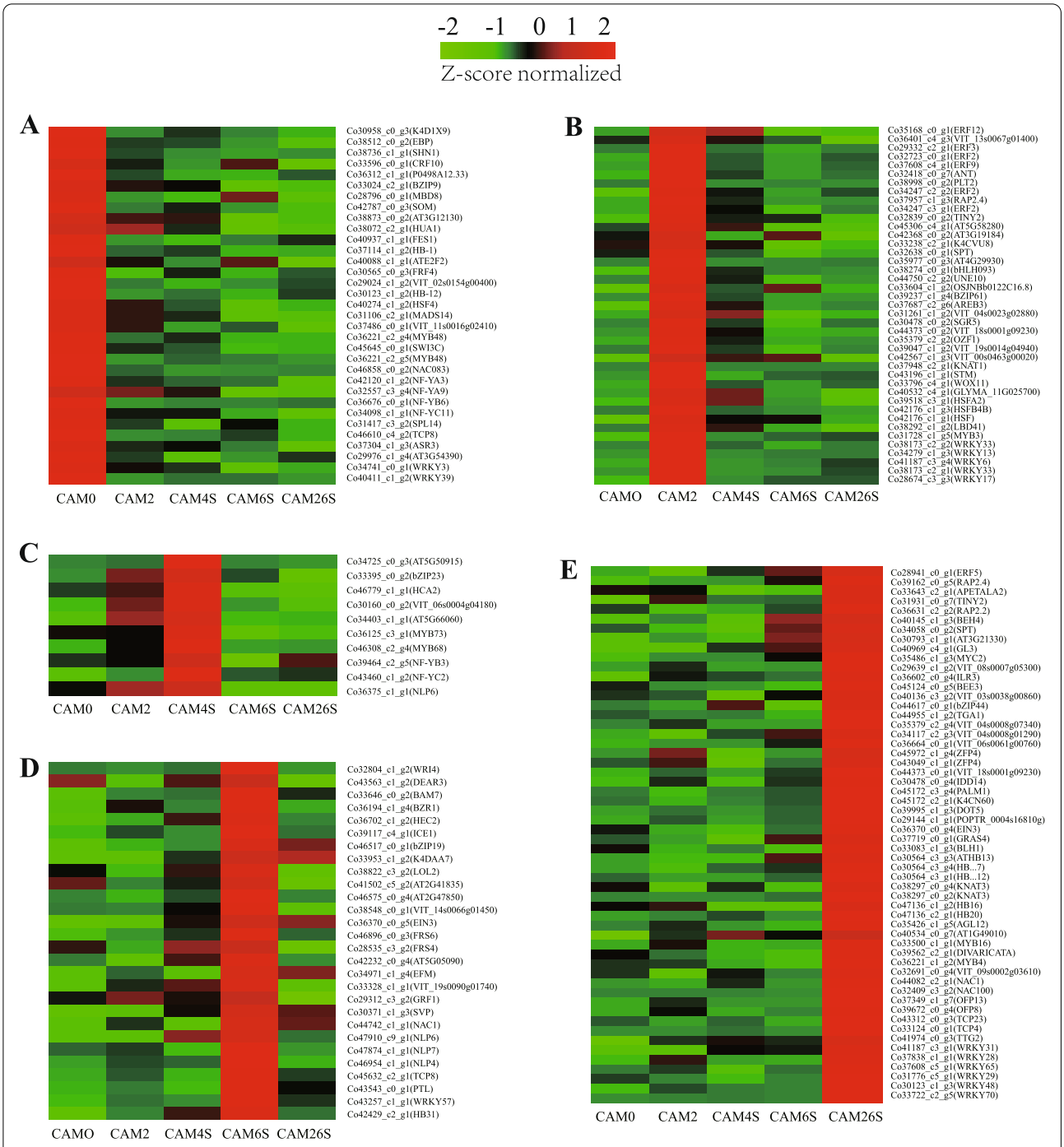

Fig. 5 Relatively highly expressed TF-encoding genes identified at specific stages. **A** Heatmap displaying the special expression patterns of TFs in samples of CAM0; **B** Heatmap displaying the special expression patterns of TFs in samples of CAM2; **C** Heatmap displaying the special expression patterns of TFs in samples of CAM4S; **D** Heatmap displaying the special expression patterns of TFs in samples of CAM6S; **E** Heatmap displaying the special expression patterns of TFs in samples of CAM26S

highly expressed in the 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].

Co-expression network analysis with WGCNA

To identify the clusters of highly interconnected genes that were specific to tissues, co-expression networks were 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 of eigengenes have been identified (Fig. 6A, Additional file 10: Table S9). Notably, 4 out of 20 co-expression modules consist of genes that are significantly relevant to hormones and growth during seed germination ($r > 0.7$, $P < 10^{-2}$) (Fig. 6B).

Here, two modules (MEyellow and MEgrey60) were listed for further analysis. The MEyellow module, including 935 genes (Additional file 10: Table S9), was highly correlated with the content of ABA (Fig. 6B, C). According to the GO analysis, all genes in the MEyellow module were highly enriched in molecular function, zinc ion binding, quercetin 7–0-glucosyltransferase activity, transcription factor activity, UDP-glycosyltransferase activity, and sugar: proton symporter activity, etc. (Additional file 7: Table S6). Those enriched in the KEGG pathways were associated with plant-pathogen interaction, plant hormone signal transduction, circadian rhythm-plant,

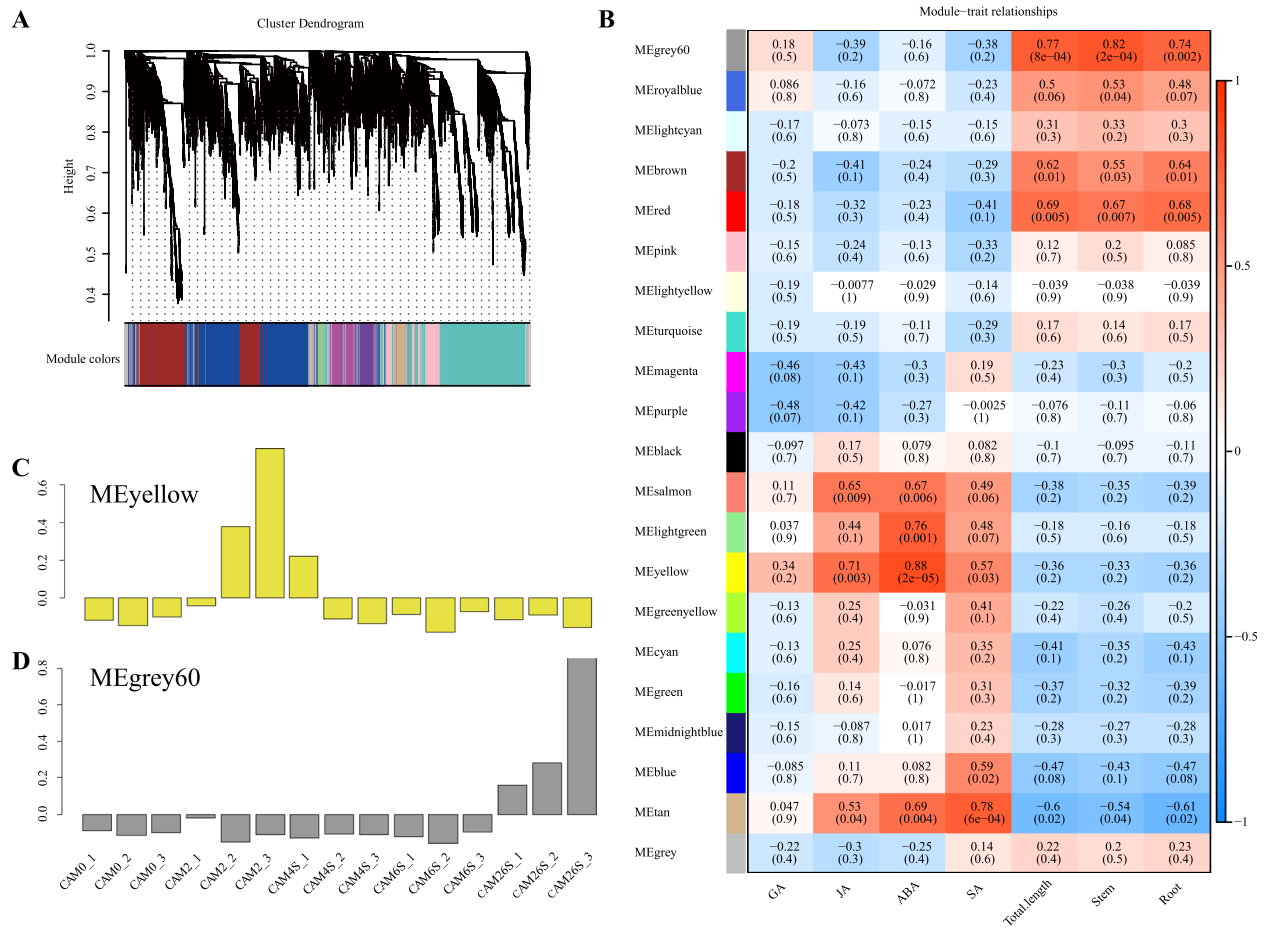

Fig. 6 Weighted Gene Co-expression Network Analysis (WGCNA) of the unigenes change in *C.oleifera*. **A** Hierarchical cluster dendrogram using unigenes identified by WGCNA in different germination stages across 20 co-expression modules. **B** Module-tissue associations. Each row corresponds module. Each column corresponds to a specific tissue. The color of each cell at the row-column intersection indicates the correlation coefficient between the module and the developmental stage. A high degree of correlation between a specific module and the development stage is indicated by the red or blue color. **C** The gene expression patterns in MEyellow module. **D** The gene expression patterns in MEgrey60 module

amino sugar, and nucleotide sugar metabolism, etc. (Additional file 10: Table S9).

WGCNA can also be employed to construct gene networks, in which each node represents a gene and the connecting lines (edges) between genes represent co-expression correlations. Hub genes refer to those that show most connections in the network. In the MEyellow module network, there were 27 genes that encode transcription factors and 4 hormone-related genes. Most of these transcription factor genes were highly expressed during seed soaking and may regulate or participate in activities operating in seed germination (Additional file 10: Table S9).

The 50 most highly connected hub genes in the MEyellow module were used for analyzing the gene expression network. Gene expression showed that the expression level in the stage of seed soaking was higher than that in other stages (Fig. 6B, 7A Additional file 11: Table S10). Co-expression network showed core hub genes, namely, 25556_c1_g1 and 26030_c3_g4. The function of the 25556_c1_g1 gene remains unknown. The 26030_c3_g4(*AT3G29970*) gene 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*), and *CoDJC66*(*29838_c0_g1*) (Fig. 7B). Interestingly, *CoPMT3* (*26833_c0_g1*) is annotated into the protein O-mannosyl transferase gene family (PMT), which is an important sugar proton symporter activity, and produces small ubiquitin-like modifiers (SUMOs). PMTs have differential tissue-specific functions in phosphatidylcholine (PC) biosynthesis and plant growth. As primary enzymes for phosphocholine (PCho) biosynthesis, *PMT3* 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 seedlings (Fig. 6B, 6D). 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, and transmembrane transport (Additional file 10: Table S9). Besides, KEGG analysis showed that those genes were enriched in pathways associated with plant-pathogen interaction, plant hormone signal transduction, phenylpropanoid biosythesis, cyanoamino acid metabolism, limonene and pinene degradation, and brassinosteroid biosynthesis (Additional file 10: Table S9). The hub genes with the maximum number of edges (81) is *CoVIT_13s0047g00260* (28624_c1_g1), a member of the *shiu*: CAMK1 protein kinase family, which 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)], and xylem development [*CoPRX52* (29405_c1_g1)] (Fig. 7C, D, Additional file 10: 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 that participates in cell wall biogenesis. *CoNPF3.1* (40839_c0_g1) is annotated into the 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 the development of buds and the growth of roots [22].

Validation of the expression of DEGs by qRT-PCR

To validate the reliability expression profiling obtained by RNA-seq, 12 DEGs with different expression patterns in cotyledons present in the seed were selected for qRT-PCR analysis. The genes chosen for qRT-PCR analysis included 9 TFs and 3 signaling-related hormones 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. 8A), confirming a high correlation between the RN-seq and qRT-PCR data (Fig. 8B).

Discussion

Cotyledons play a critical role in regulating seed germination and morphogenesis

Seed germination refers to some biological processes, including the radicle prominence of the seed coat and the resumption of embryo growth, accompanied by nutrient metabolism, DNA repair, and transcription [23, 24]. Cotyledon and endosperm, the primary nutrient storage site, play an important role during seed germination. In *Arabidopsis*, the transcriptomes of the endosperm and embryo were quite similar. Thus, the two tissues are likely to express similar genetic programs that are intrinsic to the seed [25, 26]. It can be assumed that genes expressed in cotyledons are similar to those in embryos, and cotyledons act as a critical partner in seed germination and subsequent growth.

This study characterized the gene expression profiling in the cotyledon of *C.oleifera* in different stages during seed germination. Several critical 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 (Additional file 1: Fig. S4). Similar to the process of seed germination involved the

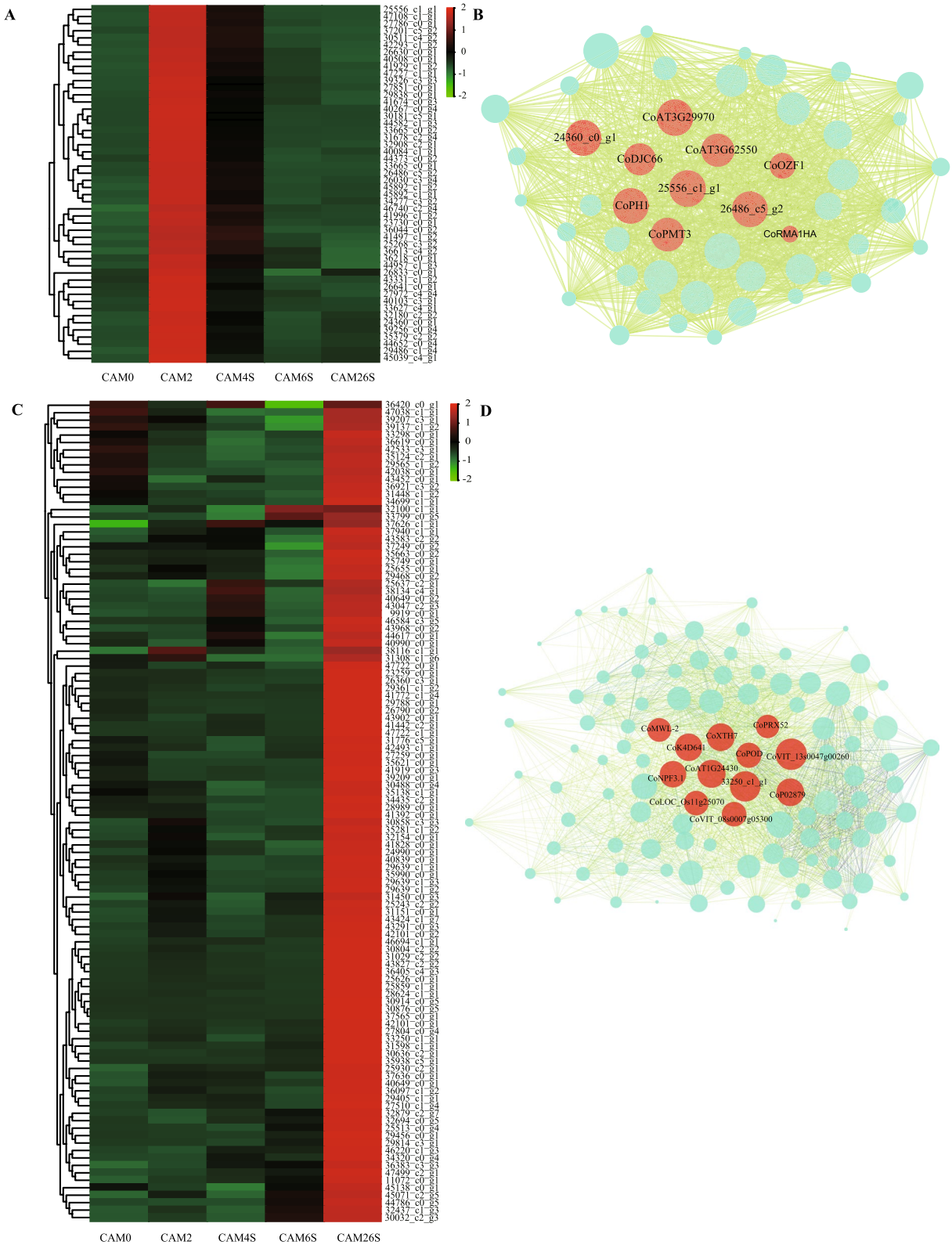

Fig. 7 Analysis of genes in MEyellow and MEgrey60 modules. **A** The heat map of the most highly connected 50 hub genes in the MEyellow module. **B** Co-expression networks of the most highly connected 50 hub genes in the MEyellow module. **C** The heat map shows the relative TPM of each gene in the MEgrey60 module. **D** Coexpression networks of the most highly connected hub genes in the MEgrey60 module

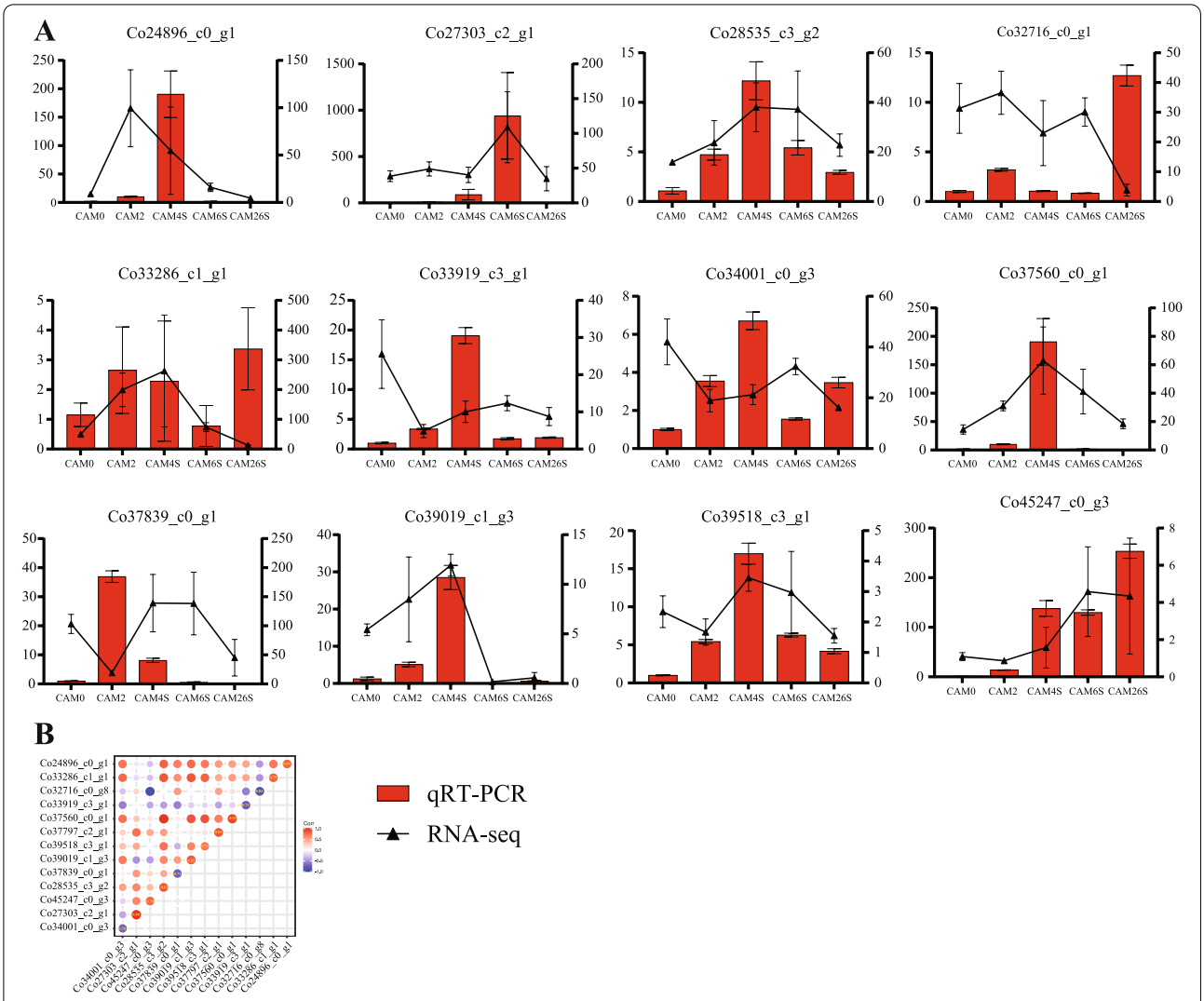

Fig. 8 Correlation between expression profiles of selected genes obtained from RNA-seq and qRT-PCR analysis. **A** Relative expressions of qRT-PCR calculated using *Tublin* and *GADPH* as the reference gene is shown in the left *y*-axis. RNA-seq expression of the unigenes is shown in the right *y*-axis. **B** Heatmaps represent expression profiles of selected genes obtained from RNA-seq (left) and qRT-PCR (bottom) analysis. The color scale at the bottom represents Z-score. The values between the two heatmaps represent the correlation value between the expression profiles obtained from RNA-seq and qRT-PCR analysis for each gene. The correlation values above 0.70 are highlighted in bold

transformation of nutrient metabolism and photosynthesis in other species, many DEGs were identified to be involved in various metabolic pathways. Most of the DEGs were related to the regulation of gene expression, followed by energy production and metastasis (Additional file 1: Fig. S4). 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 overlapping roles in governing the seed germination and the seedling emergence of *C.oleifera*, compared with the whole gene set throughout the seed development.

Candidate hormone associated with *C.oleifera* seed germination

In the present study, many genes were highly expressed in CAM26S, including *GASA3*, *GASA9* *GASA11* and *OS08G0141300*. One particular example is the gibberellin-regulated protein that may function in hormonal development control, such as seed germination, flowering, and seed maturation [27]. Meantime, there were many small auxin up RNA (SAUR) genes highly expressed in CAM26S, including *CoSAUR32*, *CoSAUR36*, *CoSAUR55*, *CoSAUR59,* and *CoSAUR74*. The genes of the SAUR family may not only play a central role in auxin-induced acid growth by regulating cell wall

acidification but also act independently of auxin.Tissue is specifically regulated by various other hormone pathways and transcription factors [28]. Molecular genetic studies in *Arabidopsis* have revealed the functions of several SAUR genes, such as *AtSARU32*, *AtSAUR36* in cell elongation [29]. Furthermore, some SAURs may perform roles in processes other than cell elongation, such as leaf senescence or cell division. Besides, The *SAUR55*, presented in our data (Fig. 4), was previously reported to be able to encode auxin response protein important for plant growth [30].

Distinct stage-specific expression patterns were observed that most of the identified genes were related to AUXIN, GA, and ETH. They were relatively highly expressed in CAM26S compared with the genes related to KARs and BR, which were highly expressed at CAM0 and CAM2 stages (Fig. 3). This suggests the potential roles of AUXIN, GA, and ETH in 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 seed germination and subsequent vegetative growth of *C.oleifera*. The cotyledons are not only the primary nutrient provider but also one of the leading instructors 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 has found about 60 kinds of TF families in higher plants [32], which were involved in regulations of various processes. In the present study, our transcriptome profiling revealed 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 (Additional file 1: Fig. S5). In CAM2, soaking seeds made the sealed seed embryo and cotyledon exchange substance and energy with the outside world and enter into an active metabolic state, which could prompt the seeds to enter the germination stage from the dormant state. Since ABA and GA act as the main plant hormones regulating seed dormancy and germination, the balance of their biosynthesis and catabolism are essential for ensuring the stability of seed dormancy and germination. It has been found in this research that *CobHLH093*(38274_c0_g1) as a transcription factor regulating gibberellin biosynthesis, *CoAREB3*(37687_c2_g6), and *MYB3*(Co31728_c1_g5) as transcription factors promoting seed dormancy and ABA activation, were highly expressed in the CAM2 stage (Fig. 5). This suggests the synergistic effect of two hormone-related genes. In addition, high expression levels 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. 5), genes regulating growth of the stem or root in meristem tissues, were found at this stage. Meantime, *CoSGR5*(30478_c0_g2) and *CoUNE10*(44750_c2_g2), which are involved in root gravity, were also highly expressed. Based on this, the switch of seed germination was triggered during imbibition.

Conclusions

In this study, a de novo assembly of transcriptome data from five stages of *C.oleifera* was performed to provide preliminary insights into the change of cotyledons during seed germination. Enrichment analysis revealed that 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* have higher expression levels in the CAM2 stage (seed soaking). The results indicate that the cotyledon-regulated program for germination had initiated to establish when the seeds were imbibition. The cotyledons play vital roles as the primary nutrient provider and one of the main instructors for seed germination and seedling growth. Taken together, our study will enrich genomic resources of *Camellia* and lay a foundation for further research on molecular mechanisms of development in seed germination and seedling growth of *C.oleifera.*

Methods

Plant materials

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 the CL18 variety were collected from Dongfanghong Forestry, Jinhua City, Zhejiang Province, and stored at 4℃ for 2 months. Matured seeds of the CL18 variety were collected from Dongfanghong Forestry Farm, Jinghua City, Zhejiang Province, China, and stored at 4℃ for 2 months. Then, in dark conditions at 25℃, seeds were flooded using deionized water for 2 days, followed by sand stratification 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 quickly removing the seed coat and the embryo, we collected cotyledon tissues of seeds that were stored at 4℃ (CAM0), soaked in purified water (CAM2), and broken in shell (CAM4S). The cotyledon tissues of seedlings had root growth of 2 cm (CAM6S) or 10 cm (CAM26S) in sand stratification (Fig. 1). The

samples were immediately frozen in liquid nitrogen and stored at -80℃ for subsequent RNA-seq or hormone determination. Each test consisted of three biological replicates. The collection of all the samples complies with institutional, national, or international guidelines and legislation. The local forestry management department authorizes the collection of all samples for this research.

Hormone determination

The content of hormones was determined using liquid chromatography-mass spectrometry (LC–MS) method [13, 34]. Approximately 0.5 g of cotyledon tissues were ground with liquid nitrogen and then suspended in 30 ml of 80% methanol at 4℃ for 24 h in darkness. The extraction mixture was centrifuged at 3500 r /min for 10 min, and the supernatant was collected. After centrifuging, the precipitation was re-suspended in 20 ml of 80% methanol at 4℃ for 1 h. The supernatant merged and evaporated at 40℃ on the rotary evaporator until there was no methanol residual. The remaining water was extracted and decolorized twice with 30 ml of petroleum ether, and the ether was discarded. The aqueous-phase pH 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℃. The dry powder was dissolved in 2 ml methanol, filtered with 0.45 m microporous membranes, and stored in a refrigerator at 4℃. Chromatographic column: Hypersil BDS C18 chromatographic column; methanol–water (containing 0.75% glacial acetic acid, 35:65, V/V) was used in the mobile phase. Detection wavelength: 254 nm; Flow rate: 0.8 ml /min; Injection volume 10 ul; Column temperature: 30℃.

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 analyzed using 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 a isolation from poly(A) mRNA by poly-T oligo attached to magnetic beads (Invitrogen). Following purification, the poly(A)- or poly(A) + RNA fractions were fragmented into small pieces using divalent cations under elevated temperatures. 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, the paired-end sequencing was performed 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 quality of raw reads was controlled 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 grouped 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 genes (DEG)

Salmon [37] was used to perform expression levels for Unigenes by calculating TPM (Transcripts Per kilobase Million) [38]. The DEGs were specified with a threshold of $|\log2(FC)|$ at ≥ 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 analyses of DEGs were implemented by GOseq R packages [40] and KOBAS software [41], respectively. Both analyses were tested at a significance cutoff FDR ≤ 0.05.

Identification of transcription factor

The TF families were identified by BLASTX against known plant TFs identified in the 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 normalized values for each transcript. The DEGs co-expression network was constructed using R package WGCNA(V1.6) with a soft threshold chosen to create networks with a scale-free topology in a way as described by Langfelder and Horvath [42, 43]. After the networks were built, modules of transcripts with similar expression patterns were created, and eigengenes for these modules were calculated. Finally, correlations between these eigengenes and the

content of hormones as well as length of seedling were calculated (Additional file 12: Table S11).

qRT-PCR verification of gene expression

The RNA samples used for qRT-PCR analysis were identical to those for the next-generation sequencing experiments. Single-strand cDNA for each sample was synthesized using the First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA)0.12 representative DEGs were selected to validate 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 (Additional file 13: Table S12). Each qRT-PCR was represented by three biological and three technical replicates. Relative transcript abundance was obtained using the $2^{-\Delta\Delta CT}$ method [45, 46]. All reactions were carried out in 96-well plate in QuantStudio™ Real-Time PCR Software (Thermo Fisher) with the TB Green® Premix Ex Taq™ (Tli RNaseH Plus) (TaKaRa) kit. The amplification procedure is 95 ℃ for 30 s, followed by 40 cycles of 95℃ for 5 s and 60℃ for 30 s.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-022-03651-4.

Additional file 1: Fig. S1. Species distribution of top BLAST hits for matched unigenes sequences. Percentage of unigenes matching the top nine species using Blastx in the NR database. **Fig. S2.** Principal component analysis (PCA) of all samples. Each color on the right indicates the meaning of legend. **Fig. S3.** Histogram presentation of GO classification in different genes. 6820 DEGs were matched by GO terms of three categories. Biological process (blue), Cellular component (green), Molecular function (red). **Fig. S4.** Classification of different genes in the KEGG pathway. The X-axis represents the value of rich factors (the ratio of annotated DEGs to all genes of the enriched pathway). The Y-axis represents the names of pathways. The color depth of each point represents q value. The size of each point represents the number of DEGs. (A) KEGG annotation of DEGs of CAM2 vs. CAM0. (B) KEGG annotation of DEGs of CAM4S vs. CAM2. (C) KEGG annotation of DEGs of CAM6S vs. CAM4S. (D) KEGG annotation of DEGs of CAM26S vs. CAM6S. (E) KEGG annotation of DEGs of CAM6S vs. CAM0. (F) KEGG annotation of DEGs of CAM4S vs. CAM0. (G) KEGG annotation of DEGs of CAM26S vs. CAM0. **Fig. S5.** The differential expression of TFs across five stages. The differential expression of TFs was depicted in heatmap based on Z-score normalized TPM values. The green color indicates lower expressed genes, while the red indicates higher expressed genes.

Additional file 2: Table S1. Sequencing statistics of the transcriptome from five stages.

Additional file 3. Table S2. Statistics of sample sequencing data evaluation and assembly results.

Additional file 4: Table S3. Annotation statistics of unigenes in publicly available databases.

Additional file 5: Table S4. The functional annotation of in unigenes.

Additional file 6: Table S5. The annotation in the eggNOG database.

Additional file 7: Table S6. The different genes in each comparison.

Additional file 8: Table S7. The functional annotation of common expression genes in the Venn diagram.

Additional file 9: Table S8. The functional annotation of GO and KEGG in all different genes.

Additional file 10: Table S9. The functional annotation of all modules in WGCNA.

Additional file 11: Table S10. The function of 50 most highly connected hub genes in the MEyellow module.

Additional file 12: Table S11. The content of hormone in seeds of C.oleifera.

Additional file 13: Table S12. Primer sequence used in qRT-PCR analysis.

Acknowledgements

Data summarized in this paper have been generated through work of several authors and we would like to thank them for their continuous efforts which contribute to the emergence of the *Camellia oleifea*.

Permission to use varieties

The varieties owner has approved us to use CL18 in current research. The local forestry management department authorizes the collection of all samples for this research.

Authors' contributions

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 of the study, data processing, and revision of the manuscript. LL assisted with the interpretation of results and manuscript preparation. XY and YS conceived the study and revised the manuscript. All authors read and approved the 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 data and materials

All used sequencing data are available from the NCBI Sequence Read Archive (SRA) database with accession number SRP349608. The transcriptome assembly data is available from the NCBI Sequence Read Archive (SRA) database with accession number PRJNA824722.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Author details

[1]Zhejiang Provincial Key Laboratory of Tree Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, Zhejiang, China. [2]Department of Hydraulic Engineering, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou 311231, Zhejiang, China.

Received: 23 January 2022 Accepted: 18 May 2022
Published online: 28 May 2022

References

1. Zhuang R. China Oil Tea. 2nd ed. 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):1764–76. https://doi.org/10.1093/pcp/pcx111.
3. Zeng W, Endo Y. Lipid characteristics of *Camellia* seed oil. J oleo sci. 2019;68(7):649–58. 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):1192–208. 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):949–60. 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):710–20. 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):501–23. 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):150–7. 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):1310–26. 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):3703–12. 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):14–7. https://doi.org/10.3969/j.issn.1672-0938.2018.04.004. (inchinese)
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):840–3. 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;9:50–1. 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 DX. Rice homeodomain protein *WOX11* recruits a histone acetyltransferase complex to establish programs of cell proliferation of crown root meristem. Plant Cell. 2017;29(5):1088–104. 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):557–63. 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):1064–75. 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):433–45. 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):1022–35. 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):127–40. 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):15694–9. 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):2366–70. 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. Plant Cell. 2006;18(8):1887–99. 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;32:759–69. 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;157:1342–62. 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):1040–5. 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 MG, 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):644–52. 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, Irizary RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417–9. 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):621–8. 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):139–40. 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;11:14. 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):W720–4. 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, Karishma D'Sa, 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):836–51. 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.

Publisher's Note