



Article

Full-Length Transcriptome Survey and Expression Analysis of *Cassia obtusifolia* to Discover Putative Genes Related to Aurantio-Obtusin Biosynthesis, Seed Formation and Development, and Stress Response

Yin Deng, Hui Zheng, Zicheng Yan, Dongying Liao, Chaolin Li, Jiayu Zhou * and Hai Liao * 

School of Life Science and Engineering, Southwest Jiaotong University, Chengdu 610031, China; DY1099438511@163.com (Y.D.); zh17828115685@163.com (H.Z.); yanzichh@163.com (Z.Y.); ldy20134237@163.com (D.L.); konekoneicl@163.com (C.L.)

* Correspondence: jiaozhou2014@163.com (J.Z.); liaohai0918@163.com (H.L.); Tel./Fax: +86-28-8760-3202 (H.L.)

Received: 28 June 2018; Accepted: 13 August 2018; Published: 21 August 2018



Abstract: The seed is the pharmaceutical and breeding organ of *Cassia obtusifolia*, a well-known medical herb containing aurantio-obtusin (a kind of anthraquinone), food, and landscape. In order to understand the molecular mechanism of the biosynthesis of aurantio-obtusin, seed formation and development, and stress response of *C. obtusifolia*, it is necessary to understand the genomics information. Although previous seed transcriptome of *C. obtusifolia* has been carried out by short-read next-generation sequencing (NGS) technology, the vast majority of the resulting unigenes did not represent full-length cDNA sequences and supply enough gene expression profile information of the various organs or tissues. In this study, fifteen cDNA libraries, which were constructed from the seed, root, stem, leaf, and flower (three repetitions with each organ) of *C. obtusifolia*, were sequenced using hybrid approach combining single-molecule real-time (SMRT) and NGS platform. More than 4,315,774 long reads with 9.66 Gb sequencing data and 361,427,021 short reads with 108.13 Gb sequencing data were generated by SMRT and NGS platform, respectively. 67,222 consensus isoforms were clustered from the reads and 81.73% (61,016) of which were longer than 1000 bp. Furthermore, the 67,222 consensus isoforms represented 58,106 nonredundant transcripts, 98.25% (57,092) of which were annotated and 25,573 of which were assigned to specific metabolic pathways by KEGG. *CoDXS* and *CoDXR* genes were directly used for functional characterization to validate the accuracy of sequences obtained from transcriptome. A total of 658 seed-specific transcripts indicated their special roles in physiological processes in seed. Analysis of transcripts which were involved in the early stage of anthraquinone biosynthesis suggested that the aurantio-obtusin in *C. obtusifolia* was mainly generated from isochorismate and Mevalonate/methylerythritol phosphate (MVA/MEP) pathway, and three reactions catalyzed by Menaquinone-specific isochorismate synthase (ICS), 1-deoxy-d-xylulose-5-phosphate synthase (DXS) and isopentenyl diphosphate (IPPS) might be the limited steps. Several seed-specific CYPs, SAM-dependent methyltransferase, and UDP-glycosyltransferase (UDPG) supplied promising candidate genes in the late stage of anthraquinone biosynthesis. In addition, four seed-specific transcriptional factors including three MYB Transcription Factor (MYB) and one MADS-box Transcription Factor (MADS) transcriptional factors and alternative splicing might be involved with seed formation and development. Meanwhile, most members of Hsp20 genes showed high expression level in seed and flower; seven of which might have chaperon activities under various abiotic stresses. Finally, the expressional patterns of genes with particular interests showed similar trends in both transcriptome assay and qRT-PCR. In conclusion, this is the first full-length transcriptome sequencing reported in Caesalpiniaceae family, and thus providing a more complete insight into aurantio-obtusin biosynthesis, seed formation and development, and stress response as well in *C. obtusifolia*.

Keywords: aurantio-obtusin biosynthesis; *Cassia obtusifolia*; organ; qRT-PCR; seed-specific genes; SMRT sequencing

1. Introduction

The seeds of many pharmaceutical plants, such as *Prunus armeniaca* [1], *Ziziphi spinosae* [2], *Lycium barbarum* [3], and *Lepidium apetalum* [4], have been widely used as traditional medicines for centuries. Many bioactive metabolites, such as alkaloid, terpenoid, flavonoid, and anthraquinone, have been reported in the seeds of plants [5,6]. Furthermore, the seeds are breeding organs in establishment and persistence of medical plants, such as *Panax ginseng*, *Forsythia suspensa*, and *Alisma plantago-aquatica* [7]. Drought, high salinity, and extreme temperature are the major limiting factors of seed formation, development, and germination [8]. Therefore, the genetic background of the seed would underlay foundation for exploring the biosynthesis of bioactive metabolites in the seed, the molecular mechanism of seed formation and development, and stress response of the seed as well.

Cassia obtusifolia L., a leguminous annual plant, is easily grown, widely cultivated in subtropical and tropical countries in Asia and Africa. The seed of *C. obtusifolia*, called as Jue-ming-zi, was introduced in “Shen nong ben cao jing”, the earliest book on Chinese herbal medicine in the world, for the first time in the early 17th century. It was used as medicine and food materials in many traditional Chinese medical (TCM) prescriptions. Meanwhile, it is a healthy tea in China and Korea. In addition, it is used as food for animals and human to supplement the inadequate protein intake in India [9]. For medical purpose, it has been used to treat dizziness and headache, as well as benefit the eyes by anchoring and nourishing the liver. Recently, it has been reported to have neuroprotective effects in brain disease models [10]. The water extract of seed also enhanced antioxidant activity in hyperlipidemic rats induced by a high-fat diet [11]. Moreover, the seed and its constituents had potential antidiabetic activity and the possible mechanism was through increasing the insulin-provoked glucose uptake and inhibiting the protein tyrosine phosphatases 1B and α -glucosidase activities in human HepG2 cells [12]. The seed of *C. obtusifolia* contained a variety of bioactive anthraquinones, which were mainly responsible for the pharmacological action ascribed to them [7,13–15]. It was reported by the Chinese Pharmacopeia in 2015 that aurantio-obtusin, a kind of anthraquinone, was the most significant active gradient and regarded as the quality marker to evaluate the pharmaceutical value of *C. obtusifolia*. As aurantio-obtusin was a specific component found in *Cassia* plants, *C. obtusifolia* might be considered as a kind of model plant to research the biosynthesis of anthraquinone. The phytochemical analysis of *C. obtusifolia* extracts showed that aurantio-obtusin distributed mainly in the seed, while only a few contents of aurantio-obtusin in the root, leaf, flower, and stem. Therefore, it was a reasonable protocol to screen genes that catalyzed the limiting steps involving with the biosynthesis of aurantio-obtusin by comparing the expression difference of candidate genes between seed and other organs.

In addition, the seed of *C. obtusifolia* was harvested and used as the breeding organ. Therefore, it is important to study the seed formation and development of *C. obtusifolia*. Previous genetic analysis indicated that the seed formation and development of plants was controlled by several transcription factors, which were possibly involved in the abiotic stimuli, secondary and hormone metabolisms, and development [16]. Finally, *C. obtusifolia* is cultivated as a kind of landscape plant due to its ability to survive in a variety of environments, and thus its response to various abiotic stresses become a noticeable issue. In higher plants, numerous genes, such as the Hsp20s family, have been reported to respond to these abiotic stresses [17]. These results provided a basis for further analysis of the molecular mechanisms of seed formation and development, and stress response.

To date, the molecular mechanism of biosynthesis of aurantio-obtusin, seed formation and development, and stress response in *C. obtusifolia* remained poorly understood due to the shortage of genomic information. Only seed transcriptome sequence information, based on short-read

next-generation sequencing (NGS) technology, has been obtained from *C. obtusifolia* by our lab [18]. However, the fundamental limitation for NGS was the short sequencing products, which required assembly and led to a very small proportion of the assembled transcripts, and misassembly as well. Furthermore, short reads from NGS made it very difficult for transcripts containing repetitive elements and high GC contents. Therefore, these efforts cannot provide more help to obtain the satisfied genetic background for further research. In recent years, single-molecule real-time (SMRT) sequencing technology, referring to the use of third generation sequencing platforms to sequence cDNA, provided researchers with an efficient way to obtain the full-length cDNA sequences and had been used in non-model plants. Here, the NGS and SMRT sequencing technology were applied to characterize the transcriptome from seed, leaf, flower, stem, and root of *C. obtusifolia*. The transcriptome data would serve as a genetic platform for the researches of *C. obtusifolia*. Furthermore, the identified transcripts provided candidate target genes to discover the biosynthesis of auranthio-obtusin, the molecular mechanism of seed formation and development, and stress response in *C. obtusifolia*.

2. Results

2.1. Combined Sequencing Approach to the Various Organs of *C. obtusifolia*

To identify and differentiate the seed transcriptome from that of the rest organs, NGS and SMRT sequencing platforms were undertaken simultaneously. First, 15 mRNA samples from five different organs (seed, root, stem, leaf, and flower, each in triplicate) were subjected to the Illumina 4000 platform, with 361,427,021 clean reads (108.13 Gb) produced. Second, full-length cDNAs from five pooled poly (A) RNA samples were normalized and subjected to an SMRT sequencing using the PacBio RS II platform. In total, 601,168 polymerase reads were generated. After filtering the adapter sequence from polymerase reads and removing sequences shorter than 50 bp, 4,315,774 subreads with sequence representing 9.66 Gb bases were obtained. Next, 282,022 ROIs were successfully extracted with mean length of 2458 bp, quality of 0.92, and 10 passes. All ROIs were further classified into 136,567 full-length nonchimeric and 114,185 non-full-length sequences with differential patterns of length distribution (Figure S2). Based on the clustering algorithm of IEC, we got 67,222 consensus isoforms with a mean length of 2250 bp, mainly generated from the 1–2 kb, 2–3 kb, and 3–6 kb libraries, as it proved to be too difficult to produce consensus isoforms from the <1 kb and >6 kb libraries, because of their shorter and larger insert lengths. The 67,222 consensus isoforms were composed of 54,533 high-quality transcripts and 12,689 low-quality transcripts which were further corrected using the Illumina reads to improve quality (Table 1). Then, after removing the redundant sequences for all high-quality transcripts and corrected low-quality transcripts using CD-HIT ($c = 0.90$), 58,106 nonredundant transcripts were produced. Besides those coding for proteins, LncRNAs (more than 200 bases) were predicted using CPC, CNCI, pfam, and CPAT analysis, 775 of predicted LncRNAs were able to match four databases. In total, 2712 alternative splicing events were detected from the nonredundant transcripts based on the de novo alignment method.

Our previous NGS result indicated that the unigenes assembled from the Illumina short reads by Trinity largely did not represent full-length cDNAs [18]. Approximately 78.36% of the assembled unigenes from NGS reads were <1000 bases, whereas only 9.23% of the consensus isoforms from the PacBio reads were <1000 bases (Table 1). Indeed, the mean full-length read lengths from the different libraries (<1, 1–2, 2–3, 3–6, and >6 kb) produced by SMRT sequencing were 897, 1371, 2457, 3485, and 9064 bases, respectively (Table 1). Nevertheless, from this study, it seems that the use of NGS data to correct the low-quality SMRT reads may be a proper strategy.

In total, from the NGS data, 40,118, 41,711, 43,288, 42,771, and 45,922 transcripts were found in seed, root, stem, leaf, and flower, respectively. Moreover, using a cut-off of FPKM > 10, expression from 5870 distinct genes was detected in the seed, with 7877 expressed in the root, 9023 in the stem, 8114 in the leaf, and 9082 in the flower. In addition, 658 genes were specially expressed in a single seed organ, 714 in the root, 432 in the stem, 889 in the leaf, and 2629 in the flower (Figure 1). Thus, it is possible to

distinguish between the transcriptomes from each of these various organs. Both SMRT sequencing data and Illumina Hiseq 4000 data have been deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under accession numbers SRP144670.

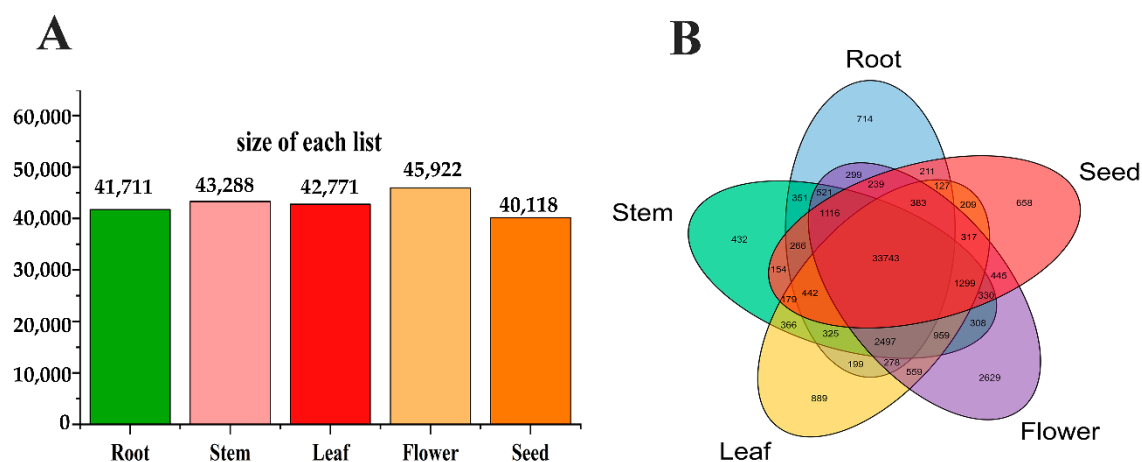


Figure 1. Specially expressed genes across five various organs of *C. obtusifolia*. (A) The gene profiles of five various organs. (B) Venn diagram for the organ-specific transcripts.

Table 1. PacBio libraries and sequencing results.

cDNA Size	Number of Consensus Isoforms	Average Consensus Isoforms Read Length	Number of Polished High-Quality Isoforms	Number of Polished Low-Quality Isoforms	Percent of Polished High-Quality Isoforms (%)
0–1 kb	6206	897	5729	477	92.31%
1–2 kb	25,179	1371	21,937	3242	87.12%
2–3 kb	17,457	2457	14,323	3134	82.05%
3–6 kb	17,616	3485	12,537	5079	71.17%
>6 kb	764	9064	7	757	0.92%
All	67,222	2250	54,533	12,689	81.12%

2.2. Functional Annotation Based on Searches against Public Databases

For annotation, distinct gene sequences were first searched using BLASTX against the Nr, Swiss-Prot, and KEGG databases. The domain/family searches were conducted against the COG, KOG, and Pfam database at NCBI using BLASTX. To functionally categorize the *C. obtusifolia* transcripts based on the Nr annotation, gene ontology analysis was conducted. Using this approach, 57,092 nonredundant transcripts (98.25%) returned a significant BLAST result (Table 2). The annotation ratio of transcripts from SMRT platform was more than that (85%) from NGS platform [19], as more full-length sequences in the former platform supply better Blast results. The numbers of best BLASTX and domain hits for the unigene sequences in each of the databases are summarized in Table 2. A total of approximately 99.28% (56,681) of the *C. obtusifolia* transcripts were found to be homologs in the NR database with an e-value smaller than the cutoff. The annotated sequences had the greatest homology with those in *Glycine max* (17,298, 30.53%), followed by *Cicer arietinum* (6216, 10.97%), *Glycine soja* (6209, 10.96%), *Phaseolus vulgaris* (6133, 10.82%), *Medicago truncatula* (4142, 7.31%), *Theobroma cacao* (1362, 2.40%), *Vitis vinifera* (1272, 2.24%), *Jatropha curcas* (946, 1.67%), *Lotus japonicas* (925, 1.63%), and *Morus notabilis* (828, 1.46%). The remaining 20.00% sequences were matched to other plants. A comparative analysis of the e-value distribution of hit unigenes shows that 75.5% of *C. obtusifolia* transcripts have the highest homology with an e-value cutoff smaller than 1×10^{-100} (Figure 2).

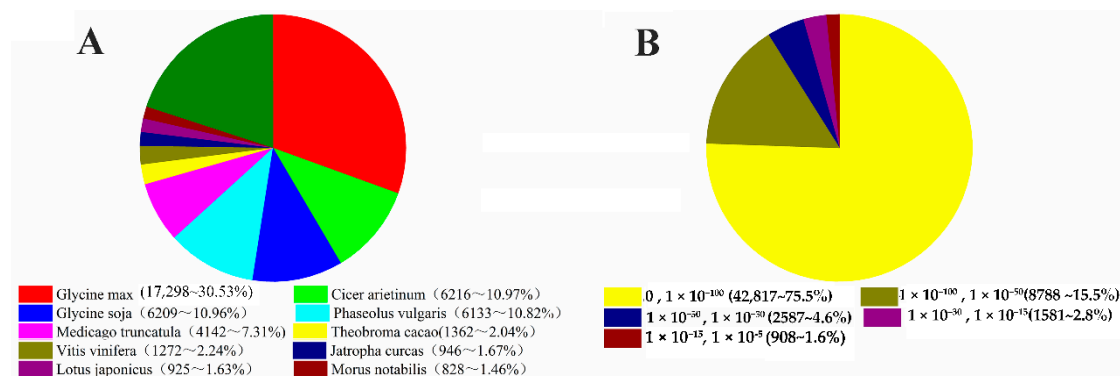


Figure 2. Characteristics of homology search of genes against the NR database. (A) Species distribution is shown as percentage of the total homologous gene hits. All of the top five species are Leguminosae. (B) E-value distribution of BLAST hits for each transcript with a cut-off of E-value of 1×10^{-5} .

Table 2. Summary of database matches for the *C. obtusifolia*.

Anno Database	Annotated Number	$300 \leq \text{Length} < 1000$	$\text{Length} \geq 1000$
COG_Annotation	24,630	1706	22,924
GO_Annotation	40,502	3344	37,158
KEGG_Annotation	25,573	2147	23,426
KOG_Annotation	36,919	2704	34,215
Pfam_Annotation	47,734	3576	44,158
Swissprot_Annotation	43,158	3383	39,775
eggNOG_Annotation	56,156	4598	51,558
Nr_Annotation	56,681	4767	51,914
All_Annotated	57,092	4800	52,292

2.3. Functional Analysis of CoDXS and CoDXR Genes

It was necessary to evaluate the accuracy of sequences from the *C. obtusifolia* transcriptome, two full-length genes annotated as *CoDXS* and *CoDXR*, and were then directly used for functional identification. We used a kind of genetically engineered *E. coli*, which provided a visible color assay to examine the biological function of genes/enzymes involved in MEP pathway, such as DXS, DXR, IPPS, PSY, etc., from various plants including *Amonum villosum* [20], *Vitis vinifera* [21], and *Tripterygium wilfordii* [22]. To confirm the function of *CoDXS*, the plasmids pTrc-*CoDXS* and pAC-BETA were cotransformed into *E. coli* Top10. The cotransformed *E. coli* cells containing pTrc-*CoDXR* and pAC-BETA showed more intense yellow color than the control containing the empty vector pTrc and pAC-BETA (Figure 3A). The results indicated that the *CoDXS* enzymes were functionally active and involved in carotenoid biosynthesis in recombinant *E. coli* cells. *E. coli* harboring the single vector pAC-BETA could not grow due to the lack of ampicillin resistance gene.

Also, the cotransformations were performed using the method mentioned above to demonstrate that the *CoDXR* encodes the anticipative functional protein. The result showed that *E. coli* containing pTrc-*CoDXR* and pAC-BETA accumulated significantly higher saffron yellow color than the control, which was similar to that of *CoDXR* (Figure 3B). These results indicated that the sequences obtained from the hybrid platform were accurate and could be used directly for genetic manipulation.

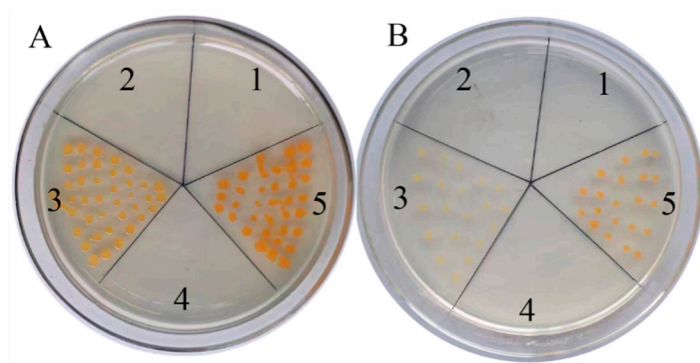


Figure 3. Function demonstration of *CoDXS* (A) and *CoDXR* (B) in *E. coli*. The *E. coli* Top 10 was transformed with pTrc-*CoDXS* and pAC-BETA harboring *CoDXS* (A5) or pTrc-*CoDXR* and pAC-BETA harboring *CoDXR* (B5) pTrc and pAC-BETA (A3, B3), pAC-BETA (A4, B4), pTrc-*CoDXS* (A2), pTrc-*CoDXR* (B2), and pTrc (A1, B1), respectively.

2.4. Gene Ontology (GO) Classification

Based on sequence homology, 40,502 transcripts can be assigned to GO classes to classify the functions of the predicted *C. obtusifolia* genes. The assigned GO terms were summarized into the three main GO categories, biological process, cellular component, and molecular function, and then into 50 functional categories (shown in Figure 4). Biological process comprised 116,967 (47.16%) GO annotations and was the largest cluster, followed by cellular component 78,824 (31.78%), and molecular function (52,256 and 21.07%). Under the molecular function category, catalytic activity (22,830 and 43.69%) represented the most abundant term. About 21,114 and 22,830 transcripts in “metabolic process” and “catalytic activity” category, respectively, suggesting this study may allow for the identification of novel genes involved in the secondary metabolite synthesis pathways. KEGG pathway mapping of *C. obtusifolia* transcriptome assembly resulted in transcripts assigned to KEGG pathway. 1214 transcripts involved in biosynthesis of secondary metabolites were found (shown in Table 3). Among them, 363, 79, and 165 transcripts were annotated from “phenylpropanoid biosynthesis”, “steroid biosynthesis”, and “terpenoid backbone biosynthesis” pathways, which provide precursors for the biosynthesis of auranthio-obtusin.

Table 3. The transcripts related to secondary metabolites.

Biosynthesis of Secondary Metabolites	Transcript Numbers
Brassinosteroid biosynthesis	8
Caffeine metabolism	14
Carotenoid biosynthesis	130
Diterpenoid biosynthesis	37
Flavone and flavonol biosynthesis	6
Flavonoid biosynthesis	177
Isoquinoline alkaloid biosynthesis	51
Limonene and pinene degradation	28
Monoterpenoid biosynthesis	5
Phenylpropanoid biosynthesis	363
Stilbenoid, diarylheptanoid, and gingerol biosynthesis	52
Terpenoid backbone biosynthesis	165
Tropane, piperidine, and pyridine alkaloid biosynthesis	85
Steroid biosynthesis	79
Zeatin biosynthesis	14
Total	1214

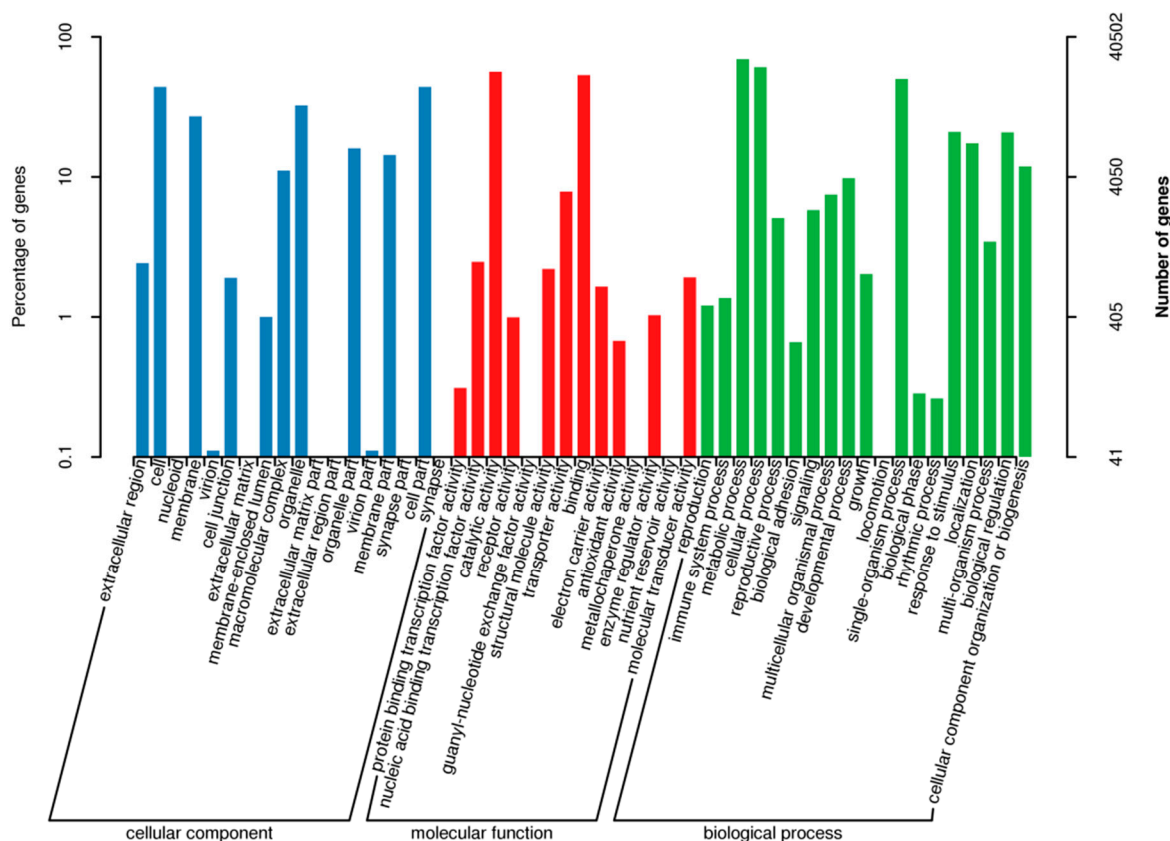


Figure 4. Gene ontology classification. Note: Gene ontology was summarized as three main categories: cellular component (left panel), molecular function (middle panel), and biological process (right panel). The percentage (left Y-axis) and number (right Y-axis) of genes were also showed.

2.5. Localization of Aurantio-Obtusin Accumulation and Gene Profiles Involved in Aurantio-Obtusin Biosynthesis

It is the seed of *C. obtusifolia* that is used in TCM, accounting for a model medicinal herb to study the aurantio-obtusin biosynthesis. Five various organs, including seed, root, leaf, flower, and stem of *C. obtusifolia*, were collected (Figure 5A). The HPLC results indicated that aurantio-obtusin was located in seed (1% of aurantio-obtusin content), whereas root, leaf, flower, and stem did not have aurantio-obtusin compound (Figure 4B). These results suggested that aurantio-obtusin biosynthesis is mainly carried out in this organ, providing a potential basis for colocation analysis.

The biosynthesis of aurantio-obtusin in plants is derived from studies on anthraquinone biosynthesis in other species. Biosynthesis of anthraquinones was studied in the Rubiaceae family such as *Lithospermum*, *Morinda*, and *Rubia* species [23–25]. One of the remarkable features of anthraquinone biosynthesis in higher plants is that they are derived from a variety of different pathways and hence difficult to elucidate. Anthraquinones in plants are derived biosynthetically by a combination of the isochorismate and mevalonic acid (MVA)/2-methyl-d-erythritol 4-phosphate (MEP) pathways [26]. Anthraquinone is made up of three benzene rings named by A, B, and C. In the early stage of anthraquinones formation, the rings of A and B are derived from 1,4-dihydroxy-2-naphthoc acid via isochorismic acid and α -ketoglutaric acid, whereas ring C is derived from isopentenyl diphosphate (IPP)/3,3-dimethylallyl diphosphate (DMAPP) via the MVA/MEP pathway. To date, though a variety of genes involving Isochorismate, MVA, and MEP pathways have been identified in plants, a limited number of genes encoding enzymes in the biosynthesis of anthraquinone have been identified from anthraquinone-containing plants, including DXS encoding 1-deoxy-D-xylulose 5-phosphate synthase from *Morinda citrifolia* [27], DXR encoding, ICS encoding isochorismate synthase from *Rubia tinctorum* [28] and *Rubia cordifolia* [28], OSBS encoding *o*-succinylbenzoate

synthase from *Rubia cordifolia* [29], SK encoding shikimate kinase from *Cassia obtusifolia* [30], OSBL encoding *o*-succinylbenzoate ligase from *Rubia cordifolia* [30], IPP1 encoding isopentenyl-diphosphate isomerase in *Cinchona robusta* [31] and *Rubia cordifolia* [29], and Polyketide synthases III from *Rheum emodi* [32] and *Aloe arborescens* [33]. In the late stage of anthraquinones biosynthesis, the backbone of anthraquinones then undergoes various modifications mediated by cytochrome P450s, SAM-dependent methyltransferases, UDP-glycosyltransferases (UDPG), and other enzymes. To date, no genes involved in the late steps of anthraquinones biosynthesis have been identified from *C. obtusifolia* or other anthraquinone-producing plants.

The biosynthesis of anthraquinone shares isochorismate pathway with phenylpropanoid and MVA/MEP pathway with sterol and terpenoids. Isochorismate pathway leads to produce isochorismate which forms a substrate for 1,4-dihydroxy-2-naphthoyl-CoA, a precursor for anthraquinone backbone. We have identified 33, 34, 34, 33, and 33 transcripts in seed, root, stem, leaf, and flower libraries, respectively, for six enzymes involved in the Isochorismate pathway. Similarly, for four enzymes in the menaquinone pathway, we have identified 42, 45, 45, 44, and 45 transcripts in seed, root, stem, leaf, and flower libraries, respectively. The MVA and MEP pathways produced dimethylallyl diphosphate, another precursor of the anthraquinone backbone. In our study, there were 40, 40, 40, 43, and 47 transcripts in seed, root, stem, leaf, and flower libraries, respectively, for six enzymes involved in the MVA pathway. In the MEP pathway, there were 46, 48, 51, 56, and 54 transcripts in seed, root, stem, leaf, and flower libraries, respectively. Biosynthesis of anthraquinone is also known to be from acetyl-CoA and malonyl-CoA through polyketide pathway in plants. Polyketide synthase III might be a key enzyme involved in the polyketide pathway [34]. In our study, there were 12, 12, 13, 13, and 13 transcripts in seed, root, stem, leaf, and flower libraries, respectively, encoding the enzymes involved in the polyketide pathway. In the late stage of anthraquinone biosynthesis, CYP450s, SAM-dependent methyltransferases and UDPG are mainly responsible for the modification of anthraquinone backbone to produce a variety of anthraquinone compounds [35]. CYP450s are member bound hemoproteins involved in biosynthesis of secondary metabolites such as phenylpropanoid, flavonoid, steroid with sterol, and terpenoids. CYP450s catalyze oxidative reactions, including hydroxylations, epoxidation, dealkylation, dehydration, and carbon—carbon bond cleavage of metabolites. According to the BLAST annotation, 135, 135, 151, 144, and 152 CYP450s were found in seed, root, stem, leaf, and flower libraries encoding for CYP450, CYP450 monooxygenase, and NADPH-CYP450 reductase enzymes, respectively. SAM-dependent methyltransferases make use of the methyl group from SAM and catalyzes the methylation of the hydroxyl group of substrate to produce methyl secondary metabolites, such as phenylpropanoids [36]. Seventy-seven, 83, 80, 76, and 91 SAM-dependent methyltransferases were discovered in seed, root, stem, leaf, and flower libraries, respectively. In general, glycosylation takes place at the end of secondary metabolites biosynthesis and results in increased stability, water solubility, and sometimes biological activity. In nature, UDPG catalyzes glycosylation at the site of hydroxyl group. In our study, there were 82, 90, 96, 96, and 99 UDPGs in seed, root, stem, leaf, and flower libraries, respectively. These CYP450s, SAM-dependent methyltransferases and UDPGs genes are therefore promising candidates for anthraquinone biosynthesis, and further enzyme assays are required to identify the function of these candidate genes (Figure 5C, Table S2).

2.6. Expression Analysis Indicates Colocalization of Aurantio-Obtusin Biosynthesis and Accumulation

Given the clear accumulation of aurantio-obtusin in seed, we hypothesize that aurantio-obtusin biosynthesis occurs mainly in this organ. Accordingly, the remainder of the genes encoding enzymes involved in the aurantio-obtusin biosynthesis might be expected to exhibit a similar (relative) high expression level in seed. Consistent with the localized production of aurantio-obtusin in the seed, analysis of expression levels revealed that 20 transcripts involved in isochorismate pathway were expressed in the seed with FPKM > 10. Of which, at least one isoform of each of the enzymes (DAHPS, DHQS, SDH, SMK, EPSP, CS, ICS, MenE, and MenB) showed somewhat higher (relatively

high) expression in the seed than other organs, whereas DHQS, SDH, EPSP, and ICS showed highest expression level in the seed. In MVA pathway, seven transcripts were expressed in the seed with FPKM > 10. Of which, HMGR also showed somewhat higher expression in seed than other organs, whereas one isoform of ACAT and MPD showed the highest expression level in seed. In MEP pathway, 15 transcripts encoding DXS, DXR, CDPMEK, ISPF, HDR, and IPPS were expressed in the seed with FPKM > 10, of which one isoform of DXS and IPPS, respectively, showed the highest expression level in seed. In polypeptide pathway, three transcripts encoding PKS III and PKC showed high expression with FPKM > 10, but not the highest expression level in the seed. It was believed that there are three limiting steps, which are catalyzed by Isopentenylidiphosphate isomerase (IPPS), 1-deoxy-Dxylulose-5-phosphate synthase (DXS), and isochorismate synthase (ICS), respectively, in the early stage of anthraquinones formation (14–16). Therefore, the observed highest expression levels of ICS, DXS, and IPPS in seed is consistent with those role in anthraquinone production and was verified by followed qRT-PCR that showed similar expression pattern with that in the transcriptome (Figure 6). Moreover, based on the differential expression analysis, ACAT, DXS, EPSP, and ICS were found to be upregulated in the seed (Table S4). As a result, comparing with IPPS, ICS, and DXS might be regarded as the important regulatory target for the aurantio-obtusin biosynthesis in *C. obtusifolia*. In addition, both EPSP and ACAT might be the candidate key enzymes involved in the aurantio-obtusin biosynthesis; the hypothesis still needs verification by further transgenic research.

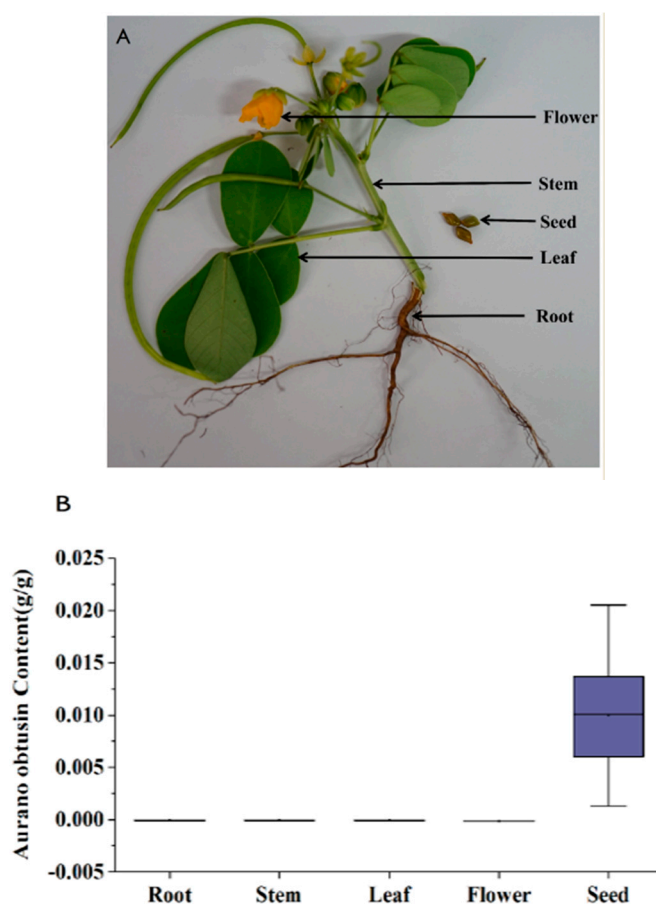


Figure 5. Cont.

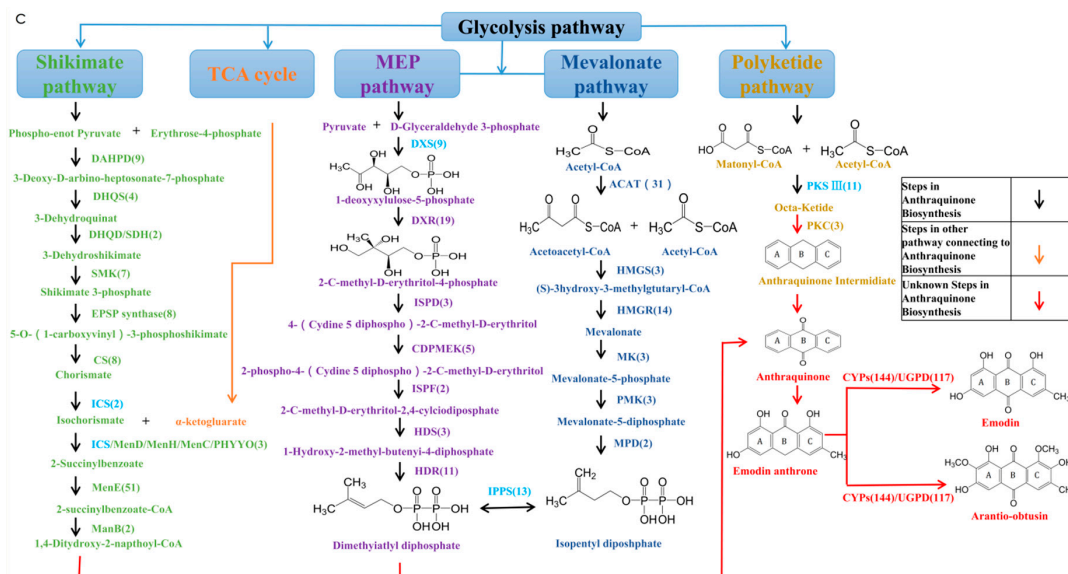


Figure 5. Various organs, aurantio-obtusin biosynthesis-related genes, and phytochemical analysis of *C. obtusifolia*. (A) Five various organs of *C. obtusifolia*. (B) Aurantio-obtusin levels in the five various organs. (C) Proposed biosynthetic pathway and transcripts involved in the biosynthesis of aurantio-obtusin in *C. obtusifolia*. Note: DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; ISPD, 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase; CDPMEK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; ISPF, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; HMGS, 3-hydroxy-3-methylglutaryl CoA synthase; HMGR, Hydroxymethylglutaryl-CoA reductase; MK, Mevalonate kinase; PMK, Phosphomevalonate kinase; MPD, Mevalonate pyrophosphate decarboxylase; IPPS, Isopentenyl diphosphate isomerase; DAHPS, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase; DHQS, 3-dehydroquininate synthase; SDH, Shikimate 5-dehydrogenase; SMK, Shikimate kinase; MenF, isochorismate synthase; PKS III, polyketide synthase III; PKC, polyketide cyclase; MenD, 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase; MenH, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase; MenB, naphthoate synthase; MenC, *o*-succinylbenzoate synthase; MenE, *o*-succinylbenzoic acid-CoA ligase; CYPs included CYP450, CYP450 monooxygenase and NADPH-CYP450 reductase. DMT, di-O-methyltransferase; UDPG, UDP-glycosyltransferase. The transcripts numbers of each gene are represented in the brackets.

CYP450s, SAM-dependent methyltransferases and UDPG are mainly involved in the late stage of anthraquinone biosynthesis. Consistent with the expanded nature of the CYP superfamily in plants, a total of 135 CYPs were identified from the seed of *C. obtusifolia*. Among them, 20 were expressed in the seed with an FPKM > 10, with nine exhibiting upregulated expression levels (and five quite specifically expressed) in seed. In present study, 17 SAM-dependent methyltransferases were found with FPKM > 10, with five members exhibited the highest expression level, in which F01.PB47348 showed specific expression in seed and was annotated as di-O-methyltransferase like protein. Di-O-methyltransferase catalyzed two sequential methylation reactions to form methyl metabolites. Aurantio-obtusin biosynthesis involved two methylation reactions after the formation of emodin, which is suggested to be the precursor of aurantio-obtusin. Therefore, F01.PB47348 was a strong candidate for a possible role in aurantio-obtusin biosynthesis; further functional characterization is necessary. 25 UDPGs were found to be highly expressed in seed with FPKM > 10, with 16 exhibiting the highest expression level (10 upregulated and seven quite specifically expressed) in the seed. Those transcripts that were related to anthraquinone biosynthesis and had high expression levels (with FPKM > 10 in seed) were summarized in Tables S3 and S4. The followed qRT-PCR results of these

seed-specific CYPs, SAM-methyltransferases and UDPGs showed similar expression pattern with those in the transcriptome (Figure 6), except CoUDPG-1, which showed different expression pattern in root, stem, leaf, and flower, in both the qRT-PCR and transcriptome assay.

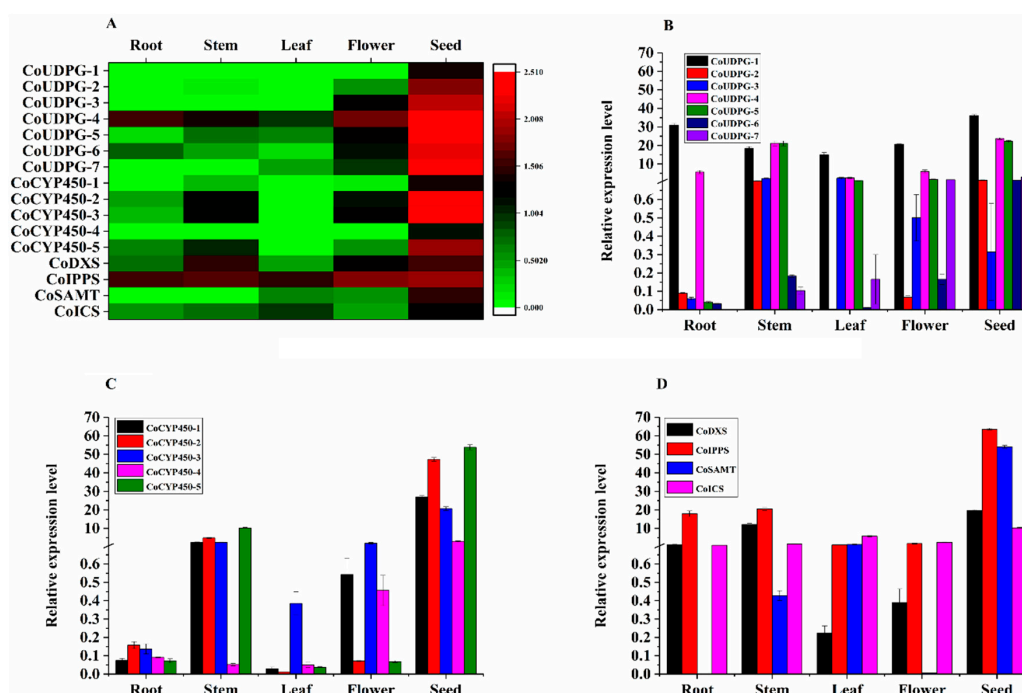


Figure 6. Heat map and following qRT-PCR analysis of the various seed-specific CYPs, SAM-methyltransferases, UDPGs, as well as ICS, DXS, and IPPS with the highest expression level in the seed and different organs, respectively. ICS, DXS, and IPPS were known to be the potential key enzymes involved in the anthraquinone biosynthesis. (A) Heat map depicting differential gene expression of seed. (B–D) qRT-PCR analysis.

Moreover, in order to predict possible function of these 20 CYPs, including 17 CYP450s, two CYP450 monooxygenases, and one CYP450 reductase, phylogenetic analysis was performed together with 85 representative CYP450s from different plant species (Figure 7, Table S5). To obtain an accurate functional prediction of CYPs, only those CYPs with sequence length over 600 bp were selected for phylogenetic analysis. As a result, 17 CYPs (the amino acid of the CYPs were listed in the supplementary data), except 35 (F01.PB1877, 3544, and 15474), were applied with phylogenetic analysis which showed nine out of the CYPs and one CYP450 monooxygenase being grouped in the CYP71-clan. Enzymes from the CYP71-clan are believed to undergo successive gene duplication events resulting in species-specific enzyme-clade, which, in turn promotes the new specialized metabolite biosynthesis in the plant. Within the CYP71-clan, 10 transcripts were clustered in four groups. In group 1, four transcripts (*CoCYP450-1*, 3, 6, and 13) showed high phylogenetic relationship with CYP71A26. The function of CYP71A26 might be involved in the ginsenosides and flavor compounds biosynthesis by catalyzing the C10 hydroxylation of α -pinene to myrtenol, and also in the response to heavy metal and salt stresses [37–39]. Interestingly, *CoCYP450-1* was found to be a potential target gene of LncRNA (F01. PB55451), suggesting F01. PB55451 might play key role in the biosynthesis of secondary metabolites by regulating the mRNA content of *CoCYP450-1*. In group 2, two putative CYP450s (*CoCYP450-9* and 14) were phylogenetically close to CYP71A1. It was demonstrated that the CYP71A1 from *Nepeta racemosa* and *Persea americana* participated in the monoterpenoids biosynthesis and was capable of oxidizing nerol and geraniol to the corresponding 2,3- and 6,7-epoxides [40]. In group 3, *CoCYP450-8* showed close phylogenetic relationship with CYP83B1 that has been shown to catalyze the initial step of indole glucosinolate biosynthesis [41]. Group 4 included

CoCYP450-10, which was phylogenetically related to CYP98A2. CYP98A2 belonged to the CYP98A family that catalyzed the meta-hydroxylation of p-coumarate derivatives, an important step in the phenylpropanoid pathway [42]. The last group in CYP71 clan, group 5 included CoCYP12 and CoCYP450 Monooxygenase-2, showing close phylogenetic relationship with CYP89A2. CYP89A2, a potential target enzyme of WAX INDUCER1 (HvWIN1) transcription factor, was reported to be involved in the pathogen and herbicide resistance by regulating the biosynthesis of the long-chain fatty acid biosynthesis [43,44]. There were two transcripts (CoCYP450-4 and 5) in the CYP72 clan. CoCYP450-4 was close to CYP734A6 and was a crucial brassinosteroid (BR) catabolic gene regulated by a KNOX transcription factor [45]. CoCYP450-5 was close to CYP714A1, a member of CYP72 clan, which has been reported to be involved in the biosynthesis of diterpenoid phytohormone GA by converting GA12 to 12 α -hydroxy GA12 [46]. In the CYP85 clan, CoCYP450-7 and 11 were close to CYP85A and CYP716B2, respectively. CYP85A1 catalyzed several essential reactions necessary for the production of castasterone and was involved in the oxidation steps necessary for the biosynthesis of BR [47]. CYP716B2 was a member of CYP716B family which was functionally characterized as taxoid 9 α hydroxylase in *Ginkgo biloba* [48]. There was only one member in CYP51 and CYP97 clan, respectively. In the CYP51 clan, CoCYP450 Monooxygenase-1 showed a close relationship with CYP51G1. CYP51G1, also named as sterol 14 α -demethylase, catalyzed the first step following cyclization in sterol biosynthesis, leading to the formation of precursors of steroid hormones, including brassinosteroids, in plants [49]. CoCYP450-2 showed a close relationship with CYP97B2, encoding enzymes downstream of β -carotene biosynthesis [50]. CoCYP450 reductase 1 was homologous with cytochrome P450 reductases that provides electrons to CYPs and takes part in the biosynthesis of a huge number of different metabolites, such as sterol and monoterpenoid [51].

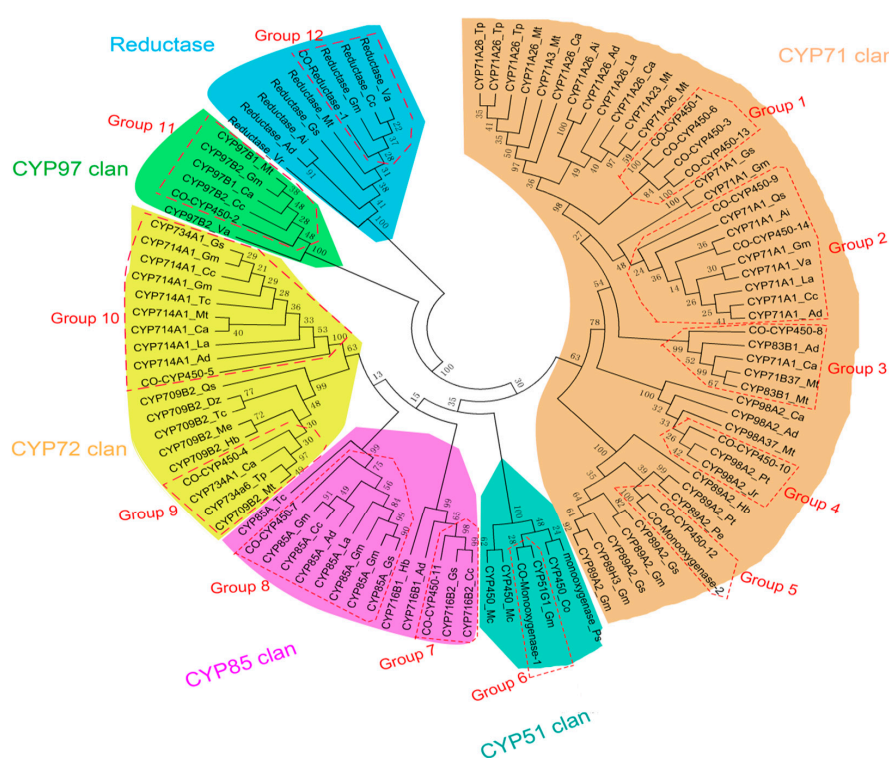


Figure 7. Phylogenetic analysis of seed-specific candidate transcripts coding CYPs associated with the late stage of anthraquinone biosynthesis with cytochrome P450s across different plant species. The nucleotide sequences of seed-specific candidate transcripts were translated into protein sequences, and subjected to multiple sequence alignment with 85 selected plant CYPs from different subfamilies (available online: <http://drnelson.uthsc.edu/CytochromeP450.html>).

2.7. Expression Pattern of Transcription Factors (TFs) in Seed

There were 303 TFs with FPKM > 10 in seed, mainly including 101 Zinc finger, 24 MYB, 24 ARF, 19 NAC, 18 Homeobox, 18 bHLH, and 15 NF TFs (Table S3). Zinc finger TFs possessed the zinc finger motif and played an extensive role in seed germination, secondary wall formation, and response to heat stress. MYB TFs were involved in controlling various processes like responses to biotic and abiotic stresses, development, differentiation, secondary metabolism, and defense [50,52,53]. In plants, the Auxin response factor (ARF) transcription factor family regulates gene expression in response to auxin and plays essential part in the root and gametophyte [54,55]. NAC TFs regulated a wide variety of processes in seed size, secondary wall biosynthesis, and secondary metabolites biosynthesis [56–58]. Homeobox TFs were important regulators of meristem function and were regulated tightly in a complex network of TFs [59]. WRKY TFs had diverse biological functions in seed and trichome development, embryogenesis, as well as additional developmental and hormone-controlled processes [60]. bHLH TFs were widely distributed in eukaryotic organisms and were thought to be one of the largest families of regulatory proteins, which played crucial roles in plant development [61]. NF TFs had been implicated in endosperm development, establishment of the legume-rhizobia symbiosis, and responding to abiotic stresses [62–64]. Meanwhile, there might be a role for alternative splicing in regulating seed formation and development more generally (Figure S3). First, nine ZF TFs exhibited alternative splicing, with only one isoform expressed with FPKM > 10 in seed. In addition, one SBP and one bHLH TFs were observed with only one isoform with FPKM > 10 in seed, suggesting that the regulation of alternative splicing may play a role in controlling the seed formation and development processes (Table S3).

Of note, 58 TFs showed the highest expression level in the seed, mainly including 18 Zinc finger, 6 MADS, 5 MYB, 4 GATA, 4 NF, 3 ARF, 3 NAC, 3 LEA, and 3 Trihelix TFs (Table S3). Among them, 19 TFs were specifically expressed in seed, including 3 MYB, 3 MADS, 3 LEA, 2 YABBY, 2 Zinc finger, 1 ARF, 1 NF, 1 HBP, 1 NAC, 1 bHLH, and 1 WRKY TFs. Besides the TFs described above, MADS played crucial role in nearly all aspects of plant development and contributed to the increasing complexity of plants [65]. Late embryogenesis abundant (LEA) protein and Zinc finger TFs regulate seed germination together [66]. YABBY-like and LOL TFs regulate fruit size together [67]. GATA TFs, which are responsive to various environmental change, regulated seed and leaf development and differentiation [68–70]. Trihelix TFs were plant-specific transcription factors, which are known to have a wide range of functions in growth and development processes involving the flower, stomata, embryo, and seed, as well as abiotic and biotic stresses [71,72]. HB29 TFs, which improved the transcriptional activities of MED25 and NAC TFs, could increase drought tolerance by upregulating several stress-inducible genes [73]. A heatmap of 19 seed-specific TF genes, represented by FPKM values in different organs, was established by R-software and validated by further qRT-PCR. As a result, 18 seed-specific TFs, except the NAC TF, showed similar expression pattern with those in the transcriptome (Figure 8).

Furthermore, those 18 verified seed-specific TFs were assigned to GO analysis to classify the predicted functions and six TFs were annotated and assigned at least one term in molecular function, cellular component, and biological process categories. The six TFs were further classified into 19 functional terms, providing an overview of ontology content (Figure 9A). In the biological process category regulation of transcription (GO:0006355) was the most highly represented term, while sequence-specific DNA binding transcription factor activity (GO:0003700) dominated the molecular function category, indicating those six seed-specific TFs which possessed extensive transcriptional regulatory activities. Among them, CoMYB1 (F01.PB3829), CoMYB2 (F01.PB48100), CoMYB3 (F01.PB6578), and CoMADS2 (F01.PB409), were suggested to be the strong candidate TFs involved in seed formation and development. F01.PB3829 might participate in organ boundary specification between lateral organs and the meristem (GO:0010199). F01.PB48100 and F01.PB6578 might take part in the vasculature (GO:0001944), xylem (GO:0010089), seed coat development (GO:0010214), and regulation of stomatal movement (GO:0010119) processes together. F01.PB409 was suggested to regulate plant-type cell wall modification (GO:0009827) and post-embryonic morphogenesis (GO:0009886) process (Figure 9B).

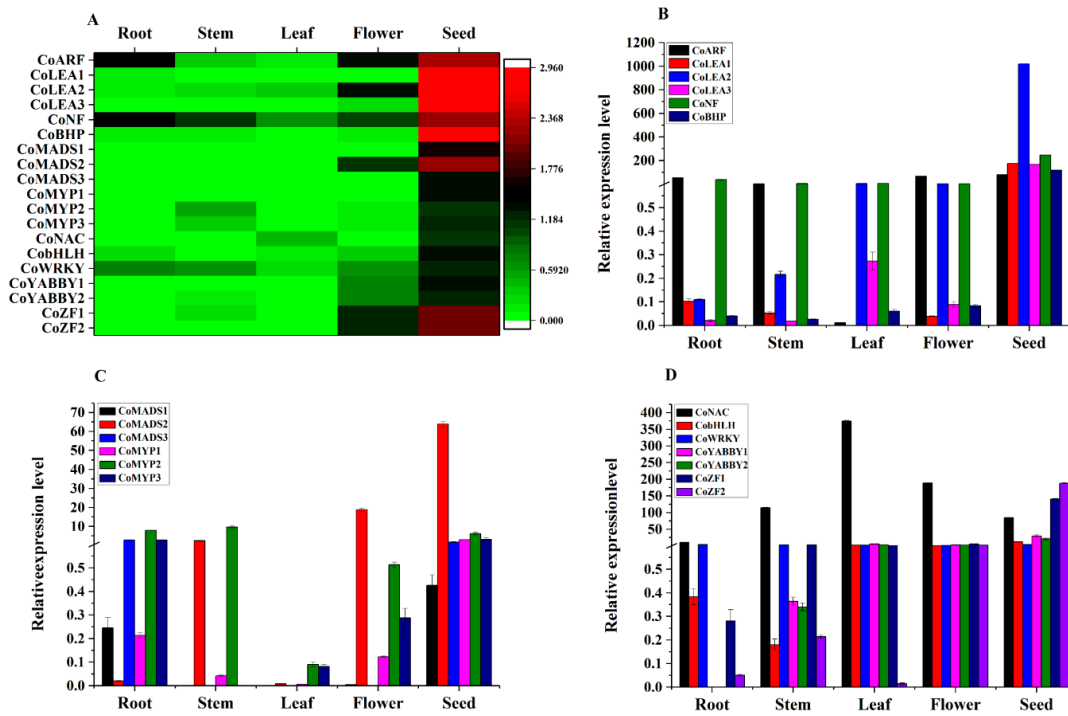


Figure 8. Transcriptome and following qRT-PCR analysis of the TFs in different organs including seed, root, stem, leaf and flower, respectively. (A) Heat map depicting of seed-specific TFs. (B–D) qRT-PCR analysis of seed-specific TFs.

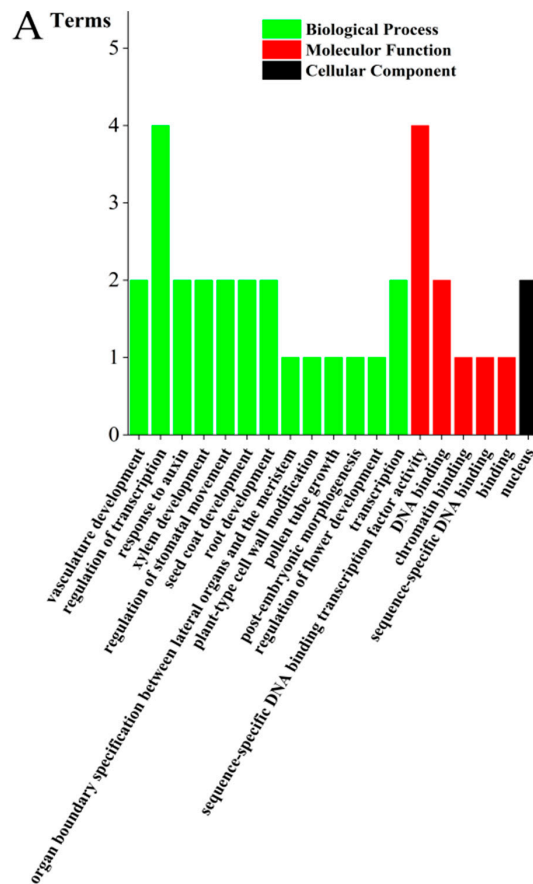


Figure 9. Cont.

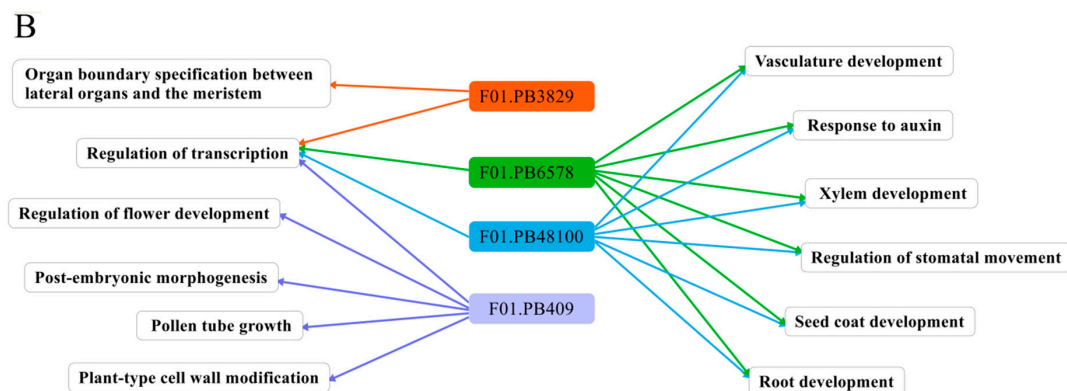


Figure 9. Gene ontology classification of seed-specific TFs. (A) The results were summarized in three main categories: cellular component, molecular function, and biological process. (B) The biological process of four TFs related to seed formation and development. F01.PB3829, F01.PB48100, F01.PB6578, and F01.PB409 represented CoMYB1, CoMYB2, CoMYB3, and CoMADS2, respectively.

2.8. Expression Pattern of CoHsp20 Genes in Various Organs and under Different Stress

It was regarded that abiotic and biotic stress would affect seed development and germination [17]. It was shown that small heat shock proteins (HSP20s) were a group of stress-responsive proteins mainly expressed in the seed and flower [74]. The HSP20s were ATP-dependent small chaperons with molecular weight ranging from 15 kDa to 42 kDa [75]. Hsp20s could avert protein denaturation, and thus maintain cellular homeostasis during abiotic stresses [76]. Hsp20s belonged to the heat shock protein family, including not only Hsp20s, but HSP100s, HSP90s, HSP70s, and HSP60s. Unlike other HSPs, the HSP20s exhibited extensive sequence variability, but a highly conserved 80–100 long amino acid sequence termed as α -crystalline C-terminal domains (ACD) [77]. The ACD contained two conserved regions, one in the N-terminal consensus region and the other is connected through a hydrophobic α 6-loop at the C-terminal common region [78]. Hsp20s played an important role in plant stress response, such as heat, drought, salt, and cold. In this study, eight putative Hsp20 (CoHsp20) genes were identified from the transcriptome of *C. obtusifolia*. A heatmap of eight CoHsp20 genes, represented by FPKM values in different organs, was established by R-software (Figure 10A). Most of CoHsp20s were expressed in one organ at least, except CoHsp20-5, which was barely expressed in any organ. Some CoHsp20 genes showed similar expression patterns in various organs. CoHsp20-1, 4, and 8 were specifically expressed in the seed. CoHsp20-6 and 7 showed relatively high expression high levels in seed and flower. CoHsp20-3 showed high expression level in most of organs, but relatively low expression level in the seed. CoHsp20-2 showed the highest expression level in the stem, but a lower expression level in seed and leaf. Among the eight CoHsp20 genes six (except CoHsp20-2 and 5) showed higher expression level in the seed and flower, suggesting their protective roles in the stress-sensitive reproductive organs. In order to demonstrate the reliability of the CoHsp20 genes expression profiles in the *C. obtusifolia* transcriptome, the qRT-PCR was used to investigate the relative expression level of eight CoHsp20 genes in five various organs using *EF1 α* as reference gene. As a result, expression trends were consistent for most of CoHsp20s, except CoHsp20-6, which showed different expression pattern in flower and stem, in both the qRT-PCR and transcriptome assay (Figure 10B).

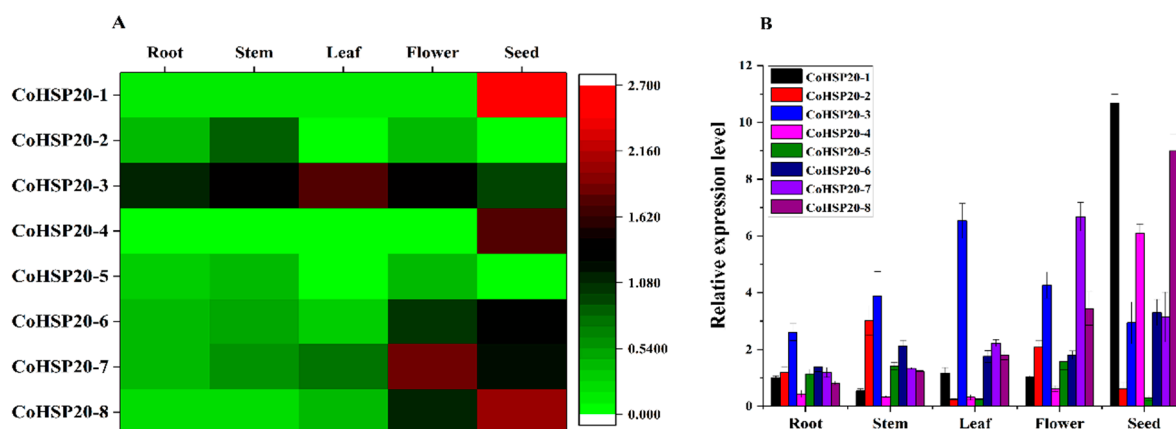


Figure 10. Heat map and following qRT-PCR analysis of the Hsp20 genes in different organs. (A) Heat map depicting differential gene expression levels. (B) qRT-PCR analysis.

Meanwhile, in order to understand the metabolism of the eight Hsp20 genes under stresses, the sterile seedlings (30 days after germination) were treated with different stresses including heat, salt, drought, cold, ABA, and MeJA, and then used as samples to analyze the expression pattern of the eight Hsp20 genes by real-time RT-PCR. Generally, the relative expression level of the CoHsp20 genes under all stresses fluctuated during 1 to 48 h treatments (Figure 11). Most of the CoHsp20 genes were upregulated to heat stress, but CoHsp20-6 was downregulated after 24 h heat treatment. The relative expression levels of three CoHsp20 genes (CoHsp20-1, 4, and 7) were extremely upregulated (more than 100 fold) after 1 h heat treatment compared with control.

Although the Hsp20 family is generally induced by heat stress, we also investigated whether the family was involved in response to salt, drought, cold, ABA, and MeJA treatments. In general, the family was less sensitive under salt, drought, cold, ABA, and MeJA stresses than that under heat stress. Under salt stress, CoHsp20-2 was evidently regulated (63-fold) after 24 h treatment. CoHsp20-1, 4, 5, and 8 reached the highest expression level after 48 h, 24 h, 6 h, and 24 h treatment, respectively, whereas CoHsp20-3 and 7 were downregulated after salt treatment. Only CoHsp20-2, 5, and 8 showed medium upregulated expression levels (2.89, 2.89, and 8.49 fold, respectively) under drought stress. CoHsp20-3, 4, and 7 were downregulated and CoHsp20-6 was not sensitive to drought stress. After cold treatment, CoHsp20-1, 2, 4, and 5 were upregulated (14.66, 6.86, 5.18, and 3.77 fold, respectively) after 48 h, 24 h, 6 h, and 6 h, respectively, compared with control. CoHsp20-3 and 6 were downregulated and CoHsp20-8 was not sensitive to cold stress. Six genes (CoHsp20-1, 2, 4, 5, 7, and 8) showed medium upregulated expression levels (5.05, 2.82, 7.91, 4.02, 2.25, and 7.06, respectively) under ABA treatment after 24 h, 24 h, 24 h, 48 h, 48 h, and 12 h, respectively, but CoHsp20-3 and 6 showed no differences. CoHsp20-1, 2, and 5 showed upregulated expression levels (11.55, 12.90, and 14.18 fold, respectively) after 24 h treatment with MeJA. CoHsp20-4 reached the highest expression level (3.88 fold) after 1 h treatment, whereas CoHsp20-3, 6, 7, and 8 were downregulated under MeJA treatment. The differential expression patterns compared with those under heat stress indicated there were different response mechanisms of the CoHsp20 family under various abiotic stress conditions.

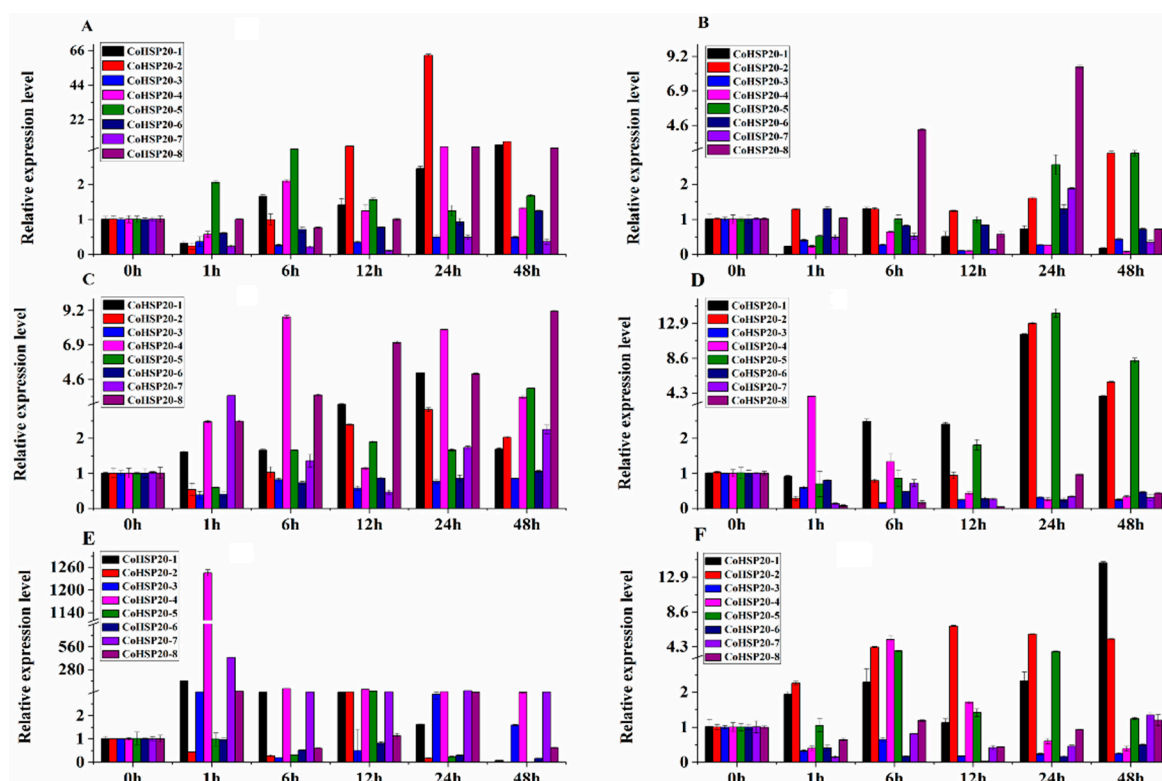


Figure 11. The relative expression level of various CoHsp20 genes under heat (A), salt (B), drought (C), cold (D), ABA (E), and MeJA (F) treatment.

3. Discussion

The long-history and widespread use of Juemingzi has led to intense interest in the biosynthesis of relevant bioactive compounds, seed formation and development, and to the molecular mechanism of the stress response as well. Aurantio-obtusin was the natural product of pharmaceutical importance and therefore regarded as the quality marker of *C. obtusifolia*. In the present study, seed-specific location of aurantio-obtusin was observed and therefore suggested that the seed was an important organ for the study of aurantio-obtusin biosynthesis and transport. Higher concentration of the purine alkaloid in Guarana seeds was reported [79]. Higher aurantio-obtusin content in the seed also suggested, as in other higher plants containing secondary metabolites, that the aurantio-obtusin precursors could be involved in some protective functions [80]. With an aim to understand the genes leading to the aurantio-obtusin biosynthesis, we carried out a short-read NGS transcriptome of the *C. obtusifolia* seed previously [81]. Meanwhile, a few NGS transcriptome databases have been obtained from *Cassia angustifolia* [82,83] and *Cassia bispularis* [84] in the Caesalpiaceae family, which belong to the Cassia genus. However, this short-read NGS study was limited by either number and/or length of the generated sequence information. In order to obtain full-length cDNA sequences, necessary further traditional gene cloning, which has been proven to be expensive, time-consuming, and inefficient, was required. The limited resources cannot provide more help to study *C. obtusifolia*.

Recently, the long-read SMRT sequencing was discovered for obtaining high percentage of full length cDNAs from the cDNA library. Moreover, the hybrid approach, combining both short-read NGS and long-read SMRT sequencing, led to the improvement of full length cDNAs of high quality, and thereby provided more accurately integrated analysis of transcriptomes than NGS. To date, hybrid transcriptome analysis followed by identification of potential candidate genes will lead to a better understanding of secondary metabolic pathways, the molecular mechanisms of organ formation and development, and the molecular mechanism of stress response as well in plants, such as *Astragalus membranaceus*, *Dendrobium officinale*, *Salvia miltiorrhiza*, *Triticum aestivum*, and Maize [18,85,86]. The transcriptome study of the

pathways leading to production of natural products such as aurantio-obtusin will help to manipulate pathways in plants and reconstitute plant pathways in microbial hosts. In this study, the 108.13 Gb NGS data and 9.66 Gb SMRT data of various organs provided the first comprehensive insight into the various organs of *C. obtusifolia*, and might serve as the genetic background for *C. obtusifolia* to improve the basic biological research of *C. obtusifolia*. Interestingly, this is the first full-length transcriptome sequencing from plants in the Caesalpiniaceae family. More than 98.25% (57,092 of 58,106) nonredundant transcripts were annotated by sequence similarity search in public databases. The percentage of the annotated transcripts was far more than those in the former NGS data from *C. obtusifolia* (85%) [87], *Ligusticum chuanxiong* (61.54%) [16], and parsley (60.84%) [88]. Two genes with full-length sequences from transcriptome, *CoDXS* and *CoDXR*, were directly used to construct recombinant vectors, transferred into *E. coli* and proven to encode functional protein by color complementation assay. The results indicated that the full-length sequences were accurate and could be used directly for functional identification and further genetic manipulation. Hybrid transcriptome sequencing, as a high throughput as well as cost-effective approach of sequence determination, has dramatically improved the quality, efficiency, and speed of gene discovery.

C. obtusifolia is widely known for its pharmaceutical important anthraquinones and hence, gaining insights into the transcriptional regulation of anthraquinone biosynthesis in general could accelerate the engineering of this pathway for production of high anthraquinone content in the future. Using GO and KEGG annotation, we have identified large number of transcripts involved in metabolism, genetic information processing, environmental information processing, cellular processes, and organizational systems. All these transcripts are important resources for genetic manipulations of *Cassia* in the future. In general, anthraquinones from higher plants are derived through combination of isochorismate and MVA/MEP, and also the polyketide pathway. In the former pathway, the backbone of anthraquinones is synthesized via the isochorismate and MVA/MEP pathway. Most of the genes encoding enzymes involved in the isochorismate and MVA/MEP pathway were present in the transcriptome of *C. obtusifolia* in our study. There was more than one transcript assigned to the same enzyme. These results also demonstrated the powerful ability of high-throughput sequencing to identify genes in metabolic pathways. Such transcripts may represent different parts of a single gene, different members of a gene family, or both. Twenty-one (including DAHPS, DHQS, SDH, SMK, EPSP, CS, ICS, MenE, and MenB), seven (including HMGR, ACAT, and MPD), and 15 (including DXS, DXR, CDPMEK, ISPF, HDR, and IPPS) transcripts were found with FPKM > 10 in isochorismate, MVA, and MEP pathways, respectively. There were three rate limiting steps, which were catalyzed by IPPS, DXS, and ICS, respectively, in the early stage of anthraquinone formation [18,20,86]. Besides IPPS, DXS, and ICS, we also found that genes encoding DHQS, SDH, EPSP, ACAT, and MPD that showed the highest expression level in the *C. obtusifolia* seed, which suggested that these steps might be rate-limiting in the formation of dimethylallyl diphosphate leading to anthraquinone formation. These genes formed likely candidates for genetic manipulation of anthraquinone biosynthesis in *C. obtusifolia*. Functional characterization of the candidate genes will not only help elucidate the biochemical mechanism for life saving compounds biosynthesis, but also provide a molecular and biochemical target for improving the content of these compounds in future. Further enzyme assays of these enzymes are required to identify the function of the candidate genes. However, although three transcripts, including PKS III and PKC, were found with FPKM > 10 in the polyketide pathway in the seed, none of them showed the highest expression level in the seed, suggesting the anthraquinone might mainly be biosynthesized from isochorismate and MVA/MEP pathway in *C. obtusifolia*.

CYP450s, SAM-dependent methyltransferases, and UDPGs mainly mediated many modifications of the backbone of anthraquinones in the late stage of anthraquinone biosynthesis. In plants, CYPs catalyzed the addition of the oxygen atom to the metabolites and many of them were involved in plant secondary metabolism. SAM-dependent methyltransferases catalyzed the methylation of the hydroxyl group of substrate to generate methyl metabolites. UDPGs catalyzed glycosylation at the site of hydroxyl group to produce glycosylated metabolites. In total, 135, 92, and 82 transcripts in seed were identified as putative CYPs, SAM-dependent methyltransferases and UDPGs, respectively, using

BLAST search. In the former seed transcriptome *C. obtusifolia* [18], only 30 CYPs, 12 SAM dependent methyltransferases, and 14 UDP-glucosyltransferase unigenes were identified [89], the number of these genes was far lower than those found in this study. These results also demonstrated the powerful ability of SMRT sequencing to identify genes in metabolic pathways. In present study, 21 CYPs, 17 SAM-dependent methyltransferases, and 25 UDPGs were found with FPKM > 10. Among them, nine CYPs, five SAM-dependent methyltransferases, and 16 UDPG exhibited the highest expression level in seed. Moreover, five CYPs, one SAM-dependent methyltransferase, and seven UDPGs were seed-specific. Therefore, further studies on higher expressed and seed-specific CYPs, SAM-dependent methyltransferases, and UDPGs genes are needed to better understand roles of these genes in the formation and development, and biosynthesis of anthraquinone in the seed; these genes could be applied to the genetic improvement of *C. obtusifolia* plants.

In addition, the potential function of several candidate CYPs was predicted by the polygenetic tree. The CYPs involved in brassinolide biosynthesis, including *CoCYP450 Monooxygenase-1*, *CoCYP450-5* and *8*, and monoterpene biosynthesis, including *CoCYP450-10*, *13*, *16*, and *CoCYP450 reductase 1* were suggested to be promising candidates associated with anthraquinone biosynthesis. In the late stage of anthraquinone biosynthesis, the backbone was hydroxylated at 1C, 2C, 3C, 6C, 7C, and 8C positions to produce various anthraquinone compounds. Therefore, *CoCYP450-10*, *13*, and *16* might be regarded as strong candidate CYP450s for the modification of anthraquinone backbone, and CYPs in group 2 are potential genes for further functional characterization.

Differentially (especially organ-specifically or tissue-specifically) expressed TFs might play an important role in recruiting nondifferentially expressed TFs to the TF—TF interaction network, offering the potential for coordinating and controlling organ/tissue gene expression across a variety of conditions [90]. The 18 verified seed-specific TFs, especially the four TFs (F01.PB3829, 48100, 6578, and 409), provided regulatory network concerning the formation, development, secondary metabolism, and stress response of the seed. Interestingly, this was the first report that YABBY and HBP TFs expressed specifically in seed. Three MYB and one MADS TFs were suggested to be the strong candidate TFs involved in seed formation and development, according to Go annotation. Both MYB and MADS TFs known to regulate organ formation and development play important role in control of seed formation and development [91,92]. Meanwhile, alternative splicing might regulate seed formation and development at the post-transcriptional level. The alternative splicing took part in several biological processes, such as plant hormone signal transduction and sugar and carbon metabolism, and play an important role in barley seed germination [93]. To date, regardless of whether or not the above genes were involved with the seed formation and development needs further transgenic evidence, those expressing a high level in the seed indicated that they might play essential role in seed formation and development. The further qRT-PCR analysis showed that the gene expression profiles of TFs were consistent with those in comparative transcriptome analysis. These results indicated that it was effective to analyze the gene expression profiles in various organs by biological repeats in high-throughput sequencing technologies and qRT-PCR.

Previous research identified 51 Hsp20 genes in the genome of soybean; most members of which were induced by stresses [94]. The low number of Hsp20 genes in transcriptome of *C. obtusifolia* is due to the fact that it grew under normal condition. The expression patterns of Hsp20 genes in different tissues and organs have been described in Arabidopsis, rice, soybean, and potato [18,95,96]. There was no uniform gene expression pattern of plant Hsp20 genes, suggesting different Hsp20 proteins might have diversified functions in the growth, development, and stress response of the plant. According to the transcriptome data of *C. obtusifolia*, six genes, mostly *CoHsp20* genes (75%), displayed relatively higher expression levels in the seed and flower. Similar to several Hsp20 genes in *Panicum virgatum*, 60% members showed specifically expressed in seed and flower under normal condition [97]. Such expression profiles were also recorded with rice and Arabidopsis Hsp20s [98,99]. For example, eight Arabidopsis Hsp20 genes were specifically expressed in leaves, and some rice Hsp20s were specifically accumulated in seeds. Considering that the flower and seed are important reproductive

organs, these relative high expression levels of CoHsp20 genes in reproductive organs, even without stress, indicated that these CoHsp20s played vital roles in maintaining cellular homeostasis during meiosis, fertilization, and seed setting. Several CoHsp20 genes, such as *CoHsp3* and *CoHsp5*, exhibited quite different expression patterns in various organs, indicating that different CoHsp proteins may have diverse functions. Furthermore, qRT-PCR was used to investigate the transcript levels of each CoHsp20 under different abiotic stresses. *CoHsp20-6* with high expression levels in the seed was not induced under various stresses. Thus, we might assume that *CoHsp20-6* is lacking of chaperone activity. Similar expression patterns in five genes (*CoHsp20-1*, 3, 4, 7, and 8) under heat stress might be caused by shared induction mechanisms. It was reported that the expression of Hsp20 family was controlled by shock transcription factors [100], thus the similar expression pattern of these five CoHsp20 genes under heat stress might be attributed the similar upstream regulating genes of heat shock transcription factors. Moreover, *CoHsp20-1*, 4, 7, and 8 were upregulated under other abiotic stresses; however, *CoHsp20-3* was not sensitive to other abiotic stresses. The differences might be due to the fact that various *cis*-elements existed in promoter regions of different CoHsp20 genes that were involved in the responses of CoHsp20 genes to other abiotic stresses. Therefore, the various responses of CoHsp20 genes reflected a complicated interconnected induction mechanism involving both *cis*-elements and heat shock transcription factors.

4. Methods and Materials

4.1. Plant Materials

The *C. obtusifolia* was cultivated in a controlled environment growth chamber under a 16 h/8 h photoperiod at 25 °C/16 °C day/night cycle. The seed, root, stem, flower, and leaf of *C. obtusifolia* 40 days after flowering (DAF) were collected. All the organs were immediately frozen in liquid nitrogen and stored at −80 °C for RNA isolation in the future. Fifteen total RNA samples (five different organs with three repetitions) were isolated using the RNeasy Plus Mini Kit (Qiagen Corporation, Hilden, Germany). The total RNA samples were treated with Dnase I to remove DNA contaminant. RNA quantity and quality were determined using Nanodrop, gel electrophoresis, and further by the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA).

4.2. Determination the Content of Aurantio-Obtusin

In order to determine the content of aurantio-obtusin, the seed, root, leaf, flower, and stem of *C. obtusifolia* were firstly extracted, respectively, by means of reflux extraction using methanol as solvent. Secondly, the extractions were centrifuged at 4 °C, 10,000 rpm, 15 min and dried by distillation. Then, the residual material was hydrolyzed by hydrochloric acid and extracted with chloroform, and dried by distillation. Moreover, the dried material was dissolved by anhydrous ethanol-ethyl acetate (2:1, volume ratio) and filtered through 0.45 μm microporous filter membrane. Finally, the content of aurantio-obtusin was determined by HPLC using Hypersil ODS2 column (4.6 mm × 200 mm, 5 μm) with flow rate of 1 mL min^{−1}, detection wavelength at 284 nm, temperature at 30 °C, acetonitrile, and 0.1% phosphoric acid in water as mobile phase.

4.3. Illumina Library Construction and Sequencing

The Poly (A) mRNA from fifteen samples of five various organs (three repetitions with each organs) was enriched from total RNA using oligo (dT) magnetic beads. Following the enrichment, the mRNA was fragmented into small pieces within fragmentation buffer. Using these short fragments as templates, the first-stand cDNA was synthesized using SuperscriptTM III reverse transcriptase and random hexamer (N6) primers. Subsequently, the RNA templates were removed and the second-strand cDNA was synthesized using dNTPs, DNA polymerase I, and RNase H. These short double cDNA fragments were purified with AMPure XP beads. After end reparation and A-tailing, the short cDNA fragments were ligated with the Illumina paired-end adaptors and purified with AMPure XP beads.

Then, PCR was used to selectively enrich DNA fragments with adapter molecules on both ends and to create the final cDNA library. The concentration of the cDNA library was assessed using Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA) and the quality of the cDNA library was measured using the Agilent 2100 Bioanalyzer. Finally, the 15 libraries were sequenced from both the 5' and 3' ends using Illumina HiSeq™ 4000 system (Illumina, San Diego, CA, USA), respectively.

4.4. Single-Molecule Real-Time Library Construction and Sequencing

Total RNA of five individual samples (seed, root, stem, leaf, and flower) was pooled to provide 42 µg of total *C. obtusifolia* RNA. Poly (A) RNA was isolated from the total RNA using the oligo (dT) magnetic bead binding method and the Poly (A) Purist™ Kit. Isolated poly (A) RNA was eluted with 20 µL of RNase-free water. One microgram of RNA was reversely transcribed using Clontech SMARTer cDNA synthesis kit. After PCR amplification, quality control, and purification, we performed size selection using BluePippin Size Selection System protocol and herein produced three fractions containing fragments of 1–2, 2–3, and 3–6 kb in length, respectively. The cDNA products were then subjected to construction of SMRTbell Template libraries using SMRTbell Template Prep Kit. The concentration and quality of the cDNA library were measured using Qubit 2.0 fluorometer and Agilent 2100 Bioanalyzer, respectively. Finally, a total of four SMRT cells were sequenced on PacBio RS II platform (Pacific Biosciences, Menlo Park, CA, USA).

4.5. Error Correction of PacBio Reads

According to the official protocol, raw polymerase reads that had full passes ≥ 0 and the predicted consensus accuracy >0.75 were selected for producing ROIs. After the ROIs shorter than 50 bp in length were discarded, they were classified into full-length (FL) and non-full-length (nFL) reads according to whether 5'/3' cDNA primers and poly (A) tail were simultaneously observed or not. We subsequently employed three strategies for improving accuracy of PacBio reads. First, all sequences were subjected to isoform-level clustering by Iterative isoform-clustering (IEC) algorithm and herein produced consensus sequences of isoform. Second, the isoform sequences were corrected to obtain high-quality isoforms and low-quality isoforms using Quiver software. Third, the low-quality isoforms were corrected in aid of Illumina short reads using Proovread tool with default setting. Finally, the longest isoform of each cluster was regarded as the transcript which was followed with correction by CD-HIT software to remove redundant information for further functional annotation.

4.6. Functional Annotation and Classification

In order to predict putative gene function, transcripts were aligned with the NR (available online: <http://www.ncbi.nlm.nih.gov/>), SwissProt (available online: <http://www.uniprot.org/>), GO (available online: <http://www.geneontology.org/>), COG (available online: <http://www.ncbi.nlm.nih.gov/COG/>), and KEGG (available online: <http://www.genome.jp/kegg/>) databases using BLASTX with E-value of 1×10^{-5} . When the results from different databases conflicted, a priority order of NR, Swiss-Prot, KEGG, and COG was obeyed. Then, based on the best BLAST hits from the KEGG database, KEGG orthology was performed using KOBAS software 2.0. At last, the amino acid sequences of all the transcripts were annotated using HMMER software (E-value $< 1 \times 10^{-10}$) with the PFAM (available online: <http://pfam.xfam.org/>) to produce the final annotation information of transcripts.

4.7. Determination of Relative Gene Expression Level

Relative gene expression levels of each transcript was determined by calculating the sum of the fragments mapping to each transcript, and then normalized by converting the fragments counts to (fragments per kilobase of transcript per million mapped reads, or FPKM) across the transcriptome.

Differential expression analysis was performed using DESeq [101]. Those genes having a fold change ratio ≥ 2 and 5, with corrected *p* value < 0.01 were considered as differentially expressed genes

and specially expressed genes, respectively [9,102]. The transcriptome analysis pipeline and strategy involved in the present study was outlined in Figure S1.

4.8. De Novo Detection and Validation of Alternative Splicing Events

The alternative splicing (AS) events were obtained according to the method of Liu et al. [103]. The prediction principle is based on pairwise alignment using BLAST. If the alignment result of two isoforms satisfy the following criterions, both are considered to be from one AS event. First, both isoforms are longer than 1000 bp and have two HSPs (High-scoring Segment pairs). Second, the AS GAP is no less than 100 bp and is at least 100 bp away from the 3' and 5' end. Third, the overlap between variable transcripts is no more than 5 bp.

4.9. Long Noncoding RNA Identification

To define a set of putative lncRNAs in the transcriptomes of *C. obtusifolia*, we applied several filtering criteria. All of the transcripts not overlapping with *C. obtusifolia* protein-coding genes and being located at least 1 kb away from the closest annotated protein were considered for our analysis. A series of filtering steps were then implemented. The first one consisted of selecting the transcripts for which Coding Potential Calculator (CPC) [104] returned no (or just partial) coding potential. Second, the Coding–Noncoding Index (CNCI) [105] was then used to search all possible protein noncoding sequences. Similarly, the Coding Potential Assessment Tool (CPAT) was used to search the possible translational products based on the length and coverage of ORF [106]. Finally, the Hmmscan tool was used to search the possible translational products of each transcript against a database of Pfam profiles and the transcripts returning an expectation value lower than 1×10^{-3} were removed.

4.10. Phylogenetic Analysis

In order to identify candidate CYPs involved in the auranlio-obtusin biosynthesis, the seed-specific CYPs with sequence length over 200 amino acids were selected and then aligned with 61 CYPs from various plants using MEGA 6.0 (Arizona State University, Tempe, AZ, USA). The unrooted phylogenetic tree was constructed using the neighbor-joining clustering method with bootstrap values obtained after 1000 replications.

4.11. Functional Analysis of CoDXS and CoDXR Genes

In order to validate the accuracy of the sequences obtained from the hybrid platform, two full-length transcripts (F01.PB17587 and F01.PB29801) encoding *C. obtusifolia* 1-deoxy-D-xylulose 5-phosphate synthase (CoDXS) and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (CoDXR) were directly chosen for functional analysis, respectively, using bacterial complementation assay [107]. In higher plant, DXS and DXR played an important role in the MEP pathway and were involved in the biosynthesis of anthraquinone [108]. The plasmids pAC-BETA and pTrc-AtIPI were purchased from ADDGENE Corporation (USA). The pAC-BETA contained all functional genes of the β -carotene synthesis, and also retained a chloramphenicol (Chl) resistance gene. Cells of *E. coli* containing pAC-BETA produced and accumulated β -carotene, resulting in yellow colonies. The pTrc-AtIPI contained a Trc promoter, ampicillin resistance gene and AtIPI gene.

The ORF of CoDXS was amplified by PCR with forward primer (5'-A TAA GAA TGC GGC CGC ATG GCT CTT TGC ACA TTC TC-3', the *Not*I site is underlined and the initiation codon is shown in italics) and reverse primer (5'-CGC GGA TCC TTA TGA CAA AAC CTC TAA TGC CTC-3', the *Bam*H1 site is underlined and the stop codon is shown in italics). The coding region of CoDXR gene was amplified by PCR using the forward primer (5'-CG GAA TTC ATG GCT CTG AAT TTG 3', the *Eco*R1 site is underlined and the initiation codon is shown in italics) and the reverse primer (5'-GA AGA TCT TCA TGC AGG AAT AGG A-3, the *Bgl*II site is underlined and the stop codon is shown in italics). The fragments were cloned into a pMD19-T vector for sequencing to confirm that no substitutions or deletions occurred. After digestion of PCR fragments and plasmid pTrc-AtIPI with

corresponding endonucleases, the recombinant plasmid pTrc-CoDXS and pTrc-CoDXR were generated and subsequently verified by sequencing, respectively.

The pTrc-AtIPI digested by PstI to remove *AtIPI* and then ligated by T4 DNA ligase was used as control. The plasmids pTrc-CoDXS and pAC-BETA, as well as the plasmids pTrc-CoDXR and pAC-BETA, were cotransformed into *E. coli* Top10, respectively. The plasmids pTrc and pAC-BETA were also cotransformed and the single plasmid pAC-BETA, which was transformed into the *E. coli* Top10 to be used as the negative controls. Transformants were cultured on solid LB medium containing ampicillin (150 mg/L) and Chl (50 mg/L) at 37 °C in dark for 2 days.

4.12. Quantitative Real-Time PCR (qRT-PCR) Assay

To analyze the expression levels of mRNA of potential genes and TFs related to anthraquinone biosynthesis, seed formation and development, and stress response in various organs of *C. obtusifolia*, the total RNA of various organs were extracted and subjected to cDNA synthesis and further qRT-PCR, respectively. Moreover, the expression pattern of 8 Hsp20 genes under heat, chilling, drought, and salt stresses was detected. For chilling and heat stress, *C. obtusifolia* were cultivated at 4 and 37 °C for 0, 1, 6, 12, 24, and 48 h, respectively. For drought and salt stresses, *C. obtusifolia* were cultivated supplemented with PEG 6000 (100 mg mL⁻¹), or NaCl (200 mM) for 0, 1, 6, 12, 24, and 48 h, respectively. The qRT-PCR was performed using SYBR Premix Ex Taq™ II (Takara, Japan) with a LightCycler 96 system (Roche Diagnostics, Mannheim, Germany). The total RNA of *C. obtusifolia* leaf under different stresses were extracted and subjected to cDNA synthesis, respectively.

In each run, 1 µL cDNA template was added to 10 µL reaction buffer using gene-specific primer pairs under the following conditions: initial denaturation at 95 °C for 30 s, 40 cycles of 5 s at 95 °C, 10 s at 60 °C, and 25 s at 72 °C. All samples were performed in triplicate. Then, based on the transcriptome data, the seed-specific CYPs, SAM-methyltransferases, UDPGs, TFs, and 8 Hsp20s were selected for verification by qRT-PCR using *EF1α* as a reference gene [109]. The relative expression levels of the selected genes were normalized to the selected reference gene and determined by the $\Delta\Delta C_t$ -method [18]. All the Primer sequences used in this study were listed in Table S1.

5. Conclusions

Compared with previous traditional cDNA cloning or NGS studies in *C. obtusifolia*, we produced a more comprehensive transcriptome data set with the following features. First, many longer (81.73%, >1000 bp) sequences with high quality and more annotation ratio (98.25%) were generated in our hybrid sequencing platform, providing a direct use for further genetic functional studies without additional RACE to get a complete ORF. Second, more sequences (1214) were involved in secondary metabolism and more sequences (607) involved in the anthraquinone backbone biosynthesis were found in the hybrid platform, whereas only 455 and 185 sequences were found in previous NGS data, respectively. Moreover, the hybrid platform provided an opportunity to detect alternative splicing and the 2712 candidate AS events suggesting a high degree of transcriptome complexity in *C. obtusifolia*. Although auranthio-obtusin biosynthesis-related genes (all with FPKM > 10 in seed) and Hsp20 genes do not seem to undergo any significant degree of alternative splicing, several potential TFs involved in the seed formation and development were identified as undergoing alternative splicing in *C. obtusifolia* (Figure S3). It should be noted that only one isoform was predominant (with FPKM > 10 in seed), suggesting that another alternative isoform may not be significant. Nevertheless, alternative splicing is clearly observed among the TFs involved in seed formation and development (Table 3), which may serve as a regulatory mechanism in controlling seed formation and development. Further studies are required to investigate the biological functions of these transcript isoforms. Although only a minority of lncRNAs could be annotated with known RNA motifs, these still provide a useful resource for understanding the potential functions of lncRNAs in *C. obtusifolia*. Finally, we characterized the strong candidate genes in the biosynthesis of auranthio-obtusin from *C. obtusifolia*, TFs involved in the seed formation and development, and stress-responsive CoHsp20 genes as well.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1422-0067/19/9/2476/s1>.

Author Contributions: Conceptualization, J.Z., H.L.; Data curation, Y.D., Z.Y., D.L., and C.L.; Funding acquisition, H.L.; Methodology, H.Z.; Project administration, H.L.; Writing—original draft, Y.D.; Writing—review & editing, J.Z.

Acknowledgments: This work was partially funded by National Natural Science Foundation of China (No. 31500276), Sichuan Science and Technology Program (No. 2017JY0222 and 2018SZ0061), Science and Technology Program for Public Wellbeing from Science and Technology Bureau of Chengdu City (No. 2015-HM01-00051-SF and No. 2016-HM01-00260-SF).

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

Abbreviations

<i>C. obtusifolia</i>	<i>Cassia obtusifolia</i>
Co	<i>Cassia obtusifolia</i>
TFs	Transcription Factor
CYPs	Cytochrome p450/Cytochrome p450 Monooxygenase/NADPH-cytochrome p450 reductase
CoCYP450	<i>Cassia obtusifolia</i> Cytochrome p450
CoHsp20	<i>Cassia obtusifolia</i> heat shock protein
CoReductase	<i>Cassia obtusifolia</i> Cytochrome p450 reductase
CoARF	<i>Cassia obtusifolia</i> Auxin-Responsive Transcription Factor
CoLEA	<i>Cassia obtusifolia</i> Late Embryogenesis Abundant Transcription Factor
CoNF	<i>Cassia obtusifolia</i> Nuclear Transcription Factor
CoHBP	<i>Cassia obtusifolia</i> HBP Transcription Factor
CoMADS	<i>Cassia obtusifolia</i> MADS-box Transcription Factor
CoMYB	<i>Cassia obtusifolia</i> MYB Transcription Factor
CoNAC	<i>Cassia obtusifolia</i> Nac Transcription Factor
CobHLH	<i>Cassia obtusifolia</i> bHLH Transcription Factor
CoWRKY	<i>Cassia obtusifolia</i> WRKY Transcription Factor
CoYABBY	<i>Cassia obtusifolia</i> YABBY Transcription Factor
CoZF	<i>Cassia obtusifolia</i> Zinc Finger Transcription Factor
CoSAMT	<i>Cassia obtusifolia</i> di-O-methyltransferase
GATA	GATA transcription factor
Trihelix TFs	Trihelix Transcription Factor
ACAT	Acetyl-CoA C-acetyltransferase
HMGS	Hydroxymethylglutaryl-CoA synthase
HMGR	Hydroxymethylglutaryl-CoA reductase
MK	Mevalonate kinase
PMK	Phosphomevalonate kinase
MPD	Diphosphomevalonate decarboxylase
DXS	1-deoxy-D-xylulose-5-phosphate synthase
DXR	1-deoxy-D-xylulose-5-phosphate reductoisomerase
ISPD	1-C-methyl-D-erythritol 4-phosphate cytidyltransferase
CDPMEK	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
ISPF	1-C-methyl-D-erythritol 2,4-cyclodiphosphate Synthase
HDS	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase
HDR	4-hydroxy-3-methylbut-2-enyl-diphosphate reductase
IPPS	Isopentenyl-diphosphate delta-isomerase
DAHPS	3-deoxy-7-phosphoheptulonate synthase
DHQS	3-dehydroquinate synthase
SDH	3-dehydroquinate dehydratase
DHQD	Shikimate dehydrogenase
SMK	Shikimate kinase
EPSP synthase	3-phosphoshikimate 1-carboxyvinyltransferase
CS	Chorismate synthase
ICS	Menaquinone-specific isochorismate synthase
menC	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase
PHYLLO	O-succinylbenzoate synthase
MenH	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase
menD	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase
menE	Acyl-activating enzyme
menB	Naphthoate synthase

PSKIII	Polyketide synthaseIII
PKC	Polyketide cyclase/dehydratase
UDPG	UDP-Glucosyl transferase

References

- Lee, H.H.; Ahn, J.H.; Kwon, A.R.; Lee, E.S.; Kwak, J.H.; Min, Y.H. Chemical composition and antimicrobial activity of the essential oil of apricot seed. *Phytother. Res.* **2014**, *28*, 1867–1872. [[CrossRef](#)] [[PubMed](#)]
- Cao, J.X.; Zhang, Q.Y.; Cui, S.Y.; Cui, X.Y.; Zhang, J.; Zhang, Y.H.; Bai, Y.J.; Zhao, Y.Y. Hypnotic effect of jujubosides from Semen *Ziziphi Spinosae*. *J. Ethnopharmacol.* **2010**, *130*, 163–166. [[CrossRef](#)] [[PubMed](#)]
- Chen, L.; Li, W.; Qi, D.; Wang, D. *Lycium barbarum* polysaccharide protects against LPS-induced ARDS by inhibiting apoptosis, oxidative stress and inflammation in pulmonary endothelial cells. *Free Radic. Res.* **2018**, *4*, 1–171. [[CrossRef](#)] [[PubMed](#)]
- Wang, S.; Shi, P.; Qu, L.; Ruan, J.; Yang, S.; Yu, H.; Zhang, Y.; Wang, T. Bioactive constituents obtained from the seeds of *Lepidium apetalum* Willd. *Molecules* **2017**, *22*, 540. [[CrossRef](#)] [[PubMed](#)]
- Luan, G.; Wang, Y.; Wang, Z.; Zhou, W.; Hu, N.; Li, G.; Wang, H. Flavonoid glycosides from Fenugreek seeds regulate glycolipid metabolism by improving mitochondrial function in 3T3-L1 adipocytes in vitro. *J. Agric. Food Chem.* **2018**, *66*, 3169–3178. [[CrossRef](#)] [[PubMed](#)]
- Maehara, S.; Agusta, A.; Kitamura, C.; Ohashi, K.; Shibuya, H. Composition of the endophytic filamentous fungi associated with *Cinchonaled geriana* seeds and production of Cinchona alkaloids. *J. Nat. Med.* **2016**, *70*, 271–275. [[CrossRef](#)] [[PubMed](#)]
- Kim, M.J.; Shim, C.K.; Kim, Y.K.; Hong, S.J.; Park, J.H.; Han, E.J.; Kim, S.C. Enhancement of seed dehiscence by seed treatment with *Talaromyces flavus* GG01 and GG04 in Ginseng (*Panax ginseng*). *Plant Pathol. J.* **2017**, *33*, 1–8. [[CrossRef](#)] [[PubMed](#)]
- Leisner, C.P.; Yendrek, C.R.; Ainsworth, E.A. Physiological and transcriptomic responses in the seed coat of field-grown soybean (*Glycine max* L. Merr.) to abiotic stress. *BMC Plant Biol.* **2017**, *17*, 242. [[CrossRef](#)] [[PubMed](#)]
- Belmonte, M.F.; Kirkbride, R.C.; Stone, S.L.; Pelletier, J.M.; Bui, A.Q.; Yeung, E.C.; Hashimoto, M.; Fei, J.; Harada, C.M.; Munoz, M.D.; et al. Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E435–E444. [[CrossRef](#)] [[PubMed](#)]
- Kim, D.H.; Kim, S.; Jung, W.Y.; Park, S.J.; Park, D.H.; Kim, J.M.; Cheong, J.H.; Ryu, J.H. The neuroprotective effects of the seeds of *Cassia obtusifolia* on transient cerebral global ischemia in mice. *Food Chem. Toxicol.* **2009**, *47*, 1473–1479. [[CrossRef](#)] [[PubMed](#)]
- Zhuang, S.Y.; Wu, M.L.; Wei, P.J.; Cao, Z.P.; Xiao, P.; Li, C.H. Changes in plasma lipid levels and antioxidant activities in rats after supplementation of obtusifolin. *Planta Med.* **2016**, *82*, 539–543. [[CrossRef](#)] [[PubMed](#)]
- Jung, H.A.; Ali, M.Y.; Choi, J.S. Promising inhibitory effects of anthraquinones, naphthopyrone, and naphthalene glycosides, from *Cassia obtusifolia* on α -glucosidase and human protein tyrosine phosphatases 1B. *Molecules* **2016**, *22*, 28. [[CrossRef](#)] [[PubMed](#)]
- Chai, M.; Zhou, C.; Molina, I.; Fu, C.; Nakashima, J.; Li, G.; Zhang, W.; Park, J.; Tang, Y.; Jiang, Q.; et al. A class II KNOX gene, KNOX4, controls seed physical dormancy. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6997–7002. [[CrossRef](#)] [[PubMed](#)]
- Griffiths, S.; Mesarich, C.H.; Saccomanno, B.; Vaisberg, A.; Wit, P.J.G.M.; Cox, R.; Collemare, J. Elucidation of cladofulvin biosynthesis reveals a cytochrome P450 monooxygenase required for anthraquinone dimerization. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6851–6856. [[CrossRef](#)] [[PubMed](#)]
- Han, Y.S.; Roytrakula, S.; Verbernea, M.C.; Heijdenb, R.V.D.; Linthorst, H.J.M.; Verpoorte, R. Cloning of a cDNA encoding 1-deoxy-D-xylulose 5-phosphate synthase from *Morinda citrifolia* and analysis of its expression in relation to anthraquinone accumulation. *Plant Sci.* **2003**, *164*, 911–917. [[CrossRef](#)]
- Li, M.Y.; Tan, H.W.; Wang, F.; Jiang, Q.; Xu, Z.S.; Tian, C.; Xiong, A.S. De novo transcriptome sequence assembly and identification of AP2/ERF transcription factor related to abiotic stress in parsley (*Petroselinum crispum*). *PLoS ONE* **2014**, *9*, e108977. [[CrossRef](#)] [[PubMed](#)]
- Van Ooijen, G.; Lukasik, E.; Van Den Burg, H.A.; Vossen, J.H.; Cornelissen, B.J.; Takken, F.L. The small heat shock protein 20 RSI2 interacts with and is required for stability and function of tomato resistance protein I-2. *Plant J.* **2010**, *63*, 563–572. [[CrossRef](#)] [[PubMed](#)]

18. Liu, Z.; Song, T.; Zhu, Q.; Wang, W.; Zhou, J.; Liao, H. De novo assembly and analysis of *Cassia obtusifolia* seed transcriptome to identify genes involved in the biosynthesis of active metabolites. *Biosci. Biotechnol. Biochem.* **2014**, *78*, 791–799. [[CrossRef](#)] [[PubMed](#)]
19. Misawa, N.; Satomi, Y.; Kondo, K.; Yokoyama, A.; Kajiwara, S.; Saito, T.; Ohtani, T.; Miki, W. Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J. Bacteriol.* **1995**, *177*, 6575–6584. [[CrossRef](#)] [[PubMed](#)]
20. Young, P.R.; Lashbrooke, J.G.; Alexandersson, E.; Jacobson, D.; Moser, C.; Velasco, R.; Melané, A. The genes and enzymes of the carotenoid metabolic pathway in *Vitis vinifera* L. *BMC Genom.* **2012**, *13*, 243. [[CrossRef](#)] [[PubMed](#)]
21. Tong, Y.R.; Su, P.; Zhao, Y.J.; Zhang, M.; Wang, X.J.; Liu, Y.J.; Zhang, X.N.; Gao, W.; Huang, L.Q. Molecular cloning and characterization of DXS and DXR genes in the terpenoid biosynthetic pathway of *Tripterygium wilfordii*. *Int. J. Mol. Sci.* **2015**, *16*, 25516–25535. [[CrossRef](#)] [[PubMed](#)]
22. Baque, M.A.; Elgirban, A.; Lee, E.J.; Paek, K.Y. Sucrose regulated enhanced induction of anthraquinone, phenolics, flavonoids biosynthesis and activities of antioxidant enzymes in adventitious root suspension cultures of *Morinda citrifolia* (L.). *Acta Physiol. Plant.* **2012**, *34*, 405–415. [[CrossRef](#)]
23. Lange, B.M.; Severin, K.; Bechthold, A.; Heide, L. Regulatory role of microsomal 3-hydroxyl-3-methylglutaryl-coenzyme A reductase for shikonin biosynthesis in *Lithospermum erythrorhizon* cell suspension cultures. *Planta* **1998**, *204*, 234–241. [[CrossRef](#)] [[PubMed](#)]
24. Perassolo, M.; Quevedo, C.; Busto, V.; Lanone, A.; Giulietti, M. Enhance of anthraquinone production by effect of praline and aminoindan-phosphonic acid in *Rubia tinctorum* suspension cultures. *Enzym. Microb. Technol.* **2007**, *41*, 181–185. [[CrossRef](#)]
25. Han, Y.S.; Heijden, R.; Lefeber, A.W.M.; Erkelens, C.; Verpoorte, R. Biosynthesis of anthraquinones in cell cultures of *Cinchona Robusta* proceeds via the methylerythritol-phosphate pathway. *Phytochemistry* **2002**, *59*, 45–55. [[CrossRef](#)]
26. Van Tegelen Leon, J.; Bongaerts, R.J.M.; Croes, A.F.; Verpoorte, R.; Wullems, G.J. Isochorismate synthase isoforms from elicited cell cultures of *Rubia tinctorum*. *Phytochemistry* **1999**, *51*, 263–269. [[CrossRef](#)]
27. Kuzuyama, T.; Takagi, M.; Takahashi, S.; Seto, H. Cloning and characterization of 1-deoxy-D-xylulose 5-phosphate synthase from *Streptomyces* sp. Strain CL190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis. *J. Bacteriol.* **2000**, *182*, 891–897. [[CrossRef](#)] [[PubMed](#)]
28. Shkryl, Y.N.; Veremeichik, G.N.; Makhazen, D.S.; Silantieva, S.A.; Mishchenko, N.P.; Vasileva, E.A.; Fedoreyev, S.A.; Bulgakov, V.P. Increase of anthraquinone content in *Rubia cordifolia* cells transformed by native and constitutively active forms of the AtCPK1 gene. *Plant Cell Rep.* **2016**, *35*, 1907–1916. [[CrossRef](#)] [[PubMed](#)]
29. Thatcher, L.F.; Powell, J.J.; Aitken, E.A.; Kazan, K.; Manners, J.M. The lateral organ boundaries domain transcription factor LBD20 functions in *Fusarium wilt* Susceptibility and jasmonate signaling in *Arabidopsis*. *Plant Physiol.* **2012**, *160*, 407–418. [[CrossRef](#)] [[PubMed](#)]
30. Ramos-Valdivia, A.C.; van der Heijden, R.; Verpoorte, R.; Camara, B. Purification and characterization of two isoforms of isopentenyl-diphosphate isomerase from elicitor-treated *Cinchona robusta* cells. *Eur. J. Biochem.* **1997**, *249*, 161–170. [[CrossRef](#)] [[PubMed](#)]
31. Pandith, S.A.; Dhar, N.; Rana, S.; Bhat, W.W.; Kushwaha, M.; Gupta, A.P.; Shah, M.A.; Vishwakarma, R.; Lattoo, S.K. Functional promiscuity of two divergent paralogs of type III plant polyketide synthases. *Plant Physiol.* **2016**, *171*, 2599–2619. [[PubMed](#)]
32. Andersen-Ranberg, J.; Kongstad, K.T.; Nafisi, M.; Staerk, D.; Okkels, F.T.; Mortensen, U.H.; Lindberg Møller, B.; Frandsen, R.J.N.; Kannangara, R. Synthesis of C-glucosylated octaketide anthraquinones in *Nicotiana benthamiana* by using a multispecies-based biosynthetic pathway. *ChemBiochem* **2017**, *18*, 1893–1897. [[CrossRef](#)] [[PubMed](#)]
33. Abdel-Rahman, I.A.; Beuerle, T.; Ernst, L.; Abdel-Baky, A.M.; Eel-D, D.; Ahmed, A.S.; Beerhues, L. In vitro formation of the anthranoid scaffold by cell-free extracts from yeast-extract-treated *Cassia bicapsularis* cell cultures. *Phytochemistry* **2013**, *88*, 15–24. [[CrossRef](#)] [[PubMed](#)]
34. Van Tegelen, L.J.; Moreno, P.R.; Croes, A.F.; Verpoorte, R.; Wullems, G.J. Purification and cDNA cloning of isochorismate synthase from elicited cell cultures of *Catharanthus roseus*. *Plant Physiol.* **1999**, *119*, 705–712. [[CrossRef](#)] [[PubMed](#)]

35. Rai, A.; Singh, R.; Shirke, P.A.; Tripathi, R.D.; Trivedi, P.K.; Chakrabarty, D. Expression of rice CYP450-like gene (Os08g01480) in *Arabidopsis* modulates regulatory network leading to heavy metal and other abiotic stress tolerance. *PLoS ONE* **2015**, *10*, e0138574. [[CrossRef](#)] [[PubMed](#)]
36. Aharoni, A.; Giri, A.P.; Verstappen, F.W.; Berteaux, C.M.; Sevenier, R.; Sun, Z.; Jongtsma, M.A.; Schwab, W.; Bouwmeester, H.J. Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* **2004**, *16*, 3110–3131. [[CrossRef](#)] [[PubMed](#)]
37. Wu, D.; Austin, R.S.; Zhou, S.; Brown, D. The root transcriptome for North American ginseng assembled and profiled across seasonal development. *BMC Genom.* **2013**, *14*, 564. [[CrossRef](#)] [[PubMed](#)]
38. Hallahan, D.L.; Lau, S.M.; Harder, P.A.; Smiley, D.W.; Dawson, G.W.; Pickett, J.A.; Christoffersen, R.E.; O'Keefe, D.P. Cytochrome P-450-catalysed monoterpene oxidation in catmint (*Nepeta racemosa*) and avocado (*Persea americana*): Evidence for related enzymes with different activities. *Biochim. Biophys. Acta* **1994**, *1201*, 94–100. [[CrossRef](#)]
39. Bak, S.; Tax, F.E.; Feldmann, K.A.; Galbraith, D.W.; Feyereisen, R. CYP83B1, a cytochrome P450 at the metabolic branch point in auxin and indole glucosinolate biosynthesis in *Arabidopsis*. *Plant Cell* **2001**, *13*, 101–111. [[CrossRef](#)] [[PubMed](#)]
40. Karamat, F.; Olry, A.; Doerper, S.; Vialart, G.; Ullmann, P.; Werck-Reichhart, D.; Bourgaud, F.; Hehn, A. CYP98A22, a phenolic ester 3'-hydroxylase specialized in the synthesis of chlorogenic acid, as a new tool for enhancing the furanocoumarin concentration in *Ruta graveolens*. *BMC Plant Biol.* **2012**, *12*, 152. [[CrossRef](#)] [[PubMed](#)]
41. Babineau, M.; Mahmood, K.; Mathiassen, S.K.; Kudsk, P.; Kristensen, M. De novo transcriptome assembly analysis of weed *Apera spica-venti* from seven tissues and growth stages. *BMC Genom.* **2017**, *18*, 128. [[CrossRef](#)] [[PubMed](#)]
42. Kumar, A.; Yogendra, K.N.; Karre, S.; Kushalappa, A.C.; Dion, Y.; Choo, T.M. WAX INDUCER1 (HvWIN1) transcription factor regulates free fatty acid biosynthetic genes to reinforce cuticle to resist Fusarium head blight in barley spikelets. *J. Exp. Bot.* **2016**, *67*, 4127–4139. [[CrossRef](#)] [[PubMed](#)]
43. Tsuda, K.; Kurata, N.; Ohyanagi, H.; Hake, S. Genome-wide study of KNOX regulatory network reveals brassinosteroid catabolic genes important for shoot meristem function in rice. *Plant Cell* **2014**, *26*, 3488–3500. [[CrossRef](#)] [[PubMed](#)]
44. Nomura, T.; Magome, H.; Hanada, A.; Takeda-Kamiya, N.; Mander, L.N.; Kamiya, Y.; Yamaguchi, S. Functional analysis of *Arabidopsis* CYP714A1 and CYP714A2 reveals that they are distinct gibberellin modification enzymes. *Plant Cell Physiol.* **2013**, *54*, 1837–1851. [[CrossRef](#)] [[PubMed](#)]
45. Pérez-España, V.H.; Sánchez-León, N.; Vielle-Calzada, J. CYP85A1 is required for the initiation of female gametogenesis in *Arabidopsis thaliana*. *Plant Signal. Behav.* **2011**, *6*, 321–326. [[CrossRef](#)] [[PubMed](#)]
46. Zhang, N.; Han, Z.; Sun, G.; Hoffman, A.; Wilson, I.W.; Yang, Y.; Gao, Q.; Wu, J.; Xie, D.; Dai, J.; et al. Molecular cloning and characterization of a cytochrome P450 taxoid 9 α -hydroxylase in *Ginkgo biloba* cells. *Biochem. Biophys. Res. Commun.* **2014**, *443*, 938–943. [[CrossRef](#)] [[PubMed](#)]
47. O'Brien, M.; Chantha, S.C.; Rahier, A.; Matton, D.P. Lipid signaling in plants. Cloning and expression analysis of the obtusifoliol 14 α -demethylase from *Solanum chacoense* Bitt., a pollination- and fertilization-induced gene with both obtusifoliol and lanosterol demethylase activity. *Plant Physiol.* **2005**, *139*, 734–749. [[CrossRef](#)] [[PubMed](#)]
48. Terracciano, I.; Cantarella, C.; Fasano, C.; Cardi, T.; Mennella, G.; D'Agostino, N. Liquid-phase sequence capture and targeted re-sequencing revealed novel polymorphisms in tomato genes belonging to the MEP carotenoid pathway. *Sci. Rep.* **2017**, *7*, 5616. [[CrossRef](#)] [[PubMed](#)]
49. Huang, F.C.; Sung, P.H.; Do, Y.Y.; Huang, P.L. Differential expression and functional characterization of the NADPH cytochrome P450 reductase genes from *Nothapodytes foetida*. *Plant Sci.* **2012**, *190*, 16–23. [[CrossRef](#)] [[PubMed](#)]
50. Lepesheva, G.I.; Nes, W.D.; Zhou, W.; Hill, G.C.; Waterman, M.R. CYP51 from *Trypanosoma brucei* is obtusifoliol-specific. *Biochemistry* **2004**, *43*, 10789–10799. [[CrossRef](#)] [[PubMed](#)]
51. Agarwal, P.; Khurana, P. Characterization of a novel zinc finger transcription factor (TaZnF) from wheat conferring heat stress tolerance in *Arabidopsis*. *Cell Stress Chaperones* **2018**, *23*, 253–267. [[CrossRef](#)] [[PubMed](#)]
52. Bogamuwa, S.; Jang, J.C. The *Arabidopsis* tandem CCCH zinc finger proteins AtTZF4, 5 and 6 are involved in light-, abscisic acid- and gibberellic acid-mediated regulation of seed germination. *Plant Cell Environ.* **2013**, *36*, 1507–1519. [[CrossRef](#)] [[PubMed](#)]

53. Huang, P.; Yoshida, H.; Yano, K.; Kinoshita, S.; Kawai, K.; Koketsu, E.; Hattori, M.; Takehara, S.; Huang, J.; Hirano, K.; et al. OsIDD2, a zinc finger and INDETERMINATE DOMAIN protein, regulates secondary cell wall formation. *J. Integr. Plant Biol.* **2018**, *60*, 130–143. [[CrossRef](#)] [[PubMed](#)]
54. Ambawat, S.; Sharma, P.; Yadav, N.R.; Yadav, R.C. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* **2013**, *19*, 307–321. [[CrossRef](#)] [[PubMed](#)]
55. Han, X.; Yin, Q.; Liu, J.; Jiang, W.; Di, S.; Pang, Y. GmMYB58 and GmMYB205 are seed-specific activators for isoflavonoid biosynthesis in *Glycine max*. *Plant Cell Rep.* **2017**, *36*, 1889–1902. [[CrossRef](#)] [[PubMed](#)]
56. Liu, Z.; Miao, L.; Huo, R.; Song, X.; Johnson, C.; Kong, L.; Sundaresan, V.; Yu, X. ARF2-ARF4 and ARF5 are Essential for Female and Male Gametophyte Development in *Arabidopsis*. *Plant Cell Physiol.* **2018**, *59*, 179–189. [[CrossRef](#)] [[PubMed](#)]
57. Orman-Ligeza, B.; Parizot, B.; Gantet, P.P.; Beeckman, T.; Bennett, M.J.; Draye, X. Post-embryonic root organogenesis in cereals: Branching out from model plants. *Trends Plant Sci.* **2013**, *18*, 459–467. [[CrossRef](#)] [[PubMed](#)]
58. Mathew, I.E.; Das, S.; Mahto, A.; Agarwal, P. Three rice NAC transcription factors heteromerize and are associated with seed size. *Front. Plant Sci.* **2016**, *7*, 1638. [[CrossRef](#)] [[PubMed](#)]
59. Zhang, T.; Song, C.; Song, L.; Shang, Z.; Yang, S.; Zhang, D.; Sun, W.; Shen, Q.; Zhao, D. RNA sequencing and coexpression analysis reveal key genes involved in α -linolenic acid biosynthesis in *Perilla frutescens* seed. *Int. J. Mol. Sci.* **2017**, *18*, 2433. [[CrossRef](#)] [[PubMed](#)]
60. Zhang, J.; Huang, G.Q.; Zou, D.; Yan, J.Q.; Li, Y.; Hu, S.; Li, X.B. The cotton (*Gossypium hirsutum*) NAC transcription factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. *New Phytol.* **2018**, *217*, 625–640. [[CrossRef](#)] [[PubMed](#)]
61. Kuijt, S.J.; Greco, R.; Agalou, A.; Shao, J.; 't Hoen, C.C.; Overnäs, E.; Osnato, M.; Curiale, S.; Meynard, D.; van Gulik, R.; et al. Interaction between the growth-regulating factor and notted1-like homeobox families of transcription factors. *Plant Physiol.* **2014**, *164*, 1952–1966. [[CrossRef](#)] [[PubMed](#)]
62. Bakshi, M.; Oelmüller, R. WRKY transcription factors: Jack of many trades in plants. *Plant Signal Behav.* **2014**, *9*, e27700. [[CrossRef](#)] [[PubMed](#)]
63. Reyero-Saavedra, M.D.R.; Qiao, Z.; Sánchez-Correa, M.D.S.; Díaz-Pineda, M.E.; Reyes, J.L.; Covarrubias, A.A.; Libault, M.; Valdés-López, O. Gene silencing of argonaute5 negatively affects the establishment of the legume-rhizobia symbiosis. *Genes* **2017**, *8*, 352. [[CrossRef](#)] [[PubMed](#)]
64. Wang, Y.; Xu, W.; Chen, Z.; Han, B.; Haque, M.E.; Liu, A. Gene structure, expression pattern and interaction of Nuclear Factor-Y family in castor bean (*Ricinus communis*). *Planta* **2018**, *247*, 559–572. [[CrossRef](#)] [[PubMed](#)]
65. Zhiguo, E.; Li, T.; Zhang, H.; Liu, Z.; Deng, H.; Sharma, S.; Wei, X.; Wang, L.; Niu, B.; Chen, C. A group of nuclear factor Y transcription factors are sub-functionalized during endosperm development in monocots. *J. Exp. Bot.* **2018**. [[CrossRef](#)]
66. Gramzow, L.; Weilandt, L.; Theißen, G. MADS goes genomic in conifers: Towards determining the ancestral set of MADS-box genes in seed plants. *Ann. Bot.* **2014**, *114*, 1407–1429. [[CrossRef](#)] [[PubMed](#)]
67. Huang, L.; Jia, J.; Zhao, X.; Zhang, M.; Huang, X.; E, J.; Ni, L.; Jiang, M. The ascorbate peroxidase APX1 is a direct target of a zinc finger transcription factor ZFP36 and a late embryogenesis abundant protein OsLEA5 interacts with ZFP36 to co-regulate OsAPX1 in seed germination in rice. *Biochem. Biophys. Res. Commun.* **2018**, *495*, 339–345. [[CrossRef](#)] [[PubMed](#)]
68. Cong, B.; Barrero, L.S.; Tanksley, S.D. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat. Genet.* **2008**, *40*, 800–804. [[CrossRef](#)] [[PubMed](#)]
69. Wang, H.; Li, W.; Qin, Y.; Pan, Y.; Wang, X.; Weng, Y.; Chen, P.; Li, Y. The Cytochrome P450 gene CsCYP85A1 is a putative candidate for super compact-1 (scp-1) plant architecture mutation in cucumber (*Cucumis sativus* L.). *Front. Plant Sci.* **2017**, *8*, 266. [[CrossRef](#)] [[PubMed](#)]
70. Hudson, D.; Guevara, D.R.; Hand, A.J.; Xu, Z.; Hao, L.; Chen, X.; Zhu, T.; Bi, Y.M.; Rothstein, S.J. Rice cytokinin GATA transcription factor1 regulates chloroplast development and plant architecture. *Plant Physiol.* **2013**, *162*, 132–144. [[CrossRef](#)] [[PubMed](#)]
71. Vadivel, V.; Janardhanan, K. Nutritional and antinutritional characteristics of seven south Indian wild legumes. *Plant Food Hum. Nutr.* **2002**, *57*, 151–164. [[CrossRef](#)]
72. Kaplan-Levy, R.N.; Brewer, P.B.; Quon, T.; Smyth, D.R. The trihelix family of transcription factors—light, stress and development. *Trends Plant Sci.* **2012**, *17*, 163–171. [[CrossRef](#)] [[PubMed](#)]

73. Zhou, Y.; Lu, D.; Li, C.; Luo, J.; Zhu, B.F.; Zhu, J.; Shanguan, Y.; Wang, Z.; Sang, T.; Zhou, B.; et al. Genetic control of seed shattering in rice by the APETALA2 transcription factor shattering abortion1. *Plant Cell* **2012**, *24*, 1034–1048. [[CrossRef](#)] [[PubMed](#)]
74. Lu, J.; Li, C.; Zhang, F.S.; Wu, W.; Zhang, Q.F.; Zhang, L.; Wang, T.; Wang, Q.; Qiu, P.H.; Liang, M.F.; et al. Expression of structural and non-structural proteins of severe fever with thrombocytopenia syndrome bunyavirus. *Chin. J. Virol.* **2011**, *27*, 515–520.
75. Steinbrecher, T.; Leubner-Metzger, G. Tissue and cellular mechanics of seeds. *Curr. Opin. Genet. Dev.* **2018**, *51*, 1–10. [[CrossRef](#)] [[PubMed](#)]
76. Yan, H.; Zhang, A.; Chen, J.; He, X.; Xu, B.; Xie, G.; Miao, Z.; Zhang, X.; Huang, L. Genome-wide analysis of the PvHsp20 family in Switchgrass: Motif, genomic organization, and identification of stress or developmental-related Hsp20s. *Front. Plant Sci.* **2017**, *8*, 1024. [[CrossRef](#)] [[PubMed](#)]
77. Zhao, P.; Wang, D.; Wang, R.; Kong, N.; Zhang, C.; Yang, C.; Wu, W.; Ma, H.; Chen, Q. Genome-wide analysis of the potato Hsp20 gene family: Identification, genomic organization and expression profiles in response to heat stress. *BMC Genom.* **2018**, *19*, 61. [[CrossRef](#)] [[PubMed](#)]
78. Van Montfort, R.; Slingsby, C.; Vierling, E. Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Adv. Protein Chem.* **2001**, *59*, 105–156. [[PubMed](#)]
79. Yu, J.; Cheng, Y.; Feng, K.; Ruan, M.; Ye, Q.; Wang, Q.; Wang, R.; Li, Z.; Zhou, G.; Yao, Z.; et al. Genome-wide identification and expression profiling of tomato Hsp20 gene family in response to biotic and abiotic stresses. *Front. Plant Sci.* **2016**, *7*, 1215. [[CrossRef](#)] [[PubMed](#)]
80. Yoon, J.; Cho, L.H.; Kim, S.L.; Choi, H.; Koh, H.J.; An, G. The BEL1-type homeobox gene SH5 induces seed shattering by enhancing abscission-zone development and inhibiting lignin biosynthesis. *Plant J.* **2014**, *79*, 717–728. [[CrossRef](#)] [[PubMed](#)]
81. He, L.; Fu, S.; Xu, Z.; Yan, J.; Zhou, H.; Chen, X.; Li, Y.; Au, K.; Yao, H. Hybrid sequencing of full-length cDNA transcripts of stems and leaves in *Dendrobium officinale*. *Genes* **2017**, *8*, 257. [[CrossRef](#)] [[PubMed](#)]
82. Li, J.; Harata-Lee, Y.; Denton, M.D.; Feng, Q.; Rathjen, J.R.; Qu, Z.; Adelson, D.L. Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis. *Cell Discov.* **2017**, *3*, 17031. [[CrossRef](#)] [[PubMed](#)]
83. Schimpl, F.C.; Kiyota, E.; Mayer, J.L.; Gonçalves, J.F.; da Silva, J.F.; Mazzafera, P. Molecular and biochemical characterization of caffeine synthase and purine alkaloid concentration in guarana fruit. *Phytochemistry* **2014**, *105*, 25–36. [[CrossRef](#)] [[PubMed](#)]
84. Mbatchou, V.C.; Tchouassi, D.P.; Dickson, R.A.; Annan, K.; Mensah, A.Y.; Amponsah, I.K.; Jacob, J.W.; Cheseto, X.; Habtemariam, S.; Torto, B. Mosquito larvicidal activity of *Cassia tora* seed extract and its key anthraquinones aurantio-obtusin and obtusin. *Parasit. Vectors* **2017**, *10*, 562. [[CrossRef](#)] [[PubMed](#)]
85. Mehta, R.H.; Ponnuchamy, M.; Kumar, J.; Reddy, N.R. Exploring drought stress-regulated genes in senna (*Cassia angustifolia* Vahl.): A transcriptomic approach. *Funct. Integr. Genom.* **2017**, *17*, 1–25. [[CrossRef](#)] [[PubMed](#)]
86. Rama Reddy, N.R.; Mehta, R.H.; Soni, P.H.; Makasana, J.; Gajbhiye, N.A.; Ponnuchamy, M.; Kumar, J. Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides from Senna (*Cassia angustifolia* Vahl.), a non-model plant with potent laxative properties. *PLoS ONE* **2015**, *10*, e0129422. [[CrossRef](#)] [[PubMed](#)]
87. Luo, Z.; Hu, J.; Zhao, Z.; Zhang, D. Transcriptomic analysis of heteromorphic stamens in *Cassia biscapsularis* L. *Sci. Rep.* **2016**, *6*, 31600. [[CrossRef](#)] [[PubMed](#)]
88. Han, Y.S.; Van der Heijden, R.; Verpoorte, R. Biosynthesis of anthraquinones in cell cultures of the Rubiaceae. *Plant Cell Tissue Organ Cult.* **2001**, *67*, 201–220. [[CrossRef](#)]
89. Rodriguez, T.J.; Verberne, M.C.; Muljono, R.A.B.; Léon, J.P.; Blanca, G.B.; Huub, J.M.; George, J.W.; John, F.B.; Roert, V. Isochorismate synthase transgenic expression in *Catharanthus roseus* cell suspension. *Plant Physiol. Biochem.* **2001**, *39*, 595–602. [[CrossRef](#)]
90. Valdivia, A.C.R.; Heijden, R.; Verpoorte, R. Elicitor-mediated induction of anthraquinone biosynthesis and regulation of isopentenyl diphosphate isomerase and farnesyl diphosphate synthase activities in cell suspension cultures of *Cinchona robusta* How. *Planta* **1997**, *203*, 155–161. [[CrossRef](#)]
91. Hu, Z.; Gallo, S.M. Identification of interacting transcription factors regulating tissue gene expression in human. *BMC Genom.* **2010**, *11*, 49. [[CrossRef](#)] [[PubMed](#)]

92. Lovisetto, A.; Masiero, S.; Rahim, M.A.; Mendes, M.A.; Casadoro, G. Fleshy seeds form in the basal *Angiosperm Magnolia grandiflora* and several MADS-box genes are expressed as fleshy seed tissues develop. *Evol. Dev.* **2015**, *17*, 82–91. [[CrossRef](#)] [[PubMed](#)]
93. Zhang, Y.; Liang, W.; Shi, J.; Xu, J.; Zhang, D. MYB56 encoding a R2R3 MYB transcription factor regulates seed size in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* **2013**, *55*, 1166–1178. [[CrossRef](#)] [[PubMed](#)]
94. Zhang, Q.; Zhang, X.; Wang, S.; Tan, C.; Zhou, G.; Li, C. Involvement of Alternative Splicing in Barley Seed Germination. *PLoS ONE* **2016**, *11*, e0152824. [[CrossRef](#)] [[PubMed](#)]
95. Lopes-Caitar, V.S.; de Carvalho, M.C.; Darben, L.M.; Kuwahara, M.K.; Nepomuceno, A.L.; Dias, W.P.; Abdelnoor, R.V.; Marcelino-Guimarães, F.C. Genome-wide analysis of the Hsp20 gene family in soybean: Comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *BMC Genom.* **2013**, *14*, 577. [[CrossRef](#)] [[PubMed](#)]
96. Sarkar, A. Heat shock factor gene family in rice: Genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol. Biochem.* **2009**, *47*, 785–795.
97. Liu, Z.; Zhu, Q.; Li, J.; Zhang, G.; Jiamahate, A.; Zhou, J.; Liao, H. Isolation, structure modeling and function characterization of a trypsin inhibitor from *Cassia obtusifolia*. *Biotechnol. Lett.* **2015**, *37*, 863–869. [[CrossRef](#)] [[PubMed](#)]
98. Scharf, K.D.; Siddique, M.; Vierling, E. The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Acid proteins). *Cell Stress Chaperones* **2001**, *6*, 225–237. [[CrossRef](#)]
99. Sarkar, N.K.; Kim, Y.K.; Grover, A. Rice sHsp genes: Genomic organization and expression profiling under stress and development. *BMC Genom.* **2009**, *10*, 393. [[CrossRef](#)] [[PubMed](#)]
100. Lang, S.; Liu, X.; Xue, H.; Wang, X. Functional characterization of BnHSFA4a as a heat shock transcription factor in controlling the re-establishment of desiccation tolerance in seeds. *J. Exp. Bot.* **2017**, *68*, 2361–2375. [[CrossRef](#)] [[PubMed](#)]
101. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)] [[PubMed](#)]
102. Liu, X.; Mei, W.; Soltis, P.S.; Soltis, D.E.; Barbazuk, W.B. Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol. Ecol. Resour.* **2017**, *17*, 1243–1256. [[CrossRef](#)] [[PubMed](#)]
103. Song, T.; Liu, Z.B.; Li, J.J.; Zhu, Q.K.; Tan, R.; Chen, J.S.; Zhou, J.Y.; Liao, H. Comparative transcriptome of rhizome and leaf in *Ligusticum chuanxiong*. *Plant Syst. Evol.* **2015**, *301*, 2073–2085. [[CrossRef](#)]
104. Kong, L.; Zhang, Y.; Ye, Z.; Liu, X.; Zhao, S.; Wei, L.; Gao, G. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **2007**, *35*, 345–349. [[CrossRef](#)] [[PubMed](#)]
105. Sun, L.; Luo, H.; Bu, D.; Guo, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [[CrossRef](#)] [[PubMed](#)]
106. Wang, L.G.; Park, H.J.; Dasari, S.; Wang, S.Q.; Kocher, J.P.; Li, W. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **2013**, *41*, e74. [[CrossRef](#)] [[PubMed](#)]
107. Flowerika, A.; Kumar, J.; Thakur, N.; Pandey, A.; Pandey, A.K.; Upadhyay, S.K.; Tiwari, S. Characterization and expression analysis of phytoene synthase from bread wheat (*Triticum aestivum* L.). *PLoS ONE* **2016**, *11*, e0162443. [[CrossRef](#)] [[PubMed](#)]
108. Liu, Z.; Zhu, Q.; Li, J.; Yu, J.; Li, Y.; Huang, X.; Wang, W.; Tan, R.; Zhou, J.; Liao, H. Selection and evaluation of reference genes for expression analysis of *Cassi*. *Biosci. Biotechnol. Biochem.* **2015**, *79*, 1818–1826. [[CrossRef](#)] [[PubMed](#)]
109. Xu, Z.; Peters, R.J.; Weirather, J.; Luo, H.; Liao, B.; Zhang, X.; Zhu, Y.; Ji, A.; Zhang, B.; Hu, S.; et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J.* **2015**, *82*, 951–961. [[CrossRef](#)] [[PubMed](#)]

