Transcriptome Analysis Provides Insights into Gingerol Biosynthesis in Ginger (Zingiber officinale)

Yusong Jiang, Mengjun Huang, Michael Wisniewski, Honglei Li, Meixia Zhang, Xiang Tao, Yiqing Liu, and Yong Zou*

ABSTRACT  Ginger (Zingiber officinale Roscoe), a perennial herb, is one of the most economically valuable plants in the Zingiberaceae family. Gingerol, as the major constituents of ginger essential oil, contributes to the unique flavor and pharmaceutical value of ginger. However, the pathway of gingerol biosynthesis has not been verified and described in ginger to help understand the biosynthesis of secondary metabolites in nonmodel species. In this study, the concentrations of gingerols were quantified at different stages of rhizome development and in different tissues. The results confirmed that rhizomes are the major source of gingerols and that accumulation of gingerols in the rhizome starts at an early developmental stage. We also assembled a reference ginger transcriptome, which is composed of 219,479 unigenes consisting of 330,568 transcripts and provides a high-quality genetic resource for further research. An analysis of differentially expressed genes (DEGs) identified 12,935 DEGs among several different comparisons. Five genes (curcumin synthase [CURS], cinnamate 4-hydroxylase [CYP73A], p-coumaroyl quinate/shikimate 3'-hydroxylase [CYP79A], caffeoyl-coenzyme A O-methyl transferase [CCoAOMT], and hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyl transferase [HCT]) associated with gingerol biosynthesis were identified as being significantly differentially expressed in the rhizome at an early developmental stage and all five genes were upregulated. Expression analysis revealed that different loci of these genes have become functionally specialized in different tissues and different developmental stages of the rhizome (subfunctionalization). Among the DEGs, CCoAOMT and HCT may act as gatekeepers and rate-limiting enzymes in the gingerol biosynthesis pathway and thus play an important role in regulating the biosynthesis of gingerol.

Keywords: CCoAOMT, caffeoyl-coenzyme A O-methyl transferase, CDS, coding sequence; CoA, coenzyme A; CURS, curcumin synthase; CYP73A, cinnamate 4-hydroxylase; CYP79A, p-coumaroyl quinate/shikimate 3'-hydroxylase; DEGs, differentially expressed unigenes; GO, Gene Ontology; HCT, hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyl transferase; KEGG, Kyoto Encyclopedia of Genes and Genomes; LC–MS, liquid chromatography–mass spectrometry; N50, sequence length of the shortest contig at 50% of the total genome length; PCR, polymerase chain reaction; Rm, rhizome at the mature stage; Rp, rhizome at the early developmental stage.

CONTACT Y. Jiang, m.huang@chongqing.3a.gov.cn.

*Corresponding author (nevernever107@126.com).

This is an open access article distributed under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Received 8 June 2018. Accepted 20 Aug. 2018.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. We can share the modified preproof only. Citation information: DOI: 10.3835/plantgenome2018.06.0034; Copyright © Crop Science Society of America 11:180034.
as a flavoring agent in beverages, bread, food, and the confectionery industry (Wohlmuth et al., 2005).

Ginger is also widely used in the pharmaceutical industry for its antioxidant and anti-inflammatory properties (Jafarzadeh et al., 2017), prevention of motion sickness (Grontved et al., 1988), neuroprotective and cognitive-enhancing effects (Sutalangka and Wattanathorn, 2017), and relief of nausea during pregnancy (Pongrojpay et al., 2007). A group of volatile phenolic compounds are grouped together as gingerols and represent the major constituents of ginger essential oil and contribute to the unique flavor of ginger. As an important medicinal component of ginger, gingerol exhibits many biological properties, including anticancer, antimicrobial, antioxidant, and anti-inflammatory properties, and also has various effects on the central nervous system (Fan et al., 2015; Roufogalis, 2014). Studies have reported that 6-gingerol treatment restores intestinal barrier functions that have been damaged and inhibits the proinflammatory response of dextran sodium sulfate-treated caco-2 monolayers (Chang and Kuo, 2015). Ajayi et al. (2018) reported that 6-gingerol effectively prevented oxidative damage to the colon by increasing the antioxidant status and lipid peroxidation in dextran sodium sulfate-treated mice. In addition, dextran sodium sulfate-induced chronic ulcerative colitis was prevented by the anti-inflammatory and antioxidant properties of 6-gingerol, which also helps to stabilize the Wingless/Integrated-β-catenin signaling pathway.

The pathway of gingerol biosynthesis has been described in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (map accession number: 00945). However, whether these genes exist in ginger, whether they are transcribed, and whether the whole pathway is valid in ginger has not been determined. In the genomic era, deciphering the transcriptome and genome is of significant importance to obtain an understanding of secondary metabolite biosynthesis and improvement of nonmodel species (Chen et al., 2017; He et al., 2016a, 2016b). Although several transcriptomic studies have been published (Gaur et al., 2016; Jiang et al., 2017; Prasath et al., 2014), there still have not been any studies on transcriptomes that are devoted to gingerol biosynthesis. In the present study, gingerol content was determined in three tissues at three developmental stages and transcriptome profiles in the same tissues were analyzed to explore the molecular biology underlying gingerol biosynthesis.

**MATERIALS AND METHODS**

**Plant Material and Cultivation**

Tissue culture seedlings of *Z. officinale* cv. Southwest were transferred to pots (40 by 21 by 25 cm) containing sterilized soil. The plants were grown in a greenhouse at 25°C with a relative air humidity of 60% and 14 h of (200 μEm-2s-1). Five tissues were collected: leaves at the mature stage (the third leaves on the main stem), stems at the mature stage (the main stems), rhizomes at the early development stage (the base parts of the first-level lateral branches, Rp), rhizomes at the mid-development stage (the base parts of the first-level lateral branches, Rd), and rhizomes at the mature stage (the base parts of the first-level lateral branches, Rm). The three development stages were designated according to the growth of the lateral branch of rhizome as follows: early development stage, the first-level lateral branch formation on the main rhizome about 2 mo after sowing; mid-development stage, second-level lateral branch formation on the first-level lateral branch rhizome about 3 mo after sowing; mature stage, third-level lateral branch formation on the second-level lateral branch rhizome about 4 mo after sowing. Each tissue collection consisted of three biological replicates and each replicate consisted of tissues from three different plants. All samples were rinsed with sterile distilled water and immediately placed and stored in liquid N for gingerol quantification and downstream transcriptome analysis.

**Preparation of Samples for Liquid Chromatography–Mass Spectrometry Analysis**

Gingerols (6-gingerol, 8-gingerol, and 10-gingerol; ChromaBio, Chengdu, Sichuan Province, China) were accurately weighed and dissolved with methanol to obtain a stock solution. The stock solution was diluted to 50, 100, 200, 500, and 1000 ng mL-1 to obtain a graded set of standard working solutions for use in generating a standard curve. To extract gingerol, 60 mg of sample was added to a 1.2-mL methanol–water solution (7:3, v/v) and cooled at -20°C for 2 min. Two steel balls were added to each tube and the sample was ground in a grinding mill (Tissuelyser-192, Jingxin, Shanghai, China) at 60 Hz for 2 min. The sample was then further exposed to an ultrasonic extractor (Scientz-500C, Scientz, Ningbo, Zhejiang Province, China) for 30 min, standing at -20°C for 20 min, and then centrifuged at 12,000 rpm at 4°C for 15 min. Subsequently, 200 μL of the supernatant was transferred into liquid chromatography–mass spectrometry (LC–MS) vials for LC–MS/mass spectrometry analysis.

**Quantification of Gingerol by LC–MS**

Ultra-high-performance liquid chromatography-linear ion trap high resolution mass spectrometry (Thermo-Fisher, Waltham, MA) equipped with an electrospray ionization source was used to identify and quantify gingerol content in the samples (Supplemental Fig. S1). The extract was loaded onto a Syncronis C18 column (100 by 2.1 mm, 1.7 μm particle size; Thermo-Fisher) with a methanol–water gradient elution containing 0.1% (v/v) formic acid for the separation and a flow rate of 0.20 mL min-1. The injection volume was 1.00 μL and the column temperature was set at 40.0°C. The gradient elution procedures are listed in Supplemental Table S1.

The mass spectrometric data were collected with a LTQ-XL linear ion trap mass spectrometer (Thermo-Fisher). The source temperature was set at 350°C, with a sheath gas flow of 30 L h-1 and an auxiliary gas flow of
10 L h⁻¹. The ion spray voltage was set at ±4000 V and the tube lens voltage was 75V. The mass spectrometry parameters are listed in Supplemental Table S2. The raw data were analyzed with Xcalibur software (ThermoFisher), which automatically integrates the data and identifies the compound. Results were checked manually.

Linear regression standard curves were plotted on the basis of the mass spectrum peak area of the analyte as the ordinate and the concentration of the analyte as the abscissa. The linear equation and the area of the mass spectrum peak of the analyte from the samples were used to calculate the concentration of the three types of gingerol in the samples.

RNA Extraction, cDNA Library Preparation, and Sequencing

RNA was isolated and cDNA libraries were constructed from all five tissues (three replicates of each tissue) according to the Illumina TruSeq (Illumina, San Diego, CA) RNA library protocol. Library sequencing was done on a HiSeq 2500 (Illumina) platform to obtain 150 bp of paired-end reads (Table 1). The raw sequencing reads were submitted to NCBI (BioProject: PRJNA477462).

RNA sequencing reads are available in the NCBI BioProject under accession PRJNA477462. Assembled transcripts, annotation, DEGs, and other information are deposited in Figshare (doi:10.6084/m9.figshare.6858347.v1).

De Novo Assembly of Transcripts and Unigenes

The raw reads from each library were first filtered by removing adapter sequences and reads with a high percentage of low-quality bases (Q-value < 20) with Trimmomatic software version 0.36 (Bolger et al., 2014). The filtered reads were corrected with Rcorrector software (Song and Florea, 2015) to erase sequencing errors. After filtering and correction, the clean reads from all of the libraries were pooled and assembled into transcripts with Trinity version 2.2.0 (Haas et al., 2013) with the default parameters. Transcripts shorter than 300 bp were discarded and the longest transcript in each cluster (unigene) was selected as the representative of the unigene (Fig. 2). The completeness of the assemblies was evaluated by benchmarking universal single-copy orthologs in BUSCO version V2 (Supplemental Table S4) (Simao et al., 2015).

Annotation of Ginger Transcripts

The coding sequence (CDS) and the translated protein sequences of unigenes were predicted with TransDecoder version r20140704 (http://transdecoder.github.io/, accessed 26 Sept. 2018). The longest CDS of each transcript was translated into an amino acid sequence based on the standard amino acid codon table. Unigenes containing CDSs were defined as protein-coding unigenes. Proteins were functionally annotated by blastp (Camacho et al., 2009) based on queries of functional databases, including the NCBI nonredundant and SwissProt database (Boeckmann et al., 2003). The e-value cutoff was set as 1 × 10⁻⁵ and low-quality alignments were removed (an identity less than 30% and high score pairs of less than 30 amino acids), and the top hit for each protein was selected as its functional annotation. Proteins were further assigned to Gene Ontology (GO) (Ashburner et al., 2000) and KEGG terms (Kanehisa and Goto, 2000) with Blast2GO software (Conesa et al., 2005) and KOBAS version 3.0 (Wu et al., 2006), respectively. A total of 67.61% of the proteins had functional annotations assigned to them (Supplemental Table S5).

Identification of Differentially Expressed Unigenes

Clean reads were aligned to the assembled transcripts with Bowtie2 (Langmead and Salzberg, 2012) and the expression levels of unigenes were calculated and normalized via Expectation Maximization in RSEM (Li and Dewey, 2011) (Supplemental Table S6). Significantly differentially expressed unigenes (DEGs) between compared tissues were identified with the edgeR package (Robinson et al., 2010) from the R language, with a false discovery rate threshold of <0.001 and a fold change of >4. Venn diagrams of DEGs from each comparison were generated at http://bioinformatics.psb.ugent.be/webtools/Venn/ (accessed 26 Sept. 2018). Genes with similar expression levels were clustered with the “unweighted pair group method with arithmetic mean” hierarchical method implemented with the ‘dist’ and ‘hclust’ functions in R language to explore the expression pattern of DEGs in each comparison. To explore the function of the DEGs, GO enrichment analyses and KEGG enrichment were performed with a hypergeometric distribution from the R language (Alexa et al., 2006). Both the GO and KEGG terms were considered significant when the corrected p-values ≤ 0.05. The pathway of gingerol biosynthesis was adopted from KEGG map00945 (stilbenoid, diarylheptanoid, and gingerol biosynthesis).

Table 1. Transcriptome sequencing data statistics for the RNA sequencing libraries made from different tissues of ginger (Z. officinale).

<table>
<thead>
<tr>
<th>Sample†</th>
<th>Replicates</th>
<th>Reads</th>
<th>Bases</th>
<th>Mapped reads</th>
<th>Mapped ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rp</td>
<td>Rp-1</td>
<td>48,691,310</td>
<td>7071,213,675</td>
<td>39,459,094</td>
<td>81.04%</td>
</tr>
<tr>
<td></td>
<td>Rp-2</td>
<td>58,240,654</td>
<td>8469,710,423</td>
<td>46,303,558</td>
<td>79.50%</td>
</tr>
<tr>
<td></td>
<td>Rp-3</td>
<td>47,617,416</td>
<td>6941,219,498</td>
<td>38,015,930</td>
<td>79.84%</td>
</tr>
<tr>
<td>Rd</td>
<td>Rd-1</td>
<td>43,738,784</td>
<td>6398,835,043</td>
<td>34,801,224</td>
<td>79.57%</td>
</tr>
<tr>
<td></td>
<td>Rd-2</td>
<td>63,477,454</td>
<td>9061,811,871</td>
<td>51,008,618</td>
<td>80.36%</td>
</tr>
<tr>
<td></td>
<td>Rd-3</td>
<td>55,165,618</td>
<td>7844,801,197</td>
<td>42,160,440</td>
<td>76.43%</td>
</tr>
<tr>
<td>Rm</td>
<td>Rm-1</td>
<td>46,997,012</td>
<td>6697,368,101</td>
<td>43,738,784</td>
<td>79.72%</td>
</tr>
<tr>
<td></td>
<td>Rm-2</td>
<td>52,422,348</td>
<td>7474,182,428</td>
<td>46,303,558</td>
<td>77.29%</td>
</tr>
<tr>
<td></td>
<td>Rm-3</td>
<td>51,979,742</td>
<td>7552,486,068</td>
<td>41,381,968</td>
<td>79.61%</td>
</tr>
<tr>
<td>S</td>
<td>S-1</td>
<td>63,587,014</td>
<td>9091,210,676</td>
<td>49,835,860</td>
<td>78.37%</td>
</tr>
<tr>
<td></td>
<td>S-2</td>
<td>49,536,470</td>
<td>7039,969,035</td>
<td>38,967,882</td>
<td>78.67%</td>
</tr>
<tr>
<td></td>
<td>S-3</td>
<td>46,807,418</td>
<td>6687,875,242</td>
<td>36,846,798</td>
<td>78.72%</td>
</tr>
<tr>
<td>L</td>
<td>L-1</td>
<td>61,622,306</td>
<td>8953,438,851</td>
<td>48,936,174</td>
<td>79.41%</td>
</tr>
<tr>
<td></td>
<td>L-2</td>
<td>59,846,328</td>
<td>8643,023,082</td>
<td>46,554,328</td>
<td>78.26%</td>
</tr>
<tr>
<td></td>
<td>L-3</td>
<td>59,802,164</td>
<td>8690,750,433</td>
<td>46,989,564</td>
<td>78.58%</td>
</tr>
</tbody>
</table>

† L, leaf at the mature stage; S, stem at the mature stage; Rp, rhizome at the early development stage; Rd, rhizome at the mid-development stage; Rm, rhizome at the mature stage.
Validation of Gene Expression Levels via Reverse Transcription–Quantitative Polymerase Chain Reaction

Fifteen DEGs were randomly selected and gene-specific primers for each unigene were designed with the Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/, accessed 26 Sept. 2018). The Actin1 gene was selected for internal controls, with primer sequences of 5'-TGGCATCTCTCAGCACATTCC-3' and 5'-TGCA-CAATGGATGGGTCAGA-3'. The total RNA of three sequenced samples (Leaf 1, Leaf 2, and Leaf 3) was treated with DNase and first-strand cDNA was synthesized from 250 ng of total RNA with the reverse transcription kit (PrimeScript RT Reagent Kit with gDNA Eraser, Takara, Dalian, Liaoning Province, China) according to the manufacturer’s instructions. Reverse transcription–quantitative real-time polymerase chain reaction (PCR) was performed via the qTOWER3G touch Real-Time PCR System (Analytik Jena AG, Jena, Germany) in a total volume of 20 μL containing 10 ng of cDNA template, 1 × TB Green Premix Ex Taq (Perfect Real Time, Takara), and 500 nM of each primer. Serial dilutions of each cDNA sample were used to generate a quantitative PCR standard curve to calculate the corresponding PCR efficiencies. The following PCR conditions were used: initial denaturation at 95°C for 30 s, followed by 40 cycles of denaturation at 95°C for 5 s and primer annealing and DNA extension at 60°C for 30 s.

RESULTS AND DISCUSSION

Quantification of Gingerol Content in Different Tissues at Different Developmental Stages

Gingerols are phenolic components and the major sources of pungent flavor in ginger. They have been reported to have various pharmacological properties, including anti-inflammatory, anticancer, and antioxidant activity (Aeschbach et al., 1994; Shukla et al., 2007; Wang et al., 2003). The rhizome is the tissue with the greatest concentrations of gingerols (Jiang et al., 2005). The concentration of three types of gingerols (6-gingerol, 8-gingerol, and 10-gingerol) was quantified in different tissues and at different developmental stages of rhizome development via LC–MS (Fig. 1, Supplemental Table S3). The results confirmed that the rhizomes (mean value of 104.39 μg g⁻¹) have much higher concentrations of gingerols than the stems (mean value of 0.84 μg g⁻¹) and leaf tissues (mean value of 4.13 μg g⁻¹). Among the three different gingerols, 6-gingerol was the major constituent with a mean value of 195.87 μg g⁻¹, whereas the concentration of other gingerols, such as 8-gingerol (mean value of 46.31 μg g⁻¹) and 10-gingerol (mean value of 70.99 μg g⁻¹), were much lower. Notably, all three gingerols began to accumulate in rhizome tissues at an early developmental stage.

Transcript Assembly and Annotation

A total of 15 libraries (five tissue samples, with three biological replicates for each tissue sample) were sequenced on the HiSeq 2500 platform (Illumina). After the removal of adaptor-contaminated, low-quality, and duplicated reads, as well as the correction of potential sequencing errors with Rcorrector (Song and Florea, 2015), 116.6 GB of clean sequence data was obtained, with read numbers ranging from 46 million to 63 million per library (Table 1). Clean reads were pooled and then assembled with Trinity software (Haas et al., 2013). The final assemblies include 219,479 unigenes consisting of 330,568 transcripts. The total size of the transcripts was 339,955,560 bp with
the sequence length of the shortest contig at 50% of the total genome length (the N50 value) being 1531 bp and the total size of the unigenes was 179,621,854 bp with a N50 value of 1077 bp (Fig. 2). BUSCO (Simao et al., 2015) was used to assess the completeness of the transcripts and the results indicated that 90.6% of 1440 complete embryo-phyte-conserved orthologs were present in the assembled transcripts (Supplemental Table S4). A larger N50 value represents more completeness within the assembly. Although the N50 value of the unigenes obtained in the present study was smaller than the value (1251 bp) reported for ginger by Gaur et al. (2016), it is greater than the N50 of 943 bp reported by Prasath et al. (2014) and the N50 of 589 bp reported by Jiang et al. (2017).

A total of 172,354 CDSs were predicted in the transcripts. For each protein-coding transcript (the transcript with a predicted CDS), only the longest CDS was retained. Finally, we obtained a total of 141,581 protein-coding transcripts (42.83% of the total transcripts), which were from 81,100 unigenes. The proteins were functionally annotated via BLASTP by querying the sequences against several functional databases, including KEGG (Kanehisa and Goto, 2000), GO (Harris et al., 2004), Swiss-Prot (Boeckmann et al., 2003), and the NCBI nonredundant protein database. Among the functionally annotated sequences, 26,476 annotated unigenes were identified in 361 KEGG pathways and 32,727 unigenes were annotated by GO terms. Totally, 54,826 (67.61% of all unigenes) protein-coding unigenes were functionally annotated (Supplemental Table S5).

Identification of DEGs in Comparisons among Different Tissue Types and Different Developmental Stages

Clean reads from the 15 sequenced libraries were mapped to the reference transcripts with Bowtie2 (Langmead and Salzberg, 2012). The expression levels of transcripts and unigenes were calculated with RSEM (Li and Dewey, 2011) software program. A Pearson correlation analysis was performed on the expression matrices of unigenes (Fig. 3A). A high degree of correlation was observed among the biological replicates, indicating that the quality and robustness of the sequenced libraries was good.

The correlation analysis also indicated that the rhizomes sampled at an early development stage had a pattern of gene expression that was quite different from the other samples. This result is consistent with the number of DEGs obtained from comparisons of different developmental stages of the rhizome (Fig. 3B). In total, 2707 DEGs were identified in the comparison between the Rp and Rd rhizome tissues, including 2332 upregulated genes and 375 downregulated genes. In total, 324 DEGs, including 209 upregulated and 115 downregulated genes, were identified in the comparison between the Rd and Rm rhizome samples. Lastly, 3629 DEGs were identified in the comparison between the Rm and Rp rhizome samples, including 628 upregulated and 3001 downregulated genes.

We also compared the DEGs from three different tissues of ginger (rhizomes, leaves, and stems) (Fig. 3C). In total, 12,681 DEGs were identified. Among these DEGs, 1349 genes were common among the three comparisons, whereas 8322 genes were differentially expressed in at least two comparisons. In total, 6197 DEGs were identified in the comparison between rhizome and stem tissues, including 3783 upregulated and 2414
downregulated genes. A total of 10,856 DEGs was identified in the comparison between rhizome and leaf tissues, including 6332 upregulated and 4524 downregulated genes. Lastly, 5299 DEGs were identified in the comparison between leaf and stem tissues, including 2731 upregulated and 2568 downregulated genes.

To validate the accuracy of our quantification and differential expression analyses, a subset of 15 DEGs was randomly selected for validation via quantitative real-time PCR. The log2-transferred fragments per kb of exon model per million mapped reads obtained from RNA sequencing were compared with the crossing threshold obtained from quantitative real-time PCR (Supplemental Table S10). An inverse relationship between fragments per kb of exon model per million mapped reads and crossing threshold values was observed, which is reasonable, as a lower crossing threshold value indicates an increased initial amount of target mRNA (Supplemental Fig. S2). A correlation coefficient of 0.87 indicates good consistency between RNA sequencing and quantitative real-time PCR.

**Enrichment Analysis of DEGs**

Enrichment analyses were conducted using GO terms and KEGG pathways to explore the function of DEGs identified in each comparison. Sixty-eight KEGG pathways were significantly enriched (false discovery rate < 0.05) based on a hypergeometric test implemented in the R programming language. Among these pathways, 31 were related to metabolism (Fig. 4A, Supplemental Table S7). Among the KEGG pathways, three (map00940: phenylpropanoid biosynthesis; map00941: flavonoid biosynthesis; map00945: stilbenoid, diarylheptanoid, and gingerol biosynthesis) overlapped at the step where cinnamoyl-coenzyme A (CoA) is converted to feruloyl-CoA by caffeoyl-CoA : O-methyl transferase encoded by a caffeoyl-CoA O-methyl transferase and cinnamoyl quinate 3'-monooxygenase, which are encoded by hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase and coumaroyl quinate 3'-monooxygenase, which are encoded by hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT, EC:2.3.1.133) and cinnamoyl-CoA by shikimate O-hydroxycinnamoyl-transferase and coumaroyl quinate 3'-monooxygenase, which are encoded by hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT, EC:2.3.1.133). Caffeoyl-CoA is then converted to feruloyl-CoA by caffeoyl-CoA O-methyl transferase encoded by a caffeoyl-CoA O-methyl transferase (CCoAOMT, EC:2.1.1.104) gene. Finally, feruloyl-CoA is converted to 1-dehydro-[6]-gingerdione and eventually to 6-gingerol by some yet uncharacterized enzymes.

In the current study, five genes associated with gingerol biosynthesis were identified as being significantly differentially expressed: CTR, CTP73A, CYP98A, CCoAOMT, and HCT. We also identified three CTR genes in ginger, including CTR1, CTR2, and CTR3. CTR1 and CTR2 prefer feruloyl-CoA as a substrate and can synthesize curcumin or desmethoxycurcumin, whereas CTR3 prefers both feruloyl-CoA and p-coumaroyl-CoA and synthesizes curcumin, bisdemethoxycurcumin, and desmethoxycurcumin (Katsuyama et al., 2009). The availability of the required substrates and the expression levels of the three enzymes may potentially influence the composition of curcuminooids. We did not focus, however, on these three CTR genes in ginger, as curcumin is not a major constituent in ginger. We rather focused on the four genes (CTP73A, CYP98A, CCoAOMT, and HCT) that are involved in the synthesis of gingerol.

Both CYP98A and CTP73A belong to the P450 gene family. CYP98A is one of the key genes associated with the phenylpropanoid pathway that is responsible for the synthesis of flavonoids, phenolics, and lignin (Rana et al.,
Karamat et al. (2012) found that overexpression of CYP98A22 positively impacted the concentration of furo- nocoumarins in *Ruta graveolens* L. *p*-Coumaroyl-CoA, a plant-specific cytochrome, is a necessary intermediate in the biosynthesis of many secondary metabolites (such as gingerol, curcumin, and methyleugenol) and is synthesized as a product of CYP73A. In our study, five genes were annotated as CYP73A genes at different loci; among them, four were significantly differentially expressed. As shown in Fig. 5B, the expression levels of two of the CYP73A genes were much higher than those of the other two CYP73A loci. Notably, the expression level
Fig. 5. The expression profile of the genes coding for enzymes in the gingerol biosynthesis pathway. (A) The metabolism pathway was adopted from Kyoto Encyclopedia of Genes and Genomes (KEGG) (map00945: stilbenoid, diarylheptanoid, and gingerol biosynthesis) and differentially expressed genes are colored orange. An asterisk indicates that a unigene is significantly differentially expressed in the comparisons. Each bar represents one biological replicate. Rhizome at the early development stage (Rp), rhizome at the mid-development stage (Rd), rhizome at the mature stage (Rm), stem (S), and leaf (L) are indicated by yellow, red, peach, blue and purple colors, respectively. (B) Heatmap of the expression level of differentially expressed unigenes involved in gingerol biosynthesis. CYP73A, cinnamate 4-hydroxylase; HCT, hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyl transferase; CYP98A, p-coumaroyl quinate/shikimate 3'-hydroxylase; CCoAOMT, caffeoyl-coenzyme A O-methyl transferase; CURS, curcumin synthase.
of these different CYP73A genes varied greatly in the stem samples. These results suggest that different CYP73A loci have become functionally specialized in different tissues and at different developmental stages of the rhizome.

CCoAOMT is responsible for the methylation of the meta-hydroxyl group in caffeoyl-CoA in the monolignols pathway (https://www.genome.jp/dbget-bin/www_bget?K00588, accessed 26 Sept. 2018). Their ring methoxylolation status is characteristic of guaiacyl or syringyl units in lignin (Walker et al., 2016). Compared with stem and leaf tissues, it had a much higher level of expression level in the rhizome at all three developmental stages. This is consistent with the fact that rhizomes have much higher concentrations of gingerols than stem and leaf tissues. Therefore, CCoAOMT may act as a gatekeeper in the gingerol biosynthesis pathway and play an important role in regulating the biosynthesis of gingerol.

HCT, a type of acyltransferase, catalyzes the hydroxylon of trans-cinnamic acid to p-coumaric acid. It plays a key role in phenylpropanoid metabolism and lignin biosynthesis and possibly anchors the phenylpropanoid enzyme complex to the endoplasmic reticulum (Ro et al., 2001). Zhang et al. (2017) found that HCT may be involved in the formation of stone cells in pear (Pyrus bretschneideri Rehder) fruit, which affect their texture. In our study, three HCT loci were identified and all of them were DEGs. Interestingly, the first two loci were expressed at a very low level in the rhizomes but highly expressed in leaf tissues, whereas the third locus was highly expressed in the early stage of rhizome development and expressed at a very low level in the subsequent developmental stages of rhizomes, as well as in mature leaf and stem tissues. Therefore, paralogous HCT genes in ginger may have undergone functional specialization in different tissues and/or different developmental stages. The HCT third locus may be expressed exclusively in the early stages of rhizome development. Gingerol began to accumulate in the rhizome at an early stage of development (Fig. 1) and HCT genes had the lowest expression level relative to other genes in the gingerol biosynthesis pathway (Fig. 5B) in rhizome tissues. Therefore, the acyltransferase encoded by HCT, especially the third locus of HCT, may be a rate-limiting enzyme in the gingerol biosynthesis pathway and thus also play an important role in regulating the biosynthesis of gingerol.

CONCLUSIONS

Ginger is an important spice and medicinal plant. Gingerol is the major constituent of ginger essential oil and contributes to the unique flavor of ginger. Gingerol also has many biological properties. The LC–MS analysis revealed that rhizomes have a much higher concentration of gingerols than stem and leaf tissues. Interestingly, all three gingerols identified in the current study began to accumulate in rhizome tissues at an early developmental stage. De novo transcriptome sequencing and assembly generated a total of 219,479 unigenes, with 81,100 protein-coding unigenes. Pearson correlations calculated for expression levels indicated that the rhizome at an early stage of development had a pattern of gene expression that was significantly different from the pattern of gene expression found in the other sampled tissues. In the rhizome at early development stage, five genes (CURS, CTP73A, CYP98A, CCoAOMT, and HCT) associated with gingerol biosynthesis were identified as being significantly differentially expressed and all of them were upregulated in this stage. By examining the expression level of these DEGs, we found evidence of subfunction- alization in different tissues and different developmental stages. Moreover, CCoAOMT and HCT exhibited characteristics that suggest they may function as gatekeepers and key rate-limiting enzymes, indicating their important role in the regulation of gingerol biosynthesis.

Supplemental Information

Supplemental Fig. S1: Chromatograms of three gingerols: A, 6-gingerol; B, 8-gingerol; C, 10-gingerol.

Supplemental Fig. S2: Correlation of RNA sequencing (RNA-Seq) and quantitative real-time PCR (qRT-PCR). The correlations between the RNA-Seq fragments per kb of exon model per million mapped reads values and the corresponding qRT–PCR crossing threshold (ΔCT) values are shown. The dashed line, the associated equation, and goodness of fit value were generated.

Supplemental Table S1: Procedures for gradient elution.

Supplemental Table S2: Parameters for mass spectrometry.

Supplemental Table S3: The concentration of gingerols determined by LC–MS analyses.

Supplemental Table S4: Assessment of the results of the transcripts assembly with BUSCO.

Supplemental Table S5: Functional annotation of the ginger transcriptome.

Supplemental Table S6: Raw counts of ginger unigenes calculated by their RSEM.

Supplemental Table S7: Enriched functional groups related to the KEGG metabolism process.

Supplemental Table S8: Enriched functional groups related to the GO secondary metabolic process.

Supplemental Table S9: Differentially expressed unigenes related to gingerol biosynthesis.

Supplemental Table S10: Experimental validation via quantitative real-time PCR (qRT-PCR) and primer sequences for qRT-PCR expression analysis.

Data Availability

RNA sequencing reads are available in the NCBI Bio-Project under accession PRJNA477462. Assembled transcripts, annotation, differentially expressed genes, and other information are deposited in Figshare (doi:10.6084/m9.figshare.6858347.v1).

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.
Author Contributions
MW, YJ, and YZ designed the experiments; YJ and MH wrote the manuscript; MW and YZ revised the manuscript; HL, XT, and YL prepared the samples, grew the ginger, and quantified gingerol contents; YJ, MH, and XT analyzed the transcriptome data.

ACKNOWLEDGMENTS
This work was supported by the grants from the National Natural Science Foundation of China (No. 31501273), the Scientific and Technological Key Research Program of Chongqing Municipal Education Commission (No. KJZD-K20180301), the Chongqing Research Program of Basic Research and Frontier Technology (Nos. CSTC2015jcyA00003, CSTC2015jcyA00004, and CSTC2017jcyAX0140), and the Chongqing Social Undertakings and the People’s Livelihood Security Program (CSTC2016shmsxsz0483).

REFERENCES


