

# Transcript Assembly and Quantification by RNA-Seq Reveals Differentially Expressed Genes between Soft-Endocarp and Hard-Endocarp Hawthorns

Hongyan Dai<sup>1</sup>\*, Guofen Han<sup>1</sup>\*, Yujiao Yan<sup>2</sup>\*, Feng Zhang<sup>2</sup>, Zhongchi Liu<sup>3</sup>, Xiaoming Li<sup>1</sup>, Wenran Li<sup>1</sup>, Yue Ma<sup>1</sup>, He Li<sup>1</sup>, Yuexue Liu<sup>1</sup>, Zhihong Zhang<sup>1\*</sup>

**1** College of Horticulture, Shenyang Agricultural University, Shenyang, China, **2** College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang, China, **3** Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland, United States of America

## Abstract

Hawthorn (*Crataegus* spp.) is an important pome with a long history as a fruit, an ornamental, and a source of medicine. Fruits of hawthorn are marked by hard stony endocarps, but a hawthorn germplasm with soft and thin endocarp was found in Liaoning province of China. To elucidate the molecular mechanism underlying the soft endocarp of hawthorn, we conducted a *de novo* assembly of the fruit transcriptome of *Crataegus pinnatifida* and compared gene expression profiles between the soft-endocarp and the hard-endocarp hawthorn varieties. *De novo* assembly yielded 52,673 putative unigenes, 20.4% of which are longer than 1,000 bp. Among the high-quality unique sequences, 35,979 (68.3%) had at least one significant match to an existing gene model. A total of 1,218 genes, represented 2.31% total putative unigenes, were differentially expressed between the soft-endocarp hawthorn and the hard-endocarp hawthorn. Among these differentially expressed genes, a number of lignin biosynthetic pathway genes were down-regulated while almost all the flavonoid biosynthetic pathway genes were strongly up-regulated, concomitant with the formation of soft endocarp. In addition, we have identified some MYB and NAC transcription factors that could potentially control lignin and flavonoid biosynthesis. The altered expression levels of the genes encoding lignin biosynthetic enzymes, MYB and NAC transcription factors were confirmed by quantitative RT-PCR. This is the first transcriptome analysis of *Crataegus* genus. The high quality ESTs generated in this study will aid future gene cloning from hawthorn. Our study provides important insights into the molecular mechanisms underlying soft endocarp formation in hawthorn.

**Citation:** Dai H, Han G, Yan Y, Zhang F, Liu Z, et al. (2013) Transcript Assembly and Quantification by RNA-Seq Reveals Differentially Expressed Genes between Soft-Endocarp and Hard-Endocarp Hawthorns. PLoS ONE 8(9): e72910. doi:10.1371/journal.pone.0072910

**Editor:** Christian Schönbach, Kyushu Institute of Technology, Japan

**Received:** April 14, 2013; **Accepted:** July 11, 2013; **Published:** September 5, 2013

**Copyright:** © 2013 Dai et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by National Natural Science Foundation of China (Grant No. 31170635). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: zhangz@syau.edu.cn

† These authors contributed equally to this work.

## Introduction

Hawthorn (*Crataegus* spp.), a genus of the Rosaceae family, is an important plant that grows in Asia, Europe, North and Central America, and northern South America. It is a small shrub or spreading tree with thorny branches, three-to five-lobed deciduous leaves, white flowers, red, orange, and yellow-green or yellow fruits [1]. Hawthorn is found widely in open woodlands, as well as in hilly montane forests. Numerous hybrids and variants exist in the *Crataegus* genus, and the total number of species is estimated at between 140 and 200 [2]. Hawthorn fruit has been consumed for 2,500 years in China, primarily to improve digestion and decrease food stasis as hawthorn fruit is rich in crataegus acid [3]. The fruit may be eaten raw or processed into sauce, slices, or juice. Besides its wide use as an ornamental, hawthorn is considered one of the oldest pharmaceutical plants, and has been used as an alternative therapy for a variety of cardiovascular conditions [1,4].

In botany, hawthorn is a pome, a type of fruit produced by flowering plants in the Rosaceae family. Fruits of hawthorn are marked by hard stony endocarps that are readily dispersed by

larger birds and small rodents [5,6]. Both the wild germplasm and cultivated varieties of *Crataegus* are abundant in China [7]. ‘Ruanheshanzha’, the soft-endocarp hawthorn, is a special germplasm of *Crataegus pinnatifida*, first found in Liaoning province of China in 1958 by the researchers of Institute of Horticulture, Liaoning Academy of Agricultural Sciences, China [8]. The endocarps of ‘Ruanheshanzha’ are soft, thin (Figure 1), and edible. The seeds can be germinated after one winter stratification. In contrast, the common hawthorn’s seeds with hard wood-like endocarps (pyrenes) usually germinate after 2 to 3 years of stratification [9]. In addition, the true seed (stones containing embryos) of ‘Ruanheshanzha’ reaches 75.5% [8], higher than those of other *Crataegus* accessions, whose true seeds are usually less than 60% [10,11].

The pyrene, usually called stone for *Prunus* species, is formed through lignification of the fruit endocarps layer, a feature of a broader class of plants [12]. Ryugo (1963) first recognized that peach stones contained lignin in the early 1960s [13]. Lignin is a compound unique to plants. Over the years, most enzymes in the lignin biosynthetic pathway and a number of potential regulatory



**Figure 1. Comparisons between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).** (a) Young fruit (23 DAB) of H8 (left) and S7 (right). (b) Ripen fruits (118 DAB) of H8 (left) and S7 (right). (c) Pyrenes (left), seeds (middle) and broken endocarp (right) of H8 (up) and S7 (down). Ba = 1 cm.

doi:10.1371/journal.pone.0072910.g001

steps in the pathway have been identified [14]. Lignin is formed from the phenylpropanoid pathway, the end products of which are coniferyl and sinapyl alcohols. These lignin monomers serve as the basis of lignification, a process lignin polymer is formed via oxidative processes guided by peroxidases and laccases [12]. Endocarp lignification in *Arabidopsis* has been well studied in relation to dehiscence [15], but the mechanism of pyrene hardening has only been investigated to a limited extent. Some enzymes in the composition and formation of stone or pyrene have been examined [16,17], and the molecular basis for stone formation during early peach fruit development was investigated with microarrays [12].

RNA-Seq is a recently developed high-throughput sequencing method that produces millions of short cDNA reads [18]. The reads are aligned to a reference genome or reference transcripts, or are *de novo* assembled without the genomic sequence information to produce a genome-scale transcription map consisting of both transcript structure and gene expression levels [19]. RNA-Seq is a powerful and accurate tool for quantifying gene expression levels, widely regarded as more superior than the microarray-based methods [20]. Specifically, RNA-Seq can reveal novel transcribed regions, splicing isoforms, single nucleotide polymorphisms (SNPs),

simple sequence repeats (SSRs), and the precise location of transcription boundaries [21]. RNA-Seq can be used in transcriptome profiling of species with no genome sequencing data.

In this work, we present a *de novo* assembly of the fruit transcriptome of *C. pinnatifida* using Illumina-based RNA-seq data. Differential gene expression between the soft-endocarp variety and the common hard endocarp variety was investigated to reveal differential regulation of key pathways.

## Results

### Temporal Pattern of Lignin Deposition in the Hawthorn Endocarp

In order to identify the critical pyrene developmental times of hard- and soft-endocarp hawthorns, the developmental time of lignin deposition was studied by staining lignin with phloroglucinol-HCl solution. Lignin deposition was first detected in the region covering seeds of fruits of H8 (hard endocarp accession) at 23 DAB. The H8 endocarps became substantially hardened by 35 DAB, after which they could not be cut with a knife. No staining was observed in tissues other than the endocarp except a few scattered vascular strands. In contrast, S7 (soft endocarp



**Figure 2. Progression of lignin deposition in developing hard-endocarp hawthorn fruit (top row) and soft-endocarp hawthorn (bottom row).** Cross-sectioned fruits were stained with phloroglucinol-HCl for lignin (red colour). Numbers indicate DAB. doi:10.1371/journal.pone.0072910.g002

accession), did not show any lignin deposition until 31 DAB, and the endocarps were very thin even at 43 DAB (Figure 2).

#### De novo Assembly and Assessment of the Illumina ESTs

For RNA-Seq analysis, two cDNA libraries, H8 and S7, were made from 23 DAB fruits of hawthorn. After removing low-quality reads and trimming adapter sequences, 11,538,395 and 14,659,624 reads (76-bp in size) were obtained for H8 and S7, respectively, encompassing 2,140,589,552 and 2,743,145,672 total nucleotides (nt) respectively (Table 1). *De novo* assembly was carried out by Trinity, an assembly software to build the reference-free full-length transcription, designed specifically for high-throughput RNA sequencing [22]. The mean length of contigs was about 130 bp, and the numbers of >200 bp contigs were 460,119 for H8 and 515,118 for S7 (Table 1). The transcripts were constructed by the Butterfly program of Trinity. We obtained 54,662 transcripts for H8, and 62,653 transcripts for S7, with average lengths of 756 and 846 bp, respectively (Table 1).

These transcripts were assembled into 39,663 putative unigenes for H8 and 41,723 putative unigenes for S7, with the mean length of 656 and 703 bp, respectively (Table 1). After combining the unigene data from H8 and S7, the unigene database of the hawthorn was established and contained 52,673 putative unigenes. The sequence information of all unigenes has been deposited in

the National Center for Biotechnology Information (NCBI) under the accession GALU000000000. The mean length of the putative unigenes was 674 bp. Among all putative unigenes of hawthorn, 10,744 (7461+3283) putative unigenes have lengths of more than 1,000 bp, representing 20.4% (10,744/52,673) of total putative unigenes (Table 2). The size distribution of the assembly unigenes is shown in Figure S1.

The putative unigenes with abundant transcripts (RPKM>100) and large expression differences between H8 and S7 [ $\log_2(S7/H8) > 5$  or  $< -5$ ], were chosen for further analysis regarding EST quality. Among the 63 putative unigenes analyzed, 52 (82.5%) have complete CDS. Both hawthorn and apple are pome fruits. A high-quality draft genome sequence of the domesticated apple (*Malus domestica*) was finished [23], and predicted CDS sequences of apple were available on website (<http://genomics.research.iasma.it/>). We compared the length and identity of CDS between hawthorn and apple. Among the 52 hawthorn putative unigenes with complete CDS, 38 have a homolog in the apple genome. The identity of CDS between hawthorn and apple gene ranges from 65.63% to 99.09%, with the mean of 90.80% (Table S1).

We detected SSRs among >1,000 bp putative unigene sequences using the MISA program (Table S2). SSRs were identified from 3,174 putative unigene sequences, which represent about 29.5% (3,174/10,744) of the analyzed unigene sequences. SSRs with mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats composed about 31.3%, 39.5%, 21.3%, 0.9%, 0.1% and 0.1% of the SSRs, respectively. And 212 putative unigene sequences with the compound SSR were identified. These SSR will serve as the basis for future marker development.

**Table 1. Summary of RNA-seq and de novo assembly of *C. pinnatifida* unigenes.**

Sequences	H8	S7
Total nucleotides	2,140,589,552	2,743,145,672
Number of clean reads	11,538,395	14,659,624
Number of >200 bp contigs	460,119	515,118
Mean length of contigs (bp)	130	131
Number of >200 bp transcripts	54,662	62,653
Mean length of transcripts (bp)	756	846
N50 length of transcripts (bp)	1,237	1,421
Number of Unigenes	39,663	41,723
Mean length of Unigenes (bp)	656	703
N50 length of Unigenes (bp)	1,083	1,227

doi:10.1371/journal.pone.0072910.t001

**Table 2. Length of *C. pinnatifida* unigenes.**

Length of unigene (bp)	No. of unigenes	Percentage (%)
200–300	20,637	39.2
300–500	12,767	24.2
500–1,000	8,525	16.2
1,000–2,000	7,461	14.2
2000+	3,283	6.2
Total	52,673	

doi:10.1371/journal.pone.0072910.t002

## Functional Annotation and Characterization of Unigenes

The entire unigene sets were annotated on the basis of similarities to known or putative sequences in the public databases. Among the 52,673 high-quality unique sequences, 35,979 (68.3%) had at least one significant match to an existing gene model in BLAST searches (Table 3). Based on sequence homology, 13,922 putative unigenes of *C. pinnatifida* were categorized into 16 functional groups, belonging to three main GO ontologies: cellular component, molecular function and biological process (Figure 3). The results show that a high percentage of genes from categories of “physiological process”, “binding activity”, “enzyme activity”, “cell” and “cellular process”, and with only a few genes related to “extracellular”, “development” and “nutrient reservoir activity”.

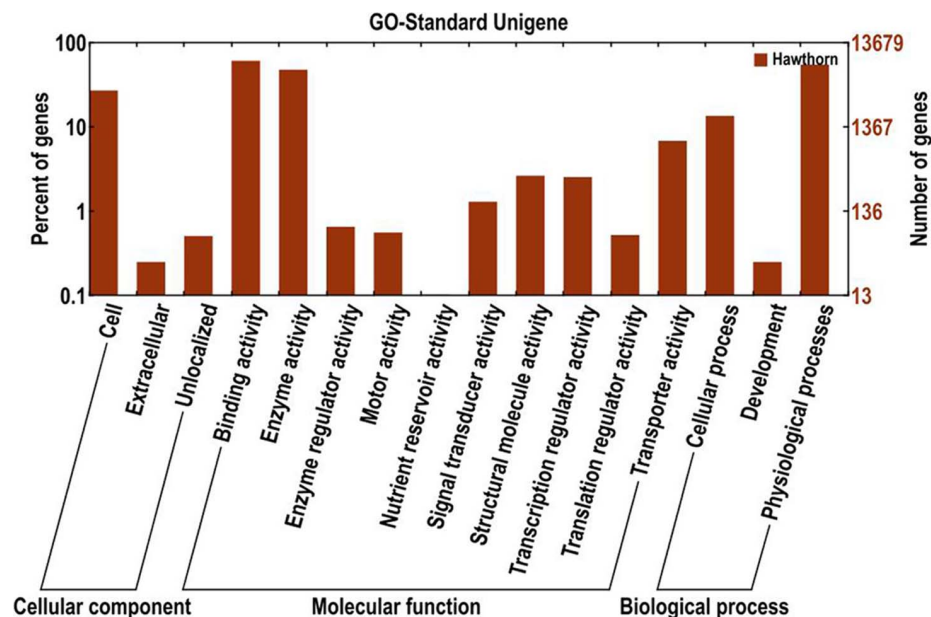
The 100 most abundant unigenes were analyzed (Table S3). Sixty-one unigenes were highly expressed both in H8 and S7. Among the 39 unigenes that were highly expressed in hard-endocarp hawthorn (H8) but not in soft-endocarp hawthorn (S7), some were lignin biosynthetic genes, including CCoA-OMT (caffeoyl-CoA O-methyltransferase) and CAD (cinnamyl-alcohol dehydrogenase) genes. By contrast, some genes encoding flavonoid biosynthetic enzymes, including CHS (chalcone synthase), CHI (chalcone isomerase) and ANR (anthocyanidin reductase) were highly expressed in S7 but not in H8.

## Transcript Differences between Hard and Soft Endocarps

We used the general chi-squared test with a random sampling model in the IDEG6 software [24] to identify genes differentially expressed in fruits between soft-endocarp hawthorn (S7) and hard endocarp control (H8). A total of 1,218 genes, represented 2.31% (1,218/52,673) total putative unigenes, were differentially expressed between H8 and S7 (twofold or more change and  $p < 0.001$ ). The detail information can be accessed through Table S4. Among the genes that were differentially expressed between the two accessions, 537 genes were up-regulated and 681 genes were down-regulated in soft-endocarp hawthorn (S7). In addition, 85.96% (1,047/1218) of the differentially expressed genes were detected in fruits of both accessions. To illustrate the differential

expression of genes detected in the fruits of hard endocarp hawthorn and soft endocarp hawthorn, we assigned GO functional classes to the differentially expressed genes with putative functions. These genes were sorted into major functional categories (Figure S2).

In an effort to identify key genes responsible for lignin deposition in the hawthorn endocarp, lignin biosynthetic pathway genes were identified from the 1,218 differentially expressed genes. Most lignin pathway genes were down-regulated in the fruits of soft-endocarp hawthorn (S7) compared to the fruits of control hard-endocarp hawthorn (H8), including the genes encoding C4H (cinnamate 4-hydroxylase) [EC 1.14.13.11], HCT (hydroxycinnamoyl-CoA shikimate/quininate hydroxycinnamoyl transferase) [EC 2.3.1.133], C3H (p-coumarate 3-hydroxylase) [EC 1.14.13.-], CCR (cinnamoyl-CoA reductase) [EC 1.2.1.44], CCoA-OMT [EC 2.1.1.104], F5H (ferulate-5-hydroxylase) [EC 1.14.-.-] and CAD [EC 1.1.1.195] (Figure 4; Table 4). Both PAL (phenylalanine ammonia-lyase) [EC 4.3.1.24] and 4CL (4-coumarate:CoA ligase) [EC 6.2.1.12] were annotated by two inversely expressed unigenes, i.e., 7\_Unigene\_BMK.17577 [ $\log_2(S7/H8) = 2.71$ ] and 8\_Unigene\_BMK.5888 [ $\log_2(S7/H8) = -2.59$ ] were annotated as PAL gene, while 7\_Unigene\_BMK.4691 [ $\log_2(S7/H8) = 3.25$ ] and 8\_Unigene\_BMK.37554 [ $\log_2(S7/H8) = -3.29$ ] were annotated as 4CL gene. Four unigenes were annotated as the gene encoding peroxidase [1.11.1.7]. Among them, two were down-regulated in the fruits of soft-endocarp hawthorn (S7) [8\_Unigene\_BMK.15024,  $\log_2(S7/H8) = -14.97$ ; 7\_Unigene\_BMK.36531,  $\log_2(S7/H8) = -3.63$ ] and two were up-regulated [7\_Unigene\_BMK.930,  $\log_2(S7/H8) = 5.51$ ; 8\_Unigene\_BMK.33978,  $\log_2(S7/H8) = 1.12$ ] (Table 4). The gene expression ratios for each selected lignin pathway gene were inserted into the phenylpropanoid biosynthetic pathway [KEGG map00940 ([http://www.genome.jp/kegg-bin/show\\_pathway?map00940](http://www.genome.jp/kegg-bin/show_pathway?map00940))], and the detail information can be accessed through the File S1. Laccases, in conjunction with peroxidase, are postulated to be involved in lignin polymeriza-



**Figure 3. Histogram of GO classifications of assembled *C. pinnatifida* unigenes.** Results are summarized in three main GO categories: cellular component, molecular function and biological process. doi:10.1371/journal.pone.0072910.g003

**Table 3.** Summary of annotations of assembled hawthorn (*C. pinnatifida*) unigenes.

Category	Number of unigenes ≥300 bp	Number of Unigenes ≥1,000 bp	Number of total annotated unigenes	Percentage (%) <sup>c</sup>
COG_Annotation	3,141	4,746	9,323	17.7
GO_Annotation	4,819	7,217	13,922	26.4
Kegg_Annotation	14,061	10,597	32,466	61.6
Swiss-Prot_Annotation	9,103	9,251	22,891	43.5
TrEMBL_Annotation	13,185	10,451	30,465	57.8
InterProScan_Annotation	6,947	9,005	18,719	35.5
Nr <sup>a</sup> _Annotation	14,270	10,602	32,757	62.2
Nt <sup>b</sup> _Annotation	11,880	10,208	28,413	53.9
All_Annotation	15,569	10,660	35,979	68.3

<sup>a</sup>Nr = NCBI non-redundant sequence database.

<sup>b</sup>Nt = NCBI nucleotide sequence database.

<sup>c</sup>Proportion of the 52,673 assembled unigenes.

doi:10.1371/journal.pone.0072910.t003

tion and cell wall lignification by catalyzing the one-electron oxidation of monolignols [25]. Eleven unigenes were annotated as laccase [EC 1.10.3.2] gene, and ten were down-regulated in the fruits of soft-endocarp hawthorn (S7) with the  $\text{Log}_2(\text{S7}/\text{H8})$  fold value from  $-17.54$  to  $-3.42$  except one that was up-regulated [7\_Unigene\_BMK.36241,  $\text{log}_2(\text{S7}/\text{H8}) = 3.94$ ] (Table 4).

While most lignin pathway genes were down-regulated in the fruits of soft-endocarp hawthorn, almost all the flavonoid biosynthetic pathway genes were strongly up-regulated in the fruits of soft-endocarp hawthorn (S7) compared to the fruits of control hard-endocarp hawthorn (H8), including the genes encoding CHS [EC 2.3.1.74], CHI [EC 5.5.1.6], F3'H (flavonoid 3'-hydroxylase) [EC 1.14.13.21], F3H (flavanone 3-hydroxylase) [EC 1.14.11.9], DFR (dihydroflavonol 4-reductase) [EC 1.1.1.234], LAR (leucoanthocyanidin reductase) [EC 1.17.1.3], LDOX (leucoanthocyanidin dioxygenase) [EC 1.14.11.19] and ANR (anthocyanidin reductase) [EC 1.3.1.77] (Figure 4; Table 4). The gene expression ratios for each selected flavonoid pathway gene were shown in the flavonoid biosynthetic pathway [KEGG map00941 ([http://www.genome.jp/kegg-bin/show\\_pathway?map00941](http://www.genome.jp/kegg-bin/show_pathway?map00941))], and the detail information can be accessed through File S2.

Endocarp is mainly composed of lignin and polysaccharides such as cellulose and hemicelluloses. The expressions of some sugar metabolism pathway genes are significantly changed between soft-endocarp hawthorn (S7) and hard-endocarp hawthorn (H8). UDP-glucose is the substrate of both cellulose and pectin synthesis, the expressions of genes encoding enzymes to catalyze UDP-glucose synthesis and metabolism, were down-regulated in the fruits of soft-endocarp hawthorn (S7), including phosphoglucomutase [EC 5.4.2.2], UGPase (UDP-glucose pyrophosphorylase, also named UTP-glucose-1-phosphate uridylyl-transferase) [EC 2.7.7.9], UGDH (UDP glucose 6-dehydrogenase) [EC 1.1.1.22] and sucrose synthase [EC 2.4.1.13]. The expressions of alpha-1,4-galacturonosyltransferase [EC 2.4.1.43], beta-fructofuranosidase [EC 3.2.1.26] and fructokinase [EC 2.7.1.4] genes were also down-regulated in the fruits of S7, while the expressions of beta-glucosidase [EC 3.2.1.21], pectinesterase [EC 3.1.1.11] and UDP-glucuronate 4-epimerase [EC 5.1.3.6] genes were up-regulated in the fruits of S7. The gene expression ratios for each selected sugar metabolism pathway gene were shown into the starch and sucrose metabolism pathway [KEGG map00500

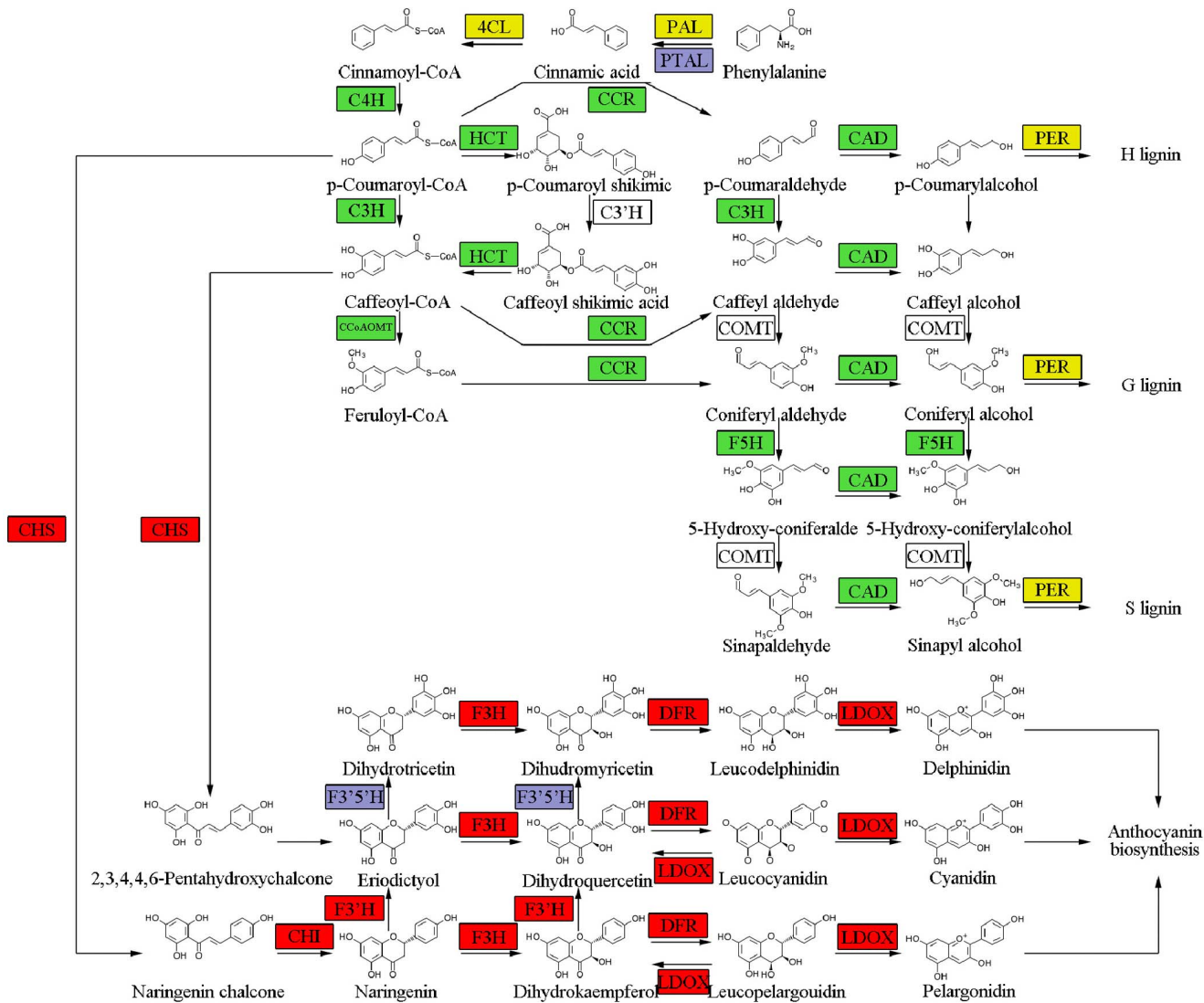
([http://www.genome.jp/kegg-bin/show\\_pathway?map00500](http://www.genome.jp/kegg-bin/show_pathway?map00500))], and the detail information can be accessed through the File S3.

To identify regulatory factors that potentially control lignin and flavonoid biosynthesis, candidate transcription factors were chosen from the transcriptome data. We initially targeted MYB proteins which are key factors in the regulation of primary and secondary metabolism [26]. Fifteen candidate MYB genes were identified from the 1,218 differentially expressed genes, among them, 12 MYB genes were down-regulated in the fruits of soft-endocarp hawthorn (S7), while 3 MYB genes were up-regulated (Table S5). The 8\_Unigene\_BMK.10344 and 8\_Unigene\_BMK.9503 were annotated as *MYB46* genes, which appears to have a more direct role in the control of secondary wall formation [15]. The 8\_Unigene\_BMK.10344 was strongly down-regulated in the fruits of soft-endocarp hawthorn (S7) [ $\text{log}_2(\text{S7}/\text{H8}) = -15.36$ ]. Four MYB genes were annotated as *R2R3-MYB* genes, and among them, 7\_Unigene\_BMK.1607 is clustered with apple *MYB10*, which is known to control anthocyanin biosynthesis [27]. The 7\_Unigene\_BMK.1607 was strongly up-regulated in the fruits of S7 [ $\text{log}_2(\text{S7}/\text{H8}) = 14.05$ ]. Secondly we targeted NAC transcription factors, which are known to control secondary wall formation in woody tissues [28]. Four candidate NAC genes were identified from the 1,218 differentially expressed genes, and all of them were strongly down-regulated in the fruits of soft-endocarp hawthorn (S7) compared to the fruits of control hard-endocarp hawthorn (H8), including 8\_Unigene\_BMK.34172 [ $\text{log}_2(\text{S7}/\text{H8}) = -14.67$ ], 8\_Unigene\_BMK.37276 [ $\text{log}_2(\text{S7}/\text{H8}) = -6.31$ ], 7\_Unigene\_BMK.19150 [ $\text{log}_2(\text{S7}/\text{H8}) = -3.98$ ] and 8\_Unigene\_BMK.30887 [ $\text{log}_2(\text{S7}/\text{H8}) = -3.00$ ].

To confirm the results of Illumina RNA-Seq analysis, the expression levels of 6 lignin pathway genes (*C3H*, *C4H*, *CAD*, *CCoA-OMT*, *CCR* and *F5H*) and Myb46 and NAC genes were measured in H8 and S7 at 19, 23, 27, 31 and 35 DAB by qRT-PCR. The qRT-PCR data for all the tested genes (Figure 5) were consistent with those obtained by Illumina RNA-Seq analysis.

## Discussion

With the application of massively parallel sequencing technologies, transcriptome information is becoming generally abundant not only for model organisms on which international research efforts and funding are concentrated, but also for non-model organisms. In this study, we generated 2.14 Gb and 2.74 Gb of



**Figure 4. Lignin and flavonoid biosynthetic pathway genes differentially expressed in fruits between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).** Green colour means the gene was significantly down-regulated in S7, red colour means the gene was significantly up-regulated in S7, yellow colour means that the expression models of annotated unigenes were inverse, and blue colour means expression difference was not significant. doi:10.1371/journal.pone.0072910.g004

raw sequence data by Illumina sequencing of hard-endocarp hawthorn and soft-endocarp hawthorn fruits, corresponding to 39,663 and 41,723 putative unigenes, respectively. For Illumina RNA-Seq, the mean size of assembled of unigenes was several hundred base pairs [18,21,29,30]. The mean size of assembled of hawthorn's putative unigenes was 674 bp, and among the total 52,673 putative unigenes, 10,744 putative unigenes were  $\geq 1,000$  bp in length (Table S2). For unigenes with abundant transcripts and large expression differences between H8 and S7, 82.5% have complete CDS. These suggest that we have obtained high quality EST data for hawthorn, facilitating future gene cloning from *Crataegus* genus for which no EST data are available in NCBI now.

Genome-wide transcriptome analysis between the soft-endocarp hawthorn and the hard-endocarp hawthorn using RNA-Seq technology revealed significant down-regulation of a number of lignin pathway genes (Table 4) concurrent with the reduction of lignin deposition and soft-endocarp forming. Down-regulation of

C4H, HCT, C3H, CCR or CCoA-OMT was previously reported to cause reduction of total lignin in transgenic alfalfa, *Arabidopsis*, and poplar [31–35]. The  $\text{Log}_2(\text{S7/S8})$  fold value of unigene annotated as F5H was lowest ( $-6.25$ ) (Table 4). F5H, a cytochrome P450-dependent monooxygenase [36,37], is a key enzyme that catalyzes the hydroxylation of ferulic acid, coniferaldehyde, and coniferyl alcohol, leading to sinapic acid and syringyl lignin (S lignin) biosynthesis [38]. F5H affects partitioning between two major traditional monolignols, coniferyl and sinapyl alcohols [39]. The *Arabidopsis* mutant deficient in F5H contained only traces of S lignin [40], while the F5H overexpressed transgenic plants displayed substantially more S lignin and was consequentially severely depleted in coniferyl alcohol-derived G lignin [39,40]. The down-regulation or overexpression of F5H by transgenic approach had little effect on the total lignin [34,41]. However, there was a substantial change in the distribution between acid-soluble and insoluble lignin fractions, with the transgenic poplar trees displaying elevated levels of acid-soluble lignin and reduced

**Table 4.** Expression profiles of lignin and flavonoid biosynthesis genes in *C. pinnatifida*.

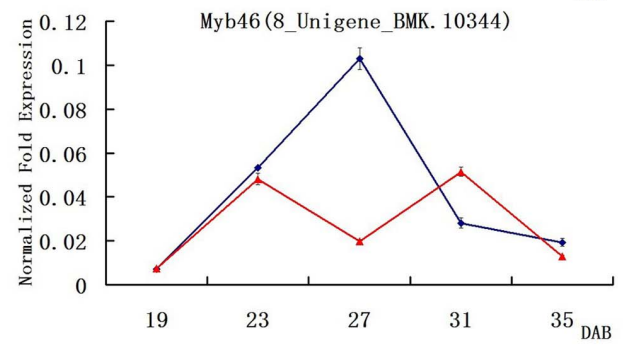
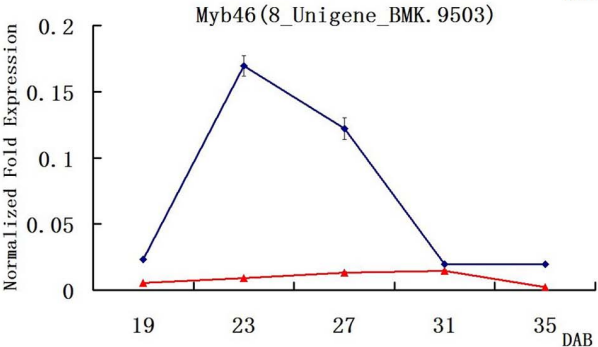
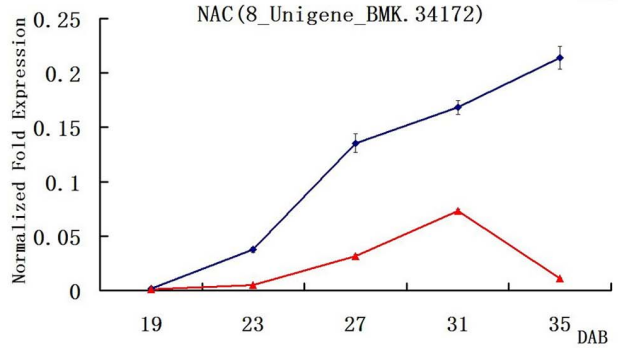
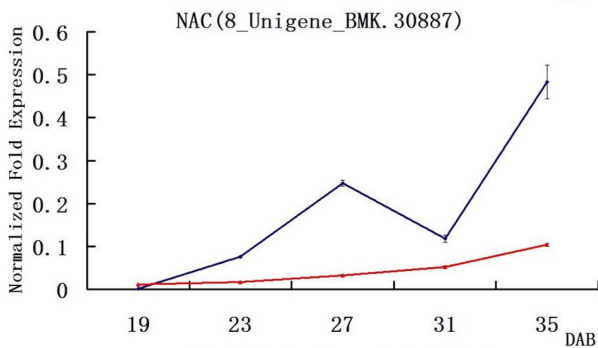
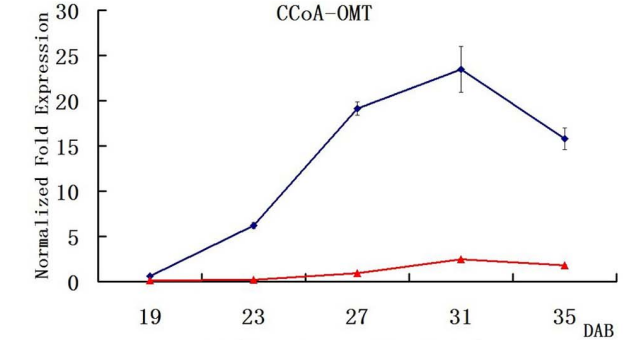
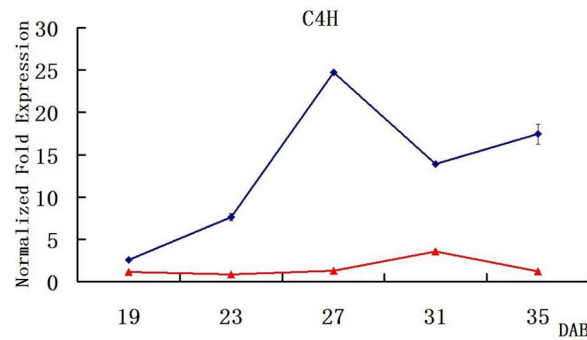
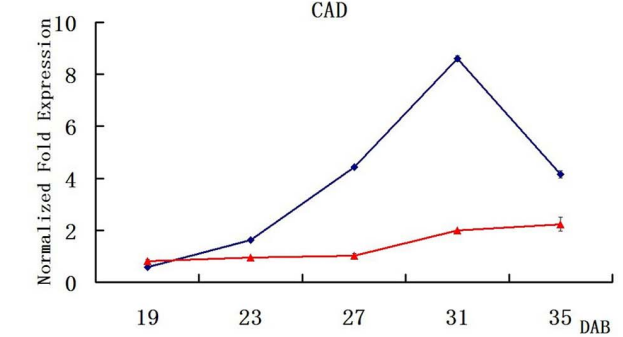
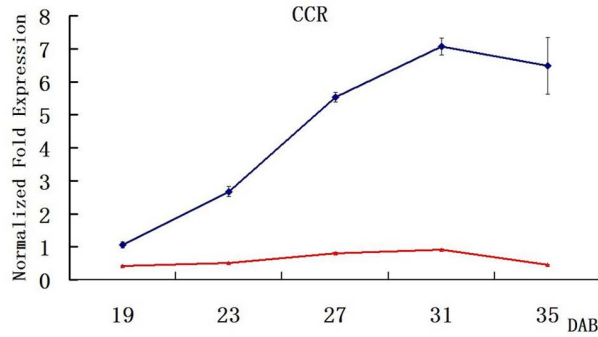
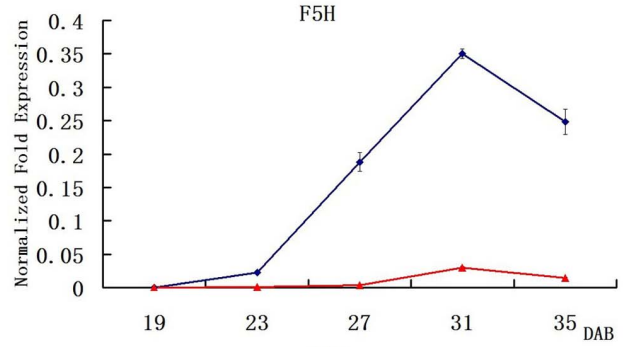
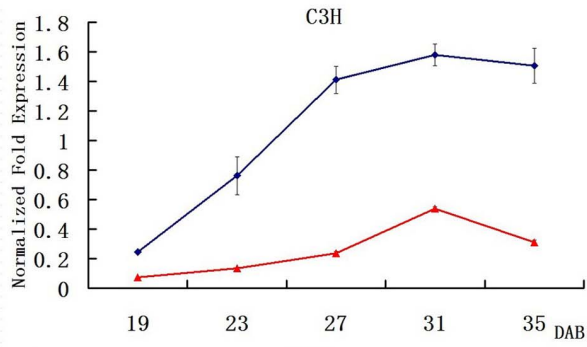
Unigene ID	Annotation	Function	log <sub>2</sub> (S7/H8)
7_Unigene_BMK.17577	PAL	Lignin and flavonoid biosynthesis	2.71
8_Unigene_BMK.5888	PAL	Lignin and flavonoid biosynthesis	-2.59
7_Unigene_BMK.4691	4CL	Lignin and flavonoid biosynthesis	3.25
8_Unigene_BMK.37554	4CL	Lignin and flavonoid biosynthesis	-3.29
8_Unigene_BMK.39561	C4H	Lignin and flavonoid biosynthesis	-1.38
7_Unigene_BMK.21369	C3H	Lignin and flavonoid biosynthesis	-3.66
8_Unigene_BMK.5677	C3H	Lignin and flavonoid biosynthesis	-5.06
8_Unigene_BMK.7232	HCT	Lignin biosynthesis	-4.58
8_Unigene_BMK.8160	CCR	Lignin biosynthesis	-3.06
7_Unigene_BMK.18764	CCoA-OMT	Lignin biosynthesis	-4.52
8_Unigene_BMK.16089	F5H	Lignin biosynthesis	-6.25
8_Unigene_BMK.34885	CAD	Lignin biosynthesis	-2.67
8_Unigene_BMK.15024	Peroxidase	Lignin biosynthesis	-14.97
7_Unigene_BMK.36531	Peroxidase	Lignin biosynthesis	-3.63
7_Unigene_BMK.930	Peroxidase	Lignin biosynthesis	5.51
8_Unigene_BMK.33978	Peroxidase	Lignin biosynthesis	1.12
8_Unigene_BMK.36210	Laccase	Lignin polymerization	-17.54
8_Unigene_BMK.8811	Laccase	Lignin polymerization	-3.42
7_Unigene_BMK.36241	Laccase	Lignin polymerization	3.94
8_Unigene_BMK.11701	CHS	Anthocyanin biosynthesis	3.82
7_Unigene_BMK.26121	CHI	Anthocyanin biosynthesis	3.34
7_Unigene_BMK.4912	CHI	Anthocyanin biosynthesis	3.29
8_Unigene_BMK.11985	F3'H	Anthocyanin biosynthesis	1.78
7_Unigene_BMK.39020	F3H	Anthocyanin biosynthesis	2.92
7_Unigene_BMK.26525	DFR	Anthocyanin biosynthesis	3.66
7_Unigene_BMK.37947	LAR	Anthocyanin biosynthesis	4.70
7_Unigene_BMK.41411	LDOX	Anthocyanin biosynthesis	3.45
7_Unigene_BMK.726	ANR	Anthocyanin biosynthesis	3.20

doi:10.1371/journal.pone.0072910.t004

levels of acid-insoluble lignin in comparison to wild-type [39]. The function of hawthorn F5H gene will be elucidated by future overexpression and silencing in hawthorn.

In vascular plants, the phenylpropanoid pathway is responsible for the biosynthesis of a variety of metabolites, including lignin and flavonoids [42]. Lignin and flavonoid biosynthesis are competitive since they presumably draw on the same precursors and diverge at the common intermediate *p*-coumaroyl CoA. For peach, during times of peak lignin deposition, genes in the lignin biosynthesis pathway were strongly induced while flavonoid pathway genes were repressed [12]. Conversely, high flavonoid gene expression was correlated with lower expression of lignin biosynthesis genes. Our gene expression data confirm the competing effect of lignin and flavonoid biosynthetic pathways. Hawthorn fruits are a rich source of flavonoids [43]. Flavonoid consumption has been documented to be negatively associated with cardiovascular conditions, so hawthorn has been used as an alternative therapy for a variety of heart disease mortality [1]. Whether the soft-endocarp hawthorn accumulates more flavonoids than common hard-endocarp hawthorns is unknown and should be tested in the future. Our work provides candidate genes for improving nutritional value of hawthorn.

The phenylpropanoid pathway genes are developmentally regulated by various classes of *cis*-acting elements [44] and *trans*-acting transcription factors [15,45]. Putative AC elements are *cis*-acting elements found in the majority promoters of phenylpropanoid biosynthetic genes, including *PAL*, *4CL*, *C3H*, *CCoAOMT*, *CCR* and *CAD* [46], and it has been proposed that coordinated expression of these genes is regulated by MYBs which bind the AC elements. Thus, regulation by lignin activators is global rather than specific for certain pathway genes [15]. In this study, 15 candidate MYB genes were identified from the differentially expressed genes, and 12 of them were down-regulated in the fruits of soft-endocarp hawthorn. It indicates that the lignin deposition in hawthorn endocarp is regulated by MYB transcription factor. MYBs are not the only transcription factors that can regulate lignin pathways. Some MYBs known to be involved in controlling lignin deposition and secondary wall formation are activated by NAC transcription factors in *Arabidopsis* [47]. For example, MYB46 and MYB83 are the direct targets of SND1, a NAC domain transcription factor [48]. In this study, four candidate NAC genes were identified from the differentially expressed genes, and all of them were strongly down-regulated in the fruits of soft-endocarp hawthorn. So, NAC domain transcription factors are





**Figure 5. The expression levels of 6 lignin pathway genes (C3H, C4H, CAD, CCoA-OMT, CCR and F5H) and Myb46 and NAC genes using qRT-PCR.** Blue lines mean the hard-endocarp hawthorn (H8), red lines mean the soft-endocarp hawthorn (S7). Y-axis represents normalized fold expression, X-axis is DAB (Days after bloom). doi:10.1371/journal.pone.0072910.g005

important candidate regulatory genes for lignin biosynthesis in endocarp of hawthorn.

This study reports the first application of RNA-Seq technology for transcriptome studies in hawthorn, a non-model species with few genomic resources. Our results demonstrate that RNA-Seq can be successfully used to obtain high quality EST data of hawthorn and to identify differentially expressed genes between the soft-endocarp hawthorn and the hard-endocarp hawthorn. This work revealed significant down-regulation of a number of lignin biosynthetic genes and up-regulation of almost all flavonoid biosynthetic genes, concomitant with the reduction of lignin deposition and soft-endocarp forming. In addition, we have identified MYB and NAC transcription factors that potentially control lignin and flavonoid biosynthesis. The function of these differentially expressed genes, in particular the MYB and NAC factors, between the soft-endocarp and the hard-endocarp hawthorn will be elucidated in the future.

## Materials and Methods

### Plant Material

The trees of *Crataegus pinnatifida* accessions H8 (hard-endocarp hawthorn) and S7 (soft-endocarp hawthorn) were maintained in the National Hawthorn Germplasm Repository at Shenyang. Bloom time was noted when 50% of flowers had opened. The fruits for analysis of lignin deposition were collected once every two days from 19 to 43 days after bloom (DAB). At each collection time, 16 fruits were collected from each accession. Ten of these were frozen in liquid N<sub>2</sub> for future RNA extraction. The remaining six were transversely cut into sections and immediately stained with phloroglucinol solution [2% phloroglucinol, 85% ethanol (v/v)] for 2 min, drained and exposed to 85% HCl for 2 min. The fruit sections were then rinsed in 95% ethanol (v/v) and photographed. The ripen fruits were collected at 118 DAB (Figure 1).

### RNA Extraction and Quality Determination

Total RNA were isolated using the modified CTAB method performed as Chang et al. [49], and the RNA samples were treated with DNase I (TaKaRa, Japan) for 4 h. The integrity of the RNA samples was examined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, USA).

### cDNA Library Preparation and Illumina Sequencing

cDNA library preparation and sequencing reactions were conducted in the Biomarker Technology Company, Beijing, China. The paired-end library preparation and sequencing were performed following standard Illumina methods using the DNA sample kit (#FC-102-1002, Illumina). The cDNA library was sequenced on the on the Illumina sequencing platform (HiSeq™ 2000).

### De Novo Assembly

Reads from each library were assembled separately. The trimming adapter sequences were removed and low-quality reads (less than 13 bp or reads with unknown nucleotides larger than 5%) were filtered with the software developed by the Biomarker Technology Company. The Trinity method [22] was used for *de novo* assembly of Illumina reads of hawthorn. Trinity consists of

three software modules: Inchworm, Chrysalis and Butterfly, applied sequentially to process large volumes of RNA-Seq reads. In the first step in Trinity, reads are assembled into the contigs by Inchworm program. The minimally overlapping contigs were clustered into sets of connected components by Chrysalis program, and then the transcripts were constructed by the Butterfly program [22]. In this study, only one k-mer length (25-mer) was chosen in Trinity, using the follow parameters: seqType fq, group\_pairs\_distance=150 and other default parameters. Finally, the transcripts were clustered by similarity of correct match length beyond the 80% of longer transcript or 90% of shorter transcript used multiple sequence alignment tool BLAT [50]. The longest transcript of each cluster was taken as the unigene. The Illumina data set has been deposited in NCBI Sequence Read Archive (SRA) under accession number SRX305204.

### SSRs Detection

SSRs were detected among the unigenes with length >1,000 bp using the software MISA (MicroSATellite; <http://pgrc.ipk-gatersleben.de/misa>) [51]. Total 7 types of SSRs were investigated, including mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats, and the compound SSR (the sequence contains two adjacent distinct SSRs separated by none to any number of base pairs).

### CDS Analysis

The coding sequence (CDS) in unigene was predicted by EMBOSS getorf program. The complete CDS sequences of hawthorn unigenes, with abundant transcripts (RPKM>100) and large expression differences between H8 and S7 [ $\log_2(S7/H8)>5$  or  $<-5$ ], were chosen to compare with the predicted CDS sequences of apple (*Malus×domestica*) (<http://genomics.research.iasma.it/>), and the identity of CDS between the hawthorn and apple gene was analyzed with DNAMAN software (version 5.2.2).

### Functional Annotation

We annotated unigenes based on a set of sequential BLAST searches [52] designed to find the most descriptive annotation for each sequence. The assembled unigenes were compared with sequences in NCBI non-redundant (Nr) protein and nucleotide (Nt) databases (<http://www.ncbi.nlm.nih.gov>), the Swiss-Prot protein database (<http://www.expasy.ch/sprot>), the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (<http://www.genome.jp/kegg>), the Cluster of Orthologous Groups (COG) database (<http://www.ncbi.nlm.nih.gov/COG>), the Translated EMBL Nucleotide Sequence database (TrEMBL) (<http://www.uniprot.org/>) and InterPro database (<http://www.ebi.ac.uk/interpro/>). The Blast2GO program [53] was used to obtain GO annotation of the unigenes. The WEGO software (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>) was then used to perform GO functional classification of all unigenes to view the distribution of gene functions.

### Digital Gene Expression Analysis

Gene expression levels were measured in the RNA-Seq analysis as reads per kilobase of exon model per million mapped reads (RPKM) [54]. The IDEG6 software [24] was used to identify differentially expressed genes in pair-wise comparison, and the

results of all statistical tests were corrected for multiple testing with the Benjamini–Hochberg false discovery rate (FDR<0.01). Sequences were deemed to be significantly differentially expressed if the adjusted *P* value obtained by this method was <0.001 and there was at least a twofold change (>1 or <-1 in log<sub>2</sub> ratio value) in RPKM between two libraries.

### Quantitative RT-PCR (qRT-PCR) Analysis

The cDNA was synthesized from total RNA using Reverse Transcriptase XL (AMV) (TaKaRa, Japan) in a 20 µL reaction system. The reverse transcription reaction mixture contained 5 µL total RNA (1 µg), 1 µL 10 mM of each dNTP, 1 µL of random primer (9 mer) (50 µM), 1 µL oligo d(T)<sub>18</sub> primer (50 µM) (TaKaRa, Japan) and 6 µL DEPC water. The mixture were incubated at 65°C for 5 min and cooled on ice for 5 min, then 4 µL 5×Reverse Transcriptase buffer, 1 µL RNasin (TaKaRa, Japan) and 1 µL AMV (5 U) were added. The mixture was incubated at 37°C for 2.5 h. The enzyme was inactivated by incubating at 72°C for 15 min. qPCR was carried out on the iQ5 Real Time PCR Detection System (BioRad, USA) with RealMasterMix SYBR Green (TIANGEN, China). Primers used in qPCR for validation of differentially expressed genes were shown in Table S6. Each gene was analyzed in triplicate, after which the average threshold cycle (CT) was calculated per sample, and an endogenous *ACTIN* gene was used for normalization. Relative fold changes in genes expression were calculated using the comparative Ct ( $2^{-\Delta\Delta C_t}$ ) method.

### Supporting Information

**Figure S1 The size distribution of the *C. pinnatifida* unigenes.**  
(EPS)

**Figure S2 Functional categories of 1,218 differentially expressed unigenes between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).**  
(EPS)

**Table S1 EST quality of *C. pinnatifida* unigenes.**  
(XLS)

### References

- Rigelsky JM, Sweet BV (2002) Hawthorn: Pharmacology and therapeutic uses. *Am J Health-Syst Pharm* 59: 417–422.
- Phipps JB, O’Kennon RJ, Lance RW (2003) Hawthorns and medlars. Royal Horticultural Society, Cambridge, UK.
- Dai H, Zhang Z, Guo X (2007) Adventitious bud regeneration from leaf and cotyledon explants of Chinese hawthorn (*Crataegus pinnatifida* Bge. var. major N.E.Br.). *In Vitro Cell. Dev –Pl* 43: 2–8.
- Kao ES, Wang CJ, Lin WL, Yin YF, Wang CP, et al. (2005) Anti-inflammatory potential of flavonoid contents from dried fruit of *Crataegus pinnatifida* in vitro and in vivo. *J Agric Food Chem* 53: 430–436.
- Courtney SP, Manzur MI (1985) Fruiting and fitness in *Crataegus monogyna*: the effects of frugivores and seed predators. *Oikos* 44: 398–406.
- Gutián P (1998) Latitudinal variation in the fruiting phenology of a bird-dispersed plant (*Crataegus monogyna*) in Western Europe. *Plant Ecol* 137: 139–142.
- Dai H, Guo X, Zhang Y, Li Y, Chang L, et al. (2009) Application of random amplified polymorphic DNA and inter-simple sequence repeat markers in the genus *Crataegus*. *Ann Appl Biol* 154: 175–181.
- Zhao H, Feng B (1996) China Fruit-plant Monograph · Hawthorn (*Crataegus*) Flora. China Forest Press, Beijing, China.
- Bujarska-Borkowska B (2002) Breaking of seed dormancy, germination and seedling emergence of the common hawthorn (*Crataegus monogyna* Jacq.). *Denbriobiology* 47: 61–70.
- Dai H (2007) Molecular identification and enhancement of germplasm in hawthorn. PhD thesis. Shenyang Agricultural University, China.
- Nas MN, Gokbunar L, Seygin N, Aydemir M, Dagli M, et al. (2012) Micropropagation of mature *Crataegus aronia* L., a medicinal and ornamental plant with rootstock potential for pome fruit. *Plant Growth Regul* 67: 57–63.
- Dardick CD, Callahan AM, Chiozzotto R, Schaffer RJ, Piagnani MC, et al. (2010) Stone formation in peach fruit exhibits spatial coordination of the lignin and flavonoid pathways and similarity to *Arabidopsis* dehiscence. *BMC Biol* 8: 13.
- Ryugo K (1963) Changes in methoxyl content in the peach endocarp and some of its soluble phenolic constituents during lignification. *Proc Amer Soc Hort Sci* 84: 110–115.
- Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin Biosynthesis and Structure. *Plant Physiol* 153: 895–905.
- Zhao Q, Dixon RA (2011) Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci* 16: 227–233.
- Alba CM, De Forchetti SM, Tugier HA (2000) Phenoloxidase of peach (*Prunus persica*) endocarp: Its relationship with peroxidases and lignification. *Physiol Plantarum* 109: 382–387.
- Morelló JR, Romero MP, Ramo T, Motilva MJ (2005) Evaluation of l-phenylalanine ammonia-lyase activity and phenolic profile in olive drupe from fruit setting period to harvesting time. *Plant Sci* 168: 65–72.
- Qiu Q, Ma T, Hu Q, Liu B, Wu Y, et al. (2011) Genome-scale transcriptome analysis of the desert poplar, *Populus euphratica*. *Tree Physiol* 31: 452–461.
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57–63.
- Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48: 249–257.
- Xia Z, Xu H, Zhai J, Li D, Luo H, et al. (2011) RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*. *Plant Mol Biol* 77: 299–308.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652.

**Table S2 Detail information of SSRs in *C. pinnatifida* unigenes.**  
(XLS)

**Table S3 The 100 most abundant unigenes in the two hawthorns sample sets.**  
(XLS)

**Table S4 All the differentially expressed unigenes between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).**  
(XLS)

**Table S5 Candidate MYB genes identified from *C. pinnatifida*.**  
(XLS)

**Table S6 Primers used to perform qPCR of lignin biosynthesis and regulate genes.**  
(XLS)

**File S1 Phenylpropanoid biosynthetic pathway (KEGG map00940) genes differentially expressed in fruits between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).**  
(RAR)

**File S2 Flavonoid biosynthetic pathway (KEGG map00941) genes differentially expressed in fruits between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).**  
(RAR)

**File S3 Starch and sucrose metabolism pathway (KEGG map00500) genes differentially expressed in fruits between the hard-endocarp hawthorn (H8) and the soft-endocarp hawthorn (S7).**  
(RAR)

### Author Contributions

Conceived and designed the experiments: HYD ZHZ. Performed the experiments: HYD GFH YJY. Analyzed the data: ZHZ HYD GFH YJY FZ XML WRL. Contributed reagents/materials/analysis tools: YM HL YXL. Wrote the paper: ZHZ ZCL.

23. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, et al. (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet* 42: 833–839.
24. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics* 12: 159–162.
25. LaFayette PR, Eriksson KEL, Dean JFD (1999) Characterization and heterologous expression of laccase cDNAs from xylem tissues of yellow-poplar (*Liriodendron tulipifera*). *Plant Mol Biol* 40: 23–35.
26. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, et al. (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 15: 573–581.
27. Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, et al. (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biol* 10: 50.
28. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, et al. (2007) NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. *Plant Cell* 19: 270–280.
29. Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, et al. (2010) *De novo* characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics* 11: 400.
30. Feng C, Chen M, Xu CJ, Bai L, Yin XR, et al. (2012) Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics* 13: 19.
31. Chen F, Srinivasa Reddy MS, Temple S, Jackson L, Shadle G, et al. (2006) Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wallbound ferulic acid in alfalfa (*Medicago sativa* L.). *Plant J* 48: 113–124.
32. Do CT, Pollet B, Thévenin J, Sibout R, Denoue D, et al. (2007) Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in *Arabidopsis*. *Planta* 226: 1117–1129.
33. Lepié JC, Dauwe R, Morreel K, Storme V, Lapiere C, et al. (2007) Downregulation of cinnamoyl-coenzyme A reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell* 19: 3669–3691.
34. Vanholme R, MorreelK RalphJ, Boerjan W (2008) Lignin engineering. *Curr Opin Plant Biol* 11: 278–285.
35. Lacombe E, Van Doorselaere J, Boerjan W, Boudet AM, Grima-Pettenati J (2000) Characterization of cis-elements required for vascular expression of the Cinnamoyl CoA Reductase gene and for protein-DNA complex formation. *Plant J* 23: 663–676.
36. Meyer K, Cusumano JC, Somerville C, Chapple CC (1996) Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases. *PNAS* 93: 6869–6874.
37. Humphreys JM, Hemm MR, Chapple C (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *PNAS* 96: 10045–10050.
38. Franke R, McMichael CM, Meyer K, Shirley AM, Cusumano JC, et al. (2000) Modified lignin in tobacco and poplar plants over-expressing the *Arabidopsis* gene encoding ferulate 5-hydroxylase. *Plant J* 22: 223–234.
39. Stewart JJ, Akiyama T, Chapple C, Ralph J, Mansfield SD (2009) The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar. *Plant Physiol* 150: 621–635.
40. Marita JM, Ralph J, Hatfield RD, Chapple C (1999) NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate 5-hydroxylase. *PNAS* 96: 12328–12332.
41. Reddy MS, Chen F, Shadle G, Jackson L, Aljoe H, et al. (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *PNAS* 102: 16573–16578.
42. Li X, Bonawitz ND, Weng JK, Chapple C (2010) The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thaliana* is independent of flavonoids. *Plant Cell* 22: 1620–1632.
43. Zhang Z, Chang Q, Zhu M, Huang Y, Ho WK, et al. (2001) Characterization of antioxidants present in hawthorn fruits. *J Nutr Biochem* 12: 144–152.
44. Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol* 133: 1051–1071.
45. Fornalé S, Shi X, Chai C, Encina A, Irar S, et al. (2010) ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *Plant J* 64: 633–644.
46. Zhong R, Ye ZH (2009) Transcriptional regulation of lignin biosynthesis. *Plant Signal Behav* 4: 1028–1034.
47. Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell* 20: 2763–2782.
48. McCarthy RL, Zhong R, Ye ZH (2009) MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell Physiol* 50: 1950–1964.
49. Chang L, Zhang Z, Yang H, Li H, Dai H (2007) Detection of Strawberry RNA and DNA Viruses by RT-PCR using total Nucleic Acid as a Template. *J Phytopathol* 155: 431–436.
50. Kent WJ (2002) BLAT\_The BLAST-like alignment tool. *Genome Res* 12: 656–664.
51. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106: 411–422.
52. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
53. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
54. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621–628.