



Correspondence

<https://doi.org/10.1631/jzus.B2100673>



Comparative transcriptome analysis of candidate genes involved in chlorogenic acid biosynthesis during fruit development in three pear varieties of Xinjiang Uygur Autonomous Region

Hao WEN¹, Xi JIANG², Wenqiang WANG¹, Minyu WU¹, Hongjin BAI², Cuiyun WU^{2✉}, Lirong SHEN^{1✉}

¹Department of Food Science and Nutrition, Zhejiang Key Laboratory for Agro-Food Processing, Zhejiang University, Hangzhou 310058, China

²College of Horticulture and Forestry, Tarim University, Alar 843300, China

Pear is one of the main fruits with thousands of years of cultivation history in China. There are more than 2000 varieties of pear cultivars around the world, including more than 1200 varieties or cultivars in China (Legrand et al., 2016). Xinjiang Uygur Autonomous Region is an important pear production region in China with 30 of varieties or cultivars. *Pyrus sinkiangensis* is the most popular variety, which is mainly distributed in Xinjiang (Zhou et al., 2018). Chlorogenic acid (CGA), *p*-coumaric acid, and arbutin are the main polyphenols in pear fruit, and their levels show great differences among different varieties (Li et al., 2014). CGA is a potential chemo-preventive agent, which possesses many important bioactivities including antioxidant, diabetes attenuating, and anti-obesity (Wang et al., 2021). Therefore, the specific CGA content of a variety is considered the embodiment of the functional nutritional value of pears.

Transcriptome analysis is an effective method to study the relationship between plant growth and gene expression (Covington et al., 2008). In 2013, Chinese scientists sequenced the first pear genome in the world, the whole genome of the *Pyrus bretschneideri* pear cultivar. A high-quality genome map with a size of 512 MB encoding 42 812 proteins was obtained (Wu et al., 2013). The transcriptomic analysis of “Suli” pear (*Pyrus pyrifolia* of white pear group) buds during

the dormancy was implemented (Liu et al., 2012). Moreover, the high-quality genome map of *Pyrus betuleafolia* (size of 532.7 MB encoding 59 552 proteins), globally the first wild pear species to be sequenced, was reported lately (Dong et al., 2020), which provided an important guarantee for transcriptome research and functional pear breeding.

Previous studies have shown that CGA is synthesized by three routes in plants, with the starting point of L-phenylalanine. The first and second routes are divided into upstream and downstream parts. The upstream part that is catalyzed by phenylalanine ammonia lyase includes phenylalanine (PAL) (MacLean et al., 2007), PAL/tyrosine ammonia lyase (PTAL) to produce *trans*-cinnamic acid, and *p*-coumaric acid through cinnamic acid-4-hydroxylase (C4H). Subsequently, *p*-coumaroyl coenzyme A (CoA) is catalyzed by 4-coumaroyl CoA ligase (4CL) (Dixon and Paiva, 1995). In the downstream part, caffeic acid and shikimic acid are involved in the CGA biosynthesis in the first pathway, while only quinic acid is present in the second pathway (Hoffmann et al., 2004). At the same time, the downstream enzymatic reaction is completed by shikimic acid/quinic acid hydroxycinnamoyl transferase (HCT) and *p*-coumaric acid 3'-hydroxylase (C3'H). Different from the top two routes, the third route has only been found in a few plant species such as sweet potato root (Villegas and Kojima, 1986). Therefore, the first and second routes are considered as the main pathways of plants for CGA biosynthesis.

In order to excavate the key genes involved in CGA biosynthesis during the development of pear fruit, the transcriptome analysis of three Xinjiang characteristic pear (XCP) varieties, Yali pear (YL; *P.*

✉ Lirong SHEN, shenlirong@zju.edu.cn

Cuiyun WU, 119900019@taru.edu.cn

✉ Lirong SHEN, <https://orcid.org/0000-0002-3197-9245>

Cuiyun WU, <https://orcid.org/0000-0002-0178-3524>

Received Aug. 1, 2021; Revision accepted Dec. 19, 2021;
Crosschecked Mar. 17, 2022

© Zhejiang University Press 2022

bretschneideri Rehd.), Korla fragrant pear (XL; *Pyrus sinkiangensis* Yü), and Yuanhuang pear (YH; *Pyrus pyrifolia* (Burm. F.) Nakai cv. Starkrimson), was conducted (Fig. S1). The pear fruit samples were collected at Luntai Plant Germplasm Resources Garden, Xinjiang Academy of Agricultural Sciences (Urumqi, China) from June to September in 2019 (Fig. 1a). CGA and arbutin contents in the pears were determined by high-performance liquid chromatography (HPLC) analysis (Fig. S2). The typical samples collected at three key development stages, the young stage (T1), the expanding stage (T2), and the mature stage (T3), were chosen for RNA-sequencing (RNA-seq) according to single fruit weight (Fig. 1a and Table S1) and the change of CGA content (Fig. 1b).

Through de novo transcriptome assembly using Trinity software ($k\text{-mer}=25$), a total of 1 356 382 096 raw reads were obtained. After the quality control of the original readings by relevant software, 1 343 719 214 clean reads including bases of 200.63 GB in total were obtained. For all 27 samples, the rate of sequencing error was 2.40‰–2.68‰ and the Phred quality score DNA sequencing of 30 (Q30) was 92.40%–95.15%, which indicated that at least 92.40% of the samples had a base sequencing error rate of less than 0.1% with a high quality of RNA-seq (Table S2). Trinity assembly was performed on all clean reads to yield a total of 119 340 transcripts. Finally, a total of 77 130 unigenes were obtained. The results showed that the N50 (50% of total scaffold length) of unigenes was 1496 bp, and the average length was 926 bp. The average alignment rate of clean reads with assembled sequences was 73.42%, and this was 74.60% for the *P. bretschneideri* reference genome (Table S2). The results met the requirements of annotation analysis.

The assembled transcriptome sequences were compared with six databases including the non-redundant protein sequence database (NR; <https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA>), an annotated protein sequence database (Swiss-Prot; <http://www.gpmaw.com/html/swiss-prot.html>), the Database of Clusters of Orthologous Genes (COG; <https://www.ncbi.nlm.nih.gov/research/cog>), a database of protein families (Pfam; <http://pfam.xfam.org>), the Gene Ontology (GO; <http://geneontology.org>), and the Kyoto Encyclopedia of Genes and Genomes (KEGG; <https://www.genome.jp/kegg/pathway.html>). A total of 47 579 unigenes accounting for 61.69% of all unigenes were annotated (Table 1).

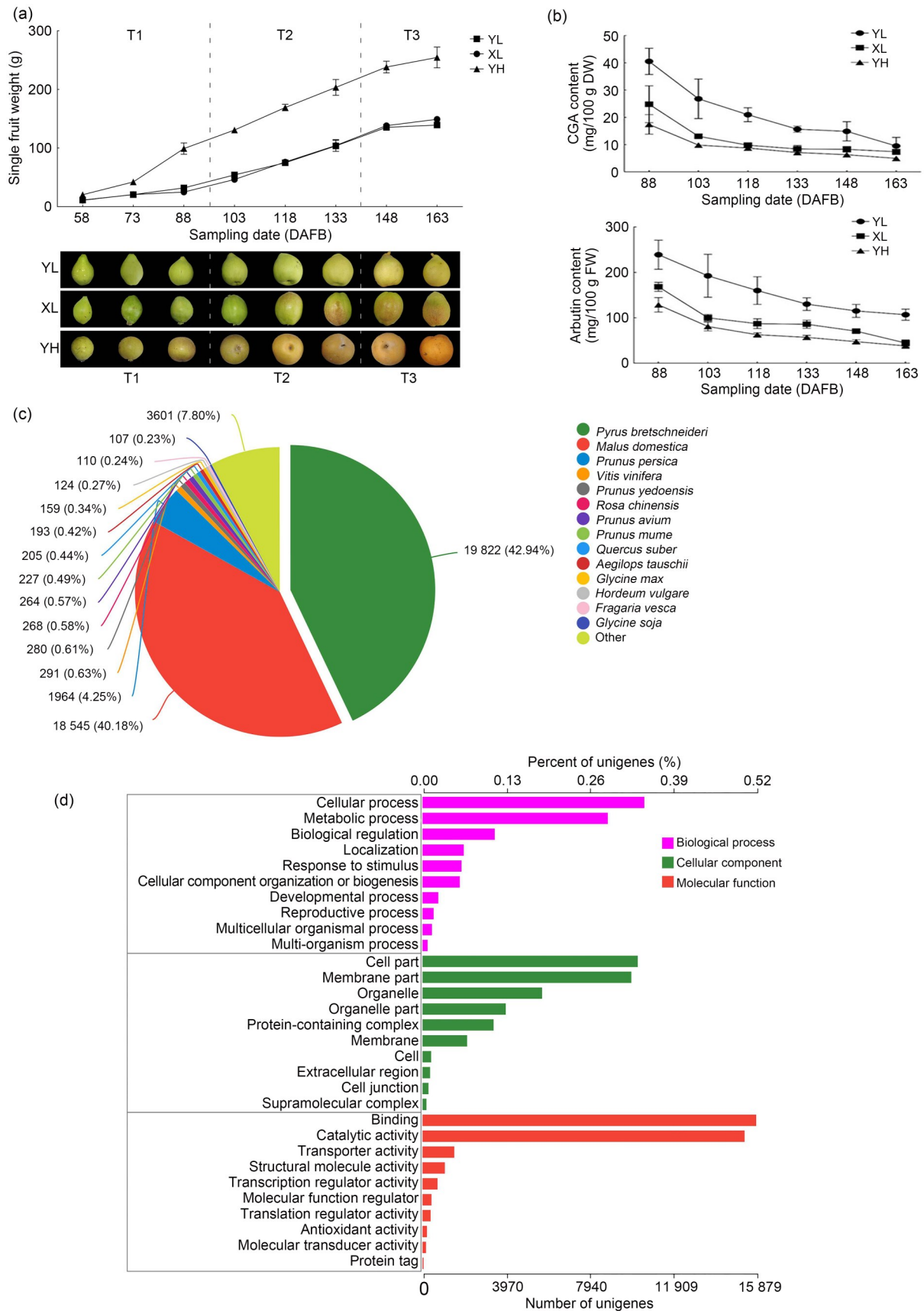
In this part, 46 160 unigenes were annotated into the NR database (Fig. 1c). Among them, 19 822 unigenes of the total NR annotation shared a homology with *P. bretschneideri*, accounting for 42.92%, followed by 18 548 unigenes of the total NR annotation that shared a homology with *Malus domestica*, accounting for 40.18%. Unigenes sharing a homology with the two varieties, accounted for 84.00% of the total NR annotations, indicating that the assembly sequence was closely related to the two varieties.

The results of GO annotation showed that 30 367 unigenes had a GO annotation, involving 52 Level 2 GO terms of three branches. In the “Molecular function” branch, the two Level 2 GO terms of “binding function” and “catalytic activity” with 15 879 and 15 329 unigenes, respectively, had the largest number of unigenes annotated. In the “Cellular component” branch with 10 242 unigenes and “Biological process” branch with 10 555 unigenes, the Level 2 GO terms with the most annotations were “cellular part” and “cell process” respectively (Fig. 1d).

The KEGG annotation resulted in a total of 19 226 annotated unigenes, of which 10 774 annotations were assigned to 146 pathways that were classified into 6 first categories and 27 second categories; 5176 unigenes (48.04%) were annotated into the “Metabolism” category, including 11 second categories. In the first category, 1648 and 1061 unigenes participated in the second categories of “Carbohydrate metabolism” and “Amino acid metabolism,” respectively (Fig. 1e).

In this study, $P<0.05$ and fold change (FC) >1.5 were used as the selection condition to screen differentially expressed genes (DEGs) between different varieties in the same developmental stage. There were 8466, 13 978, 10 202, 13 723, 11 156, and 15 638 DEGs selected from the YL_T1 vs. XL_T1, YH_T1 vs. XL_T1, YL_T2 vs. XL_T2, YH_T2 vs. XL_T2, YL_T3 vs. XL_T3, and YH_T3 vs. XL_T3 groups, respectively (Fig. 1f). To further analyze the function of DEGs in metabolic pathways, path enrichment analysis was carried out. A total of 140 and 143 pathways were found that were enriched in the YL_T1 vs. XL_T1 and YH_T1 vs. XL_T1 groups. The “Phenylpropanoid biosynthesis” pathway was highly enriched in the two groups and contained more DEGs, which was the key pathway for CGA biosynthesis (Fig. S3).

A total of 27 DEGs including 1 *PAL*, 5 *C4H*, 4 *4CL*, 2 *C3'H*, and 15 *HCT* genes involved in CGA



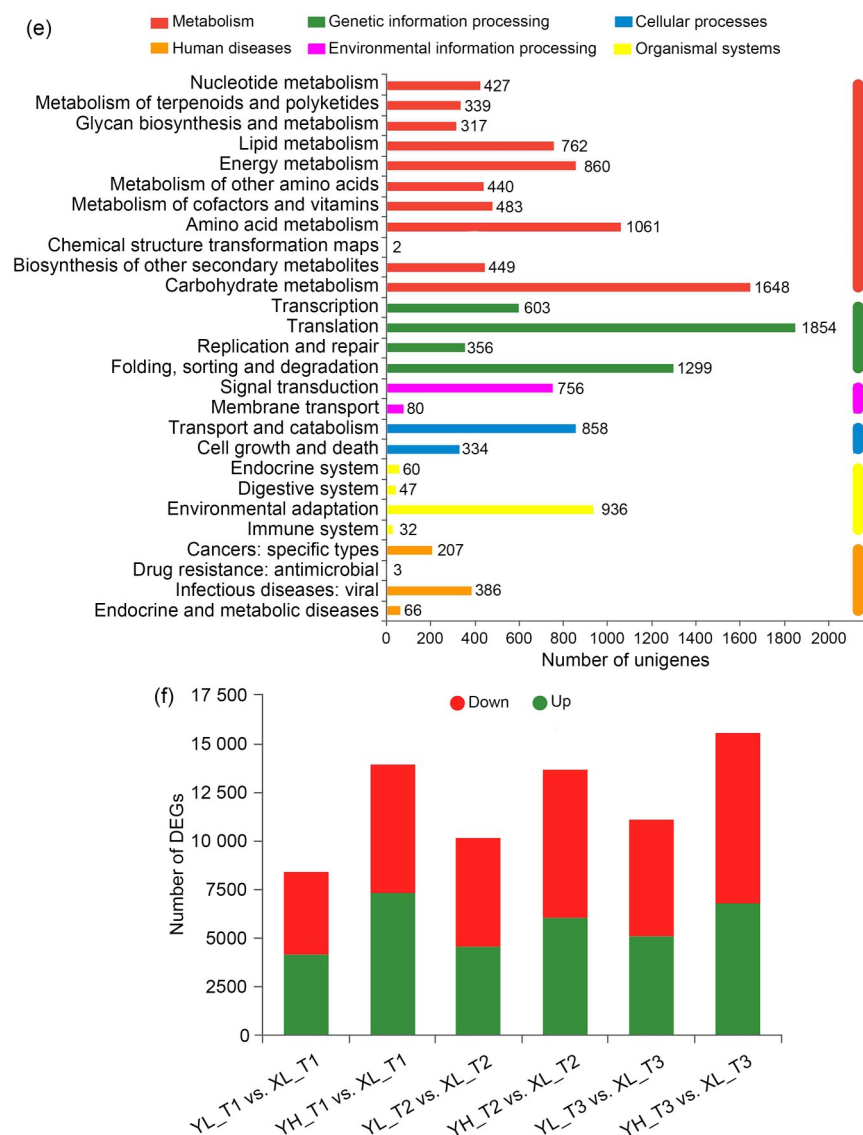


Fig. 1 Summary of comparative transcriptome analysis of candidate genes involved in CGA biosynthesis in fruits during development of three pear varieties. (a) The single fruit weight change during development periods of three pear varieties (data are expressed as mean \pm SD, $n=10$); (b) The content changes of CGA and arbutin in fruits during development of three pear varieties (data are expressed as mean \pm SD, $n=3$); (c) Species distribution of the NR annotation; (d) Summary of the GO annotation; (e) Pathway classification based on the KEGG annotation; (f) Statistical chart of DEGs in six different groups. CGA: chlorogenic acid; SD: standard deviation; NR: non-redundant protein sequence database; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEG: differentially expressed gene; YL: Yali pear; XL: Korla fragrant pear; YH: Yuanhuang pear; DAFB: days after full bloom; DW: dry weight; FW: fresh weight; T1: young fruit stage; T2: expanding stage; T3: mature stage.

biosynthesis were screened out from the two groups of YL_T1 vs. XL_T1 and YH_T1 vs. XL_T1 in T1. Among them, the expression patterns of 19 DEGs including 5 *C4H*, 4 *4CL*, 2 *C3'H*, and 8 *HCT* genes were consistent with the differences in the content of CGA between the pear varieties. In addition, a total of 31 DEGs including 2 *PAL*, 4 *C4H*, 4 *4CL*, 3 *C3'H*, and 18 *HCT* genes involved in the CGA biosynthesis

were screened in the two groups of YL_T3 vs. XL_T3 and YH_T3 vs. XL_T3 in T3. There were 20 DGEs including 1 *C4H*, 3 *4CL*, 3 *C3'H*, and 13 *HCT* genes with the same expression pattern. After combining 19 DEGs of T1 with 20 DEGs of T3, 10 DEGs including 2 *4CL*, 1 *C4H*, 2 *C3'H*, and 5 *HCT* genes were found in both development stages. Therefore, these DEGs are considered as the key regulatory genes

Table 1 Functional annotation of unigenes

Database	Unigene number (%)
GO	30 338 (39.42)
KEGG	19 212 (24.96)
COG	27 420 (35.63)
NR	46 118 (59.92)
Swiss-Prot	29 251 (38.00)
Pfam	26 382 (34.28)

GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; COG: Cluster of Orthologous Groups of proteins; NR: non-redundant protein sequence; Swiss-Prot: annotated protein sequence; Pfam: protein families.

of CGA biosynthesis, and C3'H and HCT in the downstream pathway of CGA biosynthesis are the key rate-limiting enzymes (Fig. 2).

Both the C3'H and C4H enzymes are members of the plant cytochrome P450 enzyme (CYP450) superfamily, with the C3'H enzyme belonging to the CYP98 subfamily (Sullivan and Zarnowski, 2010), and the C4H enzyme belonging to the CYP73 subfamily (Liu et al., 2018). In the process of plant growth and development, CYP450 as an oxidase participates in redox reactions including the biosynthesis and degradation of terpenes, plant hormones, signal molecules, flavonoids, and other secondary metabolites of various substrates (Han et al., 2012; McLean et al., 2012). HCT belongs to the BAHD acyltransferase family,

and uses acyl-CoA as a donor to catalyze the formation of esters or amide compounds from a variety of substrates (Tuominen et al., 2011).

In order to further explore the candidate genes related to CGA biosynthesis, the weighted gene co-expression network analysis (WGCNA) was used to construct the co-expression gene network module (Fig. 4S). The results showed that the co-expression network constructed by 14 213 preprocessed DEGs had ten modules in total. Among the ten DEGs related to CGA biosynthesis, five DEGs were located in the turquoise module and two DEGs in the pink module. The related DEGs in these modules might play an important role in CGA biosynthesis. WGCNA analysis provides a lot of gene information for the future analysis of the co-expression of related genes in CGA biosynthesis. Six unigenes encoding key enzymes were randomly selected for quantitative real-time polymerase chain reaction (qRT-PCR) to verify the accuracy of RNA-seq (the designed primers were shown in Table S3). The expression patterns were consistent with the data of RNA-seq (Fig. 5S).

In summary, this study determined the contents of CGA and arbutin in fruits during development of three pear varieties. Ten typical DEGs including two *4CL*, one *C4H*, two *C3'H*, and five *HCT* genes were identified, which are regarded as the key regulatory

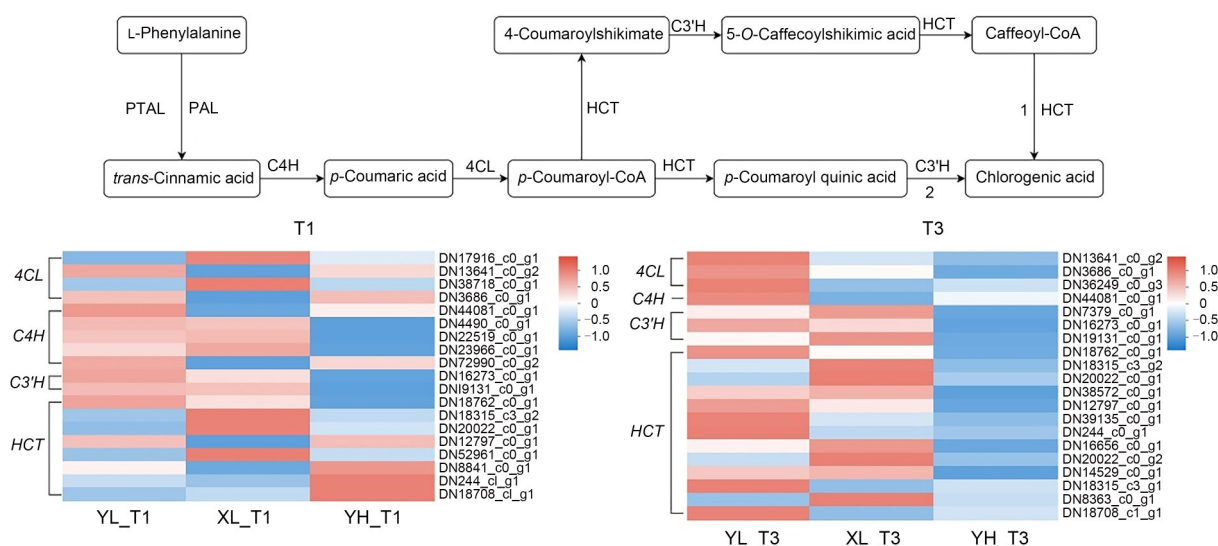


Fig. 2 Schematic representation of CGA biosynthesis pathway and heatmap of DEGs in two developmental stages. PAL: phenylalanine; PTAL: PAL/tyrosine ammonia lyase; CoA: coenzyme A; C4H: cinnamic acid-4-hydroxylase; 4CL: 4-coumaroyl CoA ligase; HCT: hydroxycinnamoyl transferase; C3'H: *p*-coumaric acid 3'-hydroxylase; YL: Yali pear; XL: Korla fragrant pear; YH: Yuanhuang pear; T1: young fruit stage; T3: mature stage; 1 and 2: key limit enzymes 1 and 2.

genes of CGA biosynthesis. The WGCNA analysis of DEGs at different development stages showed that C3'H and HCT in the downstream pathway of CGA biosynthesis are the key rate-limiting enzymes. These results revealed the potential mechanism of CGA biosynthesis.

Advocating a well-balanced diet or promoting healthy eating, especially increasing fruit intake, is becoming the most important approach to prevent or reduce diet-related non-communicable diseases, such as obesity and diabetes in China (Li, 2019). Exploring and understanding the activity and benefits of nutrients and bioactive compounds in food including fruits and their applications is one aspect of nutritional science (Durazzo and Lucarini, 2019). Pear is one of the more commonly consumed fruits in the world. The varieties or cultivars of pear containing rich polyphenols, especially CGA and arbutin with important bioactivity, have been paid more attention. This new tendency will greatly impact the pear breeding research in agriculture. Compared with previous reports, our study systematically measured the contents of CGA and arbutin, as well as their level changes during fruit development in different varieties, which provides important data for nutritional evaluation and guidance in the future breeding of pear fruits in China.

Acknowledgments

This work was supported by the Major Scientific and Technological Projects of the Xinjiang Production and Construction Corps (Nos. 2017DB006 and 2020KWZ-012). We thank Prof. Liang LIU from Department of Statistics, the University of Georgia, the United States, for his helps on designing research and revising manuscript.

Author contributions

Hao WEN performed the experimental research and data analysis, wrote and edited the manuscript. Xi JIANG, Wenqiang WANG, Minyu WU, and Hongjin BAI collected and analyzed the data. Cuiyun WU and Lirong SHEN designed the experimental research, and wrote and revised the manuscript. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Hao WEN, Xi JIANG, Wenqiang WANG, Minyu WU, Hongjin BAI, Cuiyun WU, and Lirong SHEN declare that they have no conflict of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Covington MF, Maloof JN, Straume M, et al., 2008. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol*, 9:R130.
<https://doi.org/10.1186/gb-2008-9-8-r130>
- Dixon RA, Paiva NL, 1995. Stress-induced phenylpropanoid metabolism. *Plant Cell*, 7(7):1085-1097.
<https://doi.org/10.1105/tpc.7.7.1085>
- Dong XG, Wang Z, Tian LM, et al., 2020. *De novo* assembly of a wild pear (*Pyrus betuleafolia*) genome. *Plant Biotechnol J*, 18(2):581-595.
<https://doi.org/10.1111/pbi.13226>
- Durazzo A, Lucarini M, 2019. Editorial: the state of science and innovation of bioactive research and applications, health, and diseases. *Front Nutr*, 6:178.
<https://doi.org/10.3389/fnut.2019.00178>
- Han JY, Hwang HS, Choi SW, et al., 2012. Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol*, 53(9):1535-1545.
<https://doi.org/10.1093/pcp/pcs106>
- Hoffmann L, Besseau S, Geoffroy P, et al., 2004. Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic acid hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell*, 16(6):1446-1465.
<https://doi.org/10.1105/tpc.020297>
- Legrand G, Delporte M, Khelifi C, et al., 2016. Identification and characterization of five BAHD acyltransferases involved in hydroxycinnamoyl ester metabolism in chicory. *Front Plant Sci*, 7:741.
<https://doi.org/10.3389/fpls.2016.00741>
- Li B, 2019. Diet-related NCDs in China: more needs to be done. *Lancet Public Health*, 4(12):E606.
[https://doi.org/10.1016/S2468-2667\(19\)30218-X](https://doi.org/10.1016/S2468-2667(19)30218-X)
- Li X, Wang TT, Zhou B, et al., 2014. Chemical composition and antioxidant and anti-inflammatory potential of peels and flesh from 10 different pear varieties (*Pyrus* spp.). *Food Chem*, 152:531-538.
<https://doi.org/10.1016/j.foodchem.2013.12.010>
- Liu F, Chen JR, Tang YH, et al., 2018. Isolation and characterization of cinnamate 4-hydroxylase gene from cultivated ramie (*Boehmeria nivea*). *Biotechnol Biotechnol Equip*, 32(2):324-331.
<https://doi.org/10.1080/13102818.2017.1418675>
- Liu GQ, Li WS, Zheng PH, et al., 2012. Transcriptomic analysis of 'Suli' pear (*Pyrus pyrifolia* white pear group) buds during the dormancy by RNA-Seq. *BMC Genomics*, 13:700.
<https://doi.org/10.1186/1471-2164-13-700>
- MacLean DD, Murr DP, Deell JR, et al., 2007. Inhibition of PAL, CHS, and ERS1 in 'Red d'Anjou' pear (*Pyrus communis* L.) by 1-MCP. *Postharvest Biol Technol*, 45(1):46-55.
<https://doi.org/10.1016/j.postharvbio.2007.01.007>
- McLean KJ, Hans M, Munro AW, 2012. Cholesterol, an

- essential molecule: diverse roles involving cytochrome P450 enzymes. *Biochem Soc Trans*, 40(3):587-593. <https://doi.org/10.1042/BST20120077>
- Sullivan ML, Zarnowski R, 2010. Red clover coumarate 3'-hydroxylase (CYP98A44) is capable of hydroxylating *p*-coumaroyl-shikimate but not *p*-coumaroyl-malate: implications for the biosynthesis of phaseolic acid. *Planta*, 231(2):319-328. <https://doi.org/10.1007/s00425-009-1054-8>
- Tuominen LK, Johnson VE, Tsai CJ, 2011. Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among *Populus* paralogues. *BMC Genomics*, 12:236. <https://doi.org/10.1186/1471-2164-12-236>
- Villegas RJA, Kojima M, 1986. Purification and characterization of hydroxycinnamoyl D-glucose: quinate hydroxycinnamoyl transferase in the root of sweet potato, *Ipomoea batatas* Lam. *J Biol Chem*, 261(19):8729-8733. [https://doi.org/10.1016/S0021-9258\(19\)84441-1](https://doi.org/10.1016/S0021-9258(19)84441-1)
- Wang D, Hou JX, Wan JD, et al., 2021. Dietary chlorogenic acid ameliorates oxidative stress and improves endothelial function in diabetic mice via Nrf2 activation. *J Int Med Res*, 49(1):300060520985363. <https://doi.org/10.1177/0300060520985363>
- Wu J, Wang ZW, Shi ZB, et al., 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res*, 23(2):396-408. <https://doi.org/10.1101/gr.144311.112>
- Zhou L, Li CJ, Niu JX, et al., 2018. Identification of miRNAs involved in calyx persistence in Korla fragrant pear (*Pyrus sinkiangensis* Yu) by high-throughput sequencing. *Sci Hortic*, 240:344-353. <https://doi.org/10.1016/j.scienta.2018.06.026>

Supplementary information

Tables S1–S3; Figs. S1–S5