

RESEARCH ARTICLE

Open Access

# Genome-wide transcriptome analysis of Chinese pollination-constant nonastringent persimmon fruit treated with ethanol

Chun Luo<sup>1,2</sup>, Qinglin Zhang<sup>1</sup> and Zhengrong Luo<sup>1\*</sup>

## Abstract

**Background:** The persimmon *Diospyros kaki* Thunb. is an important commercial and deciduous fruit tree. The fruits have proanthocyanidin (PA) content of >25% of the dry weight and are astringent. PAs cause astringency that is often undesirable for human consumption; thus, the removal of astringency is an important practice in the persimmon industry. Soluble PAs can be converted to insoluble PAs by enclosing the fruit in a polyethylene bag containing diluted ethanol. The genomic resource development of the persimmon is delayed because of its large and complex genome. Second-generation sequencing is an efficient technique for generating huge sequences that can represent a large number of genes and their expression levels.

**Results:** We used 454 sequencing for the *de novo* transcriptome assembly of persimmon fruit treated with 5% ethanol (Tr library) and without treatment as the control (Co library) to investigate the genes and pathways that control PA biosynthesis and other secondary metabolites. We obtained 374.6 Mb in clean nucleotides comprising 624,690 and 626,203 clean sequencing reads from the Tr and Co libraries, respectively. We also identified 83,898 unigenes; 54,719 (~65.2%) unigenes were annotated based on similarity searches with known proteins. Up to 14,954 of the unigenes were assigned to the protein database Clusters of Orthologous Groups (COG), 24,337 were assigned to the term annotation database of Gene Ontology (GO), and 45,506 were assigned to 200 pathways in the database of Kyoto Encyclopedia of Genes and Genomes (KEGG). The two libraries were compared to identify the differentially expressed unigenes. The expression levels of genes involved in PA biosynthesis and tannin coagulation were analysed, and some of them were verified using quantitative real time PCR (qRT-PCR).

**Conclusions:** This study provides abundant genomic data for persimmon and offers comprehensive sequence resources for persimmon research. The transcriptome dataset will improve our understanding of the molecular mechanisms of tannin coagulation and other biochemical processes in persimmons.

**Keywords:** Persimmon, Transcriptome analysis, 454 sequencing

## Background

The persimmon *Diospyros kaki* Thunb. ( $2n = 6X = 90$ ) originated in China and was principally cultivated in China, Korea and Japan [1]. Persimmon cultivars are classified into four types, including pollination-constant nonastringent (PCNA), pollination-constant astringent (PCA), pollination-variant nonastringent (PVNA), and pollination-variant astringent (PVA); this classification is based on the effect of pollination on flesh colour and

the natural loss of astringency at the harvest time on the tree [1]. The PCNA type includes Japanese PCNA (JPCNA) and Chinese PCNA (CPCNA), which differ in their genetic characteristic of PCNA trait [2]. The natural loss of astringency is a trait that is qualitatively inherited and recessive in JPCNA cultivars [3,4] but dominant in CPCNA cultivars. When the CPCNA cultivar 'Luotian-shi' is crossed as the maternal parent to a JPCNA or non-PCNA type, the  $F_1$  offspring are segregated into a 1:1 ratio for PCNA:non-PCNA types [5,6]. CPCNA cultivars have attracted attention in the breeding industry because of their natural ability to lose astringency, which is a dominant trait. In addition, CPCNA has the potential

\* Correspondence: luozhr@mail.hzau.edu.cn

<sup>1</sup>Key Laboratory of Horticultural Plant Biology (MOE), Huazhong Agricultural University, 430070 Wuhan, China

Full list of author information is available at the end of the article

to be an important parent in PCNA persimmon breeding in the future.

Persimmon resources are widely distributed in China. However, almost all traditional cultivars native to China are of the PCA type; some of these cultivars include 'Mopanshi', 'Fuping-jianshi', and 'Gongcheng-shuishi' [7]. 'Luotian-tianshi' (*D. kaki* Thunb.;  $2n = 6X = 90$ ) is the first PCNA persimmon native to China, and it is only distributed in Dabieshan Mountain around the junction of Hubei, Henan, and Anhui provinces in central China [4,8].

Most persimmon fruits accumulate proanthocyanidins (PAs) in their flesh during development, causing the sensation of astringency due to the coagulation of oral proteins [9]. PAs or condensed tannins are synthesised *via* the shikimate and flavonoid biosynthetic pathways [10-12]. To date, many genes encoding the structural proteins and transcription factors involved in PA biosynthesis, transportation, and polymerisation have been isolated by homology-based cloning [13-23]. However, the primary genes involved in PA biosynthesis have not yet been determined.

High-throughput sequencing technologies developed in recent years provide a convenient way of establishing a rapid and efficient molecular research platform. Next-generation sequencing (NGS) is related to the Sanger sequencing method, which is represented by first-generation sequencing technologies. Currently, the three mainstream NGS technologies are Roche/454 pyrosequencing (developed in 2005, <http://www.454.com>), Illumina/Solexa sequencing (developed in 2006, <http://www.illumina.com>), and ABI/SOLiD sequencing (developed in 2007, <http://www.appliedbiosystems.com>). These NGS technologies vary in their input requirements and sequence output with regard to the total bases sequenced, length of each sequence read, and price per megabase of sequence information [24]. Among these technologies, 454 sequencing, which generates a minimum number of sequence reads, produces the longest reads (i.e. from 100 bp to ~800–1000 bp). Long reads are optimal for initial genome and transcriptome characterisation because longer pieces are assembled more efficiently than shorter pieces [25]. Given their rapid processing, high throughput, and cost effectiveness, NGS technologies have been successfully used to study genomes and transcriptomes of species with and without sequenced genomes. Many novel and functional genes can be obtained from massive amounts of data.

Abundant genetic resources for persimmons are currently available. However, genomic information and EST sequences for this fruit tree are lacking. In addition, molecular data on persimmons are insufficient when compared to those of other fruit trees, such as apple, pear, peach, citrus, and grape. Accordingly, we performed

large-scale transcriptome sequencing of CPCNA persimmon fruit using Roche/454 technology to create a transcript sequence database of the persimmon and identify candidate genes involved in PA biosynthesis and tannin coagulation. We used IDEG6 to filter the differentially expressed genes in the treatment (Tr) and control (Co) libraries. We also verified the differentially expressed unigenes by quantitative real time PCR (qRT-PCR). The present study provides a platform for studying the genes involved in persimmon tannin coagulation and tannin biosynthesis to analyse the relationship between differentially expressed genes and persimmon fruit deastringency and clarify the mechanism of astringency loss for CPCNA.

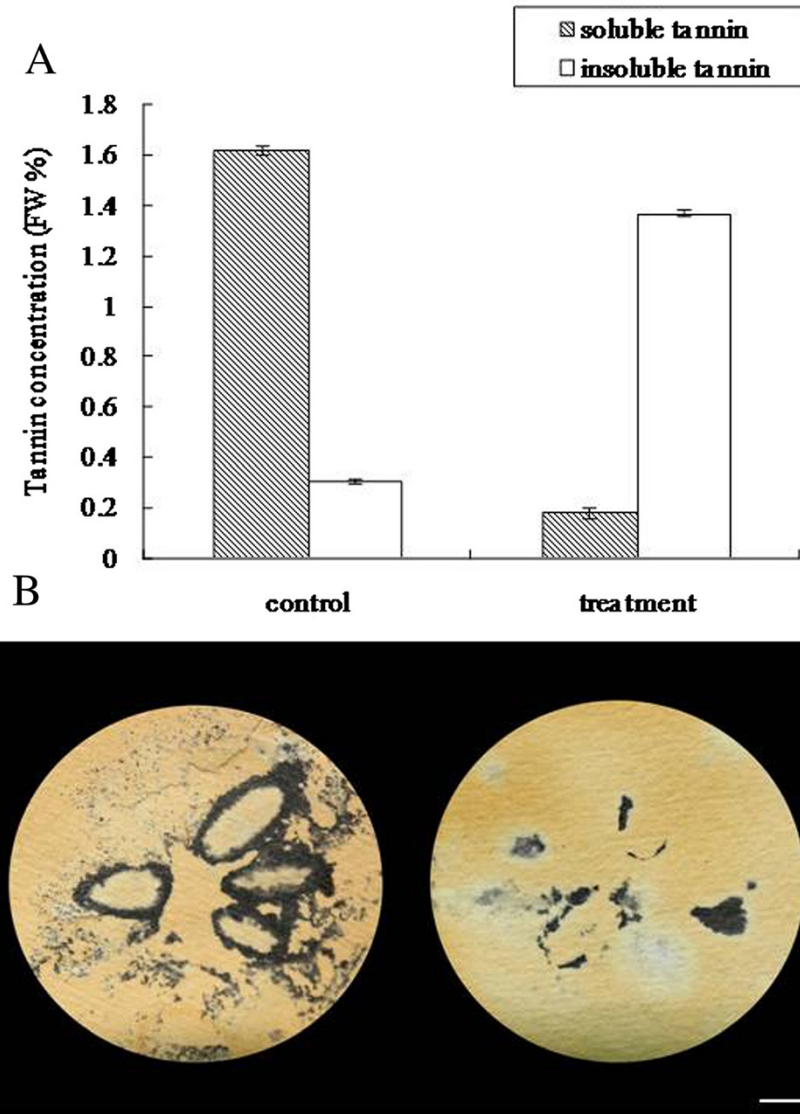
## Results

### Sequencing and assembly

The soluble PA concentration of the CPCNA persimmon fruit was <0.2% in the group treated with 5% ethanol after 3 d but was still very high (1.4%) in the control group (Figure 1A), suggesting that astringency was successfully removed in the treated fruit. In the printing test, the colour change observed after the reaction between FeCl<sub>2</sub> and soluble tannin was dark in the control group but light in the treatment group, suggesting that the amount of soluble tannin decreased in the treated fruit (Figure 1B). A half-plate run using the 454 GS FLX Titanium platform was carried out on the cDNA that was generated by SMART technology with the total RNA from the Tr (5% ethanol-treated) and Co libraries. A total of 624,690 and 626,203 high-quality reads were generated from the Tr and Co libraries, with average sequence lengths of 319 and 309 bp, respectively [National Center for Biotechnology Information (NCBI) Short Read Archive, accession SRA091427]. After trimming the adaptor sequences and removing those shorter than 100 bp, the clean reads of the two libraries were assembled into 83,898 unique sequences using Mimicking Intelligent Read Assembly (MIRA), with an average size of 579 bp. A summary of the 454 sequencing and assembly is presented in Table 1, and the size distributions for these reads and unigenes are presented in Figure 2.

### Sequence annotation

Estimating the number of genes and the level of transcript coverage is difficult because of the lack of genetic or genomic information on persimmons. We performed BLASTX alignments (E-value <  $10^{-5}$  or  $10^{-10}$ ) against the databases of NCBI-nonredundant (Nr), SwissProt, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) to identify the putative functions of the unigene sequences. A total of 54,719 unigenes, accounting for 65.2% of the total unigenes, were annotated to Nr database (Table 2).

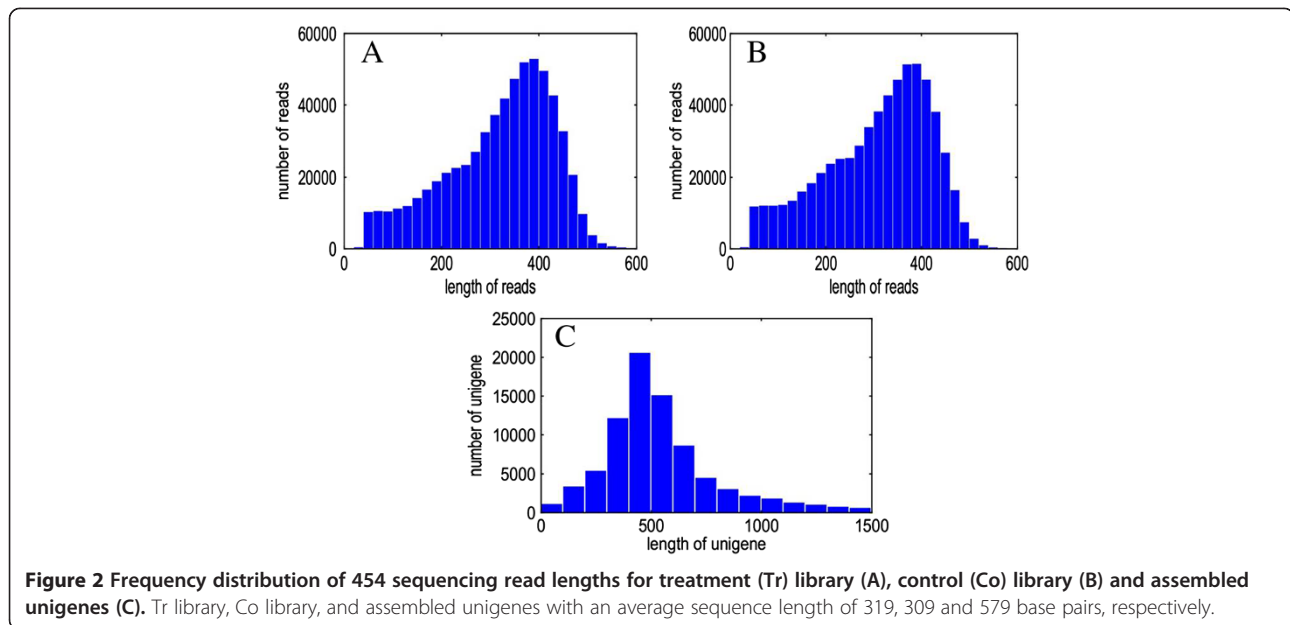


**Figure 1** Effect of ethanol treatment on deastringency. **A:** Effect of ethanol treatment on soluble tannin content and insoluble tannin content of 'Luotian-tianshi'. FW% means the tannin concentration per fresh weight. When it is lower than 0.2%, the persimmon fruit is not astringent. **B:** Analysis of soluble tannin content by FeCl<sub>2</sub> blot. FeCl<sub>2</sub> reacts with the soluble PAs, and the darker the resulting product is, the more astringent the fruit are. Bar = 1 cm.

**Table 1** Summary of the sequencing and assembly

|                                    | Treatment library | Control library |
|------------------------------------|-------------------|-----------------|
| No. of HQ reads                    | 624,690           | 626,203         |
| Total nucleotides (nt)             | 199,375,526       | 193,382,890     |
| Mean length of read (bp)           | 319               | 309             |
| No. of unigenes of combined data   | 83,898            |                 |
| Total nucleotides (nt) of unigenes | 48,556,834        |                 |
| Mean length of unigenes (bp)       | 579               |                 |
| No. of contigs above 500 bp        | 41,451            |                 |

We used GO to obtain a functional annotation of the persimmon unigenes [26]. A total of 24,337 unigenes (29.0%) were assigned to at least one GO term (Figure 3). Of the unigenes assigned to GO terms, 27,406 were assigned to cellular location, 30,629 to molecular function, and 39,037 to biological processes (Additional file 1). These unigenes were further classified into 47 functional subcategories. Within each of the three main categories of the GO classification scheme (i.e. cellular location, molecular function, and biological process), the dominant subcategories were 'cell' (9106), 'cell part' (9106), 'binding' (13,212), 'catalytic activity' (11,774), 'cellular process' (11,929), and 'metabolic process' (12,712).



The protein database Clusters of Orthologous Group (COG) has been designed to classify proteins from completely sequenced genomes based on orthology [27]. We searched annotated sequences for genes involved in COG classification to predict and classify the possible functions of all oriental persimmon unigenes. Of the 54,719 sequences that returned a hit with the Nr database, 14,954 could be assigned to 24 COG categories (Figure 4). Among these categories, the cluster for 'general function prediction only' was the largest group (2456, 18.83%), followed by 'post-translational modification, protein turnover, chaperones' (1900, 14.57%) and 'translation, ribosomal structure and biogenesis' (1811, 13.89%). The clusters for 'nuclear structure' (6, 0.05%), 'cell motility' (10, 0.08%), and 'cell cycle control, cell division, chromosome partitioning' (46, 0.35%) were the smallest groups.

We mapped the annotated sequences to the reference canonical pathways contained in the KEGG database to identify the biological pathways in persimmons. This approach is an alternative way of categorising gene functions by emphasising biochemical pathways [28]. According to the KEGG results, 31,211 unigenes were

mapped to 200 predicted metabolic pathways (Additional file 2). The transcripts identified as related to the following components or processes were the most abundant: translation (4638); folding, sorting, and degradation (4324); carbohydrate metabolism (3181); amino acid metabolism (2283); energy metabolism (2220); and transcription (1962). The largest category was metabolism (13,838), which included carbohydrate metabolism (3181), amino acid metabolism (2283), energy metabolism (2220), enzyme families (1179), lipid metabolism (1038), glycan biosynthesis and metabolism (804), metabolism of other amino acids (729), nucleotide metabolism (609), and other subcategories (Figure 5A). In the secondary metabolism category, 10 subcategories comprised 318 unigenes, the most represented of which were phenylpropanoid biosynthesis (123); flavonoid biosynthesis (62); stilbenoid, diarylheptanoid, and gingerol biosyntheses (25); streptomycin biosynthesis (25); caffeine metabolism (21); and tropane, piperidine, and pyridine alkaloid biosyntheses (19) (Figure 5B). In addition to the metabolic pathways, the genetic information processing pathways (12,838) and cellular processes pathways (2606) were also highly represented categories.

**Table 2** Annotation of non-redundant unigenes

| Database  | Number of annotated unigenes | Percentage of annotated unigenes (%) |
|-----------|------------------------------|--------------------------------------|
| Nr        | 54,719                       | 65.2                                 |
| Swissprot | 31,390                       | 37.4                                 |
| KEGG      | 31,211                       | 37.2                                 |
| COG       | 14,954                       | 17.8                                 |
| Interpro  | 33,287                       | 39.7                                 |
| GO        | 24,337                       | 29.0                                 |

#### Detection of differentially expressed unigenes in the Tr and Co libraries

A previous study proposed that comparing the number of reads for a gene between different libraries or different genes in the same library could be a reliable indicator of relative gene expression [29]. Thus, IDEG6 was used to identify unigenes that show a statistically significant difference in terms of relative abundance (as reflected by the total count of individual sequence reads) between the two

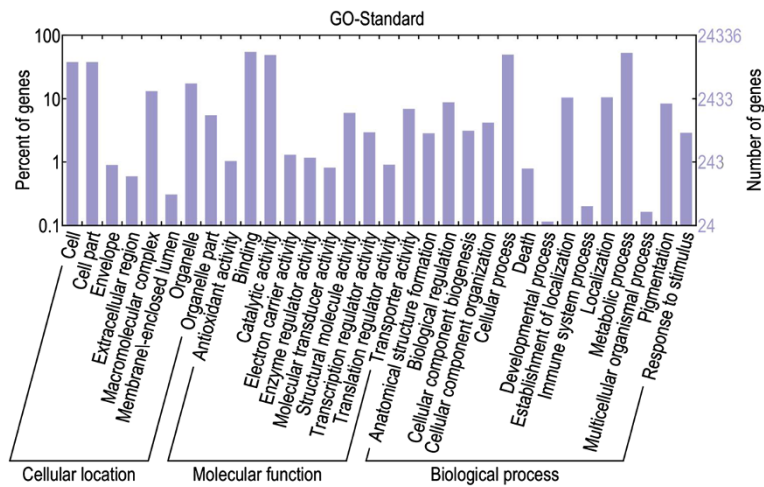
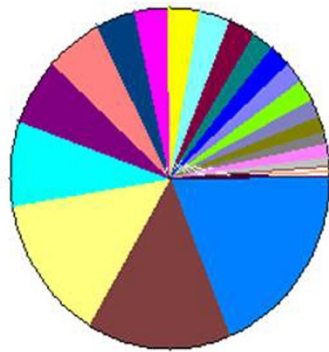


Figure 3 Gene Ontology classifications of assembled non-redundant unigenes.

libraries. A total of 3639 unigenes that were differentially expressed in the Tr and Co libraries were identified. Of these 3639 unigenes, 1560 were upregulated and 2079 were downregulated in the treated fruit. More genes were expressed in the treated fruit (15,331) than the control fruit (12,502) (Table 3).

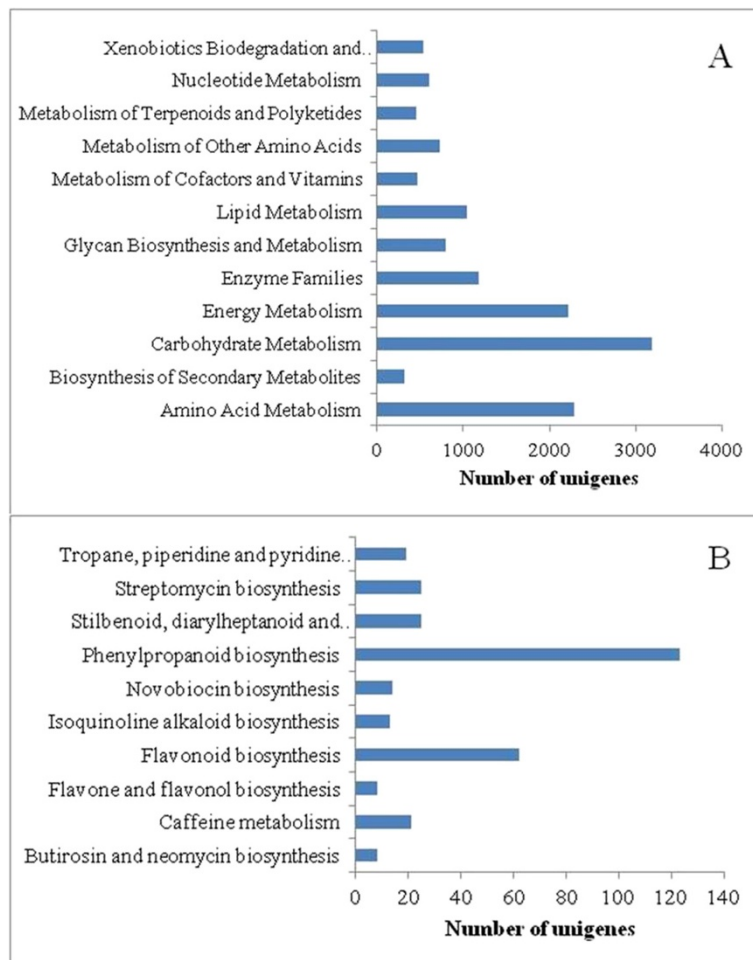
### Quantitative real-time PCR confirmation

Fifty differentially expressed unigenes were selected for qRT-PCR analysis (primers are shown in Additional file 3) to confirm the expression of the unigenes from the sequencing and computational analyses. cDNA fragments of the control and flesh treated for 3 d (15 July 2011;



Clusters of Orthologous Groups

Figure 4 Clusters of orthologous groups (COG) classification.



**Figure 5** Pathway assignment based on KEGG. (A) Classification based on metabolism categories; (B) Classification based on secondary metabolism categories.

73 d after bloom) were used as templates. The results showed that the qRT-PCR assessments (relative expressed level) of 34 unigenes (68%) were consistent with those of the 454 sequencing analysis (Figure 6). These results suggest that our transcriptome data have a high coverage.

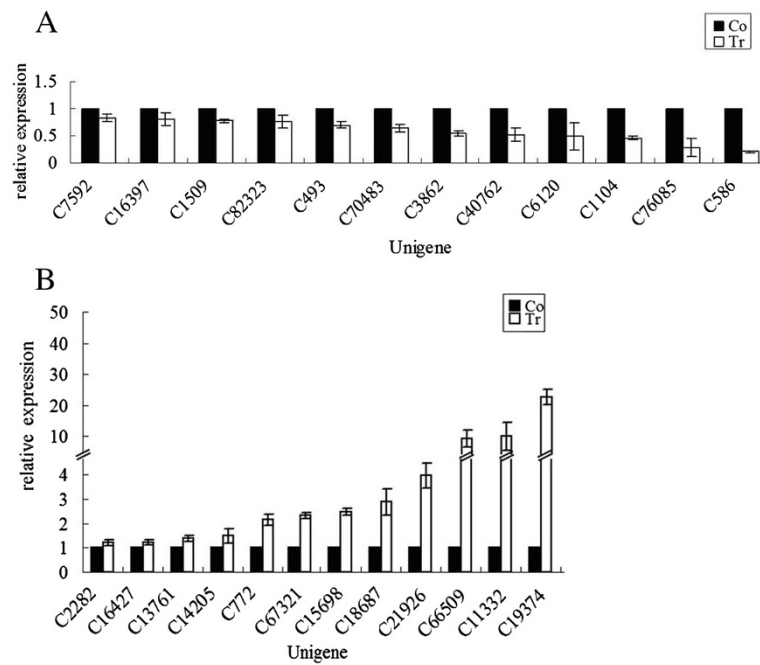
**Table 3** Summary of differentially expressed unigenes

|   | Number of unigenes | Percentage (%) |
|---|--------------------|----------------|
| Total unigene                           | 83898              |                |
| Differentially expressed in two library | 3639               | 4.34           |
| Up (Tr vs Co)                           | 1560               | 1.86           |
| Down (Tr vs Co)                         | 2079               | 2.48           |
| Expressed both in Tr and Co             | 56065              | 66.83          |
| Expressed only in Tr                    | 15331              | 18.27          |
| Expressed only in Co                    | 12502              | 14.9           |

## Discussion

Currently, the most common application of NGS in non-model species is transcriptome characterisation [30-34]. Among the currently available NGS technologies, 454 pyrosequencing produces the longest reads; thus, it has emerged as a powerful tool for transcriptome sequencing. In addition, many studies have used *de novo* assembly of such data to produce genome-level resources for non-model organisms [35-38].

Genetic studies of the persimmon are difficult to implement because of the hexaploid nature of the species and lack of linkage maps and whole genome sequences [i.e. only 14,189 EST sequences deposited in GenBank (accessed on 10/11/2013)]. Previous genetic studies have focused on the diversity and phylogeny of cultivated persimmons and related wild species. The transcriptome characterisation described in the present study will provide the initial information needed for the functional



**Figure 6 Validation of part of differentially expressed unigenes by qRT-PCR. (A)** down-regulated unigenes in treated fruit. C7592: aldehyde dehydrogenase; C16397: aldehyde dehydrogenase family 2 member B7; C1509: NADH dehydrogenase; C82323: Transcription factor bHLH49; C493: 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1; C70483: 3-dehydroquinate dehydratase/shikimate 5-dehydrogenase; C3862: *Diospyros kaki* anthocyanidin synthase; C40762: *D. kaki* flavonoid 3'5' hydroxylase; C6120: VTC2-like protein; C1104: *D. kaki* DkMyb2 mRNA for putative MYB transcription factor; C76085: *D. kaki* DKSCPL1 mRNA for serine carboxypeptidase-like protein 1; C586: DKSCPL1 mRNA for serine carboxypeptidase-like protein 1. **(B)** up-regulated unigenes in treated fruit. C2282: *D. kaki* MADS-box protein (MADS1) mRNA; C16427: dehydroascorbate reductase 1; C13761: aldehyde oxidase; C14205: *D. kaki* DkSK mRNA for putative shikimate kinase; C772: WD-repeat protein; C67321: 4-hydroxyphenylpyruvate dioxygenase; C15698: glutathione peroxidase; C18687: alcohol dehydrogenase (Adh3) mRNA; C21926: *D. kaki* beta-carotene hydroxylase mRNA; C66509: tetrahydroxychalcone 2'-glucosyltransferase; C11332: tonoplast intrinsic protein; C19374: ferulate 5-hydroxylase (F5H) mRNA.

study of persimmons. A total of 624,690 and 626,203 reads were generated from the Tr and Co libraries, with average sequence lengths of 319 and 309 bp, respectively. We obtained 83,898 unigenes from these raw reads (Table 1). The average unigene length for persimmons in this study was comparable to that observed in other species such as *Oncidium* (493 bp) [39], *Pinus contorta* (500 bp) [40], *Fraxinus* (649 bp) [41], and *Vicia faba* (615 bp) [42]. Further, the unigene was longer than that observed in *Ziziphus celata* (408 bp) [43] and *Olea europaea* (355 bp) [44] but shorter than that in *Castanea mollissima* (731 bp) [31] and *Pyrus pyrifolia* (853 bp) [45]. The length of unigenes may be related to the sequencing technique and assembly tools used. However, the unigenes assembled in the present study could provide resources for future persimmon genetic and genomic research.

BLAST searches against public databases, such as NR, SwissProt, GO, and KEGG, provided annotation data for the persimmon, with 54,719 (65.2%) unique hits (Table 2). The annotation of persimmon gained more descriptive information than that observed in other species such as *Conyza canadensis* (51.3%) [46], *Bupleurum chinense* (52.6%) [47], *Dendrocalamus latiflorus* (54.9%) [48], *Lens*

*culinaris* (55.6%) [49], and *Panax quinquefolius* (63.6%) [50]. However, it was lower than that observed in *Fagopyrum* (66.7%) [51], *Taxus cuspidata* (68.6%) [52], *Capsicum annuum* (72.04%) [53], *Olea europaea* (73%) [44], *Pyrus pyrifolia* (74.1%) [45], *Dendrocalamus latiflorus* (78.9%) [54], *Eichhornia paniculata* (87.0%) [55], and *Fraxinus* (99%) [41]. However, comparing this information across species studies is difficult because of the different sequencing depths or BLAST parameters utilized in each report [55].

A total of 31,211 persimmon unigenes were mapped into 200 KEGG pathways (Additional file 2). The genetic information processing pathway (12,838) and cellular processes pathway (2606) were highly represented categories in the metabolic pathways. Most persimmon fruits accumulate PAs in their flesh during development; PAs cause astringency due to coagulation of oral proteins and are synthesised from metabolites *via* the shikimate and flavonoid pathways [10-12]. Therefore, we focused on the pathways pertaining to phenylpropanoid biosynthesis (ko00940, 123 unigenes) and flavonoid biosynthesis (ko00941, 62 unigenes). Most of the genes related to PA biosynthesis in these two pathways were found in

our transcriptome sequencing data. PAs cause astringency that is often undesirable for human consumption; thus, the removal of astringency is an important practice in the persimmon food industry. Soluble PAs can be converted to insoluble PAs by enclosing the fruit in a polyethylene bag containing diluted ethanol [56]. Acetaldehyde formed *in situ* from ethanol is involved in the direct insolubilisation of soluble PAs, causing a loss of astringency [57]. Pyruvate decarboxylase (PDC) and alcohol dehydrogenase (ADH) are two important enzymes in this process, which is involved in the glycolysis/gluconeogenesis pathway (ko00010, 496 unigenes). Moreover, six and 31 unigenes were classified into the PDC and ADH families, respectively.

The expression levels of the unigenes were reflected by the number of reads used to characterise the differences in gene expression between the Tr and Co transcriptome libraries. A total of 3639 unigenes were found to be differently expressed (Table 3). The expression levels of the unigenes involved in PA biosynthesis and tannin coagulation are shown in Table 4. The expression levels of *ADH1*, *4CL*, *ANS*, and *F3'5'H* were significantly downregulated, whereas those of *ADH3*, *PDC*, *CHS*, *F3H*, and *LAR* were significantly upregulated after the removal of astringency. These results suggest that the expression of genes involved in PA biosynthesis might be affected by ethanol treatment, which is consistent with the findings of Ikegami et al. [14].

**Table 4 Summary of genes involved in PA biosynthesis and tannin coagulation**

| Gene name     | Number of unigenes | Number of reads/Co | Number of reads/Tr |
|---------------|--------------------|--------------------|--------------------|
| <i>ADH 1</i>  | 27                 | 217                | 163                |
| <i>ADH 2</i>  | 2                  | 8                  | 8                  |
| <i>ADH 3</i>  | 2                  | 3                  | 21                 |
| <i>PDC</i>    | 6                  | 9                  | 24                 |
| <i>ALDH2</i>  | 11                 | 1,409              | 545                |
| <i>PAL</i>    | 12                 | 48                 | 40                 |
| <i>4CL</i>    | 15                 | 249                | 152                |
| <i>CHS</i>    | 2                  | 1                  | 5                  |
| <i>CHI</i>    | 4                  | 14                 | 26                 |
| <i>F3H</i>    | 2                  | 5                  | 11                 |
| <i>F3'H</i>   | 2                  | 5                  | 8                  |
| <i>F3'5'H</i> | 4                  | 34                 | 26                 |
| <i>DFR</i>    | 4                  | 6                  | 8                  |
| <i>LAR</i>    | 5                  | 5                  | 52                 |
| <i>ANS</i>    | 11                 | 140                | 109                |
| <i>ANR</i>    | 2                  | 2                  | 2                  |
| <i>MYB</i>    | 12                 | 104                | 91                 |
| <i>bHLH</i>   | 3                  | 86                 | 90                 |
| <i>WD40</i>   | 2                  | 14                 | 32                 |

Acetaldehyde produced from ethanol is involved in the direct insolubilisation of soluble PAs [57]. The synthesis of acetaldehyde is generally catalysed by PDC, which converts pyruvate to acetaldehyde; meanwhile, ADH is involved in the potentially reversible interconversion of ethanol and acetaldehyde [58]. Furthermore, *ADH1* and *PDC* are suggested as the key genes involved in persimmon astringency removal [59]. In the present study, the *PDC* gene was upregulated, which resulted in the production of more acetaldehyde. However, the *ADH1* gene was downregulated, which may result in the reduction of the conversion of acetaldehyde into ethanol. This result suggests that acetaldehyde accumulated in the ethanol-treated fruit, which resulted in the loss of astringency. However, this result was not consistent with that reported in Min's study [59], where *ADH1* was upregulated by ethylene. This inconsistency in the results can be attributed to the different cultivars used (i.e. the non-astringent type 'Luotian-tianshi' was used in our study, while the astringent type 'Mopanshi' was used in Min's research). The use of different materials might cause different gene expression patterns; however, this hypothesis needs to be validated by further experimentation.

We found that the gene aldehyde dehydrogenase family 2 (*ALDH2*) was highly expressed (1409 reads) in the Co library and downregulated in the Tr library (545 reads), with a total of 11 unigenes in both libraries (Table 4). *ALDH2* has a broad expression pattern and is most notably involved in the second step of ethanol metabolism, (i.e. acetaldehyde oxidation). The decrease in *ALDH2* in the Tr library might have inhibited the conversion of acetaldehyde to acetic acid, which, consequently, led to acetaldehyde accumulation. Large amounts of acetaldehyde triggered the coagulation of tannins (insolubilisation of soluble PAs) causing the loss of astringency in the treated persimmon fruits. This result suggests that the *ALDH2* gene, together with the *ADH* and *PDC* genes, might have important functions in tannin coagulation.

The present study on persimmon transcriptome has several biological implications. First, the plant material persimmon, which accumulates PAs (condensed tannins) in its flesh during development, can be considered a model plant for tannin research. Second, the loss of astringency in CPCNA fruits treated with ethanol is an imitation of the natural loss of astringency, especially for tannin coagulation. This imitation helped us to understand mechanism of astringency loss in CPCNA. Third, the current study, based on the present transcriptome data (even in the absence of complete genome sequences for persimmons), will facilitate the advancement future genetic studies.

## Conclusions

This work presents the first *de novo* transcriptome sequencing analysis of the CPCNA persimmon fruit using



the 454 GS FLX Titanium platform. A total of 374.6 Mb of data were generated and assembled into 83,898 unigenes. Persimmon unigenes related to PA biosynthesis were characterised, and differentially expressed unigenes in the two libraries were verified using qRT-PCR. *ADH*, *PDC*, and the newly discovered persimmon gene *ALDH2* were found to have important functions in tannin coagulation. To the best of our knowledge, this study is the first to employ the 454 sequencing technology to investigate the whole transcriptome of the persimmon fruit. The assembly of the reads was also conducted without a reference genome. The transcriptome characterisation described in the present study will provide the initial information needed for the functional study of persimmons to elucidate the molecular mechanisms of tannin coagulation and other biochemical processes in this fruit tree.

## Methods

### Sample preparation

In previous analysis of tannin concentration per fruit, JPCNA and CPCNA varied considerably. Both types accumulate PA in their fruits at an early stage. PA accumulation is halted in JPCNA at 7–9 weeks after bloom (WAB), and a low concentration of PA is observed at 10 WAB. On the other hand, CPCNA continuously accumulates PA until the late stages and maintains a very high PA concentration [22,23]. Thus, it appears that at 9–10 WAB, JPCNA and CPCNA exhibit different PA accumulation patterns. In the present study, 30 young fruits on a CPCNA-type persimmon tree (*D. kaki* 'Luotian-tianshi',  $2n = 6X = 90$ ) grown in the Persimmon Repository of Huazhong Agricultural University, China, were enclosed with polyethylene bags containing 10 mL of 5% ethanol on 12 July 2011 (10 WAB). Control (untreated) fruits were enclosed with polyethylene bags containing 10 mL of water. Three days later, all treated and control fruits were sampled and peeled. The flesh of the fruits was diced into small pieces, frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  until use for RNA isolation.

The concentrations of soluble and insoluble tannins in the control and treated fruit flesh were measured by the Folin–Ciocalteu method after 3 d of treatment [60]. Soluble tannins of the fruit flesh were also examined after 3 d of treatment by the printing method [61], which is a convenient way of identifying persimmon astringency loss.  $\text{FeCl}_2$  reacts with the soluble PAs; thus, the darker the resulting product, the more astringent the fruits.

### RNA extraction, cDNA library construction, and 454 sequencing

For each sample (5% ethanol treated and control), approximately 10 g of mixed flesh (10 individuals) was used for RNA preparation and tannin concentration measurements. Total RNA was extracted using TRIzol Reagent (Invitrogen,

USA) following the manufacturer's protocol. The quality and quantity of the total RNA was analysed using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) and gel electrophoresis.

Approximately 1  $\mu\text{g}$  of RNA was used to generate double-stranded cDNA using the SMART<sup>TM</sup> cDNA Library Construction Kit (Clontech, USA). Finally,  $\sim 5$  mg of cDNA was used to construct a 454 library. Roche GS-FLX 454 pyrosequencing was conducted by the Oebiotech Company in Shanghai, China.

### 454 de novo transcriptome assembly and analysis

A Perl program was written to remove vector sequences and the PolyA (T) tail from sequences; reads with lengths  $<100$  bp were removed before assembly. Then, high-quality reads were assembled with MIRA [62] to construct unique consensus sequences. The 454 setting parameters were used by MIRA (--job = denovo,est,normal, 454; -SK:mnr = yes; -SK:rt = 2; 454\_SETTINGS -LR: mxti = no; -CL:qc = no).

The functions of the unigenes were annotated by BLASTX with an E-value threshold of  $10^{-5}$  to the protein databases, including NCBI-NR, Swiss-Prot, KEGG [28], and COG [63]. InterPro domains [64] were annotated by InterProScan [65] Release 16.0, and functional assignments were mapped onto GO [66]. WEGO [67] was used for GO classification and GO tree construction.

### Differentially expressed unigene detection

A freely available web tool IDEG6 [68] was used to identify unigenes showing statistically significant differences in relative abundance (as reflected by the total count of individual sequence reads) between the Tr and Co libraries. The general Chi-squared method was used because it was the most efficient analytical method [68]. Finally, unigenes with  $P \leq 0.01$  were deemed significantly different between the two libraries.

### RNA extraction, first-strand cDNA synthesis, and qRT-PCR analysis

A total of 50 unigenes generated by 454 sequencing were selected for experimental validation. The total RNA used for qRT-PCR analysis was extracted from the flesh of the Tr and Co fruits. After RNA extraction, first-strand cDNA was synthesised from 1.0  $\mu\text{g}$  of RNA using the PrimeScript<sup>®</sup> RT Reagent Kit with gDNA Eraser (TaKaRa, Dalian, China) according to the manufacturer's protocol. The cDNA was diluted threefold and used as the template for qRT-PCR. qRT-PCR was performed on a LightCycler<sup>®</sup> 480 II System (Roche Diagnostics) using SYBR<sup>®</sup> Premix Ex Taq<sup>TM</sup> II (TaKaRa). The reaction was composed as described in the manual and was performed in quadruplicate. A negative control (no template) was included in each run. The standard amplification protocol consisted of

an initial denaturing step of 95°C for 30 s, followed by 45 cycles of 95°C for 5 s, 60°C for 10 s, 72°C for 15 s, and a melting temperature cycle with constant fluorescence data acquisition from 65°C to 95°C. The gene quantification method was based on the relative expression of the target gene versus the reference gene (*DkActin*), and the ratio was calculated with the LightCycler® 480 software. All primers are listed in Additional file 3.

## Additional files

**Additional file 1:** GO classifications of CPCNA persimmon unigenes.

**Additional file 2:** KEGG pathway annotation of CPCNA persimmon unigenes.

**Additional file 3:** Information for primers used in qRT-PCR analysis.

The primer sequences of 34 unigenes verified successfully by qRT-PCR were listed in this table.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

Conceived and designed the experiments: L-ZR, LC and Z-QL. Performed the experiments and analyzed the data: LC. Wrote the manuscript: LC. Read and approved the final manuscript: LC, Z-QL and L-ZR.

## Acknowledgments

This research was supported by the Natural Science Foundation of China (31171929) and the Special Scientific Research Fund of Agricultural Public Welfare Profession of China (201203047).

## Author details

<sup>1</sup>Key Laboratory of Horticultural Plant Biology (MOE), Huazhong Agricultural University, 430070 Wuhan, China. <sup>2</sup>Key Laboratory of Tropical Fruit Biology, Ministry of Agriculture, South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang 524091, China.

Received: 8 August 2013 Accepted: 4 February 2014

Published: 8 February 2014

## References

1. Yonemori K, Sugiura A, Yamada M: **Persimmon genetics and breeding.** In *Plant Breeding Reviews. Volume 19*. Edited by John JJ. Oxford, UK: Wiley & Sons, Inc; 2010.
2. Akagi T, Katayama-Ikegami A, Yonemori K: **Proanthocyanidin biosynthesis of persimmon (*Diospyros kaki* Thunb.) fruit.** *Sci Horti* 2011, **130**:373–380.
3. Ikeda I, Yamada M, Kurihara A: **Inheritance of astringency in Japanese persimmon.** *J Jpn Soc Hortic Sci* 1985, **54**(1):39–45.
4. Yamada M, Sato A, Yakushiji H, Yoshinaga K, Yamane H, Endo M: **Characteristics of "Luo Tian Tian Shi", a non-astringent cultivar of oriental persimmon (*Diospyros kaki* Thunb.) of Chinese origin in relation to non-astringent cultivars of Japanese origin.** *Bull Fruit Tree Res Sta* 1993, **25**:19–32.
5. Ikegami A, Yonemori K, Sugiura A, Sato A, Yamada M: **Segregation of astringency in F1 progenies derived from crosses between pollination-constant, nonastringent persimmon cultivars.** *HortScience* 2004, **39**:371–374.
6. Ikegami A, Eguchi S, Yonemori K, Yamada M, Sato A, Mitani N, Kitajima A: **Segregations of astringent progenies in the F1 populations derived from crosses between a Chinese pollination-constant nonastringent (PCNA) 'Luo Tian Tian Shi' and Japanese PCNA pollination-constant astringent (PCA) cultivars of Japanese origin.** *HortScience* 2006, **41**:561–563.
7. Luo Z, Wang R: **Persimmon in China: domestication and traditional utilizations of genetic resources.** *Adv Horti Sci* 2008, **22**:239–243.
8. Wang R: **The origin of 'Luo Tian Tian Shi'.** *Chin Fruit Tree* 1982, **2**:16–19.
9. Porter LJ, Woodruffe J: **Haemalysis: the relative astringency of proanthocyanidin polymers.** *Phytochemistry* 1984, **23**:1255–1256.
10. Herrmann KM: **The shikimate pathway as an entry to aromatic secondary metabolism.** *Plant Physiol* 1995, **107**:7–12.
11. Xie DY, Dixon RA: **Proanthocyanidin biosynthesis—still more questions than answers?** *Phytochemistry* 2005, **66**:2127–2144.
12. Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M: **Genetics and biochemistry of seed flavonoids.** *Annu Rev Plant Biol* 2006, **57**:405–430.
13. Ikegami A, Kitajima A, Yonemori K: **Inhibition of flavonoid biosynthetic gene expression coincides with loss of astringency in pollination-constant, non-astringent (PCNA)-type persimmon fruit.** *J Horti Sci Biotech* 2005, **80**:225–228.
14. Ikegami A, Eguchi S, Kitajima A, Inoue K, Yonemori K: **Identification of genes involved in proanthocyanidin biosynthesis of persimmon (*Diospyros kaki*) fruit.** *Plant Sci* 2007, **172**:1037–1047.
15. Nakagawa T, Nakatsuka A, Yano K, Yasugahira S, Nakamura R, Sun N, Itai A, Suzuki T, Itamura H: **Expressed sequence tags from persimmon at different developmental stages.** *Plant Cell Rep* 2008, **27**:931–938.
16. Akagi T, Ikegami A, Suzuki Y, Yoshida J, Yamada M, Sato A, Yonemori K: **Expression balances of structural genes in shikimate and flavonoid biosynthesis cause a difference in proanthocyanidin accumulation in persimmon (*Diospyros kaki* Thunb.) fruit.** *Planta* 2009, **230**:899–915.
17. Akagi T, Ikegami A, Tsujimoto T, Kobayashi S, Sato A, Kono A, Yonemori K: **DkMyb4 is a Myb transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit.** *Plant Physiol* 2009, **151**:2028–2045.
18. Akagi T, Ikegami A, Yonemori K: **DkMyb2 wound-induced transcription factor of persimmon (*Diospyros kaki* Thunb.), contributes to proanthocyanidin regulation.** *Planta* 2010, **232**:1045–1059.
19. Akagi T, Tsujimoto T, Ikegami A, Yonemori K: **Effects of seasonal temperature changes on DkMyb4 expression involved in proanthocyanidin regulation in two genotypes of persimmon (*Diospyros kaki* Thunb.) fruit.** *Planta* 2011, **233**:883–894.
20. Akagi T, Katayama-Ikegami A, Kobayashi S, Sato A, Kono A, Yonemori K: **Seasonal abscisic acid signal and a basic leucine zipper transcription factor, DkbZIP5, regulate proanthocyanidin biosynthesis in persimmon fruit.** *Plant Physiol* 2012, **158**:1089–1102.
21. Wang Y, Zhang Q, Luo Z: **Isolation and expression of gene encoding leucoanthocyanidin reductase from *Diospyros kaki* during fruit development.** *Biol Plantarum* 2010, **54**:707–710.
22. Su F, Hu J, Zhang Q, Luo Z: **Isolation and characterization of a basic Helix–Loop–Helix transcription factor gene potentially involved in proanthocyanidin biosynthesis regulation in persimmon (*Diospyros kaki* Thunb.).** *Sci Horti* 2012, **136**:115–121.
23. Hu Q, Luo C, Zhang Q, Luo Z: **Isolation and characterization of a laccase gene potentially involved in proanthocyanidin polymerization in Oriental persimmon (*Diospyros kaki* Thunb.) fruit.** *Mol Biol Rep* 2013, **40**:2809–2820.
24. Brautigam A, Gowik U: **What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research.** *Plant Biol* 2010, **12**:831–841.
25. Glenn TC: **Field guide to next-generation DNA sequencers.** *Mol Ecol Resour* 2011, **11**:759–769.
26. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, et al: **The Gene Ontology (GO) database and informatics resource.** *Nucleic Acids Res* 2004, **32**:D258–D261.
27. Tatusov RL: **A genomic perspective on protein families.** *Science* 1997, **278**:631–637.
28. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucleic Acids Res* 2004, **32**:D277–D280.
29. Audic SCJ: **The significance of digital gene expression profiles.** *Genome Res* 1997, **7**:986–995.
30. Novaes E, Drost DR, Farmerie WG, Pappas GJ Jr, Grattapaglia D, Sederoff RR, Kirst M: **High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome.** *BMC Genomics* 2008, **9**:312.
31. Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE: **Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection.** *BMC Plant Biol* 2009, **9**:51.
32. Wall PK, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H, Landherr L, Tomsho LP, Hu Y, Carlson JE: **Comparison of next generation**

- sequencing technologies for transcriptome characterization. *BMC Genomics* 2009, **10**:347.
33. Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, Li Y: **De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.).** *BMC Genomics* 2012, **13**:90.
34. Xie F, Burklew CE, Yang Y, Liu M, Xiao P, Zhang B, Qiu D: **De novo sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome.** *Planta* 2012, **236**:101–113.
35. Wheat CW: **Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing.** *Genetica* 2010, **138**:433–451.
36. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing.** *Mol Ecol* 2008, **17**:1636–1647.
37. Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, Colbourne JK, Willis BL, Matz MV: **Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFLX.** *BMC Genomics* 2009, **10**:219.
38. Der JP, Barker MS, Wickett NJ, dePamphilis CW, Wolf PG: **De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*.** *BMC Genomics* 2011, **12**:99.
39. Chang YY, Chu YW, Chen CW, Leu WM, Hsu HF, Yang CH: **Characterization of *Oncidium* 'Gower Ramsey' transcriptomes using 454 GS-FLX pyrosequencing and their application to the identification of genes associated with flowering time.** *Plant Cell Physiol* 2011, **52**:1532–1545.
40. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA: **Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery.** *BMC Genomics* 2010, **11**:180.
41. Bai XD, Rivera-Vega L, Mamidala P, Bonello P, Herms DA, Mittapalli O: **Transcriptomic signatures of ash (*Fraxinus* spp.) phloem.** *PLoS ONE* 2011, **6**:e16368.
42. Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonforte T, Paull J, Materne M, Forster JW: **Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers.** *BMC Genomics* 2012, **13**:104.
43. Edwards CE, Parchman TL, Weekley CW: **Assembly, gene annotation and marker development using 454 floral transcriptome sequences in *Ziziphus celata* (Rhamnaceae), a highly endangered, Florida endemic plant.** *DNA Res* 2012, **19**:1–9.
44. Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G: **Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development.** *BMC Genomics* 2009, **10**:399.
45. Liu G, Li W, Zheng P, Xu T, Chen L, Liu D, Hussain S, Teng Y: **Transcriptomic analysis of 'Suli' pear (*Pyrus pyrifolia* white pear group) buds during the dormancy by RNA-Seq.** *BMC Genomics* 2012, **13**:700.
46. Peng Y, Abercrombie LL, Yuan JS, Riggins CW, Sammons RD, Tranel PJ, Stewart CN Jr: **Characterization of the horseweed (*Conyza canadensis*) transcriptome using GS-FLX 454 pyrosequencing and its application for expression analysis of candidate non-target herbicide resistance genes.** *Pest Manag Science* 2010, **66**:1053–1062.
47. Sui C, Zhang J, Wei J, Chen S, Li Y, Xu J, Jin Y, Xie C, Gao Z, Chen H, Yang C, Zhang Z, Xu Y: **Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins.** *BMC Genomics* 2011, **12**:539.
48. Zhang XM, Zhao L, Larson-Rabin Z, Li DZ, Guo ZH: **De novo sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae).** *PLoS ONE* 2012, **7**:e42082.
49. Kaur S, Cogan NO, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW: **Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery.** *BMC Genomics* 2011, **12**:265.
50. Sun C, Li Y, Wu Q, Luo H, Sun Y, Song J, Lui EM, Chen S: **De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis.** *BMC Genomics* 2010, **11**:262.
51. Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ, Penin AA: **De novo sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*).** *BMC Genomics* 2011, **12**:30.
52. Wu Q, Sun C, Luo H, Li Y, Niu Y, Sun Y, Lu A, Chen S: **Transcriptome analysis of *Taxus cuspidata* needles based on 454 pyrosequencing.** *Planta Med* 2011, **77**:394–400.
53. Lu FH, Yoon MY, Cho Yi, Chung JW, Kim KT, Cho MC, Cheong SR, Park YJ: **Transcriptome analysis and SNP/SSR marker information of red pepper variety YCM334 and Taean.** *Sci Horti* 2011, **129**:38–45.
54. Liu M, Qiao G, Jiang J, Yang H, Xie L, Xie J, Zhuo R: **Transcriptome sequencing and de novo analysis for ma bamboo (*Dendrocalamus latiflorus* Munro) using the Illumina platform.** *PLoS ONE* 2012, **7**:e46766.
55. Ness RW, Siol M, Barrett SC: **De novo sequence assembly and characterization of the floral transcriptome in cross- and self-fertilizing plants.** *BMC Genomics* 2011, **12**:298.
56. Sugjura A, Harada H, Tomana T: **Studies on the removability of astringency in Japanese persimmon fruits, 1.** *J Jpn Soc Hortic Sci* 1975, **44**:265–272.
57. Tanaka T, Takahashi R, Kouno I, Nonaka G: **Chemical evidence for the de-astringency (insolubilization of tannins) of persimmon fruit.** *J Chem Soc, Perkin Trans 1* 1994, **20**:3013–3022.
58. Strommer J: **The plant ADH gene family.** *Plant J* 2011, **66**:128–142.
59. Min T, Yin X, Shi YN, Luo ZR, Yao YC, Grierson D, Ferguson IB, Chen KS: **Ethylene-responsive transcription factors interact with promoters of ADH and PDC involved in persimmon (*Diospyros kaki*) fruit de-astringency.** *J Exp Bot* 2012, **63**:6393–6405.
60. Oshida M, Yonemori K, Sugiura A: **On the nature of coagulated tannins in astringent-type persimmon fruit after an artificial treatment of astringency removal.** *Postharvest Biol Tec* 1996, **8**:317–327.
61. Eaks I: **Ripening and astringency removal in persimmon fruits.** *Proc Amer Soc Hort Sci* 1967, **91**:868–875.
62. Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, Suhai S: **Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs.** *Genome Res* 2004, **14**:1147–1159.
63. Tatusov RL, Galperin M, Natale DA, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution.** *Nucleic Acids Res* 2008, **28**:33–36.
64. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, Bateman A, Binns D, Biswas M, Bradley P, Bork P, Bucher P, Copley RR, Courcelle E, Das U, Durbin R, Falquet L, Fleischmann W, Griffiths-Jones S, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lopez R, Letunic I, Lonsdale D, Silventoinen V, Orchard SE, Pagni M, et al: **The InterPro Database, 2003 brings increased coverage and new features.** *Nucleic Acids Res* 2003, **31**:315–318.
65. Zdobnov EM, Apweiler R: **InterProScan an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847–848.
66. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nat Genet* 2000, **25**:25–29.
67. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L: **WEGO: a web tool for plotting GO annotations.** *Nucleic Acids Res* 2006, **34**:W293–W297.
68. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA: **IDE6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments.** *Physiol Genomics* 2003, **12**:159–162.

doi:10.1186/1471-2164-15-112

**Cite this article as:** Luo et al.: Genome-wide transcriptome analysis of Chinese pollination-constant nonastringent persimmon fruit treated with ethanol. *BMC Genomics* 2014 **15**:112.