Genetics and Molecular Research

De novo assembly, functional annotation, and marker development of Asian pear (*Pyrus pyrifolia*) fruit transcriptome through massively parallel sequencing

J.F. Li, Z. Gao, Y.S. Lou, M. Luo, S.R. Song, W.P. Xu, S.P. Wang and C.X. Zhang

Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China

Corresponding author: C.X. Zhang
E-mail: acaizh@sjtu.edu.cn

Genet. Mol. Res. 14 (4): 18344-18355 (2015)
Received August 29, 2015
Accepted October 18, 2015
Published December 23, 2015
DOI http://dx.doi.org/10.4238/2015.December.23.22

ABSTRACT. This study investigated the Asian pear transcriptome using the RNA-Seq normalized fruit cDNA library to create a transcriptomic resource for unigene and marker discovery. Following the removal of lowquality reads, 127,085,054 trimmed reads were assembled *de novo* to yield 37,649 non-redundant unigenes with an average length of 599 bp. Alternative splicing events were detected in 4121 contigs. A total of 30,560 single nucleotide polymorphisms (SNPs) and 7443 simple sequence repeat (SSR) makers were obtained. Approximately 21,449 (56.9%) unigenes were categorized into three gene ontology groups; 3682 (9.8%) were classified into 25 cluster of orthologous groups; and 10,451 (27.8%) were assigned to six Kyoto Encyclopedia of Genes and Genomes pathways. Differentially expressed genes were investigated using the reads per kilobase of the exon model per million reads methodology. A total of 546 unigenes showed significant differences in expression levels at different fruit developmental stages. Gene ontology categories associated with various aspects, including carbohydrate metabolic processes, transmembrane transport, and signal

transduction, were enriched with genes with divergent expressions. These *Pyrus pyrifolia* transcriptome data provide a rich resource for the discovery and identification of new genes. Furthermore, the numerous putative SSRs and SNPs detected in this study will be important resources for the future development of a linkage map or of marker-assisted breeding programs for the Asian pear.

Key words: *Pyrus pyrifolia*; Asian pear; *De novo* assembly; Differential gene expression; Single nucleotide polymorphism; Simple sequence repeat

INTRODUCTION

The pear (*Pyrus* sp) is widely cultivated in 88 countries of the world, according to calculations by the Food and Agriculture Organization (FAO), and its global annual yield continues to increase. In China in 2013, a 1,276,000-Ha area was harvested to yield 17,440,751 metric tons of pear representing, respectively, 69.2 and 72.2% of the global total that year (FAO, 2015). Plants such as the pear display significant inter-cultivar genetic and phenotypic variation (Wu et al., 2013). The Asian pear (*P. pyrifolia*) is a pear tree native to Asian (Wang et al., 2014). The edible fruit is known by many names, including sand (Wang et al., 2014), Chinese (Liu et al., 2009), Korean (Lee et al., 2013), and Japanese (Bai et al., 2013) pear. Due to self-incompatibility and interspecies compatibility, pear is highly heterozygous (Wu et al., 2013). There is no public Asian pear genome platforms reported to date (Wu et al., 2013; Chagne et al., 2014).

As for many other agronomically significant horticultural crops, the development of genomics resources for the Asian pear has lagged far behind those for model organisms (Liang et al., 2015). The transcriptomic sequences obtained in the current study will be invaluable in understanding the genetic makeup of the pear whole transcriptome. To facilitate this, we investigated the pear transcriptome using Illumina and 454 pyrosequencing, with the aim to create a genomic resource for gene and marker discovery. The development of next-generation (NGS) and RNA sequencing technologies provide new approaches to elucidate molecular mechanisms regulating fruit development (Wang et al., 2014). Understanding the genetic principle of growth and reproduction traits is helpful for pear production and has scientific significance for basic biology.

In the present study, we conducted large-scale transcriptome sequencing from Asian pear cultivar ‘Cuiguan’ fruit using RNA-Seq technology. We generated more than 3.7 billion nucleotides of high quality cDNA sequence by high throughput sequencing; an RNA-Seq project for *P. pyrifolia* was initiated (NCBI BioProject accession No. PRJNA227906). Reads were assembled into 37,649 well-annotated unigenes.

MATERIAL AND METHODS

Plant material

Experiments were performed using fruits from 24 Asian pear trees, aged 6 years. The pear cultivar ‘Cuiguan’ (*P. pyrifolia* Nakai) was grafted onto *P. betulaefolia* Bunge rootstocks

and grown in a flat-canopied pergola training system as described by Zhang et al. (2005). At least eight fruits were harvested at each of the early, middle, and late stages of expansion and maturation, and stored at -80°C until further use. Three biological replicates, each replicate contained four pear trees.

cDNA library construction and sequencing

Total RNA was extracted from each sample using RNA Plant Plus reagent according to the manufacturer protocol (Tiangen, Shanghai, China). Then, RNA was quantified using a Nano drop ND-2000 spectrophotometer (Wilmington, DE, USA); the RNA integrity number value was >8.5. The transcriptome library was constructed by mixing equal quantities of RNA from each fruit stage, including the mesocarp and core. Messenger RNA enrichment, fragment interruption, adapter addition, size selection, PCR amplification, and RNA-Seq were performed by staff at the Beijing Genomics Institute (BGI) in Shenzhen, China, using approved institute protocols.

Bioinformatic analysis

Raw reads obtained from HiSeq 2000 (Illumina Inc., San Diego, CA, USA) were filtered to exclude low-complexity reads and reads containing adaptor sequences. Only clean reads were used for further analysis. Sequences obtained using Illumina sequencing were deposited in the NCBI sequence read archive (accession No. SRR1269627). The resulting clean reads were assembled using Trinity (Grabherr et al., 2011) and were used to optimize the Trinity original assembly result by removing sequences that could not be extended on either end. Unigenes were aligned to the NCBI non-redundant (Nr) protein database using BLASTx (Altschul et al., 1990) and a cut-off E-value of 10^{-5}). Assembled unigenes were also annotated using Blast2GO (Conesa et al., 2005) with gene ontology (GO) (Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2008). After obtaining GO annotations for every unigene, the online tool WEGO (Wheeler et al., 2007) was used to classify GO functions for all unigenes, as well as to understand the distribution of gene functions at the macro level. Single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) detection methods used were as described by Chagne et al. (2012). The statistical significance of differential expression profile was determined for each gene using the method described by Audic and Claverie (1997).

RESULTS AND DISCUSSION

Transcriptome sequencing and assembly

Transcripts were constructed using a *de novo* assembly strategy. Overall, 127,085,054 reads gave rise to 4,729,035,340 bp total residues. *De novo* assemblies yielded 22,581,776 bp total residues and 37,649 total unigenes. The smallest unigene was 250 and the largest 6595 nucleotides (nt) in length; the average length distributed in the 200-1000 bp region (Figure 1). The overall GC content of the pear fruit transcriptome was 44.4% for assembled contigs. Data sets are available at the NCBI short read archive (accession No. SRR1269627).

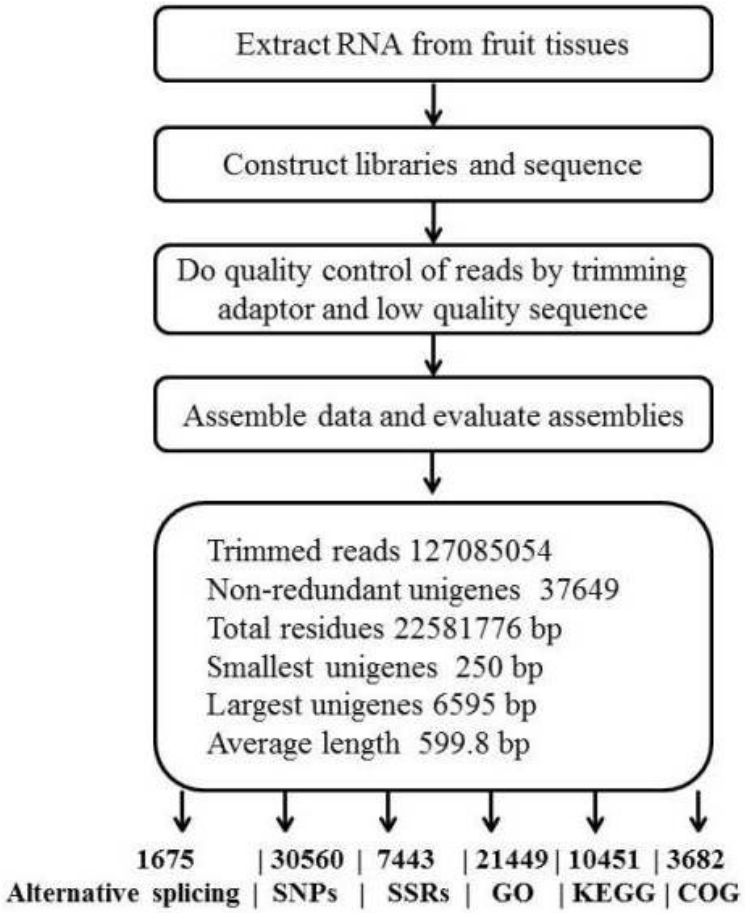

Figure 1. Schematic representation of data processing pipeline for the *de novo* assembly and annotation of Asian pear transcriptome.

Annotation of predicted proteins in the Nr database

Using BLASTx as described in the methodology, 28,550 genes (76.6% of total unigenes) were determined to be above the prescribed E-value cut-off. Among these, 27,597 matched to known genes, and 953 matched to unknown genes. A large portion of the unknown genes was made up of hypothetical proteins. The remaining 9099 (24.2%) unigenes resulted in no significant hits. This was attributed mainly to the short length of these unigenes and is observed particularly in studies involving large-scale sequencing (Venturini et al., 2013). In characterizations of apple and strawberry unigene sequences, approximately 40% of unique transcripts from the apple assembly (Park et al., 2006), and 43.9% for strawberry (Bombarely et al., 2010) were without significant similarity. In our study, the high percentage of pear transcripts with no significant similarity indicates an enormous potential for the discovery of new genes in this plant, besides providing the possibility of identification of new gene networks (Meyer et al., 2009).

Organism distribution based on BLASTx analysis showed that the unigenes hit a range of plant species (Figure 2). Among various plants with protein sequences in GenBank, pear unigenes had the highest degree of similarity with 30.5% hits with the grapevine (*V. vinifera*) genome. This was followed by castor (*Ricinus communi*, 18.2%), cottonwood (*Populus trichocarpa*, 16.3%), and soybean (*Glycine max*, 12.0%). Hits on pear (*P. pyrifolia*/*P. communis*) were only 0.9%, while hits to non-plant organisms made up 2.7%.

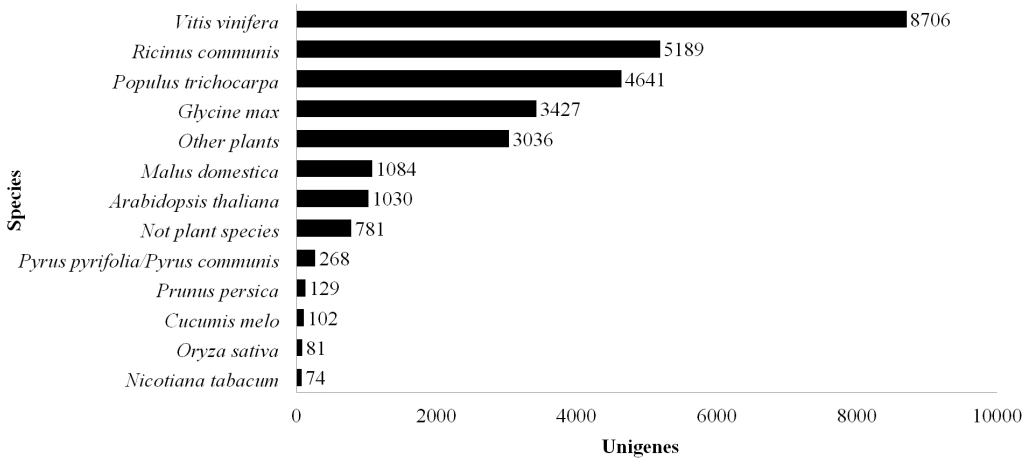

Figure 2. Species distribution of pear unigenes from *de novo* assembly.

Assessment of fruit transcriptome assembly

In order to test the accuracy of assembly, Asian pear sequences were aligned to the *P. bretschneideri* pear genome (Wu et al., 2013). The results show that 33,887 (90.0%) unigenes mapped to genome (Table 1), and the expected value reached 10^{-5} when 37,649 sequences were retrieved via BLASTn. In conclusion, the cover degree of assembled Asian pear sequences was reliable.

Table 1. Summary of mapping *Pyrus pyrifolia* unigenes to *Pyrus bretschneideri* genome.

Item	Number
Unigenes	37,649
Unigenes map to Genome	33,887
Isogroup ≥2	10,076
Isogroup ≥3	1,858
Isogroup ≥4	565
Isogroup ≥5	315

Protein coding sequence (CDS) prediction

In total, CDS were generated for 37,649 unigenes using the BLASTx protein database search. A total of 2664 genes containing a complete open reading frame (ORF) was obtained by analyzing splicing sequences. Some ORF genes were found to be involved in signal transduction and included E3 ubiquitin-ligase, S-phase-kinase-associated protein, and auxin influx carrier. Some,

such as sorbitol and sugar transporter, were involved in carbohydrate transmembrane transport activities. All sequences were uploaded in transcriptome shotgun assembly, NCBI (http://www.ncbi.nlm.nih.gov/Traces/wgs/fdump.cgi?GBED01,1,,GB~http://www.ncbi.nlm.nih.gov/Traces/wgs/fdump.cgi?GBED01,2661,,GB).

Gene ontology term classification

Gene ontology assignments were used to classify functions of the predicted Asian pear genes. Based on sequence homology, 21,449 sequences were categorized into 48 functional groups (Figure 3). In each of the three main categories of GO classification (biological processes, cellular components, and molecular function), cellular process, cell components, and binding terms, respectively, were predominant. As many as 15,959 transcripts were involved in biological processes, including cellular (13,240, 35.2%) and metabolic (12,305; 32.7%) processes, as well as signaling (1878, 5.0%). Moreover, 16,373 transcripts are subject to a cellular component and could be divided into cell component (16,118; 42.8%), organelle (11,705; 31.1%), and symplast (1286; 3.4%). Gene ontology analysis also demonstrated that 17,766 transcripts had potential molecular functions, including binding (12,651; 33.6%) and catalytic (11,073; 29.4%) and transporter (1596; 4.2%) activity.

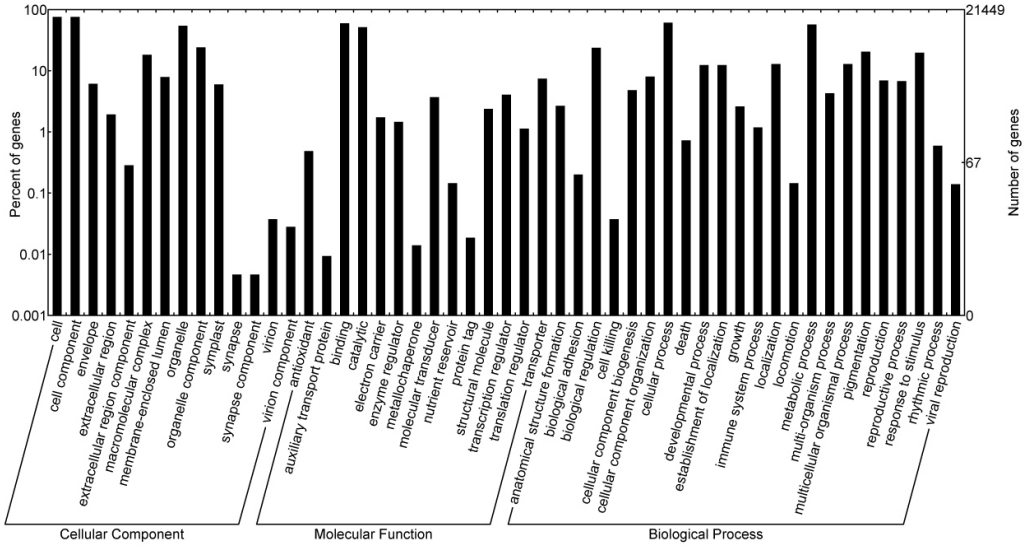

Figure 3. Histogram presentation of Gene Ontology classification. Results are summarized in three main categories: biological process, cellular component, and molecular function. The right y-axis indicates the number of genes in a category. The left y-axis indicates the percentage of a specific category of genes in the main category.

Results are summarized in three main categories: biological process, cellular component, and molecular function. The right y-axis indicates the number of genes in a category. The left y-axis indicates the percentage of a specific category of genes in the main category.

Gene pathway classification

Both gene annotation and pathway analyses were helpful for us to predict potential

genes and their functions at a whole-transcriptome level. In addition, we predicted overall 10,451 unigenes involved in 291 predicted KEGG metabolic pathways. Two major pathways (biosynthesis of secondary metabolites and oxidative phosphorylation) comprised over 1000 unigenes. The pathways with most representation were metabolic (1330 members; 12.7%) and biosynthesis of secondary metabolites (601 members; 5.8%) pathways (Table 2). These findings will provide a valuable resource for investigating specific processes, functions, and pathways for Asian pear research. Interestingly, 117 unigenes involved in plant hormone signal transduction were found, which contained nine pathways controlling the signal transduction of several plant growth regulators, such as auxin, cytokinin, gibberellin, abscisic acid, ethylene, brassinosteroid, jasmonic acid, and salicylic acid. These hormones regulate plant vegetative and reproductive growth, fruit ripening senescence, and stress responses. These annotations provide a valuable resource for investigating specific processes, functions, and pathways in fruit development research.

Table 2. Top 20 pathways with the highest unigene numbers.

ID	Pathway definition	Number
ko01100	Metabolic pathways	1330
ko01110	Biosynthesis of secondary metabolites	601
ko00190	Oxidative phosphorylation	133
ko04075	Plant hormone signal transduction	117
ko00010	Glycolysis / Gluconeogenesis	104
ko04120	Ubiquitin mediated proteolysis	99
ko00520	Amino sugar and nucleotide sugar metabolism	84
ko00500	Starch and sucrose metabolism	84
ko04146	Peroxisome	72
ko00620	Pyruvate metabolism	71
ko00710	Carbon fixation in photosynthetic organisms	68
ko00195	Photosynthesis	67
ko00360	Phenylalanine metabolism	64
ko00940	Phenylpropanoid biosynthesis	63
ko04110	Cell cycle	61
ko04810	Regulation of actin cytoskeleton	58
ko00900	Terpenoid backbone biosynthesis	55
ko04111	Cell cycle - yeast	52
ko00020	Citrate cycle (TCA cycle)	52
ko00051	Fructose and mannose metabolism	51

Clusters of orthologous groups (COG) function classification

In COG functional classification, 3682 unigenes could be annotated in 25 categories (Figure 4). Among them, the general function prediction was the largest group (729 members; 22.2%), followed by the functions of posttranslational modification, protein turnover, chaperones (369 members; 11.2%); transcription (342 members; 10.4%); and signal transduction mechanisms (293 members; 8.9%). Interestingly, we did not find any genes related to extracellular structures.

Molecular marker detection

Next-generation sequencing technologies have increased the opportunity for molecular marker development in non-model study organisms on an unprecedented scale (Torales et al., 2013). To date, only a few genetic maps have been developed for *Pyrus* spp (Montanari et al., 2013). In the current study, more contigs than isogroups were found because some contigs (also

called isocontigs) were attributed to the same isogroups as a result of alternative splicing events. There were 40.6% (1675 of 4121) isogroups having no less than two contigs in pear fruit tissues (Table 3). Alternative-spliced isogroups in the pear fruit had 2.46 unigenes.

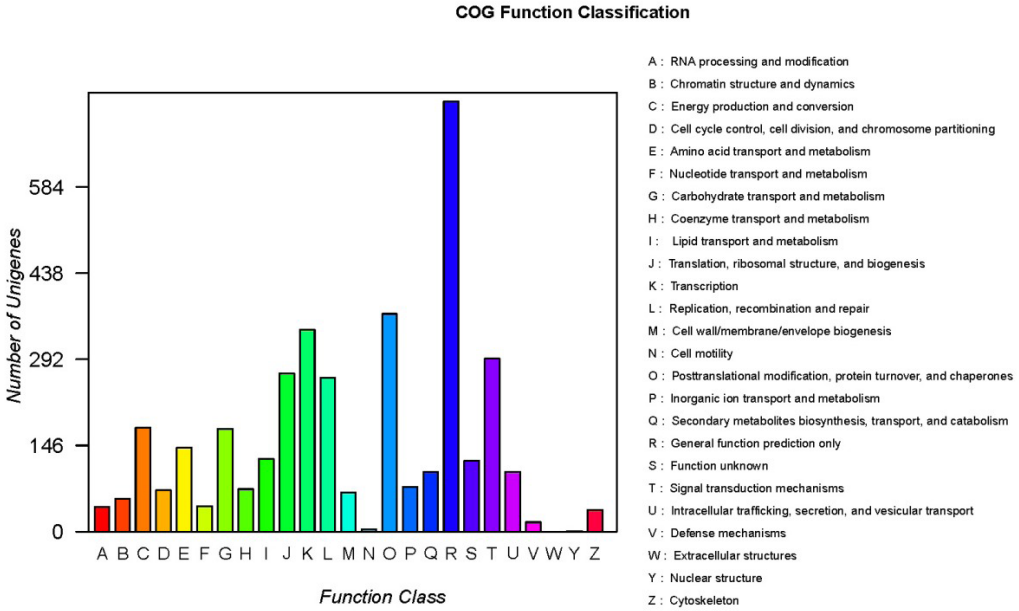

Figure 4. Histogram presentation of clusters of orthologous groups (COG) classification. 3682 sequences have a COG classification among the 25 categories.

Table 3. Variant transcripts by assemble analysis.

Number of contigs per isogroup	Number of isogroups (contigs)
2	1148 (2296)
3	367 (1101)
4	105 (420)
5	37 (185)
6	10 (60)
7	5 (35)
8	3 (34)
In total (≥2)	1675 (4121)

Plentiful expression SNPs (eSNPs) identified in this study will be valuable molecular markers for further research on the Asian pear (**Table S1**). By excluding those that could not specifically match the pear genome and had minor allele frequencies lower than 20%, we obtained a total of 30,560 SNPs, which comprised various substitutions of A-G (8442), C-T (8500), A-C (2405), A-T (2662), C-G (2399), and G-T (2369) (Table 4). The ratio of transitions (16,942) to transversions (9835) was approximately 1.72. Expression SNPs are useful for the identification of candidate genes or quantitative trait loci underlying quantitative traits, such as body composition (Montanari et al., 2013). The density of SNPs in *P. pyrifolia* was similar to that observed in the *P. trilocularis* transcriptome (Liang et al., 2015).

Table 4. Distribution of SNPs in the Asian pear transcriptome.

Type	Count	Frequency per kb
Transition		
C/T	8500	0.4
A/G	8442	0.4
Transversion		
A/T	2662	0.12
A/C	2405	0.11
T/G	2369	0.11
C/G	2399	0.11
Indel		
Insertion	1971	0.09
Deletion	1812	0.08

Simple sequence repeats, or microsatellites, are neutral molecular markers widely distributed in the genome. We obtained 7443 SSRs, of which 52.1% were mononucleotide repeats (3873), followed by 26.9% trinucleotide repeats (2005) and 18.4% dinucleotide repeats (1368), as well as 0.5% pentanucleotide repeats (5) (Table 5). In addition to (CT)n and (AG)n dinucleotide repeats, and (CCT)n and (AAG)n trinucleotide, (CT)n and (CTT)n also had high frequencies (**Table S2**), an observation in the pear that differs from that in the human genome (Lander et al., 2001). In the current study, there were six types of dinucleotide repeats; among them (CT)n, (AG)n, and (AT)n were predominant types and had frequencies of 47.6, 41.4, and 4.8%, respectively. Among the trinucleotide repeats, the frequencies of 20 SSR types seemed to vary moderately from 0.80 to 12.7%, and the most common repeats were (CTT)n (12.7%) and (AAG)n (11.8%). As many as 163 SSRs were present for tetranucleotide repeats, and five (13.5%) for (ATGT)n; 11.1% for (ATTT)n; 7.9% for (AAAT)n; 4.9% for (AAAG)n had frequencies greater than 4%. Among 35 pentanucleotide repeats types, (CTTTT)n (17.1%) and (CCTCT)n (11.4%) were the predominant types, followed by 8.6% for (ATTTT)n, and the rest less than 8%. It had been formerly proven that SSRs comprise 3% of the human genome, with the greatest contribution from dinucleotide repeats (0.5%) (Lander et al., 2001). In this study, we used NGS to detect SNPs in the pear transcriptome, to enable the design of a medium throughput SNP assay. Finally, 3864 SSR markers and 30,560 SNPs were identified in *P. pyrifolia*, which will be used in future research on the genetic diversity of *Pyrus* spp (Liang et al., 2015).

Table 5. Summary of microsatellite loci in pear transcriptome.

Type	Number of repeats	%
Mononucleotide	3873	52.04
Trinucleotide	2005	26.94
Dinucleotide	1368	18.38
Tetranucleotide	128	1.72
Pentanucleotide	35	0.47
Hexanucleotide	34	0.46

Differential expression analysis

A total of 546 unigenes were significantly changed at different fruit developmental stages (**Table S3**). Results of GO enrichment analysis are shown in Figure 5. We also performed metabolic pathways enrichment analysis, and identified the primary biochemical pathways and

signal transduction pathways in which DEGs were involved. A total of 70 DEGs were related to nine metabolic pathways, including genes involved in nitrogen compound metabolic process; catabolic, biosynthetic, secondary metabolic, pigment metabolic, macromolecule metabolic, cellular metabolic, and primary metabolic processes; and oxidation reduction.

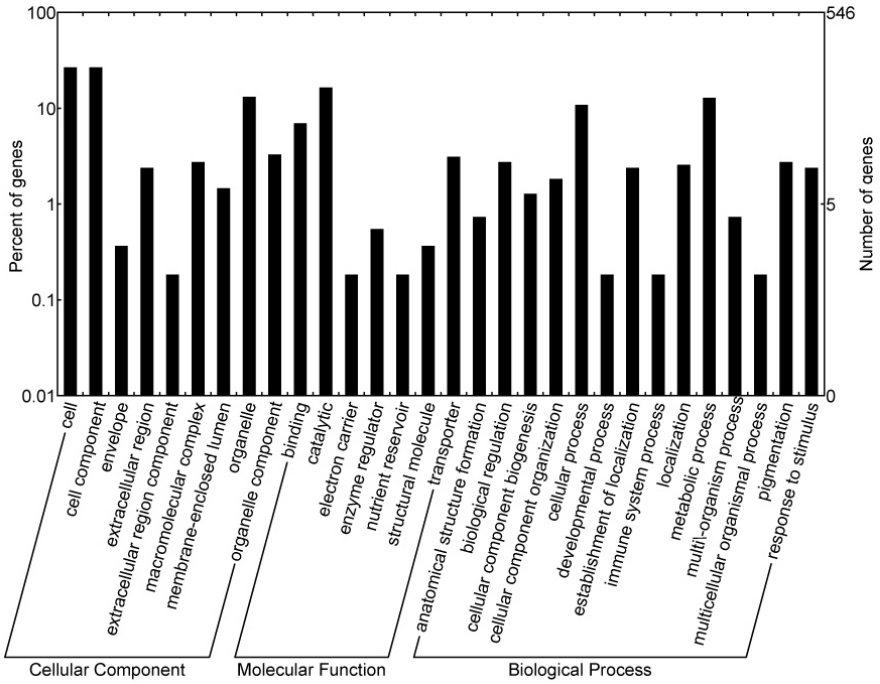

Figure 5. Functional categorization of the DEGs detected in Asian pear fruit. The functional categorization was performed using WEGO (Web Gene Ontology Annotation Plot).

CONCLUSION

This study generated a successful global analysis of the *P. pyrifolia* fruit transcriptome. Data will contribute greatly to genomics resources for *Pyrus* spp functional analysis and genetics. Finally, it will also potentially contribute to the development of population-based genome studies of the genera. Our study obtained a set of 37,649 transcripts or unigenes and demonstrated some important features of the Asia pear transcriptome, such as GC content, gene annotation, and pathways across the whole transcriptome. In addition, we identified reliable markers of 30,560 SNPs and 7443 SSRs. Along with various existing pear transcriptome databases (Wu et al., 2013; Chagne et al., 2014; Liang et al., 2015); these results can aid the investigation of pear fruit development. This work highlights the applicability of NGS for marker and gene discovery; we isolated thousands of genetic markers for parentage analysis, and identified several members of a specific gene family in *P. pyrifolia*. The unique transcripts derived from this work have greatly increased the number of pear fruit mRNA transcripts available in public databases. This information can be further utilized in gene expression, genomics, and other functional genomics studies in pear. This research will be helpful for understanding the genetic architecture of pear transcriptome,

and provides useful resource and markers for future functional genomic research. Since all our data are publicly accessible, we hope they may contribute to Asian pear research, and in any case will be available for reanalysis by the wider scientific community.

Conflicts of interest

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (Grant #31070245), Program for New Century Excellent Talents in University (NCET) from the Ministry of Education, China (#NCET-10-0582) and the SMC-Morningstar Young Scholars (#AE1500006).

Supplementary material

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25-29.

Audic S and Claverie JM (1997). The significance of digital gene expression profiles. *Genome Res.* 7: 986-995.

Bai S, Saito T, Sakamoto D, Ito A, et al. (2013). Transcriptome analysis of Japanese pear (*Pyrus pyrifolia* Nakai) flower buds transitioning through endodormancy. *Plant Cell Physiol.* 54: 1132-1151.

Bombarely A, Merchante C, Csukasi F, Cruz-Rus E, et al. (2010). Generation and analysis of ESTs from strawberry (*Fragaria xananassa*) fruits and evaluation of their utility in genetic and molecular studies. *BMC Genomics* 11: 503.

Chagne D, Crowhurst RN, Troggio M, Davey MW, et al. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS One* 7: e31745.

Chagne D, Crowhurst RN, Pindo M, Thrimawithana A, et al. (2014). The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS One* 9: e92644.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Food and Agricultural Organization of the United Nation FAO (2015). Available at [http://faostat.fao.org/site/342/default.aspx].

Grabherr MG, Haas BJ, Yassour M, Levin JZ, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644-652.

Kanehisa M, Araki M, Goto S, Hattori M, et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36: D480-484.

Lander ES, Linton LM, Birren B, Nusbaum C, et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lee HS, Isse T, Kawamoto T, Baik HW, et al. (2013). Effect of Korean pear (*Pyrus pyrifolia* cv. Shingo) juice on hangover severity following alcohol consumption. *Food Chem. Toxicol.* 58: 101-106.

Liang TT, Liu YY, Tong R, Xu XR, et al. (2015). Transcriptome sequencing and analysis of wild pear (*Pyrus trilocularis*) using the Illumina platform. *Vegetos* 28: 1-8.

Liu W, Liu J, Xie M, Liu C, et al. (2009). Characterization and high-pressure microfluidization-induced activation of polyphenoloxidase from Chinese pear (*Pyrus pyrifolia* Nakai). *J. Agric. Food Chem.* 57: 5376-5380.

Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, et al. (2009). Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics* 10: 219.

Montanari S, Saeed M, Knabel M, Kim Y, et al. (2013). Identification of *Pyrus* single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific *Pyrus* hybrids. *PLoS One* 8: e77022.

Park S, Sugimoto N, Larson MD, Beaudry R, et al. (2006). Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags. *Plant Physiol.* 141: 811-824.

Torales SL, Rivarola M, Pomponio MF, Gonzalez S, et al. (2013). *De novo* assembly and characterization of leaf transcriptome for the development of functional molecular markers of the extremophile multipurpose tree species *Prosopis alba*. *BMC Genomics* 14: 705.

Venturini L, Ferrarini A, Zenoni S, Tornielli GB, et al. (2013). *De novo* transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity. *BMC Genomics* 14: 41.

Wang YZ, Zhang S, Dai MS and Shi ZB (2014). Pigmentation in sand pear (*Pyrus pyrifolia*) fruit: biochemical characterization, gene discovery and expression analysis with exocarp pigmentation mutant. *Plant Mol. Biol.* 85: 123-134.

Wheeler DL, Barrett T, Benson DA, Bolton E, et al. (2007). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 35: D5-12.

Wu J, Wang Z, Shi Z, Zhang S, et al. (2013). The genome of the pear (*Pyrus bretschneideri* Rehd). *Genome Res.* 23: 396-408.

Zhang CX, Tanabe K, Tamura F and Itai A (2005). Spur characteristics, fruit growth, and carbon partitioning in two late-maturing Japanese pear (*Pyrus pyrifolia* Nakai) cultivars with contrasting fruit size. *J. Am. Soc. Hortic. Sci.* 130: 252-260.