Scientia Horticulturae 240 (2018) 397–404

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Comparative transcriptome analysis between interspecific hybridization (Huaren apricot ♀ × almond ♂) and intraspecific hybridization (Huaren apricot) during young fruit developmental stage

Lin Wang[1], Pan-Feng Liu[1], Han Zhao*, Gao-Pu Zhu*, Ta-Na Wuyun

Non-Timber Forestry Research and Development Center, Chinese Academy of Forestry, and China Paulownia Research and Development Center, Zhengzhou 450003, PR China

ARTICLE INFO

Keywords:
Armeniaca cathayana
Prunus dulcis
Hybrids
Comparative gene expression
Transcription factors

ABSTRACT

Huaren apricot and almond are two tree species with nuts as the main economically important products. Crossbreeding these two species might be a good approach for combining the traits of large seed kernel of almond with the wide adaptability of Huaren apricot. However, the molecular mechanism of hybridization incompatibility between Huaren apricot and almond has remained unclear until date. In the present study the transcriptomes of developing fruits after interspecific hybridization between Huaren apricot and almond and after intraspecific hybridization of Huaren apricot were compared to elucidate the possible molecular mechanisms of fruit development at the gene expression level. Approximately 562,535,240 clean reads were assembled, and 25,692 unigenes were identified. Totally, 270 differentially expressed genes (DEGs) were identified, of which 221 were up-regulated and 49 were down-regulated in almond as compared with Huaren apricot. Functional annotation analysis indicated that the expression of *cinnamyl alcohol dehydrogenase* and *EREBP* were significantly up-regulated, whereas *GA2ox* and *MADS* were down-regulated in the intraspecific hybridization (PN) as compared with the interspecific hybridization (MN). Moreover, these genes might play an important role in fruit development. This study provides novel insights into the molecular mechanisms of fruit development in interspecific and intraspecific hybrids.

1. Introduction

Interspecific hybridization is an important approach for obtaining and utilizing novel agronomic traits from related species for genetic improvement. It has played an important role in selectively creating new germplasm and types of cultivars, improving forest yield, quality, and resilience, and has resulted in important achievements in forestry (Chen et al., 2016; Tuyl, 1997; Wang et al., 2012, 2017; Zhao et al., 2016).

Huaren apricot (*Armeniaca cathayana* D. L), a new species of the genus *Armeniaca* Scop., was obtained from Zhangjiakou, Hebei Province, China (Fu et al., 2010, 2011). It is an economically and ecologically important forest species in arid and semi-arid areas in northeast, northwest, and north China (Xu et al., 2016). As compared with that in traditional fresh apricots, the flesh of Huaren apricot is thin and cracks when ripe; however, the nuts are full and sweet, suitable for food and food processing industries based on the kernel. Due to the smaller breeding population and long-term artificial breeding, a more uniform genetic background of Huaren apricot has formed, and the potential for intraspecific genetic improvement has become limited (Bao et al., 2017; Liu et al., 2015). Almond (*Prunus dulcis*) is one of the most important nut tree crops for its kernels (seeds) and the arts used are the same as those of Huaren apricot (International Nut and Dried Fruit Council, 2016; Yada et al., 2011). Because almond has larger seed kernels and thinner shells than those of Huaren apricot, the former provides a better means of genetic improvement of Huaren apricot by introducing the fine traits of almond through interspecific hybridization.

Because of reproductive isolation, several species are found to be incompatible during interspecific hybridization, owing to factors such as abnormal pollen tube growth, abnormal fertilization, and embryo abortion (Lattier et al., 2017; Tonosaki et al., 2013). The strength of this incompatibility is positively correlated with the degree of kinship (Li and Chen, 2003). Preliminary studies have shown that the incompatibility of almond and Huaren apricot mainly occurs during the embryo development stage after fertilization. However, the underlying reason

* Corresponding authors.
E-mail addresses: Zdh707@163.com (H. Zhao), zhugaopu@163.com (G.-P. Zhu).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.scienta.2018.06.038
Received 26 March 2018; Received in revised form 23 May 2018; Accepted 13 June 2018

and mechanism is unclear. Some studies on other species have shown that the main reason for this phenomenon is abnormal endosperm development, which causes inadequate supply of nutrients to the embryo (Friedman, 1998; Johnston and Hanneman, 1982) other researchers argued that abortion occurred owing to the imbalance of embryo and endosperm development (Tanaka and Watanabe, 1972). However, few studies on the molecular mechanism of interspecific hybrid abortion were reported. Because abortion is a gradual process, analysis of gene expression of hybrid fruit between interspecific and intraspecific hybrids will provide novel insights into mechanisms determining incompatibility in interspecific hybridization.

Next-generation sequencing offers unique opportunities for exploring molecular mechanisms in genetic analysis, comparative genomics, and functional genomics owing to its cost-effectiveness and high-throughput (Strickler et al., 2012). In the present study, we performed RNA sequencing (RNA-seq) to compare the transcriptomes of young Huaren apricot fruits obtained by hybridization with Huaren apricot or almond as the male parent, and monitored the early changes in gene expression level after interspecific hybridization and before abortion. The DEGs were identified, annotated, and analyzed. Functional categorization of DEGs was done to reveal various metabolic pathways involved in fruit development. This study provides a comprehensive overview of transcriptional regulation of fruit development in interspecific hybridization, thus providing a better understanding of molecular mechanisms of incompatibility between Huaren apricot and almond.

2. Materials and methods

2.1. Plant materials

The experiment was carried out in the Collection of Apricot Genetic Resources of Luoyang Academy of Agriculture, Luoyang, Henan Province, China. All the plants were grown in the same experimental field and under the same conditions. Female parent of hybrid combination was the Huaren apricot cultivar Zhongren1. Interspecific hybridization was performed using a mixture of pollen from almond clones Italian No. 1, Italian No. 2, Italian No. 3, and the cultivar Shuang ruan. Male parent of intraspecific hybridization was a mixture of the pollen of Huaren apricot clones 11D05 and 108, and the varieties Youyi and Bai Yu Bian. Three biological replicates were set up for each cross combination. Forty days after pollination, the fruits were collected and divided into normally growing and aborted groups according to their growth status. Normally growing fruits obtained by hybridization with Huaren apricot and almond as male parents were designated as PN and MN, respectively. The aborted fruits with Huaren apricot and almond as male parents were designated as PU and MU, respectively. The fruits were immediately frozen in liquid nitrogen and stored at −80 °C for RNA extraction.

2.2. RNA extraction, cDNA library construction and illumina sequencing

Total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. The total RNA quantity and purity were analyzed using Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, San Diego, CA, USA) with RIN number > 7.0. Approximately 10 μg total RNA representing a specific adipose type was subjected to isolation of poly (A) mRNA using poly-T oligo-attached magnetic beads (Invitrogen). Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. Thereafter, the cleaved RNA fragments were reverse-transcribed to prepare the final cDNA library in accordance with the protocol for mRNA-seq sample preparation kit (Illumina, San Diego, CA, USA). The average insert size for the paired-end libraries was 300 bp (± 50 bp). Subsequently, we performed the paired-end sequencing on an Illumina Hiseq 4000 at the (LC Sceiences, Houston, TX, USA), following the vendor's recommended protocol.

2.3. Sequence assembly, annotation, and analysis

Raw reads were pre-processed to remove low quality regions and adapter sequences. The Q20, Q30, GC content, and sequence duplication level of the clean data were calculated. All downstream analyses were based on high-quality clean data. The resulting high-quality cleaned reads were assembled *de novo* into contigs using Trinity with strand-specific option "-SS_lib_type" set to "F," "min_kmer_cov" set to 2 (Grabherr et al., 2011), and all other parameters set at default values. The assembled unigenes were aligned using BLASTx (E value $\leq 10^{-5}$) against the following databases: National Center for Biotechnology Information (NCBI) Non-Redundant database (NR), Pfam, Swiss-Prot, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and the Karyotic Orthologous Groups (KOG).

2.4. Identification and functional annotation of DEGs

Quantification of transcript expression was performed using the reads per kilobase per million reads (RPKM) method (Mortazavi et al., 2008). Differential expression analysis was performed using the Ballgown R package. Genes with both $|\log_2(\text{fold change})| \geq 1$ and $P < 0.05$ were considered differentially expressed in PN than in PU, PU than in PN, PN than in MN, and PU than in MU.

3. Results

3.1. Transcriptome sequencing, assembly, and identification of unigenes

RNA from PN, PU, MN, and MU were used for RNA-seq, with three replicates per fruit, respectively, which generated 577,821,402 raw reads. After removing low-quality reads and trimming adapter sequences, 562,535,240 remained (Table 1).

All high-quality clean reads were assembled into 25,692 unigenes with a minimum unigene size of 201 bp, a maximum size of 11,445 bp, an average length of 903 bp, and an N50 value of 1,298 bp. Among these genes, 17,264 (67.21%) were in the range of 200–1,000 bp, 8,428 (32.82%) had lengths of ≥1,000 bp, and 2,093 (8.15%) were longer than 2,000 bp. Length distribution of all the genes are shown in Fig. 1.

3.2. Functional annotation

Based on sequence similarity, 25,692 genes were aligned using BLASTx (E values $\leq 10^5$) searches against the Swiss-Port, NR, Pfam, GO, KOG protein, and KEGG pathway databases. All unigenes (100%) in at least one data base (Table 2).

Among the annotated genes in the NR database, 90.01% (23,125/25,692) had an E value $\leq 1.0\,e^{-5}$ and showed very strong homology to the gene sequence in the database.

The assembled genes were searched against the KOG database to estimate the gene function. In general, 19,189 unigenes were clustered into 25 KOG categories (Fig. 2). Among these categories, the largest category was "general function prediction only" (2,264, 8.81%), followed by "posttranslational modification, protein turnover, chaperones" (1,471, 5.73%) and "signal transduction mechanisms" (1,369, 5.33%). The smallest categories were "cell motility," with only ten unigenes (0.039%).

In addition, 12,436 (48.40%) genes annotated in the GO database were categorized into 57 GO terms, belonging to three main GO ontologies: biological processes (BP), cellular components (CC), and molecular functions (MF) (Fig. 3). "Transcription, DNA-dependent" (858 genes, 50.93%), "integral to membrane" (3,005 genes, 50.93%), and "ATP binding" (2,890 genes, 50.93%) were dominant among the functional groups for BP, CC, and MF, respectively.

We used the KEGG pathway database to search the functional