Genomics Data 6 (2015) 260–261

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

De novo transcriptome assembly of two different peach cultivars grown in Korea

Yeonhwa Jo [a,1], Hyosub Chu [a,1], Jin Kyong Cho [b], Hoseong Choi [a], Sen Lian [a], Won Kyong Cho [a,*]

[a] Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea
[b] Department of Fruit Tree, Korea National College of Agriculture and Fisheries, Jeonju, 560-500, Republic of Korea

ARTICLE INFO

Article history:
Received 16 October 2015
Accepted 19 October 2015
Available online 19 October 2015

Keywords:
Cultivar
Peach
RNA-Seq
Transcriptome

ABSTRACT

Peach (*Prunus persica*) is one of the most popular stone fruits worldwide. Next generation sequencing (NGS) has facilitated genome and transcriptome analyses of several stone fruit trees. In this study, we conducted *de novo* transcriptome analyses of two peach cultivars grown in Korea. Leaves of two cultivars, referred to as Jangtaek and Mibaek, were harvested and used for library preparation. The two prepared libraries were paired-end sequenced by the HiSeq2000 system. We obtained 8.14 GB and 9.62 GB sequence data from Jangtaek and Mibaek (NCBI accession numbers: SRS1056585 and SRS1056587), respectively. The Trinity program was used to assemble two transcriptomes *de novo*, resulting in 110,477 (Jangtaek) and 136,196 (Mibaek) transcripts. TransDecoder identified possible coding regions in assembled transcripts. The identified proteins were subjected to BLASTP search against NCBI's non-redundant database for functional annotation. This study provides transcriptome data for two peach cultivars, which might be useful for genetic marker development and comparative transcriptome analyses.

Specifications	
Organism/cell line/tissue	Peach (*Prunus persica*)/leaves
Sex	N.A.
Sequencer or array type	HiSeq2000
Data format	Raw and processed
Experimental factors	Transcriptome profiling of two different peach cultivars
Experimental features	Leaves of two different peach cultivars, Jangtaek and Mibaek, were harvested for total RNA extraction. Prepared libraries were paired-end sequenced using the HiSeq 2000 system. The obtained data was subjected to *de novo* transcriptome assembly using Trinity, and coding regions were predicted by TransDecoder. We performed BLASTP against the NCBI non-redundant (nr) dataset to annotate identified proteins.
Consent	N/A
Sample source location	Hoengseong, South Korea (37°28′49.6″N 127°58′34.3″E)

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRS1056585 for Peach cultivar Jangtaek.

http://www.ncbi.nlm.nih.gov/sra/SRS1056587 for Peach cultivar Mibaek.

2. Experimental design, materials, and methods

2.1. Plant materials

Peach (*Prunus persica*) is one of the most popular stone fruit trees worldwide and a member of the genus *Prunus*. Peach is a model plant for many *Prunus* species, and the genome of peach was recently reported [1]. In addition, several genetic markers for peach have been developed [2,3]. For genetic marker development and comparative transcriptome analyses, we performed transcriptome analyses of two peach cultivars. We selected two peach cultivars, referred to as Jangtaek and Mibaek, which are commercially important cultivars in Korea. Two peach cultivars were grown in an orchard located in Kadam-ri, Hoengseong-up, Korea. Five leaves were harvested from a single tree of each species and immediately frozen in liquid nitrogen for further experiments.

2.2. RNA isolation, library preparation, and sequencing

Pooled leaves were used for total RNAs extraction using Fruit-mate for RNA Purification (Takara, Shiga, Japan) and the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). For mRNA library preparation, we used

* Corresponding author.
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gdata.2015.10.014

Table 1
Summary of *de novo* assembled two peach transcriptomes.

Index	Jangtaek	Mibaek
Total trinity transcripts	110,477	136,196
Total trinity components	68,656	69,391
Percent GC	42.17	41.54
Contig N50	2051	1957
Median contig length	785	980
Average contig	1234.86	1267.29
Total assembled bases	136,423,350	172,599,271

a TruSeq RNA Library Prep Kit v2 according to manufacturer's instructions (Illumina, San Diego, U.S.A.). In brief, the poly-A containing mRNAs were isolated using poly-T oligo-attached magnetic beads. The first strand cDNA followed by second strand cDNA were synthesized from purified mRNAs. End repair was performed followed by adenylation of 3′ ends. Adapters were ligated and PCR was conducted to selectively enrich DNA fragments with adapters and to amplify the amount of DNA in the library, respectively. The quality control of generated libraries was conducted using the 2100 Bioanalyzer (Agilent, Santa Clara, U.S.A.). The libraries were paired-end sequenced by Macrogen Co. (Seoul, South Korea) using HiSeq 2000 platform.

2.3. De novo transcriptome assembly, identification protein coding regions, and annotation

We obtained a total of 8.14 GB and 9.62 GB sequence data from Jangtaek and Mibaek, respectively. *De novo* transcriptome assembly was performed using Trinity, which uses the de Bruijn graphs algorithm [4]. Detailed information of assembled transcriptome is summarized in Table 1. The numbers of total transcripts for Jangtaek and Mibaek were 110,477 and 136,196, respectively. N50 values for Jangtaek and Mibaek were 2051 and 1957. Next, we identified possible protein coding regions within the assembled transcripts using the TransDecoder program implemented in the Trinity software distribution. We identified 72,337 and 107,557 proteins from Jangtaek and Mibaek, respectively. The identified proteins were blasted against the NCBI non-redundant (nr) protein database. Except a few proteins that might be novel, most proteins were matched to known proteins. In case of Jangtaek, only 4.8% of proteins were not homologous to known proteins in two different databases. In summary, this study provides transcriptome data for two peach cultivars, which might be useful for genetic marker development and comparative transcriptome analyses.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ00976401)" Rural Development Administration, Republic of Korea.

References

[1] I. Verde, A.G. Abbott, S. Scalabrin, S. Jung, S. Shu, F. Marroni, T. Zhebentyayeva, M.T. Dettori, J. Grimwood, F. Cattonaro, The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat. Genet. 45 (2013) 487–494.
[2] B. Sosinski, M. Gannavarapu, L. Hager, L. Beck, G.J. King, C. Ryder, S. Rajapakse, W. Baird, R. Ballard, A. Abbott, Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. Theor. Appl. Genet. 101 (2000) 421–428.
[3] T. Yamamoto, K. Mochida, T. Imai, Y. Shi, I. Ogiwara, T. Hayashi, Microsatellite markers in peach [*Prunus persica* (L.) Batsch] derived from an enriched genomic and cDNA libraries. Mol. Ecol. Notes 2 (2002) 298–301.
[4] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (2011) 644–652.