Genomics Data 6 (2015) 262–263

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 *Prunus salicina* cultivars

Yeonhwa Jo [a,1], Sen Lian [a,1], Jin Kyong Cho [b], Hoseong Choi [a], Hyosub Chu [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
Plum
*Prunus salicina*
RNA-Seq
Transcriptome

ABSTRACT

Plum is a globally grown stone fruit and can be divided into several species. In particular, the *Prunus salicina*, which is native to China, is widely grown in many fruit orchards in Korea and Japan, as well as the United States and Australia. The transcriptome data for *Prunus salicina* has not been reported to our knowledge. In this study, we performed *de novo* transcriptome assembly for two selected *P. salicina* cultivars referred to as Akihime and Formosa (commercially important plum cultivars in Korea) using next generation sequencing. We obtained a total of 9.04 GB and 8.68 GB raw data from Akihime and Formosa, respectively. *De novo* transcriptome assembly using Trinity revealed 155,169 and 160,186 transcripts for Akihime and Formosa. Next, we identified 121,278 and 116,544 proteins from Akihime and Formosa using TransDecoder. We performed BLASTP against the NCBI non-redundant (nr) dataset to annotate proteins. Taken together, this is the first transcriptome data for *P. salicina* to our knowledge.



| Specifications | |
|---|---|
| Organism/cell line/tissue | Plum (*Prunus salicina*)/leaves |
| Sex | N.A. |
| Sequencer or array type | HiSeq2000 |
| Data format | Raw and processed |
| Experimental factors | Transcriptome profiling of two different Japanese plum cultivars |
| Experimental features | Leaves of two different Chinese cultivars referred to as Formosa and Akihime were harvested for total RNA extraction. Prepared libraries were paired-end sequenced by 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/SRX1187015 for *Prunus salicina* cultivar Akihime.

http://www.ncbi.nlm.nih.gov/sra/SRX1186990 for *Prunus salicina* cultivar Formosa.

## 2. Introduction

Plum is a globally grown stone fruit and can be divided into several species. Of known cultivated plum species, the *Prunus salicina* species, known as Japanese plum or Chinese plum, and *Prunus domestica* species, known as European plum, are commercially grown. In particular, the *P. salicina*, which is native to China, is widely grown in many fruit orchards in Korea and Japan, as well as the United States and Australia. Transcriptome data for the *P. domestica* species has been recently released [1]; however, the genome and transcriptome data for *P. salicina* have not been reported to our knowledge. In addition, available genetic markers for *P. salicina* are limited [2]. In this study, we performed *de novo* transcriptome analyses for two selected *P. salicina* cultivars referred to as Akihime and Formosa, which are commercially important plum cultivars in Korea, using next generation sequencing.

## 3. Experimental design, materials, and methods

### 3.1. Plant materials

Two plum cultivars were grown in an orchard located in Kadam-ri, Hoengseong-up, South Korea (Hoengseong, South Korea (37°28′49.6″N 127°58′34.3″E)). Five leaves from a single tree were harvested and immediately frozen in liquid nitrogen for further experiments.

* Corresponding author.
*E-mail address:* wonkyong@gmail.com (W.K. Cho).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gdata.2015.10.015

**Table 1**
Summary of *de novo* assembled two *Prunus salicina* transcriptomes.

| Index | Akihime | Formosa |
|---|---|---|
| Total trinity transcripts | 155,169 | 160,186 |
| Total trinity components | 71,901 | 74,663 |
| Percent GC | 41.75 | 41.80 |
| Contig N50 | 2,080 | 1973 |
| Median contig length | 995 | 906 |
| Average contig | 1,334.67 | 1,256.18 |
| Total assembled bases | 207,099,385 | 201,221,789 |

### *3.2. RNA isolation, library preparation, and sequencing*

Five leaves from a single tree were pooled and 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 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 a second strand cDNA was 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 the HiSeq 2000 platform.

### *3.3. De novo transcriptome assembly, identification protein coding regions, and annotation*

We obtained a total of 9.04 GB and 8.68 GB raw data from Akihime and Formosa, respectively. *De novo* transcriptome assembly was performed using Trinity, which uses the de Bruijn graphs algorithm [3]. Detailed information of assembled transcriptome was summarized in Table 1. The numbers of total transcripts for Akihime and Formosa were 155,169 and 160,186, respectively, and N50 values for Akihime and Formosa were 2051 and 1957, respectively. Next, we identified candidate coding regions within the assembled transcripts using the TransDecoder program implemented in the Trinity software distribution. We identified 121,278 and 116,544 proteins from Akihime and Formosa, respectively. To annotate proteins, we performed BLASTP against the NCBI non-redundant (nr) dataset. Taken together, this is the first transcriptome data for *P. salicina* to our knowledge.

## Conflict of interest

The authors declare that they have no competing interests.

## Acknowledgments

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

## References

[1] B. Rodamilans, D. San León, L. Mühlberger, T. Candresse, M. Neumüller, J.C. Oliveros, J.A. García, Transcriptomic analysis of *Prunus domestica* undergoing hypersensitive response to *Plum pox virus* infection. PLoS One 9 (2014) e100477.
[2] M. Mnejja, J. Garcia-Mas, W. Howad, M. Badenes, P. Arús, Simple-sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* lindl.) are highly polymorphic and transferable to peach and almond. Mol. Ecol. Notes 4 (2004) 163–166.
[3] 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.