Genomics Data 6 (2015) 271–272

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

De novo transcriptome assembly of a sour cherry cultivar, Schattenmorelle

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 18 October 2015
Available online 19 October 2015

Keywords:
Cherry
Cultivar
RNA-Seq
Transcriptome

ABSTRACT

Sour cherry (*Prunus cerasus*) in the genus *Prunus* in the family *Rosaceae* is one of the most popular stone fruit trees worldwide. Of known sour cherry cultivars, the Schattenmorelle is a famous old sour cherry with a high amount of fruit production. The Schattenmorelle was selected before 1650 and described in the 1800s. This cultivar was named after gardens of the Chateau de Moreille in which the cultivar was initially found. In order to identify new genes and to develop genetic markers for sour cherry, we performed a transcriptome analysis of a sour cherry. We selected the cultivar Schattenmorelle, which is among commercially important cultivars in Europe and North America. We obtained 2.05 GB raw data from the Schattenmorelle (NCBI accession number: SRX1187170). *De novo* transcriptome assembly using Trinity identified 61,053 transcripts in which N50 was 611 bp. Next, we identified 25,585 protein coding sequences using TransDecoder. The identified proteins were blasted against NCBI's non-redundant database for annotation. Based on blast search, we taxonomically classified the obtained sequences. As a result, we provide the transcriptome of sour cherry cultivar Schattenmorelle using next generation sequencing.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications	
Organism/cell line/tissue	Sour cherry (*Prunus cerasus* subsp. acida)/leaves
Sex	N.A.
Sequencer or array type	HiSeq2000
Data format	Raw and processed
Experimental factors	Transcriptome profiling of a sour cherry cultivar Schattenmorelle
Experimental features	Leaves of a sour cherry cultivar Schattenmorelle were harvested for total RNA extraction. Prepared libraries were paired-end sequenced by HiSeq 2000 system. The obtained data was subjected for *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/SRX1187170 for Sour cherry cultivar Schattenmorelle.

2. Introduction

Sour cherry (*Prunus cerasus*) in the genus *Prunus* in the family *Rosaceae* is one of the most popular stone fruit trees worldwide [1]. Of known sour cherry cultivars, the Schattenmorelle is a famous old sour cherry with a high amount of fruit production. The Schattenmorelle was selected before 1650 and described in the 1800s. This cultivar was named after gardens of the Chateau de Moreille in which the cultivar was initially found. In order to identify new genes and to develop genetic markers for sour cherry, we performed transcriptome analysis of a sour cherry. We selected the cultivar Schattenmorelle, which is among commercially important cultivars in Europe and North America.

3. Experimental design, materials and methods

3.1. Plant materials

The sour cherry cultivar Schattenmorelle was grown in an orchard located in Kadam-ri, Hoengseong-up, South Korea (37°28′49.6″N

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

http://dx.doi.org/10.1016/j.gdata.2015.10.013
2213-5960/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1
Summary of *de novo* assembled a sour cherry cultivar Schattenmorelle.

Index	Schattenmorelle
Total trinity transcripts	61,053
Total trinity components	51,176
Percent GC	42.98
Contig N50	611
Median contig length	340
Average contig	506.61
Total assembled bases	30,929,971

127°58′34.3″E). Five leaves from a single tree were harvested and immediately frozen in liquid nitrogen for further experiments.

3.2. RNA isolation, library preparation, and sequencing

Leaves 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 the 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 were 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 2.05 GB raw data from the Schattenmorelle. *De novo* transcriptome assembly was performed using Trinity, which uses the de Bruijn graphs algorithm [2]. Detailed information of assembled transcriptome was summarized in Table 1. The number of total transcripts was 61,053. N50 was 611. Next, we identified candidate coding regions within the assembled transcripts using the TransDecoder program implemented in the Trinity software distribution. We identified 25,585 protein coding sequences. Based on blast search, we taxonomically classified the obtained sequences using the MEGAN program [3]. Most sequences (97.27%, 24,889 sequences) were derived from eukaryota, followed by viruses (28 sequences), not assigned (28 sequences), and bacteria (three sequences). We identified 607 proteins that are not homologous to known protein sequences.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgment

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] F. Blando, C. Gerardi, I. Nicoletti, Sour cherry (*Prunus cerasus* L) anthocyanins as ingredients for functional foods. Biomed. Res. Int. 2004 (2004) 253–258.
[2] 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.
[3] D.H. Huson, A.F. Auch, J. Qi, S.C. Schuster, MEGAN analysis of metagenomic data. Genome Res. 17 (2007) 377–386.