Genomics Data 6 (2015) 273–274

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 mume* 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 18 October 2015
Available online 19 October 2015

Keywords:
Cultivar
Prunus mume
RNA-Seq
Transcriptome

ABSTRACT

Prunus mume, belonging to the *Prunus* genus, is an Asian tree, and its common names are Chinese plum and Japanese plum. *P. mume* are cultivated for fruit production as well as ornamental purposes. In this study, we conducted *de novo* transcriptome assembly for two selected *P. mume* cultivars referred to as Takada and Wallyoung (commercially important cultivars for fruit production and ornamental trees, respectively) by RNA-sequencing. We obtained 9.14 GB and 9.48 GB sequence data from Takada and Wallyoung (NCBI accession numbers: SRX1187101 and SRX1187169), respectively. *De novo* transcriptome assembly identified 130,989 and 116,941 transcripts for Takada and Wallyoung, respectively. In addition, we identified 96,681 and 91,429 proteins from Takada and Wallyoung, respectively, by TransDecoder program. We performed BLASTP against the NCBI non-redundant (nr) datasets to annotate identified proteins. This study provides transcriptomes and proteomes for two different *P. mume* cultivars, which might be useful for comparative transcriptome analyses and assist development of genetic markers.

© 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	*Prunus mume*/leaves
Sex	N/A
Sequencer or array type	HiSeq2000
Data format	Raw and processed
Experimental factors	Transcriptome profiling of two different *Prunus mume* cultivars
Experimental features	Leaves of two different *Prunus mume* cultivars, Takada and Wallyoung, 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	National Institute of Horticultural and Herbal Science in Suwon, South Korea (37°18′35.0″N 126°58′43.4″E)

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRX1187101 for *Prunus mume* cultivar Takada.

http://www.ncbi.nlm.nih.gov/sra/SRX1187169 for *Prunus mume* cultivar Wallyoung.

2. Introduction

Prunus mume belonging to the *Prunus* genus is an Asian tree, and its common names are Chinese plum and Japanese plum. The origin of *P. mume* is known as the south of mainland China around the Yangtze River. Currently, *P. mume* is cultivated in Japan, Korea, Taiwan, and Vietnam. *P. mume* are cultivated for fruit production as well as ornamental purposes. Recently, next generation sequencing (NGS) has facilitated genome and transcriptome analyses of diverse plant species. The genome of *P. mume* has been reported [1], and many genetic markers have been recently generated [2,3]. In this study, we conducted *de novo* transcriptome assembly for two selected *P. mume* cultivars, referred to as Takada and Wallyoung (commercially important cultivars for fruit production and ornamental trees, respectively), by RNA-sequencing.

3. Experimental design, materials and methods

3.1. Plant materials

Two *P. mume* cultivars, Takada and Wallyoung, were harvested in the National Institute of Horticultural and Herbal Science in Suwon, South Korea (37°18′35.0″N 126°58′43.4″E). Five leaves from a single

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

http://dx.doi.org/10.1016/j.gdata.2015.10.011
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 two *Prunus mume* transcriptomes.

Index	Takada	Wallyoung
Total trinity transcripts	130,989	116,941
Total trinity components	64,777	62,142
Percent GC	42.32	42.32
Contig N50	1864	2044
Median contig length	963	984
Average contig	1229.60	1320.65
Total assembled bases	161,063,714	154,438,141

tree were harvested and immediately frozen in liquid nitrogen for further experiments.

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 the TruSeq RNA Library Prep Kit v2 according to manufacturer's instruction (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.

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

We obtained 9.14 GB and 9.48 GB raw data from Takada and Wallyoung, respectively. *De novo* transcriptome assembly was performed using Trinity, which uses the de Bruijn graphs algorithm. Detailed information of assembled transcriptome is summarized in Table 1. The numbers of total transcripts for Takada and Wallyoung were 130,989 and 116,941, respectively, while the numbers of unigenes for Takada and Wallyoung were 64,777 and 62,142, respectively. N50 values for Takada and Wallyoung were 2027 and 2155 bp, respectively. Next, we identified candidate coding regions within the assembled transcripts using the TransDecoder program implemented in the Trinity software distribution [4]. We identified 96,681 and 91,429 proteins from Takada and Wallyoung, respectively. To annotate proteins, we performed BLASTP against the NCBI non-redundant (nr) datasets. In summary, this study provides transcriptomes and proteomes for two different *P. mume* cultivars, which might be useful for comparative transcriptome analyses and development of genetic markers.

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] Q. Zhang, W. Chen, L. Sun, F. Zhao, B. Huang, W. Yang, Y. Tao, J. Wang, Z. Yuan, G. Fan, Z. Xing, C. Han, H. Pan, X. Zhong, W. Shi, X. Liang, D. Du, F. Sun, Z. Xu, R. Hao, T. Lv, Y. Lv, Z. Zheng, M. Sun, L. Luo, M. Cai, Y. Gao, J. Wang, Y. Yin, X. Xu, T. Cheng, J. Wang, The genome of *Prunus mume*. Nat. Commun. 3 (2012) 1318.
[2] L. Sun, W. Yang, Q. Zhang, T. Cheng, H. Pan, Z. Xu, J. Zhang, C. Chen, Genome-wide characterization and linkage mapping of simple sequence repeats in mei (*Prunus mume* Sieb. et Zucc.). PLoS One 8 (2013), e59562.
[3] L. Sun, Q. Zhang, Z. Xu, W. Yang, Y. Guo, J. Lu, H. Pan, T. Cheng, M. Cai, Genome-wide DNA polymorphisms in two cultivars of mei (*Prunus mume* Sieb. et Zucc.). BMC Genet. 14 (2013) 98.
[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.