



Data in Brief

De novo transcriptome assembly of *Setaria italica* variety Taejin

Yeonhwa Jo ^{a,1}, Sen Lian ^{b,1}, Jin Kyong Cho ^c, Hoseong Choi ^a, Sang-Min Kim ^d, Sun-Lim Kim ^d,
Bong Choon Lee ^d, Won Kyong Cho ^{a,c,*}

^a Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea

^b College of Crop Protection and Agronomy, Qingdao Agricultural University, Qingdao, Shandong 266109, China

^c The Taejin Genome Institute, Gadam-gil 61, Hwangseong 25239, Republic of Korea

^d Crop Foundation Division, National Institute of Crop Science, RDA, Wanju 55365, Republic of Korea

ARTICLE INFO

Article history:

Received 28 April 2016

Received in revised form 2 May 2016

Accepted 3 May 2016

Available online 5 May 2016

Keywords:

Foxtail millet

RNA-Seq

Setaria italica

Transcriptome

Variety

ABSTRACT

Foxtail millet (*Setaria italica*) belonging to the family Poaceae is an important millet that is widely cultivated in East Asia. Of the cultivated millets, the foxtail millet has the longest history and is one of the main food crops in South India and China. Moreover, foxtail millet is a model plant system for biofuel generation utilizing the C₄ photosynthetic pathway. In this study, we carried out *de novo* transcriptome assembly for the foxtail millet variety Taejin collected from Korea using next-generation sequencing. We obtained a total of 8.676 GB raw data by paired-end sequencing. The raw data in this study can be available in NCBI SRA database with accession number of SRR3406552. The Trinity program was used to *de novo* assemble 145,332 transcripts. Using the TransDecoder program, we predicted 82,925 putative proteins. BLASTP was performed against the Swiss-Prot protein sequence database to annotate the functions of identified proteins, resulting in 20,555 potentially novel proteins. Taken together, this study provides transcriptome data for the foxtail millet variety Taejin by RNA-Seq.

© 2016 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	Foxtail millet (<i>Setaria italica</i> variety Taejin)/leaves
Sex	N.A.
Sequencer or array type	HiSeq2000
Data format	Raw and processed
Experimental factors	<i>De novo</i> transcriptome assembly of foxtail millet variety Taejin
Experimental features	Leaves of five foxtail millet plants (variety Taejin) were harvested for total RNA extraction. A prepared library was paired-end sequenced using the HiSeq 2000 system. The obtained data were subjected to <i>de novo</i> transcriptome assembly using Trinity, and coding regions were predicted by TransDecoder. Finally, BLASTP was performed against the Swiss-Prot protein database to annotate the 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/SRR3406552> for *Setaria italica* variety Taejin.

2. Introduction

Foxtail millet (*Setaria italica*) belonging to the family Poaceae is an important millet that is widely cultivated in East Asia. Of the cultivated millets, the foxtail millet has the longest history and is one of the main food crops in South India and China [1]. In addition, foxtail millet is a small diploid crop plant utilizing the C₄ photosynthetic pathway. Therefore, foxtail millet is an ideal model plant system for biofuel generation along with switchgrass and pearl millet [2]. The genome of foxtail millet consists of nine chromosomes encoding approximately 38,000 genes [3]. With an available draft genome, foxtail millet is an ideal experimental crop in many research areas, such as comparative genome evolution among members of the Poaceae family, development of bioenergy crops, and identification of genes associated with C₄ photosynthesis. In this study, we carried out *de novo* transcriptome assembly for foxtail millet variety Taejin collected from Korea by next-generation sequencing.

* Corresponding author.

¹ These authors contributed equally to this work.

3. Experimental design, materials, and methods

3.1. Plant materials

Plants for foxtail millet variety Taejin were grown in a field located in Gadam-ri, Hoengseong-up, South Korea. Leaves from five such plants were harvested and immediately frozen in liquid nitrogen for further experiments.

3.2. RNA isolation, library preparation, and sequencing

Ten leaves collected from five plants were pooled and used for total RNA extraction using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). For mRNA library preparation, we used a TruSeq RNA Library Prep Kit v2 according to the 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 of cDNA, followed by a second strand of 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 a 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 of protein coding regions, and annotation

We paired-end sequenced a library generated from foxtail millet variety Taejin using the Illumina HiSeq 2000 system, resulting in 8.676 GB of raw data. *De novo* transcriptome assembly for foxtail millet variety Taejin was performed using Trinity program ver. 2.0.6 [4]. We summarized the detailed information on the *de novo* transcriptome assembly for foxtail millet variety Taejin in Table 1. The numbers of total transcripts and components for foxtail millet variety Taejin were 145,332 and 81,613, respectively. The N50 value was 1853 bp, and the median contig length was 771 bp. We further predicted coding regions within the assembled transcripts by the TransDecoder program implemented in the Trinity software distribution. As a result, we predicted a total of 82,925 proteins. BLASTP was performed against the Swiss-Prot protein sequence database using the predicted 82,925 protein sequences to annotate the functions of identified proteins. The BLASTP results showed

Table 1

Summary of *de novo* assembled transcriptomes for foxtail millet variety Taejin.

Index	Taejin
Total trinity transcripts	145,332
Total trinity components	81,613
Percent GC	46.84
Contig N50	1853
Median contig length	771
Average contig	1138.59
Total assembled bases	165,473,745

that 20,555 proteins were novel without any hits to known protein sequences, and most proteins (55,617) were not assigned to any known organism. Of the assigned proteins, most proteins were derived from eukaryotes (5163 proteins) followed by bacteria (278 proteins) and viruses (71 proteins). To our knowledge, this is the first transcriptome data for foxtail millet collected from Korea. The transcriptome data for foxtail millet variety Taejin might be useful to develop genetic markers for the molecular breeding of foxtail millet in near future.

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 Agriculture Science & Technology Development (Project No. PJ01186102)” conducted by the Rural Development Administration, Republic of Korea. This work is dedicated to the memory of my father, Tae Jin Cho (1946–2015).

References

- [1] Y. Li, S. Wu, Traditional maintenance and multiplication of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces in China. *Euphytica* 87 (1996) 33–38.
- [2] A.N. Doust, E.A. Kellogg, K.M. Devos, J.L. Bennetzen, Foxtail millet: a sequence-driven grass model system. *Plant Physiol.* 149 (2009) 137–141.
- [3] G. Zhang, X. Liu, Z. Quan, S. Cheng, X. Xu, S. Pan, M. Xie, P. Zeng, Z. Yue, W. Wang, Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30 (2012) 549–554.
- [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.