Genomics Data 10 (2016) 33–34

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Transcriptome profiling of *Curcuma longa* L. cv. Suvarna

Ambika Sahoo, Basudeba Kar, Suprava Sahoo, Asit Ray, Sanghamitra Nayak *

Centre for Biotechnology, Siksha O Anusandhan University, Bhubaneswar 751003, India

ARTICLE INFO

Article history:
Received 31 August 2016
Accepted 7 September 2016
Available online 9 September 2016

Keywords:
Curcuma longa
Suvarna
RNA-seq
Transcriptome

ABSTRACT

Turmeric is an economically valued crop, because of its utility in the food, pharmaceutical industries and Ayurvedic medicine, attracts the attention in many areas of research work. In the present study, we executed resequencing through transcriptome assembly of the turmeric cultivar Suvarna (CL_Suv_10). Resequencing of Suvarna variety has generated 5 Gbases raw data with 75 bp paired-end sequence. The raw data has been submitted to SRA database of NCBI with accession number SRR4042181. Reads were assembled using Cufflinks-2.2.1 tool which ended up with 42994 numbers of transcripts. The length of transcripts ranged from 83 to15565, with a N50 value 1216 and median transcript length 773. The transcripts were annotated through number of databases. For the first time transcriptome profiling of cultivar Suvarna has been done, which could help towards identification of single nucleotide polymorphisms (SNPs) between Suvarna and other turmeric cultivars for its authentic identification.

Specification	
Organism/cell line/tissue	Turmeric (*Curcuma longa* cv. Suvarna) mature rhizome
Sex	NA
Sequencer or array type	Illumina Nextseq 500
Data format	Raw data in fastaq
Experimental factors	Resequencing of *Curcuma longa* cv. Suvarna through transcriptome profiling.
Experimental features	Fresh and healthy mature rhizome of *Curcuma longa* cv. Suvarna were taken, transcriptome assembly by re-sequencing (75 bp paired end) and gene annotations has been done.
Consent	N/A
Sample source location	Centre for Biotechnology, Siksha O Anusandhan University, Bhubaneswar, Odisha

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA339387

2. Introduction

Turmeric (*Curcuma longa* L.) of family Zingiberaceae is industrially as well as pharmaceutically important crop extensively cultivated in India. It is a vegetatively propagated, polyploid crop commonly known as "Golden spice". Because of its multipurpose use as food preservative, spice, therapeutic agent and natural dye, it is acquiring more importance in cosmetics, food and pharmaceutical industries [1,2]. In turmeric number of cultivars has been developed with different characteristics like high curcumin content, high essential oil content, rhizome yield and disease tolerance etc. but their asexual mode of reproduction and poor genetic makeup has creates confusion towards authentic identification. Single nucleotide polymorphism (SNP) study could be useful to solve these problems by comparing sequence of each variety with each other. In this current study, we conducted reference based transcriptome assembly of Suvarna, a well-known turmeric cultivar which could latter on used to get potential SNPs by comparing with other varieties towards proper identification.

3. Experimental design, materials and methods

3.1. Plant material

Fresh mature rhizomes of Suvarna variety has been collected from Centre for Biotechnology, Siksha O Anusandhan University, Bhubaneswar, India. These rhizomes were washed properly with distilled water and suspended in RNA later solution for further analysis.

3.2. RNA extraction and transcriptome sequencing

RNA extraction and library preparation has been done by using Illumina TruSeq RNA library tool kit. Sequencing was done using Illumina Nextseq500 platform.

* Corresponding author.

http://dx.doi.org/10.1016/j.gdata.2016.09.001

Table 1
Summary of assembly statistics.

Sample name	*Curcuma longa* cv. Kedaram
Fastq file size	5GB
Total transcripts generated	42994
Maximum transcript length (bp)	15565
Median transcript length (bp)	773
Total transcripts ≥500 bp	29515
Total transcripts ≥200 bp	42588
Total transcripts >1 kb	13762
N50 value	1216

3.3. Transcriptome assembly and annotation

75 bp paired-end sequencing has been done using Illumina Nextseq500 sequencing platform which developed5GB of raw data. Raw reads were mapped with the reference sequence using Tophat-2.0.13[3] tool. Cufflinks-2.2.1 [4] tool was used for transcriptome assembly. We have precisely given detail information about the reference based transcriptome assembly of Suvarna cultivar (Table 1). The transcriptome assembly of Suvarna developed 42994 numbers of transcripts, with median transcript length 773 and N50 value 1216. We annotated the transcripts using various databases like GO, KEGG, KOG, PlantCyc etc. As per our understanding transcriptome analysis of cultivar Suvarna has been done for the first time and this result could be efficiently used for development of markers such as SSRs and SNPs for identification of Suvarna from its closely similar turmeric cultivars.

Acknowledgement

We are thankful to our President Prof. Manoj Ranjan Nayak for support and providing lab facility.

References

[1] M.T. Selvan, K.G. Thomas, K. Manojkumar, Ginger (ZingiberofficinaleRosc.), Indian Spices-Production and Utilization. Coconut Development Board, India, 2002 110–131.
[2] B. Sasikumar, Genetic resources of Curcuma: diversity, characterization and utilization. Plant Genetic Resources: Characterization and Utilization, 3, 2005, pp. 230–251.
[3] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25 (9) (2009) 1105–1111.
[4] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nat. Biotechnol. 28 (5) (2010) 511.