Genomics Data 9 (2016) 160–161

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Resequencing of *Curcuma longa* L. cv. Kedaram through transcriptome profiling reveals various novel transcripts

Ambika Sahoo, Sudipta Jena, Suprava Sahoo, Sanghamitra Nayak, Basudeba Kar *

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

ARTICLE INFO

Article history:
Received 8 August 2016
Accepted 11 August 2016
Available online 17 August 2016

Keywords:
Curcuma longa
Kedaram
RNA-seq
Next seq 500
Transcriptome

ABSTRACT

Curcuma longa L. (Turmeric), of the family Zingiberaceae, is one of the economically as well as medicinally important plant species. It is a sterile, polyploid and vegetatively propagated spice crop cultivated usually in Southeast Asia. In the current study, we carried out re-sequencing through transcriptome profiling of *Curcuma longa* cv. Kedaram (Cl_Ked_6). We acquired a total of 1 GB raw data by resequencing through paired-end sequencing using Nextseq 500 platform. The raw data obtained in this study can be accessible in NCBI database with accession number of SRR3928562 with bioproject accession number PRJNA324755. Cufflinks-2.2.1 tool was used for transcriptome assembly which resulted in 39,554 numbers of transcripts. The transcript length ranged from 76 to 15,568, having N50 value of 1221 and median transcript length of 860. We annotated the transcripts using multiple databases. This data will be beneficial for studying sequence variations particularly SNPs between cultivars of turmeric towards authentic identification and discovery of novel functional transcripts in Kedaram.

Specification	
Organism/tissue	Turmeric (*Curcuma longa* cv. Kedaram) mature rhizome
Sex	NA
Sequencer	Illumina Nextseq 500
Data format	Raw data
Experimental factors	Transcriptome profiling through resequencing of elite turmeric cultivar cv. Kedaram (Cl_Ked_6).
Experimental features	Fresh and healthy rhizome of *Curcuma longa* cv. Kedaram (Cl_Ked_6), were harvested for RNA extraction, transcriptome profiling by re-sequencing (75 bp paired end) and annotations with different databaaes.
Consent	N/A
Sample source location	Centre of Biotechnology, Siksha O Anusandhan University, Bhubaneswar, Odisha

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA324755

2. Introduction

Curcuma longa L. (belonging to the family Zingiberaceae) is a vegetatively propagated, polyploid crop cultivated mostly in Southeast Asia, commonly known as “Indian saffron” and “Golden spice”. Turmeric is widely used as a spice, natural food dye and preservative in Asian countries [1]. India is the highest producer, exporter and consumer of turmeric [2]. Turmeric has multifunctional use in aromatherapy, cosmetics, medicine, dye and also in food industry [2]. Several cultivars in turmeric has been released having different phytochemical properties like high essential oil yield, high curcumin content, and rhizome yield but their genetic analysis has yet to be done. In this study, we accomplished reference based transcriptome assembly of Kedaram, an elite cultivar of turmeric collected from IISR, Kerala for its high yielding and high curcumin content through next generation sequencing [3].

3. Experimental design, materials and methods

3.1. Plant material

Fresh rhizomes of cultivar Kedaram was collected from Centre of Biotechnology, Siksha O Anusandhan University, Bhubaneswar, India. Fresh rhizome was harvested, washed properly with distilled water followed by ice cold 95% ethanol. After washing rhizome was chopped and suspended in RNA later solution and shifted to −80 °C till further analysis.

3.2. RNA extraction and transcriptome sequencing

RNA extraction and transcriptome library preparation were accomplished using Illumina TruSeq RNA library protocol summarized in

* Corresponding author at: Center for Biotechnology, School of Pharmaceutical Sciences (SPS), Siksha ‘O’ Anusandhan University, Kalinga Nagar, Ghatikia, Bhubaneswar 751003, Odisha, India.

http://dx.doi.org/10.1016/j.gdata.2016.08.010

Table 1
Summary of primary analysis and assembly statistics.

Sample name	*Curcuma longa* cv. Kedaram
Fastq file size	1,054,782 kb
Total transcripts generated	39,554
Maximum transcript length (bp)	15,568
Median transcript length (bp)	860
Total transcripts ≥500 bp	26,216
Total transcripts ≥200 bp	38,840
Total transcripts >1 kb	12,332
GC percent	46%
N50 value	1221

"TruSeq RNA Sample Preparation Guide". Illumina Nextseq 500 platform was used for resequencing.

3.3. Transcriptome assembly and annotation

We have done resequencing having 75 bp paired-end sequencing using Illumina Nextseq 500 platform. A total of 1 GB of raw data was resulted. Mapping of raw reads with the reference sequence (previous de novo transcriptome sequence of turmeric) was done using Tophat-2.0.13 [4] tool. Transcriptome assembly was performed using **Cufflinks-2.2.1** [5] tool. We summarized detailed information on the reference based transcriptome assembly of an elite turmeric cultivar Kedaramin (Table 1). The transcriptome assembly produced 39,554 transcripts, having N50 value of 1221 and Median transcript length of 860. We annotated the transcripts using multiple databases like Viridiplantae kingdom protein sequences (from UniProt Protein Database), Eukaryotic Orthologous Groups of proteins (KOG), PlantCyc enzyme sequences (from PMN Database), Kyoto Encyclopedia of Genes and Genomes (KEGG) database. To our understanding, this is the first transcriptome data for an elite turmeric cultivar Kedaram collected from IISR, Kerala. The transcriptome data of Kedaram can be effectively used for developing genetic markers like SSRs and SNPs which could be used for identification of Kedaram from its closely related cultivars.

Acknowledgement

We are thankful for the financial assistance from the DST, Government of India (SERB/SR/SO/PS/138/2012), New Delhi. We are also thankful to our President Prof. Manoj Ranjan Nayak for his support and for providing the lab facility.

References

[1] B. Sasikumar, Genetic resources of curcuma: diversity, characterization and utilization. Plant Genet. Resour. Charact. Util. 3 (2005) 230–251.
[2] M.T. Selvan, K.G. Thomas, K. Manojkumar, Ginger (*Zingiber officinale* Rosc.). Indian Spices-Production and Utilization, Coconut Development Board, India 2002, pp. 110–131.
[3] B. Sasikumar, K.J. George, K.V. Saji, T.J. Zachariah, Two new high yielding, high curcumin, turmeric (*Curcuma longa* L.) varieties-'IISR Kedaram' and 'IISR Alleppey Supreme'. J. Spices Aromat. Crops 14 (1) (2014).
[4] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25 (9) (2009) 1105–1111.
[5] 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.