



Data in Brief

De Novo transcriptome assembly of *Zingiber officinale* cv. Suruchi of Odisha



Mahendra Gaur, Aradhana Das, Rajesh Kumar Sahoo, Basudeba Kar, Sanghamitra Nayak, Enketeswara Subudhi *

Center of Biotechnology, Siksha 'O' Anushandhan University, Bhubaneswar, Odisha, India

ARTICLE INFO

Article history:

Received 9 June 2016

Accepted 18 June 2016

Available online 20 June 2016

Keywords:

Zingiber officinale

Suruchi

Illumina

Transcriptome

ABSTRACT

Zingiber officinale Rosc., known as ginger, is an Asian crop, popularly used in every household kitchen and commercially used in bakery, beverage, food and pharmaceutical industries. The present study deals with *de novo* transcriptome assembly of an elite ginger cultivar Suruchi by next generation sequencing methodology. From the analysis 10.9 GB raw data was obtained which can be available in NCBI accession number SAMN03761185. We identified 41,969 transcripts using *Trinity RNA-Seq* from ginger rhizome of Suruchi variety from Odisha. The transcript length varied from 300 bp to 8404 bp with a total length of 3,96,40,526 bp and N50 of 1251 bp. To the best of our knowledge, this is the first transcriptome data of an elite ginger cultivar Suruchi released for Odisha state of India which will help molecular biologists to develop genetic markers for identification of cultivars.

© 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	Ginger (<i>Zingiber officinale</i> cv. Suruchi) rhizome
Sex	N/A
Sequencer or array type	Illumina Nextseq 500
Data format	Raw data
Experimental factors	Transcriptome profiling of elite ginger cv. Suruchi
Experimental features	Fresh and healthy rhizome of <i>Zingiber officinale</i> cv. Suruchi, grown in High Altitude Research Station, Koraput, Odisha were harvested for RNA isolation, <i>de novo</i> transcriptome assembly and protein annotations.
Consent	N/A
Sample source location	High Altitude Research Station of Orissa University of Agriculture & Technology, Pottangi-764,039, Koraput (Dist), Orissa

1. Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/biosample/?term=SAMN03761185> for ginger cultivar Suruchi.

* Corresponding author.

E-mail address: enketeswarasubudhi@soauniversity.ac.in (E. Subudhi).

2. Introduction

Ginger (*Zingiber officinale*, Rose), is an herb, belongs to family Zingiberaceae. This crop has been spread across the globe for its multi-tude use, although it is native to South East Asia. India is the second largest producer of ginger and many elite cultivars have been released basing on phenotypic and phytochemical characteristics and initiations are taken to understand them at the genetic level by developing suitable markers and screen out disease tolerant genes and genotypes [1]. The information on the use of suitable marker and genes responsible different useful traits are limited. In the present study we conducted *de novo* transcriptome assembly for one of the most elite ginger cultivar Suruchi of Odisha for its low fiber and high dry recovery using next generation sequencing [2].

3. Experimental design, materials and methods

3.1. Plant materials

Fresh, healthy rhizome of *Zingiber officinale*, Rose. cv. Suruchi, grown in High Altitude Research Station, Koraput, Odisha were harvested from underground soil, rinsed thoroughly with sterile distilled water, immediately dipped into RNA stabilizer solution (Xcelris Genomics, India) and stored in liquid nitrogen until further experimentation.

3.2. RNA isolation, library preparation and sequencing

RNA isolation and transcriptome library construction was performed according to the Illumina TruSeq RNA library protocol and sequencing was done using Illumina Nextseq 500 at Genotypic Technologies Genomics facility, Genotypic Technology (P) Limited, Bangalore.

3.3. Transcriptome de novo assembly, annotation and classification

Raw data of size 10.9 GB was obtained from ginger variety Suruchi. *De novo* assembly of Illumina Nextseq 500 processed data was performed using trinityrnaseq [3] for k-mers = 25 has been selected for downstream analysis. Detail statistics of transcriptome *de novo* assembly is presented in Table 1. The number of total generated transcripts (≥ 300 bp) was 41,969 with a median transcript length of 599 bp and N50 value of 1251 for the Suruchi ginger cultivar. Transcripts were annotated using NCBI BLAST v2.2.29 [4] with the proteins *viridiplantae* against UniProt database. For annotation we have considered transcripts having length ≥ 300 bp, followed by clustering these transcripts with 95% indent using CD-HIT [5] which resulted into COG's. Unannotated transcripts were considered for Pfam domain analysis. We obtained 29,893 proteins of which only 23,416 are annotated. So far our best knowledge goes, this is the first transcriptome data for Suruchi ginger variety derived from Odisha, India, which can be utilized for development of suitable genetic markers for identification of elite cultivars.

Conflict of interest

The authors declare that they have no competing interests.

Transparency document

The [Transparency document](#) associated with this article can be found, in online version.

Table 1

Summary of de novo assembled cv. Suruchi transcriptome.

Features	cv. Suruchi
Total trinity transcripts generated	41,969
Maximum transcript length (bp)	8404
Median transcript length (bp)	599
Total transcripts ≥ 500 bp	37,791
Total transcripts > 1 Kb	20,305
Average transcript length (bp)	944.5 \pm 676.7
Total transcripts length	3,96,40,526
GC percent	45.85
N50 value	1251

Acknowledgement

The encouragement and support by Siksha 'O' Anusandhan University, Bhubaneswar, to carry out the present work is highly acknowledged.

References

- [1] E. Subudhi, A. Das, R.K. Joshi, S. Mohanty, S. Nayak, Genetic diversity analysis and redundant identification in 48 core collections of *Zingiber officinale* Rosc. (Zingiberaceae). *Braz. J. Bot.* (2016) 1–15.
- [2] D.C. Mohanty, B.S. Panda, Genetic resources of ginger. in: K.L. Chadha, P. Rethinam (Eds.), *Adv. Hortic. Vol. 9 - Plant. Spice Crop. Part 1*, vol. 9, Malhotra Publishing House, New Delhi 1994, pp. 151–167.
- [3] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29 (2011) 644–652.
- [4] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* 215 (1990) 403–410.
- [5] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22 (2006) 1658–1659.