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

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/).

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.

1. Direct link to deposited data


* Corresponding author.

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

<table>
<thead>
<tr>
<th>Features</th>
<th>cv. Suruchi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total trinity transcripts generated</td>
<td>41,969</td>
</tr>
<tr>
<td>Maximum transcript length (bp)</td>
<td>8404</td>
</tr>
<tr>
<td>Median transcript length (bp)</td>
<td>599</td>
</tr>
<tr>
<td>Total transcripts ≥500 bp</td>
<td>37,791</td>
</tr>
<tr>
<td>Total transcripts &gt;1 Kb</td>
<td>20,305</td>
</tr>
<tr>
<td>Average transcript length (bp)</td>
<td>944.5 ± 676.7</td>
</tr>
<tr>
<td>Total transcripts length</td>
<td>3,964,526</td>
</tr>
<tr>
<td>GC percent</td>
<td>45.85</td>
</tr>
<tr>
<td>N50 value</td>
<td>1251</td>
</tr>
</tbody>
</table>

Acknowledgement

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

References