Genomics Data 9 (2016) 42–43

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Comparative transcriptome analysis of ginger variety Suprabha from two different agro-climatic zones of Odisha

Mahendra Gaur, Aradhana Das, Rajesh Kumar Sahoo, Sujata Mohanty, Raj Kumar Joshi, Enketeswara Subudhi *

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

ARTICLE INFO

Article history:
Received 22 June 2016
Accepted 23 June 2016
Available online 24 June 2016

Keywords:
Zingiber officinale
Suprabha
Agro-climatic condition
Illumina
Transcriptome

ABSTRACT

Ginger (*Zingiber officinale* Rosc.), a well-known member of family Zingiberaceae, is bestowed with number of medicinal properties which is because of the secondary metabolites, essential oil and oleoresin, it contains in its rhizome. The drug yielding potential is known to depend on agro-climatic conditions prevailing at the place cultivation. Present study deals with comparative transcriptome analysis of two sample of elite ginger variety Suprabha collected from two different agro-climatic zones of Odisha. Transcriptome assembly for both the samples was done using next generation sequencing methodology. The raw data of size 10.8 and 11.8 GB obtained from analysis of two rhizomes S1Z4 and S2Z5 collected from Bhubaneswar and Koraput and are available in NCBI accession number SAMN03761169 and SAMN03761176 respectively. We identified 60,452 and 54,748 transcripts using trinity tool respectively from ginger rhizome of S1Z4 and S2Z5. The transcript length varied from 300 bp to 15,213 bp and 8988 bp and N50 value of 1415 bp and 1334 bp respectively for S1Z4 and S2Z5. To the best of our knowledge, this is the first comparative transcriptome analysis of elite ginger cultivars Suprabha from two different agro-climatic conditions of Odisha, India which will help to understand the effect of agro-climatic conditions on differential expression of secondary metabolites.

Specification:	
Organism/cell line/tissue	Ginger (*Zingiber officinale* cv. Suprabha) rhizome
Sex	N/A
Sequencer or array type	Illumina Nextseq 500
Data format	Raw data
Experimental factors	Transcriptome profiling of elite ginger cv. Suprabha from two different agro climatic zones of Odisha.
Experimental features	Fresh and healthy rhizomes of *Zingiber officinal* Rose. cv. Suprabha, grown in two different agro climatic zones of Odisha were harvested for RNA isolation, *de novo* transcriptome assembly and protein annotations.
Consent	N/A
Sample source location	S1Z4: Center of Biotechnology, Siksha 'O' Anusandhan University, Kalinga Nagar, Ghatikia, Bhubaneswar, Odisha. S2Z5: High Altitude Research Station of Orissa University of Agriculture & Technology, Pottangi-764039, Koraput (Dist), Orissa.

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

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/biosample/?term=SAMN03761176 for Ginger cultivar S1Z4.
http://www.ncbi.nlm.nih.gov/biosample/?term=SAMN03761169 for Ginger cultivar S2Z5.

2. Introduction

Ginger (*Zingiber officinale*, Rose.), a pan-tropical plant of South East Asian origin, belongs to family Zingiberaceae. Even though, the crop is able to grow in different climatic conditions, essential oil and oleoresin synthesized in its rhizomes are reported to vary with climate and soil type of the area of cultivation [1]. Agro-climatic conditions at different localities are known to vary greatly across a state like Odisha, Eastern India and thus classified into ten different zones. Agro-climatic conditions are known to influence the production of secondary metabolites in ginger rhizome when same cultivar is grown in two different locations [2]. Therefore, in the present study we conducted *de novo* transcriptome assembly for two ginger cv. Suprabha rhizome samples S1Z4 and S2Z5 collected respectively from two different locations of

http://dx.doi.org/10.1016/j.gdata.2016.06.014

the state; (i) Bhubaneswar of agro-climatic zone 4 (Climate: Hot and humid and Soil type: Saline, lateritic alluvial, red, mixed Red and Black); (ii) Koraput belonging to agro-climatic zone 5 (Climate: Hot and moist sub humid and Soil type: Brown forest, lateritic alluvial, red, mixed Red and Black) using next generation sequencing.

3. Experimental design, materials and methods

3.1. Plant materials

Fresh, healthy rhizome of *Zingiber officinale*, Rose. (cv. Suprabha) sample S1Z4 and S2Z5 were harvested from the underground soil of the High Altitude Research Station, Koraput and medicinal plant garden of Center of Biotechnology, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha. Rhizomes are rinsed thoroughly with sterile distilled water, immediately dipped into RNA stabilizer solution (Xcelris Genomics, India) and stored in liquid nitrogen until further experiments.

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 Technology's Genomics facility, Genotypic Technology (P) Limited Bangalore.

3.3. Transcriptome de novo assembly, annotation and classification

Raw data of size 10.8 GB and 11.8 GB approximately was obtained from both the ginger variety S1Z4 and S2Z5. *De novo* assembly of Illumina Nextseq 500 processed data was performed using trinityrnaseq [3] for k-mers = 25 has been selected for downstream analysis. Detailed statistics of transcriptome *de novo* assembly are presented in Table 1. The number of total generated transcripts (≥300 bp) was 60,452 and 54,748 with a median transcript length of 393 bp and 1164 bp and N50 value of 1415 and 1334 respectively for ginger cultivar S1Z4 and S2Z5. Transcripts were annotated using NCBI BLAST 2.2.29 [4] with the proteins viridiplantae taken from 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 54,322 and 48,483 proteins of which only 38,243 and 36,678 are annotated for sample S1Z4 and S2Z5 respectively. The first comparative transcriptome analysis of elite ginger cultivars S1Z4 and S2Z5 from two different agro-climatic conditions of Odisha, India will help to understand the effect of agro-climatic conditions on differential expression of secondary metabolites in addition to genetic marker development.

Table 1
Summary of *de novo* assembled cv. Suprabha transcriptome.

Features	S1Z4	S2Z5
Total trinity transcripts generated	60,452	54,748
Maximum transcript length (bp)	15,213	8988
Median transcript length (bp)	393	1164
Average transcript Length (bp)	1009.9 ± 830.6	986.7 ± 716.5
Total transcripts Length (bp)	6,10,51,850	5,40,19,323
Total transcripts ≥ 500 bp	40,026	37,791
Total transcripts > 1 Kb	21,743	20,305
Total transcripts > 5 Kb	232	36
N50	1415	1334
GC (%)	45.06	44.81

Note: S1Z4-Sample 1 of cv. Suprabha harvested from zone 4; S2Z5-Sample 2 of cv. Suprabha harvested from zone 5.

Acknowledgements

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] V.S. Govindarajan, D.W. Connell, Ginger-chemistry, technology, and quality evaluation: part 2. CRC Crit. Rev. Food Sci. Nutr. 17 (1983) 189–258.
[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.