Data Article

Data on small cardamom transcriptome associated with capsule rot disease

K. Mary Mathew a, Renjanan Reshma a, M. Geethu a, Varghese Rithin a, K.K. Sabu b, *, F. Nadiya b, Muhammad Ali Noushad b, Soumya S. Dharan b, Y.S. Rao a, A.B. Remashree a

a Indian Cardamom Research Institute (Spices Board Min Commerce & Industry, Government of India), Myladumpara, Idukki, Kerala 685553, India
b Jawaharlal Nehru Tropical Botanic Garden and Research Institute, Palode, Thiruvananthapuram 695562, India

Article history:
Received 6 August 2019
Received in revised form 26 September 2019
Accepted 27 September 2019
Available online 8 October 2019

Keywords:
Small cardamom
RNA sequencing
Transcriptome
Differential expression

Small cardamom (Elettaria cardamomum (L.) Maton, also known as the ‘Queen of Spices’) is a rhizomatous herbaceous monocot from the family Zingiberaceae. In the present study, using HiSeq™ 2000 RNA sequencing technology, transcriptome sequencing was performed for both control and disease stressed small cardamom leaf tissues. RNA-seq generated 46,931,637 (101 base) and 31,682,496 (101 base) raw reads and totally 9.93GB and 6.63GB of sequence data for cardamom control and stressed samples respectively. The raw data were submitted to NCBI SRA database of under the accession numbers SRX2512359 and SRX2512358 for the control and diseased samples respectively. The raw reads were quality filtered and assembled using TRINITY de novo assembler which created 1,11,495 (control) and 91,096 (diseased) contigs with N50 values 3013 (control) and 2729 (stressed). The data was further used to identify significantly differentially expressed unigenes between control and stressed samples. Assembled unigenes were further annotated and evaluated in silico to predict the function using publicly available databases and gene annotation tools.

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Data

Data shared in this article includes RNA-seq generated paired end strand specific 46,931,637 (101 base) and 31,682,496 (101 base) raw reads and totally 9.93GB and 6.63GB of sequence data for cardamom control and stressed samples respectively.

2. Experimental design, materials, and methods

2.1. Plant material

Leaf tissues from both sets, i.e., naturally infected capsule rot and non-infected control plants were collected followed by immediate freezing in liquid nitrogen. Ten biological replicates were pooled from leaf tissues under these two conditions [3].

2.2. Total RNA isolation and transcriptome sequencing

RNA extraction was done using a modified protocol of RNeasy Plant Mini Kit (Qiagen) and CTAB method [4]. RNA integrity and quality analysis was done using 2100 BioAnalyzer (Agilent Technologies). Illumina sequencing was performed using the HiSeq™ 2000 platform as per the manufacturer’s instructions (Illumina, San Diego, CA). RNA-seq generated paired end strand specific 46,931,637 (101 base) and 31,682,496 (101 base) raw reads and totally 9.93GB and 6.63GB of sequence data for cardamom control and stressed samples respectively.

2.3. De novo transcriptome assembly and functional annotation

The raw reads were pre-processed to remove adapter sequences, low quality bases, tRNAs and rRNAs. De novo transcriptome assembly was performed with TRINITY program [5] to generate the
assembled contigs. The assembler created 1,11,495 and 91,096 contigs for control and stressed cardamom samples (Table 1). The assembled unigenes were used for further downstream analysis such as annotation to publicly available databases, Gene Ontology (GO) enrichment and finally validation of differentially expressed genes using qPCR. Additionally, the reads from both pairs were combined and assembled together to generate a reference transcriptome (1,62,589 contigs, 310.7 MB). The information provided by the current study might be useful in developing molecular markers, SNPs, screening of R genes and marker assisted selection to develop superior cultivar varieties in cardamom.

**Acknowledgments**

We acknowledge the financial support of Secretary, Spices Board of India, Ministry of Commerce and Industry, Government of India.

**Conflict of interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**References**

[1] R.S. Bhai, Y.R. Sarma, Characterization of Phytophthora de Bary inciting capsule rot/Azhukal and leaf blight disease of cardamom (Elettaria cardamomum L. Maton), J. Plant. Crops 33 (2005) 193–203.

[2] J. Thomas, R.S. Bhai, Fungal and bacterial diseases of cardamom (Elettaria cardamomum Maton) and their management, J. Spices Aromat. Crop. 4 (1) (1995) 24–31.

[3] C. Zou, P. Wang, Y. Xu, Bulked sample analysis in genetics, genomics and crop improvement, Plant Biotechnol. J 14 (10) (2016) 1941–1955, https://doi.org/10.1111/pbi.12559.

[4] F. Nadiya, N. Anjali, A. Gangaprasad, K.K. Sabu, High-quality RNA extraction from small cardamom tissues rich in polysaccharides and polyphenols, Anal. Biochem. 485 (2015) 25–27, https://doi.org/10.1016/j.ab.2015.05.017.

[5] Brian J. Haas, Papanicolaou Alexie, Yassour Moran, Manfred Grabherr, D. Philip, Blood, Joshua Bowden, Matthew Brian Couger et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494, https://doi.org/10.1038/nprot.2013.084.

**Table 1**

| Plant Material | Control | Diseased |
|----------------|---------|----------|
| Total number of raw reads | $46,931,637^2 = 93,863,274$ | $31,682,496^2 = 63,364,992$ |
| Total number of bases | 9.93 GB | 6.63 GB |
| Initial GC% | 43 | 44 |
| Read length | 101 | 101 |
| GC% after trimming | 43 | 43.5 |
| Reads after adapter removal and quality trimming | $46,097,664^2 = 92,195,328$ | $31,183,779^2 = 62,367,558$ |
| Total contigs | 111,495 | 91,096 |
| Max Contig Length | 17,667 | 16,840 |
| N50 | 3013 | 2729 |
| Total Length | 243,651,614 | 185,125,249 |
| GC% after assembly | 39.90 | 40.35 |
| Total size of assembly | 274.5 MB | 210.6 MB |