Genomics Data 11 (2017) 102–103

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

# Transcriptome profiling of *Elettaria cardamomum* (L.) Maton (small cardamom)

F. Nadiya [a], N. Anjali [a], Jinu Thomas [a], A. Gangaprasad [b], K.K. Sabu [a,*]

[a] Biotechnology and Bioinformatics Division, Jawaharlal Nehru Tropical Botanic Garden and Research Institute (JNTBGRI), Palode, Thiruvananthapuram 695562, India
[b] Department of Botany, University of Kerala, Thiruvananthapuram 695581, India

ARTICLE INFO

*Article history:*
Received 16 December 2016
Accepted 24 December 2016
Available online 31 December 2016

*Keywords:*
Transcriptome profiling
*Elettaria cardamomum*
Cardamom
Ion proton sequencing
RNA seq

ABSTRACT

*Elettaria cardamomum* (L.) Maton, known as 'queen of spices, is a perennial herbaceous monocot of the family Zingiberaceae, native to southern India. Cardamom is an economically valuable spice crop and used widely in culinary and medicinal purposes. In the present study, using Ion Proton RNA sequencing technology, we performed transcriptome sequencing and *de novo* transcriptome assembly of a wild and five cultivar genotypes of cardamom. RNA-seq generated a total of 22,811,983 (92 base) and 24,889,197 (75 base) raw reads accounting for approximately 8.21GB and 7.65GB of sequence data for wild and cultivar genotypes of cardamom respectively. The raw data were submitted to SRA database of NCBI under the accession numbers SRX1141272 (wild) and SRX1141276 (cultivars). The raw reads were quality filtered and assembled using MIRA assembler resulted with 112,208 and 264,161contigs having N50 value 616 and 664 for wild and cultivar cardamom respectively. The assembled unigenes were functionally annotated using several databases including PlantCyc for pathway annotation. This work represents the first report on cardamom transcriptome sequencing. In order to generate a comprehensive reference transcriptome, we further assembled the raw reads of wild and cultivar genotypes which might enrich the plant transcriptome database and trigger advanced research in cardamom genomics.



| Specifications | |
|---|---|
| Organism/cell line/tissue | Wild and cultivar *Elettaria cardamomum* (L.) Maton |
| Sex | Monoecious |
| Sequencer or array type | Ion Proton System™ |
| Data format | Raw data in BAM file |
| Experimental factors | Transcriptome sequencing of tissues of wild and cultivar cardamom |
| Experimental features | Freshly collected leaf, stem, flower, flower buds and young fruits were used for RNA isolation, tissues were pooled, RNA seq libraries representing two genotypes of cardamom were sequenced and *de novo* assembled |
| Consent | NA |
| Sample source location | JNTBGRI, Cardamom germplasm conservatory |

## 1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRX1141276[accn]
http://www.ncbi.nlm.nih.gov/sra/SRX1141272[accn]

## 2. Introduction

*Elettaria cardamomum* (L.) Maton or small cardamom, belongs to the family Zingiberaceae, is a potential spice crop due to its unique aromatic flavor and multiple health benefits. Cardamom seeds and fruits are the economically significant parts and effectively used as a traditional medicine, food additive and flavoring agent [1]. The essential oil component extracted from the fruits of cardamom was reported to possess antibacterial, anti-inflammatory, analgesic, and antispasmodic activities [2,3]. In the cardamom essential oil, the bioactive component d- limonene was reported to possess chemopreventive property towards colon cancer, mammary, lung, liver, skin and stomach cancers in rodents [4–6]. Hence better understanding of genes and pathways associated with the biosynthesis of the active compounds in cardamom might be beneficial for therapeutic purposes as well as selection of superior genotypes. Since there are no genomic and transcriptome data of cardamom available in any of the publicly available databases, the data we presented

* Corresponding author.
*E-mail addresses:* nadiyashiyas@yahoo.com (F. Nadiya), anjalinair20@gmail.com (N. Anjali), jinutchirayath@gmail.com (J. Thomas), agangaprasad@yahoo.com (A. Gangaprasad), sabu@jntbgri.res.in (K.K. Sabu).

http://dx.doi.org/10.1016/j.gdata.2016.12.013

**Table 1**
Read and assembly statistics of cardamom transcriptome.

| Plant material | Cultivar | Wild |
|---|---|---|
| Total number of raw reads | 24,889,197 | 22,811,983 |
| Total number of bases | 1,856,105,873 | 2,105,322,150 |
| >20 | 1,518,466,470 | 1,719,212,155 |
| GC% | 55 | 55 |
| Read length | 73 bp | 92 bp |
| Reads after adapter removal and quality trimming | 18,921,421 | 17,515,399 |
| Total contigs | 264,161 | 112,208 |
| Largest contig | 1489 | 1263 |
| N50 contig size | 616 | 664 |
| Average consensus quality | 26 | 26 |
| Contigs with reads without quality values | 0 | 0 |
| Maximum coverage (total) | 24,835 | 17,774 |
| Total size of assembly | 16 MB | 43 MB |

here provide information on potential genes coding for enzymes responsible for pharmaceutically valuable bioproducts.

## 3. Experimental design, materials and methods

### 3.1. Plant material

Fresh leaf, stem, flower, flower bud and young fruit were collected from wild and five cultivar accessions (including landraces and released varieties) of cardamom growing in the JNTBGRI cardamom conservatory. The tissues were immediately frozen in liquid nitrogen and kept as such until processing.

### 3.2. Total RNA isolation and transcriptome sequencing

Total RNA was isolated from all tissues of cardamom using RNeasy plant mini kit combined with modified CTAB method [7]. Quantification and quality analysis were done using 2100 BioAnalyzer (Agilent Technologies). The tissues of wild cardamom were pooled as one sample and the tissues from all cultivars were pooled as second sample representing wild and cultivar genotypes of cardamom. Two sets of total RNA were purified and cDNA library was constructed using Ion Total RNA seq Kit V2 according to manufacturer's instructions. The purified libraries were submitted to RNA sequencing with the Ion Proton sequencer. RNA seq generated a total of 22,811,983 (92 base) and 24,889,197 (75 base) raw reads accounting for approximately 8.21 and 7.65 GB of sequence data for wild and cultivar genotypes of cardamom respectively.

### 3.3. De novo transcriptome assembly and functional annotation

The raw reads were preprocessed to remove low quality bases, tRNAs and rRNAs. *De novo* assembly was performed with MIRA (Mimicking Intelligent Read Assembly) program [8] which assembles read pairs from partial path into contigs. The transcriptome assembly generated 112,208 and 264,161 contigs for wild and cultivar genotypes and to generate a comprehensive reference transcriptome a combined assembly was performed which yielded 178,745 contigs (Table 1). We annotated the assembled transcripts to NCBI non redundant (NR) protein database, Swissprot, TrEMBL, pfam, PlantCyc, KOG and TAIR10 databases. We also identified the differential expression analysis of significant genes in wild and cultivar genotypes of cardamom. The information we provided might be useful in molecular marker development, SNP identification, genetic manipulation of specific genes to enhance the production of desired compounds and breeding for developing novel elite cultivars.

## Conflict of interest

The authors declare that they have no competing interests.

## Acknowledgements

The authors thank the Director of the Jawaharlal Nehru Tropical Botanic Garden & Research Institute (JNTBGRI) for providing the necessary facilities and financial assistance through Plan Project (P114) to carry out this research work. Thanks are due to the Kerala State Council for Science, Technology, and Environment (KSCSTE) for granting Research Fellowship to Nadiya, F. (010-52/FSHP/10/CSTE) to conduct this study. We acknowledge the permission granted by Kerala Forest Department for collecting the plant samples. Thanks are due to Shafeek, S. and Vishnu, J. S. for assistance rendered for the study. We are grateful to the staff at the Centre for Cellular and Molecular Platforms (C-CAMP, Bangalore, India) for performing high throughput sequencing. We are also thankful to Indian Cardamom Research Institute and Cardamom Research Centre in Kerala, India for providing necessary cardamom cultivars used in the present study.

## References

[1] M. Valko, C.J. Rhodes, J. Moncol, M.M. Izakovic, M. Mazur, Free radicals, metals and antioxidants in oxidative stress-induced cancer. Chem. Biol. Interact. 160 (2006) 1–40.
[2] H. Al-Zuhair, B. El-Sayeh, H.A. Ameen, H. Al-Shoora, Pharmacological studies of cardamom oil in animals. Pharmacol. Res. 34 (1996) 79–82.
[3] M. Elgayyar, F.A. Draughon, D.A. Golden, J.R. Mount, Antimicrobial activity of essential oils from plants against selected pathogenic and saprophytic microorganisms. J. Food Prot. 64 (2001) 1019–1024.
[4] A. Acharya, I. Das, S. Singh, T. Saha, Chemopreventive properties of indole-3-carbinol, diindolylmethane and other constituents of cardamom against carcinogenesis. Recent Pat. Food Nutr. Agric. 2 (2010) 166–177.
[5] P.L. Crowell, M.N. Gould, Chemoprevention and therapy of cancer by D-limonene. Crit. Rev. Oncog. 5 (1994).
[6] M. Asamoto, T. Ota, H. Toriyama-Baba, N. Hokaiwado, A. Naito, H. Tsuda, Mammary carcinomas induced in human c-Ha-ras proto-oncogene transgenic rats are estrogen-independent, but responsive to D-limonene treatment. Jpn. J. Cancer Res. 93 (2002) 32–35.
[7] 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, http://dx.doi.org/10.1016/j.ab.2015.05.017.
[8] B. Chevreux, T. Pfisterer, B. Drescher, A.J. Driesel, W.E.G. Müller, T. Wetter, S. Suhai, Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14 (2004) 1147–1159.