Data in Brief 14 (2017) 789–792

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Data on identification of conserved and novel miRNAs in *Elettaria cardamomum*

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*

A R T I C L E I N F O

Article history:
Received 14 May 2017
Received in revised form
2 August 2017
Accepted 28 August 2017
Available online 5 September 2017

Keywords:
miRNA
Cardamom
Cultivar
Spice crop
Wild genotype

A B S T R A C T

Elettaria cardamomum (L.) Maton, or small cardamom referred as 'queen of spices', is a perennial herbaceous rhizomatous monocot of the family Zingiberaceae. Cardamom seeds and fruits are the economically significant parts and effectively used as a traditional medicine, food additive and flavoring agent. In the present study, using Ion Proton next generation sequencing technology we performed the small RNA sequencing, conserved and novel miRNA predictions of a wild and five cultivar genotypes of cardamom. Small RNA sequencing generated a total of 5,451,328 and 2,756,250 raw reads for wild and cultivar cardamom respectively. The raw data was submitted to SRA database of NCBI under the accession numbers and SRX2273863 (wild) and SRX2273862 (cultivars). The raw reads were quality filtered and predicted conserved and novel miRNAs for wild and cultivar cardamom. The predicted miRNAs, miRNA-targets and functional annotations might provide valuable insights into differences between wild progenitor and cultivated cardamom.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* 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.dib.2017.08.037
2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications Table:

Subject area	Biology
More specific subject area	Plant molecular biology
Type of data	Small RNA sequence data
How data was acquired	Next generation sequencing using Ion Proton System$^{(TM)}$
Data format	Raw data in FASTQ file
Experimental factors	Small RNA 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, Small RNA seq libraries representing two genotypes of cardamom were sequenced and predicted known and novel miRNAs
Data source location	JNTBGRI, cardamom germplasm conservatory
Data accessibility	NCBI, SRA

Value of the data

- The data presented in the article provides information on genome wide conserved and novel miRNA profiling and differential expression analysis of wild and cultivar genotypes of cardamom.
- This data identifies miRNAs participating in the negative regulation of significant transcription factors, enzymes and other mRNAs involved in plant development, stress responses and pathogenic resistance.
- The identified miRNAs were also compared with other monocots and *Arabidopsis* to determine the extent of conservation of known miRNAs between cardamom, other monocots and *Arabidopsis*.
- The data allows other researchers to study miRNA regulations in various biosynthetic pathways.

1. Data

The data shared in this article provides information on conserved and novel miRNAs and differential miRNA expression analysis between wild and cultivar genotypes of cardamom.

https://www.ncbi.nlm.nih.gov/sra/SRX2273862[accn].

https://www.ncbi.nlm.nih.gov/sra/SRX2273863[accn].

2. Experimental design, materials and methods

2.1. Plant material

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

2.2. Total RNA isolation and small RNA sequencing

Total RNA was isolated from leaves, stem, flower, flower buds and young fruits of cardamom using miRNeasy mini kit combined with modified CTAB method [1]. 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

Table 1
Read and assembly statistics.

Plant material	Wild	Cultivar
Raw reads obtained after sequencing	5,451,328	2,756,250
Reads after adapter removal	5,375,277 (98.6%)	2,720,430 (98.7%)
Reads after other RNAs removed	3,524,114 (64.6%)	1,794,905 (65.1%)
Reads after size filtering	2,066,216 (37.9%)	1,125,109 (40.8%)
Unique sequences obtained after collapsing	1,080,645 (19.8%)	588,044 (21.3%)
Number of reads aligned to miRBase database	1559	1368
Number of known miRNAs	138	141
Reads retained for novel miRNA prediction	1,015,420 (18.6%)	547,739 (19.8%)
Number of reads aligned to *Curcuma longa*	487,401	277,428
Novel miRNAs predicted	9	5

cultivar cardamom. Two sets of total RNA enriched with miRNAs 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 small RNA sequencing with the Ion Proton Sequencer.

2.3. Identification of conserved and novel miRNAs

The raw reads were preprocessed to remove adapter sequences, rRNAs, tRNAs, other small RNAs and sequences below 17 and 27nt. The identical sequences were collapsed and known miRNAs were identified by performing blastn search against plant mature miRNAs from miRBase database. The remaining reads were mapped to the *Curcuma longa* (Zingiberaceae family) transcriptome sequence (the whole genome or transcriptome sequence of cardamom was not available at the time of the study) to predict putative novel miRNAs using miRDeep-P software [2] satisfying Mayer's criteria [3] of miRNA prediction (Table 1). Using the identified miRNAs we also performed the differential expression analysis, target prediction and Gene Ontology (GO) annotations for wild and cultivar genotypes of cardamom. The information we provided might furnish awareness on genetic resources of wild cardamom, molecular marker development and breeding for developing novel elite cultivars using RNAi technology. The study also provides insight into the genetic variation in wild and domesticated cardamom.

2.4. Differential expression analysis and target prediction

Count of each miRNA was normalized to Tags per million (TPM) to compare the abundance of miRNAs in wild and cultivar library. Fold change of each miRNA between the two libraries was calculated employing the formula: $\log_2$ (cultivars/wild) [4]. Unique potential targets were predicted for both novel and conserved miRNAs using psRNA target [5].

Acknowledgements

The authors thank the Director of the Jawaharlal Nehru Tropical Botanic Garden & Research Institute (JNTBGRI) for providing the necessary facilities 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.

Transparency document. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.08.037.

References

[1] 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.
[2] X. Yang, L. Li, miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants, Bioinformatics 27 (2011) 2614–2615.
[3] B.C. Meyers, M.J. Axtell, B. Bartel, D.P. Bartel, D. Baulcombe, J.L. Bowman, X. Cao, J.C. Carrington, X. Chen, P.J. Green, S. Griffiths-Jones, S.E. Jacobsen, A.C. Mallory, R.A. Martienssen, R.S. Poethig, Y. Qi, H. Vaucheret, O. Voinnet, Y. Watanabe, D. Weigel, J.-K. Zhu, Criteria for annotation of plant microRNAs, Plant Cell. 20 (2008) 3186–3190.
[4] A.M. Perez-de-Castro, S. Vilanova, J. Cañizares, L. Pascual, J.M. Blanca, M.J. Diez, J. Prohens, B. Pico, Application of genomic tools in plant breeding, Curr. Genom. 13 (2012) 179–195.
[5] X. Dai, P.X. Zhao, psRNATarget: a plant small RNA target analysis server, Nucleic Acids Res. 39 (2011) W155–W159.