Data Article

Data on occurrence of miRNA precursors in the *Cucurbita maxima* phloem sap

Eugeny A. Tolstyko a, Alexander A. Lezzhov b, Anna D. Solovieva a, Andrey G. Solovyev a, c, d, *

a Department of Virology, Biological Faculty, Moscow State University, Moscow 119234, Russia
b Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow 119991, Russia
c Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119992, Russia
d Sechenov First Moscow State Medical University, Institute of Molecular Medicine, Moscow 119991, Russia

** Article history:**
Received 22 October 2019
Received in revised form 16 December 2019
Accepted 23 December 2019
Available online 3 January 2020

**Keywords:**
miRNA
pri-miRNA
miRNA precursor
Phloem
Phloem RNA

**Abstract**
The phloem sieve elements (SEs), enucleate cells, contain RNAs, which are imported from surrounding tissues and cells, mostly companion cells tightly associated with SEs, and transported via the phloem over the whole plant body. The RNA phloem transport is essential for plant individual development and responses to environmental cues. Recently, we identified primary miRNA (pri-miRNA) sequences in *de novo* assembled transcriptome of *Cucurbita maxima* phloem sap and reported 11 most abundant pri-miRNAs [1]. Here, we provide the output of this analysis in complete detail. For the full set of pri-miRNAs identified in the *C. maxima* phloem sap transcriptome, data on relative abundance are provided along with annotated sequence data.

© 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 description

Sequence analysis of transcriptome sequencing data for *Cucurbita maxima* phloem sap revealed the presence of 35 contigs showing significant (e-value less than 1e-15) sequence similarity to known pri-miRNAs of *Cucumis melo* [1]; raw data on sequences of identified contigs are presented in Supplementary Fig. S1. In a further analysis of the raw data, a relative abundance of each of these contigs in the phloem sap was characterized by the number of primary reads that could be aligned to this contig, read count normalized per 100 nucleotides of contig sequence, and average coverage (Table 1).

Sequences showing significant (e-value less than 1e-15) sequence similarity to the following *C. melo* miRNA precursors were not found in *C. maxima* phloem sap transcriptome: miR156a, miR156b, miR156c, miR156e, miR156f, miR156h, miR156i, miR156j, miR159b, miR160c, miR160d, miR164a, miR164b, miR164c, miR164d, miR166a, miR166c, miR166g, miR166h, miR166i, miR167a, miR167b, miR167d, miR169a, miR169b, miR169c, miR169d, miR169e, miR169f, miR169g, miR169h, miR169i, miR169j, miR169k, miR169l, miR169m, miR169o, miR169p, miR169q, miR169s, miR169t, miR171a, miR171b, miR171c, miR171d, miR171e, miR171f, miR171g, miR171h, miR172a, miR172c, miR172d, miR172e, miR172f, miR1863, miR2111b, miR390c, miR390d, miR393b, miR393c, miR394a, miR394b, miR395a, miR395b, miR395c, miR395d, miR395e, miR395f, miR396a, miR396c, miR396d, miR396e, miR397, miR398a, miR398b, miR399a, miR399b, miR399c, miR399d, miR399e, miR399f, miR399g, miR408, miR477a, miR477b, miR530a, miR7129, miR7130, miR828, miR845, miR858.

2. Experimental design, materials, and methods

Transcriptome sequencing data for *C. maxima* phloem sap (SRX146322) were downloaded using fastq-dump tool of NCBI SRA Toolkit 2.9.0. (http://ncbi.github.io/sra-tools/). Reads quality was checked with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). De novo assembly of phloem sap transcriptome was carried out using SPAdes 3.12.0 [2] in “RNA mode”. The obtained C.
maxima phloem sap transcriptome assembly included 96318 contigs. Sequences of C. melo pre-miRNAs annotated at miRBase [3] were downloaded in fasta format and used as queries for BLAST [4] searches for C. maxima assembled contigs containing related sequences. In order to obtain coverage values, primary reads were aligned with assembled contigs using Bowtie2 [5].

**CRediT authorship contribution statement**

**Eugeny A. Tolstyko:** Methodology, Investigation. **Alexander A. Lezzhov:** Methodology, Investigation. **Anna D. Solovieva:** Investigation. **Andrey G. Solovyev:** Conceptualization, Funding acquisition, Writing - original draft.

**Acknowledgments**

This work was supported by the Russian Science Foundation (grant 17-14-01032).

**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.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.105083.

References

[1] E.A. Tolstyko, A.A. Lezzhov, A.G. Solovyev, Identification of miRNA precursors in the phloem of Cucurbita maxima, PeerJ 7 (2019) e8269, https://doi.org/10.7717/peerj.8269.
[2] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.
[3] A. Kozomara, S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data, Nucleic Acids Res. 42 (2014) D68–D73, https://doi.org/10.1093/nar/gkt1181.
[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, https://doi.org/10.1016/S0022-2836(05)80360-2.
[5] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.