Genomics Data 11 (2017) 60–61

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

De novo transcriptome assembly of the mycoheterotrophic plant *Monotropa hypopitys*

Alexey V. Beletsky, Mikhail A. Filyushin, Eugeny V. Gruzdev, Alexander M. Mazur, Egor B. Prokhortchouk, Elena Z. Kochieva, Andrey V. Mardanov, Nikolai V. Ravin *, Konstantin G. Skryabin

Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, 119071 Moscow, Russia

ARTICLE INFO

Article history:
Received 16 November 2016
Accepted 28 November 2016
Available online 30 November 2016

Keywords:
Parasitic plant
Monotropa hypopitys
Transcriptome
RNA-Seq

ABSTRACT

Monotropa hypopitys (pinesap) is a non-photosynthetic obligately mycoheterotrophic plant of the family *Ericaceae*. It obtains the carbon and other nutrients from the roots of surrounding autotrophic trees through the associated mycorrhizal fungi. In order to understand the evolutionary changes in the plant genome associated with transition to a heterotrophic lifestyle, we performed *de novo* transcriptomic analysis of *M. hypopitys* using next-generation sequencing. We obtained the RNA-Seq data from flowers, flower bracts and roots with haustoria using Illumina HiSeq2500 platform. The raw data obtained in this study can be available in NCBI SRA database with accession number of SRP069226. A total of 10.3 GB raw sequence data were obtained, corresponding to 103,357,809 raw reads. A total of 103,025,683 reads were filtered after removing low-quality reads and trimming the adapter sequences. The Trinity program was used to *de novo* assemble 98,349 unigens with an N50 of 1342 bp. Using the TransDecoder program, we predicted 43,505 putative proteins. 38,416 unigenes were annotated in the Swiss-Prot protein sequence database using BLASTX. The obtained transcriptomic data will be useful for further studies of the evolution of plant genomes upon transition to a non-photosynthetic lifestyle and the loss of photosynthesis-related functions.

Specifications

Organism/cell line/tissue	*Monotropa hypopitys*/flowers, flower bracts, and roots
Sex	N/A
Sequencer or array type	Illumina HiSeq2500
Data format	Raw data: FASTQ file
Experimental factors	*De novo* transcriptome assembly of *Monotropa hypopitys*
Experimental features	Flowers and leaves (flower bracts) of two individual *Monotropa hypopitys* plants and two pooled samples of roots with haustoria were harvested for total RNA extraction, sequencing, *de novo* transcriptome assembly and annotation
Consent	N/A
Sample source location	Sample was collected in Kaluga region, Russia

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/sra/SRP069226.

2. Introduction

Monotropa hypopitys (pinesap) is a non-photosynthetic obligately mycoheterotrophic plant of the family *Ericaceae*, subfamily *Monotropoideae* [1]. It obtains the carbon and other nutrients from the roots of surrounding autotrophic trees through the associated mycorrhizal fungi [2]. The above-ground part of *M. hypopitys* plant is up to 35 cm tall unbranched adventitious raceme-like inflorescence with several flowers on the top and scale-like flower bracts which cover most of the inflorescence. Short roots are invested by a net of sheathing mycorrhizas forming haustorium-like structures used to attach the fungal partner.

Transition to a heterotrophic lifestyle and the loss of photosynthesis is expected to relax the selective pressure on photosynthetic machinery which becomes unnecessary. The most evident consequence is a reduction in the size and gene content of the chloroplast genome, which correlates with the loss of genes related to photosynthesis [3–5]. Recently we sequenced the chloroplast genome of *M. hypopitys* and found that it is highly reduced in size and lacks genes encoding NADH dehydrogenase, photosynthesis-related proteins, the plastid-encoded RNA polymerase and some other functions [6,7]. Transition to heterotrophy is expected to be associated also with the changes in the nuclear genome of a parasitic plant since it encodes a number of functions related to autotrophic lifestyle. However, these changes are poorly known and

* Corresponding author.
E-mail addresses: nravin@biengi.ac.ru, nravin@mail.ru (N.V. Ravin).

http://dx.doi.org/10.1016/j.gdata.2016.11.020

limited to few studies of transcriptomes of parasitic plants (e.g. [8–10]). In this study we carried out sequencing and *de novo* assembly of transcriptome for *M. hypopitys*.

3. Experimental design, materials and methods

3.1. Plant materials

Several *Monotropa hypopitys* plants, including the underground parts with a piece of soil, were collected in the forest in Kaluga region, Russia and quickly transported to the laboratory for isolation of nucleic acids.

3.2. RNA isolation, library preparation, and sequencing

Fresh flowers and leaves (often considered as flower bracts) of two individual *Monotropa hypopitys* plants and two samples of pooled roots with haustoria were used for RNA isolation. Roots were carefully rinsed in water to eliminate attached soil immediately before RNA extraction. Total RNA was isolated from ~300 mg tissue for each six samples using the RNeasy Plant Mini kit (Qiagen, Valencia, CA).

mRNA library preparation was performed using a NEBNext® mRNA Library Prep Reagent Set for Illumina® according to the manufacturer's instructions (New England BioLabs Inc., Ipswich, MA, USA). The libraries were sequenced by MyGene Co. (Moscow, Russia) using the HiSeq 2500 platform. About 17 million of 100-bp single end reads were generated for each sample.

3.3. De novo transcriptome assembly, identification of protein coding regions, and annotation

We obtained a total of 10.3 GB raw sequence data, corresponding to 103,357,809 raw reads (18.5 mln for flowers of plant 1.7 mln for leaves of plant 1, 14.3 mln for flowers of plant 2, 17.7 mln for leaves of plant 2, 16.1 mln for root sample 1, and 18.9 mln for root sample 2). A total of 103,025,683 high quality reads were filtered after removing of adapter sequences and quality trimming with Cutadapt [11] and Sickle (https://github.com/najoshi/sickle), respectively.

The information on the transcriptome sequencing and assembly is summarised in Table 1.

De novo assembly of transcriptome with clean reads from the combined six RNA-seq datasets was carried out using Trinity 2.1.1 with default parameters [12]. The assembly generated 98,349 unigens ranging in length from 201 to 12,993 bp, with an N50 of 1342 bp. Coding regions prediction in the assembled transcripts was performed by the TransDecoder program (http://transdecoder.github.io) implemented in the Trinity software. As a result, we predicted a total of 43,505 proteins. Trinotate (https://trinotate.github.io/) was used to assign hits from TrEMBL and Swiss-Prot databases (http://www.uniprot.org/uniprot/), and to assign GO terms and pfam domains. 37,974 unigenes were annotated in the TrEMBL protein database using predicted protein sequences and 38,416 unigenes were annotated in the Swiss-Prot database using BLASTX.

The relative levels of transcription of protein-coding genes in each of four tissue samples were calculated by mapping of cleaned reads on the assembled transcripts employing Trinity scripts, RSEM [13] and Bowtie 2 program [14]. Cross sample normalization of transcription levels was done using Trinity scripts (TMM method). The data on the relative expression of annotated genes are available in Supplementary File 1.

822 of the assigned proteins were probably derived from root-associated fungi since (i) they have best BLASTP hits in fungal proteomes, (ii) the corresponding transcripts were expressed only in root samples but not in flowers and leaves, and (iii) nucleotide sequences of these genes were not found in the genome sequences derived from the above-ground parts of *M. hypopitys* plants (to be published elsewhere).

In conclusion, we have sequenced and *de novo* assembled the transcriptome of the non-photosynthetic parasitic plant *M. hypopitys*. The obtained transcriptome data will be useful for further studies of the evolution of plant genomes upon transition to a non-photosynthetic lifestyle and the loss of photosynthesis-related functions.

Table 1
Summary of the sequencing and *de novo* assembly.

Index	*Monotropa hypopitys*
No of assembled reads	103,025,683
Number of unigenes	98,349
N50 length of unigene (bp)	1342
Average unigene length (bp)	780
Predicted proteins	43,505
Unigens annotated in TrEMBL	37,974
Unigens annotated in Swiss-Prot	38,416

Conflict of interest

The authors declare no conflicts of interest in this study.

Acknowledgments

This work was funded by Russian Science Foundation grant 14-24-00175.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.11.020.

References

[1] K.A. Kron, W.S. Judd, P.F. Stevens, D.M. Crayn, A.A. Anderberg, P.A. Gadek, C.J. Quinn, J.L. Luteyn, Phylogenetic classification of Ericaceae: molecular and morphological evidence. Bot. Rev. 68 (2002) 335–423.
[2] J.R. Leake, The biology of myco-heterotrophic ('saprophytic') plants. New Phytol. 127 (1994) 171–216.
[3] C.W. de Pamphilis, J.D. Palmer, Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348 (1990) 337–339.
[4] K.H. Wolfe, C.W. Morden, J.D. Palmer, Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 10648–10652.
[5] C.F. Barrett, J.V. Freudenstein, J. Li, D.R. Mayfield-Jones, L. Perez, J.C. Pires, C. Santos, Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. Mol. Biol. Evol. 31 (2014) 3095–3112.
[6] E.V. Gruzdev, A.V. Mardanov, A.V. Beletsky, E.Z. Kochieva, N.V. Ravin, K.G. Skryabin, The complete chloroplast genome of parasitic flowering plant *Monotropa hypopitys*: extensive gene losses and size reduction. Mitochondrial DNA 1 (2016) 212–213.
[7] N.V. Ravin, E.V. Gruzdev, A.V. Beletsky, A.M. Mazur, E.B. Prokhortchouk, M.A. Filyushin, E.Z. Kochieva, V.V. Kadnikov, A.V. Mardanov, K.G. Skryabin, The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. BMC Plant Biol. 16 (Suppl. 3) (2016) 238, http://dx.doi.org/10.1186/s12870-016-0929-7.
[8] M.J. Torres, A.A. Tomilov, N. Tomilova, R.L. Reagan, J.I. Yoder, Pscroph, a parasitic plant EST database enriched for parasite associated transcripts. BMC Plant Biol. 5 (2005) 24.
[9] S. Yoshida, J.K. Ishida, N.M. Kamal, A.M. Ali, S. Namba, K. Shirasu, A full-length enriched cDNA library and expressed sequence tag analysis of the parasitic weed, *Striga hermonthica*. BMC Plant Biol. 10 (2010) 55.
[10] N.J. Wickett, L.A. Honaas, E.K. Wafula, M. Das, K. Huang, B.A. Wu, L. Landherr, M.P. Timko, J. Yoder, J.H. Westwood, C.W. de Pamphilis, Transcriptomes of the parasitic plant family *Orobanchaceae* reveal surprising conservation of chlorophyll synthesis. Curr. Biol. 21 (2011) 2098–2104.
[11] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17 (2011) 10–12.
[12] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. Leduc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (2013) 1494–1512.
[13] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinforma. 12 (2011) 323.
[14] B. Langmead, S. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (2012) 357–359.