Genomics Data 7 (2016) 18–19

Contents lists available at ScienceDirect

Genomics Data

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

RNA-seq analysis for plant carnivory gene discovery in *Nepenthes* × *ventrata*

Wan Nor Adibah Wan Zakaria, Kok-Keong Loke, Hoe-Han Goh, Normah Mohd Noor

Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 29 October 2015
Accepted 7 November 2015
Available online 10 November 2015

Keywords:
Carnivorous plant
Digestive enzyme
Nepenthes
Pitcher
Transcriptome

ABSTRACT

Carnivorous plants have the ability to capture and digest insects for nutrients, which allows them to survive in land deprived of nitrogenous nutrients. *Nepenthes* spp. are one of the carnivorous plants, which uniquely produce pitcher from the tip of an elongated leaf. This study provides the first transcriptome resource from pitcher of a *Nepenthes ventricosa* × *Nepenthes alata* hybrid, *Nepenthes* × *ventrata* to understand carnivory mechanism in *Nepenthes* spp., as well as in other carnivorous species. Raw reads and the transcriptome assembly project have been deposited to SRA database with the accession numbers SRX1389337 (day 0 control), SRX1389392 (day 3 longevity), and SRX1389395 (day 3 chitin-treated).

Specifications	
Subject area	Biology, Plant Molecular Biology
Type of data	Transcriptome sequences
Organism/Cell line/tissue	*Nepenthes* × *ventrata* (Pitcher)
Sequencer type	Illumina HiSeq™ 2500
Data format	Raw and processed
Experimental factors	Experimental plot, control and treatments of pitchers
Experimental features	RNA-seq dataset for gene discovery in a pitcher plant
Sample source location	Malaysia
Data accessibility	SRA database accession SRX1389337 (http://www.ncbi.nlm.nih.gov/sra/SRX1389337) SRX1389392 (http://www.ncbi.nlm.nih.gov/sra/SRX1389392) SRX1389395 (http://www.ncbi.nlm.nih.gov/sra/SRX1389395)

1. Value of the data

- *Nepenthes* spp. plants are one of passive carnivorous genus which lack in molecular genetics information.
- The lack of a complete transcript database from this genus hinders new protein discovery through proteomics approach. Hence, these data sets help in the exploration of novel genes/proteins to understand carnivory in pitcher plants, and more generally in carnivorous plants.
- These data are also important for the identification of unique digestive enzymes and aspartic proteinases from pitcher plant. This will improve our understanding on the evolutionary history of this family of carnivorous plants.

2. Data

Transcriptome profile of *N.* × *ventrata* were generated from the polyA-enriched cDNA libraries prepared from total RNA extracted from its pitcher. The short reads were filtered, processed, assembled and analyzed as describe in the next section. Raw data for this project were deposited at SRA database with the accession numbers SRX1389337 (http://www.ncbi.nlm.nih.gov/sra/SRX1389337) for day 0 control, SRX1389392 (http://www.ncbi.nlm.nih.gov/sra/SRX1389392) for day 3 longevity experiment, and SRX1389395 (http://www.ncbi.nlm.nih.gov/sra/SRX1389395) for day 3 chitin-treatment experiment.

3. Experimental design, materials and methods

3.1. Plant materials

N. × *ventrata* pitcher plants were grown under shady environment in experimental plot (2°55′09.0″N 101°47′04.8″E) at Universiti Kebangsaan Malaysia, Bangi. Whole pitchers were collected and freeze in liquid nitrogen before stored in −80 °C for further use.

Three different pitcher samples were collected, namely day 0 control, day 3 longevity and day 3 chitin-treated. Day 0 control sample was collected with 24 h of pitcher opening. For longevity experiment to understand the effect of time and protein depletion after pitcher opening on gene expression, day 0 pitcher fluids were syringe filtered through 0.22 μm PVDF membrane followed by protein concentration

E-mail address: gohhh@ukm.edu.my (H.-H. Goh).

http://dx.doi.org/10.1016/j.gdata.2015.11.007

Table 1
Statistics of *N.* × *ventrata* assembly.

Attributes	Value
Pre-assembly	
Total raw reads	131,938,236
Total processed reads	124,906,521
Post-assembly	
Number of unigenes	181,810
Number of unique transcripts	170,214
N50 (bp)	1207
Size range (bp)	224–13,720

Table 2
Functional annotation of *N.* × *ventrata* unique transcript.

Annotation/Tools	Number of unique transcripts
Total transdecoder peptides	57,833
BLASTX — SwissProt	59,335
BLASTP — SwissProt	11,854
PFAM — TMHMM	36,497
eggNOG	25,031
Gene Ontology (GO)	52,722
SignalP	2511
RNAMMER	0

at 10,000 molecular weight cutoff (MWCO). The pitchers were then replenished by the filtrate with depleted protein bigger than 10 kDa, sealed by parafilm, and collected after 3 days. For chitin treatment experiment, 30 mM (w/v) of chitin were added into the pitcher fluid upon day 0 pitcher opening, sealed by parafilm, and collected after 3 days.

3.2. Total RNA extraction and quality control, library preparation and RNA-seq

For RNA works, RNA from all samples were extracted using modified method of CTAB [1]. Quantity and integrity of extracted total RNA were determined using NanoDrop (Thermo Fisher Scientific Inc., USA) and Agilent 2100 bioanalyzer (Agilent Technologies, USA), respectively.

One pitcher for each treatment of *N.* × *ventrata* was sequenced using the Illumina HiSeq 2500 sequencing platform. Paired end reads of 125 bp was generated through the standard polyA-enriched library preparation protocol implemented by Macrogen, South Korea.

3.3. Transcriptome de novo assembly, annotation and classification

Raw reads from all three data sets were filtered to remove adapter sequences with sequence pre-processing tool, Trimmomatic [2]. High quality Illumina raw reads with phred score ≥ 25 were kept for assembly. De novo assembly of these processed reads was performed with Trinity (v2.0.6) [3]. Statistics of the assembly is showed in Table 1.

Protein coding sequences of unique transcripts were analyzed via Transdecoder version v2.0.1 as a part of Trinity analysis pipeline. Standard Trinotate (v2.0.0) annotation pipeline (https://trinotate.github.io/) was carried out to annotate the assembled unique transcripts against Swissprot [4], Pfam [5], eggNOG [6], Gene Ontology [7], SignalP [8], and Rnammer [9]. Summary of the annotation is showed in Table 2.

Conflict of interest

All the authors have approved submission and there are no conflicts of interest.

Acknowledgements

This research was supported by Fundamental Research Grant Scheme UKM-RB-06-FRGS0259-2010 & FRGS/2/2014/SG05/UKM/02/4 from the Malaysian Ministry of High Education (MOHE) and Research University Grant DIP-2014-008.

References

[1] S.H. Kim, T. Hamada, Biotechnol. Lett. 27 (2005) 1841–1845.
[2] A.M. Bolger, M. Lohse, B. Usadel, Bioinformatics 30 (2014) 2114–2120.
[3] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, Nat. Protoc. 8 (2013) 1494–1512.
[4] B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M.J. Martin, K. Michoud, C. O'Donovan, I. Phan, Nucleic Acids Res. 31 (2003) 365–370.
[5] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, Nucleic Acids Res. (2013) D222–D230.
[6] S. Powell, K. Forslund, D. Szklarczyk, K. Trachana, A. Roth, J. Huerta-Cepas, T. Gabaldón, T. Rattei, C. Creevey, M. Kuhn, Nucleic Acids Res. (2013) D231–D239.
[7] C. Gene Ontology, Nucleic Acids Res. 32 (2004) D258–D261.
[8] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, Nat. Methods 8 (2011) 785–786.
[9] K. Lagesen, P. Hallin, E.A. Rødland, H.-H. Stærfeldt, T. Rognes, D.W. Ussery, Nucleic Acids Res. 35 (2007) 3100–3108.