Genomics Data 12 (2017) 130–131

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Iso-Seq analysis of *Nepenthes ampullaria*, *Nepenthes rafflesiana* and *Nepenthes* × *hookeriana* for hybridisation study in pitcher plants

Muhammad Mu'izzuddin Zulkapli, Muhammad Aqil Fitri Rosli, Faris Imadi Mohd Salleh, Normah Mohd Noor, Wan Mohd Aizat, Hoe-Han Goh*

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

ARTICLE INFO

Keywords:
Carnivorous plant
Iso-Seq
Nepenthes
Pitcher
PacBio SMRT
Transcriptome

ABSTRACT

Tropical pitcher plants in the species-rich Nepenthaceae family of carnivorous plants possess unique pitcher organs. Hybridisation, natural or artificial, in this family is extensive resulting in pitchers with diverse features. The pitcher functions as a passive insect trap with digestive fluid for nutrient acquisition in nitrogen-poor habitats. This organ shows specialisation according to the dietary habit of different Nepenthes species. In this study, we performed the first single-molecule real-time isoform sequencing (Iso-Seq) analysis of full-length cDNA from *Nepenthes ampullaria* which can feed on leaf litter, compared to carnivorous *Nepenthes rafflesiana*, and their carnivorous hybrid *Nepenthes* × *hookeriana*. This allows the comparison of pitcher transcriptomes from the parents and the hybrid to understand how hybridisation could shape the evolution of dietary habit in Nepenthes. Raw reads have been deposited to SRA database with the accession numbers SRX2692198 (*N. ampullaria*), SRX2692197 (*N. rafflesiana*), and SRX2692196 (*N.* × *hookeriana*).

Specifications	
Organism/cell line/tissue	*Nepenthes ampullaria*, *N. rafflesiana and N.* × *hookeriana* (whole pitcher tissue)
Sex	Not applicable
Sequencer or array type	PacBio RS II
Data format	Raw sequences (HDF5)
Experimental factors	Experimental terrace, pitcher within 24 h of opening
Experimental features	Iso-Seq dataset for 3 *Nepenthes* spp.
Consent	Not applicable
Sample source location	Bangi, Malaysia (2°55′11.5″N 101°47′01.4″E)

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRX2692198.
http://www.ncbi.nlm.nih.gov/sra/SRX2692197.
http://www.ncbi.nlm.nih.gov/sra/SRX2692196.

2. Value of the data

- There is still limited molecular genetics information on different species of *Nepenthes* pitcher plants.
- The lack of transcriptomes from this genus hinders further studies on the molecular mechanism and evolution of their carnivory habit.
- This dataset provides the first full-length transcriptome sequences from pitcher tissues of three well-studied *Nepenthes* species, which is important for guiding functional genomics and proteomics studies.
- This will further improve our understanding on the Nepenthaceae evolutionary history and contribute to gene mining for useful digestive enzymes for industrial applications.

3. Data

Full-length cDNA transcriptome profiles of three *Nepenthes species* (*N. ampullaria*, *N. rafflesiana and N.* × *hookeriana*) were generated from the polyA-enriched cDNA libraries prepared from total RNA extracted from whole pitchers. The sequences generated using PacBio RS II platform from SMRTbell libraries were processed using the SMRT Analysis Server. Raw data for this project were deposited in the SRA database with the accession numbers SRX2692198 (http://www.ncbi.nlm.nih.gov/sra/SRX2692198) for *N. ampullaria*, SRX2692197 (http://www.ncbi.nlm.nih.gov/sra/SRX2692197) for *N. rafflesiana*, and

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

http://dx.doi.org/10.1016/j.gdata.2017.05.003
Received 4 April 2017; Received in revised form 26 April 2017; Accepted 2 May 2017
Available online 03 May 2017
2213-5960/ © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

Table 1
Statistics of overall read filtering.

Metrics	Pre-filter	Post-filter
Total polymerase read bases	5,208,565,586	4,796,143,944
Number of polymerase reads	450,876	273,200
Polymerase read N50 (bp)	27,495	28,186
Mean polymerase read length (bp)	11,552	17,555
Mean polymerase read quality	0.56	0.83
Total subread bases	–	4,680,734,229
Number of subreads	–	2,704,918
Subread N50 (bp)	–	1747
Mean subread length (bp)	–	1730

Table 2
Statistics of Iso-Seq of three *Nepenthes* species.

Metrics	*N. ampullaria*	*N. rafflesiana*	*N. × hookeriana*
Number of reads of insert	86,407	90,076	86,246
Read bases of insert	154,845,182	166,851,165	164,740,830
Mean read length of insert (bp)	1792	1852	1910
Mean read quality of insert	93.6%	93.7%	93.8%
Mean number of passes	9.23	9.05	9.29
Number of five prime reads	58,433	58,647	60,760
Number of three prime reads	59,448	63,228	63,815
Number of poly-A reads	49,162	61,347	61,339
Number of filtered short reads	5834	5631	3613
Number of non-full-length reads	32,103	35,443	31,027
Number of full-length reads	48,470	49,002	51,606
Number of full-length non-chimeric reads	48,147	48,552	51,265
Mean full-length non-chimeric read length (bp)	1590	1623	1668
Number of consensus isoforms	26,130	30,558	33,279
Number of polished high-quality (≥ 0.99) isoforms	17,221	20,254	21,739
Number of polished low-quality (< 0.99) isoforms	8813	10,304	11,540
Mean length of consensus isoforms (bp)	1625	1680	1722

SRX2692196 (http://www.ncbi.nlm.nih.gov/sra/SRX2692196) for *N. × hookeriana*.

4. Experimental design, materials and methods

4.1. Plant materials

All three species of pitcher plants were growing together on a terrace (2°55′11.5″N 101°47′01.4″E) next to experimental plots at Universiti Kebangsaan Malaysia, Bangi. Whole mature pitchers were collected in the morning within 24 h of pitcher opening in June 2015, emptied and frozen in liquid nitrogen before stored in − 80 °C for further use.

4.2. Total RNA extraction and quality control, library preparation and Iso-Seq

Total 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 sample for each species was sequenced using on the Pacific Biosciences RS II platform with one SMRT cell v3 each based on P6-C4 chemistry after standard full-length cDNA (1–3 kb) library preparation protocol (SMRTbell Template Preparation Kit 1.0) at Icahn Medical Institute (Mount Sinai, New York City, USA) [2].

4.3. Read analysis

Sequence movie files from all three data sets were processed and analysed through Iso-Seq pipeline (RS_IsoSeq protocol) using PacBio SMRT Analysis Server v2.3.0 (http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/) to filter out polymerase read reads < 50 bp and quality < 0.75 with 0 minimum full passes (Table 1). Filtered reads were further classified (≥ 300 bp, full-length reads do not require poly-A tails, and 0 maximum number of paths per isoform/read), and clustered using ICE algorithm (estimated cDNA size 1–2 kb) with quiver polishing (quality ≥ 0.99) to generate consensus isoform sequences. Further information on the different reads generated can be found in PacBio wiki (https://github.com/PacificBiosciences/cDNA_primer/wiki/Understanding-PacBio-transcriptome-data). Statistics of the filtered sequences from each transcriptome library is showed in Table 2. The consensus isoform sequences can be used as full-length transcriptome references for the three species of carnivorous pitcher plants for further studies.

Conflict of interest

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

Acknowledgements

We thank Prof Dr Jumaat Haji Adam for contributing to the pitcher samples. This research was supported by FRGS/2/2014/SG05/UKM/02/4 from the Ministry of Higher Education, Malaysia and UKM Research University Grant DIP-2014-008.

References

[1] S.-H. Kim, T. Hamada, Rapid and reliable method of extracting DNA and RNA from sweetpotato, *Ipomoea batatas* (L). Lam, Biotechnol. Lett. 27 (2005) 1841–1845.
[2] S.E. Abdel-Ghany, M. Hamilton, J.L. Jacobi, P. Ngam, N. Devitt, F. Schilkey, A. Ben-Hur, A.S.N. Reddy, A survey of the sorghum transcriptome using single-molecule long reads, Nat. Commun. 7 (2016) 11706.