Genomics Data 12 (2017) 134–135

Contents lists available at ScienceDirect

Genomics Data

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

Data in brief

SMRT sequencing data for *Garcinia mangostana* L. variety Mesta

Mohd Razik Midin[a,d], Kok-Keong Loke[b], Maria Madon[c], Mohd Shukor Nordin[d], Hoe-Han Goh[b,*], Normah Mohd Noor[b]

[a] *Department of Biotechnology, Kulliyyah of Science, International Islamic University Malaysia (IIUM) Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia*
[b] *Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[c] *Advanced Biotechnology and Breeding Centre (ABBC), Malaysian Palm Oil Board (MPOB), No. 6, Persiaran Institusi Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia*
[d] *Department of Plant Science, Kulliyyah of Science, International Islamic University Malaysia (IIUM) Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia*

ARTICLE INFO

Keywords:
Genome sequencing
PacBio
SMRT sequencing
Mangosteen, Mesta

ABSTRACT

The "Queen of Fruits" mangosteen (*Garcinia mangostana* L.) produces commercially important fruits with desirable taste of flesh and pericarp rich in xanthones with medicinal properties. To date, only limited knowledge is available on the cytogenetics and genome sequences of a common variety of mangosteen (Abu Bakar et al., 2016 [1]). Here, we report the first single-molecule real-time (SMRT) sequencing data from whole genome sequencing of mangosteen of Mesta variety. Raw reads of the SMRT sequencing project can be obtained from SRA database with the accession numbers SRX2718652 until SRX2718659.

Specifications

Organism/cell line/ tissue	*Garcinia mangostana* var. Mesta (leaf tissue)
Sex	Female
Sequencer or array type	PacBio RS II
Data format	Raw sequences (HDF5)
Experimental factors	Experimental plot
Experimental features	SMRT-seq dataset for mangosteen variety Mesta
Consent	Not applicable
Sample source location	Bangi, Malaysia (2°55′09.0″N 101°47′04.8″E)

1. Direct link to deposited data

www.ncbi.nlm.nih.gov/sra/SRX2718652
www.ncbi.nlm.nih.gov/sra/SRX2718653
www.ncbi.nlm.nih.gov/sra/SRX2718654
www.ncbi.nlm.nih.gov/sra/SRX2718655
www.ncbi.nlm.nih.gov/sra/SRX2718656
www.ncbi.nlm.nih.gov/sra/SRX2718657
www.ncbi.nlm.nih.gov/sra/SRX2718658
www.ncbi.nlm.nih.gov/sra/SRX2718659

2. Value of the data

- The first SMRT sequence data of mangosteen to help in genome assembly.
- Provide important long sequences spanning repeats for scaffolding of draft genome assembly from short reads.
- Allow better assessment on the genome composition of *G. mangostana* with haplotype information such as GC content, repeats, heterozygosity and genome size.
- Improve genetic information on *G. mangostana* for gene annotation and further studies.

3. Data

Genome sequences of *G. mangostana* var. Mesta were generated from DNA extract of young leaf tissues, size fractionated at 10 kb cutoff and sequenced using PacBio RS II platform. Raw reads for 8 SMRT cells of sequencing from this project were deposited at SRA database with accession numbers SRX2718652 until SRX2718659 (www.ncbi.nlm.nih.gov/sra/SRX2718652) under the BioSample accession SAMN06698961.

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

http://dx.doi.org/10.1016/j.gdata.2017.04.003
Received 7 April 2017; Received in revised form 26 April 2017; Accepted 27 April 2017
Available online 05 May 2017

Table 1
Statistics of Garcinia mangostana var. Mesta SMRT sequencing.

SMRT cell	Accession number	Polymerase reads		Reads Of Insert (ROI)	
		Average read length (bp)	Quality	Average read length (bp)	Quality
C01_1	SRX2718652	14,361	0.83	11,546	0.94
D01_1	SRX2718653	13,737	0.83	11,767	0.95
A01	SRX2718654	14,598	0.83	11,222	0.95
B01	SRX2718655	14,674	0.84	11,440	0.95
C01	SRX2718656	14,551	0.83	11,791	0.95
D01	SRX2718657	14,280	0.83	11,806	0.95
E01	SRX2718658	14,527	0.84	11,696	0.95
F01	SRX2718659	15,064	0.83	11,847	0.95

Table 2
Overall sequencing statistics of *G. mangostana* var. Mesta from 8 SMRT cells.

Metrics	Pre-filter	Post-filter
Polymerase reads		
Total number	1,202,336	660,009
Total bases (bp)	10,258,700,063	9,550,515,555
N50 (bp)	19,480	19,773
Average read length (bp)	8532	14,470
Average quality	0.478	0.833
Subreads		
Total number		1,006,751
Total bases (bp)		9,533,971,571
N50 (bp)		12,947
Average subread length (bp)		9470
Reads of insert (ROI)		
Total number		21,680
Total bases (bp)		252,689,571
Average read length (bp)		11,655
Average quality		0.947
Average number of passes		4.0

4. Experimental design, materials and methods

4.1. Plant materials

Mangosteen plants of Mesta variety were grown under shady environment in experimental plot (2°55′09.0″N 101°47′04.8″E) at Universiti Kebangsaan Malaysia, Bangi. Red young leaf tissues from 4 months old plant were collected in June 2015 and frozen in liquid nitrogen before stored at − 80 °C for DNA extraction.

4.2. DNA extraction and quality control, library preparation and SMRT-Seq

DNA from leaf samples were extracted using CTAB method [2] and further cleanup using AMPure PB magnetic beads. Quantity and quality of extracted total DNA were determined using NanoDrop 1000 (Thermo Fisher Scientific Inc., USA), Qubit Fluorimeter (Thermo Fisher Scientific Inc., USA) and Agilent 2100 bioanalyzer (Agilent Technologies, USA). DNA quality is visualized on Pippin Pulse electrophoresis before performing BluePippin (Sage Science, USA) size selection with 10 kb cut off. SMRT libraries were prepared using the size fractionated DNA using standard manufacturer's protocols (http://www.pacb.com/documentation/guide-pacific-biosciences-template-preparation-and-sequencing/) with DNA Template Prep Kit 3.0 and Binding Kit P6, and loaded at 0.2 nM with MagBead (MagBead OneCellPerWell v1) into SMRT Cell 8Pac v3 for sequencing on PacBio RS II system with P6C4 chemistry by Treecode (Malaysia).

4.3. Raw reads processing

Sequence movie files of 240 min from all 8 SMRT cell runs were processed and analysed through secondary analysis protocols (RS_Subreads and RS_ReadsOfInsert) using PacBio SMRT Analysis Server v2.3.0 (http://www.pacb.com/products-and-services/analytical-software/smrt-analysis/) with default settings. Table 1 shows the SMRT sequencing statistics of each SMRT cell. Table 2 shows the overall statistics of combined analysis of reads from all 8 SMRT cells. Further information on the different reads generated can be found in PacBio wiki (https://github.com/PacificBiosciences/cDNA_primer/wiki/Understanding-PacBio-transcriptome-data).

Conflict of interest

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

Acknowledgements

This research was supported by Research University Grant AP-2012-018. We thank Malaysia Genome Institute for providing the Qubit analysis.

References

[1] S. Abu Bakar, S. Sampathrajan, K.-K. Loke, H.-H. Goh, N. Mohd Noor, DNA-seq analysis of *Garcinia mangostana*, Genomics Data 7 (2016) 62–63.
[2] A. Healey, A. Furtado, T. Cooper, R.J. Henry, Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species, Plant Methods 10 (2014) 21.