Data in Brief

DNA shotgun sequencing analysis of *Garcinia mangostana* L. variety Mesta

CrossMark

Syuhaidah Abu Bakar[a], Suresh Kumar[b], Kok-Keong Loke[a], Hoe-Han Goh[a,*], Normah Mohd Noor[a]

[a] *Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[b] *Department of Diagnostic & Allied Health Science, Faculty of Health and Life Sciences, Management and Science University, 40100 Shah Alam, Selangor, Malaysia*

ARTICLE INFO

ABSTRACT

Mangosteen (*Garcinia mangostana* Linn.) is an ultra-tropical tree characterized by its unique dark purple fruits with white flesh. The xanthone-rich purple pericarp tissue contains valuable compounds with medicinal properties. Following previously reported genome sequencing of a common variety of mangosteen [1], we performed another whole genome sequencing of a commercially popular variety of this fruit species (var. Mesta) for comparative analysis of its genome composition. Raw reads of the DNA sequencing project were deposited to SRA database with the accession number SRX2709728.

Specifications

Organism/cell line/tissue	*Garcinia mangostana* var. Mesta (leaf tissue)
Sex	Female
Sequencer or array type	Illumina HiSeq™ 2000
Data format	Raw sequences (Fastq)
Experimental factors	Experimental plot
Experimental features	DNA-seq dataset for mangosteen genome survey
Consent	Not applicable
Sample source location	Bangi, Malaysia (2°55′09.0″N 101°47′04.8″E)

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRX2709728

2. Value of the data

Limited Whole genome data on of *Garcinia mangostana* plant are available.

Additional sequence data on a variety species of *Garcinia mangostana* add important information for genome surveys of the GC content, heterozygosity and genome size.

Genomics data will contribute to genetic studies and crop improvement of this commercially important fruit tree.

3. Data

Genome sequences of *G. mangostana* var. Mesta were generated from DNA extract of young leaf tissues. The short reads were filtered and processed as described in the next section. Raw data for this project were deposited in SRA database with the accession number SRX2709728 (http://www.ncbi.nlm.nih.gov/sra/SRX2709728) under the BioSample accession SAMN06688777.

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 to 5 months old plant were collected in September 2013 and frozen in liquid nitrogen before stored at − 80 °C for DNA extraction.

4.2. DNA extraction and quality control, library preparation and DNA-seq

DNA from leaf samples were extracted using DNeasy Plant mini kit (QIAGEN) based on manufacturer's protocol. Quantity and quality of extracted total DNA were determined using NanoDrop 1000 (Thermo Fisher Scientific Inc., USA) and Agilent 2100 bioanalyzer (Agilent Technologies, USA), respectively.

Sequencing library was prepared using TruSeq DNA Sample Prep kit (Illumina, FC-121-2001/FC-121-2002) with the standard Illumina DNA shotgun library preparation protocol. Briefly, 1 μg of genomic DNA was

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

Attributes	Value
Raw reads	
Total number	497,249,484
Total bases (bp)	50,222,197,884
Filtered reads	
Total number	397,005,854
Total bases (bp)	40,097,591,254
N (%)	0.0016
GC (%)	38.0
Q20 (%)	99.1
Q30 (%)	95.0

fragmented by Covaris M220 Focused-ultrasonicator, end repaired, 3′ end 'A' tailing before Illumina adapter ligation and size selection for 400–500 bp products. PCR amplification was performed as followed: 30 s at 98 °C; 9 cycles of 10 s at 98 °C 30 s at 60 °C, 30 s at 72 °C; and 5 min final extension at 72 °C. Paired end reads of 101 bp was generated through the Illumina HiSeq 2000 sequencing platform by Macrogen, Korea.

4.3. Raw reads processing

Raw reads 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 further analysis. Table 1 shows the sequencing statistics.

Conflict of interest

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

Acknowledgements

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, http://dx.doi.org/10.1016/j.gdata.2015.11.018.

[2] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics (2014) btu170, , http://dx.doi.org/10.1093/bioinformatics/btu170.