



Data Article

Draft genome sequencing data of a pathogenic *Pantoea stewartii* subspecies *stewartii* strain SQT1 causing bronzing disease of jackfruit in Malaysia

Rohaya Ibrahim^a, Noor Wahida Ismail-Suhaimy^a, Tan Shu-Qing^a, Siti Izera Ismail^a, Md Yasin Ina-Salwany^b, Mohd Termizi Yusof^e, Mansor Hakiman^c, Daljit Singh Karam^d, Dzarifah Zulperi^{a,f,*}

^a Department of Plant Protection, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

^b Department of Aquaculture, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

^c Department of Crop Science, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

^d Department of Land Management, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

^e Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

^f Laboratory of Sustainable Resources Management, Institute of Tropical Forestry and Forest Products, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 3 April 2020

Revised 20 April 2020

Accepted 22 April 2020

Available online 30 April 2020

Keywords:

Malaysia

bronzing disease

jackfruit

Pantoea stewartii subspecies *stewartii*

genome sequencing

Illumina Hiseq

virulence factors

T6SS

ABSTRACT

A Gram-negative bacterium, *Pantoea stewartii* subspecies *stewartii* (*P. stewartii* subsp. *stewartii*) has been recognized as the causative agent for jackfruit bronzing disease in Malaysia. Here, we report the whole genome sequencing dataset of *P. stewartii* subsp. *stewartii* strain SQT1 isolated from local infected jackfruit. The paired-end libraries with an insert size of 350 bp was subjected to the Illumina Hiseq 4000, generating a genome size of 4,783,993 bp with a G+C content of 53.7%. A total protein of 4,671 was identified including virulence factors, resistance factors and secretion systems. *Pantoea stewartii* subsp. *stewartii* strain DC283 (NCBI accession no. CP017581.1) was used as a reference genome, where the query hit 72% coverage and average sequencing depth of 68.

* Corresponding author. Dzarifah Zulperi. .

E-mail address: dzarifah@upm.edu.my (D. Zulperi).

In total, 28,717 nucleotide polymorphisms, 520 small insertion/deletions and 142 structure variants were identified. The complete genome was deposited at the European Nucleotide Archive under the sample accession number ERP119356 and study accession number PRJEB36196.

© 2020 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY license.
(<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Molecular Biology
Specific subject area	Microbial genomics
Type of data	Table Figure
How data were acquired	Completed genome sequence in FASTA format
Data format	Illumina Hiseq sequencing platform
Parameters for data collection	Raw and analyzed
Description of data collection	Genomic DNA from bacterial pure culture
Data source location	Bacterial isolation from pure culture, sequencing the whole genome, genome assembly by reference method and annotation City/Town/Region: Muadzam Shah State: Pahang Country: Malaysia
Data accessibility	Repository name: European Nucleotide Archive Data identification number: ERP119356 Direct URL to data: 1) https://www.ebi.ac.uk/ena/browser/view/PRJEB36196 2) https://www.ebi.ac.uk/ena/browser/text-search?query=PRJEB36196
Related research article	Zulperi, D., Manaf, N., Ismail, S. I., Karam, D. S., & Yusof, M. T. (2017). First report of <i>Pantoea stewartii</i> subspecies <i>stewartii</i> causing fruit bronzing of jackfruit (<i>Artocarpus heterophyllus</i>), a new emerging disease in Peninsular Malaysia. <i>Plant Disease</i> , 101(5), 831-831. https://doi.org/10.1094/PDIS-11-16-1689-PDN 1 Ibrahim, R., Ismail-Suhaimy, N. W., Shu-Qing, T., Ismail, S. I., Abidin, N., Hakiman, M., Karam, D. S., Ahmad-Hamdani, M. S., Yusof, M. T. and Zulperi, D. (2019). Molecular characterization and phylogenetic analysis of <i>Pantoea stewartii</i> subspecies <i>stewartii</i> causing bronzing disease of jackfruit in Malaysia based on cps and hrp gene sequences. <i>Journal of Plant Pathology</i> , 1-7. https://doi.org/10.1007/s42161-019-00383-7 .

Value of the Data

The draft genome sequence data of *P. stewartii* subsp. *stewartii* strain SQT1 provides additional information to the available limited number of *P. stewartii* genomes in the public databases and helps for better understanding on bacterial interaction with the host genes resulting in enhanced breeding programs with resistance to jackfruit bronzing disease.

Virulence and antibiotic resistance gene analysis may predict an organism's likelihood of being a multidrug resistance pathogen.

Data can be used as reference to other *P. stewartii* genomes to provide insights into the possible mechanisms of virulence.

1. Data Description

Bronzing disease is a serious problem in Malaysia's jackfruit industry, as it decreases the price of fresh jackfruit and consumer preferences. The bronzing disease causal bacterium, *P. stew-*

Table 1Overview of *Pantoea stewartii* subspecies *stewartii*, strain SQT1.

Attribute	Value
Genome size (bp)	4,783,993
G + C content (%)	53.7
N50	307,002
Contigs	67
Subsystems	517
Coding sequences	4609
RNAs	71

artii subsp. *stewartii* strain SQT1 was isolated from infected jackfruit in Muadzam Shah, Pahang, Malaysia [1,2]. In this data, genomic sequences were generated and assembled before proceeded to annotation process. Annotation by Rapid Annotation using Subsystem Technology (RAST) revealed that the final assembly of the draft genome consists of 4,783,993 bp genome length with 53.7% G+C content, 307,002 N50, and 67 contigs (Table 1).

RAST also disclosed the subsystem coverage, category distribution and features counts of the draft genome as depicted in Figure 1. The subsystem coverage comprised 52% of the protein, with 517 subsystem, 4609 protein coding sequences and 71 RNAs (Table 1). The subsystem feature counts pointed out that 86 of the coding sequences related to virulence, disease and defense. Under the adhesion subdivision, a conserved genomic region-encoded YidE gene was presented and acting as a mediator on hyper-adherence.

The genome organism is found to be tolerant towards colicin E2, a type of bacteriocin that causes DNA breakdown as their modes of action [3]. It also encodes for antimicrobial resistance gene against fluoroquinolone. The genome has a beta-lactamase which is a defining feature of the organism. RAST server also uncovered the resistance towards heavy metals; cobalt-zinc-cadmium, arsenic and copper. The mycobacterium virulence operon was also identified to be responsible for the invasion and intracellular resistance of the bacterial genome.

There were 153 coding sequences (CDS) related to protein secretion systems (Figure 1), and type VI secretion system (T6SS) was identified as the main bacterial secretion machinery. Of 34 T6SS, four VgrG proteins were found on the genome sequences at locations 2657, 2667, 2671 and 3078. Together with Hcp, VgrG protein forms a tube of stacked hexamers in vitro, a hallmark of T6SS [4]. T6SS has been reported in plant pathogen of *Pantoea* species [5–7] and within α , β and γ -proteobacteria are mostly reported [8]. Previous studies have shown that T6SS engages in manipulation of host cell and interbacterial competition by injecting toxins effectors into prokaryotic or eukaryotic cells. The role is not restricted to virulence and antibacterial activity since the system is dispersing among the commensal and pathogenic phytobacteria, thus adding the colonization benefits and fitness *in planta* [9]. T6SS could be a dynamic system for bacterial communication by cell-to cell signalling system [10], since the T6SS components that have been used for destroying cells can be reused for new T6SS assembly [11].

2. Experimental design, materials, and methods

2.1. DNA extraction

Following sample preparation, total genomic DNA of strain SQT1 was extracted using the Presto™ mini gDNA Bacteria Kit (Geneaid Biotech Ltd., Taiwan) following manufacturer's instructions. For the sample quality control, DNA quality was measured using a NanoDrop spectrophotometer (Thermo Scientific, USA) and proceeded to library preparation and sequencing analyses.

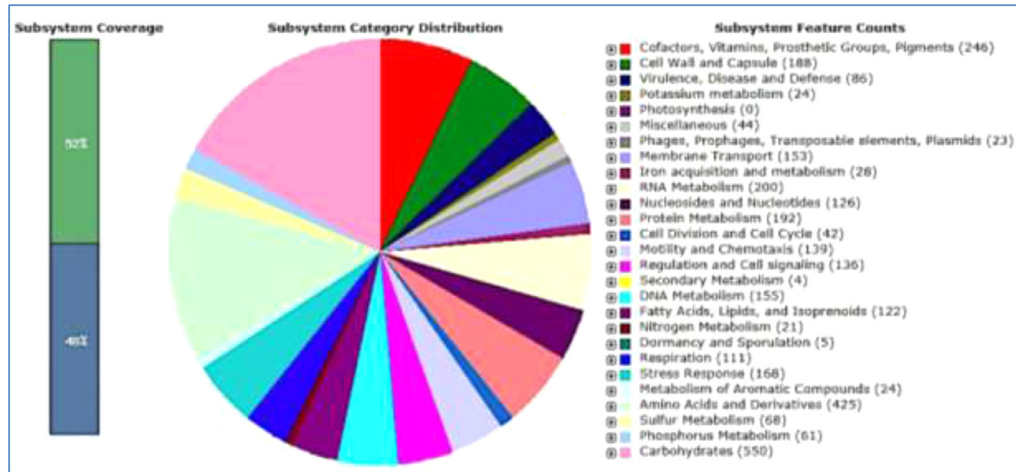


Fig. 1. Subsystem coverage, category distribution and features counts of the *Pantoea stewartii* subspecies *stewartii* strain SQT1.

2.2. Library preparation and sequencing

The method of DNA library was performed according to Wang et al. [12]. Fragmentation of the genomic DNA by random nebulisation was achieved using a paired-end DNA sample preparation kit. The process resulted in double-stranded DNA fragments consisting of 3' or 5' overhangs. The overhangs were transformed to blunt ends, where the 3' to 5' exonuclease activity removed 3' overhangs and the polymerase activity filled in the 5' overhangs. Using the polymerase activity, an 'A' base has been linked to the 3' end. Ligation of the DNA adaptors and the DNA fragments took place prior to purification to remove unligated adaptors, as well as self-ligated adaptors. The products were subjected to quality control and detected the fragments size and yield. The generated library was analysed using Hiseq 4000 platform (Illumina, Inc.) that utilises sequencing-by-synthesis technology.

2.3. Bioinformatics analysis

The sequencing data (raw data) generated from the Illumina pipeline was analysed using CASAVA base calling. Data quality control was carried out where the adapter sequence and low quality reads were cut off and generated clean data, where Burrows-Wheeler Aligner (BWA) [13] was used to align reads to the reference sequence, *P. stewartii* subs. *stewartii* strain DC283, whilst coverage was computed by SAMTOOLS software. The aligned data was stored in BAM files and ready for variant calling analyses.

2.4. Data quality control

Quality control was performed and the adapter and low quality sequences were removed. Also, the reads with unknown bases of greater than 10% were removed. Statistical analysis was conducted afterwards to get the raw FASTQ data as well as clean data. The obtained clean data was used for subsequent analysis.

2.5. SNPs/InDel analysis

Not only detecting individual single nucleotide polymorphisms (SNPs), SAMtools software was also used to detect insertion and deletion (InDel) of small fragments that are less than 50bp. Then, the position of SNP/InDel in the functional regions of the SQT1 genome was annotated.

2.6. SV analysis

The insertion, deletion, inversion, intra-chromosomal translocation, and inter-chromosomal translocation among the reference and sample were performed by BreakDancer software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank staffs from the Bacteriology Laboratory, Department of Plant Protection, Universiti Putra Malaysia and Department of Agriculture of Malaysia for assistance in sampling and isolation procedures. This genome project was supported by Fundamental Research Grant Scheme from Ministry of Education Malaysia (grant 5540090) and UPM-IPM fund (grant 9494600). Rohaya Ibrahim is the recipient of Graduate Research Fellowship from Universiti Putra Malaysia.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.105634.

References

- [1] R. Ibrahim, N.W. Ismail-Suhaimy, T. Shu-Qing, S.I. Ismail, N. Abidin, M. Hakiman, D.S. Karam, M.S. Ahmad-Hamdani, M.T. Yusof, D. Zulperi, Molecular characterization and phylogenetic analysis of *Pantoea stewartii* subspecies *stewartii* causing bronzing disease of jackfruit in Malaysia based on cps and hrp gene sequences, *Journal of Plant Pathology* (2019) 1–7 <https://doi.org/10.1007/s42161-019-00383-7>.
- [2] D. Zulperi, N. Manaf, S.I. Ismail, D.S. Karam, M.T. Yusof, First report of *Pantoea stewartii* subsp. *stewartii* causing fruit bronzing of jackfruit (*Artocarpus heterophyllus*), a new emerging disease in Peninsular Malaysia, *Plant Disease* 101 (5) (2017) 831–831 <https://doi.org/10.1094/PDIS-11-16-1689-PDN>.
- [3] E. Cascales, S.K. Buchanan, D. Duché, C. Kleanthous, R. Lloubes, K. Postle, M. Riley, S. Slatin, D. Cavard, Colicin biology, *Microbiol. Mol. Biol. Rev.* 71 (1) (2007) 158–229 <https://doi.org/10.1128/MMBR.00036-06>.
- [4] E. Cascales, C. Cambillau, Structural biology of type VI secretion systems, *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1592) (2012) 1102–1111 <https://dx.doi.org/10.1098/rstb.2011.0209>.
- [5] D.Y. Shyntum, J. Theron, S.N. Venter, L.N. Moleleki, I.K. Toth, T.A. Coutinho, *Pantoea ananatis* utilizes a type VI secretion system for pathogenesis and bacterial competition, *Molecular Plant-Microbe Interactions* 28 (4) (2015) 420–431 <http://dx.doi.org/10.1094/MPMI-07-14-0219-R>.
- [6] D.Y. Shyntum, S.N. Venter, L.N. Moleleki, I. Toth, T.A. Coutinho, Comparative genomics of type VI secretion systems in strains of *Pantoea ananatis* from different environments, *BMC genomics* 15 (1) (2014) 163 <https://doi.org/10.1186/1471-2164-15-163>.
- [7] Medrano, E. G., & Bell, A. A. (2012). Genome sequence of *Pantoea* sp. strain Sc 1, an opportunistic cotton pathogen. <http://dx.doi.org/10.1128/JB.00450-12>.
- [8] F. Boyer, G. Fichant, J. Berthod, Y. Vandenbrouck, I. Attree, Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC genomics* 10 (1) (2009) 104 <https://doi.org/10.1186/1471-2164-10-104>.
- [9] P. Bernal, M.A. Llamas, A. Filloux, Type VI secretion systems in plant-associated bacteria, *Environmental microbiology* 20 (1) (2018) 1–15 <https://doi.org/10.1111/1462-2920.13956>.
- [10] M. Gallique, M. Bouteiller, A. Merieau, The type VI secretion system: a dynamic system for bacterial communication? *Frontiers in microbiology* 8 (2017) 1454 <https://doi.org/10.3389/fmicb.2017.01454>.
- [11] A. Vettiger, M. Basler, Type VI secretion system substrates are transferred and reused among sister cells, *Cell* 167 (1) (2016) 99–110 <https://doi.org/10.1016/j.cell.2016.08.023>.
- [12] X. Wang, W. Sui, W. Wu, X. Hou, M. Ou, Y. Xiang, Y. Dai, Whole-genome resequencing of 100 healthy individuals using DNA pooling, *Experimental and therapeutic medicine* 12 (5) (2016) 3143–3150 <https://doi.org/10.3892/etm.2016.3797>.
- [13] A.E. Minoche, J.C. Dohm, H. Himmelbauer, Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems, *Genome biology* 12 (11) (2011) R112 <https://doi.org/10.1186/gb-2011-12-11-r112>.