Whole-genome shotgun sequence of phenazine-producing endophytic
Streptomyces kebangsaanensis SUK12

Juwairiah Remali\textsuperscript{a}, Kok-Keong Loke\textsuperscript{b}, Chyan Leong Ng\textsuperscript{b}, Wan Mohd Aizat\textsuperscript{b}, John Tiong\textsuperscript{c}, Noraziah Mohamad Zin\textsuperscript{d,}\textsuperscript{*}

\textsuperscript{a} School of Bioscience and Biotechnology, Faculty of Science and Technology, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
\textsuperscript{b} Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
\textsuperscript{c} School of Pharmacy, Taylor’s University, No. 1, Jalan Taylor’s, 47500 Subang Jaya, Selangor, Malaysia
\textsuperscript{d} School of Diagnostic and Applied Health Sciences, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia

ARTICLE INFO
Keywords:
Genome
Phenazine
Streptomyces kebangsaanensis SUK12
Antibiotics

ABSTRACT
Streptomyces sp. produces bioactive compounds with a broad spectrum of activities. Streptomyces kebangsaanensis SUK12 has been identified as a novel endophytic bacteria isolated from ethnomedicinal plant Portulaca oleracea, and was found to produce the phenazine class of biologically active antimicrobial metabolites. The potential use of the phenazines has led to our research interest in determining the genome sequence of Streptomyces kebangsaanensis SUK12. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number PRJNA269542. The raw sequence data are available [https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP105770].

Specifications

Organism/cell line/tissue
Streptomyces kebangsaanensis

Strain
SUK12\textsuperscript{T} (= DSM 42048\textsuperscript{T} = NRRL B-24860\textsuperscript{T})

Sex
N/A

Sequencer or array type
Illumina HiSeq 2000

Data format
Raw and processed

Experimental factors
S. kebangsaanensis SUK12 was cultured on International Streptomyces Project-2 (ISP-2) agar for 7 days. Genomic DNA from agar culture was isolated as previously described [1]. Sequencing was performed according to Illumina specific protocols for library preparation and DNA-seq.

Experimental features

Consent
N/A

Sample source location
Streptomyces kebangsaanensis SUK12 has been isolated from Portulaca oleracea L. plant collected from the Nenasi Reserve Forest, Pahang, Malaysia (2° 53.852’ N 103° 25.507’ E) [2].

1. Direct link to deposited data
Data have been deposited in repository [https://www.ncbi.nlm.nih.gov/bioproject/269542] and raw sequence files (s_CGATGT_1.fastq.gz and s_CGATGT_2.fastq.gz) are also available under the accession number SRP105770 ([https://www.ncbi.nlm.nih.gov/sra/?term=SRP105770]).

2. Introduction
In this 21st century, bacterial resistance towards antibiotics is one of the most challenging problems in the medical field. As pathogenic bacteria are increasingly resistant to antibiotics, newer sources of such compounds need to be investigated [3]. It is well accepted that the Streptomyces have a great capability to produce diverse bioactive compounds that have a wide spectrum of activity. Streptomyces kebangsaanensis SUK12, an endophytic bacteria isolated from plant Portulaca oleracea was found to contain a phenazine antibiotic [2]. The first phenazine isolated from Streptomyces griseoluteus was an antibiotic called griseolutein [4]. In recent years, Streptomyces sp. has been a rich source of several diverse and complex phenazine. These include lomofungin from Streptomyces lomondensis [5] and endophenazines from Streptomyces anulatus [6,7]. Compared to Pseudomonas, Streptomyces species are considered to be more promising sources for bioactive compounds from

* Corresponding author.
E-mail addresses: kloke@ukm.edu.my (K.-K. Loke), clng@ukm.edu.com (C.L. Ng), wma@ukm.edu.my (W.M. Aizat), noraziah.zin@ukm.edu.my (N.M. Zin).

http://dx.doi.org/10.1016/j.gdata.2017.05.015
Received 5 April 2017; Received in revised form 17 May 2017; Accepted 21 May 2017
Available online 24 May 2017
phenazine owing to their structural complexity [8]. Thus, the genome sequence of *S. kebangsaanensis* SUK12 would provide an entry point to elucidate the antibiotic-producing capability of this *Streptomyces* species to identify new sources of novel phenazine based antibacterial compounds. The data set will be valuable for the scientific community working in the area of bacteriology, molecular biology, drug discovery and other related researches. Plus, it also would provide impetus for research involving the development of important genetic tools which can optimize the yield of useful antibiotic(s) and/or other secondary metabolites whilst facilitating the development of new antibiotic analogues via semi-synthetic approach.

3. Experimental design, materials and methods

3.1. Library preparation and sequencing

The strain was obtained from the stock culture of Novel Antibiotic Research Laboratory, Universiti Kebangsaan Malaysia (UKM). Genomic DNA extraction was performed according to Kieser et al. [1] with slight modifications. For library preparation, genomic DNA was fragmented with the targeted size 400–600 bp using Covaris S220 (Covaris Inc., USA). The selected sized DNA was then ligated to Illumina TruSeq adapters and amplified using TruSeq DNA Sample Preparation Kit (Illumina, USA). Quantification was carried out using KAPA kit (KAPA Biosystem, USA) on Agilent Stratagene Mx-3005p quantitative PCR machine (Agilent, USA). Library size was then verified via Agilent Bioanalyzer High Sensitivity DNA Chip. Whole genome sequencing finally was performed using Illumina Genome Analyzer (Illumina, San Diego, CA).

3.2. Preprocessing and genome assembly

In order to evaluate read quality before and after pre-processing, FastQC assessment plots were generated for sequenced reads using Babraham Bioinformatics programme (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The filtered reads were then assembled using an assembler pipeline called MaSuRCA [9]. Both paired-end and single reads were utilized in the assembly process. 560 assembled contigs were further scaffolded using paired-end library information which forms 170 scaffolds.

4. Data description

The presented data summarizes information computed from DNA library preparation (Table 1), the raw DNA sequence reads (Table 2), preprocessing (Fig. 1) as well as assembly and scaffolding (Table 3). The data shown corresponds to the filenames and obtained sequence content following the de-multiplexing of the data, and following the data conversion to the FASTQ format (Table 1). The sequencing produces 2.6 Gbp raw reads (Table 2). These reads were filtered to remove sequences with ambiguous bases, duplicated reads as well as adapter bases (Fig. 1). FastQC assessment plots were generated for sequenced reads to evaluate read quality before (Fig. 1A) and after preprocessing (Fig. 1B). De novo assembly yielded a total of 560 contigs and 170 scaffolds (Table 3). The longest scaffold has 453,879 bp in length and the shortest length in 1072 bp (Table 3).

| Sample QC info                    | Final library | Multiplexing |
|-----------------------------------|---------------|--------------|
| Vol. (μl) | Conc. (ng/μl) | Total conc. (μg) | Library conc. (nM) | Average base pair (bp) | Size range (bp) | Adapter Index |
| 60 | 31.8 | 1.9 | 15.83 | 590 | 448-921 | 2 | CGATGT |

Table 2
Overall raw read information of paired-end reads for *S. kebangsaanensis*.

| Strain           | Adaptor | Average sample size (bp) | Total raw reads (bp) | Read length (bp) | Total bases (bp) | Sequence files |
|------------------|---------|--------------------------|----------------------|-----------------|----------------|----------------|
| *S. kebangsaanensis* | CGATGT | 482 | 13,112,394 | 101 | 1,324,351,794 | s_CGATGT_1.fastq.gz |
|                  |         |                          | 13,112,394           |                 | 1,324,351,794   | s_CGATGT_2.fastq.gz |
Fig. 1. Per base quality score distribution raw reads of *S. kebungsasensis* sample before (A) and after (B) preprocessing (generated by FastQC).
Acknowledgements

Financial support was obtained from the Ministry of Higher Education of Malaysia under the grant number ERGS/1/2013/SKK04/UKM/02/2. We would like to thank MGRC for the genome sequencing service.

References

[1] T. Kieser, M.J. Bibb, M.J. Buttner, D.A. Hopwood, Practical Streptomyces Genetics, John Innes Cent Ltd., 2000.

[2] N.I.M. Sarmin, G.Y.A. Tan, C.M.M. Franco, R. Edrada-Ebel, J. Latip, N.M. Zin, Streptomyces kebangsaanensis sp. nov., an endophytic actinomycete isolated from an ethnomedicinal plant, which produces phenazine-1-carboxylic acid, Int. J. Syst. Evol. Microbiol. 63 (2013) 3733–3738.

[3] G.D. Wright, Antibiotics: A new hope, Chem. Biol. 19 (2012) 3–10.

[4] H. Umezawa, S. Hayano, K. Maeda, On a new antibiotic, griseolutein, produced by Streptomyces, Jpn. Med. 3 (1950) 111–117.

[5] L.E. Johnson, A. Dietz, Lomofungin, a new antibiotic produced by Streptomyces lomondensis sp. n, Appl. Microbiol. 17 (1969) 755–759.

[6] K. Gebhardt, J. Schimana, P. Krastel, K. Dettner, J. Rheinheimer, A. Zeeck, H.-P. Fiedler, Endophenazines A–D, new phenazine antibiotics from the arthropod associated endosymbiont Streptomyces anulatus. I. Taxonomy, fermentation, isolation and biological activities, J. Antibiot. (Tokyo) 55 (2002) 794–800.

[7] P. Krastel, A. Zeeck, Endophenazines A similar to D, new phenazine antibiotics from the arthropod associated endosymbiont Streptomyces anulatus II. Structure elucidation, J. Antibiot. (Tokyo) 55 (2002) 801–806.

[8] J. Laursen, J. Nielsen, Phenazine natural products: biosynthesis, synthetic analogues, and biological activity, Chem. Rev. 104 (2004) 1663–1686.

[9] A.V. Zimin, G. Marçais, D. Puiu, M. Roberts, S.L. Salzberg, J.A. Yorke, The MaSuRCA genome assembler, Bioinformatics 29 (2013) 2669–2677.

Table 3
Statistics of contigs and scaffolds assembled for S. kebangsaanensis.

|                  | Sequences (n) | Bases (nt) | Length range (nt) | Mean length (nt) | N50 (nt) |
|------------------|---------------|-----------|-------------------|------------------|---------|
| Clean reads      | 15,474,727    | 1,483,526,984 | 75–101           | –                | –       |
| Assembled contigs| 560           | 8,308,573  | 124–78,384        | 14,837           | 24,540  |
| Assembled scaffolds | 170       | 8,328,719  | 1072–453,879      | 48,992           | 110,454 |