Genomics Data 14 (2017) 44–46

Contents lists available at ScienceDirect

Genomics Data

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

# Draft genome sequence of an endophytic bacterium, *Paenibacillus tyrfis* strain SUK123, isolated from *Santiria apiculata* stem

Emmanuel Haruna[a,b], Noraziah M. Zin[a,*], Jonathan M. Adams[c,*]

[a] *Programme of Biomedical Science, School of Diagnostic and Applied Health Sciences, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia*

[b] *Department of Biochemistry Kaduna State University, Kaduna, Nigeria*

[c] *Division of Agrifood and Environment, School of Water, Energy and Environment, Building 52a, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK*

## ABSTRACT

Here we report the draft genome sequence of an endophytic *Paenibacillus tyrfis* strain isolated from the Universiti Kebangsaan Malaysia reserve forest, Malaysia. The genome size was approximately 8.04 Mb, and the assembly consisted of 107 scaffolds with 168 contigs, and had a G + C content of 53%. Phylogenetic analysis of strain SUK123 using the 16S rRNA gene revealed that it belonged to the family *Paenibacillaceae* with the highest similarity to *Paenibacillus elgii* SD$^T$ (99%). Whole genome comparison of SUK123 with related species using average nucleotide identity (ANI) analysis revealed a similarity of 98% to *Paenibacillus tyrfis* Mst1$^T$, 94% to *Paenibacillus elgii* B69$^T$, 91% to *Paenibacillus ehimensis* A2$^T$, 68% to *Paenibacillus polymyxa* SC2$^T$ and 69% to *Paenibacillus alvei* DMS29$^T$. The draft genome was deposited at the European Nucleotide Archive (PRJEB21373).

| Specifications | |
|---|---|
| Organism/cell line/tissue | *Paenibacillus tyrfis* |
| Sex | Not applicable |
| Sequencer or array type | Illumina Miseq |
| Data format | Raw data and analyzed i.e. assembled |
| Experimental factors | *Paenibacillus tyrfis strain was isolated from Santiria apiculata stem* |
| Experimental features | Isolation of bacteria, Genome sequencing, de novo assembly |
| Consent | Not applicable |
| Sample source location | *Universiti Kebangsaan Malaysia reserve forest, Latitude & Longitude* – 2.9125 & 101.7872 |

## 1. Direct link to deposited data

http://www.ebi.ac.uk/ena/data/view/PRJEB21373.

## 2. Introduction

The bacterial genus *Paenibacillus* has been isolated from many environments, mostly from plant organs and their surrounding soil [1]. The species of *Paenibacillus* was included in the genus *Bacillus* until 1993 when it was proposed that the member of the "16S rRNA group 3" bacilli be distinguished from it [2]. Presently, the genus *Paenibacillus* consists of 395 known species [3]. Most members of the genus *Paenibacillus* are gram variable, either aerobic or facultatively anaerobic, rod-shaped, and endospore-forming with peritrichous flagella for motility [4]. The DNA G + C content of the various species of *Paenibacillus* ranges between 39 and 54 mol% whilst anteiso-$C_{15:0}$ is the major cellular fatty acid and *meso*-diaminopimelic acid is the cell wall peptidoglycan diamino acid [5]. Most members of this genus have been reported to be producers of either active antimicrobial metabolites such as lipopeptides [6,7], plant-growth promoting hormones [8,9], or insecticides [10].

## 3. Experimental design, materials and methods

The endophytic *Paenibacillus tyrfis*, strain SUK123, was isolated from

* Corresponding authors.
*E-mail addresses:* noraziah.zin@ukm.edu.my (N.M. Zin), j.m.adams@cranfield.ac.uk (J.M. Adams).

http://dx.doi.org/10.1016/j.gdata.2017.08.005
Received 11 July 2017; Received in revised form 9 August 2017; Accepted 11 August 2017
Available online 17 August 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 assembled sequence length.

| | |
|---|---|
| No. of all scaffolds | 107 |
| Bases in all scaffolds | 8,041,385 |
| No. of large scaffolds (> 1000 bp) | 73 |
| Bases in large scaffolds | 8,024,609 |
| Largest length | 505,699 |
| Scaffold N50 | 257,089 |
| Scaffold N90 | 62,404 |
| G + C content (%) | 53.061 |
| No. of all contigs | 168 |
| Bases in all contigs | 8,041,324 |
| No. of large contigs (> 1000 bp) | 114 |
| Bases in large contigs | 8,013,898 |
| Largest length | 501,150 |
| Contig N50 | 197,909 |
| Contig N90 | 48,401 |

the *Santiria apiculata* stem located at the Universiti Kebangsaan Malaysia forest reserve, whilst screening for endophytes with antimicrobial potential against ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter species*) pathogens. After sample collection, the stem surface was sterilized to avoid surface-bound epiphytes as described previously [11]. The sterilized stem was cut into a 2 cm section and plated on water agar plate medium from where SUK123 (an antibiotic producing bacterium) was isolated.

Endophytic *Paenibacillus tyrfis* DNA was extracted using an Ultraclean Microbial DNA Isolation Kit as described by the manufacturer (Mo Bio Laboratories, 2746 Loker Ave W # A, Carlsbad, CA 92010, USA). The bacterium was identified by sequencing its 16S rRNA gene. Sequence was performed on Illumina Miseq platform (Majorbio, China) by 2 × 400 bp paired-end libraries. The raw sequence quality was assessed using PRINSEQ lite version 0.20.4 and the genome was assembled using SOAPdenovo v2.04 with all parameters set by default [12] and GapCloser v1.12 was used to fill local inner gaps. The open reading frames (ORFs) were predicted using Glimmer 3.02 [13]. The biological functions of these predicted ORFs were annotated using various databases.

## 4. Data description

The analyses of the assembled genome revealed a genome size of about 8,041,385 bp made of 107 scaffolds, 168 contigs, and with a G + C content of 53.06% and N50 contigs size of 197,909 (Table 1). The genome was annotated using Rapid Annotation Subsystems Technology server [14].

The RAST server prediction revealed 7368 coding sequences (CDS) with a total of 2880 CDS (40%) constituting 2747 and 133 of non-hypothetical and hypothetical proteins respectively in the subsystem coverage. A total of 4488 CDS (60%) comprising 2003 and 2485 of non-hypothetical and hypothetical proteins respectively were outside the subsystem coverage (Fig. 1).

The average nucleotide identity (ANI) [15] analysis of SUK123 with closely related species, revealed a similarity index with *Paenibacillus tyrfis* Mst1$^T$ (98.06%), *Paenibacillus elgii* B69$^T$ (93.98%), *Paenibacillus ehimensis* A2$^T$ (91.17%), *Paenibacillus alvei* DMS29$^T$ (68.81%) and *Paenibacillus polymyxa* SC2$^T$ (68.27%), suggesting our isolate most closely related to *Paenibacillus tyrfis* Mst1$^T$.

## Declaration of interest

We declare no conflict of interest with respect to this article titled '**Draft genome sequence of an endophytic *Paenibacillus tyrfis* strain SUK123 isolated from *Santiria apiculata* stem**' submitted to Genomics data journal.

## Acknowledgement

We appreciate University Kebangsaan University Forest Reserve and Herbarium staff and grateful to Muhanna Al-Shabanni, Aishah Ismail, Radhiah Binti Khairon, and Nur Faizah Abu Bakar, for their assistance during sampling. This work was funded by the Universiti Kebangsaan Malaysia Research Grant (GUP-2015-042).

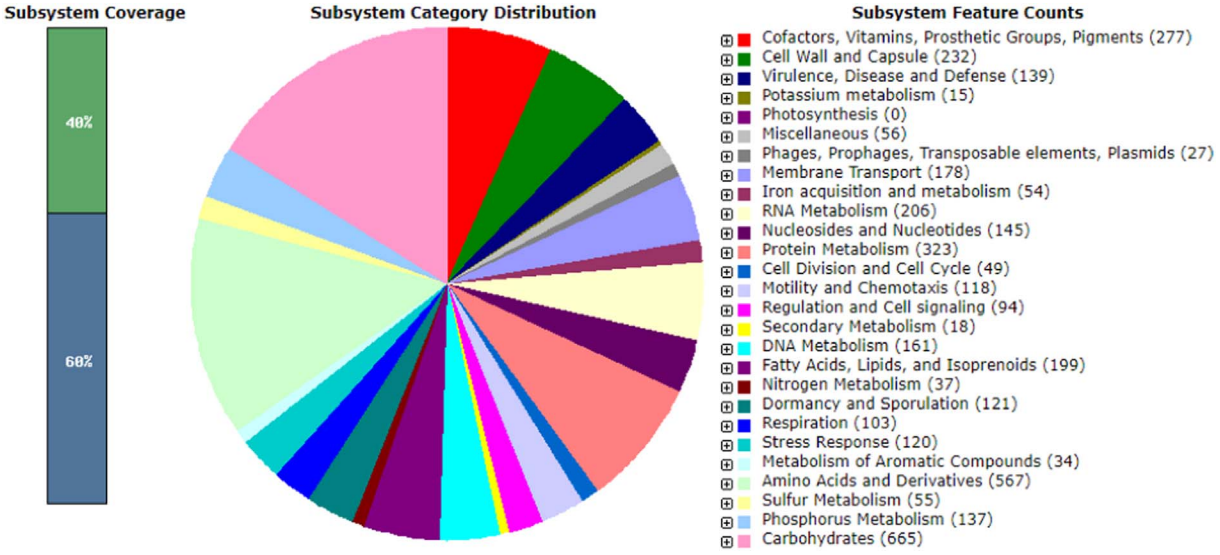


**Fig. 1.** A diagrammatic representation of *Paenibacillus tyrfis* strain genome subsystem coverage as annotated by RAST.

## References


[1] E.N. Grady, J. MacDonald, L. Liu, A. Richman, Z.-C. Yuan, E. Novelli, Current knowledge and perspectives of *Paenibacillus*: a review, Microb. Cell Factories 15 (2016) 203, http://dx.doi.org/10.1186/s12934-016-0603-7.

[2] C. Ash, F.G. Priest, M.D. Collins, Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*, Antonie Van Leeuwenhoek 64 (1993) 253–260.

[3] S.-H. Yoon, S.-M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun, Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies, Int. J. Syst. Evol. Microbiol. 67 (2017) 1613–1617, http://dx.doi.org/10.1099/ijsem.0.001755.

[4] Y.-K. Aw, K.-S. Ong, L.-H. Lee, Y.-L. Cheow, C.M. Yule, S.-M. Lee, Newly isolated *Paenibacillus tyrfis* sp. nov., from Malaysian tropical peat swamp soil with broad spectrum antimicrobial activity, Front. Microbiol. 7 (2016) 219, http://dx.doi.org/10.3389/fmicb.2016.00219.

[5] O. Shida, H. Takagi, K. Kadowaki, L.K. Nakamura, K. Komagata, Transfer of *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*, Int. J. Syst. Bacteriol. 47 (1997) 289–298, http://dx.doi.org/10.1099/00207713-47-2-289.

[6] Z. Huang, Y. Hu, L. Shou, M. Song, Isolation and partial characterization of cyclic lipopeptide antibiotics produced by *Paenibacillus ehimensis* B7, BMC Microbiol. 13 (2013) 87, http://dx.doi.org/10.1186/1471-2180-13-87.

[7] G. Aktuganov, A. Melentjev, N. Galimzianova, E. Khalikova, T. Korpela, P. Susi, Wide-range antifungal antagonism of *Paenibacillus ehimensis* IB-X-b and its dependence on chitinase and beta-1,3-glucanase production, Can. J. Microbiol. 54 (2008) 577–587, http://dx.doi.org/10.1139/w08-043.

[8] K. Zhou, M. Yamagishi, M. Osaki, *Paenibacillus* BRF-1 has biocontrol ability against *Phialophora gregata* disease and promotes soybean growth, Soil Sci. Plant Nutr. 54 (2008) 870–875, http://dx.doi.org/10.1111/j.1747-0765.2008.00308.x.

[9] B. Pichard, D. Thouvenot, Effect of *Bacillus polymyxa* seed treatments on control of black-rot and damping-off of cauliflower, Seed Sci. Technol. 27 (1999) 445–465.

[10] S. Neung, X. Nguyen, K. Naing, Y. Lee, Insecticidal potential of *Paenibacillus elgii* HOA73 and its combination with organic sulfur pesticide on diamondback moth, *Plutella xylostella*, Appl. Biol. 55 (2014) 181–186.

[11] E. Haruna, N.M. Zin, D. Kerfahi, J.M. Adams, Extensive overlap of tropical rainforest bacterial endophytes between soil, plant parts, and plant species, Microb. Ecol. (2017) 1–16, http://dx.doi.org/10.1007/s00248-017-1002-2.

[12] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, Gigascience 1 (2012) 18.

[13] A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg, Identifying Bacterial Genes and Endosymbiont DNA With Glimmer, 23 (2007), pp. 673–679, http://dx.doi.org/10.1093/bioinformatics/btm009.

[14] R. Aziz, D. Bartels, A. Best, The RAST Server: rapid annotations using subsystems technology, BMC 9 (2008) 75.

[15] S.-H. Yoon, S. Ha, J. Lim, S. Kwon, J. Chun, A large-scale evaluation of algorithms to calculate average nucleotide identity, Antonie Van Leeuwenhoek (2017) 1–6, http://dx.doi.org/10.1007/s10482-017-0844-4.

https://dspace.org/

# Draft genome sequence of an endophytic bacterium, Paenibacillus tyrfis strain SUK123, isolated from Santiria apiculata stem

**Haruna, Emmanuel; Zin, Noraziah M.; Adams, Jonathan M.**

**2017-07-17**
Elsevier
https://doi.org/10.1016/j.gdata.2017.08.005
Emmanuel Haruna, Noraziah M. Zin, Jonathan M. Adams, Draft genome sequence of an endophytic bacterium, Paenibacillus tyrfis strain SUK123, isolated from Santiria apiculata stem, Genomics Data, Volume 14, December 2017, Pages 44-46

Downloaded from a DSpace repository. https://dspace.org/