Data in Brief 57 (2024) 110807

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Analysis of the genome of *Bacillus safensis* strain WOB3 KX774195, a Linamarin-utilizing bacterium (LUB) isolated from Cassava wastewater (CWW), Lagos State, Nigeria

Adewale K. Ogunyemi[a,b,*], Olanike M. Buraimoh[c,e], Wadzani P. Dauda[f], Olufunmilayo O. Akapo[g], Bukola C. Ogunyemi[d], Titilola A. Samuel[d,e], Matthew O. Ilori[c], Olukayode O. Amund[c]

[a] *Department of Microbiology, Trinity University, Yaba, Lagos State, Nigeria*
[b] *Department of Biological Sciences (Microbiology Unit), Lagos State University of Science & Technology, Ikorodu, Lagos State, Nigeria*
[c] *Department of Microbiology, University of Lagos, Akoka, Lagos State, Nigeria*
[d] *Department of Biochemistry, University of Lagos, Id-Araba, Lagos State, Nigeria*
[e] *TETFund Centre of Excellence on Biodiversity Conservation and Ecosystem Management (TCEBCEM), University of Lagos, Akoka, Lagos State, Nigeria*
[f] *Department of Agronomy (Crop Science Unit), Federal University Gashua, Gashua, Yobe State, Nigeria*
[g] *Department of Biochemistry and Microbiology, University of Zululand, KwaDlangezwa Main Campus, KwaDlangezwa, 3886, South Africa*

ARTICLE INFO

Article history:
Received 23 February 2024
Revised 8 May 2024
Accepted 2 August 2024
Available online 13 August 2024

Dataset link: Analysis of the genome of Bacillus safensis, a Linamarin-utilizing Bacterium (LUB) isolated from Cassava wastewater (CWW), Lagos-State, Nigeria (Original data)

ABSTRACT

Linamarin-utilizing bacterium (LUB) is a microorganism that uses and breaks down cassava's principal cyanogenic compound, linamarin. Here, we present the draft genome sequence of *Bacillus safensis* strain WOB3 (previously *Bacillus pumilus* strain WOB3) sequenced and assembled with a total reads of 8,750,054 bp. The genome has 1,269 contigs and, G+C content of 41.55%. The genome has 4,749 total genes, 4,614 protein-coding sequences (CDSs), 3, 8 and 10 rRNA genes, 74 tRNA genes, and 5 ncRNA genes. This whole genome shotgun project has been deposited in GenBank under accession number JAYSGU000000000

* Corresponding author.
E-mail address: adewale.ogunyemi@trinityuniversity.edu.ng (A.K. Ogunyemi).
Social media: @K24619Adewale (A.K. Ogunyemi), @milori (M.O. Ilori)

https://doi.org/10.1016/j.dib.2024.110807

Keywords:
Bacillus safensis
Cassava wastewater (CWW)
Whole genome sequence
Linamarin
Linamarin-utilizing bacterium

Specifications Table

Subject area	Biology
More specific subject area	Microbiology, Microbial Genomics, Molecular Biology
Type of data	Whole genomic sequence data represented by Tables and Figures
How data was acquired	The complete genome sequence was determined using the Illumina HiSeq4000X platforms
Data format	Raw and analyzed
Parameters for data collection	Pure culture of *Bacillus safensis* WOB3 was grown in nutrient agar (NA) at a temperature of 37°C and pH of 7.0
Description of data collection	The genomic DNA was sequenced and annotation was done by using Prokaryotic Genome Annotation Pipeline (PGAP)
Experimental features	Raw sequence reads generated using Illumina MiHiSeq4000X platform
Data source location	Cassava wastewater (CWW) samples were collected from a processing factory, in Odogunyan, Ikorodu, Lagos State, Nigeria
Data accessibility	The complete genome sequence of *Bacillus safensis* WOB3 was deposited in NCBI GenBank under accession JAYSGU000000000 Direct URLto data:numberhttps://www.ncbi.nlm.nih.gov/nuccore/JAYSGU000000000.1 Database link: Bioproject: PRJNA1043453 Biosample: SAMN38323292

1. Value of the Data

- This study reports the genomic insights of *Bacillus safensis* strain WOB3 KX774195, a linamarin-utilizing bacterium (LUB), isolated from cassava wastewater in Odogunyan, Ikorodu, Lagos State, Nigeria, using whole-genome sequencing.
- The data provides important information about the detoxification of cyanogens with valuable insights on the prospects of linamarase and other associated genes.
- These genomic data provide information on *Bacillus safensis*' metabolic strategies for detoxifying cassava wastewater.
- Data of the draft genome sequencing of a *Bacillus safensis is* available for download without restrictions. This data can benefit bioinformaticians and environmental microbiologists as it can be used as reference sample or for testing. Moreover, it can also be used for educational purposes.
- The linamarin residues present in cassava products are causing severe environmental problems in many cassava processing plants in Nigeria, the largest cassava producer worldwide. These residues are a significant concern, given cassava products' critical role in the human diet.
- The study's data will help policymakers develop environmentally friendly policies and mitigate cyanogen pollution.

2. Background

Linamarin occurs naturally in cassava and other plants [1,2]. As linamarin is metabolized, hydrogen cyanide (HCN), a toxic compound, is released [2,3]. A linamarin-utilizing bacterium is an organism that utilizes and degrades linamarin, a major cyanogenic compound in cassava. Microbes that degrade linamarin use it as a sole carbon source by breaking down the com-

pound [4]. They possess specific enzymes such as linamarase, which can hydrolyze linamarin to produce hydrogen cyanide and glucose. The release of hydrogen cyanide during the degradation process is a potential hazard, as it is highly toxic to most living organisms [4]. Researchers have identified bacterial strains like *Bacillus*, *Pseudomonas* and *Stenotrophomonas* as linamarin degraders, demonstrating their successful use of linamarin [5]. *Bacillus* is a genus that belongs to the phylum Firmicutes, with diverse bacterial species that are Gram-positive, rod-shaped, and spore formers [6]. *Bacillus* species are ubiquitous and have been isolated from numerous environments such as plants, animals, freshwater, and soil [7]. Some strains of *Bacillus* genus promote growth of different plants through various mechanisms, such as biofertilization, increasing accessibility of primary nutrients such as sulphur, potassium, phosphorus, nitrogen, iron, aromatic compounds for the strain, virulence (associated with disease and defense), dormancy (with sporulation), RNA and DNA metabolism, stress response, miscellaneous, through the production of phytohormones such as indole acetic acid (IAA), auxin and ethylene, as well as biocontrol by production of antimicrobial metabolites [8-10]. In addition, *Bacillus* species can form spores, an advantage that allows this group of bacteria to survive in unfavorable conditions [11]. This work aims to establish data on the genome sequence of *Bacillus safensis* WOB3 and its linamarase system to provide genomic information for further evaluations regarding detoxification of cyanogens.

3. Data Description

Bacillus safensis WOB3 was isolated from CWW collected from a processing factory, Odogunyan, Ikorodu, Lagos State, Nigeria. This study presents the draft whole genome sequence of *Bacillus safensis* WOB3. The genome sequencing was performed using the Illumina Hiseq4000X platform. The assembled genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (version 6.6), with default parameters [12]. The results show that the draft genome sequence of *Bacillus safensis* strain WOB3 (previously *Bacillus pumilus* strain WOB3) was sequenced and assembled with 8,750,054 bp of total reads. There are 1,269 contigs in the strain WOB3 genome, which has a G+C content of 41.55%. The strain WOB3 genome has a genomic size of 4.0 x 10^6 bp (4.0 Mb) with a corresponding genome coverage of 328X. Furthermore, strain WOB3 has 4,749 total genes, 4,614 protein-coding sequences (CDSs), 3, 8 and 10 rRNA genes, 74 tRNA genes, and 5 ncRNA genes. The assembly statistics and genomic features of *Bacillus safensis* WOB3 were summarized in Table 1. *Bacillus safensis* WOB3 whole genome sequence was used

Table 1
Genome features of *Bacillus safensis* strain WOB3.

Attribute	Value
Total reads	8,750,054
Genome coverage (X)	328
Genome size (Mb)	4.0
CDSs (total)	4,649
CDSs (with protein)	4,614
CDSs (without protein)	35
Genes (total)	4,749
Number of contigs	1,269
GC content (%)	41.55
Contig N50 (kb)	299.4
Contig L50	5
tRNA	74
rRNA	3,8,10
ncRNA	5
Accession number	JAYSGU000000000

CDS- coding sequence; GC - guanine-cytosine content; tRNA-transfer RNAs; rRNA- ribosomal RNA; ncRNA- non-coding RNAs.

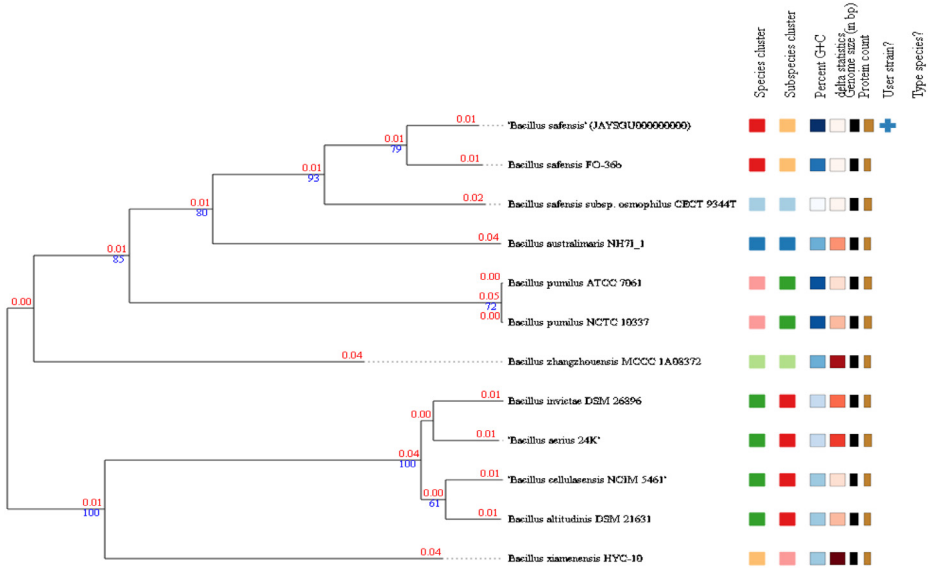

Fig. 1. Whole genome phylogenetic tree constructed by Type Strain Genome Server, using Maximum Likelihood Method based on Generalised Time Reversible (GTR) model. The tree shows the close relationship between *Bacillus safensis* WOB3 with the closed species.

Table 2
Comparison of several *Bacillus* isolates based on genomic metrics including digital DNA-DNA hybridization (dDDH).

Subject strain	dDDH (d4, in %)	C.I. (d4, in %)	dDDH (d6, in %)	C.I. (d6, in %)	G+C content difference (in %)
Bacillus safensis FO-36b	84.2	[81.4 - 86.6]	88	[85.1 - 90.5]	0.26
Bacillus safensis subsp. osmophilus CECT 9344T	68.4	[65.4 - 71.2]	79.1	[75.7 - 82.2]	0.98
Bacillus australimaris NH7I 1	52	[49.4 - 54.7]	78.2	[74.8 - 81.3]	0.52
Bacillus pumilus NCTC 10337	44.7	[42.1 - 47.3]	73.5	[70.0 - 76.7]	0.15
Bacillus pumilus ATCC 7061	44.6	[42.1 - 47.2]	73.2	[69.7 - 76.4]	0.2
Bacillus zhangzhouensis MCCC 1A08372	42.1	[39.6 - 44.6]	70.6	[67.1 - 73.8]	0.49
Bacillus xiamenensis HYC-10	37.4	[35.0 - 39.9]	64.4	[61.0 - 67.6]	0.57
Bacillus aerius 24K	36.6	[34.1 - 39.1]	67.7	[64.3 - 70.9]	0.66
Bacillus invictae DSM 26896	36.5	[34.1 - 39.0]	69.2	[65.7 - 72.4]	0.76
Bacillus cellulasensis NCIM 5461	36.3	[33.9 - 38.8]	72	[68.5 - 75.2]	0.54
Bacillus altitudinis DSM 21631	36.3	[33.9 - 38.8]	70.8	[67.4 - 74.1]	0.6

C.I-Confidence interval.

to construct an accurate evolutionary relationship with other bacterial whole genomes closely related to *Bacillus safensis* species using the Type Strain Genome Server (TYGS) [13]. Fig. 1 shows that *Bacillus safensiis* WOB3 is closely related to *Bacillus safensis* FO-36b and forms a clade with *Bacillus safensis*. Tables 2 and 3 give further information on Genome Blast Distance (GBDP) phylogeny of *Bacillus safensis* WOB3.

To confirm the phylogenetic relationship of WOB3, Digital DNA-DNA hybridization (dDDH) values between *Bacillus safensis* and closely related species were calculated by TYGS. In Table 2, *Bacillus safensis* FO-36b had a while *Bacillus safensis* subsp. *osmophilus* CECT 9344T had 68,4% (value that fall within the species boundary value) [13], indicating the consistency of the phylogenetic relationship of *Bacillus safensis* WOB3.

Table 3
GBDP phylogeny based on genome data.

Undelying genomes	Data type	Distance formula	Distance algorithm	%G+C	δ-statistics	Genome size (bp)	N of proteins	SSU length (bp)	Finished time
1155, 11802, 17444, 17597,18..	genome	D5	Greedy With Trimming	40.88-41.87	0.027-0.054	3,611,490-4,022,069	3,668-4,863	1,058-1,557	2024-D5 - 06134437+0200

N-Number; bp-base pair; GG+C-Guanine+cysteine content; GBDP- Genome Blast Distance Phylogeny approach; SSU-Small subunit ribosomal RNA.
Generated from https://tygs.dsmz.de/user_phylogenies/230581?guid=ef701ce9-fccc-47b4-be50-e4e88002d9c2.

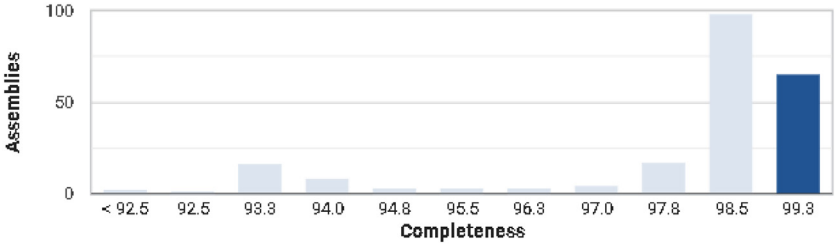

Fig. 2. CheckM analysis (v1.2.2) of completeness of *Bacillus safensis* RefSeq assemblies. calculated on the Prokaryotic Genome Annotation Pipeline (PGAP) gene set with the *Bacillus* CheckM marker set.

Bacillus is a distinctive genus with G + C% content ranging from 34 to 35% (*Bacillus cereus* and other *Bacillus* related species) to 44–46% (*Bacillus subtilis* and other *Bacillus* related species) and genome size ranges from 3.7 to 6.4 Mb [14,15]. *Bacillus safensis* strain WOB3 genome size and G + C% content are within the range of most sequenced genomes of *Bacillus cereus* species [16,17]. Fig. 1 shows the subsystem statistics information of *Bacillus safensis* WOB3. The bar chart on the left side of the figure depicts the percentage coverage of subsystems. The pie chart generated by the RAST server and viewed in SEED viewer depicts the distribution of the most common subsystem categories among 1,883 subsystem categories. The most abundant subsystem categories were amino acids and derivatives (337), carbohydrates (269), protein metabolism (175), cofactors, vitamins, prosthetic groups, pigments (166), nucleosides, and nucleotides (117). Other genes associated with CWW detoxification activities were dormancy and sporulation (93), respiration (82), cell wall and capsule (78), DNA metabolism (78), virulence, diseases and defense (59), iron acquisition (56), RNA metabolism (56), stress response (41), membrane transport (38), regulation and cell signaling (29), metabolism of phosphorus (21), nitrogen (17), aromatic compounds (12), potassium (10), sulphur (6), motility and chemotaxis (9) secondary metabolism (9), cell division (6). Similar genes were previously identified in *Bacillus safensis* B204-B1-5 (1326975.9) [18] and *Bacillus safensis* subsp. *safensis* (2490859.4) [19].

Fig. 2 presents the quality check (CheckM) analysis) [20] on WOB3 genome sequence. The quality analysis shows that completeness of 99.41% (100th Percentile, dark blue bar) with very minimal contamination of 1.65%.

4. Experimental Design, Materials and Methods

4.1. Bacterial isolation

Bacillus safensis strain WOB3 was isolated from cassava wastewater (CWW) containing linamarin as a major cyanogenic compound using the method described by Reynold et al. [21], with some modifications. Briefly, immediately after CWW collection from the processing factory, in

the cassava processing factory at Odogunyan, IKorodu, Lagos State, Nigeria, transported to the Nigerian Institute of Medical Research laboratory. The samples were subjected to 10-fold serial dilutions [21]. The pour plate and spread methods were used to plate the appropriate dilutions of each sample into triplicates onto nutrient agar [22]. These were incubated at 37 ºC for 24 hours. Pure bacteria colonies were maintained on a nutrient agar slant and stored at 4°C until needed. The plates were monitored for growth daily, grown colonies were sub-cultured several times on fresh media, preserved in nutrient agar slant for storage, and stored at 4°C, 30% glycerol stock solution, and stored at -80°C for long storage and future use.

4.2. DNA extraction and genome sequencing

Bacillus safensis strain WOB3 was cultured aerobically on nutrient agar plates at 37°C for 24 hours. Extraction of genomic DNA was performed using the Zymo Research Fungal/Bacterial DNA MiniPrep Kit as per the manufacturer's instructions. The quality of the DNA was assessed with a Nanodrop spectrophotometer determining A260/280 ratio. The DNA was sent to a commercial service provider, Laragen Inc., 10061, Culver City, Virginia, USA for sequencing. Illumina libraries were generated using xGen DNA kit, and then 2x150 bp paired-end sequencing was performed with an Illumina HiSequencing platform (HiSeq4000X). The obtained raw reads from the run were quickly checked using FastQC (v.1.0.0, BaseSpace Illumina) and subsequently trimmed using Fast XTool (v.2.2.5, BaseSpace Illumina) [23].

4.3. Genome assembly and annotation

Genome denovo assembly and assembly quality were performed using SPAdes (v 3.9.0, BaseSpace Illumina) [24,25] The genome assemblies were then improved using Pilon (version 1.24.) [23,26], and Busco (v 5.4.7) [24,27]. Also, the CheckM analysis (v1.2.2) [20] was used for quality check of the WOB3 genome. The final genome assembly was annotated through the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [12] and RAST [25,28].

4.4. Phylogenomic classification

The genome sequence data was uploaded to the National Centre for Biotechnology Information Server (NCBI) (https://www.ncbi. nlmnih.gov), for a whole genome-based taxonomic analysis with other validly published type strains [26]. The construct of an accurate evolutionary relationship with other bacterial whole genomes that are closely related to *Bacillus safensis* species was carried out using the Type Strain Genome Server (TYGS) [13].

Limitations

None.

Ethics Statement

The authors have read and followed the ethical requirements for publication in Data in Brief and confirmed that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT Author Statement

Ogunyemi, A.K, Conceptualization, Investigation, Methodology, Project administration, Writing – original draft. **Buraimoh, M.O,** Investigation, Writing – review and editing. **Dauda W.P,** Investigation, Data curation, Writing – review and editing. **Akapo O.O,** Investigation, Writing – review and editing. **Ogunyemi, B.C,** Investigation, Writing – review and editing. **Samuel T.A,** Supervision, Writing – review and editing. **Ilori M.O,** Supervision, Writing – review and editing. **Amund, O.O,** Supervision, Writing – review and editing.

Data Availability

Analysis of the genome of Bacillus safensis, a Linamarin-utilizing Bacterium (LUB) isolated from Cassava wastewater (CWW), Lagos-State, Nigeria (Original data) (NCBI).

Acknowledgements

Our gratitude goes to the Scientists at Laragen In., 1061 Culver City, Virginia, USA, for their technical assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

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.

References

[1] C.I. Harbor, E.C. Ogundu, Effect of processing on cyanide reduction in different cassava products, Niger. J. Biochem. Mol. Biol. 24 (2009) 35–37.
[2] A. Hidayat, N. Zuaraida, I. Hanarida, D.S. Damardjati, Cyanogenic content of cassava root of 179 cultivars grown in Indonesia, J. Food Comp. Anal. 13 (2000) 71–82.
[3] D. Miles, E. Jansson, M.C. Mai, M. R Azer, P. Day, C. Shadbolt, V. Stitt, A. Kiermeier, E. Elizabeth Szabo, A survey of total hydrocyanic acid content in ready-to-eat cassava-based chips obtained in the Australian market in 2008, J. Food Prot. 74 (6) (2011) 980–985 2011, doi:10.4315/0362-028X.JFP-10-557.
[4] S.P. Vasconcellos, M.P. Cereda, J.R. Cagnon, M.A. Foglio, R.A. Rodrigues, G.P. Manfio, V.M. Oliveira, In vitro degradation of linamarin by microorganisms isolated from cassava wastewater treatment lagoons, Braz. J. Microbiol. 40 (4) (2009) 879–883, doi:10.1590/S1517-838220090004000019.
[5] S. Islam, A.M. Akanda, A. Prova, M.T. Islam, M.M. Hossain, Isolation and identification of plant growth promoting rhizobacteria from cucumber rhizosphere and their effect on plant growth promotion and disease suppression, Front. Microbiol. 6 (1360) (2016) 1e12.
[6] S. Johnson Card, L. Teasdale, J. Caradus, Deciphering endophyte behaviour: the link between endophyte biology and efficacious biological control agents, FEMS Microbiol. Ecol. 92 (2016) 1e19.
[7] H. Maughan, G. Van der Auwera, *Bacillus* taxonomy in the genomic era finds phenotypes to be essential though often misleading, Infect. Genet. Evol. 11 (5) (2011) 789e797 2011.
[8] Z. Xie, Y.Y. Chu, W. Zhang, D. Lang, X. Zhang, *Bacillus pumilus* alleviates drought stress and increases metabolite accumulation in Glycyrrhiza uralensis Fisch, Environ. Exp. Bot. 158 (2019) 99e106.
[9] R. Maheshwari, N. Bhutani, A. Bhardwaj, P. Suneja, Functional diversity of cultivable endophytes from *Cicer arietinum* and *Pisum sativum*: bioprospecting their plant growth potential, Biocatal. Agric. Biotechnol 20 (2019) 101229.
[10] I.C.P. Paz, R.C.M. Santin, A.M. Guimaraes, O.P.P.D. Rosa, M.C. Quecine, M.C.P.E. Silva, J.L. Azevedo, A.T.S. Matsumura, Biocontrol of Botrytis cinerea and *Calonectria gracilis* by eucalypts growth promoters *Bacillus* spp, Microb. Pathog. 121 (2018) 106e109.
[11] A. Hashem, B. Tabassum, E. Fathi Abd_Allah, *Bacillus subtilis*: a plant-growth promoting *Rhizobacterium* that also impacts biotic stress, Saudi J. Biol. Sci. 26 (6) (2019) 1291–1297, doi:10.1016/j.sjbs.2019.05.004.
[12] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucleic Acids Res. 44 (14) (2016) 6614–6624, doi:10.1093/nar/gkw569.

[13] J.P. Meier-Kolthoff, H.P. Klenk, M. Göker, Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age, Int. J. Syst. Evol. Microbiol. 64 (2014) 352–356, doi:10.1099/ijs.0.056994-0.
[14] M. Eppinger, et al., Genome sequences of the biotechnologically important Bacillus megaterium strains QM B1551 and DSM319, J. Bacteriol. 193 (2011) 4199–4213.
[15] Q. Zeng, J. Xie, Y. Li, T. Gao, C. Xu, Q. Wang, Comparative genomic and functional analyses of four sequenced Bacillus cereus genomes reveal conservation of genes relevant to plant-growth-promoting traits, Sci. Rep. 8 (1) (2018) 1–10.
[16] I. Anderson, A. Sorokin, V. Kapatral, G. Reznik, A. Bhattacharya, N. Mikhailova, H. Burd, V. Joukov, D. Kaznadzey, T. Walunas, M. D'Souza, N. Larsen, G. Pusch, K. Liolios, Y. Grechkin, A. Lapidus, E. Goltsman, L. Chu, M. Fonstein, S.D. Ehrlich, R. Overbeek, N. Kyrpides, N. Ivanova, Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis*, FEMS Microbiol. Lett. 250 (2) (2005) 175–184.
[17] L.D. Alcaraz, G. Moreno-Hagelsieb, L.E. Eguiarte, V. Souza, L. Herrera-Estrella, G. Olmedo, Understanding the evolutionary relationships and major traits of Bacillus through comparative genomics, BMC Genomics 11 (1) (2010).
[18] M. Satomi, M.T. La Duc, K. Venkateswaran, Bacillus safensis sp. nov., isolated from spacecraft and assembly-facility surfaces, Int. J. Syst. Evol. Microbiol. 56 (2006) 1735–1740.
[19] L.M. Ludueña, M.S. Anzuay, J.G. Angelini, M. McIntosh, A. Becker, O. Rupp, et al., Genome sequence of the endophytic strain *Enterobacter* sp. J49, a potential biofertilizer for peanut and maize, Genomics 111 (2018) 913–920.
[20] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. 25 (7) (2015) 1043–1055, doi:10.1101/gr.186072.114.
[21] J. Reynolds, Serial Dilution Protocols, ASM Microbe Library, 2005 http://www.microbelibrary.org/component/resource/laboratory-test/2884-serial-dilution-protocols.
[22] A. Lateef, The microbiology of a pharmaceutical effluent and its public health implications, World J. Microbiol. Biotechnol. 22 (2004) (2004) 167–171.
[23] B.J. Walker, T. Thomas Abeel, T. Terrance Shea, M. Margaret Priest, A. Amr Abouelliel, S. Sharadha Sakthikumar, C.A. Cuomo, Q. Qiandong Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One 9 (11) (2014) e112963, doi:10.1371/journal.pone.0112963.
[24] M. Manni, M.R. Berkeley, M. Seppey, F.A. Simão, M. Evgeny, E.M. Zdobnov, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, Mol. Biol. Evol. 38 (1) (2021) 4647–4654 0.
[25] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology, BMC Genomics 8 (9) (2008) 75.
[26] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, Nat. Commun. 10 (2019) 2182, doi:10.1038/s41467-019-10210-3.
[27] G.J. Hannon, FASTX-Toolkit Package. Available online at: http://hannonlab.cshl.edu/fastx_toolkit/ (2010), accessed on 7th March, 2024.
[28] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, doi:10.1089/cmb.2012.0021.