

Whole Genome Sequencing of Sunflower Root-Associated *Bacillus cereus*

Olubukola Oluranti Babalola , Bartholomew Saanu Adeleke and Ayansina Segun Ayangbenro 

Food Security and Safety Niche Area, Faculty of Natural and Agricultural Sciences, North-West University, Mmabatho, South Africa.

Evolutionary Bioinformatics
Volume 17: 1–6
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769343211038948



ABSTRACT: In recent times, diverse agriculturally important endophytic bacteria colonizing plant endosphere have been identified. Harnessing the potential of *Bacillus* species from sunflower could reveal their biotechnological and agricultural importance. Here, we present genomic insights into *B. cereus* T4S isolated from sunflower sourced from Lichtenburg, South Africa. Genome analysis revealed a sequence read count of 7 255 762, a genome size of 5 945 881 bp, and G + C content of 34.8%. The genome contains various protein-coding genes involved in various metabolic pathways. The detection of genes involved in the metabolism of organic substrates and chemotaxis could enhance plant-microbe interactions in the synthesis of biological products with biotechnological and agricultural importance.

KEYWORDS: Bacterial endophytes, plant-microbe interactions, root architecture, secondary metabolites, siderophore

RECEIVED: May 1, 2021. **ACCEPTED:** July 9, 2021.

TYPE: Metagenomics and the analysis of microbiomes – Short Report

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the National Research Foundation of South Africa grants (UID: 123634; 132595 to OOB).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Olubukola Oluranti Babalola, Food Security and Safety Niche Area, Faculty of Natural and Agricultural Sciences, North-West University, Private Bag X2046, Mmabatho 2735, South Africa. Email: olubukola.babalola@nwu.ac.za

Introduction

Insights into plant endosphere communities have provided several opportunities in ensuring sustainable agriculture.¹ Endosphere represents the entire internal tissue of plants colonized by microorganisms called ‘microbial endophytes’, which is beneficial to the host plants without eliciting any pathological effect.² The role of microbial endophytes has been postulated with significant interest in agricultural biotechnology for plant growth and protection against environmental stress adaptors.³ Remarkably, identifiable bacterial species in the genus *Bacillus* from plant roots with plant growth-promoting characteristics and application in the greenhouse and field trials upon inoculation that enhances cucumber yield has suggested their potential in the formulation of bioinoculants in maintaining plant health and disease control.⁴

Genetic modification of *Bacillus* species has contributed immensely to ecological balance with multifaceted significance in mitigating the excesses of chemical fertilizer application, and suppression of stress adaptors.⁵ The synthesis of bioinoculants from agriculturally important microbes has been employed as an alternative to develop environmentally friendly agriculture and to minimize threats posed by agrochemical usage in food production.⁶ However, the rationale of bacterial dynamics and their mechanisms in plant growth-promotion has been explained, but the actual mechanisms are less understood.

The genomic analysis of *Bacillus* species associated with the medicinal plant, *Costus igneus* has been reported,⁷ nevertheless, genomic insights into *Bacillus* species associated with sunflower in Southern Africa has not been studied. Hence, studying the entire genomic structure of strain T4S from sunflower can pinpoint a delivery roadmap in the identification of novel genes participating in the metabolic pathways of biomolecules

as the whole genome reports on *Bacillus* species in the endosphere of sunflower are rarely studied. Hence, integration of genomic data analysis will help predict important genes involved in metabolic pathways expressed by root-associated bacterial endophytes. Thus, this study presents the genomic analysis of endophytic bacterium *B. cereus* T4S earlier isolated from the sunflower root endosphere for improved sunflower yield.

The bacteria isolated from sunflower roots were sourced from farmlands in Lichtenburg, South Africa (26°4’31.266”S, 25°58’44.442”E) in February 2020. The healthy sunflower plants were carefully uprooted, placed inside sterile zip-lock bags, transported into the laboratory and stored at 4°C. The root samples were cut into small sizes with a sterile scalpel and washed in sterile distilled water. To ensure complete removal of the epiphytic bacteria, surface sterilization was achieved by soaking the samples in 70% ethanol for 3 minutes, followed by 3% sodium hypochlorite for 3 minutes, 70% ethanol for 30 seconds and rinsed with sterile distilled water. The level of sterility of the samples was assessed by pour plating on Luria-Bertani (LB) agar using the last water used to rinse plant samples. One gram (1 g) of plant material was weighed, suspended in 1 M phosphate-buffered solution (PBS) and manually macerated in a mortar and pestle until a smooth suspension was obtained. Sample suspensions were serially diluted up to 10⁻⁹ dilutions and 0.1 mL of an aliquot from dilutions 10⁻⁵ and 10⁻⁶ were pipetted into Petri dishes and pour plated with sterilized LB agar. Inoculated Petri plates were incubated at 28°C for 24 hours. Distinct bacterial colonies formed on the plates were counted and selected based on morphological appearance. Pure culture of the bacterial isolate was obtained by repeated streaking onto sterile LB agar and incubated at 28°C for 24 hours. The pure bacterial strains were kept on an agar slant and stored at 4°C for further use.



The stored purified *B. cereus* T4S in 30% glycerol at -80°C were sub-cultured on LB agar (g/L; NaCl – 10 g, tryptone – 10 g, and yeast extract – 5 g) and incubated at 28°C for 24 hours, then used for DNA extraction. Bacterial DNA was extracted by employing a commercial Quick-DNA™ Miniprep Kit specific for fungi or bacteria (Zymo Research, Irvine, CA, USA; Cat. No. D6005), following the manufacturer's protocol. To quantify the purity of DNA, a NanoDrop spectrophotometer (Thermo Fischer Scientific, CA, USA) was used, while DNA quality was evaluated on 2% agarose gel and electrophoresed. The WGS of endophytic bacterial strain T4S was performed on Illumina's Nextseq platform following the standard Illumina methods at Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa.

Genomic sequences were analysed on the KBase platform.⁸ The quality assessment of the sequence reads was evaluated by FastQC (version 0.11.5),⁹ while the removal of sequence adaptor and low-quality bases was processed with trimmomatic (version 0.36).¹⁰ Furthermore, sequence reads were assembled with SPAdes (version 3.13.0).¹¹ Gene annotation and prediction were performed on RASTtk (Rapid Annotations using Subsystems Technology toolkit – version 1.073) and the publicly available NCBI (<https://www.ncbi.nlm.nih.gov/>) Prokaryotic Genome Annotation Pipeline (PGAP).¹² All analyses were performed using default parameters unless otherwise specified. Secondary metabolites were determined by antiSMASH (version 6.0.0).¹³ The circular genome visualization was obtained from KBase (<https://kbase.us/>), while the phylogeny analysis was performed using MrBayes (http://www.phylogeny.fr/one_task.cgi?task_type=mrbayes) (version 3.2.6).^{8,14} The genomic comparison of strain T4S with other *B. cereus* was achieved as described by Zeng et al.¹⁵ In addition,

in vitro screening of the plant growth-promoting potential of the strain T4S was performed using plate assay.

The strain T4S responded positively to almost all the plant growth-promoting tests (Table 1).

The comparison of genomic features of *B. cereus* strain T4S with other *B. cereus* is presented in Table 2. Notably, the whole-genome sequence analysis of strain T4S had a sequence read

Table 1. Plant growth-promoting features of *B. cereus* T4S.

PLANT GROWTH-PROMOTING TESTS		
ISOLATE SCREENING	QUALITATIVE	QUANTITATIVE
IAA	+	11.29 ± 0.01 ^a
Siderophore	++	87.30 ± 0.38 ^c
Phosphate	+	30.43 ± 0.18 ^b
Exopolysaccharide	+++	–
Enzyme assay		ZOC (mm)
Amylase	–	0 ± 0.00 ^a
Cellulase	++	50.00 ± 0.01 ^d
Xylanase	++	180.00 ± 0.01 ^e
Mannanase	++	35.00 ± 0.01 ^c
Protease	+	4.00 ± 0.01 ^b

Abbreviations: IAA, indole acetic acid; ZOC, zone of clearance measurement. Mean ± standard deviation values with different superscripts (small letters) within the same column represent a significant difference.
– = negative reaction.
+ = positive reaction.

Table 2. Comparison of genomic features of *B. cereus* strain T4S with other *B. cereus*.

GENOME	<i>B. CEREUS</i> T4S	<i>B. CEREUS</i> LCR12	<i>B. CEREUS</i> 905	<i>B. CEREUS</i> PGBE311	<i>B. CEREUS</i> AR156	<i>B. CEREUS</i> UW85
Genome size (bp)	5 945 881	6 031 570	5 386 583	5 744 144	5 671 798	6 404 077
G + C value	34.8	34.8	35	35	35.25	34.81
N_{50}	65 078	3706	91	5 230 609	5 160 326	285
L_{50}	32	1	17	1	1	6
Number of contigs	198	40	127	3	4	54
Number of genes	6086	6078	5560	5801	5320	5321
NCS	6277	6005	5412	5625	5725	6257
tRNA genes	56	108	83	108	99	70
rRNA genes	5	48	6	42	39	38
Bio-project numbers	SAMN18138757	SAMN05436732	SAMN04496851	SAMN09783485	SAMN04924832	SAMN05171626
Bio-sample	PRJNA706601	PRJNA331062	PRJNA312259	PRJNA485192	PRJNA320205	PRJNA322638
Accession number	JAFNAY000000000	MCAX000000000	LSTW000000000	SAUQ000000000	CP015589	LYVD000000000

Abbreviation: NCS, number of the coding sequence.

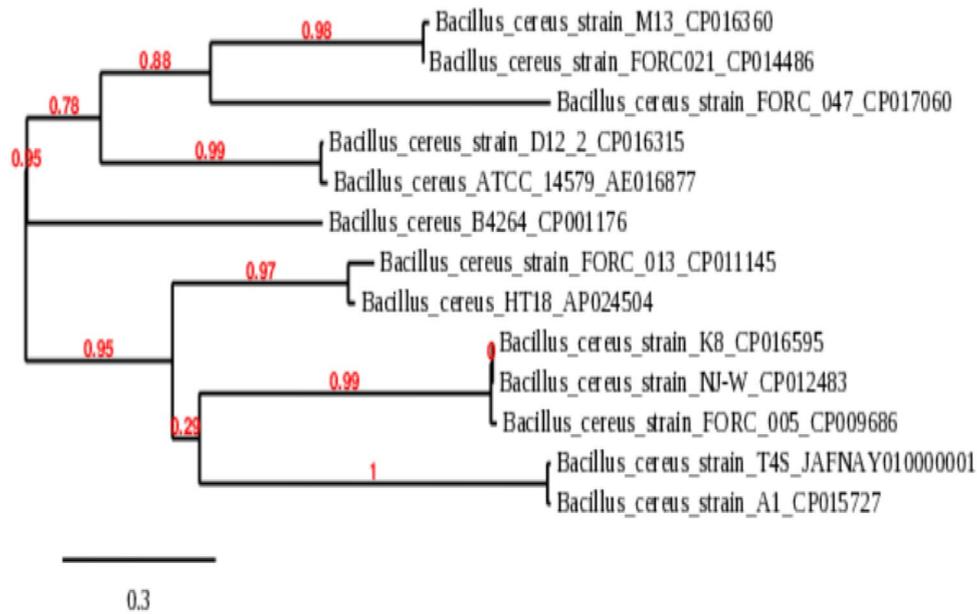


Figure 1. Phylogeny sequence data of *B. cereus* T4S.

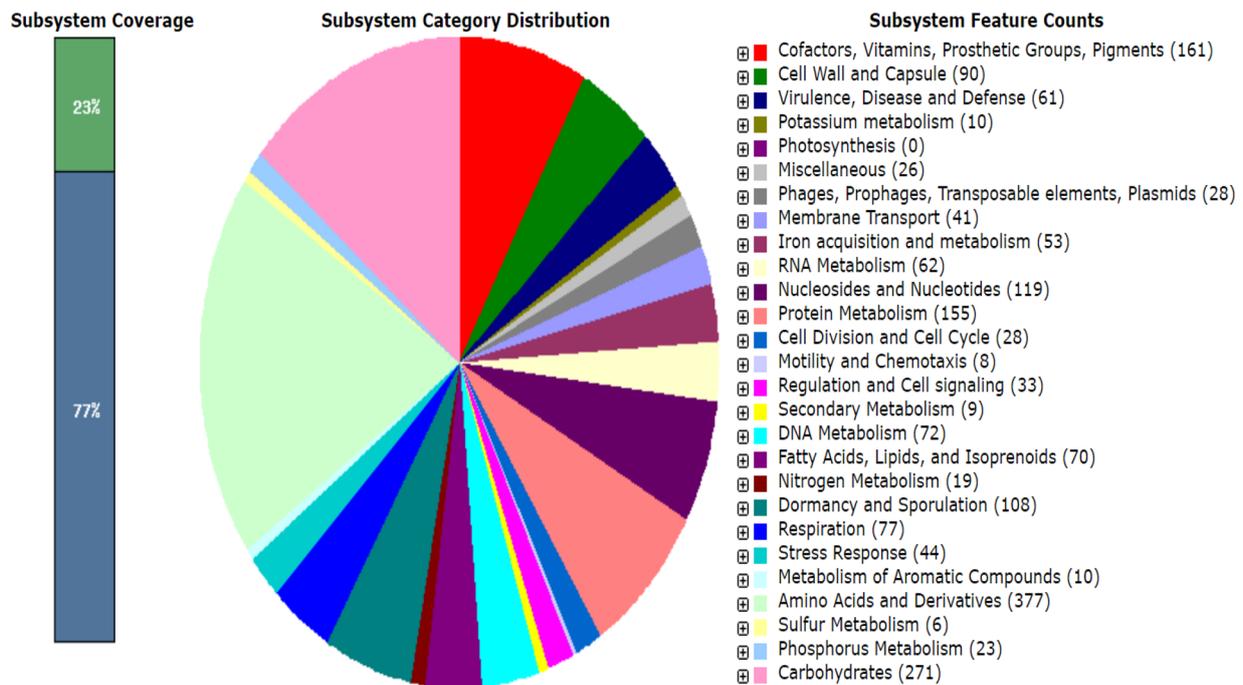


Figure 2. Subsystem category distribution of key PCG of *B. cereus* strain T4S annotated in the RAST SEED viewer annotation online server. The green/blue bar represents the subsystem coverage in percentage. Blue bar correlates with the percentage (%) of proteins present.

count of 7 255 762, a genome size of 5 945 881 bp, and a G + C value of 34.8%. The average read length was 151 bp. The N_{50} and L_{50} values of the raw sequences were 65 078 and 32 bp, while the number of contigs and subsystems were 198 and 341, respectively (Table 2). Also, genome analysis revealed 6277 coding sequences and 63 RNAs. Figure 1 shows the phylogeny of the sequence data of *B. cereus* T4S, while the subsystem category distribution of the key PCG is presented in Figure 2. The subsystem statistics showed 27 subsystem feature counts of the coding protein into functional groups with a total of 5159

PCG. The 1961 genes annotated by the SEED viewer were grouped into molecular function, cellular components and biological processes. The topmost 5 groups were protein metabolism (155 genes), vitamins, prosthetic groups, pigments (161 genes), cofactor, amino acids and derivatives (377 genes), carbohydrates (271 genes) and nucleosides and nucleotides (119 genes). The circular representation of the draft genome of *B. cereus* strain T4S is presented in Figure 3.

The secondary metabolite gene clusters, namely, non-ribosomal peptides (NRPS) and siderophore of strain T4S

Accession: JAFNAY000000000
 Size: 5,945,881 bp



Figure 3. Circular representation of the draft genome of *Bacillus cereus* strain T4S. Each color from the external to internal circle depicts green (GC skew +), and red (ORF). The ring black coloration (GC content) at the peak indicated higher or lower values than average GC content. The GC skew (-/+) in purple/green peaks in/outside the circle indicated values greater or smaller than 1. The GC skew is calculated as $G-C/G + C$.

Table 3. Information on the siderophore similar known gene cluster.

SIDEROPHORE						
Genes	Other	Other	Other	Additional biosynthetic	Core synthetic	Core synthetic
Metabolites	Petrobactin biosynthesis (PB) protein AsbF	PB protein AsbE	PB aryl carrier protein AsbD	3,4 dihydroxybenzoic acid-AMP ligase AsbC	PB protein AsbB	PB protein AsbA
BA	1986	1985	1984	1983	1982	1981
Location	7090-6248	6210-5227	5203-4928	4931-3693	3706-1868	1809-1
BLAST hit with	J0658-00175 (30 471-31 320)	J0658-00180 (31 357-32 341)	J0658-00185 (32 364-32 631)	J0658-00190 (32 627-33 866)	J0658-00195 (33 852-35 691)	J0658-00200 (35 751-37 560)
Percentage identity	88	93	93	93	84	92
Percentage coverage	100	100	103	100	100	100
BLAST hit score	506	631	158	784	1064	1052
E-value	2.4e-143	5.9e-181	5.1e-39	6.8e-227	6.6e-311	3.1e-307
Function	Sugar phosphate isomerase/epimerase	Petrobactin biosynthesis protein AsbE	Petrobactin biosynthesis protein AsbD	AMP-building protein	Siderophore biosynthesis protein	lucA/lucC family siderophore biosynthesis protein

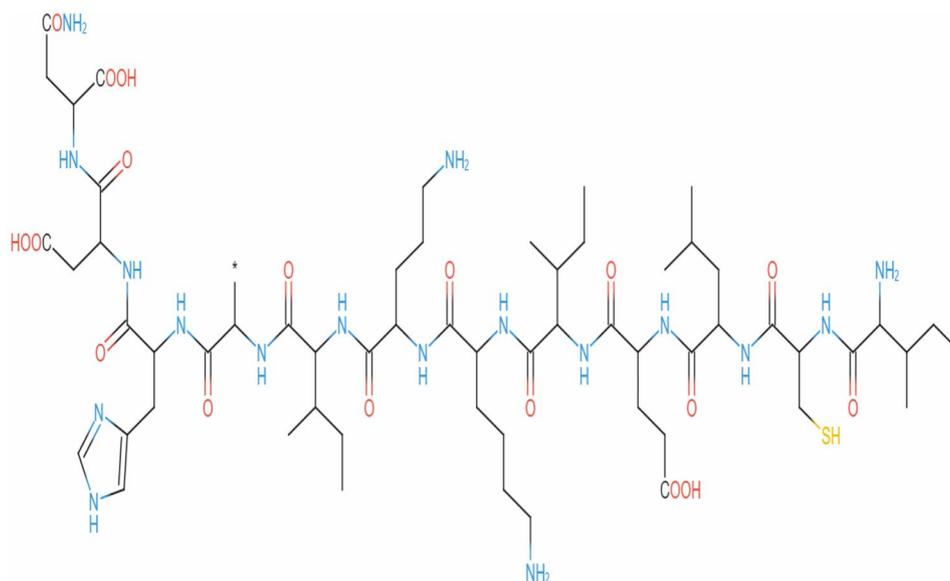


Figure 4. Predicted core structure of NRPS synthesis by *B. cereus*.

Table 4. Genes involved in plant growth promotion (PGP).

PGP	IDENTIFIABLE GENES
Nitrogen fixation	<i>nif3</i> -like, <i>nifU</i> , <i>gltPX</i> , <i>glnAHR</i> , <i>nadR</i> , <i>nirBD</i> , <i>narHIJK</i>
Siderophore	<i>fbpA</i> , <i>feoB</i> , <i>fetB</i>
ACC deaminase	<i>acdA</i>
phosphate	<i>kdpA</i> , <i>pstABCS</i> , <i>phoUH</i> , <i>gltP</i> , <i>ugpC</i>
IAA	<i>trpBCDES</i> , <i>dhaS</i>
Cytokinin	<i>miaA</i>
ammonia	<i>nadE</i>
Induction of system resistance	<i>cspABC</i>
Attachment organelles	<i>fliCSEMYNPHQLFRK</i> , <i>flgKGEBD</i> , <i>flhFAB</i> , <i>cheVW</i> , <i>efp</i> , <i>hfq</i> , <i>motAB</i>
Antibiosis/Biocontrol	<i>acoAB</i> , <i>ilvADCNBE</i> , <i>ispG</i> , <i>mind</i>
Oxidative stress	<i>hcp</i> , <i>grx</i> , <i>bsaA</i> , <i>ahpF</i> , <i>ahpC</i> , <i>osmC</i> , <i>bcp</i> , <i>tpx</i> , <i>katB</i> , <i>soda</i>

with 100% similarity selected and information on the functions of notable secondary metabolite (siderophore) genes (AsbABCDEF) that code for petrobactin synthesis/petrobactin-synthetic aryl carrier protein and 3,4 dihydroxybenzoic acid-AMP ligase, specific for additional biosynthetic, core synthetic and other genes were presented (Table 3).

Also, in Figure 4, the core structure of NRPS with a functional group, thiol (SH), carboxyl (COOH), amine (NH₂), carbonyl (C=O) is represented. Genomic analysis revealed various genes involved in nitrogen fixation and nitrogen metabolism, iron transport and siderophore production, growth hormone synthesis, phosphate solubilization and transport, motility, biofilm formation and chemotaxis, biological control, and oxidative and nitrosative stress (Table 4).

The results obtained from this study revealed the genetic capacity of strain T4S, which suggests its bioprospecting in

agriculture in enhancing plant growth. The predicted genes on the other hand may contribute to plant-bacterial interactions and ensure endosphere competence. Therefore, genome analysis of strain T4S provides information on the genes involved in sustainable plant growth and health.

Sequence Data Information

From the NCBI database output, the Bioproject is <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA706601>; BioSample number is <https://www.ncbi.nlm.nih.gov/biosample/SAMN18138757> while the whole genome accession number is JAFNAY000000000.

Acknowledgements

OOB acknowledges the National Research Foundation of South Africa for the grants (Grant numbers: 123634; 132595),

supporting research in her laboratory. BSA thanked the National Research Foundation of South Africa and The World Academy of Science (TWAS) for the NRF-TWAS African Renaissance Doctoral scholarship (UID:116100). ASA is grateful to North-West University for a postdoctoral fellowship award.

Author Contributions

All the mentioned authors contributed substantially and intellectually to this work. OOB designed the research, revised the work critically for important intellectual content, performed quality assurance, provided funding, project administration and resources. BSA was involved in data curation, formal analysis, investigation, visualization of data and writing of the original draft of the manuscript. ASA was involved in data curation, visualization of data, reviewing and thoroughly editing of the initial draft, validation and formal analysis.

ORCID iDs

Olubukola Oluranti Babalola  <https://orcid.org/0000-0003-4344-1909>

Ayansina Segun Ayangbenro  <https://orcid.org/0000-0002-3220-1873>

REFERENCES

1. Adeleke BS, Babalola OO. The endosphere microbial communities, a great promise in agriculture. *Int Microbiol.* 2020; 4:1-17.
2. Zhang L, Zhang W, Li Q, et al. Deciphering the root endosphere microbiome of the desert plant *Alhagi sparsifolia* for drought resistance-promoting bacteria. *Appl Environ Microbiol.* 2020;86:2863-2819.
3. Santoyo G, Moreno-Hagelsieb G, del Carmen Orozco-Mosqueda M, Glick BR. Plant growth-promoting bacterial endophytes. *Microbiol Res.* 2016;183:92-99.
4. Samaras A, Nikolaidis M, Antequera-Gómez ML, et al. Whole genome sequencing and root colonization studies reveal novel insights in the biocontrol potential and growth promotion by *Bacillus subtilis* MBI 600 on cucumber. *Front Microbiol.* 2020;11:600393.
5. Gond SK, Bergen MS, Torres MS, White JF Jr. Endophytic *Bacillus* spp produce antifungal lipopeptides and induce host defence gene expression in maize. *Microbiol Res.* 2015;172:79-87.
6. Pandey G. Challenges and future prospects of agri-nanotechnology for sustainable agriculture in India. *Environ Technol Innov.* 2018;11:299-307.
7. Polter SJ, Caraballo AA, Lee YP, et al. Isolation, identification, whole-genome sequencing, and annotation of four *Bacillus* species, *B. anthracis* RIT375, *B. circulans* RIT379, *B. altitudinis* RIT380, and *B. megaterium* RIT381, from internal stem tissue of the insulin plant *Costus igneus*. *Genome Announc.* 2015;3:847-815.
8. Arkin AP, Cottingham RW, Henry CS, et al. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol.* 2018;36:566-569.
9. Bioinformatics B. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham Institute; 2011.
10. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114-2120.
11. Nurk S, Bankevich A, Antipov D, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol.* 2013;20:714-737.
12. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44:6614-6624.
13. Weber T, Blin K, Duddela S, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 2015; 43:237-243.
14. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 2001;17:754-755.
15. Zeng Q, Xie J, Li Y, Gao T, Xu C, Wang Q. Comparative genomic and functional analyses of four sequenced *Bacillus cereus* genomes reveal conservation of genes relevant to plant-growth-promoting traits. *Sci Rep.* 2018;8:17009.