BMC Genomic Data

DATA NOTE Open Access

Unveiling the complete genome sequence of *Paenibacillus taichungensis*: genomic features and biocontrol potential

Shanwen Ding[1], Zijun Ma[1], Yafei Tang[1], Lin Yu[1], Zifu He[1] and Xiaoman She[1*]

Abstract

Objectives The genus *Paenibacillus* encompasses a diverse group of Gram-positive bacteria with significant biotechnological potential. However, the research data and application cases of *Paenibacillus taichungensis* were still poorly understood. In this study, we isolated a *P. taichungensis* strain BB507, which demonstrated antibacterial effect on *Ralstonia solanacearum* species complex, and provided data and analysis of its complete genome.

Data description Strain BB507 was isolated from a pine rhizosphere in Meizhou city, Guangdong province of China, and showed antibacterial activity against *Ralstonia solanacearum* species complex. Complete genome was sequenced using Illumina NovaSeq (second-generation) and Oxford Nanopore (third-generation) platforms. The genome of BB507 comprised of a 6,974,531 bp circular chromosome and a 352,197 bp circular plasmid, encoding a total of 6,649 gene with an average gene length of 950 bp, 103 transfer RNAs, 2 sRNAs, and 36 rRNAs. Three candidate CRISPRs, 6 genomic islands and 14 prophages were predicted. The bacterial orthologous average nucleotide identity (OAT) and the type genome server (TYGS) analysis highlighted the strain BB507 was clustered into a subgroup with *P. taichungensis*. antiSMASH v7.0 predicted the presence of 10 secondary metabolite gene clusters in the genome. These findings will serve as a useful resource for further research in industrial and agricultural biotechnology.

Keywords *Paenibacillus taichungensis*, Antibacterial activity, Complete genome, Secondary metabolites

Objective

The genus *Paenibacillus* was designated in 1993, comprising 11 species that were originally classified under the genus *Bacillus* [1, 2]. Novel species of this genus have been rapidly discovered, and currently, more than 150 named species have been identified. Members of this genus are facultative anaerobic, endospore-forming, motile, rod-shaped and gram-positive bacteria [3]. It has been widely used in industry and agriculture because of its unique ability in plant growth promotion, biocontrol and biodegradation [4]. *P. taichungensis* was firstly isolated and identified from a sample of soil from Taichung, a city in Taiwan of China [5]. It was subsequently isolated from the endophytic bacteria of *Arabidopsis thaliana* and found to have significant growth-promoting effects [6]. *P. taichungensis* strain NC1 displaying high arsenic resistance was isolated and identified from the Zijin gold-copper mine in Fujian, mainland of China [7]. This suggests that *P. taichungensis*, like most *Paenibacillus*, has strong application value in agricultural and industrial fields. In agriculture, arsenic resistance could be utilized to develop bioinoculants that enable plants to thrive in

*Correspondence:
Xiaoman She
shexiaoman@gdppri.com
[1]Guangdong Provincial Key Laboratory of High Technology for Plant Protection, Plant Protection Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong 510640, China

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	The circular maps of the BB507 genomes chromosome and plasmid	Figure file (.tif)	Figshare (https://doi.org/10.6084/m9.figshare.28013507) [17]
Data file 2	Genomic features of BB507	Word file (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.28281965) [18]
Data file 3	Summary of gene annotation of the strain BB507	Word file (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.28281971) [19]
Data file 4	Pairwise average nucleotide identity (ANI) comparisons of whole genomes of BB507 and other strains	Figure file (.tif)	Figshare (https://doi.org/10.6084/m9.figshare.28281974) [20]
Data file 5	Genome-Blast Distance Phylogeny (GBDP) tree inferred with FastME 2.1.6.1	Figure file (.tif)	Figshare (https://doi.org/10.6084/m9.figshare.28283099) [21]
Data file 6	Pairwise comparisons of strain BB507 vs. type strain genomes	Table file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.28283114) [22]
Data file 7	Comparison of predicted and known secondary metabolites of BB507	Word file (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.28013507) [17]
Dataset 8	Genome assembly of *Paenibacillus taichungensis* strain BB507 Chromosome	Fasta files (.fasta)	NCBI GenBank (http://identifiers.org/nucleotide:CP175536.1) [23]
Dataset 9	Genome assembly of *Paenibacillus taichungensis* strain BB507 plasmid	Fasta files (.fasta)	NCBI GenBank (http://identifiers.org/Nucleotide:CP175537.1) [24]

arsenic-contaminated soils by reducing arsenic uptake and enhancing overall plant health. In industrial applications, such as mining and wastewater treatment, this resilience to arsenic toxicity could be harnessed to mitigate environmental pollution and promote sustainable practices. There were no complete-level genome data of *P. taichungensis*, only nine contig/scaffold-level data were included in the NCBI database.

Bacterial wilt, attributed to *Ralstonia solanacearum* species complex (RSSC) infection, has caused destructive impacts and colossal economic losses on agricultural production [8, 9]. Chemical control, which was ubiquitously used, cannot manage this disease as expected, so biocontrol has been followed with interest to date. We found a *P. taichungensiss* strain BB507 that had a potent antagonistic activity against RSSC, and the genomic sequencing and analysis revealed biocontrol mechanisms might be involved in the strain BB507.

Data description

Strain BB507 was isolated from the rhizosphere of a pine tree in Meizhou city, Guangdong province, China, and demonstrated potent antibacterial activity against RSSC using a plate confrontation assay, as previously described in our published work [10]. BB507 was sequenced using the combination of Oxford Nanopore PromethION and Illumina NovaSeq PE150 platform at Novogene Bioinformatics Technology Co., Ltd in Beijing, China. Initially, library preparation was performed according to Oxford Nanopore Technologies (ONT, Oxford, United Kingdom) protocol for multiplexing samples (1D native barcoding genomic DNA with EXP-NBD103 and SQK-LSK108). Genome DNA was sonicated to a fragment size of 350 bp, followed by A-tailing, end-polishing, and ligation with full-length adaptors for Illumina sequencing. Polymerase Chain Reaction (PCR) products were purified using the AMPure XP system, and library size distribution was analyzed by Agilent2100 Bioanalyzer. Quantification was performed using real-time PCR. A total of 1,044 Mb Nanopore clean data were generated with an estimated 145×average depth of sequencing coverage. Post-quality control, paired reads were assembled into a complete genome using SMRT Link v5.1.0, with further refinement using Illumina data and Pilon. The complete genome of BB507 was 7.33 Mb in total with GC content of 45.84%, which comprised a 6.97 Mb circular chromosome and a 0.35 Mb circular plasmid (Table 1, Data file 1). CheckM2 analysis revealed that the assembled genome of BB507 was complete, with a genome completeness of 99.85% and a contamination level of 1.15%. Gene prediction identified 6,649 genes in BB507 using GeneMarkS v4.17 [11]. Genomic component analysis of strain BB507 revealed the presence of 103 transfer RNAs (tRNAs), identified using tRNAscan-SE v1.3.1 [12], and 36 ribosomal RNAs (rRNAs) (5 S, 16 S, 23 S), identified using rRNAmmer v1.2 [13]. Additionally, 3 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) were detected using CRISPRdigger v1.0 [14], 6 genomic islands were identified using IslandPath-DIMOB v0.2 [15], and 14 prophages were identified using phiSpy v2.3 [16]. The general features of the BB507 genome were listed in Data file 2 (Table 1). The complete genome sequence of BB507 has been deposited in GenBank dataset under the accession number CP175536 (Table 1, Dataset 8) and CP175537 (Table 1, Dataset 9).

Functional annotation revealed that 4,296 genes were assigned into 47 Gene Ontology (GO) categories [25], 6016 genes were enriched in 106 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [26], and 4,483 genes were classified into 23 Clusters of Orthologous

Groups (COG) categories [27]. Annotation identified 244 putative virulence factors, 498 pathogen–host interaction genes, 604 transport proteins, 484 carbohydrate-active enzymes, and 2 antibiotic resistance proteins using databases such as Virulence Factors of Pathogenic Bacteria (VFDB) [28], Pathogen Host Interactions Database (PHI) [29], Transporter Classification Database (TCDB) [30], Carbohydrate-Active enZYmes Database (CAZyme) [31], and Antibiotic Resistance Genes Database (ARDB) [32], respectively (Table 1, Data file 3).

Orthologous Average Nucleotide Identity Tool (OAT) [33] analysis revealed that the eight *P. taichungensis* strains shared the indexes of Average Nucleotide Identity (ANI) ranging from 90.24 to 100%, and 67.06–69.25% ANI indexes with strains of *Paenibacillus graminis* and *Paenibacillus chitinolyticus*. Notably, the highest ANI index of 99.02% was found between strain BB507 and strain NC1, and they were also clustered together on the phylogenetic branch (Table 1, Data file 4). In addition, the type genome server (TYGS) [34]was used as a means of confirming the previous results shown by both the multigene phylogenetic approach and the ANI calculations. The TYGS results indicated that BB507 formed a monophyletic relationship with closely related *P. taichungensis* (Table 1, Data file 5). The digital DNA-DNA hybridization (dDDH) values with both *P. taichungensis* strain were 66.7–93.2% (Table 1, Data file 6). BB507 had a dDDH of 93.2% with G + C% content difference of 0.34 with *P. taichungensis* strain NC1. The ANI indexes < 95% and dDDH < 75% observed in some strains within the clade of *P. taichungensis* were due to *P. taichungensis* strain DB-4 (ANI index = 92.94%, dDDH = 69.4%) and VTTE13328 (ANI index = 90.28%, dDDH = 66.7%). These discrepancies suggested that the taxonomic status of these two strains might be deviated. Therefore, the taxonomic definition of *P. taichungensis* might be limited by the current sample size. However, the above results still indicated that BB507 belonged to the *P. taichungensis*.

The secondary metabolite gene clusters of BB507 were predicted using antiSMASH v7.0 [35], identifying a total of ten potential gene clusters, nine of which were chromosomal and one plasmid-borne. Seven of these clusters showed similarity to known biosynthetic gene clusters of Paeninodin, Staphyloferrin B, Corynecin, Bacillopaline, Carotenoid, Bacillibactin, and Aurantinin, with similarities ranging from 10 to 100% (Table 1, Data file 7). Additionally, three gene clusters of unknown function were identified, suggesting the presence of novel secondary metabolites.

Limitations

While this study provides valuable insights into the genomic features and biocontrol potential of *P. taichungensis* strain BB507, several limitations should be considered. First, the functional roles of the predicted secondary metabolite gene clusters require further experimental validation. Although antiSMASH v7.0 predicted ten secondary metabolite gene clusters, including both known biosynthetic pathways and several novel ones, the actual production and biological activity of these metabolites have not been confirmed. Additionally, the taxonomic analysis relies on the available genome data, which may have limitations due to the current sample size and the presence of closely related strains with similar ANI and dDDH values. The phylogenetic relationships within the *P. taichungensis*, particularly concerning strains with deviating taxonomic status (e.g., strains DB-4 and VTTE13328), highlight the need for further genomic and phenotypic investigations to refine the classification.

Acknowledgements
Not applicable.

Authors' contributions
ZJM, YFT and LY: performed strain isolation, cultivation and DNA extraction. SWD: performed the genome analysis. SWD: prepared the manuscript draft. ZFH and XMS: supervised the project, designed the experiments, and edited the manuscript. The authors read and approved the final manuscript.

Funding
The Special Fund for Guangdong Action Plan for Seeds Industry-Excavation, creation and application of special high-efficiency crop biocontrol agent ***Paenibacillus*** (2023-WPY-00-001). The Special Fund for Scientific Innovation Strategy-Construction of the construction of Main Force of agricultural research (R2023PY-JG011). The Special Fund for Scientific Innovation Strategy-Construction of High-Level Academy of Agriculture Science (R2022YJ-YB3019). Guangdong Vegetable Industry Technology System Innovation Team (2024CXTD08).

Data availability
The genome assembly data that support the findings of this study have been deposited in NCBI GenBank under the accession numbers CP175536-CP175537.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 10 December 2024 / Accepted: 10 March 2025
Published online: 17 March 2025

References

1. Grady EN, MacDonald J, Liu L, Richman A, Yuan ZC. Current knowledge and perspectives of *Paenibacillus*: a review. Microb Cell Fact. 2016;15(1):203.
2. Genersch E, Forsgren E, Pentikäinen J, Ashiralieva A, Rauch S, Kilwinski J, Fries I. Reclassification of *Paenibacillus Larvae* subsp. *Pulvifaciens* and *Paenibacillus Larvae* subsp. *Larvae* as *Paenibacillus Larvae* without subspecies differentiation. Int J Syst Evol MicroBiol. 2006;56(3):501–511.
3. Priest FG. Paenibacillus. In: Bergey's Manual of Systematics of Archaea and Bacteria. 2015:1–40.

4. Chávez-Ramírez B, Kerber-Díaz JC, Acoltzi-Conde MC, Ibarra JA, Vásquez-Murrieta M-S, Estrada-de los Santos P. Inhibition of *Rhizoctonia Solani* RhCh-14 and *Pythium ultimum* PyFr-14 by *Paenibacillus polymyxa* NMA1017 and *Burkholderia cenocepacia* CACua-24: A proposal for biocontrol of phytopathogenic fungi. Microbiol Res. 2020;230:126347.
5. Lee FL, Tien CJ, Tai CJ, Wang LT, Liu YC, Chern LL. *Paenibacillus taichungensis* Sp. nov., from soil in Taiwan. Int J Syst Evol MicroBiol. 2008;58(11):2640–5.
6. Mayer E, Dörr de Quadros P, Fulthorpe R. Plantibacter flavus, *Curtobacterium herbarum*, *Paenibacillus taichungensis*, and *Rhizobium selenitireducens* endophytes provide Host-Specific growth promotion of *Arabidopsis thaliana*, Basil, lettuce, and Bok Choy plants. Appl Environ Microbiol. 2019;85(19):e00383–00319.
7. Yu Y, Su J, Xu J, Li Yuan P, Alwathnani Hend A, Wu Z, Ji C, Feng R, Rensing C, Herzberg M. As(III) exposure induces a zinc scarcity response and restricts iron uptake in High-Level Arsenic-Resistant *Paenibacillus taichungensis* strain NC1. Appl Environ Microbiol. 2022;88(9):e00312–00322.
8. Ding S, Ma Z, Yu L, Lan G, Tang Y, Li Z, He Z, She X. Comparative genomics and host range analysis of four *Ralstonia pseudosolanacearum* strains isolated from sunflower reveals genomic and phenotypic differences. BMC Genom. 2024;25(1):191.
9. Jiang GF, Wei Z, Xu J, Chen HL, Zhang Y, She XM, Macho AP, Ding W, Liao BS. Bacterial wilt in China: history, current status, and future perspectives. Front Plant Sci. 2017;8:1549.
10. Ding S, Li P, Tang Y, He Z, She X. Identification and genomic insights into *Bacillus siamensis* strains with host colonization potential and activity against tomato bacterial wilt. Pest Manag Sci. 2025;81:1547-1561n/a(n/a).
11. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 2001;29(12):2607–18.
12. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.
13. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.
14. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35(suppl2):W52–7.
15. Bertelli C, Brinkman FSL. Improved genomic Island predictions with IslandPath-DIMOB. Bioinformatics. 2018;34(13):2161–7.
16. Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res. 2012;40(16):e126–126.
17. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the complete genome sequence of Paenibacillus taichungensis: genomic features and biocontrol potential. Figshare. 2025. https://doi.org/10.6084/m9.figshare.28013507.
18. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the complete genome sequence of Paenibacillus taichungensis: genomic features and biocontrol potential. Figshare. 2025. https://doi.org/10.6084/m9.figshare.28281965.
19. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the Complete Genome Sequence of Paenibacillus taichungensis: Genomic Features and Biocontrol Potential. *Figshare* 2025. https://doi.org/10.6084/m9.figshare.28281971.
20. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the complete genome sequence of Paenibacillus taichungensis: genomic features and biocontrol potential. Figshare. 2025. https://doi.org/10.6084/m9.figshare.28281974.
21. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the complete genome sequence of Paenibacillus taichungensis: genomic features and biocontrol potential. Figshare. 2025. https://doi.org/10.6084/m9.figshare.28283099.
22. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the complete genome sequence of Paenibacillus taichungensis: genomic features and biocontrol potential. Figshare. 2025. https://doi.org/10.6084/m9.figshare.28283114.
23. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the complete genome sequence of Paenibacillus taichungensis: genomic features and biocontrol potential. NCBI GenBank 2025. http://identifiers.org/nucleotide:CP175536.1.
24. Ding S, Ma Z, Tang Y, Yu L, He Z. Unveiling the Complete Genome Sequence of Paenibacillus taichungensis: Genomic Features and Biocontrol Potential. *NCBI GenBank* 2025. http://identifiers.org/Nucleotide:CP175537.1.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25(1):25–9.
26. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006;34(suppl1):D354–7.
27. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43(D1):D261–9.
28. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. 2011;40(D1):D641–5.
29. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J, Hammond-Kosack KE. PHI-base: a new database for pathogen host interactions. Nucleic Acids Res. 2006;34(suppl1):D459–64.
30. Saier MH Jr, Reddy VS, Tamang DG, Västermark Å. The transporter classification database. Nucleic Acids Res. 2013;42(D1):D251–8.
31. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2008;37(suppl1):D233–8.
32. Liu B, Pop M. ARDB—Antibiotic resistance genes database. Nucleic Acids Res. 2008;37(suppl1):D443–7.
33. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol MicroBiol. 2016;66(2):1100–3.
34. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 2019;10(1):2182.
35. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, et al. AntiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51(W1):W46–50.

Publisher's note