DATA NOTE Open Access



Complete genome sequences of two *Pantoea* stewartii strains ATCC 8199 from maize and PSCN1 from sugarcane

Na Chu¹, Tian-Tian Liu¹, Hui-Li Zhang¹, Dong Cui¹, Mei-Ting Huang¹, Hua-Ying Fu¹, Jun-Bo Su^{2*} and San-Ji Gao^{1*}

Abstract

Objectives The pathogen of *Pantoea stewartii* (*Ps*) is the causal agent of bacterial disease in corn and various graminaceous plants. *Ps* has two subspecies, *Pantoea stewartii* subsp. *stewartia* (*Pss*) and *Pantoea stewartii* subsp. *indologenes* (*Psi*). This study presents two complete genomes of *Ps* strains including ATCC 8199 isolated from maize and PSCN1 causing bacterial wilt in sugarcane. The two bacterial genomes information will be helpful for taxonomy analysis in this genus *Pantoea* at whole-genome levels and accurately discriminated the two subspecies of *Pss* and *Psi*.

Data description The reference strain ATCC 8199 isolated from maize was purchased from Beijing Biobw Biotechnology Co., Ltd. (China) and the strain of PSCN1 was isolated from sugarcane cultivar YZ08-1095 in Zhanjiang, Guangdong province of China. Two complete genomes were sequenced using Illumina Hiseq (second-generation) and Oxford Nanopore (third-generation) platforms. The genome of the strain ATCC 8199 comprised of 4.78 Mb with an average GC content of 54.03%, along with five plasmids, encoding a total of 4,846 gene with an average gene length of 827 bp. The genome of PSCN1 comprised of 5.03 Mb with an average GC content of 53.78%, along with two plasmids, encoding a total of 4,725 gene with an average gene length of 913 bp. The bacterial pan-genome analysis highlighted the strain ATCC 8199 was clustered into a subgroup with a *Pss* strain CCUG 26,359 from USA, while the strain PSCN1 was clustered into another subgroup with a *Ps* strain NRRLB-133 from USA. These findings will serve as a useful resource for further analyses of the evolution of *Ps* strains and corresponding disease epidemiology worldwide.

Keywords Pantoea stewartia, Stewart's bacterial wilt, Complete genome, Sugarcane (Saccharum spp.)

Objective

The genus *Pantoea* includes 19 species and three uncertain species (*P. endophytica*, *P. latae*, *P. mediterraneensis*, and *P. persica*), which occurs in association with plant and animal hosts, and environmental samples, including soil and rivers [1]. Moreover, Mergaert et al. (1993) proposed that *P. stewartia* (*Ps*) is constituted by two subspecies, *P. stewartii* subsp. *stewartii* (*Pss*) and *P. stewartii* subsp. *indologenes* (*Psi*) [2]. The strains from *P. stewartia* infect various *Poaceae* plants, including leek, onion, chive, Japanese bunching onion, rice, corn, common wheat, sugarcane, foxtail millet, pearl millet, oat, lucky

¹National Engineering Research Center for Sugarcane, Fujian Agriculture and Forestry University, Fuzhou 350002, China

²South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Science, Zhanjiang 524091, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

^{*}Correspondence: Jun-Bo Su junbosu@126.com San-Ji Gao gaosanji@fafu.edu.cn

Chu et al. BMC Genomic Data (2024) 25:86 Page 2 of 5

bamboo, and so on. The *Pss* is the agent of Stewart's vascular wilt in corn and bacterial wilt in sugarcane [3, 4].

A PSCN1 strain was isolated from the cultivar YZ08-1095 in 2017 and this strain forms yellow-colored colonies on solid nutrient agar (NA) medium and straight rods and nonencapsulated cells were observed under transmission electron microscopy [3]. Additionally, the pathogenicity of this bacterial strain has been verified by the Koch's postulate [3]. The strain PSCN1 was proposed as the Pss based on the phylogenetic analysis of bacterial 16S rDNA sequences [3]. To further exploring taxonomic classification on PSCN1, complete genome sequences of this pathogen together with the reference strain of Pss ATCC 8199 (provided by Beijing Biobw Biotechnology Co., Ltd., China) hosted in corn were sequenced and assembly based on the combination of the Illumina Hiseq and Oxford Nanopore platforms. Although the genome sequence derived from strain CCUG 26,359 (=ATCC 8199) has been assembled at the contig level by the Illumina MiSeq platform, this contig-level genome assembly includes 352 contigs and no assembled chromosomes (NCBI dataset: PRJNA563568). Thus, the two whole genome sequences will enable us to illustrate more accurate taxonomic classification of these pathogens at a pan-genome level. Additionally, our data provide some reference value for the prevention and control of bacterial wilt in corn and sugarcane.

Data description

A pure culture of two strains (ATCC 8199 and PSCN1) were grown in liquid nutrient broth (NB) medium with constant shaking at 200 rpm and 28 °C for 24 h. After extraction of bacterial genomic DNA, the 10-Kb DNA library was constructed using the SQK-LSK109 linker kit according to the manufacturer's instructions, and then sequenced by Oxford Nanopore PromethION (third-generation). Meanwhile, a 350-bp library was constructed using another sequence platform of Illumina Hiseq (second-generation). A total of 1,325 Mb and 1,915 Mb Nanopore clean data were generated in two strains ATCC 8199 and PSCN1 with an estimated 276.73× and 380.43× average depth of sequencing coverage, respectively (Table 1, Data file1) [5]. After quality control of the sequencing data, subreads from the Nanopore platform were assembled with Canu (version 1.5) [21]. The assembly results were further corrected with Illumina data using Pilon [22]. The genome of ATCC 8199 strain is 4.78 Mb in total with GC content of 54.03%, including one circular chromosome (4,526,106 bp) and five plasmids (P1-P5, with 25,199, 35,601, 62,815, 65,465, and 73,276 bp, respectively). The genome PSCN1 strain is 5.03 Mb in total with GC content of 53.78%, containing one circular chromosome (4,511,897 bp) and two plasmids (pCN101- pCN102 with 310,867 and 211,928 bp, respectively) (Data file1) [5]. Gene prediction identified 4,846 genes in ATCC 8199 strain and 4,725 genes in PSCN1 strain using Prodigal [23]. The genome of ATCC 8199 strain includes 71 tRNAs, 21 rRNAs, 6 ncRNA, 11 CRISPR numbers, 17 genomic islands, and 4 prophages. Genomic component analysis revealed that PSCN1 contained 78 tRNAs, 22 rRNAs, 18 ncRNA, 8 CRISPR numbers, 17 genomic islands, and one prophage (Data file1) [5]. The complete genome sequences of PSCN1 and ATCC 8199 have been deposited in GenBank dataset under the accession numbers CP046585-CP046587 and CP046558-CP046563, respectively.

Gene annotation was determined with the BLAST program [24] and with 12 different databases. The overview of the two complete genomes were presented to the annotation information using Circos [25] (Data file2) [6]. For the ATCC 8199 strain, GO analysis [26] revealed that 3,337 genes were assigned into 41 GO categories, with the most genes in catalytic activity (2,090 genes). KEGG analysis [27] revealed that 2,612 genes were significantly enriched in 106 pathways. A total of 559 putative virulence factors, 797 pathogen-host interaction genes, 404 transport proteins, 144 carbohydrate active enzymes, and 3 antibiotic resistance proteins were annotated based on the VFDB database [28], PHI-base [29], TCDB database [30], CAZyme database [31] and ARDB database [32], respectively (Data file3) [7]. For the PSCN1 strain, 2,790 genes were assigned into 41 GO categories. The largest category was assigned to catalytic activity (1,772 genes). A total of 2785 genes were significantly enriched in 107 KEGG pathways. A total of 695 putative virulence factors, 911 pathogen-host interaction genes, 418 transport proteins, 164 carbohydrate active enzymes, and 3 antibiotic resistance proteins were annotated based on those above-mentioned datasets (Data file4) [8].

The two genome sequences obtained in this study along with 45 Ps strains retrieved from NCBI library were used for sequence analysis. A strain of P. agglomerans CFSAN047153 was used an outgroup (Data file5) [9]. The average nucleotide identities (ANI) were calculated by pairwise genome comparison based on BLAST+and FastANI [33]. The 47 Ps strains worldwide shared 98.40-99.99% ANI indexes and shared 80.89-81.38% ANI index with the strain CFSAN047153 of P. agglomerans. The two strains ATCC 8199 and PSCN1 had 98.56% ANI index with each other and shared 98.41-99.97% and 98.30-99.07% with other *Ps* strains, respectively (Data file6) [10]. Notably, ATCC 8199 and PSCN1 had highest ANI indexes with IPV-BO 2766 (NCBI dataset: PRJNA856801) and HR3-48 (NCBI dataset: PRJNA844595) strains, respectively.

To exploring the phylogenetic relationship between PSCN1 and ATCC 8199 with other strains, the gene family clustering was carried out based on the alignment Chu et al. BMC Genomic Data (2024) 25:86 Page 3 of 5

Table 1 Overview of data files/data sets

Label	Name of datafile/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	General features of two genomes from PSCN1 and ATCC 8199 strains	Word file (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.26166673) [5]
Data file 2	Genome organization and gene distribution in two strain PSCN1(A) and ATCC 8199(B)	Word file (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.26170231) [6]
Data file 3	Summary of gene annotation of the strain ATCC 8199	Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.26170141) [7]
Data file 4	Summary of gene annotation of the strain PSCN1	Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.26170150) [8]
Data file 5	Characteristics of strains used in this study	Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.26170171) [9]
Data file 6	Average nucleotide identity based on the entire ge- nome sequence of 47 strains of <i>Pantoea stewartia</i> . and one strain of <i>Pantoea agglomerans</i>	Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.26170210) [10]
Data file 7	Phylogeny tree analysised with core-gene sequence of 42 strains of <i>Pantoea stewartii</i> and one strain of <i>Pantoea agglomerans</i>	Word file (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.26170219) [11]
Dataset 8	Genome assembly of <i>Pantoea stewartia</i> subsp. <i>stewartia</i> strain ATCC 8199 chromosome.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046558) [12]
Dataset 9	Genome assembly of <i>Pantoea stewartia</i> subsp. <i>stewartia</i> strain ATCC 8199 plasmid p1.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046559) [13]
Dataset 10	Genome assembly of <i>Pantoea stewartia</i> subsp. <i>stewartia</i> strain ATCC 8199 plasmid p2.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046560) [14]
Dataset 11	Genome assembly of <i>Pantoea stewartia</i> subsp. <i>stewartia</i> strain ATCC 8199 plasmid p3.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046561) [15]
Dataset 12	Genome assembly of <i>Pantoea stewartia</i> subsp. <i>stewartia</i> strain ATCC 8199 plasmid p4.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046562) [16]
Dataset 13	Genome assembly of <i>Pantoea stewartia</i> subsp. <i>stewartia</i> strain ATCC 8199 plasmid p5.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046563) [17]
Dataset 14	Genome assembly of <i>Pantoea stewartia</i> strain PSCN1 chromosome.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046585) [18]
Dataset 15	Genome assembly of <i>Pantoea stewartia</i> strain PSCN1 plasmid pCN101.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046586) [19]
Dataset 16	Genome assembly of <i>Pantoea stewartia</i> strain PSCN1 plasmid pCN102.	GenBank file (.bp)	GenBank (https://identifiers.org/ncbi/insdc:CP046587) [20]

of single copy genes identified with OrthoMCL [34]. The phylogenetic tree was constructed with core-gene sequence of 43 Pantoea strains using maximum likelihood method and 1,000 bootstrap replications using PhyML [35]. Ten Ps strains including ATCC 8199 and PSCN1 were clustered into one subclade, which would be further separated into three groups. Furthermore, the strains HR3-48 (Ps) and LMG 2671 (Psi) were clustered together in the group (I) The strains PSCN1 and NRRLB-133 (Ps) were clustered together in the group (II). Other six Pss strains including ATCC 8199 were clustered in the group III (Data file7) [11]. The PSCN1 was proposed as a Pss strain based on bacterial 16S rDNA sequences [3], but this pathogen might be a Psi strain at pan-genome level. However, an accurate classification of two subspecies Pss and Psi within this genus Pantoea need be further confirmed based on numerous complete genome sequences and biological experiments.

Limitations

This data note presented two complete genome sequences, one from reference strain ATCC 8199 hosted in corn and another strain PSCN1 isolated from infected sugarcane plants showing bacterial wilt symptoms. However, only a single strain genome sequence from sugarcane in China. More strains need be collected at a global context and used for the whole genome sequencing. Thus, taxonomic classification of this bacterial species would be further accurately illustrated.

Abbreviations

ANI Average Nucleotide Identity

Bp Base pair

BLAST Basic Local Alignment Search Tool

NCBI National Center for Biotechnology Information

NB Nutrient broth

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-024-01268-0.

Chu et al. BMC Genomic Data (2024) 25:86 Page 4 of 5

Supplementary Material 1

Supplementary Material 2

Supplementary Material

Supplementary Material

Supplementary Material

Supplementary Material

Supplementary Material 7

Acknowledgements

We thank the Beijing Biomarker Technology Co., Ltd. (China) for wholegenome sequencing and Beijing Biobw Biotechnology Co., Ltd. (China) providing the strain ATCC 8199.

Author contributions

DC, MTH, HYF, and JBS: performed strain isolation, cultivation and DNA extraction. NC and HLZ: performed the genome analysis. NC and TTL: prepared the manuscript draft. SJG and JBS: supervised the project, designed the experiments, and edited the manuscript. The authors read and approved the final manuscript.

Funding

This research was supported by the earmarked fund for China Agriculture Research System (grant no. CARS-17).

Data availability

The genome assembly data that support the findings of this study have been deposited in NCBI GenBank under the accession numbers CP046558-CP046563 and CP046585-CP046587 (Table 1).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 4 July 2024 / Accepted: 20 September 2024 Published online: 08 October 2024

References

- Palmer M, Coutinho TA. Pantoea. In: Anonymous NN, editors, Bergey's Manual of Systematics of Archaea and Bacteria. 2022;1–6. https://doi. org/10.1002/9781118960608.gbm01157.pub2
- Mergaert J, Verdonck L, Kersters K. Transfer of Erwinia ananas (synonym, Erwinia uredovora) and Erwinia stewartii to the Genus Pantoea emend. As Pantoea ananas (Serrano 1928) comb. nov. and Pantoea stewartii (Smith 1898) comb. nov., respectively, and description of Pantoea stewartii subsp. indologenes subsp.nov. Int J Syst Evol Micr. 1993;43(1):162–73. https://doi. org/10.1099/00207713-43-1-162.
- Cui D, Huang MT, Hu CY, Su JB, Lin LH, Javed T, Deng ZH, Gao SJ. First report of *Pantoea stewartii* subsp. stewartii causing bacterial leaf wilt of sugarcane in China. Plant Dis. 2020;105(4). https://doi.org/10.1094/PDIS-09-20-2015-PDN.
- Ren L, Zhang S, Xu ZY, Hu HQ. Complete genome sequence of *Pantoea stewartii* subsp. *indologenes* ZJ-FGZX1, a lucky bamboo pathogen. Mol Plant-Microbe in. 2020;33(11):1274–6. https://doi.org/10.1094/MPMI-05-20-0111-A.
- Chu N. General features of two genomes from PSCN1 and ATCC 8199 strains.
 2024. Figshare. https://doi.org/10.6084/m9.figshare.26166673

- Chu N. Genome organization and gene distribution in two strain PSCN1(A) and ATCC 8199(B). figshare. 2024. https://doi.org/10.6084/ m9.figshare.26170231.
- Chu N. Summary of gene annotation of the strain ATCC 8199. figshare. 2024. https://doi.org/10.6084/m9.figshare.26170141
- Chu N. Summary of gene annotation of the strain PSCN1. figshare. 2024. https://doi.org/10.6084/m9.figshare.26170150.
- Chu N. Characteristics of strains used in this study. 2024. Figshare. https://doi. org/10.6084/m9.figshare.26170171
- Chu N. Average nucleotide identity based on the entire genome sequence of 47 strains of Pantoea stewartia. And one strain of Pantoea agglomerans. Figshare. 2024. https://doi.org/10.6084/m9.figshare.26170210.
- Chu N. Phylogeny tree analysised with core-gene sequence of 42 strains of *Pantoea stewartii and* one strain of *Pantoea agglomerans*. 2024. Figshare. https://doi.org/10.6084/m9.figshare.26170219
- Zhang HL, Gao SJ, Cui D. Genome assembly of Pantoea stewartia subsp. stewartia strain ATCC 8199 chromosome. 2024. NCBI GenBank. https://identifiers. org/ncbi/insdc:CP046558
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* subsp. stewartia strain ATCC 8199 plasmid p1. 2024. NCBI GenBank. https://identifiers. org/ncbi/insdc:CP046559
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* subsp. stewartia strain ATCC 8199 plasmid p2. 2024. NCBI GenBank. https://identifiers. org/ncbi/insdc:CP046560
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* subsp. stewartia strain ATCC 8199 plasmid p3. 2024. NCBI GenBank. https://identifiers. org/ncbi/insdc:CP046561
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* subsp. stewartia strain ATCC 8199 plasmid p4. 2024. NCBI GenBank. https://identifiers. org/ncbi/insdc:CP046562
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* subsp. stewartia strain ATCC 8199 plasmid p5. 2024. NCBI GenBank. https://identifiers. org/ncbi/insdc:CP046563
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* strain PSCN1 chromosome. 2024. NCBI GenBank. https://identifiers.org/ncbi/ insdc:CP046585
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* strain PSCN1 plasmid pCN101. 2024. NCBI GenBank. https://identifiers.org/ncbi/ insdc:CP046586
- Zhang HL, Gao SJ, Cui D. Genome assembly of *Pantoea stewartia* strain PSCN1 plasmid pCN102. 2024. NCBI GenBank. https://identifiers.org/ncbi/ insdc:CP046587
- Koren S, Walenz BP, Konstantin B, Miller JR, Bergman NH, Phillipy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36. https://doi.org/10.1101/ gr.215087.116.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9(11):e112963. https://doi.org/10.1371/journal. pone.0112963.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11(1):119. https://doi.org/10.1186/1471-2105-11-119.
- Altschul SF, Madden TL, Schäffer AA, Zhang JH, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein databases search programs. Nucleic Acids Res. 1997;25(17):3389–402. https://doi.org/10.1093/ nar/25.17.3389.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45. https://doi.org/10.1101/gr.092759.109.
- Ashburner M, Ball CA, Blake JA, Botstein D, Cherry JM. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9. https://doi.org/10.1038/75556.
- Minoru K, Susumu G, Shuichi K, Yasushi O, Masahiro H. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32:D277. https://doi. org/10.1093/nar/gkh063.
- Chen L, Yang J, Yu J, Yao Z, Jin Q, Chen L. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33(suppl 1):D325–8. https://doi.org/10.1093/nar/gki008.
- 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(suppl 1):D459–464. https://doi.org/10.1093/nar/gkj047.

Chu et al. BMC Genomic Data (2024) 25:86 Page 5 of 5

- Saier MH, Tran CV, Barabote RD. TCDB: the transporter classification database for membrane transport protein analyses and information. Nucleic Acids Res. 2006;34(suppl 1):D181–6. https://doi.org/10.1093/nar/gkj001.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009;37(suppl 1):D233–8. https://doi.org/10.1093/nar/gkn663.
- 32. Bo L, Mihai P. ARDB—antibiotic resistance genes database. Nucleic Acids Res. 2009;37(suppl 1):D443–7. https://doi.org/10.1093/nar/gkn656.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinos KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114. https://doi.org/10.1038/ s41467-018-07641-9.
- Chen F, Mackey AJ, Stoeckert CJ, Roos D. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006;34(Database issue):D363–8. https://doi.org/10.1093/nar/gkj123.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21. https://doi.org/10.1093/sysbio/syq010.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.