BMC Genomic Data

DATA NOTE

Open Access

Complete genome sequence of *Paracidovorax avenae* causing red stripe in sugarcane

Hai-Hong Duan[1], Jian-Ying Zhao[1], Wei-Ting Liu[1], Quan Xie[1], Chu-Yang Zheng[1] and San-Ji Gao[1*]

Abstract

Objective *Paracidovorax avenae* (*Pa*) is the causative agent of red stripe disease in sugarcane and belongs to the Gram-negative β-Proteobacteria. Red stripe is a major bacterial disease of sugarcane worldwide. Limited genome sequences of *Pa* can be used for exploring the phylogenetic and genetic diversity analysis in this pathogen at the complete genome level. In this study, a whole genome sequence of *Pa* CNGX08 strain isolated from sugarcane in China was assembled and annotated.

Data description Genome assembly data from second- and third-generation sequencing revealed that the entire genomic sequence of *Pa* CNGX08 strain causing red stripe in sugarcane, consisted of a 5,625,582 bp circular chromosome with a GC content of 68.97%. In total, 4,915 protein-coding genes were annotated. Additionally, 9 ribosomal RNAs and 52 transfer RNAs were identified. This genomic resource will facilitate the genome-based taxonomic classification of the genus *Paracidovorax* and the exploration of pathogenic mechanisms underlying sugarcane red stripe disease caused by *Pa*.

Keywords *Paracidovorax avenae*, Genome assembly, Sugarcane (*Saccharum* spp.), Red stripe disease

Objective

Red stripe is one of the three major bacterial diseases in sugarcane and is prevalent in many sugarcane-growing regions worldwide [1]. This disease was first reported in the 1920s in Hawaii [2] and now has commonly occurred in Argentina [3] and China [4]. Two symptoms of this disease are leaf stripe and top rot, appearing separately or simultaneously in the field [2]. Economic losses are caused by red stripe, especially the top rot form [5]. Red stripe is caused by *Acidovorax avenae* subsp. *avenae* belonging to the Gram-negative β-Proteobacteria. Recently, Du et al. (2023) [6] proposed the division of the genus *Acidovorax* into the two novel genera *Paracidovorax* gen. nov. and *Paenacidovorax* gen. nov. Thus, the *Acidovorax avenae* subsp. *avenae* was renamed as *Paracidovorax avenae* (*Pa*). Here, we report the complete and annotated genome sequence of *Pa* strain CNGX08 causing sugarcane red stripe in Guangxi province, China. This work provides valuable insights into the comparative genome analysis among global *Pa* strains and the molecular mechanisms of interaction between *Pa* and host sugarcane.

Data description

The *Pa* CNGX08 was originally isolated from sugarcane leaf showing red stripe symptoms in the Guangxi province, China (24.5285°N, 109.2613°E). The CNGX08 forms white-cream colored colonies on nutrient agar (NA) medium [7]. Bacterial genome DNA was extracted using the SDS method, then detected by agarose gel

*Correspondence:
San-Ji Gao
gaosanji@fafu.edu.cn
[1]National Engineering Research Center for Sugarcane, Fujian Agriculture and Forestry University, Fuzhou 350002, China

electrophoresis and quantified using a Qubit® 2.0 Fluorometer (Thermo Scientific, China). The Blue Pippin system was utilized to automatically isolate large DNA fragments, which were subsequently subjected to a repair process. The incorporation of a unique barcode was facilitated by the PCR-free EXP-NBD104 kit from Oxford Nanopore Technologies (ONT), ensuring minimal bias and maintaining the integrity of the DNA. Following this, the Advanced Analytical Technologies, Inc. (AATI) capillary electrophoresis device was employed to precisely measure the fragment sizes, ensuring they met the required specifications. The samples were then mixed to achieve an isomolar concentration, allowing for an equal representation of each DNA fragment. The final step involved the use of the SQK-LSK109 kit (Oxford Nanopore Technologies, Oxford, UK) to attach the necessary adapters, culminating in the successful assembly of the 10 K sequencing library.

Sequencing libraries were prepared using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA), with index codes added to attribute sequences to each sample. The whole genome of CNGX08 was sequenced using the Nanopore PromethION platform and the Illumina NovaSeq PE150 at the Beijing Novogene Bioinformatics Technology Co., Ltd. The genome was assembled with all processed long-reads and short-reads using Unicycler (version 0.4.8) [8]. Briefly, the 150-bp paired-end reads were filtered and trimmed using fastp (v0.23.0). A frame diagram genome was assembled using SPAdes, and long-reads were added to the frame map using miniasm and Racon. A complete genome of CNGX08 was 5,625,582 bp long with a GC content of 68.97% (Data file 1, Table 1) [9] and was generated using Unicycler [8].

Subsequently, the BUSCO v4.0.2 tool (http://busco.ezlab.org/) was used to assess the final genome assembly. BUSCO (98.4%) analyses suggest that our genome assembly exhibits a high degree of completeness. The CNGX08 genome was generated into one circular chromosome harboring a length of 5,625,582 bp with 68.97% GC content (Data file 1) [9]. As shown in Table 1, circos (version 0.64) was used to display the genome structure and annotation of CNGX08 (Data file 2) [10]. The genomic sequence was uploaded to the National Center for Biotechnology Information (NCBI) GenBank database with the accession number CP156079 under the Bioproject number PRJNA1006162, as shown in Data file 3 [11].

Moreover, the coding genes were predicted using GeneMarkS (version 4.17). The signal peptides and transmembrane structures of proteins were analyzed by the SignalP (version 4.1) [12], while the secreted protein was comprehensively predicted with the TMHMM (version 2.0c). Among these annotation results of the protein sequence function database, the secretion system proteins and the type III secretion system (T3SS) effector proteins were predicted using EffectiveT3 (version 1.0.1) [13] (Data file 1) [9]. A total of 310 T3SS effective proteins were found. The genome of CNGX08 contains 4,915 protein-coding genes, 41 repeat regions, 8 genomic islands, 9 bacteriophages, and 4 clustered regularly spaced short palindromic repeats (CRISPR) (Data file 1) [9]. Gene functions were predicted with GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Clusters of Orthologous Groups), NR (Non-Redundant Protein Database), TCDB (Transporter Classification Database), and Swiss-Prot. A BLAST search (E-value less than 1×10^{-5}, minimal alignment length percentage larger than 40%) was performed against these databases. Pathogenicity and drug resistance analyses were carried out using the PHI (Pathogen Host Interactions) database. These results are shown in Supplementary Data file 4 [14].

The genomic sequences of CNGX08 and 19 worldwide strains of *Pa* along with *P. facilis* strain DSM 649 (as an outgroup) were used for collinearity analysis, population evolution analysis, SNP/INDEL analysis, and Average nucleotide identity (ANI) analysis. A phylogenetic tree was constructed using TreeBeST with the 1,000 bootstraps based on the core pan-genome level. All tested *Pa* strains were clustered into several clades, while the CNGX08 was clustered in a unique clade. ANI analysis showed that the CNGX08 were 93.32-94.04% identities

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	Genome statistics, genomic features, protein prediction, and components of the *Paracidovorax avenae* strain CNGX08	Word file (.docx)	Figshare https://doi.org/10.6084/m9.figshare.26054092 [9]
Data file 2	ONT sequencing read quality control, circular genome representation of the *Paracidovorax avenae* strain CNGX08	Word file (.docx)	Figshare https://doi.org/10.6084/m9.figshare.26054230 [10]
Data file 3	Genome assembly of *Paracidovorax avenae* strain CNGX08	Fasta file (.fasta)	NCBI GenBank: https://identifiers.org/ncbi/insdc:CP156079 [11]
Data file 4	Annotation of protein-coding genes in the *Paracidovorax avenae* strain CNGX08	Excel file (.xlsx)	Figshare https://doi.org/10.6084/m9.figshare.26054257 [14]
Data file 5	Phylogenetic analysis and average nucleotide identities of worldwide *Paracidovorax avence* strains including CNGX08	Word file (.docx)	Figshare https://https://doi.org/10.6084/m9.figshare.26054461 [15]

with other *Pa* strains (Data file 5) [15]. Notably, the CNGX08 shared the sequence identities of 93.43% and 93.45% with the T10_61 strain infecting sugarcane in Argentina and the reference strain of ATCC 19860, respectively. These results suggested that CNGX08 is a distinct *Pa* strain from T10_61 and ATCC 19860 strains. However, it is necessary that more genomic sequences of *Pa* strains from various plant hosts and geographical origins were sequenced and assembly for further illustrating the divergence and taxonomic classification of this pathogen.

In conclusion, this study demonstrated genomic information of the *Pa* strain CNGX08 infecting sugarcane in China, providing an understanding of the evolutionary history of the *Paracidovorax* genus. The availability of this data serves as a reference for comparative genomics of *Pa* across different hosts.

Limitations

This Data Note presents the whole-genome sequencing of the *Pa* CNGX08 strain based on the combination of the Nanopore PromethION (third-generation) and the Illumina NovaSeq PE150 (second-generation) platforms, providing a high-quality and gap-free reference genome. Therefore, the authors believe that there are no limitations in the data.

Abbreviations

ANI	Average nucleotide identity
bp	Base pair
BLAST	Basic Local Alignment Search Tool
DNA	Deoxyribonucleic acid
NCBI	National Center for Biotechnology Information
ONT	Oxford Nanopore Technology

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-024-01271-5.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Author contributions

HHD, WTL, QX and CYZ: performed strain isolation, cultivation and DNA extraction. JYZ and HHD: performed the genome analysis. HHD and JYZ: prepared the manuscript draft. SJG: 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

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 18 June 2024 / Accepted: 11 October 2024
Published online: 22 October 2024

References

1. Chu N, Sun HD, Zhou JR, Fu HY, Li XY, Gao SJ. Research advances of sugarcane red stripe disease and its pathogeny biology. Sugar Crops China. 2020;42(1):66–70. https://doi.org/10.13570/j.cnki.scc.2020.01.012.
2. Rott P, Davis MJ. Red stripe (top rot). In: Rott P, Bailey RA, Comstock JC, Croft BJ, Saumtally AS, editors. A guide to Sugarcane diseases. Montpellier. France: CIRAD/ISSCT; 2000. pp. 58–62.
3. Fontana PD, Rago AM, Fontana CA, Vignolo GM, Cocconcelli PS, Mariotti JA. Isolation and genetic characterization of *Acidovorax avenae* from red stripe infected sugarcane in northwestern Argentina. Eur J Plant Pathol. 2013;137(3):525–34. https://doi.org/10.1007/s10658-013-0263-y.
4. Li XY, Sun HD, Rott PC, Wang JD, Huang MT, Zhang QQ, Gao SJ. Molecular identification and prevalence of *Acidovorax avenae* subsp. *avenae* causing red stripe of sugarcane in China. Plant Pathol. 2018;67:929–37. https://doi.org/10.1111/ppa.12811.
5. Martin JP, Wismer CA. Red stripe. In: Ricaud C, Egan BT, Gillaspie AG, Hughes CG, editors. Disease of Sugarcane. Major diseases. Amsterdam, Netherlands: Elsevier Science Publishers BV; 1989. pp. 81–95. https://doi.org/10.1016/B978-0-444-42797-7.50009-0.
6. Du J, Liu Y, Zhu H. Genome-based analyses of the genus Acidovorax: proposal of the two novel genera *Paracidovorax* gen. nov., *Paenacidovorax* gen. nov. and the reclassification of *Acidovorax antarcticus* as *Comamonas Antarctica* comb. nov. and emended description of the genus *Acidovorax*. Arch Microbiol. 2023;205(1):42. https://doi.org/10.1007/s00203-022-03379-7.
7. Zhao JY, Chen J, Hu ZT, Li J, Fu HY, Rott PC, Gao SJ. Genetic and morphological variants of *Acidovorax avenae* subsp. *avenae* cause red stripe of sugarcane in China. Front Plant Sci. 2023;14:1127928. https://doi.org/10.3389/fpls.2023.1127928.
8. Wick RR, Judd LM, Gorrie C, Holt KE, Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
9. Data file 1: Genome statistics, Genomic features, protein prediction and components of the *Paracidovorax avenae* strain CNGX08. Figshare. (2024).
10. Data file 2: ONT sequencing read quality control, circular genome representation of the *Paracidovorax avenae* strain CNGX08. Figshare. (2024).
11. Data file 3: NCBI GenBank. 2024. https://identifiers.org/ncbi/insdc:CP156079. Accessed 21 Jul 2024.
12. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8(10):785–6. https://doi.org/10.1038/nmeth.1701.
13. Eichinger V, Nussbaumer T, Platzer A, Jehl MA, Arnold R, Rattei T. EffectiveDB-updates and novel features for a better annotation of bacterial secreted proteins and type III, IV, VI secretion systems. Nucleic Acids Res. 2016;44(D1):D669–74. https://doi.org/10.1093/nar/gkv1269.
14. Data file 4. Annotation of protein-coding genes in the *Paracidovorax avenae* strain CNGX08. Figshare. (2024).
15. Data file 5. Phylogenetic analysis and average nucleotide identities of worldwide *Paracidovorax avence* strains including CNGX08. Figshare. (2024).

Publisher's note