RESOURCE ANNOUNCEMENT

Complete Genome Sequence Resource of *Pantoea anthophila* CL1 Causing Soft Rot Disease in *Clausena lansium* (Wampee) in China

Qian He, Juming Li, Qingmei Liu, Yonglin Li, Xinyue Liao, Jianuan Zhou, and Xiaofan Zhou†

Guangdong Laboratory for Lingnan Modern Agriculture, Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou 510642, China

Abstract

Pantoea anthophila CL1 is a causal agent of soft rot disease in *Clausena lansium* (wampee) in China and has inhibitory activity against the bacterial wilt pathogen *Ralstonia solanacearum*. Here we report the genome sequencing and analysis of *P. anthophila* CL1, representing the first complete genome resource of the species. The CL1 genome consists of four circular replicons (one chromosome and three plasmids), with a total size of 4,594,065 bp, and contains 4,109 protein-coding genes and 106 RNA genes. Our bioinformatic analysis of CL1 predicted 228 virulence factors, two Type VI Secretion Systems, and six secondary metabolite biosynthesis gene clusters producing saccharides, siderophores, and terpene. The complete genome sequence of *P. anthophila* CL1 provides a solid foundation for further investigation of its pathogenesis and antimicrobial activity and also represents a valuable resource for the comparative genomics of *Pantoea*.

Funding

Funding was provided by the Guangdong General University Key Area Research Project (2021KTSCX013), Key-Area Research and Development Program of Guangdong Province (grant numbers 2018B020205003, 2020B0202090001), and the Guangdong Forestry Science and Technology Innovation Project (grant numbers 2018KJCX009, 2020KJCX009).

Keywords

complete genome sequence, *Pantoea anthophila*, virulence factor, wampee soft rot

Resource Announcement

The genus *Pantoea* is a highly diverse group of gram-negative bacteria that are distributed globally, ubiquitous in various environments, and associated with a wide range of animals and plants (Walterson and Stavrinides 2015). In particular, *P. agglomerans*, *P. ananatis*, *P. stewartia*, and other members of *Pantoea* have been recognized as pathogens of many economically important plants (e.g., rice, maize, and cotton) and are of important agricultural relevance (Lv et al. 2022; Silva-Rojas et al. 2012; Tufail et al. 2020). At the same time, *Pantoea* also contains beneficiary strains that can be utilized for environmental modification, biocontrol of plant diseases, and plant growth promotion (Ishimaru et al. 2017; Pileggi et al. 2012). In our previous work, we identified *P. anthophila* CL1 as a causal agent of soft rot disease and cracking on *Clausena lansium* (wampee), a fruit widely cultivated in Asia (Zhou et al. 2015). Other *P. anthophila* strains have also been reported to infect flowering plants naturally (Brady et al. 2009). Additionally, our further investigation found that *P. anthophila* CL1 could also exhibit antagonistic activity against *Ralstonia solanacearum*, one of the most notorious bacterial plant pathogens (Mansfield et al. 2012). However, the study of *P. anthophila* is hindered by its rather limited genomic information; currently, only three *P. anthophila* genomes are available in the NCBI GenBank database, all of which are incomplete with tens of contigs (Table 1).

Here, we sequenced the genome of *P. anthophila* CL1 using both second- and third-generation high-throughput sequencing technologies. Genomic DNA was extracted from CL1 in Luria-Bertani medium culture using EasyPure Bacteria Genomic DNA Kit (Transgen, Beijing, China). The harvested DNA was detected by agarose gel electrophoresis and

†Corresponding author: X. Zhou; xiaofan_zhou@scau.edu.cn

Q. He, J. Li, and Q. Liu contributed equally to this work.

The author(s) declare no conflict of interest.

Accepted for publication 27 November 2022.

Table 1. Genome features of *Pantoea anthophila* CL1 and three other *P. anthophila* strains

	Pantoea anthophila strains			
Genome features	**CL1**	**11-2**	**LMG 2558**	**EKM101A**
Genome size (bp)	4,593,785	4,600,679	4,699,058	4,687,858
Number of scaffolds	4	16	66	75
N50 (bp)	3,909,012	453,410	238,228	537,232
GC content	56.88%	56.77%	56.76%	56.62%
Genes (total)	4,257	4,312	4,437	4,452
Coding genes	4,109	4,171	4,265	4,298
Pseudo-genes	42	53	76	72
rRNAs	22	15	25	9
tRNAs	77	67	75	66
ncRNAs	7	6	6	7

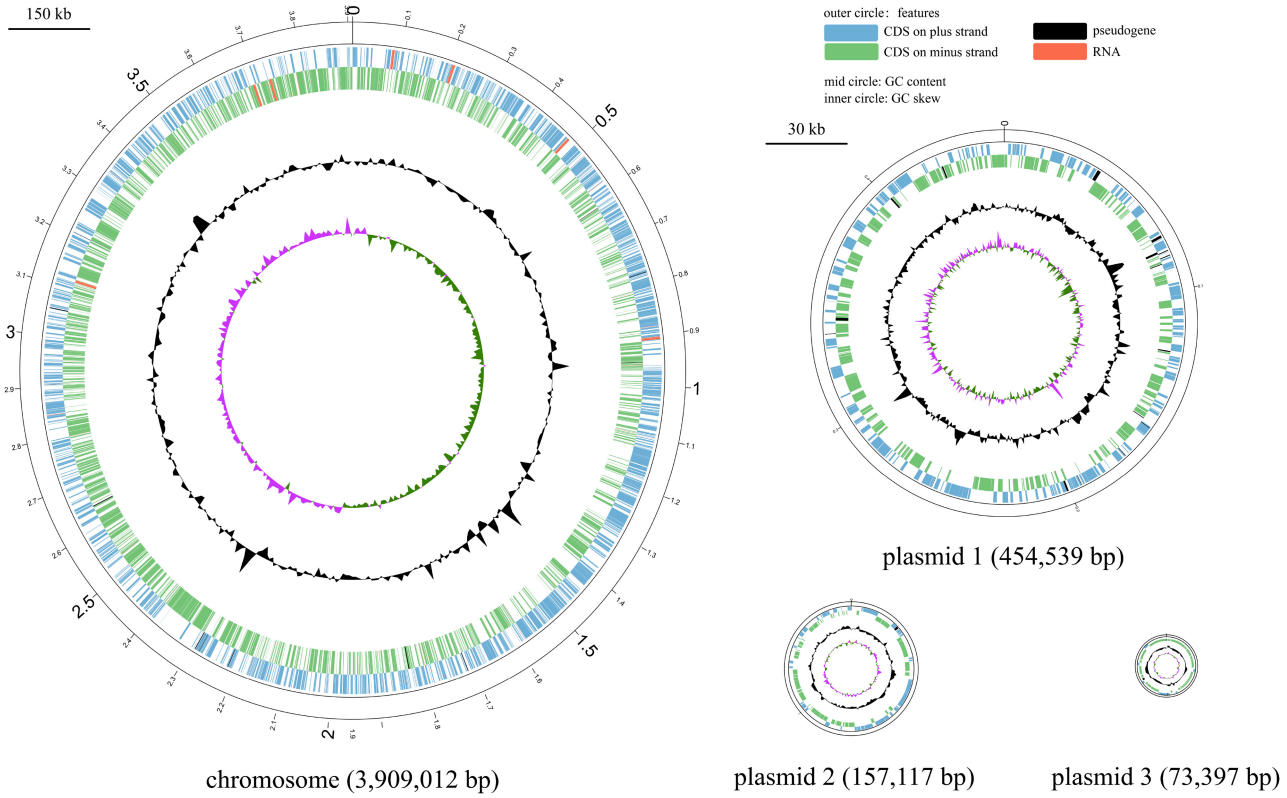

Fig. 1. Circular maps of the chromosome and plasmids of *Pantoea anthophila* CL1. In each map, the circles from outside to inside represent the following: (i) features of the positive strand, showing coding sequence (CDS) (blue), rRNA (red), and pseudogenes (black); (ii) features of the negative strand, showing CDS (green), rRNA (red), and pseudogenes (black); (iii) GC content; and (iv) GC-skew (pink and green indicate values higher and lower than the mean value, respectively). It should be noted that the plasmid maps are displayed at a different scale from the chromosome.

quantified by Qubit 2.0 Fluorometer (Thermo Scientific, Waltham, MA, U.S.A.). Fragmented genomic DNA with an average size of 300 bp and long DNA fragments with an average size of 23 kbp were selected for library preparation and sequenced on BGISEQ-500 and Nanopore PromethION platforms, respectively, by Nextomics (Wuhan, China). In total, 1.53 Gb of short-read sequencing data (~5.1M pairs of PE150 reads) and 1.54 Gb of long-read sequencing data (82,964 reads with a mean length of 18,498 bp) were generated. The Nanopore sequencing data was assembled into four circular replicons using Flye v2.9.1 (Kolmogorov et al. 2019) with default options (same for all bioinformatic analyses unless specified otherwise), and the resulting assembly was polished with both long- and short-read data using NextPolish v1.4.1 (Hu et al. 2019). Almost identical results were obtained from a long-read-only assembly using NextDenovo v2.5.0 (https://github.com/Nextomics/NextDenovo) and a hybrid assembly of both short- and long-read sequencing data using Unicycler v0.5.0 (Wick et al. 2017).

The complete genome sequence of *P. anthophila* CL1 consists of one large chromosome (3,909,012 bp in size) and three smaller plasmids (454,539, 157,117, and 73,397 bp in size, respectively), all of which have a similar GC content of ~57% (Fig. 1). QUSAT v5.2.0 (Gurevich et al. 2013) assessment with sequencing data did not detect any assembly errors. Evaluations using BUSCO v5.4.2 (Manni et al. 2021) (with the options "-m geno" and "-l enterobacterales_odb10") and CheckM v1.1.6 (Parks et al. 2015) both indicate that the CL1 genome has a completeness level of 99.6%. Genome-based taxonomy classification using GTDB-Tk v2.1.0 (Chaumeil et al. 2022) confirmed that CL1 belongs to *P. anthophila* as it revealed a high Average Nucleotide Identity value (98.42%) between CL1 and the *P. anthophila* type strain LMG 2558. So, the CL1 genome represents the first complete genome assembly in *P. anthophila*, which is an important addition to the genomic resource of *Pantoea*.

Annotation of the *P. anthophila* CL1 genome using PGAP v2022-10-03. build6384 (Tatusova et al. 2016) identified 4,109 protein-coding genes, 42 pseudogenes, 22 tRNAs, 77 rRNAs, and seven ncRNAs (Table 1). Functional annotations of CL1 protein-coding genes were generated using the FACoP web server (de Jong et al. 2022). In addition, we compared the annotated proteins of CL1 against the VFDB database and predicted 228 potential virulence factors (BLASTp hits with e-value < 1e−20, identity ≥ 50%, and coverage ≥ 70%), including 52 proteins related to motility, 46 proteins related to immune modulation, 44 potential nutritional/metabolic factors, 35 proteins related to effector delivery system, 20 proteins related to adherence, and 31 proteins involved in various other processes such regulation, antimicrobial activity, and stress survival.

Notably, the annotation of CL1 did not predict any Type II or III Secretion Systems, which are important pathogenicity determinants in many plant pathogens. As T2SS and T3SS were also found to have patchy distributions in several other *Pantoea* species (De Maayer et al. 2017; Moretti et al. 2021), further genomic and experimental investigations are necessary to determine their presence/absence and functional roles in *Pantoea*. On the other hand, CL1 harbors two Type VI Secretion Systems (T6SSs), which might have important roles in the pathogenicity and bacterial competition of CL1. We also predicted in the CL1 genome six secondary metabolite biosynthesis gene clusters using antiSMASH v6.1.1 (Blin et al. 2019), three of which are highly similar to clusters producing known compounds (i.e., carotenoid, desferrioxamine E, and aryl polyenes), while the other three might produce novel saccharides and non-ribosomal peptides. These compounds, the two T6SSs, and other genes related to antimicrobial activity might collectively contribute to the inhibitory effect of CL1 against *R. solanacearum*. Interestingly, the other three *P. anthophila* strains are also predicted to have the same set of biosynthesis gene clusters and one to two T6SSs, suggesting that they might also have the same antagonistic activity as CL1.

In summary, the complete genome sequence of *P. anthophila* CL1 we present in this study provides a solid foundation for further investigation of its pathogenesis and antimicrobial activity and represents a valuable resource for the comparative genomics of *Pantoea*.

Data Availability

The complete genome sequence and raw genome sequencing data of *P. anthophila* CL1 have been deposited at National Center for Biotechnology Information (NCBI) under the GenBank accessions CP110471.1 to CP110474.1, Bioproject accession PRJNA893243, and Biosample accession SAMN31416257. The annotation data generated in this study has been deposited in the figshare repository at https://figshare.com/articles/online_resource/Complete_genome_assembly_and_annotation_of_Pantoea_anthophila_CL1/21385029.

Literature Cited

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., and Weber, T. 2019. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47:W81-W87.

Brady, C. L., Venter, S. N., Cleenwerck, I., Engelbeen, K., Vancanneyt, M., Swings, J., and Coutinho, T. A. 2009. *Pantoea vagans* sp. nov., *Pantoea eucalypti* sp. nov., *Pantoea deleyi* sp. nov and *Pantoea anthophila* sp. nov. Int. J. Syst. Evol. Microbiol. 59:2339-2345.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. 2022. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925-1927.

de Jong, A., Kuipers, O. P., and Kok, J. 2022. FUNAGE-Pro: Comprehensive web server for gene set enrichment analysis of prokaryotes. Nucleic Acids Res. 50: W330-W336.

De Maayer, P., Aliyu, H., Vikram, S., Blom, J., Duffy, B., Cowan, D. A., Smits, T. H. M., Venter, S. N., and Coutinho, T. A. 2017. Phylogenomic, pan-genomic,

pathogenomic and evolutionary genomic insights into the agronomically relevant enterobacteria *Pantoea ananatis* and *Pantoea stewartii*. Front. Microbiol. 8:1755.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

Hu, J., Fan, J., Sun, Z., and Liu, S. 2019. NextPolish: A fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36:2253-2255.

Ishimaru, C. A., Lansdell, T. A., Clardy, J., Duffy, B., and Smits, T. H. M. 2017. The histidine-reversible antibiotic herbicolin O produced by *Pantoea vagans* C9-1 is pantocin A. J. Plant Pathol. 99:91-97.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37:540-546.

Lv, L., Luo, J., Ahmed, T., Zaki, H. E. M., Tian, Y., Shahid, M. S., Chen, J., and Li, B. 2022. Beneficial effect and potential risk of *Pantoea* on rice production. Plants 11:2608.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. 2021. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. 38:4647-4654.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S. V., Machado, M. A., Toth, I., Salmond, G., and Foster, G. D. 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol. Plant Pathol. 13:614-629.

Moretti, C., Rezzonico, F., Orfei, B., Cortese, C., Moreno-Pérez, A., van den Burg, H. A., Onofri, A., Firrao, G., Ramos, C., Smits, T. H. M., and Buonaurio, R. 2021. Synergistic interaction between the type III secretion system of the endophytic bacterium *Pantoea agglomerans* DAPP-PG 734 and the virulence of the causal agent of olive knot *Pseudomonas savastanoi* pv. *savastanoi* DAPP-PG 722. Mol. Plant Pathol. 22:1209-1225.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25:1043-1055.

Pileggi, M., Pileggi, S. A. V., Olchanheski, L. R., da Silva, P. A. G., Munoz Gonzalez, A. M., Koskinen, W. C., Barber, B., and Sadowsky, M. J. 2012. Isolation of mesotrione-degrading bacteria from aquatic environments in Brazil. Chemosphere 86:1127-1132.

Silva-Rojas, H. V., Esker, P. D., Mahuku, G., Peralta-Nava, J. A., Cortes-Cueto, A. L., and Rodriguez-Herrera, J. A. 2012. *Pantoea ananatis*, a maize plant pathogenic bacterium associated with chlorotic streaks and rolling upper leaves. Phytopathology 102:6.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., and Ostell, J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44:6614-6624.

Tufail, M. R., Fayyaz, A., Sarfraz, S., Waqas, M., Ahmad, M. Z., Hameed, A., Masood, A., Abdin, Z. U., Amrao, L., Sahi, S. T., and Abbas, A. 2020. First report of Pantoea leaf blight of cotton caused by *Pantoea spp*. in Punjab, Pakistan. Plant Dis. 104:1534.

Walterson, A. M., and Stavrinides, J. 2015. *Pantoea*: Insights into a highly versatile and diverse genus within the Enterobacteriaceae. FEMS Microbiol. Rev. 39: 968-984.

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13:e1005595.

Zhou, J. N., Liu, S. Y., Chen, Y. F., and Liao, L. S. 2015. First report of *Pantoea anthophila* causing soft rot disease in *Clausena lansium* (wampee) in China. Plant Dis. 99:416.