Complete Genomic Data of *Enterobacter asburiae* Strain SD4L Associated with Bacterial Palea Browning of Rice in China

Peng Cao,¹ Chenxu Li,¹ Kefei Tan,² Junwei Zhao,^{1,†} Xiangjing Wang,¹ and Wensheng Xiang^{1,3,†}

¹ Key Laboratory of Agricultural Microbiology of Heilongjiang Province, Northeast Agricultural University, No. 600 Changjiang Road, Xiangfang District, Harbin 150030, P.R. China

² Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar, P.R. China

³ State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, P.R. China

Abstract

Enterobacter asburiae is the causal agent of rice bacterial palea browning disease. Here, we report the complete genome of *E. asburiae* strain SD4L, which represents the first whole genome sequence of an isolate from rice seedlings in China. The assembled genome consisted of two contigs, with a circular chromosome of 4,574,166 bp, and a plasmid of 85,271 bp, respectively. This complete genome will provide a valuable resource for further studies on bacterial palea browning worldwide.

Genome Announcement

Rice (*Oryza sativa* L.) is a staple food in different areas of world, especially in China. In recent years, rice seedlings have been affected seriously by bacterial palea browning (BPB) in Heilongjiang Province. BPB, typically caused by *Pantoea agglomerans* (Gavini et al. 1983, 1989; Kim et al. 1989; Tabei et al. 1988), *P. ananatis* (Azegami 2013; Cortesi and Pizzatti 2007), or *P. ananas* (Hasegawa et al. 2003; Mergaert et al. 1993), has been reported in previous studies. Infection of plants by these pathogens at the heading and flowering stages result in quality reduction, while infection at an early stage causes severe damage (Azegami 2013; Cao et al. 2020; Hasegawa et al. 2003). The most destructive phase of the disease is the wilt phase resulting from early systemic infection in the nursery or from seed infection (Cao et al. 2020; Tabei et al. 1988). In 2020, BPB caused by strain SD4L (representative strain) was first reported in Heilongjiang Province, northeastern China, by our laboratory and preliminarily identified as *Enterobacter cloacae* (Cao et al. 2020). This pathogen may present a threat to the production of rice in this area, and the highest incidence can reach approximately 30%. However, the complete genome sequence of the pathogenic bacteria, *E. cloacae* strain SD4L isolated from rice in China, remains to be determined. Interestingly, whole genome sequence analysis showed that strain SD4L should be classified into *E. asburiae* in this study. Although more than 100 *E. asburiae* strains have been sequenced and 16 strains of them have high-quality complete genomes based on National Center for Biotechnology Information (NCBI) accession, the complete genome sequences of *E. asburiae* strains that served as plant pathogenic bacteria have not been reported up to now. To boost research on etiology and pathogenesis of such bacterial infection, we first

Funding

This work was supported by grants from the National Natural Youth Science Foundation of China (no. 31701858), the China Postdoctoral Science Foundation (2018M631907), and the Heilongjiang Postdoctoral Fund (LBH-Z17015).

Keywords

Enterobacter asburiae strain SD4L, complete genome, rice, bacterial palea browning

†Corresponding authors: J. Zhao; guyan2080@126.com, and W. Xiang; xiangwensheng@neau.edu.cn

Peng Cao and Chenxu Li contributed equally to this work.

*The e-Xtra logo stands for “electronic extra” and indicates that supplementary materials are published online.

The author(s) declare no conflict of interest.

Accepted for publication 23 June 2021.

Table 1. Genomes comparison between strain SD4L and the type strain *E. asburiae* ATCC 35953^T

Features	SD4L	ATCC 35953 ^T
Accession	CP063441	CP011863
BioSample	SAMN16515958	SAMN03742638
BioProject	PRJNA224116	PRJNA224116
Assembly	GCF_015168595	GCF_001521715
Size	4,574,166	4,713,742
Genes	4,398	4,746
CDSs	4,231	4,633
tRNAs	83	83
rRNAs	25	5
ncRNAs	7	5
Pseudogenes	52	214
Host	Rice	<i>Homo sapiens</i>

report the complete whole-genome sequence of the *E. asburiae* strain SD4L isolated from rice seedlings in China in the present study.

Genomic DNA of *E. asburiae* strain SD4L was extracted with the sodium dodecyl sulfate (SDS) method (Peršoh et al. 2008). The harvested DNA was detected by agarose gel electrophoresis and quantified by a Qubit 2.0 Fluorometer (Thermo Scientific, Waltham, MA). Two whole-genome sequencing technologies, Nanopore PromethION platform (Oxford Nanopore Technologies, Oxford, UK) and Illumina NovaSeq platform (Illumina, San Diego, CA), were used to generate long reads and short reads, respectively. For PromethION, libraries were constructed with an insert size of 10 kb. First, large fragments of DNA were recovered by Blue Pippin (Sage Science, Beverly, MA) automatic nucleic acid fragment recovery system, and then repaired. Barcodes were added by the PCR-free method with the EXP-NBD104 kit (Oxford Nanopore Technologies). Afterward, the size of fragments was detected by AATI automatic capillary electrophoresis instrument (Advanced Analytical Technologies, Inc., Ankeny, IA), and the sample was isomolar mixed with other genome samples. Finally, the SQK-LSK109 connection kit (Oxford Nanopore Technologies) was used to link the adapter, and the 10K library was constructed. For Illumina NovaSeq platform, a total amount of 1 µg DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, Ipswich, MA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, the DNA sample was fragmented by sonication to a size of 350 bp, then DNA fragments were end-polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. Finally, PCR products were purified (AMPure XP system, Beckman Coulter, Brea, CA) and libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. After that, the whole genome of strain SD4L was sequenced using Nanopore PromethION platform and Illumina NovaSeq PE150 (pair-end 150 bp reads) at the Beijing Novogene Bioinformatics Technology Co., Ltd. For whole genome assembly, Unicycler software was used (<https://github.com/rnwick/Unicycler>) to combine PE150 data and Nanopore data; then the reads were compared with the assembled sequence to count the distribution of sequencing depth. Finally, the assembled sequence was determined as chromosomal sequences and/or plasmid sequences according to sequence length and alignment, and checked whether it is circular.

A total of 80,107 reads were obtained after sequencing by Nanopore PromethION platform, with an N50 value of 17,058 bp and average read length of 13,639 bp, consisting of 1,092,595,251 bases in total, and a genome coverage of 240x. For Illumina NovaSeq platform, it generated 8,977,876 high-quality reads, with a genome coverage of 310x. The final complete genome of *E. asburiae* SD4L was composed of a single, continuous, circular chromosome of 4,574,166 bp in size, with a 55.97% GC content and a plasmid consisted of 85,271 bp, with a 51.32% GC content. It was automatically annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Haft et al. 2018) and the RAST tool kit (RASTtk) (Overbeek et al. 2014). Functional analysis was done by using Clusters of Orthologous Groups of Proteins (COGs) (Galperin et al. 2015), Gene Ontology (GO) (Ashburner et al. 2000), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa

et al. 2016) annotation. Identification of genomic islands (GIs) was performed using the IslandPath-DIMOB (Version 0.2), and Islander prediction methods (Bertelli et al. 2017). Average nucleotide identity (ANI) was calculated using the ChunLab online ANI calculator (<https://www.ezbiocloud.net/tools/ani>) (Yoon et al. 2017). Strain SD4L was preliminarily identified as *E. cloacae* based on the multilocus sequence analysis and phenotypic and biochemical characterization in previous study (Cao et al. 2020). In this study, the ANI score was greater than 95%, to *E. asburiae* strain ATCC 35953^T (NCBI accession no. GCA_001521715) isolated from *Homo sapiens*. Therefore, strain SD4L was further identified as *E. asburiae*. Moreover, a total of 4,398 genes were annotated in the *E. asburiae* SD4L strain genome, of which 4,231 genes were assigned as putative protein-coding genes (CDSs). In addition, 115 RNA-encoding genes including 25 ribosomal RNA (rRNA) genes, 83 transfer RNA (tRNA) genes, and 7 noncoding RNA (ncRNA) genes, as well as 52 pseudogenes were predicted. By using functional protein alignments, 3,674 proteins were assigned to COGs database, and 3,155 proteins were assigned to GO terms, and 4,189 proteins were mapped to KEGG pathways. By using the IslandPath-DIMOB, 9 GIs were identified. The annotated features are summarized in Table 1, and the circular representation of the complete genome of *E. asburiae* strain SD4L is shown in Supplementary Fig. S1.

To the best of our knowledge, this is the first reported genome of *E. asburiae* strain SD4L isolated from rice. Our data provide a starting platform for comparative genomics of other *Enterobacter* spp., which are responsible for rice bacterial palea browning disease in the main rice cultural region, and clarify the pathogenesis of *E. asburiae* SD4L in rice.

Data Availability

The whole-genome sequences reported here have been deposited in DDBJ/ENA/GenBank, under the accession number CP063441 for chromosome 1, and CP063442 for plasmid 1. The versions described in this paper are the first version. The Nanopore and Illumina raw reads have been deposited in the Sequence Read Archive (SRA) under the accession numbers SRR13212344 and SRR13180197, respectively.

Literature Cited

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25:25-29.
- Azegami, K. 2013. Suppressive effect of bacteriophage on bacterial palea browning of rice caused by *Pantoea ananatis*. *J. Gen. Plant Pathol.* 79: 145-154.
- Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., Winsor, G. L., and Brinkman, F. S. L. 2017. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45:W30-W35.
- Cao, P., Li, C., Tan, K., Liu, C., Xu, X., Zhang, S., Wang, X., Zhao, J., and Xiang, W. 2020. Characterization, phylogenetic analyses, and pathogenicity of *Enterobacter cloacae* on rice seedlings in Heilongjiang Province, China. *Plant Dis.* 104:1601-1609.
- Cortesi, P., and Pizzatti, C. 2007. Palea browning, a new disease of rice in Italy caused by *Pantoea ananatis*. *J. Plant Pathol.* 89:S76.
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43:D261-D269.
- Galvani, F., Lefebvre, B., and Leclerc, H. 1983. Etude taxonomique de souches appartenant ou apparentées au genre *Erwinia*, groupe *Herbicola*, et à l'espèce *Enterobacter agglomerans*. *Syst. Appl. Microbiol.* 4:218-235.
- Galvani, F., Mergaert, J., Beji, A., Mielcarek, C., Izard, D., Kersters, K., and De Ley, J. 1989. Transfer of *Enterobacter agglomerans* (Beijerinck 1888) Ewing and Fife 1972 to *Pantoea* gen. nov. as *Pantoea agglomerans* comb. nov. and Description *Pantoea dispersa* sp. nov. *Int. J. Syst. Bacteriol.* 39:337-345.
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetverin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., et al. 2018. RefSeq: An Update on Prokaryotic Genome Annotation and Curation. *Nucleic Acids Res.* 46:D851-D860.
- Hasegawa, M., Azegami, K., Yoshida, H., and Otani, H. 2003. Behavior of *Erwinia ananas* transformed with bioluminescence genes on rice plants. *J. Gen. Plant Pathol.* 69:267-270.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44:D457-D462.
- Kim, Y. C., Kim, K. C., and Choi, B. H. 1989. Palea browning disease of rice caused by *Erwinia herbicola* and ice nucleation activity of the pathogenic bacterium. *Korean J. Plant Pathol.* 5:72-79.
- Mergaert, J., Verdonck, L., and Kersters, K. 1993. 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. Bacteriol.* 43:162-173.
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., et al. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42:D206-D214.
- Peršoh, D., Theuerl, S., Buscot, F., and Rambold, G. 2008. Towards a universally adaptable method for quantitative extraction of high-purity nucleic acids from soil. *J. Microbiol. Methods* 75:19-24.
- Tabai, H., Azegami, K., and Fukuda, T. 1988. Infection site of rice grain with *Erwinia herbicola*, the causal agent of bacterial palea browning of rice. *Ann. Phytopathol. Soc. Jpn.* 54:637-639.
- Yoon, S. H., Ha, S. M., Lim, J., Kwon, S., and Chun, J. 2017. A large scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 110:1281-1286.