Genome sequence of the rice pathogenic bacterium Acidovorax avenae subsp. avenae RS-1

Guan-Lin Xie1, Guo-Qing Zhang1, He Liu1, Miao-Miao Lou1, Wen-Xiao Tian1, Bin Li1, Xue-Ping Zhou1, Bo Zhu1§, Gu-Lei Jin2§

1State Key Laboratory of Rice Biology and Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Ministry of Agriculture, Institute of Biotechnology, Zhejiang University, Hangzhou 310029, China
2Institute of Bioinformatics, Zhejiang University, Hangzhou 310029, China

§Corresponding author:
Bo Zhu: bzhu@zju.edu.cn; Phone: +86-571-86971412
Gu-Lei Jin: guleijin@zju.edu.cn; Phone: +86-571-88982731

Key words: Genome sequencing; Acidovorax avenae subsp. avenae; Rice; Pathogenic bacterium
Abstract

*Acidovorax avenae* subsp. *avenae* is a phytobacterium which is the causal agent of several economic plant diseases. Here, we present the draft genome sequence of strain RS-1, which was isolated from rice shoot in rice field of China. This strain can cause bacterial stripe of rice.

*Acidovorax avenae* subsp. *avenae*, formerly *Pseudomonas avenae* (10), can cause diseases in many economic plants, including rice, corn, oats, sugarcane, millet and foxtail (9). In rice, this pathogen can cause bacterial brown stripe and has been reported in many countries in Asia, Africa, the Americas and Europe (11). The symptoms started as brown stripes on the bottom of stems 5 days after emergence and frequently extended into the sheaths, then spreading along the leaf midrib and throughout the entire seedling at the one-leaf stage (3). As *A. avenae* subsp. *avenae* is a widely distributed seed-born pathogen of rice (8-9), the rice seeds contaminated with this pathogen are important sources of primary inoculum and a means of dissemination of the pathogen to new areas (1, 8). So, this pathogen has gained more and more attention in China.

We sequenced and annotated the draft genome of *A. avenae* subsp. *avenae* strain RS-1, a strain isolated from diseased rice shoot at Zhejiang province in 2010 (3). The genomic DNA, isolated using Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) was whole-genome sequenced by using Illumina HiSeq 2000. This resulted in 17,300,242 high quality filtered reads of 96 bp average read length.
and coverage equivalent to about 200 times. Quality filtered reads were assembled in silico with SOAPdenovo and GapCloser program (4). Based on the reference genome of *A. avenae* subsp. *avenae* ATCC 19860 (sequenced by JGI; isolated from maize leaf), a draft genome of RS-1 was completed. By subsequent PCR and re-sequencing, 62 genome gaps were closed, but remaining 156 contigs.

The draft genome sequence of strain RS-1 comprises 5,522,282 bases representing approximately more than 99.9% of the estimated genome size of the RS-1. The genome of this strain has a high G+C content of 68.7%. A total of 5,043 coding sequences (CDSs) were predicted using GLIMMER (7). Putative functions of encoding genes were automatically identified using GenDB annotation engine (6). The chromosome has three rRNA operons and 43 tRNAs predicted by RNAmmer and tRNAscan (2, 5). Furthermore, 90.4% of the ORFs have orthologs in the reference strain *A. avenae* subsp. *avenae* ATCC 19860 (BLASTP < 1e-5), but 301 ORFs were not found in the released genomes of *Acidovorax* genus. Interestingly, many of these ORFs are clustered together. The result suggested that these regions may be genomic islands in *A. avenae* subsp. *avenae* RS-1.

Pathogenicity related genes (such as those involved in hypersensitive response and the type III secretion system related proteins), which are essential for many plant pathogenic bacteria, are found in RS-1. Meanwhile, several loci encoding proteins for fimbrial biosynthesis, adhesion like proteins and RND efflux system proteins are also found. Overall, the genome sequence of *A. avenae* subsp. *avenae* RS-1 provides a foundation for both basic and agricultural applied researches.
Nucleotide sequence accession numbers

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AFPT00000000. The version described in this paper is the first version, AFPT01000000.

Acknowledgements

This work was supported by the Special Fund for Agro-scientific Research in the Public Interest (201003029 ; 201003066), the National Natural Science Foundation of China (30871655), and Zhejiang Agri. Dep. (Ji-fa 2008-97).
