Pantoea ananatis belongs within the Enterobacteriaceae family and the genus Pantoea, which hosts a number of important plant pathogens and clinically relevant species. It has been isolated from a range of plant hosts worldwide, and its ability to cause disease in several important staple food crops, including maize, onion and rice, has led to its being recognized as an important emerging plant pathogen (3). In South Africa, P. ananatis causes severe blight and dieback of Eucalyptus hybrid and clone seedlings/cuttings, which affects the pulp and paper industry (4). For this reason, we chose to sequence the genome of the highly virulent Eucalyptus-pathogenic strain P. ananatis LMG20103. Previous analyses have indicated that this strain lacks genes encoding type II and type III secretion systems (3), which represent two major pathogenicity determinants in most plant- and animal-pathogenic bacteria (7). The genome sequence may resolve how P. ananatis LMG20103 causes disease in the absence of these secretion mechanisms. This also constitutes the first genome of a member of the genus Pantoea and the first bacterial phytopathogen genome sequenced in Africa. The genome was sequenced at Inqaba Biotec, South Africa, using a Roche 454 GS20 sequencer. A total of 991,246 reads, 80 to 120 nucleotides in length, yielding 97,152,414 nucleotides were sequenced to give a 20.9-times coverage. A draft assembly of 117 contigs was produced using Newbler Assembler 2.0.0.0. A scaffolding approach with 12 genomes of closely related Enterobacteriaceae, combined with PCR gap closure, was utilized for final genome assembly. Protein-coding sequences (CDSs) were predicted using a combination of Glimmer 3.0 (5) and AMIGene (2). Annotation was performed using the automated annotation engines BASys (9) and MaGe (8) and validated manually against the NCBI (1) and COG databases (6).

The genome of P. ananatis LMG20103 consists of a single chromosome ~4.69 Mb in size, with a G+C content of 53.69%. There are 4,272 predicted CDSs with an average length of 969 nucleotides, giving a coding intensity of 88.27%. Seventy-two tRNAs and 6 rRNA operons were identified. Putative functions have been assigned to 75% of the encoded proteins. The presence of potential homologues to genes involved in pathogenesis in other plant- and animal-pathogenic bacteria suggests that the genome sequence of P. ananatis LMG20103 will provide considerable insights into the pathogenesis of this organism on Eucalyptus and into diseases caused by its plant- and animal-pathogenic relatives within the genus Pantoea.

Nucleotide sequence accession number. The genome sequence was deposited in GenBank with the accession number CP001875.

This study was partially supported by the National Research Foundation (NRF), the Tree Protection Co-operative Programme (TCP), the NRF/Dept. of Science and Technology Centre of Excellence in Tree Health Biotechnology (CTHB), and the THRIP support program of the Department of Trade and Industry, South Africa.

We acknowledge Genoscope, BaSYS, and Gary van Domselaar for their help with the annotation process and Leigh A. Riley at NCBI for assisting with the genome submission.

REFERENCES


