Genome Sequence of *Pantoea* sp. Strain Sc 1, an Opportunistic Cotton Pathogen

Enrique G. Medrano and Alois A. Bell

U.S. Department of Agriculture—Agricultural Research Service, Cotton Pathology Research Unit, College Station, Texas, USA

***Pantoea* is comprised of a broad spectrum of species, including plant pathogens. Here, we provide an annotated genome sequence of *Pantoea* sp. strain Sc 1, which was isolated from a diseased cotton boll. This research provides the first genome sequence of a bona fide *Pantoea* sp. insect-vectored cotton pathogen.**

Pantoea spp. are members of the family *Enterobacteriaceae* that have traditionally been regarded as saprophytes with a cosmopolitan distribution. Recent taxonomic studies have separated the genus into groups of biocontrol, plant, and clinical strains (10, 11). Previously, we demonstrated that *Pantoea* sp. strain Sc 1 is a causal agent of an emerging disease of immature green cotton bolls in southeastern Cotton Belt states and is vectored by boll-piercing insects such as stink bugs (6, 8, 9).

The draft genome sequence (50-fold coverage) of strain Sc 1 was generated using a Roche 454 GS-Junior DNA analyzer (454 Life Sciences, Branford, CT). A GS Titanium shotgun library (Roche) was prepared, and two runs were pyrosequenced, yielding 76 Mb (187,123 reads with an average length of 410 bases) and 57 Mb (129,858 reads with an average length of 445 bases). Using the same genomic DNA stock, 3- and 8-kb paired-end Titanium libraries were constructed and pyrosequenced, producing 30 Mb (96,198 reads with 70,100 paired reads) and 51 Mb (173,866 reads with 118,148 paired reads), respectively. The genome was constructed using GS De Novo Assembler 454 (Life Sciences, Branford, CT), resulting in four scaffolds totaling 4.4 Mb. Only the largest scaffold (3.8 Mb) contained contig gaps that were closed by cloning and sequencing (BigDye Chemistry) PCR products and employing an ABI PRISM 3100 genetic analyzer (Applied Biosystems, Carlsbad, CA).

Putative coding sequences were predicted by the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP), which uses a combination of GLIMMER (2, 3), GeneMark (1), and tRNAscan-SE (7) and searches several protein databases (Entrez Protein Clusters, Conserved Domain Databases, and Clusters of Orthologous Groups). The PGAAP results were then manually curated. Manual annotation was also conducted using the NCBI BLAST program.

The G-C content was 57%, and extrachromosomal DNA was not detected. The genome contained a total of 4,061 predicted coding sequences (CDS), 5 rRNA operons, and 67 tRNAs. Predicted open reading frames (ORFs) showed considerable homology with components of both type IV and VI secretion systems that are reportedly involved in plant pathogenesis (4, 5) and may be involved in cotton disease. Sequencing the genome of strain Sc 1 is an important step in identifying putative genes involved in cotton boll infection.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AJFP000000000](https://www.ncbi.nlm.nih.gov/nuclseq/AJFP000000000). The version described in this paper is the first version, AJFP000000000.1.

ACKNOWLEDGMENTS

Thanks are due to Cyndy Menn for technical assistance in sequencing, Shannon Ng for editing the manuscript, and Sandra Truong for providing key annotation support.

This work was supported by the USDA CRIS project number 6202-22000-026-00D.

REFERENCES

1. Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* 33:W451–W454.
2. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27:4636–4641.
3. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
4. De Maayer P, et al. 2011. Comparative genomics of the type VI secretion systems of *Pantoea* and *Erwinia* species reveals the presence of putative effector islands that may be translocated by the VgrG and Hcp proteins. *BMC Microbiol.* 12:576.
5. Engledow AS, Medrano EG, Mahenthiralingam E, LiPuma JJ, Gonzalez CF. 2004. Involvement of a plasmid-encoded type IV secretion system in the plant tissue water-soaking phenotype of *Burkholderia cenocepacia*. *J. Bacteriol.* 186:6015–6024.
6. Esquivel JF, Medrano EG, Bell AA. 2010. Southern green stink bugs (Hemiptera: Pentatomidae) as vectors of pathogens affecting cotton bolls—a brief review. *Southwest. Entomol.* 35:457–461.
7. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964.
8. Medrano EG, Bell AA. 2007. Role of *Pantoea agglomerans* in opportunistic bacterial seed and boll rot of cotton grown in the field. *J. Appl. Microbiol.* 102:134–143.
9. Medrano EG, Esquivel JF, Bell AA. 2007. Transmission of cotton seed and boll rotting bacteria by the southern green stink bug (*Nezara viridula* L.). *J. Appl. Microbiol.* 103:436–444.
10. Rezzonico F, Smits TH, Montesinos E, Frey JE, Duffy B. 2009. Genotypic comparison of *Pantoea agglomerans* plant and clinical strains. *BMC Microbiol.* 9:204.
11. Rezzonico F, Vogel G, Duffy B, Tonolla M. 2010. Application of whole-cell matrix-assisted laser desorption/ionization–time of flight mass spectrometry for rapid identification and clustering analysis of *Pantoea* species. *Appl. Environ. Microbiol.* 76:4497–4509.

Received 21 March 2012 Accepted 27 March 2012

Address correspondence to Enrique G. Medrano, gino.medrano@ars.usda.gov.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00450-12