

Genome Sequence of *Pectobacterium carotovorum* subsp. *carotovorum* Strain PCC21, a Pathogen Causing Soft Rot in Chinese Cabbage

Tae-Ho Park,^a Beom-Soon Choi,^b Ah-Young Choi,^b Ik-Young Choi,^b Sunggi Heu,^c and Beom-Seok Park^d

Department of Horticulture, Daegu University, Gyeongsan, Republic of Korea^a; National Instrumentation Center for Environmental Management, Seoul National University, Seoul, Republic of Korea^b; Microbial Safety Division, National Academy of Agricultural Science, Rural Development Administration, Suwon, Republic of Korea^c; and Genomics and Functional Bio-Material Division, National Academy of Agricultural Science, Rural Development Administration, Suwon, Republic of Korea^d

***Pectobacterium carotovorum* is a plant-pathogenic enterobacterium responsible for soft rot in various commercially important plants. Here we report the complete genome sequence and automatic annotation of strain PCC21.**

Pectobacterium carotovorum subsp. *carotovorum* is a Gram-negative phytopathogen causing soft rot and the degradation of pectin, a major component of the primary cell wall, in a wide range of plant species, including various commercially important plants. One of the *P. carotovorum* subsp. *carotovorum* strains isolated in Korea has been used to investigate bacteriocin (7, 8), and the present study uses this strain for complete genome sequencing.

Initially, the genome sequence of *P. carotovorum* subsp. *carotovorum* strain PCC21 was obtained using the Roche 454 GS FLX pyrosequencing platform at the National Instrumentation Center for Environmental Management (NICEM). Here 367,410 good-quality reads with an average read length of 231 bp, amounting to over 85 million nucleotides and representing 17.01-fold genome coverage computed based on the genome size of 5 Mbp, were generated and processed. Short read sequences were assembled using a Newbler assembler, version 2.3 (454 Life Science), and 68 contigs with a total length of 4,802,357 bp were generated. The minimum and maximum contig lengths were 118 bp and 923,611 bp, respectively. These contigs were compared with the genome sequence (BX950851) of *Erwinia carotovora* subsp. *atroseptica* (1), which had many gaps and different orientations between and within sequences. Therefore, secondary sequence information derived from the fosmid library of the strain was included in this study. The fosmid library of 76,800 clones, with an average insert size estimated to be approximately 40 kbp, was constructed as described by Park et al. (5). This represented approximately 15-fold coverage based on a *P. carotovorum* genome size of 5 Mbp. A portion of the fosmid library was end sequenced, and 3,192 good-quality fosmid end sequences with an average read length of 675 bp, amounting to over 2 million nucleotides and representing 0.43-fold genomic coverage computed based on a genome size of 5 Mbp, were generated. Combined data from GS FLX and the fosmid library produced 9 contigs with a total length of 4,816,356 bp. The comparison between the sequences and those of *E. carotovora* subsp. *atroseptica* showed that gaps between these contigs were expected to be smaller than 10 kbp. These gaps were completely filled with full sequences of selected fosmid clones or sequences generated using genome walking based on the PCR method. All tRNA genes were predicted by tRNAscan-SE (3), and rRNA genes were identified by comparing them with other enterobacterial sequences through BLASTN. Protein-coding open reading frames (ORFs) were predicted using Glimmer 3.0 (2).

In summary, the analysis of the unclosed genome sequence of *P. carotovorum* subsp. *carotovorum* showed a genome size of 4,842,771 bp, with a mean GC content of 52.18%, and it is comparable to the published genome sizes of closely related species, which range from 3.8 to 5.1 Mbp (1, 4, 6, 9, 10). The coding regions accounted for 86.1% of the total sequence, and 4,263 annotated coding sequences (CDSs) had an average length of 978 bp. Finally, 22 rRNA operons and 76 tRNA sequences were identified in the chromosome and, among the 4,263 CDSs, 82 (2.3%) had no matches in the current NCBI database.

Nucleotide sequence accession number. The sequence has been deposited in GenBank under accession number CP003776.

ACKNOWLEDGMENT

This work was supported by a grant from the Next-Generation BioGreen 21 Program (no. PJ008944) of the Rural Development Administration in Republic of Korea.

REFERENCES

1. Bell KS, et al. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* 101:11105–11110.
2. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
3. 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.
4. Park DH, et al. 2011. Complete genome sequence of Japanese *Erwinia* strain Ejp617, a bacterial shoot blight pathogen of pear. *J. Bacteriol.* 193: 586–587.
5. Park T-H, et al. 2011. Construction of random sheared fosmid library from Chinese cabbage and its use for *Brassica rapa* genome sequencing project. *J. Genet. Genomics* 38:47–53.
6. Powney R, et al. 2011. Genome sequence of an *Erwinia amylovora* strain with pathogenicity restricted to *Rubus* plants. *J. Bacteriol.* 193:785–786.

Received 28 August 2012 Accepted 10 September 2012

Address correspondence to Tae-Ho Park, thzoo@daegu.ac.kr, or Beom-Seok Park, pbeom@korea.kr.

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

doi:10.1128/JB.01583-12

7. Roh E, et al. 2009. Diverse antibacterial activity of *Pectobacterium carotovorum* subsp. *carotovorum* isolated in Korea. *J. Microbiol. Biotechnol.* 19:42–50.
8. Roh E, et al. 2010. Characterization of a new bacteriocin, carocin D, from *Pectobacterium carotovorum* subsp. *carotovorum* PCC21. *Appl. Environ. Microbiol.* 76:7541–7549.
9. Sebailhia M, et al. 2010. Complete genome sequence of the plant pathogen *Erwinia amylovora* strain ATCC 49946. *J. Bacteriol.* 192:2020–2021.
10. Smits THM, et al. 2010. Complete genome sequence of the fire blight pathogen *Erwinia pyrifoliae* DSM 12163^T and comparative genomic insights into plant pathogenicity. *BMC Genomics* 11:2. doi:10.1186/1471-2164-11-2.