JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of *Pectobacterium* sp. Strain SCC3193

J. Patrik Koskinen,[a] Pia Laine,[b] Outi Niemi,[a] Johanna Nykyri,[c] Heidi Harjunpää,[a] Petri Auvinen,[b] Lars Paulin,[b] Minna Pirhonen,[c] Tapio Palva,[a] and Liisa Holm[a,b]

Division of Genetics Department of Biosciences, University of Helsinki, Helsinki, Finland[a]; Institute of Biotechnology, University of Helsinki, Helsinki, Finland[b]; and Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland[c]

We report the complete and annotated genome sequence of the plant-pathogenic enterobacterium *Pectobacterium* sp. strain SCC3193, a model strain isolated from potato in Finland. The *Pectobacterium* sp. SCC3193 genome consists of a 516,411-bp chromosome, with no plasmids.

Members of the genus *Pectobacterium* cause soft rot and blackleg disease in a wide range of plants, including economically important crop species such as potato (1). *Pectobacterium* sp. strain SCC3193 has been a model strain in the research of soft rot molecular biology for over 2 decades. It was isolated from potato stem on a Finnish field in the 1980s (8). SCC3193 was originally characterized as *Pectobacterium carotovorum* but the species status is, at this point, unclear. Previously sequenced *Pectobacterium* species are *P. atrosepticum* strain SCRI1043, *P. carotovorum* strain WPP14, *P. brasiliensis* strain PBR1692, *P. carotovorum* subsp. *carotovorum* strain PC1, and *P. wasabiae* strain WPP163.

The genome was sequenced using a hybrid approach combining data from three different technologies. First, a fragment library was produced and sequenced with the 454 Genome Sequencer GS20 (6), yielding 712,702 reads, with a mean read length of 101 bp and a total of 71,982,902 bp (ca. 14× coverage). A fosmid library was constructed by shearing genomic DNA cloned in the CopyControl pCC1FOS vector (Epicentre, Madison, WI) mechanically with a needle. End sequences were determined from 865 fosmids using BigDye chemistry and analyzed on ABI 3730 (Applied Biosystems, Foster City, CA), yielding a total of 1,047,795 bp. In addition, we constructed a fragment library for sequencing using SOLiD 2 chemistry (Applied Biosystems, Foster City, CA), with a read length of 35 bp. A total of 52,618,885 reads were obtained, of which 26,309,443 were mapped to the assembly (920,830,488 bp), giving a coverage of 184.

The assembly of 454 reads was performed using Newbler (Roche) and fosmid end sequences were used for the scaffolding of the obtained contigs. The genome was closed using PCR or linker PCR on genomic DNA or fosmids and direct sequencing of the obtained products. Homopolymer errors obtained by 454 sequencing and assembly were corrected by mapping the SOLiD reads to closed genome sequence. The editing and closing of gaps were done in Gap4 from the Staden Package (9).

Coding sequences (CDSs) were predicted using the Prodigal gene prediction program (2). GenePRIMP (7) was run to correct systematic errors made by Prodigal and to reanalyze the remaining intergenic regions for missed CDSs. Functional annotation was performed by using the PANNZER annotation tool (J. P. Koskinen, P. Törönen, J. Nokso-Koivisto, and L. Holm, unpublished data). The tRNA and rRNA genes were annotated using the tRNAscan-SE 1.23 (5) and RNAmmer 1.2 software programs (3). Orthologous groups between the different closely related proteomes were identified using OrthoMCL (4).

The genome of *Pectobacterium* sp. SCC3193 consists of a single circular chromosome that is 5.16 Mbp in size, with an overall G+C content of 50%, without any plasmids. The chromosomal genome contains 4,705 predicted protein-coding sequences, 76 tRNA genes, 7 rRNA operons, and 2 CRISPR loci. Based on the orthologous grouping, 772 (16%) of the SCC3193 CDSs have no detectable homologs in any of the complete *Pectobacterium* or *Dickeya* proteomes published to date.

Nucleotide sequence accession number. The genome sequence of *Pectobacterium* sp. SCC3193 was deposited in GenBank under the accession number CP003415.

ACKNOWLEDGMENTS

We acknowledge the support of the Academy of Finland (Center of Excellence program 2006-2011, grants 213509 and 129628 and grants 136470, 120821, and 128566), Biocentrum Helsinki, Biocenter Finland, the Finnish Doctoral Program in Computational Sciences FICS, the Viikki Doctoral Program in Molecular Biosciences, and the Finnish Doctoral Program in Plant Science.

REFERENCES

1. **Czajkowski R, Pérombelon MCM, van Veen JA, van der Wolf JM.** 2011. Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. Plant Pathol. **60**:999–1013.
2. **Hyatt D, et al.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119.
3. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
4. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.
5. **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.
6. **Margulies M, et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380.
7. **Pati A, et al.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat. Methods **7**:455–457.
8. **Pirhonen M, Heino P, Helander I, Harju P, Palva ET.** 1988. Bacteriophage T4 resistant mutants of the plant pathogen *Erwinia carotovora*. Microb. Pathog. **4**:359–367.
9. **Staden R, Beal KF, Bonfield JK.** 2000. The Staden package, 1998. Methods Mol. Biol. **132**:115–130.

Received 25 April 2012 **Accepted** 24 August 2012

Address correspondence to Liisa Holm, liisa.holm@helsinki.fi.

doi:10.1128/JB.00681-12