genomeA
Journals.ASM.org

Genome Sequence of *Pectobacterium atrosepticum* Strain 21A

Yevgeny Nikolaichik,[a] Vladimir Gorshkov,[b] Yuri Gogolev,[b] Leonid Valentovich,[a,c] Anatoli Evtushenkov[a]

Department of Molecular Biology, Faculty of Biology, Belarus State University, Minsk, Belarus[a]; Kazan Institute of Biochemistry and Biophysics, Kazan Scientific Center, Russian Academy of Sciences, Kazan, Russia[b]; Institute of Microbiology, National Academy of Sciences, Minsk, Belarus[c]

We report the annotated genome sequence of the enterobacterial plant pathogen *Pectobacterium atrosepticum* strain 21A, isolated in Belarus from potato stem with blackleg symptoms.

Received 17 August 2014 **Accepted** 20 August 2014 **Published** 18 September 2014

Citation Nikolaichik Y, Gorshkov V, Gogolev Y, Valentovich L, Evtushenkov A. 2014. Genome sequence of *Pectobacterium atrosepticum* strain 21A. Genome Announc. 2(5): e00935-14. doi:10.1128/genomeA.00935-14.

Address correspondence to Yevgeny Nikolaichik, nikolaichik@bio.bsu.by.

Pectobacterium spp. are capable of causing disease in a wide spectrum of plant species. However, *Pectobacterium atrosepticum* is characterized by a rather narrow host range and is mostly associated with blackleg and soft rot diseases in potato. The *P. atrosepticum* strain 21A was isolated in Belarus in 1978 (1) from potato stem with blackleg symptoms. The strain is virulent in potato and differs from the well-characterized *P. atrosepticum* strain SCRI 1043 (2) in its ability to cause hypersensitive reactions in nonhost plants.

The data for the genome assembly were generated using Illumina MiSeq and the Nextera XT library preparation protocol. After quality filtering with Prinseq (http://prinseq.sourceforge.net), 12,238,721 reads were retained, of which 98% mapped to the final assembly, giving a coverage of 268. Filtered reads were assembled into 35 large (>1,000 bp) contigs using SPAdes (3) with BayesHammer (4) error correction. SSPACE (5) and GapFiller (6) were used for initial gap closure, followed by manual resolution of repeats, with the genome sequence of *P. atrosepticum* strain SCRI 1043 used to assist in scaffolding.

The complete genome of *P. atrosepticum* strain 21A consists of a 4,991,806-bp chromosome with a GC content of 51.1% and a 32,444-bp plasmid with a GC content of 47%. Based on the difference in coverage of the two replicons, the plasmid is present in 3 to 4 copies per cell.

Genome annotation was performed using the Prokka annotation pipeline (7). Coding sequences were predicted using Prodigal (8), signal peptides by SignalP (9). tRNA genes and transfer-messenger RNA (tmRNA) were predicted by ARAGORN (10), rRNA genes by Barrnap (http://www.vicbioinformatics.com/software.barrnap.shtml), and noncoding RNAs- by Infernal (11). Clustered regularly interspaced short palindromic repeats (CRISPRs) were detected by MinCED (https://github.com/ctSkennerton/minced). The genome contains 4,424 protein coding sequences and 22 rRNA genes organized into 7 operons, 77 tRNAs, and 2 CRISPR loci. The genome codes for a set of extracellular hydrolases typical for pectolytic bacteria, including 9 pectate lyases, 4 polygalacturonases, 1 cellulase, 2 hemicellulases, and an extracellular protease. All six known types of protein secretion systems are present.

Organization of the *P. atrosepticum* 21A chromosome is very similar to that of three other known *P. atrosepticum* genomes. Overall, gene content and order are the same in the four strains, with the exception of horizontally transferred sequences (mostly phage related), which account for <100 genes unique for *P. atrosepticum* 21A. Another notable difference between *P. atrosepticum* chromosomes is a large (1.35- Mb) inversion in *P. atrosepticum* strains 21A and CFBP6276 (12) relative to SCRI 1043 and the recently sequenced genome of strain JG10-08 (GenBank accession no. CP007744).

The plasmid in *P. atrosepticum* 21A has weak similarity to the plasmid-like sequence integrated into the SCRI 1043 chromosome. The similarity, however, is restricted to the type IV secretion genes that might be responsible for conjugative transfer and the antirestriction gene. Compared to the plasmid-like element in SCRI 1043, the *P. atrosepticum* 21A plasmid lacks arsenical resistance genes but has genes that might be related to pathogenicity, including genes coding for a phospholipase and an H-NS-like protein.

Nucleotide sequence accession numbers. The nucleotide sequence accession numbers are CP009125 and CP009126 for the chromosome and the plasmid, respectively.

ACKNOWLEDGMENTS

We are grateful to Anton Korobeynikov for helpful advice on BayesHammer and SPAdes usage.

This work was supported by the Ministry of Education of Belarus (task 2.52 within the State Program "Fundamentals of Biotechnology"), the EurAsEC program "Innovative Biotechnologies" (task 1.9), and by the Russian Foundation for Basic Research (research project numbers 14-04-01750_A and 14-04-01828_A).

REFERENCES

1. **Evtushenkov AN.** 1981. Ph.D. Thesis. Belarus State University, Minsk, Belarus.
2. **Bell KS, Sebaihia M, Pritchard L, Holden MT, Hyman LJ, Holeva MC, Thomson NR, Bentley SD, Churcher LJ, Mungall K, Atkin R, Bason N, Brooks K, Chillingworth T, Clark K, Doggett J, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Norbertczak H, Ormond D, Price C, Quail MA, Sanders M, Walker D, Whitehead S, Salmond GP, Birch PR, Parkhill J, Toth IK.** 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. http://dx.doi.org/10.1073/pnas.0402424101.

3. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
4. **Nikolenko SI, Korobeynikov AI, Alekseyev MA.** 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics **14:**S7. http://dx.doi.org/10.1186/1471-2164-14-S1-S7.
5. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
6. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol. **13:**R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.
7. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.
8. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119. http://dx.doi.org/10.1186/1471-2105-11-119.
9. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8:**785–786. http://dx.doi.org/10.1038/nmeth.1701.
10. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32:**11–16. http://dx.doi.org/10.1093/nar/gkh152.
11. **Kolbe DL, Eddy SR.** 2011. Fast filtering for RNA homology search. Bioinformatics **27:**3102–3109. http://dx.doi.org/10.1093/bioinformatics/btr545.
12. **Kwasiborski A, Mondy S, Beury-Cirou A, Faure D.** 2013. Genome sequence of the *Pectobacterium atrosepticum* strain CFBP6276, causing blackleg and soft rot diseases on potato plants and tubers. Genome Announc. **1**(3):e00374-13. http://dx.doi.org/10.1128/genomeA.00374-13.