AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Pantoea agglomerans* ASB05 Using Illumina and PacBio Sequencing

Sang In Lee,[a] Thao D. Tran,[b] Steven Huynh,[c] Craig T. Parker,[c] Robert Hnasko,[c] Jeffery A. McGarvey[a]

[a]Foodborne Toxin Detection and Prevention Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Albany, California, USA
[b]Department of Wildlife, Fish, and Conservation Biology, University of California, Davis, California, USA
[c]Produce Safety and Microbiology Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Albany, California, USA

ABSTRACT We present the complete genome sequence of *Pantoea agglomerans* ASB05 and three associated plasmids, generated using a combination of the Illumina and PacBio platforms. *P. agglomerans* ASB05 was isolated from fresh cherries purchased in Albany, CA, in 2016.

Pantoea agglomerans, formerly known as *Enterobacter agglomerans*, is a Gram-negative, rod-shaped bacterium that belongs to the family *Enterobacteriaceae* (1, 2). *P. agglomerans* is ubiquitously found in environmental samples such as water, soil, dust, plant surfaces, and confined animal feeding operations (3, 4). Its ability to cause disease in healthy humans is uncertain (1, 5); however, it has been isolated from immunocompromised individuals along with other bacteria such as *Mycobacterium* spp. and *Pseudomonas* spp. (6). In this study, we present the genome sequence of *P. agglomerans* strain ASB05, which was isolated from cherries purchased from a grocery store in Albany, CA, in 2016.

P. agglomerans strain ASB05 was isolated as described by McGarvey et al. with modifications (7). Briefly, store-bought whole cherries were washed in phosphate-buffered saline with 0.01% Tween 80 for 1 h at 25°C with shaking at 200 rpm. The liquid was decanted and plated onto Reasoner's 2A (R2A) agar (Remel, KS, USA) that was incubated for 24 h at 37°C. Single colonies were struck on tryptic soy agar (Oxoid, Basingstoke, Hampshire, England), incubated for 24 h at 37°C, and cryopreserved for further use.

Prior to DNA extraction, a single colony of *P. agglomerans* strain ASB05 was inoculated into 100 ml of tryptic soy broth (TSB) (Oxoid, Basingstoke, Hampshire, England) and incubated aerobically for 24 h at 37°C with shaking at 200 rpm. Genomic DNA was extracted from harvested cells by sucrose-Tris with phenol-chloroform cleanup extractions as described by Miller et al. (8).

P. agglomerans strain ASB05 was primarily sequenced via the Pacific Biosciences (PacBio, Menlo Park, CA) RS II platform and produced sequences that were compared and confirmed with sequences generated with the Illumina (San Diego, CA) MiSeq platform. Both sequencing methods were performed by following the standard library construction protocol described previously by Parker et al. (9). For the PacBio platform, the SMRTbell library was prepared from 10 μg of bacterial genomic DNA fragmented using G-tube (Covaris, Woburn, MA) following the standard PacBio 20-kb library preparation procedure (10) but with 1$\times$ AMPure bead (PacBio) cleanup and an extra DNA repair step after BluePippin size selection with a 0.75% DF Marker S1 high-pass 6- to 10-kb vs3 cassette (Sage Science, Beverly, MA). The library was run in one single-molecule real-time (SMRT) cell with the 0.1 nM on-plate concentration, P6/C4 sequencing chemistry, MagBead One Cell Per Well v1 collection protocol, and 360-min data collection mode. For the Illumina platform, the library was prepared from 1.5 μg of bacterial genomic DNA fragmented by microtube (Covaris) to 700- to 770-bp fragments at 30 lb/in^2 for 40 s following the LTP library preparation kit manufacturer's protocol (KAPA Biosystems, Wilmington, MA) (9). Sequencing was performed using a 2 $\times$ 250-cycle paired-end v2 reagent kit on a MiSeq instrument (Illumina). Among the 53,017 total reads produced by

Citation Lee SI, Tran TD, Huynh S, Parker CT, Hnasko R, McGarvey JA. 2021. Complete genome sequence of *Pantoea agglomerans* ASB05 using Illumina and PacBio sequencing. Microbiol Resour Announc 10:e00501-21. https://doi.org/10.1128/MRA.00501-21.

Editor Irene L. G. Newton, Indiana University, Bloomington

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Jeffery A. McGarvey, jeffery.mcgarvey@usda.gov.

Received 12 May 2021
Accepted 6 July 2021
Published 29 July 2021

TABLE 1 antiSMASH-predicted secondary metabolite gene clusters

Predicted compound	Genes (%)[a]	Function	Location[c]
Chromosome			
Stewartan	92	EPS[b]	1459712–1493750
Aryl polyene	94	Antioxidant/UV protection	2585609–2599670
Amonabactin P 750	57	Siderophore	3619385–3629104
pASB05p1			
Terpene carotenoid	100	UV protection	154479–160707
Desferrioxamine E	100	Siderophore	354749–361266
pASB05p2			
Phenazine iodinin	45	Antimicrobial	169167–173831
pASB05p3			
None			

[a] Percentage of genes present.
[b] EPS, extracellular polysaccharide.
[c] Nucleotide position.

the PacBio RS II platform, 50,314 reads were mapped, or assembled, whereas the Illumina platform produced 1,879,022 total reads and used 1,857,038 reads to assemble.

The reads (N_{50} read length, 24,284 bp) generated from the PacBio platform were initially processed and assembled via the Hierarchical Genome Assembly Process (HGAP) v3.0 in Single-Molecule Real-Time (SMRT) Analysis v2.2.0 (Pacific Biosciences, Menlo Park, CA). PacBio DNA internal control complex P6 was used as an internal sequencing control, and the read quality control was conducted using FastQC (Pacific Biosciences). Illumina MiSeq reads (read length, 251 bp) were trimmed using a quality score threshold of 30 or higher (Q30) and assembled to the PacBio contigs within Geneious Prime software v2019.2.3 (Biomatters, Ltd., Auckland, New Zealand). The final validation process was carried out by comparing MiSeq reads and the PacBio assembly using the find variation/single nucleotide polymorphisms (SNPs) module in Geneious Prime v2020.0.4 (Biomatters, Ltd.), with a minimum coverage parameter of 50 and minimum variant frequency parameter of 0.8. Finally, genes were both manually and automatically annotated based on *P. agglomerans* strain L15 (GenBank accession no. CP034148) and via the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (11), respectively.

According to the assembly and validation processes, the PacBio platform produced a total of 4 contigs for strain ASB05, consisting of a single circular chromosome with a length of 4,022,781 bp, and 3 plasmids with lengths of 563,807 bp, 207,454 bp, and 64,606 bp. In addition, 3,615 coding sequences (CDS), 22 rRNA operons, and 77 tRNA genes were identified and annotated from the circular chromosome of ASB05. The sequences were also scanned for bacteriophage via PHASTER (http://www.phaster.ca) (12); one intact prophage and three incomplete prophages were identified from the chromosome, and one incomplete prophage was identified in plasmid pASB05p3. Finally, 157 insertion sequences (IS) were identified using ISfinder (https://www-is.biotoul.fr/) (13).

The ASB05 genome and plasmids were also examined for predicted secondary metabolite biosynthetic gene clusters using the antiSMASH v5.1.0 software (https://antismash.secondarymetabolites.org/) (14) (Table 1). The chromosome contained gene clusters associated with the production of extracellular polysaccharide stewartan (15), an aryl polyene (16), and the siderophore amonabactin P 750 (17). The plasmid pASB05p1 contained gene clusters associated with the production of a terpene carotenoid (18) and the siderophore desferrioxamine E (19). The plasmid pASB05p2 contained a gene cluster associated with the production of phenazine iodinin (20), and the plasmid pASB05p3 did not contain any gene clusters associated with secondary metabolites.

About 74 either draft or complete *P. agglomerans* genome sequences are published on PubMed (https://pubmed.ncbi.nlm.nih.gov), and 28 have been described in *Microbiology Resource Announcements* (https://journals.asm.org/journal/mra). Most of them have a genome size of ~4 to 6 Mb and contain 3 plasmids, which the *P. agglomerans* strain ASB05 in this study is consistent with.

Data availability. The whole-genome sequence is available at DDBJ/ENA/GenBank under the accession numbers CP046722 (ASB05 chromosome), CP046723 (pASB05p1), CP046724 (pASB05p2), and CP046725 (pASB05p3), BioProject PRJNA594723, and BioSample SAMN13527266. The PacBio and Illumina raw data are accessible from the SRA under the accession numbers SRR10665756 (PacBio) and SRR11187857 (Illumina).

ACKNOWLEDGMENT

This work was supported by the United States Department of Agriculture, Agricultural Research Service, National Program 108: Food Safety (animal and plant products).

REFERENCES

1. Büyükcam A, Tuncer Ö, Gür D, Sancak B, Ceyhan M, Cengiz AB, Kara A. 2018. Clinical and microbiological characteristics of *Pantoea agglomerans* infection in children. J Infect Public Health 11:304–309. https://doi.org/10.1016/j.jiph.2017.07.020.
2. Cruz AT, Cazacu AC, Allen CH. 2007. *Pantoea agglomerans*, a plant pathogen causing human disease. J Clin Microbiol 45:1989–1992. https://doi.org/10.1128/JCM.00632-07.
3. Andersson A, Weiss N, Rainey F, Salkinoja-Salonen M. 1999. Dust-borne bacteria in animal sheds, schools and children's day care centres. J Appl Microbiol 86:622–634. https://doi.org/10.1046/j.1365-2672.1999.00706.x.
4. Walterson AM, Stavrinides J. 2015. *Pantoea*: insights into a highly versatile and diverse genus within the *Enterobacteriaceae*. FEMS Microbiol Rev 39:968–984. https://doi.org/10.1093/femsre/fuv027.
5. Cheng A, Liu CY, Tsai HY, Hsu MS, Yang CJ, Huang YT, Liao CH, Hsueh PR. 2013. Bacteremia caused by *Pantoea agglomerans* at a medical center in Taiwan, 2000–2010. J Microbiol Immunol Infect 46:187–194. https://doi.org/10.1016/j.jmii.2012.05.005.
6. Flores Popoca EO, Miranda García M, Romero Figueroa S, Mendoza Medellín A, Sandoval Trujillo H, Silva Rojas HV, Ramírez Durán N. 2012. *Pantoea agglomerans* in immunodeficient patients with different respiratory symptoms. ScientificWorldJournal 2012:156827. https://doi.org/10.1100/2012/156827.
7. McGarvey JA, Tran TD, Hnasko R, Gorski L. 2019. Use of phyllosphere-associated lactic acid bacteria as biocontrol agents to reduce *Salmonella enterica* serovar Poona growth on cantaloupe melons. J Food Prot 82:2148–2153. https://doi.org/10.4315/0362-028XJFP-19-246.
8. Miller WG, On SL, Wang G, Fontanoz S, Lastovica AJ, Mandrell RE. 2005. Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*. J Clin Microbiol 43:2315–2329. https://doi.org/10.1128/JCM.43.5.2315-2329.2005.
9. Parker CT, Cooper KK, Huynh S, Smith TP, Bono JL, Cooley M. 2018. Genome sequences of eight Shiga toxin-producing *Escherichia coli* strains isolated from a produce-growing region in California. Microbiol Resour Announc 7:e00807-18. https://doi.org/10.1128/MRA.00807-18.
10. PacBio. 2015. Procedure and checklist: 20 kb template preparation using BluePippin size-selection system. https://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-20-kb-Template-Preparation-Using-BluePippin-Size-Selection.pdf.
11. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
12. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
13. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.
14. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.
15. Langlotz D, Schollmeyer M, Coplin DL, Nimtz M, Geider K. 2011. Biosynthesis of the repeating units of the exopolysaccharides amylovoran from *Erwinia amylovora* and stewartan from *Pantoea stewartii*. Physiol Mol Plant Pathol 75:163–169. https://doi.org/10.1016/j.pmpp.2011.04.001.
16. Grammbitter GLC, Schmalhofer M, Karimi K, Shi Y-M, Schöner TA, Tobias NJ, Morgner N, Groll M, Bode HB. 2019. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. J Am Chem Soc 141:16615–16623. https://doi.org/10.1021/jacs.8b10776.
17. Esmaeel Q, Chevalier M, Chataigné G, Subashkumar R, Jacques P, Leclère V. 2016. Nonribosomal peptide synthetase with a unique iterative-alternative-optional mechanism catalyzes amonabactin synthesis in *Aeromonas*. Appl Microbiol Biotechnol 100:8453–8463. https://doi.org/10.1007/s00253-016-7773-4.
18. Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K. 1990. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. J Bacteriol 172:6704–6712. https://doi.org/10.1128/jb.172.12.6704-6712.1990.
19. Barona-Gomez F, Wong U, Giannakopulos AE, Derrick PJ, Challis GL. 2004. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J Am Chem Soc 126:16282–16283. https://doi.org/10.1021/ja045774k.
20. Shi YM, Brachmann AO, Westphalen MA, Neubacher N, Tobias NJ, Bode HB. 2019. Dual phenazine gene clusters enable diversification during biosynthesis. Nat Chem Biol 15:331–339. https://doi.org/10.1038/s41589-019-0246-1.