Draft Genome Sequence of *Pantoea anthophila* Strain 11-2 from Hypersaline Lake Laysan, Hawaii

Xuehua Wan,[a] Shaobin Hou,[a] Nolwenn Phan,[b] Jennifer S. Malone Moss,[b] Stuart P. Donachie,[b] Maqsudul Alam[a,b]†

Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA[a]; Department of Microbiology, University of Hawaii, Honolulu, Hawaii, USA[b]

† Deceased 20 December 2014.

This paper is dedicated to our mentor and colleague, Maqsudul Alam, in memory of his contributions to genome science.

Most *Pantoea* spp. have been isolated from plant sources or clinical samples. However, we cultivated *Pantoea anthophila* 11-2 from hypersaline water from the lake on Laysan, Northwestern Hawaiian Islands. Draft genome sequencing of 11-2 provides a molecular basis for studies in evolution and pathogenicity in *Pantoea* spp.

Address correspondence to Xuehua Wan, xuehua@hawaii.edu.

*P*antoea spp. are Gram-negative bacteria generally isolated from agricultural or clinical samples, and marine sediment (1–3). Strain 11-2 from hypersaline Lake Laysan shared 100% nucleotide identity in its 16S rRNA gene with the 16S rRNA gene in *Pantoea anthophila* BD 871^T (LMG 2558) from *Impatiens balsamina* (4–6). We report the draft genome sequence of *P. anthophila* 11-2.

Shotgun and 8-kb-span paired-end libraries prepared and sequenced in the Roche 454 GS FLX+ platform generated 126.7 Mb of shotgun sequences and 116.3 Mb of 8-kb-span paired-end sequences, providing ~50× genome coverage. Assembly was performed in Newbler 2.8 in a two-step strategy. Shotgun reads were assembled into contigs, after which paired-end reads were added to build 5 scaffolds containing 4,609,867 bp (N_{50} = 3,885,396 bp) and 16 contigs containing 4,600,679 bp assembled in the 5 scaffolds. Gap regions in scaffolds were estimated to cover ~9 kb. The genome's G+C content is 56.8%.

The genome was annotated in the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (7) and the Rapid Annotation Using Subsystem Technology (RAST) server (8, 9). PGAP identified 4,069 protein-coding open reading frames, 67 tRNA coding regions, and 15 rRNA coding regions. RAST identified 502 function-related subsystems and 7 phage components, including 3 phage baseplate proteins, 2 phage packaging machineries, and 1 each of phage tail protein and phage tail fiber protein. Six virulence regulation components were also identified in the BarA-UvrY two-component regulatory system. Ten proteins in the type III secretion system (T3SS) apparatus, and 64 T3SS-related transcription regulators were predicted using BLASTp against the T3SS database (T3DB) (E < 1e^{-10}) (10). Eleven diguanylate cyclases and nine diguanylate phosphodiesterases (E < 1e^{-10}) were predicted, which synthesize and degrade the ubiquitous second messenger, cyclic-di-GMP, involved in bacterial morphology and biofilm formation. The N-terminal PAS domain and GAF domain sensors of several diguanylate cyclases and diguanylate phosphodiesterases were predicted by the SMART server (11). The draft genome sequence reported here provides a molecular basis for studies in pathogenicity and evolution of *Pantoea* spp.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JXXL00000000. The version described in this paper is version JXXL01000000.

REFERENCES

1. **Mergaert J, Verdonck L, Kersters K.** 1993. Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend. as *Pantoea ananas* (Serrano 1928) comb. nov. and *Pantoea stewartii* (Smith 1898) comb. nov., respectively, and description of *Pantoea stewartii* subsp. *indologenes* subsp. nov. Int J Syst Bacteriol 43:162–173. http://dx.doi.org/10.1099/00207713-43-1-162.
2. **Brady CL, Cleenwerck I, Venter SN, Engelbeen K, De Vos P, Coutinho TA.** 2010. Emended description of the genus *Pantoea*, description of four species from human clinical samples, *Pantoea septica* sp. nov., *Pantoea eucrina* sp. nov., *Pantoea brenneri* sp. nov. and *Pantoea conspicua* sp. nov., and transfer of *Pectobacterium cypripedii* (Hori 1911) Brenner et al. 1973 emend. Hauben et al. 1998 to the genus as *Pantoea cypripedii* comb. nov. Int J Syst Evol Microbiol 60:2430–2440. http://dx.doi.org/10.1099/ijs.0.017301-0.
3. **Silvi S, Barghini P, Aquilanti A, Juarez-Jimenez B, Fenice M.** 2013. Physiologic and metabolic characterization of a new marine isolate (BM39) of *Pantoea* sp. producing high levels of exopolysaccharide. Microb Cell Fact 12:10. http://dx.doi.org/10.1186/1475-2859-12-10.
4. **Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, Park SC, Jeon YS, Lee JH, Yi H, Won S, Chun J.** 2012. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol 62:716–721. http://dx.doi.org/10.1099/ijs.0.038075-0.
5. **Brady CL, Venter SN, Cleenwerck I, Engelbeen K, Vancanneyt M, Swings J, Coutinho TA.** 2009. *Pantoea vagans* sp. nov., *Pantoea eucalypti*

sp. nov., *Pantoea deleyi* sp. nov. and *Pantoea anthophila* sp. nov. Int J Syst Evol Microbiol **59:**2339–2345. http://dx.doi.org/10.1099/ijs.0.009241-0.

6. **Donachie SP, Hou S, Lee KS, Riley CW, Pikina A, Belisle C, Kempe S, Gregory TS, Bossuyt A, Boerema J, Liu J, Freitas TA, Malahoff A, Alam M.** 2004. The Hawaiian Archipelago: a microbial diversity hotspot. Microb Ecol **48:**509–520. http://dx.doi.org/10.1007/s00248-004-0217-1.

7. **Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O.** 2008. Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. Omics **12:**137–141. http://dx.doi.org/10.1089/omi.2008.0017.

8. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V,** **Wilke A, Zagnitko O.** 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.

9. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R.** 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res **42:**D206–D214. http://dx.doi.org/10.1093/nar/gkt1226.

10. **Wang Y, Huang H, Sun M, Zhang Q, Guo D.** 2012. T3DB: an integrated database for bacterial type III secretion system. BMC Bioinformatics **13:**66. http://dx.doi.org/10.1186/1471-2105-13-66.

11. **Schultz J, Milpetz F, Bork P, Ponting CP.** 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA **95:**5857–5864. http://dx.doi.org/10.1073/pnas.95.11.5857.