Draft Genome Sequence of a *Pantoea* sp. Isolated from a Preterm Neonatal Blood Sepsis Patient

K. A. Kropp,[a] A. Lucid,[a] J. Carroll,[b] V. Belgrudov,[c] P. Walsh,[b] B. Kelly,[b] K. Templeton,[d,e] C. Smith,[d] P. Dickinson,[d] A. O'Driscoll,[c] P. Ghazal,[d] R. D. Sleator[a]

Department of Biological Sciences, Cork Institute of Technology, Bishopstown, Cork, Ireland[a]; NSilico, Cork, Ireland[b]; Department of Computing, Cork Institute of Technology, Bishopstown, Cork, Ireland[c]; Division of Pathway Medicine, University of Edinburgh, Edinburgh, United Kingdom[d]; Microbiological Diagnostic Unit, Royal Infirmary, University of Edinburgh, Edinburgh, United Kingdom[e]

Herein, we report the draft genome sequence of *Pantoea* sp. ED-NGS-1003, cultivated from a blood sample taken from a neonatal sepsis patient at the Royal Infirmary, Edinburgh, Scotland, United Kingdom.

*P*antoea spp. are Gram-negative, rare, opportunistic pathogens that can infect immune-compromised patients. They cause urinary infections and blood sepsis (1–3) and specifically *Pantoea agglomerans* has been linked to several outbreaks in neonatal units (1, 4, 5). Preterm neonates are a highly susceptible patient group for bacterial infections (6–8) and rapid detection of blood sepsis and identification of the causative agent are critical to enable proper treatment (9–11). The ClouDx-i project aims to extend current knowledge on circulating pathogenic strains linked with neonatal blood sepsis to inform the development of new and improved molecular diagnostic assays. Herein, we present the draft genome of a *Pantoea* sp. strain, isolated from a preterm neonate at the Royal Infirmary, Edinburgh in 2013. Positivity for blood sepsis and species identification were confirmed by classical microbiological identification and characterization techniques.

The isolate was grown overnight at 37°C on Luria broth (LB) agar, and genomic DNA was isolated using Qiagen genomic tips (Venlo, Limburg, Netherlands). Genomic DNA fragments were produced ranging in size from 2 to 10 kb using sonication and subsequently used to produce a non-size-selected genome library using the Nextera mate pair kit (Illumina, San Diego, CA). This library was sequenced on an Illumina MiSeq using MiSeq Reagent kit v3. Genomic sequence assembly, analysis and automated reporting was carried out using Simplicity (12). This approach produced 2,596,947 total reads, resulting in an average 114-fold coverage. The average G+C content was 58.80%. For sequence assembly, we used a *de novo* assembly pipeline based on the Spades 3.10 assembly tool with k-mers K21, K33, K55, K77, K99, and K127 nucleotides, resulting in a total of 133 contigs, of which 49 were >1,000 bp representing 99.02% of sequence information. Post assembly processing was performed by Spades and only scaffolds of length greater than 1,000 bp were considered when estimating genome length as 4,822,832 bp. We annotated the genome with Prokka (13) and used the identified 16S rRNA gene to confirm the species as *Pantoea* sp. A scaffold of the genome was produced with Contiguator2 by mapping the contigs back to several *Pantoea* reference genomes. However, BLASTing the scaffold against the NCBI database could not identify a closely related strain. The genome was then screened using Glimmer3 (14) identifying 4,670 open reading frames (ORFs). The predicted ORFs were compared to the Uniprot-Trembl database using BLASTp for strain identification, mapping 3,495 ORFs to the database. To identify potential virulence factors in the genome we compared a local database built from the VFDB (15) and Victors databases with the BLASTp tool, using a 75% amino-acid sequence identity cut-off while only considering alignments longer than 100 amino-acids, identifying 93 hits.

Samples were handled in accordance with local ethical approval by the ethics committees of the NHS Lothian SAHSC Bioresource and NHS R&D office, Project ID 2011/R/NE/01 and the HSS BioResource Request ID 13/ES/0126.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession no. JPQA00000000. The version described in this paper is version JPQA01000000.

REFERENCES

1. **Aly NYA, Salmeen HN, Lila RAA, Nagaraja PA.** 2008. *Pantoea agglomerans* bloodstream infection in preterm neonates. Med. Princ. Pract. 17:500–503. http://dx.doi.org/10.1159/000151575.
2. **Mehar V, Yadav D, Sanghvi J, Gupta N, Singh K.** 2013. *Pantoea dispersa*: an unusual cause of neonatal sepsis. Braz. J. Infect. Dis. 17:726–728. http://dx.doi.org/10.1016/j.bjid.2013.05.013.
3. **Van Rostenberghe H, Noraida R, Wan Pauzi WI, Habsah H, Zeehaida M, Rosliza AR, Fatimah I, Nik Sharimah NY, Maimunah H.** 2006. The clinical picture of neonatal infection with *Pantoea* species. Jpn. J. Infect. Dis. 59:120–121.
4. **Cheng A, Liu C-Y, Tsai H-Y, Hsu M-S, Yang C-J, Huang Y-T, Liao C-H, Hsueh P-R.** 2013. Bacteremia caused by *Pantoea agglomerans* at a medical center in Taiwan, 2000–2010. J. Microbiol. Immunol. Infect. 46:187–194. http://dx.doi.org/10.1016/j.jmii.2012.05.005.
5. **Mardaneh J, Dallal MMS.** 2013. Isolation, identification and antimicrobial susceptibility of *Pantoea* (*Enterobacter*) *agglomerans* isolated from consumed powdered infant formula milk (PIF) in NICU ward: first report from Iran. Iran. J. Microbiol. 5:263–267.

6. **Kaufman D, Fairchild KD.** 2004. Clinical microbiology of bacterial and fungal sepsis in very-low-birth-weight infants. Clin. Microbiol. Rev. **17:** 638–680. http://dx.doi.org/10.1128/CMR.17.3.638-680.2004.

7. **Shah BA, Padbury JF.** 2014. Neonatal sepsis: an old problem with new insights. Virulence **5:**170–178. http://dx.doi.org/10.4161/viru.26906.

8. **Ghazal P, Dickinson P, Smith CL.** 2013. Early life response to infection. Curr. Opin. Infect. Dis. **26:**213–218. http://dx.doi.org/10.1097/QCO.0b013e32835fb8bf.

9. **Labib AZ, Mahmoud AB, Eissa N, El Gendy FM, Soliman MA, Aly AA.** 2013. Early diagnosis of neonatal sepsis: a molecular approach and detection of diagnostic markers versus conventional blood culture. Int. J. Microbiol. Res. **4:**77–85.

10. **Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M.** 2010. The era of molecular and other non-culture-based methods in diagnosis of sepsis. Clin. Microbiol. Rev. **23:**235–251. http://dx.doi.org/10.1128/CMR.00043-09.

11. **Sibley CD, Peirano G, Church DL.** 2012. Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. Infect. Genet. Evol. **12:**505–521. http://dx.doi.org/10.1016/j.meegid.2012.01.011.

12. **Walsh P, Carroll J, Sleator RD.** 2013. Accelerating *in silico* research with workflows: a lesson in simplicity. Comput. Biol. Med. **43:**2028–2035. http://dx.doi.org/10.1016/j.compbiomed.2013.09.011.

13. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

14. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:** 4636–4641. http://dx.doi.org/10.1093/nar/27.23.4636.

15. **Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q.** 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. **33:** D325–D328. http://dx.doi.org/10.1093/nar/gki177.