

Genome Sequence of *Burkholderia pseudomallei* NCTC 13392

Jason W. Sahl,^{a,b} Joshua K. Stone,^a H. Carl Gelhaus,^c Richard L. Warren,^d Caroline J. Cruttwell,^e Simon G. Funnell,^e Paul Keim,^{a,b} Apichai Tuanyok^a

Northern Arizona University, Center for Microbial Genetics and Genomics, Flagstaff, Arizona, USA^a; Translational Genomics Research Institute, Flagstaff, Arizona, USA^b; Battelle Biomedical Research Center (BBRC), Columbus, Ohio, USA^c; Tunnell Consulting, Bethesda, Maryland, USA^d; Health Protection Agency, Porton Down, Salisbury, United Kingdom^e

Here, we describe the draft genome sequence of *Burkholderia pseudomallei* NCTC 13392. This isolate has been distributed as K96243, but distinct genomic differences have been identified. The genomic sequence of this isolate will provide the genomic context for previously conducted functional studies.

Received 12 March 2013 Accepted 5 April 2013 Published 23 May 2013

Citation Sahl JW, Stone JK, Gelhaus HC, Warren RL, Cruttwell CJ, Funnell SG, Keim P, Tuanyok A. 2013. Genome sequence of *Burkholderia pseudomallei* NCTC 13392. *Genome Announc.* 1(3):e00183-13. doi:10.1128/genomeA.00183-13.

Copyright © 2013 Sahl et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Apichai Tuanyok, Apichai.Tuanyok@nau.edu.

Burkholderia pseudomallei is the causative agent of the potentially fatal disease melioidosis (1). Melioidosis is endemic to Southeast Asia and northern Australia, where *B. pseudomallei* is found in wet soil (2). *B. pseudomallei* is a phylogenetically diverse pathogen, as revealed by whole-genome sequence analysis (3).

A recent study demonstrated that *B. pseudomallei* NCTC 13392 (cited at the time of the study by the United Kingdom National Collection of Type Cultures as strain K96243), provided by the Health Protection Agency (HPA) in the United Kingdom, was shown to cause lethal inhalational melioidosis in a marmoset model (4). We sequenced the isolate used in that study and compared its genome to that of the published version of K96243 (5). Based on a comparison of single nucleotide polymorphisms (SNPs) and phylogenetic position, we determined that NCTC 13392 is substantially different from K96243. Subsequent investigation identified that although NCTC 13392 was received from the same laboratory in Thailand at the same time as K96243, it was unexpectedly a different strain; the authors of the marmoset study have since published an erratum to describe this discrepancy (6). Here, we describe the genome to provide the context for the functional results obtained from the marmoset model study.

Genomic DNA was isolated with Wizard Genomic DNA purification kit (Promega, Inc.). The genomic sequence was generated on the Illumina GA IIx platform, using both 300-bp and 1,000-bp insert sizes. A comparative assembly was performed with AMOScmp (7) with the 1,000-bp insert sequences, using K96243 as the reference sequence, based on preliminary phylogenetic analyses. The AMOS assembly was then processed with the PAGIT pipeline (8) to improve the contiguity of the assembly. A *de novo* assembly was also performed with Velvet (9) in conjunction with the VelvetOptimiser (<http://bioinformatics.net.au/software.velvetoptimiser.shtml>). The reference-based assembly and the *de novo* assembly were then aligned with Mugsy (10). Regions unique to the *de novo* assembly were parsed from the alignment and concatenated to the comparative assembly. Errors were corrected in the final assembly with iCORN (11). The raw reads were also mapped to the final assembly with the Burrows-Wheeler Aligner

(BWA) (12), and SNPs were called with the Genome Analysis Toolkit (GATK) (13).

The final assembly consisted of 48 contigs, with an assembled genome length of 7.16 Mbp and an N_{50} of ~418 Kbp. The average coverage across the genome was $140\times$ (1,000-bp library). An *in silico* multilocus sequence type (MLST) analysis classified NCTC 13392 as sequence type 23 (ST23), with a single nucleotide difference to K96243 in the *gltB* gene. There are 7 ST23 isolates in the MLST database (<http://bpseudomallei.mlst.net/>), all of which were isolated in Thailand. An SNP analysis of raw reads against the finished genome of K96243 identified 13,256 SNPs, including 3,050 nonsynonymous SNPs. In addition, ~121 Kbp of sequence data was identified in NCTC 13392 that was absent from K96243; genes from these regions were associated with bacteriophages in previously sequenced *B. pseudomallei* genomes. This analysis demonstrates that the genome of NCTC 13392 is distinct from that of K96243 and that genomic elements from NCTC 13392 might be used to correlate with the results from animal model studies.

Nucleotide sequence accession numbers. This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AOUG000000000](https://www.ncbi.nlm.nih.gov/nuclink/AOUG000000000). The version described in this paper is the first version, accession no. [AOUG010000000](https://www.ncbi.nlm.nih.gov/nuclink/AOUG010000000).

ACKNOWLEDGMENTS

This work was funded by the Biomedical Advanced Research and Development Authority (BARDA) through Battelle Memorial Institute (BMI) contract no. HHSO10020110000051 to H.C.G. and A.T.

REFERENCES

1. White NJ. 2003. Melioidosis. *Lancet* 361:1715–1722.
2. Dance DA. 2000. Melioidosis as an emerging global problem. *Acta Trop.* 74:115–119.
3. Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, Price EP, Glass MB, Leadem B, Beckstrom-Sternberg JS, Allan GJ, Foster JT, Wagner DM, Okinaka RT, Sim SH, Pearson O, Wu Z, Chang J, Kaul R, Hoffmaster AR, Brettn TS, Robison RA, Mayo M,

- Gee JE, Tan P, Currie BJ, Keim P. 2009. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol.* 7:78.
4. Nelson M, Dean RE, Salguero FJ, Taylor C, Pearce PC, Simpson AJ, Lever MS. 2011. Development of an acute model of inhalational melioidosis in the common marmoset (*Callithrix jacchus*). *Int. J. Exp. Pathol.* 92:428–435.
 5. Holden MT, Titball RW, Peacock SJ, Cerdeño-Tárraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD, Sebaihia M, Thomson NR, Bason N, Beacham IR, Brooks K, Brown KA, Brown NF, Challis GL, Cherevach I, Chillingworth T, Cronin A, Crossett B, Davis P, DeShazer D, Feltwell T, Fraser A, Hance Z, Hauser H, Holroyd S, Jagels K, Keith KE, Maddison M, Moule S, Price C, Quail MA, Rabinowitsch E, Rutherford K, Sanders M, Simmonds M, Songsivilai S, Stevens K, Tumapa S, Vesaratchaveest M, Whitehead S, Yeats C, Barrell BG, Oyston PC, Parkhill J. 2004. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc. Natl. Acad. Sci. U. S. A.* 101:14240–14245.
 6. Nelson DM, Dean RE, Salguero FJ, Taylor C, Pearce PC, Simpson AJ, Lever M. 2013. Erratum. *Int. J. Exp. Pathol.* 94:74.
 7. Pop M, Phillippy A, Delcher AL, Salzberg SL. 2004. Comparative genome assembly. *Brief. Bioinform.* 5:237–248.
 8. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat. Protoc.* 7:1260–1284.
 9. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
 10. Angiuoli SV, Salzberg SL. 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* 27:334–42.
 11. Otto TD, Sanders M, Berriman M, Newbold C. 2010. Iterative correction of reference nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* 26:1704–1707.
 12. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
 13. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43:491–498.