

Complete Genome Sequences for 59 *Burkholderia* Isolates, Both Pathogenic and Near Neighbor

Shannon L. Johnson,^a Kimberly A. Bishop-Lilly,^{c,d} Jason T. Ladner,^e Hajnalka E. Daligault,^a Karen W. Davenport,^a James Jaissle,^b Kenneth G. Frey,^{c,d} Galina I. Koroleva,^e David C. Bruce,^c Susan R. Coyne,^b Stacey M. Broomall,^f Po-E Li,^a Hazuki Teshima,^a Henry S. Gibbons,^f Gustavo F. Palacios,^e C. Nicole Rosenzweig,^f Cassie L. Redden,^{c,d} Yan Xu,^a Timothy D. Minogue,^b Patrick S. Chain^a

Los Alamos National Laboratory (LANL), Los Alamos, New Mexico, USA^a; United States Army Medical Research Institute of Infectious Diseases (USAMRIID) Diagnostic Systems Division (DSD), Fort Detrick, Maryland, USA^b; Naval Medical Research Center (NMRC)-Frederick, Fort Detrick, Maryland, USA^c; Henry M. Jackson Foundation, Bethesda, Maryland, USA^d; United States Army Medical Research Institute of Infectious Diseases (USAMRIID) Center for Genome Sciences (CGS), Fort Detrick, Maryland, USA^e; United States Army Edgewood Chemical Biological Center (ECBC), Aberdeen Proving Ground, Aberdeen, Maryland, USA^f

The genus *Burkholderia* encompasses both pathogenic (including *Burkholderia mallei* and *Burkholderia pseudomallei*, U.S. Centers for Disease Control and Prevention Category B listed), and nonpathogenic Gram-negative bacilli. Here we present full genome sequences for a panel of 59 *Burkholderia* strains, selected to aid in detection assay development.

Received 6 February 2015 Accepted 30 March 2015 Published 30 April 2015

Citation Johnson SL, Bishop-Lilly KA, Ladner JT, Daligault HE, Davenport KW, Jaissle J, Frey KG, Koroleva GI, Bruce DC, Coyne SR, Broomall SM, Li P-E, Teshima H, Gibbons HS, Palacios GF, Rosenzweig CN, Redden CL, Xu Y, Minogue TD, Chain PS. 2015. Complete genome sequences for 59 *Burkholderia* isolates, both pathogenic and near neighbor. *Genome Announc* 3(2):e00159-15. doi:10.1128/genomeA.00159-15.

Copyright © 2015 Johnson et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Shannon L. Johnson, shannonj@lanl.gov.

Burkholderia mallei and *Burkholderia pseudomallei* are among the bacterial species considered to be potential bioweapons, along with *Bacillus anthracis*, *Brucella melitensis*, *Brucella abortus*, and *Yersinia pestis* (1, 2). *B. pseudomallei* causes melioidosis, often a respiratory infection mimicking tuberculosis, while *B. mallei* generally infects horses, causing glanders. The listing of these bacteria as potential biothreats is due to their easy availability (*B. pseudomallei* is often recovered from soils in regions where it is endemic), their ability to cause severe and often fatal disease, multiple routes of infection, native antibiotic resistance, lack of available vaccines, wide host range, and ability to persist in the environment for weeks to years (3–9). *B. mallei* was reportedly used as a biological weapon on several occasions (10–14); however, while *B. pseudomallei* was investigated for its use as a bioweapon, there are no reports that it has been employed in this fashion (5, 11). Other *Burkholderia* species are opportunistic pathogens (e.g., the *Burkholderia cepacia* complex [Bcc] that adversely affects cystic fibrosis patients [including 7 species sequenced here]), plant pathogens (such as *Burkholderia gladioli*) and/or common soil microorganisms. Here we present full genome sequences of 59 strains useful for detection assay development, including both species that should be detected (inclusivity) and those that should not be (exclusivity).

Draft genome assemblies included two or more data sets (specific data types and coverages are listed in the NCBI records): Illumina (short- and/or long-insert paired data), Roche 454 (long-insert paired data), and PacBio long reads. Short- and long-insert paired data were assembled together in both Newbler and Velvet, and computationally shredded into 1.5-kbp overlapping shreds. If the PacBio coverage was 100× or greater, the data were assembled using PacBio's Hierarchical Genome Assembly Process (HGAP) (15). All data were additionally assembled together in Allpaths whenever possible (16). Consensus sequences from both

HGAP and Allpaths were computationally shredded into 10-kbp overlapping pieces. All shreds were integrated using Phrap. Possible misassemblies were corrected and repeat regions verified using in-house scripts and manual editing in Consed (17–19). All of the genomes were assembled into finished-quality complete genomes (20). Each genome assembly was annotated using an Ergatis-based (21) workflow with minor manual curation.

Genome assemblies range from 5.4 to 9.7 Mb (Table 1, mean 6.96 ± 0.014 Mb), with two or three chromosomes and up to three plasmids. As expected for the genus, the G+C content was high, averaging 67.7%.

Nucleotide sequence accession numbers. Accession numbers for all 59 genomes are listed in Table 1.

ACKNOWLEDGMENTS

Funding for this effort was provided by the Defense Threat Reduction Agency's Joint Science and Technology Office (DTRA J9-CB/JSTO) and Department of Homeland Security Science and Technology Directorate, award HSHQDC-08-X-00790. This article is approved by LANL for unlimited release (LA-UR-14-29605).

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the United States Government.

Bacterial strains were obtained from the Department of Defense's Unified Culture Collection (<http://www.usamriid.army.mil/ucc/>) or BEI Resources (<http://beiresources.org/>).

REFERENCES

- Thavaselvam D, Vijayaraghavan R. 2010. Biological warfare agents. *J Pharm Bioallied Sci* 2:179–188. <http://dx.doi.org/10.4103/0975-7406.68499>.
- Lim DV, Simpson JM, Kearns EA, Kramer MF. 2005. Current and developing technologies for monitoring agents of bioterrorism and bio-warfare. *Clin Microbiol Rev* 18:583–607. <http://dx.doi.org/10.1128/CMR.18.4.583-607.2005>.

3. Arun S, Neubauer H, Gürel A, Ayyildiz G, Kuşçu B, Yesildere T, Meyer H, Hermanns W. 1999. Equine glanders in Turkey. *Vet Rec* 144:255–258. <http://dx.doi.org/10.1136/vr.144.10.255>.
4. Cheng AC, Currie BJ. 2005. Melioidosis: epidemiology, pathophysiology, and management. *Clin Microbiol Rev* 18:383–416. <http://dx.doi.org/10.1128/CMR.18.2.383-416.2005>.
5. Lafontaine ER, Zimmerman SM, Shaffer TL, Michel F, Gao X, Hogan RJ. 2013. Use of a safe, reproducible, and rapid aerosol delivery method to study infection by *Burkholderia pseudomallei* and *Burkholderia mallei* in mice. *PLoS One* 8:e76804. <http://dx.doi.org/10.1371/journal.pone.0076804>.
6. Lazar Adler NR, Govan B, Cullinane M, Harper M, Adler B, Boyce JD. 2009. The molecular and cellular basis of pathogenesis in melioidosis: how does *Burkholderia pseudomallei* cause disease? *FEMS Microbiol Rev* 33: 1079–1099. <http://dx.doi.org/10.1111/j.1574-6976.2009.00189.x>.
7. Neubauer H, Meyer H, Finke E-J. 1997. Human glanders. *Rev Int Serv Santé Forces Armées* 70:258–265.
8. Whitlock GC, Estes DM, Torres AG. 2007. Glanders: off to the races with *Burkholderia mallei*. *FEMS Microbiol Lett* 277:115–122. <http://dx.doi.org/10.1111/j.1574-6968.2007.00949.x>.
9. Wiersinga WJ, van der Poll T, White NJ, Day NP, Peacock SJ. 2006. Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *Nat Rev Microbiol* 4:272–282. <http://dx.doi.org/10.1038/nrmicro1385>.
10. Waag D, DeShazer D. 2005. Glanders: new insights into an Old disease, p 209–237. In Lindler L, Lebeda F, Korch G (ed), *Biological weapons defense: infectious diseases and counterbioterrorism*. Humana Press, Totowa, NJ.
11. Wheelis M. 1998. First shots fired in biological warfare. *Nature* 395:213. <http://dx.doi.org/10.1038/26089>.
12. Sharrer GT. 1995. The great glanders epizootic, 1861–1866: a Civil War legacy. *Agric Hist* 69:79–97.
13. Christopher GW, Cieslak TJ, Pavlin JA, Eitzen EM, Jr. 1997. Biological warfare. A historical perspective. *JAMA* 278:412–417. <http://dx.doi.org/10.1001/jama.1997.03550050074036>.
14. Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T, Ulrich RL, Ronning CM, Brinkac LM, Daugherty SC, Davidsen TD, Deboy RT, Dimitrov G, Dodson RJ, Durkin AS, Gwinn ML, Haft DH, Khouri H, Kolonay JF, Madupu R, Mohammoud Y, Nelson WC, Radune D, Romero CM, Sarria S, Selengut J, Shamblyn C, Sullivan SA, White O, Yu Y, Zafar N, Zhou L, Fraser CM. 2004. Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci USA* 101: 14246–14251. <http://dx.doi.org/10.1073/pnas.0403306101>.
15. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
16. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. *Genome Res* 18:810–820. <http://dx.doi.org/10.1101/gr.7337908>.
17. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* 8:186–194.
18. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces UsingPhred. I. Accuracy assessment. *Genome Res* 8:175–185. <http://dx.doi.org/10.1101/gr.8.3.175>.
19. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202. <http://dx.doi.org/10.1101/gr.8.3.195>.
20. Chain PS, Graffham DV, Fulton RS, FitzGerald MG, Hostetler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Genomic Standards Consortium Human Microbiome Project Jumpstart Consortium, Detter JC. 2009. Genomics. Genome project standards in a new era of sequencing. *Science* 326:236–237. <http://dx.doi.org/10.1126/science.1180614>.
21. Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q. 2010. An Ergatis-based prokaryotic genome annotation web server. *Bioinformatics* 26:1122–1124. <http://dx.doi.org/10.1093/bioinformatics/btq090>.

TABLE 1 Listing of *Burkholderia* isolate genomes released to NCBI

Species and isolate	Accession no. (no. of contigs) ^a	Panel ^b	Genome (bp)	No. of plasmids	No. of CDSs ^c	G+C content (%)
<i>B. ambifaria</i>						
AMMD	CP009797–CP009800	E	7,528,578	1	6,602	67
<i>B. cepacia</i>						
LMG 16656	JTDP00000000 (5)	E	7,923,342	1	7,278	68
<i>B. dolosa</i>						
AU0158	CP009793–CP009795	E	6,409,095	2	5,657	67
<i>B. fungorum</i>						
ATCC BAA-463	CP010024–CP010027	E	9,058,983	1	8,206	62
<i>B. gladioli</i>						
ATCC 10248	CP009319–CP009322	E	8,899,459	3	7,561	68
<i>B. glumae</i>						
ATCC 33617	CP009432–CP009435	E	6,820,727	2	5,864	68
<i>B. mallei</i>						
6	CP008710–CP008711	I	5,647,769	0	4,872	68
11	CP009587–CP009588	I	5,913,134	0	5,083	68
NCTC 10247	CP007801–CP007802	I	5,827,656	0	5,001	68
2000031063	CP008731–CP008732	I	5,874,930	0	5,067	68
2002721276	CP010065–CP010066	I	5,780,439	0	4,954	69
2002734299	CP009337–CP009338	I	5,740,115	0	4,966	68
2002734306	CP009707–CP009708	I	5,409,162	0	4,703	68
China5	JPNX00000000 (2)	I	5,869,855	0	5,043	68
FMH 23344	CP008704–CP008705	I	5,625,292	0	4,883	68

(Continued on following page)

TABLE 1 (Continued)

Species and isolate	Accession no. (no. of contigs) ^a	Panel ^b	Genome (bp)	No. of plasmids	No. of CDSs ^c	G+C content (%)
India86-567-2	CP009642–CP009643	I	5,686,446	0	4,911	68
KC_1092	CP009942–CP009943	I	5,661,851	0	4,868	68
<i>B. multivorans</i>						
BAA-247	CP009830–CP009832	E	6,322,746	0	5,607	67
<i>B. oklahomensis</i>						
C6786	CP009555–CP009556	E	7,135,022	0	6,083	67
EO147	CP008726–CP008727	E	7,313,673	0	6,312	67
<i>B. pseudomallei</i>						
9	CP008753–CP008755	I	7,228,737	1	5,978	68
576	CP008777–CP008778	I	7,266,604	0	5,944	68
1026b	CP004379–CP004380	I	7,450,511	0	6,113	68
1106a	CP008758–CP008759	I	7,086,433	0	5,758	68
7894	CP009535–CP009536	I	7,381,912	0	6,036	68
PB08298010	CP009550–CP009551	I	7,375,551	0	6,023	68
K96243	CP009537–CP009538	I	7,247,614	0	5,933	68
MSHR 146	CP004042–CP004043	I	7,313,103	0	5,963	68
MSHR 1655	CP008779–CP008780	I	7,027,950	0	5,798	68
MSHR 2543	CP009477–CP009478	I	7,446,569	0	6,183	68
MSHR 305	CP006469–CP006470	I	7,428,072	0	6,105	68
MSHR 346	CP008763–CP008764	I	7,354,416	0	6,015	68
MSHR 406e	CP009297–CP009298	I	7,271,506	0	5,927	68
MSHR 491	CP009484–CP009485	I	7,356,376	0	6,080	68
MSHR 511	CP004023–CP004024	I	7,316,085	0	5,964	68
MSHR 520	CP004368–CP004369	I	7,447,511	0	6,113	68
MSHR 668	CP009545–CP009546	I	7,042,714	0	5,793	68
MSHR 840	CP009473–CP009474	I	7,129,813	0	5,860	68
NAU 20B-16	CP004003–CP004004	I	7,313,851	0	5,969	68
NAU 35A-3	CP004377–CP004378	I	7,204,083	0	5,844	68
NCTC 13178	CP004001–CP004002	I	7,408,007	0	6,133	68
NCTC 13179	CP003976–CP003977	I	7,337,157	0	6,085	68
Pasteur 52237	CP009898–CP009899	I	7,325,318	0	6,015	68
PHLS 112	CP009585–CP009586	I	7,202,363	0	5,868	68
<i>B. thailandensis</i>						
2002721643	CP009601–CP009602	E	6,722,801	0	5,649	68
2002721687	CP009547–CP009549	E	7,285,824	1	6,327	67
2002721723	CP004097–CP004098	E	6,577,133	0	5,533	68
2003015869	CP008914–CP008915	E	6,728,980	0	5,679	68
34	CP010016–CP010018	E	7,120,198	1	6,129	67
E254	CP004381–CP004382	E	6,676,730	0	5,591	68
E264	CP008785–CP008786	E	6,722,099	0	5,655	68
E444	CP004117–CP004118	E	6,651,696	0	5,571	68
H0587	CP004089–CP004090	E	6,768,375	0	5,629	68
Malaysia 20	CP004383–CP004384	E	6,684,359	0	5,620	68
MSMB 121	CP004095–CP004096	E	6,731,379	0	5,758	68
Phuket 4W-1	AQJ00000000 (3)	E	6,674,944	0	5,635	68
<i>B. ubonensis</i>						
MSMB 22	CP009486–CP009488	E	7,189,071	0	6,257	67
<i>B. vietnamiensis</i>						
LMG 10929	CP009629–CP009632	E	6,930,496	1	6,120	67
<i>B. xenovorans</i>						
LB400	CP008760–CP008762	E	9,702,951	0	8,684	63

^a Contig count is listed only for genomes at Improved High Quality Draft (IHQD) quality; all others are finished (20).

^b E, exclusivity strain; I, inclusivity strain.

^c CDS, coding sequence.