




Genome Sequences of *Burkholderia thailandensis* Strains E421, E426, and DW503

Catherine M. Mageeney,^a Anupama Sinha,^a Kelly P. Williams,^a  Steven S. Branda^a

^aSandia National Laboratories, Livermore, California, USA

ABSTRACT We present the draft genome sequences of three *Burkholderia thailandensis* strains, E421, E426, and DW503. E421 consists of 90 contigs of 6,639,935 bp and 67.73% GC content. E426 consists of 106 contigs of 6,587,853 bp and 67.73% GC content. DW503 consists of 102 contigs of 6,458,767 bp and 67.64% GC content.

B*urkholderia thailandensis*, occurring naturally in soil and water, serves as a surrogate for the study of its more pathogenic congeners, select agent bacteria *Burkholderia pseudomallei* and *Burkholderia mallei*, with which it shares many genome features and pathogenesis mechanisms, and it provokes similar immune responses in mammalian systems (1). We report the genome sequences of three *B. thailandensis* strains obtained from BEI Resources (Manassas, VA, USA). *B. thailandensis* E421 and E426 were isolated independently from a rice field in Northeast Thailand. *B. thailandensis* DW503 is an allelic exchange mutant of *B. thailandensis* E264 in which the gene cluster comprising *amrR*, *amrA*, *amrB*, and *oprA* is deleted (2).

E421, E426, and DW503 were grown in Luria broth in Erlenmeyer flasks at 37°C for ~16 h. From each culture, 2 ml was centrifuged at 4,000 × *g* for 4 min, and DNA was extracted from the pelleted bacteria using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany). DNA sequencing libraries were generated using the Nextera DNA library prep and DNA sample preparation index kits (Illumina, San Diego, CA, USA). Library concentrations were measured using the Qubit fluorometer with its high-sensitivity DNA assay kit (Thermo Fisher Scientific, Waltham, MA, USA). Equal amounts of the three libraries were mixed to generate a multiplexed library; its concentration was measured using the Qubit device, and fragment sizes were measured using the Bioanalyzer electrophoresis system with its high-sensitivity DNA analysis kit (Agilent, Santa Clara, CA, USA). The multiplexed library was sequenced in paired-end mode using a NextSeq 500 instrument with its high-output (300 cycles) kit (Illumina).

Reads were quality filtered using BBDuk (27 June 2016 release with the following parameters: *ktrim* = *r*, *k* = 21, *mink* = 11, *hdist* = 1) (<http://jgi.doe.gov/data-and-tools/bb-tools/>). Assembly of the sequencing reads was completed using the default settings for SPAdes v. 3.9.0 (3) (Table 1). FastANI alignment (4), run using default parameters, revealed that E421 and E426 share 99.8% average nucleotide identity (ANI) with E264. DW503 shares 99.9978% ANI with E264; its deletion at nucleotide position 129447 (GenBank accession number [JAAAQG010000016](https://www.ncbi.nlm.nih.gov/nuccore/JAAAQG010000016)) spans 148 kbp (CP000086.1, coordinates 2668007 to 2816380) and 99 genes (chromosomal loci BTH_I2364 to BTH_I2461), including the gene cluster comprising *amrR*, *amrA*, *amrB*, and *oprA*, as confirmed by analysis of the raw reads with readStepper (5). Sequence annotation was performed using our tater.pl pipeline (8), which calls Prokka v. 1.11 (7), tFind, and rFind software (8); counts for coding regions, miscellaneous RNA (miscRNA), rRNA sequences, and tRNA sequences (6) are listed in Table 1. All the strains had hits to the following Rfam (9) miscRNA profiles: 6S, Bacteria_large_SRP, Bacteria_small_SRP, Betaproteobacteria_toxic_sRNA, *cspA*, FMN, Intron_ *gpl*, mini-*ykkC*, P9, *pfl*, RNaseP_bact_a, SAH_riboswitch,

Citation Mageeney CM, Sinha A, Williams KP, Branda SS. 2020. Genome sequences of *Burkholderia thailandensis* strains E421, E426, and DW503. Microbiol Resour Announc 9:e00312-20. <https://doi.org/10.1128/MRA.00312-20>.

Editor Julia A. Maresca, University of Delaware

Copyright © 2020 Mageeney et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Steven S. Branda, sbranda@sandia.gov.

Received 24 March 2020

Accepted 29 April 2020

Published 21 May 2020

TABLE 1 Assembly statistics and genome sequence features

Feature ^a	Data for strain:		
	E421	E426	DW503
No. of reads	133,200,415	136,919,906	61,127,893
Coverage (fold)	2,824	2,997	1,440
No. of contigs	90	106	102
Total size (bp)	6,639,935	6,587,853	6,458,767
<i>N</i> ₅₀ (bp)	159,164	121,401	156,901
GC content (%)	67.73	67.73	67.64
No. of ambiguous bases	10	924	0
No. of CDS	5,608	5,524	5,427
No. of rRNA operons	3	2	2
No. of tRNA/tmRNA genes	52	47	50
No. of miscRNA genes	27	25	21
No. of CRISPR-Cas arrays	1	3	4
No. of GIs	7	5	8
Prophage coordinates ^b	45oatA selB 02:159872–204745 17mutS 10:52214–68931 40F 16:5517–45277	41Hyp 04:49998–75818	16mutS 12:52204–68122 44pcaK 33:9582–54013 11R 49:24269–32704, 84:10537–2253
Other GI coordinates ^b	26dapA sbp 04:49955–75760 33T 19:84650–117520 22S 21:23808–45519 24P 32:39753–64195	8G 01:381563–389876 3Phage_GPD 14:8710–49502 19A 16:4992–8027 26dapA sbp 23:49376–68235	49CbiA HBT 08:149573–198632 19Hyp 20:23740–42637 22A 55:4198–4233, 68:1–10395, 82:2548–1, 89:1–1497, 55:1–4200 48R 50:2569–32098, 89:1–1497, 67:1–10868, 83:2372–1, 50:2517–2568 3Acyl_transf_3 42:53140–55361, 05:2308–2523, 17:137203–136908
GenBank accession no.	JAAAQH000000000	JAAAWT000000000	JAAAQG000000000
NCBI SRA project no.	PRJNA600278	PRJNA600279	PRJNA600280
NCBI run no.	SRR11144399	SRR11142218	SRR11073226

^a CDS, coding DNA sequences; tmRNA, transfer-messenger RNA.
^b Names of prophages and other genomic islands (GIs) are in bold, followed by their corresponding contig numbers and nucleotide positions (note that DW503 GIs 11R, 22A, 48R, and 3Acyl_transf_3 are split between two or more contigs).

SECIS_3, *sucA*, and TPP. In addition, E421 and E426 had hits to C4, cobalamin, and *isrK*. Additionally, sequences were analyzed to identify CRISPR-Cas arrays (CRISPRCasFinder [10]), prophages, and other genomic islands (6, 8). The TIGER package (6), including tater.pl, genomic island prediction software, and prophage calling software, is available at github.com/sandialabs/TIGER.

Names of prophages and other genomic islands (GIs), contig numbers, and coordinates are listed in Table 1. DW503 is a deletion mutant of E264 (GenBank assembly number [GCA_000012365](https://www.ncbi.nlm.nih.gov/assembly/GCA_000012365)) (2); we previously reported the coordinates for E264 prophages and GIs (6). Some of the GIs in E264 were found to be split between two or more DW503 contigs (separated by commas in Table 1).

Data availability. The assemblies have been deposited at GenBank under the following accession numbers: [JAAAQH000000000](https://www.ncbi.nlm.nih.gov/assembly/JAAAQH000000000) (E421), [JAAAWT000000000](https://www.ncbi.nlm.nih.gov/assembly/JAAAWT000000000) (E426), and [JAAAQG000000000](https://www.ncbi.nlm.nih.gov/assembly/JAAAQG000000000) (DW503). The raw data have been deposited at the NCBI Sequence Read Archive (SRA) under the BioProject and run accession numbers [PRJNA600278](https://www.ncbi.nlm.nih.gov/sra/PRJNA600278) and [SRR11144399](https://www.ncbi.nlm.nih.gov/sra/SRR11144399) (E421), [PRJNA600279](https://www.ncbi.nlm.nih.gov/sra/PRJNA600279) and [SRR11142218](https://www.ncbi.nlm.nih.gov/sra/SRR11142218) (E426), and [PRJNA600280](https://www.ncbi.nlm.nih.gov/sra/PRJNA600280) and [SRR11073226](https://www.ncbi.nlm.nih.gov/sra/SRR11073226) (DW503), respectively.

ACKNOWLEDGMENTS

Sandia National Laboratories is a multitechnology laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.

Strains E421 (catalog number NR-9909), E426 (catalog number NR-9908), and DW503 (catalog number NR-4075) were obtained through BEI Resources, NIAID, NIH.

REFERENCES

- Wiersinga WJ, de Vos AF, de Beer R, Wieland CW, Roelofs JJ, Woods DE, van der Poll T. 2008. Inflammation patterns induced by different *Burkholderia* species in mice. *Cell Microbiol* 10:81–87. <https://doi.org/10.1111/j.1462-5822.2007.01016.x>.
- Burntack M, Bolton A, Brett P, Watanabe D, Woods D. 2001. Identification of the acid phosphatase (*acpA*) gene homologues in pathogenic and non-pathogenic *Burkholderia* spp. facilitates *TnpHoA* mutagenesis. *Microbiology* 147:111–120. <https://doi.org/10.1099/00221287-147-1-111>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell se-

- quencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
4. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
 5. Hudson CM, Bent ZW, Meagher RJ, Williams KP. 2014. Resistance determinants and mobile genetic elements of an NDM-1 encoding *Klebsiella pneumoniae* strain. *PLoS One* 9:e99209. <https://doi.org/10.1371/journal.pone.0099209>.
 6. Mageeney CM, Lau BY, Wagner JM, Hudson CM, Schoeniger JS, Krishnakumar R, Williams KP. 2020. New candidates for regulated gene integrity revealed through precise mapping of integrative genetic elements. *Nucleic Acids Res* 48:4052–4065. <https://doi.org/10.1093/nar/gkaa156>.
 7. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 8. Hudson CM, Lau BY, Williams KP. 2015. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res* 43:D48–D53. <https://doi.org/10.1093/nar/gku1072>.
 9. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, Petrov AI. 2018. Non-coding RNA analysis using the Rfam database. *Curr Protoc Bioinformatics* 62:e51. <https://doi.org/10.1002/cpbi.51>.
 10. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCas-Finder, an update of CRISRFinder [sic], includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 46:W246–W251. <https://doi.org/10.1093/nar/gky425>.