AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete and Circularized Genome Assemblies of the *Kroppenstedtia eburnea* Genus Type Strain and the *Kroppenstedtia pulmonis* Species Type Strain with MiSeq and MinION Sequence Data

Christopher A. Gulvik,[a] Dhwani Batra,[b] Lori A. Rowe,[b] Milli Sheth,[b] Ben W. Humrighouse,[a] Dakota T. Howard,[b] Justin Lee,[b] John R. McQuiston,[a] Brent A. Lasker[a]

[a]Bacterial Special Pathogens Branch, Division of High-Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[b]Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

ABSTRACT *Kroppenstedtia eburnea* DSM 45196^T and *Kroppenstedtia pulmonis* W9323^T are aerobic, Gram-positive, filamentous, chemoorganotrophic thermoactinomycetes. Here, we report on the complete and circular genome assemblies generated using Illumina MiSeq and Oxford Nanopore Technologies MinION reads. Putative gene clusters predicted to be involved in the production of secondary metabolites were also identified.

Isolates within the *Kroppenstedtia* genus are characterized as Gram-positive, nonmotile, aerobic, filamentous chemotrophs capable of producing heat-resistant endospores (1–3). The four species of the genus include *Kroppenstedtia eburnea* (1), *Kroppenstedtia guangzhouensis* (2), *Kroppenstedtia pulmonis* (3), and *Kroppenstedtia sanguinis* (3). *K. eburnea* DSM 45196^T was isolated from a plastic surface in Germany; subsequently, clinical isolates of the same species were identified in the United States (1, 4). *K. pulmonis* W9323^T was isolated from a lung biopsy sample from a 78-year-old male patient from the United States (3). The genomes for *Kroppenstedtia eburnea* DSM 45196^T and *Kroppenstedtia pulmonis* W9323^T described here were sequenced because of their potential sources of genes encoding secondary metabolites, as well as adaptation of an environmental thermoactinomycete isolated from soil to clinical sources.

Kroppenstedtia eburnea DSM 45196^T was purchased from the German Collection of Microorganisms and Cell Cultures (catalogue number DSM45196), and *Kroppenstedtia pulmonis* W9323^T was obtained from the Special Bacteriology Reference Laboratory, Centers for Disease Control and Prevention (CDC) (Atlanta, GA) (3). Cells were grown in Trypticase soy broth from single colonies, and genomic DNA used for both libraries was purified using the MasterPure DNA purification kit (Epicentre, Madison, WI) according to the manufacturer's protocol (5). MinION libraries were made with the rapid barcoding kit (Oxford Nanopore Technologies), and sequences were generated with R9.4.1 flow cells and Guppy v3.2.8+bd97289. The numbers of raw reads for *K. eburnea* from the MiSeq and MinION instruments were 13,310,694 and 148,000, respectively, and those for *K. pulmonis* from the MiSeq and MinION instruments were 2,248,184 and 404,000, respectively. The MinION sequence N_{50} values were 5,681 bp and 5,607 bp for *K. eburnea* and *K. pulmonis*, respectively. Default parameters were used for all software unless otherwise specified. Flye v2.6 with the setting "-g 4m" formed initial assemblies (6). Three sequential rounds of assembly corrections were performed with minimap

Citation Gulvik CA, Batra D, Rowe LA, Sheth M, Humrighouse BW, Howard DT, Lee J, McQuiston JR, Lasker BA. 2020. Complete and circularized genome assemblies of the *Kroppenstedtia eburnea* genus type strain and the *Kroppenstedtia pulmonis* species type strain with MiSeq and MinION sequence data. Microbiol Resour Announc 9:e00650-20. https://doi.org/10.1128/MRA.00650-20.

Editor David Rasko, University of Maryland School of Medicine

Address correspondence to Brent A. Lasker, blasker@cdc.gov.

Received 10 June 2020
Accepted 4 October 2020
Published 29 October 2020

v2.17-r941, using the setting "-x map-ont," and racon v1.3.2 (7, 8). A final long-read correction was accomplished with medaka v0.11.1, with 131× and 362× coverages for *K. eburnea* and *K. pulmonis*, respectively. Illumina MiSeq sequences (2 × 250 bp) that had been filtered with Trimmomatic v0.35 to scores of ≥Q30 were used with the Unicycler v0.4.8 polishing function, which used Bowtie v2.3.4.3, SAMtools v1.3.1, and Pilon v1.23 (9–13), with 642× and 89× coverages for *K. eburnea* and *K. pulmonis*, respectively. Polishing continued until the assembly likelihood scores no longer improved according to ALE v20180904 (four rounds for *K. eburnea* fixed 20,779 variants, and two rounds for *K. pulmonis* fixed 22,201 variants) (14). The 3,564,999-bp (54% GC content) and 3,345,811-bp (46% GC content) circular chromosomes for *K. eburnea* and *K. pulmonis*, respectively, were reoriented to start with *dnaA*, which was located with BLAST 2.9.0+ (15) using locus tag D1G38_003660 in *Kroppenstedtia sanguinis* (5). Pseudogenes were inferred by comparing proteins aligned with DIAMOND v0.9.22 to the UniProtKB/TrEMBL database v2019_10 (16–20). The numbers of best-match (based on bit scores) alignments with >10% deviation in length were 88 in *K. eburnea* and 298 in *K. pulmonis*. CheckM v1.0.13 estimated both assemblies as 100.00% complete (21). The genome assemblies were annotated using PGAP v4.11, which predicted 4 and 6 CRISPR sequences in *K. eburnea* and *K. pulmonis*, respectively (22). AntiSMASH v5.1.0 and Prism v4.4.3 predicted ectoine biosynthesis capability in both genomes (23, 24). AntiSMASH found 10 additional putative biosynthetic gene clusters (BGCs) of interest (5 nonribosomal peptide synthetases [NRPSs] and putative genes associated with siderophore production, which may enhance pathogenicity [25]) for *K. pulmonis* and 6 additional BGCs of interest (2 NRPSs) for *K. eburnea*. Thermonucleases can serve as virulence factors (26, 27); 2 were predicted for *K. eburnea* (protein accession numbers QKI81670.1 and QKI83414.1) and 1 was predicted for *K. pulmonis* (protein accession number QKG85853.1). These complete type strain genomes will be valuable for taxonomic assignments and will aid in biosynthesis and natural product research.

Data availability. The whole-genome sequences of *Kroppenstedtia eburnea* DSM 45196^T and *Kroppenstedtia pulmonis* W9323^T have been deposited in DDBJ/ENA/GenBank under the accession numbers CP048103 and CP048104, respectively. The BioProject accession number for *Kroppenstedtia eburnea* DSM 45196^T and *Kroppenstedtia pulmonis* W9323^T is PRJNA602730, and the associated BioSample accession numbers are SAMN13905711 and SAMN13905715, respectively. The *Kroppenstedtia eburnea* DSM 45196^T Illumina MiSeq reads and Nanopore reads are available in the NCBI Sequence Read Archive (SRA) under the accession numbers SRX7624950 and SRX7624951, respectively. The *Kroppenstedtia pulmonis* W9323^T Illumina MiSeq reads and Nanopore reads are available in the NCBI SRA under the accession numbers SRX7624960 and SRX7624961, respectively.

ACKNOWLEDGMENT

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC. The mention of company names or products does not constitute endorsement by the CDC.

REFERENCES

1. von Jan M, Riegger N, Pötter G, Schumann P, Verbarg S, Spröer C, Rohde M, Lauer B, Labeda DP, Klenk H-P. 2011. *Kroppenstedtia eburnea* gen. nov., sp. nov., a thermoactinomycete isolated by environmental screening, and emended description of the family *Thermoactinomycetaceae* Matsuo *et al.* 2006 emend. Yassin *et al*. 2011. Int J Syst Evol Microbiol 61:2304–2310. https://doi.org/10.1099/ijs.0.026179-0.
2. Yang G, Qin D, Wu C, Yuan Y, Zhou S, Cai Y. 2013. *Kroppenstedtia guangzhouensis* sp. nov., a thermoactinomycete isolated from soil. Int J Syst Evol Microbiol 63:4077–4080. https://doi.org/10.1099/ijs.0.051011-0.
3. Bell ME, Lasker BL, Klenk H-P, Hoyles L, Spröer C, Schumann P, Brown JM. 2016. *Kroppenstedtia pulmonis* sp nov. and *Kroppenstetia sanguinis* sp nov., isolated from human patients. Antonie Van Leeuwenhoek 109:603–610. https://doi.org/10.1007/s10482-016-0663-z.
4. Barker AP, Simmon KE, Cohen S, Slechta ES, Fisher MA, Schlaberg R. 2012. Isolation and identification of *Kroppenstedtia eburnea* isolates from multiple patient samples. J Clin Microbiol 50:3391–3394. https://doi.org/10.1128/JCM.01186-12.
5. Arthur RA, Nicholson AC, Humrighouse BW, McQuiston JR, Lasker BL. 2019. Draft genome sequence of *Kroppenstedtia sanguinis* X0209^T, a clinical isolate recovered from human blood. Microbiol Resour Announc 8:e00354-19. https://doi.org/10.1128/MRA.00354-19.
6. Kolmogorov M, Yuan J, Lin Y, Pevzner P. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
7. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
8. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.

9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
10. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
11. Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
13. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
14. Clark SC, Egan R, Frazier PI, Wang Z. 2013. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics 29:435–443. https://doi.org/10.1093/bioinformatics/bts723.
15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
16. Watson M, Warr A. 2019. Errors in long-read assemblies can critically affect protein prediction. Nat Biotechnol 37:124–126. https://doi.org/10.1038/s41587-018-0004-z.
17. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
18. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
19. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
20. UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049.
21. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
22. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
23. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.
24. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. 2017. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res 45:W49–W54. https://doi.org/10.1093/nar/gkx320.
25. Miethke M, Marahiel MA. 2007. Siderophore-based iron acquisition and pathogen control. Microbiol Mol Biol Rev 71:413–451. https://doi.org/10.1128/MMBR.00012-07.
26. Hu Y, Meng J, Shi C, Hervin K, Fratamico PM, Shi X. 2013. Characterization and comparative analysis of a second thermonuclease from *Staphylococcus aureus*. Microbiol Res 168:174–182. https://doi.org/10.1016/j.micres.2012.09.003.
27. Juneau RA, Stevens JS, Apicella MA, Criss AK. 2015. A thermonuclease of *Neisseria gonorrhoeae* enhances bacterial escape from killing by neutrophil extracellular traps. J Infect Dis 212:316–324. https://doi.org/10.1093/infdis/jiv031.