Microbiology Resource Announcements

| Bacteriology | Announcement

Complete genome sequence of the environmental *Burkholderia pseudomallei* strain 22-10884_313#20 from Guadeloupe, French West Indies

Bernice J. M. Klotoe,[1] Mégane Gasqué,[1,2] Fabien Vorimore,[3] Emma Newall-Rochelle,[2] Vanina Guernier-Cambert,[2,4] Karine Laroucau[1]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Burkholderia pseudomallei* is the causative agent of melioidosis. Here, we present the complete genome sequence of the strain 22-10884_313#20, isolated from a soil sample in Guadeloupe (French West Indies).

KEYWORDS *Burkholderia pseudomallei*, environmental, Guadeloupe

Burkholderia pseudomallei is a bacterium responsible for melioidosis, a serious tropical disease endemic to Southeast Asia and northern Australia. It thrives in contaminated soil or water and infects humans or animals by inoculation, inhalation, or ingestion (1). Clinical manifestations range from pneumonia to sepsis and skin infections, primarily affecting adults with comorbidities, particularly diabetes (2). Outside endemic areas, melioidosis is often underdiagnosed due to limited awareness and diagnostic challenges (3).

In the French West Indies, 25 cases of melioidosis have been reported since 1993 (4). Local contamination was suspected in patients without a travel history, but environmental *B. pseudomallei* was only recently confirmed with the isolation of strain 22-10884_313#20 from soil in Guadeloupe (4). Isolation involved liquid enrichment in erythritol and Ashdown media, plating on a modified chromogenic *Burkholderia cepacia* agar, and subculture on blood agar with 5% horse serum at 37°C for 4 days (4).

Genomic DNA from this strain was extracted for sequencing using two kits. For MinION sequencing, DNA was extracted using the Lucigen kit (Biosearch Technologies, UK), and libraries were prepared using the Rapid Barcoding SQK-RBK-004 kit (Oxford Nanopore Technologies, UK) without DNA shearing or size selection. Sequencing was performed on a MinION MK1B system using Flowcell R9.4.1, generating 25,025 reads with an N50 read length of 10,575 bp. Base calling was performed with Guppy (v6.5.7) using the model dna_r9.4.1_450bps_sup, and read analysis was performed with NanoPlot (v1.44.0) (5). Poor quality reads (10%) were filtered using Filtlong (v0.2.1) with parameters --keep_percent 90 and --mean_q_weight 10 (6). For MiSeq Illumina sequencing, DNA was extracted using the Roche kit (Roche Diagnostics, France). Libraries were prepared using the Illumina DNA Prep (M) Tagmentation kit (Illumina, USA) and sequenced on a MiSeq Illumina platform using the MiSeq Reagent kit v3 (2 × 300). Sequencing generated 2,784,709 paired-end reads with 100× depth coverage. Quality assessment was performed using FastQC (v0.12.1) (7), and low-quality reads (<Q20) and adapters were removed using Fastp (v0.23.4) with the parameters --corrections and --cut_right (8).

Hybrid assembly using the Flye nano-raw model (v2.9.5) (9) produced two circular contigs with an N50 of 4,122,689 bp. Each contig formed a single unique graph edge, which was visualized using Bandage v0.9.0 (10) to confirm the circularization of the contigs. The draft assembly was corrected with long reads using the Medaka v1.11.3 (Oxford Nanopore Technologies Ltd). Illumina reads were aligned to the draft

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Karine Laroucau, karine.laroucau@anses.fr.

The authors declare no conflict of interest.

Received 7 October 2024
Accepted 1 February 2025
Published 28 March 2025

genome using BWA-MEM (v0.7.18) (11), with further polishing using Polypolish (12) and Pypolca (13). Genome quality was assessed using QUAST (v5.2.0) (14). Default parameters were used for the BWA-MEM aligner, Polypolish, Pypolca, and QUAST. Annotation was performed using NCBI's PGAP (v2024-07-18.build7555) (15) with GeneMarkS-2+ (16).

The assembled genome of the strain is 7,364,047 bp in length, organized into two circular contigs (4,123,479 bp and 3,240,568 bp), corresponding to its two chromosomes. The annotation reveals 6,245 coding sequence genes, 235 pseudogenes, and 78 RNA genes. The genome has a G+C content of 67.94%, an N50 of 4,123,479 bp, and no undetermined bases (N).

Comparative genomic analysis of this strain with local and global strains may provide insights into the ecological and genomic adaptations of *B. pseudomallei* in the French West Indies.

ACKNOWLEDGMENTS

This study was supported by the Ile de France Region (Dim1Health). We thank the French Institut de Recherche pour le Développement (IRD) and the French Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES) for their financial support.

AUTHOR AFFILIATIONS

[1]Animal Health Laboratory, Bacterial Zoonosis Laboratory, Paris Est University, ANSES, Maisons-Alfort, Île-de-France, France
[2]Institute of Ecology and Environmental Sciences-Paris (iEES-Paris), Sorbonne Université, UPEC, IRD, CNRS, INRAE, Paris, Île-de-France, France
[3]Identypath, Paris Est University, ANSES, Maisons-Alfort, France
[4]Faculty of Veterinary Technology, Kasetsart University, Bangkok, Bangkok, Thailand

AUTHOR ORCIDs

Karine Laroucau http://orcid.org/0000-0001-7605-8602

AUTHOR CONTRIBUTIONS

Bernice J. M. Klotoe, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Mégane Gasqué, Investigation, Writing – original draft, Writing – review and editing | Fabien Vorimore, Data curation, Formal analysis, Methodology | Emma Newall-Rochelle, Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review and editing | Vanina Guernier-Cambert, Conceptualization, Project administration, Writing – original draft, Writing – review and editing | Karine Laroucau, Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The genome sequence data of strain 22-10884_313#20 has been deposited at ENA under accession number GCA_964417475.1. MinION and Illumina reads have been deposited at ENA under accession numbers ERR13717737 and ERR13718582, respectively.

REFERENCES

1. Meumann EM, Limmathurotsakul D, Dunachie SJ, Wiersinga WJ, Currie BJ. 2024. *Burkholderia pseudomallei* and melioidosis. Nat Rev Microbiol 22:155–169. https://doi.org/10.1038/s41579-023-00972-5
2. Gassiep I, Armstrong M, Norton R. 2020. Human melioidosis. Clin Microbiol Rev 33:e00006-19. https://doi.org/10.1128/CMR.00006-19
3. Savelkoel J, Dance DAB, Currie BJ, Limmathurotsakul D, Wiersinga WJ. 2022. A call to action: time to recognise melioidosis as a neglected tropical disease. Lancet Infect Dis 22:e176–e182. https://doi.org/10.1016/S1473-3099(21)00394-7
4. Gasqué M, Guernier-Cambert V, Manuel G, Aaziz R, Terret J, Deshayes T, Baudrimont X, Breurec S, Rochelle-Newall E, Laroucau K. 2024. Reassessing the distribution of *Burkholderia pseudomallei* outside known endemic areas using animal serological screening combined with environmental surveys: the case of Les Saintes (Guadeloupe) and French Guiana. PLoS Negl Trop Dis 18:e0011977. https://doi.org/10.1371/journal.pntd.0011977

5. De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39:btad311. https://doi.org/10.1093/bioinformatics/btad311
6. Wick R. 2021. Filtlong is a tool for filtering long reads by quality. Available from: https://github.com/rrwick/Filtlong
7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
8. Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. Imeta 2:e107. https://doi.org/10.1002/imt2.107
9. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
10. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383
11. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2.
12. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802
13. Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. 2024. How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. Microb Genom 10:001254. https://doi.org/10.1099/mgen.0.001254
14. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150. https://doi.org/10.1093/bioinformatics/bty266
15. 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
16. Lomsadze A, Gemayel K, Tang S, Borodovsky M. 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. Genome Res 28:1079–1089. https://doi.org/10.1101/gr.230615.117