CrossMark
click for updates

Draft Genome Sequence of the *Pandoraea apista* LMG 16407 Type Strain

A. D. Millard,[a] L. F. Westblade,[b] J. J. LiPuma,[c] K. Vavikolanu,[d] T. D. Read,[d,e] M. J. Pallen,[a] E. M. Burd,[e,f] C. I. Constantinidou[a]

Microbiology and Infection Unit, Warwick Medical School, University of Warwick, Coventry, United Kingdom[a]; Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York, USA[b]; Department of Pediatrics, University of Michigan, Ann Arbor, Michigan, USA[c]; Department of Medicine, Department of Infectious Diseases, Emory University School of Medicine, Atlanta, Georgia, USA[d]; Emory Antibiotic Resistance Center, Atlanta, Georgia, USA[e]; Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia, USA[f]

Pandoraea species, in particular *Pandoraea apista*, are opportunistic, multidrug-resistant pathogens in persons with cystic fibrosis (CF). To aid in understanding the role of *P. apista* in CF lung disease, we used Illumina MiSeq and nanopore MinION technology to sequence the whole genome of the *P. apista* LMG 16407^T.

Address correspondence to C. I. Constantinidou, C.I.Constantinidou@warwick.ac.uk.

The genus *Pandoraea* is a group of multidrug-resistant, nonglucose fermenting, Gram-negative rods that have been recovered from the respiratory tract of cystic fibrosis (CF) patients (1–3). Members of the genus are closely related to other CF pathogens, most notably *Burkholderia* species (1, 2). Therefore, accurate genus- and species-level identification can be challenging. *Pandoraea apista* is one of the most frequently isolated *Pandoraea* species from CF patients (3) and has been associated with transmission between patients (4). To better understand its pathogenic traits and assist phylogenetic analysis of the genus, we sequenced the genome of the type strain, *P. apista* LMG 16407^T, originally isolated from a CF patient (1).

Bacterial genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen, USA) and sequenced using MiSeq (Illumina, USA) and MinION (Oxford Nanopore Technologies, UK) instruments. A total of 1 ng of genomic DNA was prepared using the Nextera XT DNA sample preparation kit (Illumina) before sequencing on the MiSeq platform using the paired-end 2 × 251-bp (version 2) protocol. Prior to analysis on the MinION instrument, 1 μg of unsheared genomic DNA was treated with the Genomic DNA sequencing kit (SQK-MAP-002; Oxford Nanopore Technologies).

Using SPAdes (5), data derived from both sequencers were assembled to produce a draft genome, with the options of -careful and -pacbio for the MinION reads. Only contigs longer than 500 bp and with greater than 5× depth of coverage were included in the analysis. To correct any miscalled bases, the assembled genome was reviewed for errors by mapping reads against the individual contigs with BWA-MEM using the default settings (6). Contigs were sorted against the *P. apista* TF81F4 genome (CP010518) using the Mauve aligner (7). The draft genome included 45 contigs greater than 500 bp with a N$_{50}$ of 223,394 bp; the longest contig was 716,258 bp. The estimated size of the whole genome is 5,537,404 bp (mean depth of coverage, 39×) with a G+C content of 63%.

Genome annotation with Prokka (8) identified 4,993 potential protein-coding sequences (CDS), 7 rRNAs, and 62 tRNAs. The CDS include predicated virulence factors also present in the related genus *Burkholderia* (http://www.mgc.ac.cn/cgi-bin/VFs/compvfs.cgi?Genus=Burkholderia) and encompass genes associated with flagella, adhesion, capsule, and type III and type IV secretion systems. Additionally, the *P. apista* LMG 16407^T strain harbors a putative 61,143-bp plasmid that encodes a complete type IV secretion system.

Phylogenetic analysis of single copy genes present in the 20 publically available *Pandoraea* species genomes was performed using RAxML version 7.2.8 (9). The resultant data revealed that *P. apista* LMG 16407^T is part of a clade composed of *P. apista* strains TF81F3, TF80G25, and AU2161. Average nucleotide identity (ANI) analysis demonstrated that the sequenced *P. apista* strains have an ANI greater than 99%, which exceeds the 95 to 96% threshold used to define a species (10), and are therefore considered members of the same species.

Nucleotide sequence accession number. This draft genome sequence of *P. apista* LMG 16047^T has been deposited in DDBJ/ENA/GenBank under the accession number CEWL00000000.

REFERENCES

1. **Coenye T, Falsen E, Hoste B, Ohlén M, Goris J, Govan J, Gillis M, Vandamme P.** 2000. Description of *Pandoraea* gen. nov. with *Pandoraea apista* sp. nov., *Pandoraea pulmonicola* sp. nov., *Pandoraea pnomenusa* sp. nov., *Pandoraea sputorum* sp. nov. and *Pandoraea norimbergensis* comb. nov. Int J Syst Evol Microbiol **50:**887–899. http://dx.doi.org/10.1099/00207713-50-2-887.
2. **Daneshvar MI, Hollis DG, Steigerwalt AG, Whitney AM, Spangler L, Douglas MP, Jordan JG, MacGregor JP, Hill BC, Tenover FC, Brenner DJ, Weyant RS.** 2001. Assignment of CDC Weak Oxidizer Group 2

(WO-2) to the genus *Pandoraea* and characterization of three new *Pandoraea* genomospecies. J Clin Microbiol **39**:1819–1826. http://dx.doi.org/10.1128/JCM.39.5.1819-1826.2001.

3. **LiPuma JJ**. 2010. The changing microbial epidemiology in cystic fibrosis. Clin Microbiol Rev **23**:299–323. http://dx.doi.org/10.1128/CMR.00068-09.

4. **Jørgensen IM, Johansen HK, Frederiksen B, Pressler T, Hansen A, Vandamme P, Høiby N, Koch C**. 2003. Epidemic spread of *Pandoraea apista*, a new pathogen causing severe lung disease in cystic fibrosis patients. Pediatr Pulmonol **36**:439–446. http://dx.doi.org/10.1002/ppul.10383.

5. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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 sequencing. J Comput Biol **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

6. **Li H**. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2. http://arxiv.org/abs/1303.3997v2.

7. **Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT**. 2009. Reordering contigs of draft genomes using the mauve aligner. Bioinformatics **25**:2071–2073. http://dx.doi.org/10.1093/bioinformatics/btp356.

8. **Seemann T**. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

9. **Stamatakis A** 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313. http://dx.doi.org/10.1093/bioinformatics/btu033.

10. **Richter M, Rosselló-Móra R**. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA **106**:19126–19131. http://dx.doi.org/10.1073/pnas.0906412106.