AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of the Bacterium *Paraburkholderia aromaticivorans* AR20-38, a Gram-Negative, Cold-Adapted Degrader of Aromatic Compounds

Caroline Poyntner,[a] Dechao Zhang,[b,c] Rosa Margesin[a]

[a]Institute of Microbiology, University of Innsbruck, Innsbruck, Austria
[b]Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China
[c]University of Chinese Academy of Sciences, Beijing, China

ABSTRACT Here, we report the draft genome sequence of *Paraburkholderia aromaticivorans* strain AR20-38, a cold-adapted Gram-negative bacterium. It was isolated from Alpine forest soil and can degrade a range of aromatic compounds.

Paraburkholderia is a genus of *Proteobacteria*, class *Betaproteobacteria*. Members of this genus have been isolated from diverse ecological niches, including pristine and contaminated soil, sediments, rocks, and plants (1, 2).

Paraburkholderia aromaticivorans strain AR20-38 was isolated from an Italian Alpine forest soil sample (3). Soil samples were surface spread onto Reasoner's 2A (R2A) agar. Growing strains were subcultured, purified, and stored at −80°C. Due to its properties, strain AR20-38 was chosen for full-genome sequencing.

The strain was grown from a single colony on R2A agar and was further inoculated in nutrient broth incubated at 10°C until the stationary growth phase. After lyophilization, genomic DNA was extracted using lysozyme, SDS, and phenol-chloroform-isoamyl alcohol. DNA quality and quantity were determined using a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and agarose gel electrophoreses. DNA was used for Oxford Nanopore and Illumina sequencing.

The one-dimensional (1D) ligation sequencing kit (SQK-LSK109 kit; Oxford Nanopore) was used with additional reagents from New England Biolabs (NEBNext FFPE repair mix, NEBNext end repair/dA-tailing module, and NEBNext quick ligation module) following the manufacturer's recommendations. No size selection or shearing was applied.

For Illumina sequencing, 1 μg DNA was used with the NEBNext Ultra DNA library prep kit (New England Biolabs) following the manufacturer's recommendations. The Nanopore library was sequenced on the PromethION instrument (PromethION flow cells, FLO-PRO002; Oxford Nanopore), and the Illumina library was sequenced on the Illumina NovaSeq PE150 instrument at the Beijing Novogene Bioinformatics Technology Co. Ltd.

For all software used, default parameters were used except where otherwise noted.

The Nanopore fast5 file was base called using Guppy (Oxford Nanopore), and qcat was applied. Nanopore quality control was achieved using NanoPlot with a threshold value (Q) of >7, resulting in 132,813 reads with a median read length of 15,994 bp and an N_{50} value of 19,781 bp. Illumina data were quality controlled using Readfq, which removed reads containing more than 40% low-quality bases (quality value, ≤20), overlaps with adapter sequences, and duplicates. The Illumina reads were assembled using SPAdes 3.10.0 (4). A hybrid assembly was created using Racon (5), miniasm (6), and Unicycler 0.4.7 (7). The contigs were controlled for overlapping end sequences and start, end, *dnaA*, and *repA* sites, resulting in three assembled, circular chromosomes and

Citation Poyntner C, Zhang D, Margesin R. 2020. Draft genome sequence of the bacterium *Paraburkholderia aromaticivorans* AR20-38, a Gram-negative, cold-adapted degrader of aromatic compounds. Microbiol Resour Announc 9:e00463-20. https://doi.org/10.1128/MRA.00463-20.

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

Address correspondence to Caroline Poyntner, caroline.poyntner@uibk.ac.at.

Received 24 April 2020
Accepted 11 June 2020
Published 2 July 2020

TABLE 1 Genome data of the three chromosomes and the plasmid

Chromosome or plasmid	Size (bp)	GC content (%)	Form	No. of rRNA genes
Chromosome 1	2,486,079	60.09	Circular	3
Chromosome 2	3,638,240	62.46	Circular	9
Chromosome 3	4,573,438	62.5	Circular	9
Plasmid	142,975	59.48	Circular	

one plasmid (Table 1). GeneMarkS 4.17 (8), RepeatMasker 4.0.5 (9), and Tandem Repeats Finder (TRF) 4.07b (10) were used to predict coding genes, interspersed repetitive sequences, and tandem repeats. Further, tRNA genes were predicted using tRNAscan-SE 1.3.1 (11), rRNA genes were predicted using RNAmmer 1.2 (12), and snRNA genes were predicted using the Rfam database (13). The assembled genome contained genomic islands (IslandPath 0.2 [14]), prophage sequences (phiSpy 2.3 [15]), and CRISPRs (CRISPRdigger 1.0 [16]).

Gene functions were determined using Gene Ontology (GO) (17, 18), KEGG (19, 20), COG (21), the transporter classification database (TCDB) (22), and SWISS-PROT (23). Additional secretory proteins (SignalP 4.1 [24]), type I to VII proteins (EffectiveT3 [25]), and secondary metabolism gene clusters (antiSMASH 2.0.2 [26]) were predicted. PHI (27), VFDB (28), ARDB 1.1 (29), and CAZy (30) were applied. The results are in line with properties observed in the lab.

Data availability. The assembled genome and sequencing reads have been deposited in GenBank under the BioProject number PRJNA624061 and the accession numbers CP051514, CP051515, CP051516, and CP051517 and in the NCBI Sequence Read Archive under the numbers SRX8492130 and SRX8492131.

ACKNOWLEDGMENT

We thank P. Thurnbichler (University of Innsbruck) for skillful technical assistance.

REFERENCES

1. Lee Y, Jeon CO. 2018. Paraburkholderia aromaticivorans sp. nov., an aromatic hydrocarbon-degrading bacterium, isolated from gasoline-contaminated soil. Int J Syst Evol Microbiol 68:1251–1257. https://doi.org/10.1099/ijsem.0.002661.
2. Lee Y, Lee Y, Jeon CO. 2019. Biodegradation of naphthalene, BTEX, and aliphatic hydrocarbons by Paraburkholderia aromaticivorans BN5 isolated from petroleum-contaminated soil. Sci Rep 9:24–30. https://doi.org/10.1038/s41598-018-36165-x.
3. França L, Sannino C, Turchetti B, Buzzini P, Margesin R. 2016. Seasonal and altitudinal changes of culturable bacterial and yeast diversity in Alpine forest soils. Extremophiles 20:855–873. https://doi.org/10.1007/s00792-016-0874-2.
4. 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. https://doi.org/10.1089/cmb.2012.0021.
5. 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.
6. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
7. 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.
8. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607.
9. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. http://repeatmasker.org.
10. Benson G. 1999. Tandem Repeats Finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580. https://doi.org/10.1093/nar/27.2.573.
11. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–W57. https://doi.org/10.1093/nar/gkw413.
12. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
13. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res 46:D335–D342. https://doi.org/10.1093/nar/gkx1038.
14. Hsiao W, Wan I, Jones SJ, Brinkman F. 2003. IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics 19:418–420. https://doi.org/10.1093/bioinformatics/btg004.
15. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:347–352. https://doi.org/10.1093/nar/gkr485.
16. Ge R, Mai G, Wang P, Zhou M, Luo Y, Cai Y, Zhou F. 2016. CRISPRdigger: detecting CRISPRs with better direct repeat annotations. Sci Rep 6:32942–32910. https://doi.org/10.1038/srep32942.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556.
18. The Gene Ontology Consortium. 2019. The Gene Ontology resource: 20 years and still GOing strong. Nucleic Acids Res 47:D330–D338. https://doi.org/10.1093/nar/gky1055.
19. Kanehisa M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277–D280. https://doi.org/10.1093/nar/gkh063.
20. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical

genomics: new developments in KEGG. Nucleic Acids Res 34:D354–D357. https://doi.org/10.1093/nar/gkj102.
21. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. https://doi.org/10.1093/nar/gku1223.
22. Saier MH, Reddy VS, Tamang DG, Västermark Å. 2014. The transporter classification database. Nucleic Acids Res 42:251–258. https://doi.org/10.1093/nar/gkt1097.
23. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48. https://doi.org/10.1093/nar/28.1.45.
24. Petersen TN, Brunak S, Von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.
25. Arnold R, Brandmaier S, Kleine F, Tischler P, Heinz E, Behrens S, Niinikoski A, Mewes HW, Horn M, Rattei T. 2009. Sequence-based prediction of type III secreted proteins. PLoS Pathog 5:e1000376. https://doi.org/10.1371/journal.ppat.1000376.
26. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0: a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. https://doi.org/10.1093/nar/gkv437.
27. Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, Lindeberg M, Hansen H, Rawlings C, Hammond-Kosack KE, Köhler J. 2008. PHI-base update: additions to the pathogen host interaction database. Nucleic Acids Res 36:D572–D576. https://doi.org/10.1093/nar/gkm858.
28. Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res 40:641–645. https://doi.org/10.1093/nar/gkr989.
29. Liu B, Pop M. 2009. ARDB—antibiotic resistance genes database. Nucleic Acids Res 37:443–447. https://doi.org/10.1093/nar/gkn656.
30. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkn663.