



Draft Genome Sequence of *Cellulomonas* sp. PS-H5, Isolated from Sekinchan Beach in Selangor, Malaysia

 Nurfatini Radzlin,^{a,b}  Kheng Oon Low,^a  Kok Jun Liew,^c  Kian Mau Goh,^c  Iffah Izzati Zakaria,^a  Ummirul Mukminin Kahar^a

^aMalaysia Genome Institute, National Institutes of Biotechnology Malaysia, Kajang, Selangor, Malaysia

^bDepartment of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

^cFaculty of Science, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

Ummirul Mukminin Kahar and Iffah Izzati Zakaria contributed equally to this work. Author order was determined on the basis of seniority.

ABSTRACT *Cellulomonas* sp. PS-H5 was isolated from Sekinchan Beach in Selangor, Malaysia, using an *ex situ* cultivation method. The present work reports a high-quality draft annotated genome sequence of this strain and suggests its potential glycoside hydrolase enzymes for cellulose, hemicellulose, and starch degradations.

Cellulomonas spp. are alkali-tolerant bacteria known to be industrial enzymes producers (1, 2). *Cellulomonas* sp. PS-H5 was isolated using an *ex situ* cultivation method (3) from wet sediment and mud (upper 15-cm layer) of Sekinchan Beach in Selangor, Malaysia (3.5029N, 101.0945E). Genomic DNA was extracted from strain PS-H5 using the Monarch genomic DNA purification kit (New England BioLabs, Ipswich, MA, USA), following the manufacturer's instructions. Then, the 16S rRNA gene was amplified by PCR using the 27F and 1492R primers (4) and sequenced. Taxonomic identification was performed by comparing the PS-H5 16S rRNA gene to the sequences in the NCBI and the EzBioCloud 16S databases (5). PS-H5 was most closely related (99.51%) to *Cellulomonas pakistanensis* NCCP-11T (NCBI GenBank accession number [AB618146.1](https://doi.org/10.1128/MRA.00956-21)). Here, we report the genome sequence of *Cellulomonas* sp. PS-H5 and analyze its potential applications.

Cellulomonas sp. PS-H5 was grown on marine agar (Condalab, Madrid, Spain) at 30°C (pH 6.5) for 24 h. PS-H5 genomic DNA was extracted from a single colony of the cells using the standard protocol of the Monarch genomic DNA purification kit. A paired-end library was prepared using the NEBNext Ultra DNA library preparation kit for Illumina (New England BioLabs) according to the manufacturer's instructions. Sequencing was performed using the NovaSeq 6000 system with 150-bp paired-end reads (Illumina, San Diego, CA, USA). The raw reads were subjected to trimming using Trimmomatic v0.40 (6), *de novo* assembled using SOAPdenovo v2.04 (7), and annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.20 (8). The annotated genes were assigned functions using eggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) v5.0 (9). Using digital DNA-DNA hybridization (dDDH) via the Genome-to-Genome Distance Calculator v2.1 (10) and the average nucleotide identity (ANI) function in the EzBioCloud server (11), the PS-H5 genome was compared against all 129 available genomes of *Cellulomonas* spp. in the NCBI genome database (September 2021). Genes encoding carbohydrate-active enzymes (CAZymes) present in the genome of strain PS-H5 were detected using dbCAN2 (12). Default parameters were used for all software tools, unless stated otherwise.

The sequence data comprised 1,425,379,800 bases from 4,751,266 paired-end reads. The PS-H5 genome was assembled into 133 contigs with a coverage of 273×. The genome was 4,481,115 bp (N_{50} , 64,382 bp) long with a G+C content of 75.3%. A total of 4,117 genes were predicted, including 4,019 protein-coding sequences, 58 RNAs, and 40 pseudogenes. Of these, 97.62% (4,019 genes) were linked to the clusters of

Citation Radzlin N, Low KO, Liew KJ, Goh KM, Zakaria II, Kahar UM. 2021. Draft genome sequence of *Cellulomonas* sp. PS-H5, isolated from Sekinchan Beach in Selangor, Malaysia. *Microbiol Resour Announc* 10:e00956-21. <https://doi.org/10.1128/MRA.00956-21>.

Editor J. Cameron Thrash, University of Southern California

Copyright © 2021 Radzlin 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 Ummirul Mukminin Kahar, ummirul@nibm.my, or Iffah Izzati Zakaria, iffahizzati@nibm.my.

Received 24 September 2021

Accepted 14 October 2021

Published 28 October 2021

orthologous group functions. The genome comparison analyses indicated that PS-H5 was closely related to *Cellulomonas pakistanensis* NCCP-11^T (dDDH, 52.0%; ANI, 93.2%). As the values for the dDDH (<70%) (13) and ANI (<96%) (11) were below the corresponding thresholds, strain PS-H5 might be a new species of *Cellulomonas*. The PS-H5 genome encoded 173 CAZymes, including 104 glycoside hydrolases (GHs), 56 glycoside transferases, 8 carbohydrate esterases, and 5 auxiliary activity enzymes. Strain PS-H5 harbors 14 GHs belonging to GH families 1, 13, and 127 (i.e., β -glucosidase, β -galactosidase, endoglucanase, and α -L-arabinofuranosidase) that are important for cellulose and hemicellulose degradation. Additionally, four GH13 enzymes (two α -amylases, an α -glucosidase, and a pullulanase) involved in starch degradation were detected in the PS-H5 genome. To date, none of the pullulanases from *Cellulomonas* spp. have been characterized. Summarily, the *Cellulomonas* sp. PS-H5 genome provides various GH candidates for potential biotechnological applications.

Data availability. The whole-genome shotgun sequence of *Cellulomonas* sp. PS-H5 has been deposited in NCBI GenBank under BioProject accession number [PRJNA716128](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA716128), BioSample accession number [SAMN18395727](https://www.ncbi.nlm.nih.gov/biosample/SAMN18395727), and GenBank accession number [JAHVCI000000000](https://www.ncbi.nlm.nih.gov/genbank/JAHVCI000000000). The version described in this paper is the first version, [JAHVCI000000000.1](https://www.ncbi.nlm.nih.gov/genbank/JAHVCI000000000.1). The raw sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession number [SRR15464914](https://www.ncbi.nlm.nih.gov/sra/SRR15464914). The 16S rRNA gene sequence of *Cellulomonas* sp. PS-H5 has been deposited in NCBI GenBank under accession number [MW786713.1](https://www.ncbi.nlm.nih.gov/genbank/MW786713.1).

ACKNOWLEDGMENTS

This work was supported by Eleventh Malaysia Plan 2016–2020 grant RMKe-11 (RP4) P30006059763005, awarded to U.M.K., I.I.Z., and K.O.L. U.M.K., and I.I.Z. are thankful for Malaysia Fundamental Research Grant Scheme (FRGS) grant FRGS/1/2020/WAB11/MESTECC/02/1. K.M.G. is grateful for funding received from Malaysia FRGS grant 5F241. K.J.L. appreciates the support from Universiti Teknologi Malaysia.

REFERENCES

- Li J, Solhi L, Goddard-Borger ED, Mathieu Y, Wakarchuk WW, Withers SG, Brumer H. 2021. Four cellulose-active lytic polysaccharide monoxygenases from *Cellulomonas* species. *Biotechnol Biofuels* 14:1–19. <https://doi.org/10.1186/s13068-020-01860-3>.
- Hammerstrom RA, Claus KD, Coghlan JW, McBee RH. 1955. The constitutive nature of bacterial cellulases. *Arch Biochem Biophys* 56:123–129. [https://doi.org/10.1016/0003-9861\(55\)90341-6](https://doi.org/10.1016/0003-9861(55)90341-6).
- Radzlin N, Mohd Omar S, Liew KJ, Goh KM, Zakaria II, Kahar UM. 2021. Draft genome sequence of *Roseovarius* sp. PS-C2, isolated from Sekinchan Beach in Selangor. *Microbiol Resour Announc* 10:e00673-21. <https://doi.org/10.1128/MRA.00673-21>.
- Chen YL, Lee CC, Lin YL, Yin KM, Ho CL, Liu T. 2015. Obtaining long 16S rDNA sequences using multiple primers and its application on dioxin-containing samples. *BMC Bioinformatics* 16:S13. <https://doi.org/10.1186/1471-2105-16-S18-S13>.
- Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67:1613–1617. <https://doi.org/10.1099/ijsem.0.001755>.
- 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>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2015. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 4:30. <https://doi.org/10.1186/s13742-015-0069-2>.
- 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>.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:309–314. <https://doi.org/10.1093/nar/gky1085>.
- Auch AF, Von Jan M, Klenk H, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2:117–134. <https://doi.org/10.4056/sigs.531120>.
- Yoon SH, Ha S, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>.
- Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>.
- Meier-Kolthoff JP, Klenk HP, Göker M. 2014. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* 64:352–356. <https://doi.org/10.1099/ijms.0.056994-0>.