



Data in Brief

Genome sequencing and annotation of *Cellulomonas* sp. HZMPatric Chua ^{*}, Zi Mei Har, Christopher M. Austin, Catherine M. Yule, Gary A. Dykes, Sui Mae Lee

School of Science, Monash University, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 28 April 2015

Accepted 7 May 2015

Available online 16 May 2015

Keywords:

Tropical peat swamp

Cellulomonas

Cellulolytic

Whole genome sequencing

ABSTRACT

We report the draft genome sequence of *Cellulomonas* sp. HZM, isolated from a tropical peat swamp forest. The draft genome size is 3,559,280 bp with a G + C content of 73% and contains 3 rRNA sequences (single copies of 5S, 16S and 23S rRNA).

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications

Organism/cell line/tissue Strain(s)	<i>Cellulomonas</i> sp. HZM
Sequencer or array type	Sequencer; Illumina MiSeq
Data format	Processed
Experimental factors	Microbial strains
Experimental features	Draft genome sequence of <i>Cellulomonas</i> sp. HZM, assembly and annotation
Consent	N/A
Sample source location	Tropical peat swamp in Pekan, Pahang, Malaysia

1. Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/bioproject/236451>.

2. Experimental design, materials and methods

Cellulomonas sp. HZM is a Gram positive, rod-shaped bacterium isolated from soil surface plant detritus in Pekan tropical peat swamp in Pahang, Malaysia. The isolate was acquired by culture-plating on Sizova's cellulose minimal salt media [1] preceded by a series of enrichment steps. Genomic DNA was extracted from 24 hour old cultures using the GF⁻¹ nucleic acid extraction kit (Vivantis, Malaysia) as per manufacturer's instructions. The genome was sequenced using an Illumina MiSeq sequencer (150-bp paired-end reads) and the raw reads were trimmed and assembled *de novo* using CLC Genomics Workbench 6 (CLC Bio, Denmark). A total of 78 contigs with an accumulated length of 3,559,280 bp (74-fold coverage, N₅₀ = 99,115 bp) and a G + C content of 73% were obtained. The genome contains 52 tRNA genes and

3 rRNA genes (5S–23S–16S) as predicted using tRNAscan 1.2 [2] and RNAmmer 1.2 [3], respectively.

A total of 3180 coding sequences in 352 subsystems were functionally annotated by Rapid Annotation using the Subsystems Technology (RAST) server Fig. 1. A gene sequence putatively encoding endoglucanase (EC 3.2.1.4), as well as six copies of β-glucosidase (EC 3.2.1.21) gene sequences were identified in the starch and sucrose metabolism subsystem using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database available in RAST [4]. No sequences for exoglucanases (EC 3.2.1.91) were identified, possibly indicating an incomplete cellulolytic system. Of the hemicellulolytic enzymes, only a single copy of the sequence of alpha-L-arabinofuranosidase B (EC 3.2.1.55) and acetyl xylan esterase (EC 3.2.1.41) were identified. These enzymes remove side chains and thus have assistive roles as compared to endoxylanases (EC 3.2.1.8) and β-xylosidases (EC 3.2.1.37) which cleave glycosidic bonds between xylose residues [5]. However, no sequence for any xylan-degrading enzymes was annotated within the genome. Other gene sequences for hydrolytic enzymes identified include those for β-mannosidase (EC 3.2.1.25) and chitinase (EC 3.2.1.14). Both these enzymes may assist in providing nutrient availability for *Cellulomonas* sp. HZM.

Functional comparison of genome sequences in the RAST server revealed the closest neighbors of *Cellulomonas* sp. HZM as *Sanguibacter keddieii* DSM 10,542 (score 555) followed by *Cellulomonas flavigena* DSM 20,109 (score 336), *Beutenbergia cavernae* DSM 12,333 (score 293), *Actinomyces odontolyticus* ATCC 17,982 (score 189) and *Xylanimonas cellulositytica* DSM 15,894 (score 186). Analysis using the complete 16S rRNA sequence (performed using EzTaxon at <http://www.ezbiocloud.net/eztaxon> [6]) is more specific by comparison, with the list of candidates being predominantly *Cellulomonas*, whereby the closest match is *Cellulomonas uda* (pairwise similarity of 98.20%)

^{*} Corresponding author.

E-mail address: tcchu4@student.monash.edu (P. Chua).

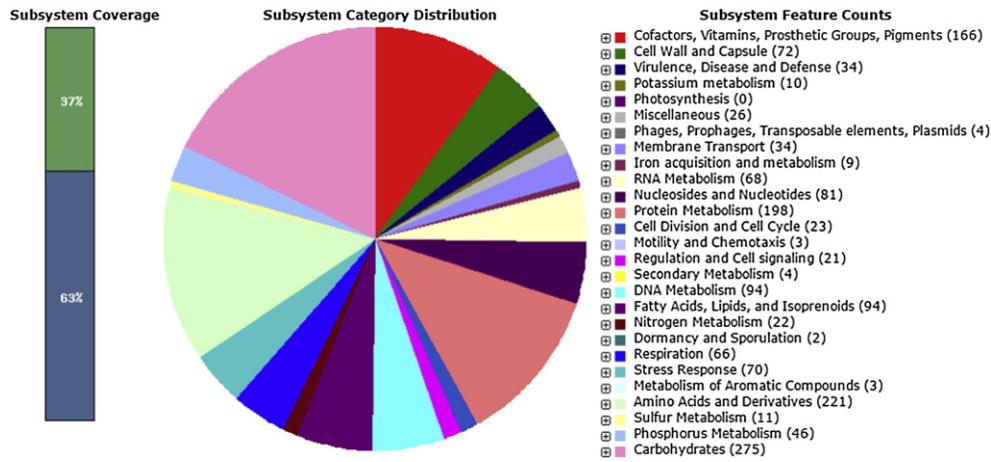


Fig. 1. Subsystem distribution of *Cellulomonas* strain HZM (based on RAST annotation server).

followed by *Cellulomonas chitinilytica* (98.15%), *Cellulomonas gelida* (98.06%) and *Cellulomonas iranensis* (97.89%).

3. Nucleotide sequence accession number

The whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. JEOE00000000.

Conflict of interest

The authors declare that there is no conflict of interests with respect to the work published in this paper.

Acknowledgments

Funding for this study was provided by the School of Science, Monash University Malaysia. We are grateful to the Pahang State Forestry Department for their assistance in procuring the samples.

References

- [1] M.V. Sizova, J.A. Izquierdo, N.S. Panikov, L.R. Lynd, Cellulose- and xylan-degrading thermophilic anaerobic bacteria from biocompost. *Appl. Environ. Microbiol.* 77 (7) (2011) 2282–2291.
- [2] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25 (1997) 955–964.
- [3] K. Lagesen, P. Hallin, E.A. Rodland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35 (2007) 3100–3108.
- [4] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9 (2008) 75.
- [5] V. Juturu, J.C. Wu, Microbial xylanases: engineering, production and industrial applications. *Biotechnol. Adv.* 30 (2012) 1219–1227.
- [6] O.S. Kim, Y.J. Cho, K. Lee, S.H. Yoon, M. Kim, H. Na, S.C. Park, Y.S. Jeon, J.H. Lee, H. Yi, S. Won, J. Chun, Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* 62 (2012) 716–721.