

Draft Genome Sequence of Cellulolytic and Xylanolytic *Cellulomonas* sp. Strain B6 Isolated from Subtropical Forest Soil

Florencia Piccinni,^{a,d} Yanina Murua,^{a,c,*} Silvina Ghio,^b Paola Talia,^{a,d} Máximo Rivarola,^{a,c,d} Eleonora Campos^{a,d}

Instituto de Biotecnología, CICVyA, INTA, Buenos Aires, Argentina^a; Instituto de Suelos, CIRN, INTA, Buenos Aires, Argentina^b; Facultad de Ingeniería y Ciencias, Exactas, UADE, Buenos Aires, Argentina^c; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina^d

* Present address: Yanina Murua, Fundación Instituto Leloir (FIL), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina.

***Cellulomonas* sp. strain B6 was isolated from a subtropical forest soil sample and presented (hemi)cellulose-degrading activity. We report here its draft genome sequence, with an estimated genome size of 4 Mb, a G+C content of 75.1%, and 3,443 predicted protein-coding sequences, 92 of which are glycosyl hydrolases involved in polysaccharide degradation.**

Received 29 June 2016 Accepted 1 July 2016 Published 25 August 2016

Citation Piccinni F, Murua Y, Ghio S, Talia P, Rivarola M, Campos E. 2016. Draft genome sequence of cellulolytic and xylanolytic *Cellulomonas* sp. strain B6 isolated from subtropical forest soil. *Genome Announc* 4(4):e00891-16. doi:10.1128/genomeA.00891-16.

Copyright © 2016 Piccinni et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](http://creativecommons.org/licenses/by/4.0/).

Address correspondence to Eleonora Campos, campos.eleonora@inta.gob.ar.

Cellulases and xylanases are widely used in textile, animal feed, food, and paper industries. They also play a key role in the production of cellulosic ethanol (1).

Cellulomonas sp. strain B6 (available from Argentine collection of microorganisms as IMIZA:CEB6) was isolated from the first 10-cm layer of a preserved native subtropical forest soil sample (26°01'34''S 54°26'59''W) (2). It is a Gram-positive, rod-shaped, aerobic isolate that can grow on lignocellulosic biomass, such as sugarcane residue, as a sole carbon source. Its secreted protein extract presented cellulose- and xylan-degrading activities (our unpublished data). Based on 16S rRNA analysis, strain B6 formed a cluster with *Cellulomonas flavigena* (accession no. AF140036.1) and *Cellulomonas persica* (accession no. NR_024913.1). The genomes of *Cellulomonas flavigena* DSM 20109 and B6 show an average nucleotide Identity (ANI) value of 81.99%, suggesting that they are different species.

Genomic DNA of *Cellulomonas* sp. strain B6 was extracted from a 24-h culture in LB broth by a commercial extraction kit (Wizard genomic DNA extraction kit; Promega) and sequenced using the Illumina MiSeq platform. The data comprised 1,532,556 paired-end reads of 500 bp, resulting in 83-fold genome coverage. The raw reads were subjected to trimming using Trimmomatic version 0.33 (3) and assembled *de novo* using Celera Assembler version 8.2 (4), followed by the SPAdes genome assembler version 3.5.0 (5), generating 279 contigs, with a total length of 4,042,435 bp (N_{50} , 24,612 bp) and a G+C content of 75.1%, consistent with the genus.

Gene prediction and functional analysis were carried out using the Rapid Annotations using Subsystems Technology (RAST) server version 2.0 (6) and the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Using the NCBI pipeline, 3,691 genes, including 3,443 protein-coding sequences, 50 tRNA, and a set of full-length 5S, 23S, and 16S rRNA gene sequences, were predicted. A noncoding RNA (ncRNA) of an RNase P (ATM99_11600) was also predicted. Similar results were obtained by RAST. A comparison of a represen-

tative set of *FigFam* protein-coding genes from *Cellulomonas* sp. B6 to other bacterial sequences available in RAST identified *Cellulomonas flavigena* DSM 20109 (score, 413) and *Sanguibacter keddieii* DSM 10542 as the closest neighbors.

Utilizing all functional annotations from CAZy (<http://www.cazy.org/>) (7) and dbCAN (<http://csbl.bmb.uga.edu/dbCAN/>) (8), 92 sequences encoding potential glycosyl hydrolases (GH) were identified, including six endo- β -1,4-glucanases (two GH5 and four GH9), two exo-glucanases (GH6 and GH48), 11 β -glucosidases (three GH1 and eight GH3), 10 endo-1,4- β -xylanases (eight GH10, one GH11, and one GH43), two β -xylosidases (two GH39 and one GH43:1), four α -L-arabinofuranosidases (two GH43 and two GH51), two endo-1,5- α -arabinosidases (GH43), and an α -glucuronidase (GH67). These results are consistent with the cellulolytic and xylanolytic activities of this bacterial isolate.

The genome information will be useful for studies of microbial enzymes for industrial application in lignocellulosic biomass utilization.

Accession number(s). This whole-genome shotgun project has been deposited at NCBI SRA database under the accession no. [LNTD00000000](https://www.ncbi.nlm.nih.gov/sra/LNTD00000000). The version described in this paper is version [LNTD01000000](https://www.ncbi.nlm.nih.gov/sra/LNTD01000000).

ACKNOWLEDGMENTS

F.P. is a Ph.D. student of the Department of Biological Chemistry (QB) of the School of Natural and Exact Sciences (FCEN) of the University of Buenos Aires (UBA) and has a doctoral fellowship from the Argentine National Council of Research (CONICET).

M.R., P.T., and E.C. are members of the Scientific Research Career of CONICET.

Sequencing services were performed at INTA, Consorcio Argentino de Tecnología Genómica (CATG) (PPL Genómica, MINCYT), and this work used computational resources from the Bioinformatics Unit, Instituto de Biotecnología, CICVyA, INTA.

FUNDING INFORMATION

This work was funded by Instituto Nacional de Tecnología Agropecuaria (INTA) (project PNAIyAV 1130034). This work, including the efforts of Eleonora Campos, was funded by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (PIP 11420110100124). This work, including the efforts of Florencia Elizabeth Piccinni, Yanina Alejandra Murua, Silvina Ghio, Paola Monica Talia, Maximo Rivarola, and Eleonora Campos, was funded by Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCyT) (PICT2011-2735).

REFERENCES

- Lennartsson PR, Erlandsson P, Taherzadeh MJ. 2014. Integration of the first and second generation bioethanol processes and the importance of by-products. *Bioresour Technol* 165:3–8. <http://dx.doi.org/10.1016/j.biortech.2014.01.127>.
- Campos E, Negro Alvarez MJ, Sabaris Di Lorenzo G, Gonzalez S, Rorig M, Talia P, Grasso DH, Saéz F, Manzanares Secades P, Ballesteros Perdices M, Cataldi AA. 2014. Purification and characterization of a GH43 beta-xylosidase from *Enterobacter* sp. identified and cloned from forest soil bacteria. *Microbiol Res* 169:213–220. <http://dx.doi.org/10.1016/j.micres.2013.06.004>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
- Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ. 2000. A whole-genome assembly of *Drosophila*. *Science* 287:2196–2204. <http://dx.doi.org/10.1126/science.287.5461.2196>.
- 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>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D95. <http://dx.doi.org/10.1093/nar/gkt1178>.
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <http://dx.doi.org/10.1093/nar/gks479>.