DR. ELEONORA CAMPOS (Orcid ID : 0000-0002-0887-6627)

Article type : -Original Article

TITLE

Secretome profile of *Cellulomonas* sp. B6 growing on lignocellulosic substrates

Florencia Elizabeth Piccinni[a], Ornella Mailén Ontañon[a], Silvina Ghio[a], Diego Herman Sauka[a,b], Paola Mónica Talia[a], Máximo Lisandro Rivarola[a], María Pia Valacco[c], Eleonora Campos[a]

a) Instituto de Agrobiotecnología y Biología Molecular (IABIMO), Instituto Nacional de Tecnología Agropecuaria (INTA), Consejo Nacional de investigaciones Científicas y Tecnológicas (CONICET). Los Reseros y Nicolas Repetto (1686), Hurlingham, Buenos Aires, Argentina.

b) Instituto de Microbiología y Zoología Agrícola, Centro de Investigación en Ciencias Veterinarias y Agronómicas (CICVyA), Instituto Nacional de Tecnología Agropecuaria (INTA). Los Reseros y Nicolas Repetto (1686), Hurlingham, Buenos Aires, Argentina.

c) Centro de Estudios Químicos y Biológicos por Espectrometría de Masa (CEQUIBIEM). Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - Ciudad Universitaria (C1428EGA), Ciudad Autónoma de Buenos Aires, Argentina.

CORRESPONDING AUTHOR

Dr. Eleonora Campos

Mailing address: campos.eleonora@inta.gob.ar / eleonoracampos@gmail.com

RUNNING HEAD

Cellulomonas sp. B6 CAZymes

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jam.14176

FUNDING

This work was supported by the Secretary of Science and Technology of Argentina by means of grants PICT2014-0791 and PICT2016-4695.

ABSTRACT

Aims

Lignocellulosic biomass deconstruction is a bottleneck for obtaining biofuels and value-added products. Our main goal was to characterize the secretome of a novel isolate, *Cellulomonas* sp. B6, when grown on residual biomass for the formulation of cost-efficient enzymatic cocktails.

Methods and Results

We identified 205 potential CAZymes in the genome of *Cellulomonas* sp. B6, 91 of which were glycoside hydrolases (GH). By secretome analysis of supernatants from cultures in either extruded wheat straw (EWS), grinded sugar cane straw (SCR) or carboximethylcellulose (CMC), we identified which proteins played a role in lignocellulose deconstruction. Growth on CMC resulted in the secretion of two exo-glucanases (GH6 and GH48) and two GH10 xylanases, while growth on SCR or EWS resulted in the identification of a diversity of CAZymes. From the 32 GHs predicted to be secreted, 22 were identified in supernatants from EWS and/or SCR cultures, including endo- and exo-glucanases, xylanases, a xyloglucanase, an arabinofuranosidase/ β-xylosidase, a β-glucosidase and an AA10. Surprisingly, among the xylanases, seven were GH10.

Conclusions

Growth of *Cellulomonas* sp. B6 on lignocellulosic biomass induced the secretion of a diverse repertoire of CAZymes.

Significance and Impact of the Study

Cellulomonas sp. B6 could serve as a source of lignocellulose degrading enzymes applicable to bioprocessing and biotechnological industries.

KEYWORDS

CELLULOMONAS; SECRETOME; CAZYMES; GLYCOSYL HYDROLASES; XYLANASES; GLUCANASES; GH10; LIGNOCELLULOSE.

INTRODUCTION

Lignocellulosic biomass, the non-edible part of crops, pastures and trees, is an abundant renewable resource for obtaining biofuels and value-added products. It is formed by the structural polysaccharides cellulose and hemicelluloses, which are the major components of plant cell walls (35 to 50% and 20 to 35% of plant dry weight, respectively), as well as lignin. Cellulose is a large linear polymer of β-1,4-linked glucose units. It is organized in fibers that are stabilized by hydrogen bonds among chains, therefore forming a crystalline, non-soluble, structure, with only some amorphous regions (Lynd *et al.* 2002; Mischnick and Momcilovic 2010). Cellulose is coated in a complex matrix of lignin and hemicellulose. Chemical bonds among all components make plant cell wall a highly recalcitrant material. While lignin is made of organic alcohols, hemicelluloses are carbohydrates. Xylan, the most abundant hemicellulose, consists of β-1,4- linked xylose units and, contrary to cellulose, is a short and heavily substituted polymer (Pauly *et al.* 2013; Zhao *et al.* 2015). Cellulose is converted to cellobiose and cello-oligomers by endo-β-1,4-glucanases (EC 3.2.1.8) and both reducing and non-reducing end-acting cellobiohydrolases (EC 3.2.1.176 and EC 3.2.1.91, respectively). Also, β-glucosidase (EC 3.2.1.21) activity is required for the final conversion to glucose (Lynd *et al.* 2002).

The deconstruction of xylan occurs in a similar fashion and it involves the concerted action of several enzymes, primarily endo-β-1,4-xylanases (EC 3.2.1.8) that release xylooligomers (XOS) from the xylan backbone and β-1,4-xylosidases (3.2.1.37) to further hydrolyze oligomers into xylose. Debranching enzymes, such as arabinofuranosidases (EC 3.2.1.55), arabinases (3.2.1.99), glucuronidases (3.2.1.131/139) and feruloyl esterases (EC 3.2.1.73) are also required due to the many substituents typically found in xylan (Biely et al., 2016). It is worth noting that the breakdown of cellulose and hemicellulose can also occur in an oxidative way by lytic polysaccharide mono-oxygenases (LPMOs),

which add to the overall activity of glycoside hydrolases (Hemsworth et al., 2016; Kruer-Zerhusen et al., 2017; Loose et al., 2016). All these enzymes, known as CAZymes (carbohydrate active enzymes), degrade, modify, or create glycosidic bonds and are classified according to their structure in the CAZy database (http://www.cazy.org) (Lombard *et al.* 2014).

In nature, cellulose and hemicellulose are mostly hydrolyzed by fungi and bacteria which are important in the recycling of carbon from dead matter in soil (Bomble *et al.* 2017). Their lignocellulolytic systems are studied for potential applications in food, paper-pulp, animal feed, laundry detergents and second generation bioethanol production (Bhat 2000; Bhattacharya *et al.* 2015; Lynd *et al.* 2017; Toushik *et al.* 2017). Anaerobic cellulolytic bacteria produce a membrane bound complex of enzymes, called the cellulosome, first described in *Clostridium thermocellum* (Lamed *et al.* 1983; Bayer *et al.* 1998). The strategy for (hemi)cellulose conversion is different in aerobic microorganisms and consists in the secretion of soluble enzymes that generate mainly large oligosaccharides from polysaccharide chains, followed by internalization of small sugars (Shallom and Shoham 2003). Many aerobic cellulolytic bacteria belong to the phylum *Actinobacteria*, order *Actinomycetales* whereas most anaerobic bacteria belong to the phylum *Firmicutes*, order *Clostridiales*. Among the *Actinobacteria*, members of the genus *Cellulomonas* are the only known and reported cellulolytic facultative anaerobes (Christopherson *et al.* 2013).

The *Cellulomonas* genus comprises Gram positive, non sporulating, coryneform actinobacteria that have the ability to degrade cellulose and other polysaccharides such as xylan (Stackebrandt and Schumann 2015). The most characterized members of the genus are *Cellulomonas fimi* and *Cellulomonas flavigena*. Their (hemi)cellulolytic system was studied by growth in defined substrates, such as xylan or cellulose, revealing a complex array of enzymatic activities involved in polysaccharides utilization (Ponce-Noyola and De La Torre 2001; Amaya-Delgado *et al.* 2006). Some individual glycoside hydrolases of both species have also been characterized, confirming their activity (Din *et al.* 1990; Mayorga-Reyes *et al.* 2002; Santiago-Hernández *et al.* 2007; Gao and Wakarchuk 2014; Lisov *et al.* 2017; Yang *et al.* 2015).

Since their first isolation, metagenomic and culture prospection of diverse cellulolytic habitats revealed novel *Cellulomonas* isolates, which suggested the important role of this genus in the degradation of structural polysaccharides (Elberson *et al.* 2000; Rusznyák *et al.* 2011; Hatayama *et al.* 2013; Zhang *et al.* 2013; Zhuang *et al.* 2015). From a subtropical forest soil cellulolytic consortium, we isolated a *Cellulomonas* sp. novel strain (Campos *et al.* 2014). This isolate, named *Cellulomonas* sp. B6, had the capacity to grow on minimal media with cellulose or biomass as sole carbon sources. The main goal of this study was to identify the (hemi)cellulose degrading enzymes secreted by *Cellulomonas* sp. B6 when grown on residual inexpensive lignocellulosic substrates in order to develop a cost-efficient enzymatic cocktail and also contribute to understand the core mechanism of aerobic degradation of structural polysaccharides (cellulose and hemicellulose) of this important genus.

MATERIAL AND METHODS

Organism

Cellulomonas sp. strain B6 was isolated from a bacterial consortium obtained from a preserved native subtropical forest soil sample (NCBI BioSample: SAMN04287673) (Campos *et al.* 2014). The isolation was achieved by culture in minimal saline medium (MM) (K_2HPO_4 1.67 g l^{-1}, KH_2PO_4 0.87 g l^{-1}, NaCl 0.05 g l^{-1}, $MgSO_4 \cdot 7H_2O$ 0.1 g l^{-1}, $CaCl_2$ 0.04 g l^{-1}, FeCl 0.004 g l^{-1}, $NaMoO4 \cdot 2H_2O$ 0.005 g l^{-1}, biotin 0.01 g l^{-1}, nicotinic acid 0.02 g l^{-1}, pantothenic acid 0.01 g l^{-1}, NH_4Cl 1.0 g l^{-1}) supplemented with 0.5% CMC (low viscosity Sigma C5678) with filter paper as sole carbon source and successive sub-cultures in solid minimal medium MM- 15 g l^{-1} agar (MM-CMC) (Ghio *et al.* 2012). For cellulase activity monitoring, 0.1 g l^{-1} Trypan Blue was added to the MM-CMC agar plates. For long term preservation, individual clones were grown in LB supplemented with 5 g l^{-1} CMC, allowed to grow overnight at 30°C and stored at −80°C with 20% glycerol.

Phylogenetic analysis and strain characterization

A phylogenetic tree was constructed using Neighbour Joining method (Saitou and Nei 1987). Statistical significances were assessed using 2000 bootstrap replications (Felsenstein 1985) with MEGA7 (Kumar *et al.* 2016). For this analysis, we used with the full 16S rRNA coding sequence from the genome of *Cellulomonas* sp. B6 (ATM99_00120) (Piccinni et al., 2016) (https://www.ncbi.nlm.nih.gov/genome/13845?genome_assembly_id=259466) and 16S rRNA sequences of all *Cellulomonas* sp. type strains available at the Ribosomal Database Project (https://rdp.cme.msu.edu/) (Cole *et al.* 2014). The tree was rooted using *Micrococcus luteus* as outgroup. Similar results were obtained using Maximum Likelihood (Tamura and Nei 1993) with 2000 bootstrap repetitions.

For *in-silico* estimation of DNA-DNA Hybridization, the genomic sequences for *Cellulomonas flavigena* ATCC 482 (GCA_000092865.1) and *Cellulomonas persica* ATCC 700642^T (GCA_001315245.1) were obtained from the NCBI Assembly database (https://www.ncbi.nlm.nih.gov/assembly) (Agarwala *et al.* 2017). Estimation of DNA-DNA hybridization (Meier-Kolthoff *et al.* 2013) with *Cellulomonas* sp. B6 genomic sequence (GCA_001462455.1) was performed using the online Genome to Genome Distance Calculator tool (http://ggdc.dsmz.de/home.php) from the Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures. Average Nucleotide Identity and Average Aminoacid Identity were calculated with the online tools (http://enve-omics.ce.gatech.edu/) developed by the Environmental Microbial Genomics Laboratory (Georgia Institute of Technology) (Goris *et al.* 2007).

Biochemical characteristics of *Cellulomonas* sp. B6 were determined using API 50CH galleries (bioMérieux), according to the manufacturer's indications.

Bacterial cultures for activity assays

Cellulomonas sp. B6 was grown on minimal medium (MM) with 0.1% yeast extract (YE) as a nitrogen source or on Lysogeny Broth (LB) (Bertani 1951), supplemented with 5 g l^{-1} CMC (low viscosity, Sigma C5678), 10 g l^{-1} grinded sugar cane straw (SCR) or 10 g l^{-1} wheat straw pre-treated by extrusion (EWS). The culture was grown at 30°C in an orbital shaker (Innova 42, NewBrunswick) with agitation (200 rpm) and harvested by centrifugation at 5000 g for 20 minutes at 4°C, at 72 h, which corresponded to late exponential phase in all cases. Cell free supernatants from the cultures grown on SCR or EWS were filtered through #32 glass fiber (1.2 µm) filter paper (Schleicher and Schnell) to remove traces of biomass. Sodium azide (NaN_3) at 0.04% was added to supernatants for preservation purposes. Bacterial pellet was stored at -20°C and then suspended in 0.1 mol l^{-1} phosphate buffer pH7 to analyze intracellular activity.

Measurement of enzymatic activity

Endoglucanase (CMCase) and xylanase activities in the cell free supernatants (extracellular enzymatic extracts) were tested using 10 g l^{-1} CMC and 5 g l^{-1} beechwood xylan (SIGMA) as substrates, respectively. The enzymatic extracts were incubated with the substrates for 30 min with agitation (400 rpm), after which the resulting reducing sugars were detected by addition of dinitrosalicylic acid (DNS) to the reaction mixture according to already described protocols (Ghose and Bisaria 1987; Ghio *et al.* 2016). Standard curves using glucose or xylose were used to determine the reducing sugars equivalents based on the absorbance at 540 nm. In all cases, controls of "substrate without enzyme" and "enzyme without substrate" were included and the reducing sugars detected were subtracted prior to further calculations. Activity was expressed as International Units (IU) per ml of enzymatic extract. One IU was defined as the amount of enzyme necessary to release 1 µmol of glucose or xylose equivalents per min, under the assay conditions. The effect of pH and temperature on CMCase and xylanase activity profiles was studied using 0.3 mol l^{-1} citrate/phosphate buffer, pH from 4.5 to 8 at 50°C, and at pH 7, with temperatures from 35°C to

60°C or 75°C (for CMCase or xylanase activity, respectively). The endoglucanase and xylanase stability of the cell free supernatants was tested by incubation at 45 and 50°C. Samples were collected at 0, 1, 3, 5, 24 and 48 h and the remaining activity was detected as described above.

β-glucosidase, cellobiohydrolase and β-xylosidase activities in the intracellular and extracellular extracts were assayed using p-nitrophenyl β-D-glucopyranoside (pNPG) 5 mmol l^{-1}, p-nitrophenyl β-D-cellobioside (pNPC) 5 mmol l^{-1} and p-nitrophenyl-β-D-xylopyranoside (pNPX) 4 mmol l^{-1} (Sigma-Aldrich) as substrates. Absorbance at 410 nm was measured and activity was determined based on p-nitrophenol (pNP) standard curve (Ghose and Bisaria, 1987). One international unit was defined as the amount of enzyme that released 1 µmol of pNP per min under the assay conditions. All experiments were conducted in triplicates. Protein concentration was determined using the commercial Micro BCA Protein Assay Kit (ThermoFischer). Significant differences of enzymatic activity assays were determined by analysis of variance (ANOVA) followed by Tukey's test ($P<0.05$) using GraphPad Prism software v6.0 (GraphPad, San Diego, CA, USA), with "growth of *Cellulomonas* sp. B6 on different substrates does not affect xylanase or CMCase activity of the cell free culture supernatant" as the null hypothesis.

Detection of enzymatic reaction products by Thin Layer Chromatography (TLC)

Xylanase and CMCase reaction products were detected by TLC using a 2:1:1 butanol: acetic acid: water mix and silica as mobile and stationary phases, respectively. Xylooligosaccharides (X2 to X6) and xylose 0.25 mg/ml each were used as standards for xylanase activity. Similarly, cellooligosaccharides (G2 to G4) and glucose 0.25 mg/ml were used as standards for CMCase activity. Oligo and monosaccharides were purchased from Sigma and Megazyme (https://www.megazyme.com/). A 3:70:20 sulphuric acid: ethanol: water solution with 1% orcinol was used for product visualization.

Biomasses compositional analysis and pretreatment

EWS straw (pre-treated by extrusion at 70ºC, NaOH 6%, 2mm) was gently provided by Dr. Mercedes Ballesteros and Dr. Paloma Manzanares Secades from CIEMAT (Madrid, Spain). SCR was gently provided by Dr. Alejandro Valeiro from INTA EEA Famaillá (Tucumán, Argentina). Pre-treated biomass composition was determined at CIEMAT following NREL specifications. EWS: 46.8% cellulose, 30.0 % hemicellulose, 19.8% acid insoluble lignin, 1.3% acid soluble lignin, 2.1 % ashes and 0.1% acetic acid. SCR: 32.6% cellulose, 23.9 % hemicellulose, 14.9% acid insoluble lignin, 5.1% acid soluble lignin, 7.7 % ashes and 15.9% extractives.

Prediction of secreted CAZymes

The predicted CAZome of *Cellulomonas* sp. B6 was constructed using dbCAN algorithm (http://csbl.bmb.uga.edu/dbCAN/), which takes into account unique signatures and HMM models of CAZYmes (Yin et al., 2012) in order to identify potential CAZymes in the genomic sequence of *Cellulomonas* sp. B6 (NCBI accession number LNTD00000000.1). Full genome annotation was performed with RAST and NCBI pipelines and specific CAZY annotation with dbCAN and confirmation of sequence homology using BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) (Altschul *et al.* 1990). SignalP v4.1 (http://www.cbs.dtu.dk/services/SignalP/) for Gram positive bacteria was used for prediction of signal peptides (Käll *et al.* 2004; Petersen *et al*. 2011) and SecretomeP 2.0 (http://www.cbs.dtu.dk/services/SecretomeP/) was used for prediction of extracellular proteins secreted by non-classical pathways (Dyrløv Bendtsen *et al*. 2005). Comparisons of secreted CAZymes of *Cellulomonas* sp. B6 and *Cellulomonas fimi* ATCC 484 (Spertino et al., 2018) was carried out by Blastp sequence alignment. An identity ≥ 90%, with a query cover ≥70%, was categorized as high.

Secretome analysis by mass spectrometry

Bacteria were grown on MM supplemented with 0.5% sucrose, 1% CMC, 1% SCR or 1% EWS as described before. Proteins from cell free supernatants were quantified by Bradford assay (Promega), then precipitated with 10% trichloroacetic acid (TCA) and resuspended in water (18 Ω) to 1 mg/ml total protein. Supernatants from cultures grown on sucrose were used as a negative control for CAZYmes secretion. Sucrose was used instead of glucose to avoid interference with DNS activity assay. Protein digestion and Mass Spectrometry analysis were performed at CEQUIBIEM (http://cequibiem.qb.fcen.uba.ar/). Proteins were reduced with dithiotreitol 10 mM for 45 min at 56°C, alkylated with iodoacetamide (55 mmol l^{-1}) for 45 min in the dark and digested with trypsin (Promega V5111) overnight at 37°C. The digests were analyzed by nano LC-MS/MS in a Thermo Scientific Q-Exactive Mass Spectrometer coupled to a nano HPLC EASY-nLC 1000 (Thermo Scientific). For the LC-MS/MS analysis, approximately 1 µg of peptides was loaded onto the column and eluted for 120 min using a reverse phase column (C18, 2 µm, 100A, 50 µm x 150 mm) Easy-Spray Column PepMap RSLC (P/N ES801) suitable for separating complex peptide mixtures with a high degree of resolution. The flow rate used for the nano column was 300 nl min^{-1} and the solvent range from 7% B (5 min) to 35% (120 min). Solvent A was 0.1% formic acid in water whereas B was 0.1% formic acid in acetonitrile. The injection volume was 2µL. A voltage of 3,5 kV was used for Electro Spray Ionization (Thermo Scientific, EASY-SPRAY). The MS equipment has a high collision dissociation cell (HCD) for fragmentation and an Orbitrap analyzer (Thermo Scientific, Q-Exactive). XCalibur 3.0.63 (Thermo Scientific) software was used for data acquisition and equipment configuration to allow peptide identification at the same time of their chromatographic separation. Full-scan mass spectra were acquired in the Orbitrap analyser. The scanned mass range was 400-2000 m/z, at a resolution of 70000 at 400 m/z and the 12 most intense ions in each cycle, were sequentially isolated, fragmented by HCD and measured in the Orbitrap analyser. Peptides with a charge of +1 or with unassigned charge state were excluded from fragmentation for MS2.

Q Exactive raw data were processed using Proteome Discoverer software (version 2.1.1.21 Thermo Scientific) and searched against *Cellulomonas* sp. B6 protein sequences database (from the genome deposited at NCBI SRA database under the accession LNTD00000000) with trypsin specificity and a maximum of one missed cleavage per peptide. Proteome Discoverer searches were performed with a precursor mass tolerance of 10 ppm and product ion tolerance to 0.05 Da. Static modifications were set to carbamidomethylation of Cys, and dynamic modifications were set to oxidation of Met and N-terminal acetylation.

Protein hits were filtered for high confidence peptide matches with a maximum protein and peptide false discovery rate of 1% calculated by employing a reverse database strategy.

For the estimation of relative abundance, we used the protein abundance index emPAI calculated by Sequest using protein identification data. The equation emPAI/Σ (emPAI) x 100 was used to calculate the protein content in mol % (emPAI%) (Ishihama *et al* 2005, Shinoda *et al*. 2010).

RESULTS

Genomic and phenotypic characterization of *Cellulomonas* sp. B6

In a previous study, we had isolated the strain B6 from a native sub-tropical forest soil sample in Argentina and subsequently sequenced its genome (Piccinni *et al.* 2016). This isolate was classified as a member of the *Cellulomonas* genus by 16S rRNA sequence analysis, closely related to type strains *C. flavigena* ATCC 482 and *C. persica* ATCC 700642^T (Figure 1). *Cellulomonas* sp. B6 shared 97.73% 16S rRNA sequence identity with *C. flavigena* ATCC 482 and 98.54% with *C. persica* which made difficult to establish whether it was a novel species. Regarding biochemical assays, we run an API 50 CH test which resulted in differences between the results obtained for *Cellulomonas* sp. B6 and what had been reported for *C. flavigena* (Ahmed *et al*. 2014) and *C. persica* (Elberson *et al*. 2000) (Table S1). Genome comparisons between *Cellulomonas* sp. B6 and *C. flavigena* ATCC 482 resulted in an *in-silico* 25.40% DNA-DNA hybridization (DDH) and 83.71% Average Nucleotide Identity (ANI). It has been reported that to consider genomes of be from the same species DDH

values should be above 70% and ANI values above 95% (Goris *et al*. 2007; Auch *et al*. 2010; Thompson *et al*. 2013) (Table S2). Average Aminoacid Identity (AAI) between *C. flavigena* and *Cellulomonas* sp. B6 was 79.14%, which further supported these differences. *In-silico* DDH with *C. persica* was also low (29.9%), but ANI and AAI could not be calculated, as the complete genome and protein sequences from this strain were not available. Therefore *Cellulomonas* sp. B6 could not be assigned to either *C. flavigena* or *C. persica*. Additional assays are being conducted in order to describe and name the new species. We also conducted the analysis with *C. fimi* type strain, as it is the most studied strain of the *Cellulomonas* genus. Percentage of 16S rRNA similarity, DDH, ANI and AAI values between *Cellulomonas* sp. B6 vs *C. fimi* were 95.07%, 21.70%, 80.95% and 70.36% respectively (Table S2).

Enzymatic activities of *Cellulomonas* sp. B6

Cellulomonas sp. B6, was able to grow on LB and minimal media (MM) supplemented with carboxymethylcelulose (CMC) and grinded sugarcane straw (SCR), achieving highest CMCase and xylanase activities at 72 h in the supernatant from MM-SCR cultures (figure S1). All cultures were in late log phase and cell density were 9.66 ± 0.19; 9.77 ± 0.23 and 10.27 ± 0.08 $\log_{10}$ ml^{-1} for CMC, SCR and EWS respectively. Optimal enzymatic activities, from MM-SRC cultures were obtained at 50°C, pH 7 for CMCase (0.16 ± 0.01 U ml^{-1}) and at 60°C, pH 6 for xylanase (0.41 ± 0.01 U ml^{-1})**.** In addition, when the supernatants were incubated at 50°C, over 60% of xylanase activity was preserved for up to 48 h while CMCase activity decreased more than 80% at 24 h (Figure 2). This could suggest that xylan-active enzymes secreted by *Cellulomonas* sp. B6 when grown on MM-SCR could be more tolerant to higher temperatures than enzymes active on CMC. Some cellobiohydrolase activity (pNP-cellobioside) was also observed in the extracellular extract (2.09 ± 0.12 mU ml^{-1}) while β-glucosidase and β-xylosidase activities were detected mainly in the intracellular fraction (28.58 ± 6.05 and 277.65 ± 0.18 mU ml^{-1} on pNPG and pNPX, respectively). Very low extracellular β-xylosidase activity (9.14 ± 0.28 mU ml^{-1}) and β-glucosidase activity (0.58 ± 0.14 mU ml^{-1}) were observed. Consistent

with these results, the main reaction products of CMCase and xylanase activities were xylo- and cello-oligosaccharides, while no glucose and only traces of xylose (X1) were detected (Figure 2).

***In silico* prediction of extracellular CAZome**

In the interest of assigning the observed activities to specific enzymes, we studied the CAZome (all potential CAZymes encoded in the genome) of *Cellulomonas* sp. B6, using the dbCAN algorithm and the CAZy database (Yin *et al*. 2012). We identified a total of 205 CAZymes, from which 91 (48%) were glycoside hydrolases (GH), 57 (28%) were glycosyl transferases, 34 (17%) were carbohydrate esterases (CE), 12 (6%) were proteins with only carbohydrate-binding modules (and not known catalytic domain) (CBM) and 10 (5%) were proteins with auxiliary activities (AA), while only one polysaccharide lyase (PL) was identified (Table S3). From the 205 CAZYmes, 48 were predicted to be secreted by classical signal peptide pathway (Petersen *et al.* 2011) and additional 28 proteins were predicted to be extracellular by non-classical secretion pathways (independent of signal peptide sequence) (Dyrløv Bendtsen *et al*. 2005). Interestingly, 12 proteins with predicted signal peptide were considered below the threshold for secretion by SecretomeP server (Table S3).

Enzymes with predicted GH activity on structural polysaccharides, as well as LPMOs, were analyzed in more detail (Figure 3, Table 1). Regarding hemicelluloses deconstruction, we identified 9 xylanases (EC 3.2.1.8), of which only one was classified as GH11 and the rest were GH10. From these, one GH10 (KSW14744.1) sequence was incomplete in the aminoterminal portion so no signal peptide was identified, but was predicted as extracellular by SecretomeP server. Another GH10 (KSW28941.1) was predicted to be intracellular (by both analysis), which may indicate a different role from xylan hydrolysis. We identified two GH39 and two GH43 β-xylosidases (EC 3.2.1.37) which were predicted to be intracellular, that could hypothetically be involved in the final degradation of xylooligosaccharides (Shallom and Shoham 2003). Two other GH43 had signal peptide and another one could be secreted by non-classical pathways. Additionally, a GH74 xyloglucanase (EC 3.2.1.151), a GH18 chitinase (EC 3.2.1.14) and a GH5 β-mannosidase (EC 3.2.1.25) were also predicted to be

secreted. Noteworthy, the GH74 xyloglucanase presented four catalytic modules, and was the only protein with multiple catalytic domains identified in this study.

Regarding the breakdown of cellulose to oligosaccharides, we identified only two extracellular exo-glucanases, GH6 and GH48 (non-reducing EC 3.2.1.91 and reducing end EC 3.2.1.176, respectively), and six extracellular endo-glucanases (EC 3.2.1.4) (two GH6 and four GH9). Furthermore, we identified two potentially secreted monooxygenases of family AA10 which could have a role on cellulose deconstruction by oxidation. Finally we identified 11 β-glucosidases (EC 3.2.1.21) from families GH1 (3) and GH3 (8) which were mostly predicted to be intracellular, except for two GH3, which were predicted to be secreted by non-classical pathways (Table S3). Also, the differences found in the CAZome of *Cellulomonas* sp. B6 and *C. flavigena* further supported the differences between both strains (Figure S2).

The GH62 (KSW17752.1) and one of the GH43 (KSW21678.1) were associated with an additional module identified as a GH5 domain by dbCAN algorithm but annotated as a ricin-type beta-trefoil domain (cd00161), involved in carbohydrate binding, by NCBI. To the best of our knowledge, no reports are available in the literature regarding similar proteins. Therefore, only experimental assays with the purified proteins would clarify the biological function of these modules.

Secretome analysis of CMC, SCR and EWS culture supernatants by shotgun proteomics

To further clarify which predicted CAZYmes were effectively being expressed and secreted, we analyzed by shotgun proteomics the secretome of *Cellulomonas* B6 supernatants from cultures grown on 10 g l^{-1} CMC, 10 g l^{-1} SCR and 10 g l^{-1} extruded wheat straw (EWS), using sucrose 5 g l^{-1} as control (negative for CAZymes secretion). Prior to proteomic analysis, protein secretion and overall activity were tested in the culture supernatants. As expected, the protein content and enzymatic activities were higher in the extracts from SCR and EWS than from CMC (0.79 ± 0.10, 0.58 ± 0.07 and 0.11 ± 0.05 mg ml^{-1}, respectively). The sucrose control resulted in similar protein content (0.13 ± 0.05 mg ml^{-1}) as CMC, but with negative results regarding enzymatic activities (Figure 4).

Extracellular proteins in the cell-free supernatants were analyzed by mass spectrometry and compared to the potentially secreted enzymes. All CAZy domain-containing proteins with signal peptide sequence detected by at least one matching unique peptide in one of the replicates were taken into consideration (Table S4). Proteins identified in the supernatants without signal peptide and not predicted to be extracellular by SecretomeP, could correspond to intracellular proteins, presumably due to cell lysis during experimental procedure, and were therefore not included in the analysis. In CMC supernatants, seven proteins with signal peptide were identified, four of which were CAZYmes. In SCR and EWS, 78 and 65 proteins with signal peptide were identified, from which 22 and 18 were CAZYmes, respectively, corresponding to approximately 28% in both cases. Although the abundance of a particular enzyme is not directly related to its activity, the differences observed in CMCase and xylanse activity levels correlated with the total proteins and CAZYmes identified. We also analyzed closely those proteins identified in culture supernatants that did not have a signal peptide sequence but were predicted to be extracellular by SecretomeP server. The mechanism (signal peptide independent or non-classical pathways) by which they could be secreted is not yet understood (Dyrløv Bendtsen *et al*. 2005).

From the 48 CAZy domain-containing proteins potentially secreted by a classical pathway (signal peptide dependent), 29 were identified (Table 2, Table S3). Among them, the exo-glucanases from families GH6 (KSW15663.1) and GH48 (KSW15916.1), as well as three GH10 xylanases (KSW23552.1 and KSW29966.1) were secreted by growth with all substrates (CMC, SCR and EWS), indicating their importance for polysaccharides utilization. Cellobiohydrolases of families GH6 and GH48 had also recently been identified in the secretome of *C. fimi* when grown on CMC and wheat straw (Spertino *et al.* 2018), supporting the relevance of these enzymes in the lignocellulose degrading activity of the *Cellulomonas* genus.

In SCR and EWS culture supernatants we also identified other three GH10 xylanases (KSW16349.1, KSW20567.1 and KSW18330.1) and the only GH11 (KSW19115.1), while two additional GH10s (KSW14744.1 and KSW21671.1) were also present in SCR (Table 2). Therefore, all extracellular

endoxylanases encoded in the genome were present in biomass culture supernatants, which could account for the xylanolytic activity observed in SCR and EWS supernatants. We also identified five of the seven predicted endoglucanases in supernatants from biomass cultures (the four GH9 and a GH6). Surprisingly, no endoglucanases were present in CMC culture supernatants, under our assay conditions. In SCR and EWS secretomes, we identified enzymes from GH62, GH18 and GH74, as well as three hypothetical proteins (KSW15560.1, KSW21470.1 and KSW21507.1), which presented putative esterases Pfam domains (CE1), and four CBM-containing proteins (KSW23491.1, KSW29693.1, KSW30160.1 and KSW29929.1) with no assigned catalytic module or enzymatic activity. The relative abundance of all proteins in each sample was estimated (Table S5). The GH10 (KSW23552.1) identified in all secretomes was also one the most abundant CAZymes, further supporting its relevance in *Cellulomonas* sp. B6 polysaccharides utilization.

Other proteins were only present in SCR culture supernatants, including an AA10 (KSW19119.1) (one of the two predicted in the genome), a GH27 α-galactosidase (KSW30136.1), a GH43 (KSW21678.1) and the only predicted PL3 (KSW13730.1), while a GH13 α-1,6-glucosidase (KSW29886.1) was identified in EWS extracts. It is worth mentioning that substrate- binding proteins from carbohydrate ABC transporters were present in all extracts, and were highly abundant (Table S5). These proteins may be involved in the uptake of the small sugars resulting from polysaccharides hydrolysis and their expression could be possibly co-regulated with some of the enzymes. For example, we identified genes coding for ABC transporter components downstream of the GH6 exo-glucanase gene, possibly in a transcriptional unit, although the role of these proteins and the regulation of their expression need to be further investigated. In the sucrose control we only identified three proteins (an esterase, a protein of undefined function with a CBM32 domain and a hypothetical protein with a recently described endogalactosaminidase GH114 domain), thus validating this culture condition as control for CAZymes secretion and confirming secretion induction by cellulosic substrates. Interestingly, we detected a GH3 β-glucosidase in supernatants from the SCR and EWS cultures and a CBM32 domain-containing protein in the extract from the EWS culture, both predicted to be secreted by signal

peptide-independent pathways. In summary, when lignocellulosic biomass was used as culture substrate, a higher number and diversity of enzymes was secreted, compared to growth on a pure soluble cellulosic substrate, such as CMC (Figure 3).

In a recent study, Spertino *et al* (2018) analysed the secretome of *C. fimi* when cultured on CMC and wheat straw. Although *Cellulomonas* sp. B6 and *C. fimi* are different species and a different methodology was used, interesting conclusions can be drawn by the comparison of results. Eight enzymes with high sequence identity were identified in both studies, corroborating their importance for the core lignocellulose degrading mechanism of the *Cellulomonas* genus. These were GH6 and GH48 exoglucanases, GH6 and GH9 endoglucanases, a GH10 xylanase, a GH74 xyloglucanase, a GH62 arabinofuranosidase and a PL3 pectate lyase (Table S6). However, other CAZymes secreted by *Cellulomonas* sp. B6 when grown on cellulosic biomass that also have homologues of high sequence identity to CAZymes encoded by *C. fimi* were not identified in its extracellular extract, such as two GH9 endoglucanases (KSW29417.1 and KSW29539.1), a GH10 xylanase (KSW29966.1), a GH13 α-1,6-glucosidase (KSW29886.1) and the only GH18 (KSW18655.1). Other CAZymes secreted by *Cellulomonas* sp. B6 had low sequence identity to proteins encoded by *C. fimi*. Conversely, there were eight enzymes identified in the secretome of *C. fimi* that did not have homologues or were not identified in culture supernatants of *Cellulomonas* sp. B6 (Table S6) (Spertino et al., 2018).

DISCUSSION

In this work, we studied a novel isolate from the *Cellulomonas* genus that presented high secretion of proteins when grown on low cost residual lignocellulosic biomass, as sole carbon source. *Cellulomonas* is a very interesting genus as it is a bacterial source of extracellular CAZYmes, which could replace or complement fungal enzymes for several applications, such as 2G-ethanol production. Its biological life cycle is shorter than those of fungi, which is an advantageous feature to lignocellulose-based industrial activities (Gomez Del Pulgar and Saadeddin 2014).

Cellulomonas sp. B6 was closely related phylogenetically to the well-known *C. flavigena* and to a rather novel species, *C. persica*. However, owing to the differences observed in the Average Aminoacid Identity (AAI) and Average Nucleotide Identity (ANI) as well as *in-silico* genome-to-genome DNA hybridization (DDH), we could not place it in one species or the other. In fact, the evidences reported here (including biochemical API results) indicated that *Cellulomonas* sp. B6 could be considered a novel species. Further studies will be undertaken to unequivocally assign a species to *Cellulomonas* sp. B6 isolate.

Enzymatic activities and thermal stability of enzymatic extracts from *Cellulomonas* sp. B6 were in the range of other previously described *Cellulomonas* species, although there were differences with the reported levels of protein secretion (Din *et al*. 1990; Ponce-Noyola and De La Torre 2001; Mayorga-Reyes *et al.* 2002; Amaya-Delgado *et al*. 2006; Sánchez-Herrera *et al.* 2007; Santiago-Hernández *et al.* 2007; Rusznyák *et al.* 2011; Yang *et al.* 2015; Kane and French 2018). In our study, the total protein secretion yield achieved by culture in lab scale flasks using biomass, either wheat straw or sugar cane straw, varied between 0.5 to 1 mg ml^{-1} opening a promissory way for improvement and standardization.

The *Cellulomonas* sp. B6 CAZome (the collection of carbohydrate-active enzymes encoded by the genome of an organism) indicated the presence of 205 CAZY proteins, from which 91 were GHs. These results further differentiated B6 from the type strains of *C. flavigena* and *C. fimi,* which had 86 and 113 predicted GHs (CAZy Database). In *C. flavigena,* the absence of GH74 and GH18 enzymes and the presence of more than one GH11 were among the most interesting differences. *C. fimi*, by contrast, had a GH74 and only one GH11 but presented a larger number of GH1 and GH3, and was more distant phylogenetically to isolate B6 (Figure S2).

By mass spectrometry of biomass-culture supernatants, we identified 21 out of the 32 GHs predicted to be secreted (29 with signal peptide and 3 by non-classical pathways). Among them, a GH6, a GH48 and two GH10 were present in all culture conditions. Remarkably, the GH48 was the only one

encoded in the genome. It was also noteworthy that two GH10 xylanases were induced by culture on CMC, a purely cellulosic substrate.

When grown on biomass, *Cellulomonas* sp. B6 secreted seven GH10 and a GH11. The nature of this apparent redundancy of GH10 xylanases could be related to the heterogeneity of xylan composition in biomass. Whether all GH10 xylanases are necessary for biomass deconstruction or have different substrate affinity, remains unclear. In consistency with our results, GH10 enzymes were also detected by shotgun proteomics in the supernatants of *C. fimi* and *C. flavigena* grown on CMC (Spertino *et al.* 2018; Wakarchuk *et al.* 2016).

The only GH11 of *Cellulomonas* sp. B6 was detected exclusively in the supernatants from cultures grown on SCR and EWS, and not CMC. Similarly, the GH11 of *C. fimi* was detected in xylan but not in CMC culture supernatants (Wakarchuk *et al*. 2016). However, it was not detected in *C. fimi* supernatants when the bacterium was grown on wheat straw (Spertino *et al.* 2018). Moreover, three GH11 xylanses were detected when *C. flavigena* was cultured on CMC and xylan (Wakarchuk *et al*. 2016) and the transcriptional upregulation of a GH11 in the presence of a lignocellulosic substrate had been previously described (Mayorga-Reyes *et al.* 2002). Therefore, the induction of xylanases by lignocellulosic biomass seems clear, although the fine-tunning of which xylanses are expressed under different culture conditions still needs further clarification.

Previous studies reported the identification of CbhA, a secreted GH6 exo-glucanase, from *C. fimi* in the presence of both cellulosic and hemicellulosic substrates (Meinke *et al*. 1994) and the presence of the CbhB exo-glucanase (GH48), from *C. fimi* and *C. flavigena,* in CMC (but not in xylan) culture supernatants (Shen *et al*. 1995). CbhA (GH6) and CbhB (GH48) were also present in *C. fimi* wheat straw culture supernantants (Spertino *et al*. 2018). It was suggested that these enzymes could be forming a stable complex and that this interaction could result in exo-exo synergy (Spertino *et al.* 2018). In this study, we identified the exo-glucanases GH6 and GH48 of *Cellulomonas* sp. B6 in supernatants of all cultures (biomass and CMC), supporting their role as the main cellulases of the

Cellulomonas sp. genus. Whether they act individually or forming a stable complex will help to establish a paradigm for the degradation of polysaccharides by aerobic bacteria soluble enzymes.

The results presented in this work contribute to establish the core enzymes used by the *Cellulomonas* genus to degrade lignocellulosic biomass in order to develop novel cost-efficient (hemi)cellulolytic bacterial enzymatic cocktails.

ACKNOWLEDGMENTS

FP 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 held a doctoral fellowship from the National Council for Research of Argentina (CONICET) during the activities described in the article. OO has a post-doctoral fellowship from CONICET. SG holds a fellowship from INTA. EC, DS, MR, PT and PV are members of the Scientific Research Career of CONICET. Authors thank Dr. Julia Sabio y García for English language editing assistance and critical reading of the manuscript.

CONFLICT OF INTEREST

No conflict of interest declared.

REFERENCES

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Charowhas, C., Clark, K., DiCuccio, M., Dondoshansky, I., Feolo, M., Funk, K., Geer, L.Y., Gorelenkov, V., Hlavina, W., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lee, J.M., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'Sullivan, C., Palanigobu, V., Panchenko, A.R., Phan, L., Pruitt, K.D., Rodarmer, K., Rubinstein, W., Sayers, E.W., Schneider, V., Schoch, C.L., Schuler, G.D., Sherry, S.T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Todorov, K., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W.J., Yaschenko, E. and Zbicz, K. (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **44**(D1), D7-19

Ahmed, I., Kudo, T., Abbas, S., Ehsan, M., Iino, T., Fujiwara, T. and Ohkuma, M. (2014). *Cellulomonas pakistanensis* sp. nov., a moderately halotolerant Actinobacteria. *Int J Syst Evol Microbiol* **64**, 2305–2311.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–10.

Amaya-Delgado, L., Vega-Estrada, J., Flores-Cotera, L.B., Dendooven, L., Hidalgo-Lara, M.E. and Montes-Horcasitas, M.C. (2006). Induction of xylanases by sugar cane bagasse at different cell densities of *Cellulomonas flavigena*. *Appl Microbiol Biotechnol* **70**, 477–481.

Auch, A.F., von Jan, M., Klenk, H.-P. and 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.

Bayer, E.A., Chanzy, H., Lamed, R. and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Curr Opin Struct Biol* **8**, 548–557.

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J. Bacteriol. 62, 293–300.

Bhat, M.K. (2000). Cellulases and related enzymes in biotechnology. *Biotechnol Adv* **18**, 355–383.

Bhattacharya, A.S., Bhattacharya, A. and Pletschke, B.I. (2015). Synergism of fungal and bacterial cellulases and hemicellulases: a novel perspective for enhanced bio-ethanol production. *Biotechnol Lett* **37**, 1117–1129.

Biely, P., Singh, S. and Puchart, V. (2016). Towards enzymatic breakdown of complex plant xylan structures: State of the art. *Biotechnol Adv* **34**, 1260–1274.

Bomble, Y.J., Lin, C.Y., Amore, A., Wei, H., Holwerda, E.K., Ciesielski, P.N., Donohoe, B.S., Decker, S.R., Lynd, L.R. and Himmel, M.E. (2017). Lignocellulose deconstruction in the biosphere. *Curr Opin Chem Biol* **41**, 61-70.

Campos, E., Negro Alvarez, M.J., Sabarís di Lorenzo, G., Gonzalez, S., Rorig, M., Talia, P., Grasso, D.H., Sáez, F., Manzanares Secades, P., Ballesteros Perdices, M. and Cataldi, A.A. (2014). Purification and characterization of a GH43 β-xylosidase from *Enterobacter* sp. identified and cloned from forest soil bacteria. *Microbiol Res* **169**, 213–220.

Coenye, T., Gevers, D., Van De Peer, Y., Vandamme, P. and Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiol Rev* **29**, 147–167.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R. and Tiedje, J.M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, 633–642.

Din, N., Beck, C.F., Miller, R.C., Kilburn, D.G. and Warren, R.A.J. (1990). Expression of the *Cellulomonas fimi* cellulase genes cex and cenA from the divergent tet promoters of transposon Tn10. *Arch Microbiol* **153**, 129–133.

Dyrløv Bendtsen, J., Kiemer, L., Fausbøll, A. and Brunak, S. (2005). Non-classical protein secretion in bacteria. *BMC Microbiol* **5**, 58-71.

Elberson, M.A., Malekzadeh, F., Yazdi, M.T., Kameranpour, N., Noori-Daloii, M.R., Matte, M.H., Shahamat, M., Colwell, R.R. and Sowers, K.R. (2000). *Cellulomonas persica* sp. nov. and *Cellulomonas iranensis* sp. nov., mesophilic cellulose-degrading bacteria isolated from forest soils. *Int J Syst Evol Microbiol* **50**, 993–996.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**, 783-791.

Gao, J. and Wakarchuk, W. (2014). Characterization of five β-glycoside hydrolases from *Cellulomonas fimi* ATCC 484. *J Bacteriol* **196**, 4103–4110.

Ghio, S., Di Lorenzo, G.S., Lia, V., Talia, P., Cataldi, A., Grasso, D. and Campos, E. (2012). Isolation of *Paenibacillus* sp. and *Variovorax* sp. strains from decaying woods and characterization of their potential for cellulose deconstruction. *Int J Biochem Mol Biol* **3**, 352–364.

Ghose, T.K. and Bisaria, V.S. (1987). International Union of Pure Commission on Biotechnology. Measurement of cellulase activities. *Pure Appl Chem* **59**, 257–268.

Gomez Del Pulgar, E.M. and Saadeddin, A. (2014). The cellulolytic system of *Thermobifida fusca*. *Crit Rev Microbiol* **40**, 236–247.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P. and Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81–91.

Hatayama, K., Esaki, K. and Ide, T. (2013). *Cellulomonas soli* sp. nov. and *Cellulomonas oligotrophica* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **63**, 60–65.

Hemsworth, G.R., Dejean, G., Davies, G.J. and Brumer, H. (2016). Learning from microbial strategies

for polysaccharide degradation. *Biochem Soc Trans* **44**, 94–108.

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265-1272.

Käll, L., Krogh, A. and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**, 1027–1036.

Kane, S.D. and French, C.E. (2018). Characterisation of novel biomass degradation enzymes from the genome of *Cellulomonas fimi*. *Enzyme Microb Technol* **113**, 9–17.

Kruer-Zerhusen, N., Alahuhta, M., Lunin, V. V., Himmel, M.E., Bomble, Y.J. and Wilson, D.B., (2017). Structure of a *Thermobifida fusca* lytic polysaccharide monooxygenase and mutagenesis of key residues. *Biotechnol Biofuels* **10**, 1–12.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870–1874.

Lamed, R., Setter, E. and Bayer, E. (1983). Characterization of a cellulose-binding, cellulose-containing complex in *Clostridium thermocellum*. *J Bacteriol* **156**, 828–836.

Lisov, A.V., Belova, O.V., Lisova, Z.A., Vinokurova, N.G., Nagel, A.S., Andreeva-Kovalevskaya, Z.I., Budarina, Z.I., Nagornykh, M.O., Zakharova, M.V., Shadrin, A.M., Solonin, A.S. and Leontievsky, A.A. (2017). Xylanases of *Cellulomonas flavigena*: expression, biochemical characterization, and biotechnological potential. *AMB Express* **7**(1),5.

Loose, J.S.M., Forsberg, Z., Kracher, D., Scheiblbrandner, S., Ludwig, R., Eijsink, V.G.H. and Vaaje-Kolstad, G. (2016). Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase. *Protein Sci* **25**, 2175–2186.

Lynd, L.R., Liang, X., Biddy, M.J., Allee, A., Cai, H., Foust, T., Himmel, M.E., Laser, M.S., Wang, M. and

Wyman, C.E. (2017). Cellulosic ethanol: status and innovation. *Curr Opin Biotechnol* **45**, 202–211.

Lynd, L.R., Weimer, P.J., van Zyl, W.H. and Pretorius, I.S. (2002). Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**(3), 506–577.

Mayorga-Reyes, L., Morales, Y., Salgado, L.M., Ortega, A. and Ponce-Noyola, T. (2002). *Cellulomonas flavigena*: Characterization of an endo-1,4-xylanase tightly induced by sugarcane bagasse. *FEMS Microbiol Lett* **214**, 205–209.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P. and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60.

Meinke, A., Gilkes, N.R., Kwan, E., Kilburn, D.G., Warren, R.A.J. and Miller, R.C. (1994). Cellobiohydrolase A (CbhA) from the cellulolytic bacterium *Cellulomonas fimi* is a beta-1,4-exocellobiohydrolase analogous to *Trichoderma reesei* CBH II. *Mol Microbiol* **12**, 413–22.

Mischnick, P. and Momcilovic, D. (2010). Chemical Structure Analysis of Starch and Cellulose Derivatives. *Adv Carbohydr Chem Biochem* **64**, 117-120.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A. and Xiong, G. (2013). Hemicellulose biosynthesis. *Planta* **238**(4), 627-642

Petersen, T.N., Brunak, S., von Heijne, G. And Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–786.

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

Ponce-Noyola, T. and De La Torre, M. (2001). Regulation of cellulases and xylanases from a derepressed mutant of *Cellulomonas flavigena* growing on sugar-cane bagasse in continuous culture. *Bioresour Technol* **78**, 285–291.

Rusznyák, A., Tóth, E.M., Schumann, P., Spröer, C., Makk, J., Szabó, G., Vladár, P., Márialigeti, K. and Borsodi, A.K. (2011). *Cellulomonas phragmiteti* sp. nov., a cellulolytic bacterium isolated from reed (*Phragmites australis*) periphyton in a shallow soda pond. *Int J Syst Evol Microbiol* **61**, 1662–1666.

Saitou, N. and Nei, M. (1987). The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. *Mol Biol Evol* **4**, 406–425.

Sánchez-Herrera, L.M., Ramos-Valdivia, A.C., De La Torre, M., Salgado, L.M. and Ponce-Noyola, T. (2007). Differential expression of cellulases and xylanases by *Cellulomonas flavigena* grown on different carbon sources. *Appl Microbiol Biotechnol* **77**, 589–595.

Santiago-Hernández, A., Vega-Estrada, J., Del Carmen Montes-Horcasitas, M. and Hidalgo-Lara, M.E. (2007). Purification and characterization of two sugarcane bagasse-absorbable thermophilic xylanases from the mesophilic *Cellulomonas flavigena*. *J Ind Microbiol Biotechnol* **34**, 331–338.

Shallom, D. and Shoham, Y. (2003). Microbial hemicellulases. *Curr Opin Microbiol* **6**, 219–228.

Shen, H., Gilkes, N.R., Kilburn, D.G., Miller, R.C. and Warren, R.A. (1995). Cellobiohydrolase B, a second exo-cellobiohydrolase from the cellulolytic bacterium *Cellulomonas fimi*. *Biochem J* **311** (Pt 1), 67–74.

Shinoda, K., Tomita, M., Ishihama, Y. (2010). emPAI Calc—for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics* **26**, 576–577.

Spertino, S., Boatti, L., Icardi, S., Manfredi, M., Cattaneo, C., Marengo, E. and Cavaletto, M. (2018). *Cellulomonas fimi* secretomes : In vivo and in silico approaches for the lignocellulose bioconversion. *J Biotechnol* **270**, 21–29.

Stackebrandt, E. and Schumann, P (2015). *Cellulomonas*. In Bergey's Manual of Systematics of Archaea and Bacteria, Bergey's Manual Trust. DOI: 10.1002/9781118960608.gbm00063.

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512–26.

Thompson, C.C., Chimetto, L., Edwards, R.A., Swings, J., Stackebrandt, E. and Thompson, F.L., (2013). Microbial genomic taxonomy. *BMC Genomics* **14**, 931.

Toushik, S.H., Lee, K.T., Lee, J.S. and Kim, K.S. (2017). Functional Applications of Lignocellulolytic Enzymes in the Fruit and Vegetable Processing Industries. *J Food Sci* **82**, 585–593.

Wakarchuk, W.W., Brochu, D., Foote, S., Robotham, A., Saxena, H., Erak, T. and Kelly, J. (2016). Proteomic analysis of the secretome of *Cellulomonas fimi* ATCC 484 and *Cellulomonas flavigena* ATCC 482. *PLoS One* **11**(3), e0151186.

Yang, Y., Zhang, L., Guo, M., Sun, J., Matsukawa, S., Xie, J. and Wei, D. (2015). Novel α- L -Arabinofuranosidase from *Cellulomonas fimi* ATCC 484 and Its Substrate-Specificity Analysis with the Aid of Computer. *J Agric Food Chem* **63**, 3725–3733.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, 445–451.

Zhang, L., Xi, L., Qiu, D., Song, L., Dai, X., Ruan, J. and Huang, Y. (2013). *Cellulomonas marina* sp. nov., isolated from deepsea water. *Int J Syst Evol Microbiol* **63**, 3014–3018.

Zhao, L., Geng, J., Guo, Y., Liao, X., Liu, X., Wu, R., Zheng, Z. and Zhang, R. (2015). Expression of the *Thermobifida fusca* xylanase Xyn11A in Pichia pastoris and its characterization. *BMC Biotechnol* **15**, 18.

Zhuang, W., Zhang, S., Xia, X. and Wang, G. (2015). Draft genome sequence of *Cellulomonas carbonis* T26^{T} and comparative analysis of six *Cellulomonas* genomes. Stand. *Genomic Sci* **10**, 104.

TABLES

Table 1 Predicted glycosyde hydrolases (GH) active on cellulose or hemicellulose and lytic polysaccharide monooxygenases (LPMO) of *Cellulomonas* sp. B6. CAZy domains were identified using dbCAN. Multiple domains are indicated in parentheses. Signal peptides were predicted using SignalPv4.1. Non-classical secretion was assigned to predicted proteins with no identified signal peptide and a score above 0.5 using the SecretomeP 2.0 server. Note: *annotation in the genome corresponds to partial sequences.

Predicted Activity	CAZy Domains	GenBank Accession	Signal Peptide	Predicted secretion mechanism
α-glucuronidase (EC 3.2.1.139)	GH67	KSW29842.1	N	-
α-L-arabinofuranosidase (EC 3.2.1.55)	GH51	KSW14247.1	N	-
	GH51	KSW29942.1	N	-
	GH62, GH5	KSW17752.1	Y	Classical
arabinan endo-1,5-α-L-arabinosidase	GH43	KSW21776.1	N	-
(EC 3.2.1.99)	GH43	KSW21922.1	N	-
β -glucosidase (EC 3.2.1.21)	GH1	KSW21931.1	N	-
	GH1	KSW29322.1	N	-
	GH1	KSW29522.1	N	-
	GH3	KSW19794.1	N	-
	GH3	KSW21537.1	N	-
	GH3	KSW21816.1	N	Non-classical
	GH3	KSW21825.1	N	Non-classical
	GH3	KSW21826.1	N	-

	GH3	KSW28963.1	N	-
	GH3	KSW29220.1	N	-
	GH3	KSW29840.1	N	-
β -mannosidase (EC 3.2.1.25)	GH2	KSW13761.1	N	-
	GH5	KSW28923.1*	Y	Classical
β -xylosidase (EC 3.2.1.37)	GH39	KSW30127.1	N	-
	GH39	KSW28373.1	N	-
β-xylosidase (EC 3.2.1.37)/ α -L-	GH43	KSW29213.1	N	-
arabinofuranosidase (EC 3.2.1.55)	GH43	KSW21730.1	N	-
	GH43	KSW29954.1	N	Non-classical
	GH43, CBM6, CBM13	KSW30137.1	Y	Classical
	GH43, GH5	KSW21678.1	Y	Classical
Chitinase (EC 3.2.1.14)	GH18, CBM2	KSW18655.1	Y	Classical
endo-β-1,4-glucanase (EC 3.2.1.4)	GH5	KSW14495.1	N	-
	GH6, CBM2	KSW30082.1	Y	Classical
	GH6, CBM2	KSW30095.1	Y	Classical
	GH9	KSW29417.1*	Y	Classical
	GH9, CBM2, CBM3	KSW29083.1	Y	Classical
	GH9, CBM4 (2)	KSW29108.1	Y	Classical
	GH9, CBM4 (2)	KSW29539.1	Y	Classical
β-1,4-xylanase (EC 3.2.1.8)	GH10, CBM22	KSW14744.1*	Incomplete sequence	Non-classical *
	GH10	KSW28941.1	N	-
	GH10	KSW18330.1	Y	Classical
	GH10	KSW29966.1*	Y	Classical

	GH10, CBM13	KSW21671.1	Y	Classical
	GH10, CBM2	KSW20567.1	Y	Classical
	GH10, CBM9, CBM22	KSW23552.1	Y	Classical
	GH10, CBM9, CBM22 (2)	KSW16349.1	Y	Classical
	GH11, CBM2	KSW19115.1	Y	Classical
exo-β-1,4-glucanase (EC 3.2.1.176)	GH48, CBM2	KSW15916.1	Y	Classical
exo-β-1,4-glucanase (EC 3.2.1.91)	GH6, CBM2	KSW15663.1	Y	Classical
Xyloglucanase (EC 3.2.1.151)	GH74 (4), CBM2	KSW29726.1	Y	Classical
Lytic Polysaccharide Monooxygenase	AA10, CBM2	KSW19119.1	Y	Classical
(LPMO)	AA10, CBM2	KSW30009.1	Y	Classical

Table 2 Extracellular proteins involved in carbohydrate utilization identified in culture supernatants of *Cellulomonas sp*. B6 grown on CMC, SCR or EWS by mass spectrometry. Multiple domains are indicated in parentheses. Note: *annotation in the genome corresponds to partial sequences.

Predicted Activity	GenBank Accession	CAZy Domains	Growth Substrate	Calculated MW[kDa]/pI
β-1,4-xylanase (EC 3.2.1.8)	KSW14744.1*	GH10, CBM22	SCR	54.8/ 4.75
	KSW29966.1*	GH10	CMC, SCR, EWS	37.8/8.51
	KSW16349.1	GH10, CBM9, CBM22 (2)	SCR, EWS	130.7/ 4.42
	KSW18330.1	GH10	SCR, EWS	65.4/5.27
	KSW19115.1	GH11, CBM2	SCR, EWS	36.5/9.36
	KSW20567.1	GH10, CBM2	SCR,EWS	50.6/8.76
	KSW23552.1	GH10, CBM9, CBM22	CMC, SCR, EWS	112.3/4.82
	KSW21671.1	GH10, CBM13	CMC, SCR	53.2/8.70
β-xylosidase (EC 3.2.1.37)/ α-L-arabinofuranosidase (EC 3.2.1.55)	KSW21678.1	GH43, GH5	SCR	76.0/9.07
endo-β-1,4-glucanase (EC 3.2.1.4)	KSW29417.1*	GH9	SCR	65.8/6.48
	KSW29083.1	GH9, CBM2, CBM3	SCR	86.4/5.68
	KSW29108.1	GH9, CBM4 (2)	SCR, EWS	114.1/4.58
	KSW29539.1	GH9, CBM4 (2)	SCR, EWS	118.8/4.58

	KSW30082.1	GH6, CBM2	EWS	47.3/8.41
exo-β-1,4-glucanase (EC 3.2.1.91)	KSW15663.1	GH6, CBM2	CMC, SCR, EWS	63.4/ 6.47
exo-β-1,4-glucanase (EC 3.2.1.176)	KSW15916.1	GH48, CBM2	CMC, SCR, EWS	90.6/ 5.33
β -glucosidase (EC 3.2.1.21)	KSW21825.1	GH3	SCR, EWS	83.5/4.48
Xyloglucanase (EC 3.2.1.151)	KSW29726.1	GH74 (4), CBM2	SCR, EWS	95.7/5.00
α-1,6-glucosidase (EC 3.2.1.20)	KSW29886.1	GH13 (2), CBM20, CBM48	SCR, EWS	207.1/4.75
α-galactosidase (EC 3.2.1.22)	KSW30136.1	GH27, CBM13	SCR	60.6/6.79
α -L-arabinofuranosidase (EC 3.2.1.55)	KSW17752.1	GH62, GH5	SCR, EWS	53.6/8.29
Chitinase (EC 3.2.1.14)	KSW18655.1	GH18, CBM2	SCR, EWS	56.3/5.77
Carbohydrate binding protein	KSW23491.1	CBM44 (2), CBM66	SCR, EWS	144.8/5.22
	KSW30160.1	CBM32	SCR, EWS	102.2/4.54
	KSW29929.1	CBM32	EWS	61.7/4.63
	KSW29693.1	CBM2	SCR, EWS	76.7/4.84
LPMO	KSW19119.1	AA10, CBM2	SCR	37.0/5.92
Esterases	KSW21470.1	CE1 (2)	SCR, EWS	93.6/4.75
	KSW21507.1	CE1	SCR, EWS	30.2/4.4
	KSW15560.1	CE1 (2)	SCR, EWS	98.0/ 4.67
Pectate lyase	KSW13730.1	PL3, CBM13 (2)	SCR	30.2/4.4
Sugar ABC transporter substrate-binding protein	KSW19791.1	-	EWS	45.1/4.65
	KSW21774.1	-	SCR, EWS	48.6/4.2
	KSW21936.1	-	SCR, EWS	59.4/4.64

	KSW28372.1	-	SCR, EWS	61.0/4.61
	KSW28971.1	-	SCR, EWS	48.2/4.26
	KSW29516.1	-	SCR, EWS	60.6/4.32
	KSW29892.1	-	SCR, EWS	39.9/4.40
	KSW30268.1	-	SCR, EWS	48.6/4.46
	KSW19409.1	-	CMC	46.7/4.33

Legends for the Supporting Information

Figure S1. a) CMCase and b) xylanase activity of *Cellulomonas* sp. B6 extracellular extracts (culture supernatants). Culture media is indicated in the X-axis. Statistically significant differences were determined by analysis of variance (one-way ANOVA) followed by Tukey´s test (**P<0.01; ***P<0.001; *** P<0.0001). c) Zymogram of CMCase and xylanase activities from CMC and SCR culture supernatants. SDS-PAGE 10–0.2% CMC or 0.2% XY were used as previously described (Ghio et al., 2016).

Figure S2. Number of proteins containing each CAZy domain in *Cellulomonas fimi* ATCC 484, *Cellulomonas flavigena* ATCC 482 and *Cellulomonas* sp. B6. Key: *A GH10 domain as well as a GH11 module were identified in one protein of *C. flavigena* and is taken into account for the total number of both GH10 and GH11 containing proteins; ** GH5 domains were found in four proteins, two of which also presented a GH62 or a GH43 domain.

Table S1. API50CH results of *Cellulomonas* sp. B6 compared to available information about *C. flavigena* and *C. persica* published in the International Journal of Systematic and Evolutionary Microbiology (Ahmed et. al, 2014; Elbertson et al., 2000) and in the Bergey's Manual of Systematics of Archea and Bacteria (Stackebrandt and Schumann, 2015). Key: + positive; - negative; ND not determined.

Table S2. Genomic taxonomy analysis of *Cellulomonas* sp. B6 in terms of computational DNA-DNA hybridization and Average Nucleotide Identity (ANI)/Average Aminoacid Identity (AAI) with *Cellulomonas flavigena* and *Cellulomonas persica* type strains. Genomic and proteomic sequences, length and G+C content were obtained from the NCBI Database. Distance, DNA-DNA hybridization (DDH) (identities/high score pairs length), probability that DDH >70% (same species) or >79% (same subspecies) and differences in %G+C (≥1 between distinct species) were estimated using the Genome to Genome Distance Calculator of the Leibniz Institute DMSZ (http://ggdc.dsmz.de). Two-way ANI and AAI were calculated with the online tools developed by the Environmental Microbial

Genomics Laboratory from Georgia Institute of Technology (http://enve-omics.ce.gatech.edu/). Key: DDH: DNA-DNA hybridization (computational), CI: confidence interval, SD: standard deviation

Table S3. CAZome of *Cellulomonas* sp. B6. All predicted CAZy domain-containing proteins were identified in the genomic sequence using the dbCAN algorithm. Multiple domains are indicated in parentheses. SignalP v4.1 (for Gram positive bacteria) was used for prediction of signal peptides and SecretomeP 2.0 was used for prediction of extracellular proteins secreted by non-classical pathways.

Table S4. Score and number of matching peptides detected by mass spectrometry in EWS, SCR and CMC supernatants for secreted CAZymes and substrate-binding proteins of carbohydrate ABC transporters.

Table S5. Relative abundance of proteins identified in the secretomes of *Cellulomonas* sp. B6, grown on cellulosic substrates. EWS: extruded wheat straw; SCR: sugar cane residue; CMC: carboxi-methyl-cellulose. Only biologically relevant proteins with signal peptide or considered extracellular by Secretome P algorithm are shown. The protein abundance index emPAI calculated by Sequest using protein identification data was used. The equation emPAI/Σ (emPAI) x 100 was used to calculate the protein content in mol % (emPAI%). The differences observed in the emPAI% correlate with the differences in the total number of proteins identified.

Table S6. CAZymes identified in secretomes of *Cellulomonas* sp. B6 and *C. fimi* ATCC 484 cultured on lignocellulosic biomass. Sequence identity was calculated using NCBI BLASTp. A sequence identity ≥ 90% (query cover ≥ 70%) was categorized as high.

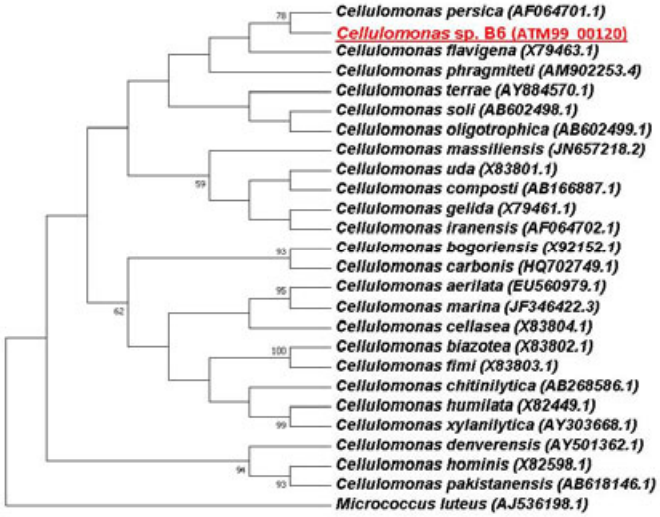

Figure 1 Phylogenetic relationship of *Cellulomonas* sp. B6 with species of the genus based on 16S rRNA sequence, using Neighbour Joining as statistical Method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) is shown next to the branches. Sequence accession numbers from the Ribosomal Database Project are indicated in parenthesis. *Micrococcus luteus* was used as outgroup to root the tree.

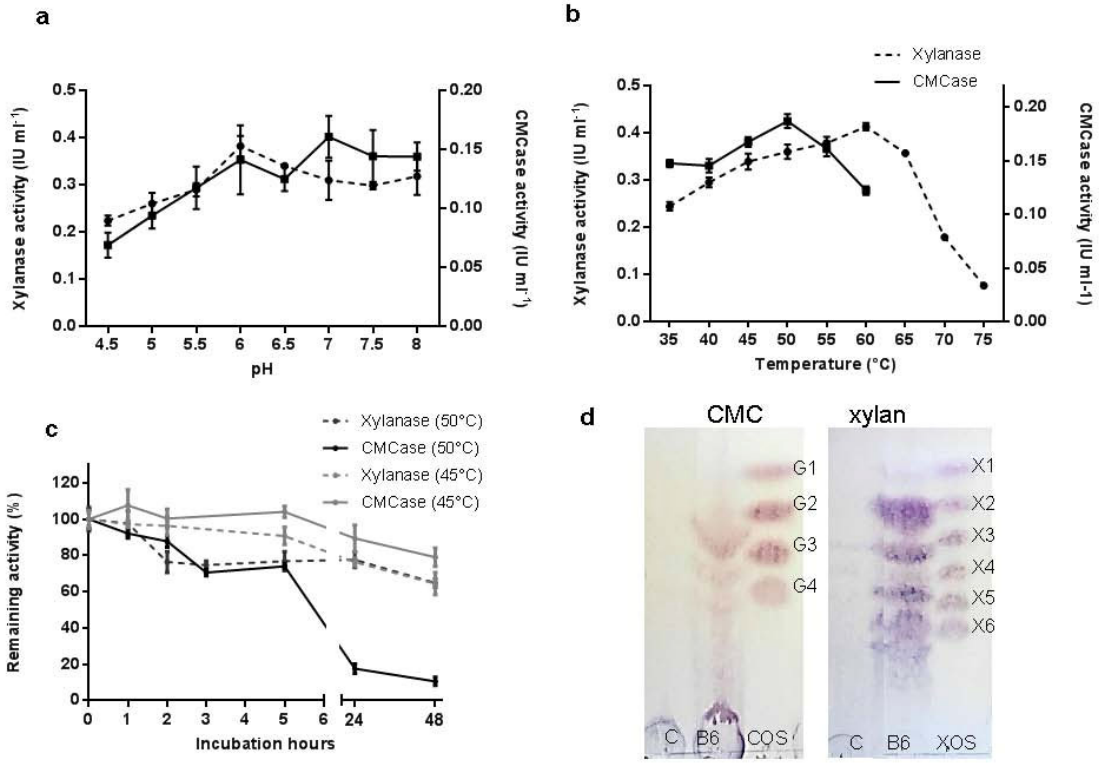

Figure 2 Extracellular CMCase and xylanase activities of *Cellulomonas* sp. B6. Activity profile in A) pH range from 4.5 to 8 and B) temperatures between 35°C and 60°C (CMCase activity) or 75°C (xylanase activity). C) Thermal stability of CMCase and xylanase activities in the extracellular extract after incubation at 45°C or 50°C. Each point is the media and the standard deviation of three biological replicates. D) TLC of CMC and xylan hydrolysis products (C: control, COS and XOS: cello- and xylo-oligosaccharides, respectively).

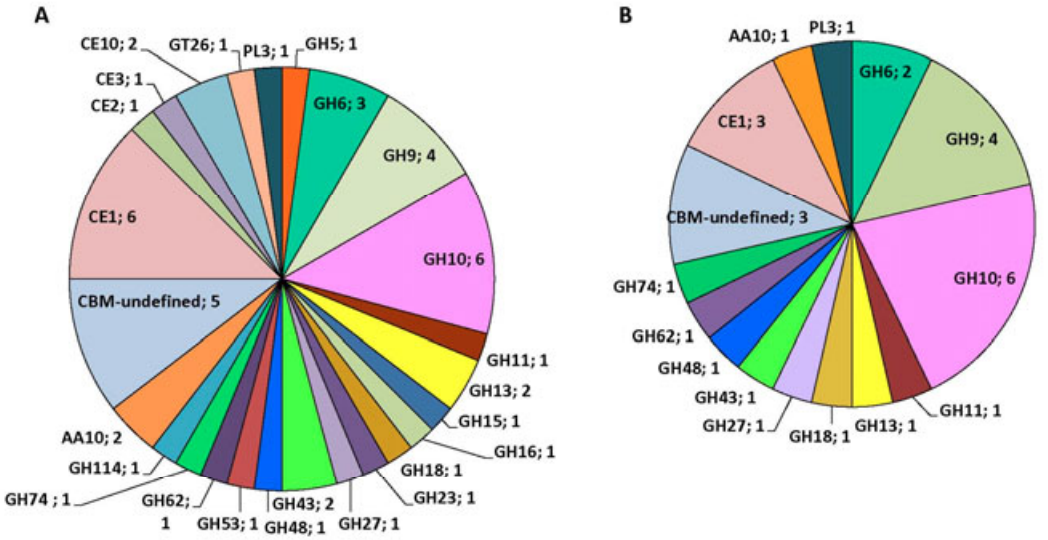

Figure 3 Analysis of *Cellulomonas* sp. B6 secretome. A) Theoretical secreted CAZymes by classical signal peptide pathway. B) Extracellular CAZYmes family distribution of proteins identified in *Cellulomonas* sp. B6 culture supernatants (from EWS, SCR or CMC cultures). Total numbers of proteins for each category are indicated in each chart. References: GH, glycoside hydrolases; GT, glycosyl transferases; CE, carbohydrate esterases; PL, polysaccharide lyases; AA, auxiliary activities; CBM, carbohydrate binding proteins. EWS, extruded wheat straw; SCR: sugar cane straw; CMC: carboximethylcellulose.

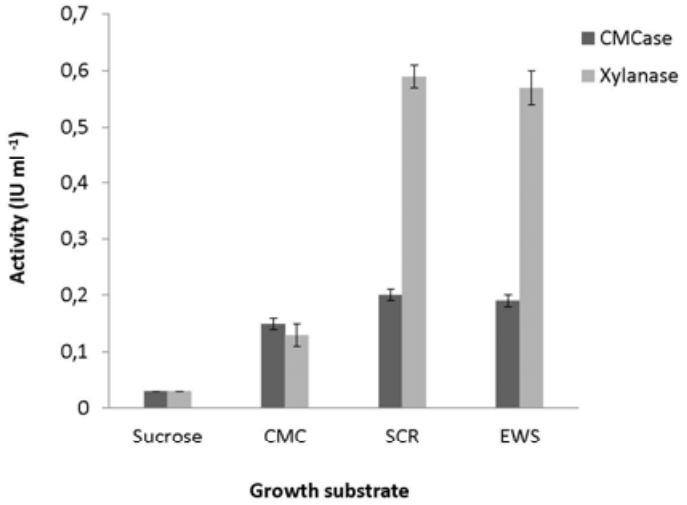

Figure 4 Enzymatic activity in culture supernatants of *Cellulomonas* sp. B6. The media and standard deviation of three biological replicates is indicated for each case. CMC: carboximethylcellulose; SCR: sugar cane residue; EWS: extruded wheat straw.