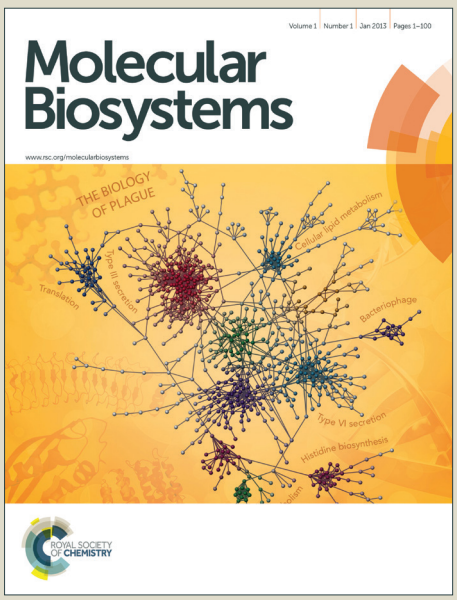
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Depiction of Carbohydrate-Active Enzymes Diversity in *Caldicellulosiruptor* sp. F32 at Genome Level Reveals Insights into Distinct Polysaccharides Degradation Feature

Dong-Dong Meng[1], Yu Ying[1,2], Kun-Di Zhang[1], Ming Lu[1], and Fu-Li Li[1*]

[1]Key Laboratory of Biofuels, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China

[2]Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, 266071, China

[*]Corresponding author: Fu-Li Li, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China. Tel.: 86-532-80662655. Fax: 86-532-80662778. E-mail: lifl@qibebt.ac.cn

Abstract

Thermophilic bacterium *Caldicellulosiruptor* sp. F32 can utilize cellulose-, hemicellulose-containing biomass, including unpretreated wheat straw. We have conducted a bioinformatics analysis of the carbohydrate-active enzyme (CAZyme) in the genome of *Caldicellulosiruptor* sp. F32, which reveals a broad substrate range of the strain. Among 2285 predicted open reading frames (ORFs), 73 (3.2%) CAZyme encoding genes, including 44 glycoside hydrolases (GHs) distributing in 22 GH families, 6 carbohydrate esterases (CEs), 3 polysaccharide lyases (PLs), 21 glycosyl transferases (GTs), and 25 carbohydrate-binding modules (CBMs) were found. An in-depth bioinformatics analysis of CAZyme families that target cellulose, hemicellulose, chitin, pectin, starch, and β-1,3-1,4-glucan degradation were performed to highlight specialized polysaccharides degrading abilities of strain F32. A great number of orthologous multimodular CAZymes of *Caldicellulosiruptor* sp. F32 were found in other strains of genus *Caldicellulosiruptor*. While, a portion of the CAZymes of *Caldicellulosiruptor* sp. F32 showed sequence identity with proteins from strains of genus *Clostridium*. A thermostable β-glucosidase BlgA synergistically facilitated enzymatic degradation of Avicel by endo-1,4-β-glucanase CelB, which indicated that synchronous action of synergism between CAZymes enhanced the lignocellulose degradation of *Caldicellulosiruptor* sp. F32.

Keywords: CAZymes; Glycoside hydrolase; Lignocellulose; Synergism; *Caldicellulosiruptor*

Introduction

Thermophilic bacteria of genus *Caldicellulosiruptor* (T_{opt} 65–78 °C) are robust and efficient carbohydrate degrading microorganisms, which secrete a series of carbohydrate active enzymes (CAZymes) to deconstruct lignocellulosic biomass[1]. Gram-positive *Caldicellulosiruptor* spp. have been isolated from a wide variety of ecological niches. Most of the species can metabolize an extensive monosaccharides, oligosaccharides, and polysaccharides, including crystalline cellulose, xylan, and switchgrass[2]. Furthermore, since no carbon catabolite repression (CCR) system exists in *C. saccharolyticus*, this strain has been shown to be able to ferment C_5 and C_6 sugars simultaneously, which was a highly desirable feature for lignocellulose conversion[3, 4]. All isolates are given a diverse set of cell wall degrading enzymes (CWDEs) as an excellent CAZymes reservoir for carbohydrate degradation, which are potential consolidated bioprocessing (CBP) platforms for biomass-to-biofuel conversion. Cellulolytic enzymes systems can be divided into two main categories, namely cellulosomal and noncellulosomal. Genus *Caldicellulosiruptor* adopts an intermediate strategy, secreting many free multi-modular cellulases containing multiple binding and catalytic domains[5].

Blumer-Schuette et al. have examined the CAZymes of eight numbers of this genus. Each strain contains a significant number of glycoside hydrolases (GHs, 37~77), carbohydrate binding modules (CBMs, 15~28), polysaccharide lyases (PLs, 1~4), carbohydrate esterases (CEs, 4~9), and glycosyl transferases (GTs, 27~35)[2], which comprise 1%~3% of the genome. Comparison analysis within the genus from genome level indicates 106 GHs distributing in 43 GH families constitute the *Caldicellulosiruptor* pangenome, while 26 GHs from 17 families are conserved in

the core genome[2]. Though more than half of the lignocellulose degradation related proteins are shared between all *Caldicellulosiruptor* species, genomic and proteomic studies in one individual species are needed to distinguish the difference of carbohydrate utilization within the genus[6]. As an efficient lignocellulolytic bacterium toward unpretreated plant biomass, analysis of CAZyme families and its architecture composition at genome level is helpful to elucidate the ability of *Caldicellulosiruptor* species for deconstruction of plant polysaccharides.

Caldicellulosiruptor sp. F32 is an extremely thermophilic anaerobic bacterium, which was isolated from biocompost with optimal growth temperature at 75 °C[7]. Compared with the type stain *C. saccharolyticus* DSM 8903, *Caldicellulosiruptor* sp. F32 grew better on unpretreated wheat straw, exhibited higher endoglucanase and xylanase activities in the secretome[8]. Three CAZy enzymes, GH11 xylanase XynA (GenBank accession number JX030400), GH10 xylanase XynB (GenBank accession number JX030401)[9], and GH5 lichenase F32EG5 (GenBank accession number KC958563, unpublished) have been characterized in detail. Further elucidation of the abundant CAZymes is thus essential to understand the complexity process of the lignocellulose deconstruction. This report focused on the CAZymes identification of strain F32 and provided a wealth of information to generate a comprehensive analysis of CWDEs in the host. Sequence and module annotation have been applied to predict functions of each CAZyme proteins. Moreover, synergism between two cellulose-degrading enzymes was investigated. Briefly, this will provide a better understanding of the plant-biomass degradation mechanisms of the genus *Caldicellulosiruptor*.

Materials and method

Genomic DNA isolation and genome sequencing

Caldicellulosiruptor sp. F32 (CGMCC 1.5183, China General Microbiological Collection Center, Beijing, China) was cultured at 75 °C in an anaerobic jar containing modified GS-2 medium[7]. Genomic DNA was isolated[10] from *Caldicellulosiruptor* sp. F32 and was sequenced by a standard shotgun strategy using GS FLX-454 pyrosequencing technology (Roche, Basel, Switzerland). The numerous reads were assembled by using the 454 Newbler assembler software (Roche) into hundreds of contigs, which were re-sorted subsequently to predict gene functions using Glimmer[11], GeneMark[12], and Zcurve[13]. The genes were annotated through the National Center for Biotechnology Information (NCBI) nr database[14] and KEGG database[15].

CAZymes annotation

All protein-encoding open reading frames (ORFs) from the *Caldicellulosiruptor* sp. F32 genome generated through nr and KEGG databases annotation, were further refined and screened for carbohydrate-active modules using Carbohydrate-active enzyme ANnotatio (dbCAN, http://csbl.bmb.uga.edu/dbCAN/index.php)[16]. To eliminate proteins identified by dbCAN but those were not really CAZymes, all proteins were individually examined using NCBI's Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)[17] and Pfam (http://pfam.sanger.ac.uk/)[18]. These data sources were combined to assert the function of each predicted proteins. When a gene appeared to be truncated and the missing sequence data couldn't be found within all the supercontig, the orthologous coding region in other *Caldicellulosiruptor* isolates were used to modify the gene sequence from *Caldicellulosiruptor*

sp. F32 genome. Further annotation of selected proteins included protein similarities and identities determination[19], molecular size and isoelectric point prediction[20], and signal peptide prediction (SignalP v4.1)[21].

Cloning, expression, and purification of *Caldicellulosiruptor* sp. F32 GHs

The primers BlgA-F 5′-ATGAGTTTCCCAAAAGGATTTTTG-3′ and BlgA-R 5′-GTATATAAAGGAAAATTCGTAA-3′ were used to amplify the β-glucosidase (BlgA) encoding gene. The primers CelB-F 5′-CAAAATACTGCGTATGAAAAGG-3′ and CelB-R 5′-TTACATCTTTCCTGTAAGTTCTAAAATTTTG-3′ were used to amplify the coding gene of endo-β-glucanase (CelB) without *N*-terminal signal peptide (1−28). The PCR product was cloned into the His tag expression vector *p*EASY-E1 (TransGen, Beijing, China) and transformed into *Escherichia coli* BL21 (DE3). The nucleotide sequence was verified via sequencing by GenScript (Nanjing, China). Protein expression was induced by using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were harvested 16 h after shaker at 16 °C and resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0) containing protease inhibitor cocktail (Amresco, Solon, USA). The resuspended cells were disrupted using ultrasonication on ice. Then, the crude enzyme was heated at 70 °C for 10 min and centrifuged at 4 °C to remove denaturated *E. coli* proteins. The enzyme solution was applied to a Ni-NTA-Sefinose column (Sangon, Shanghai, China). The active fraction was concentrated and replaced buffer by ultrafiltration using a 10-kDa cutoff membrane (Millipore, Billerica, MA) at 4 °C. Finally, the purified recombinant proteins were analyzed by using SDS-PAGE. Protein concentration was measured using the Bradford method with bovine serum albumin as the

standard[22].

Biochemical characterization of the recombinant enzymes

Two buffers, acetate buffer (0.2 M, pH 4.0−5.6) and PC buffer (50 mM phosphate, 12 mM citrate, pH 6.0−8.0), were used for pH profiling. For determination of optimal temperature, enzymatic reaction was carried out at different temperature ranging from 65 to 85 °C with a 5 °C interval at pH 5.6. The half-life ($t_{1/2}$) was obtained by measuring the residual activity of the enzyme at 75 °C and pH 5.6. The specific activity of the purified recombinant using of 1% (W/V) carboxymethyl-cellulose (CMC), Avicel PH-101, and cellobiose as substrate was determined. The reducing sugars released were measured using the dinitrosalicylic acid (DNS) method by reading the absorbance at a wavelength of 540 nm with glucose used as standard. *p*-Nitrophenyl-β-D-cellobioside (*p*NPC) was used at a final concentrations of 1 mM to determine the substrate-specific activity of the enzymes. *p*-Nitrophenol was measured by reading the absorbance at 405 nm to calculate activity. All cellulosic substrates were purchased from Sigma-Aldrich (St Louis, MO, USA). One unit (U) of enzyme activity was defined as the amount of enzyme that liberated 1 μmol of reducing sugar or *p*-nitrophenol per min. Unless otherwise stated, the reaction was performed at triplicate.

Each cello-oligosaccharide (cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose) at a final concentration of 10 mg mL^{-1}, were incubated with 0.5 μg BlgA in a total volume of 10 μL for 12 h at 75 °C, respectively. The hydrolysis products were analyzed using thin-layer chromatography (TLC) method by using silica gel plates 60F254 (Merck, Darmstadt, Germany). The plates were developed with a butan-1-ol−acetic acid−water (2:1:1, v/v) solvent

system. The color reaction was carried out in an oven at 105 °C for 30 min after spraying the plates with a methanol−sulfuric acid mixture (4:1, v/v). Oligosaccharides were purchased from Megazyme (Wicklow, Ireland) and used as the standards. The *p*NPCase of BlgA was performed with various glucose in a final concentration of 0−1000 mM at 75 °C and pH 5.6 to determinate the glucose-tolerance of BlgA.

Time course hydrolysis of Avicel by CelB was performed at a volume of 1 mL containing a portion (50 mg) of Avicel and 10 μg CelB at 75 °C and pH 5.6. At different time intervals (0 h, 1 h, 3 h, 6 h, 12 h, 24 h, and 36 h), aliquots were taken and subjected to high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD, Dionex Corporation, Sunnyvale, CA, USA) analysis to detect end products as described[23]. Synergy experiment was performed in triplicate in 1 mL reaction mixture containing 50 mg Avicel and 10 μg CelB with the supplementation dosage of BlgA at 0.1 μg, 0.5 μg, and 2 μg, respectively. Reaction was performed at 75 °C and pH 5.6 for 36 h, and reducing sugars were detected by DNS method. The degree of synergy (DOS) was calculated as reported before[24].

Nucleotide sequence accession numbers

The *Caldicellulosiruptor* sp. F32 genome sequence and annotation data have been deposited in GenBank under accession number APGP00000000. The GenBank accession numbers for the coding sequences of β-glucosidase BlgA and endo-β-glucanase CelB are JX030398 and JX030399, respectively.

Results and discussion

Identification of CAZyme modules in *Caldicellulosiruptor* sp. F32

A total of 127 contigs constituting 24 scaffolds were generated, yielding a total sequence of 2,378,643 bp of the *Caldicellulosiruptor* sp. F32 draft genomic DNA, which is shorter than the genome of type strain *C. saccharolyticus* DSM 8903 (2,970,275 bp). The protein coding genes of *Caldicellulosiruptor* sp. F32 were 2285, which account for 85.3% ORFs of *C. saccharolyticus* DSM 8903. The genome of *Caldicellulosiruptor* sp. F32 has a G+C content of 35.2%, and no plasmid was found. Our analysis indicated that 3.2% (73 of 2285) predicted proteins encoded in the genome of *Caldicellulosiruptor* sp. F32 are CAZymes involved in lignocellulosic biomass deconstruction. This value is similar to the numbers of CAZymes found in most genus of *Caldicellulosiruptor*[2]. A total of 44 ORFs containing one or two GHs have been found in the genome of *Caldicellulosiruptor* sp. F32, while in the type strain *C. saccharolyticus* DSM 8903 there are 59 ORFs containing GH domain. We noticed that the number of GHs in strain F32 is greater than that in *C. kristjanssonii* (37). Other than GH containing ORFs, the carbohydrate-binding modules (23), polysaccharide lyases (2), carbohydrate esterases (6), glycosyl transferase (21), and ABC transporters (22) are also identified in *Caldicellulosiruptor* sp. F32 (Table 1). In addition, twenty-one glycosyl transferases, distributing in GT families of 2, 4, 5, 28, and 36, are identified.

Cellulose deconstruction

Cellulose formed by β-1,4-linked glucose residues can be degraded by a synergistic action of three GH enzymes, namely, endo-β-1,4-glucanase (EC 3.2.1.4), cellobiohydrolase (EC 3.2.1.91), and β-glucosidase (EC 3.2.1.21). In addition to GHs, there are auxiliary activities (AA) proteins[25]

and CBMs[26] that assist to break the crystal structure of cellulose and act to concentrate enzyme by targeting and proximity effects, respectively. In *Caldicellulosiruptor* sp. F32, there are three endo-β-1,4-glucanases, distributing in GH families of 5, 9, and 48, respectively. Multi-modular enzyme CelA (Csac_1076 of *C. saccharolyticus* or Athe_1867 of *C. bescii*) represented 34.7% of total secreted proteins in a cellulose based enrichment of *C. bescii* secretome, and it was considered as one of the most important secreted cellulolytic enzymes in the species[27]. The polypeptide is composed of an *N*-terminal GH9 and a *C*-terminal GH48 catalytic domain that are separated by a CBM3c and two CBM3b module, which is also present in *Caldicellulosiruptor* sp. F32 (Fig. 1). All these two catalytic modules of CelA can synergistically hydrolyze crystalline cellulose[28], and can excavate extensive cavities into the surface of the substrate[5]. In the genome of strain F32, there are two predicted GH5 endo-β-1,4-glucanases, which may play an important role in cellulose degradation. One GH5 cellulase CelB of strain F32 containing a CBM28 module and three S-layer homology domains showed an amino acid sequence identity of 99% with Csac_0678 from *C. saccharolyticus* DSM 8903 (Fig. 1). It has been reported that Csac_0678 is bound to the *C. saccharolyticus* S-layer and contributed mainly to the hydrolysis of complex polysaccharides and deconstruction of lignocellulosic biomass[29]. An intracellular GH5 protein, located at 228572 to 229558 (reverse) of *Caldicellulosiruptor* sp. F32 genome, was predicted as cellulase, which might hydrolyze oligosaccharides that were transferred into cells. In the genome of *Caldicellulosiruptor* sp. F32, five putative β-glucosidases belonging to GH1 and GH3 were predicted (Table 2). In *Caldicellulosiruptor* sp. F32, 25 CBMs distribute in 12 families, and 18 out of these CBMs may be related with cellulose- or glucan-binding. The above

analysis suggests that the minimal set of CAZy-related proteins, a multi-modular GH9-GH48, 2 GH5 enzymes, and 5 β-glucosidases, endow *Caldicellulosiruptor* sp. F32 the ability to fully degrade cellulose into glucose, which exhibited 0.31 U mg^{-1} of filter paper activity (FPA), 0.57 U mg^{-1} of exo-glucanase, and 0.0035 U mg^{-1} of endo-glucanase in the Avicel induced secreted protein[8]. This fits well with core genome analysis of eight sequenced genomes of *Caldicellulosiruptor*[2].

Hemicellulose degradation

Hemicellulose, the second most abundant polysaccharides on earth, accounts for 20%~35% of lignocellulosic biomass[30], which contains pentose (β-D-xylose, α-L-arabinose), hexose (β-D-mannose, β-D-glucose, α-D-galactose) and/or uronic acid (α-D-glucuronic, α-D-4-*O*-methylgalacturonic and α-D-galacturonic acid)[31]. Hemicellulose is a heterogeneous group of branched and linear polysaccharides including xyloglucans, xylans, mannans and glucomannans, which degradation is mediated by several kinds of enzymes. Xylan is the most common hemicellulose and its degradation requires endo-1,4-β-xylanase (EC 3.2.1.8) and β-xylosidase (EC 3.2.1.37). In addition, α-L-arabinofuranosidase, α-glucuronidase, acetylxylan esterase, ferulic acid esterase, and *p*-coumaric acid esterase are necessary for the hydrolysis of various substituted xylans[30].

Xylanase (endo-1,4-β-xylanase) hydrolyzes the β-1,4 bond in the xylan backbone of hemicellulose, yielding short xylooligosaccharides, which are further degraded by β-xylosidase. Most xylanases belong to GH families 10 and 11[32]. *Caldicellulosiruptor* sp. F32 exhibits distinctive properties in growth and xylan hydrolysis, which was facilitated by the thermostable

and efficient set of xylanases in the secreted proteins[7]. The *Caldicellulosiruptor* sp. F32 genome contains two multi-modular xylanases, namely GH10 XynA and GH11 XynB[9]. In addition, another ORF encoding GH10 xylanase without signal peptide was found in the genome of strain F32 (Fig. 1) and it might degrade oligosaccharides with a low degree of polymerization. Furthermore, candidate β-xylosidases including 7 putative GH2 and 5 putative GH43 proteins were detected in the genome, and they were annotated as xylan β-1,4-xylosidase or α-L-arabinofuranosidase which might hydrolyze oligosaccharides and assist xylanases for the complete degradation of xylan to xylose. Interestingly, homology sequences of these two GH10 xylanases were found in all *Caldicellulosiruptor* species, while GH11 xylanase were only found in *C. bescii* (ACM59249.1), *C. owensensis* (ADQ03663.1), and *Caldicellulosiruptor* sp. Rt69B.1 (AAB95327.1). It is worth noting that, multi-modular extracellular GH11 xylanase possessed 9 fold higher specific activity than GH10 xylanase, and they also showed a degree of synergy of approximately 1.3[9]. It has been reported that GH11 xylanases are exclusively active on D-xylose containing substrates, while, GH10 xylanases show mainly activity towards xylooligosaccharide and cellulose substrates with low molecular mass[32]. In the genome of the type strain *C. saccharolyticus* DSM 8903, six GH10 enzymes encoding ORFs including Csac_1078, Csac_0204, Csac_0696, Csac_2405, Csac_2408, and Csac_2410 were found, but no GH11 xylanase encoding sequence present. Due to the synergistic activity of GH11 and GH10 xylanases found in strain F32, 2.5 fold higher xylanase activity of the secreted proteins in strain F32 has been detected than that in *C. saccharolyticus* DSM 8903[7]. As we know, high xylanase activity can lead to improved utilization of lignocellulose during initial and mid-exponential

phage of microorganisms[33], therefore, we can explain at genome level now why strain F32 showed better growth (8.3×10^7 cells/mL) than DSM 8903 (2.5×10^7 cells/mL) when unpretreated wheat straw was used as the sole carbon source. The secreted protein encoded by Csac_2410 is a GH5 and GH10 containing multi-modular enzyme, activity towards CMC, Avicel, barley β-glucan, and xylan has been proved[34].

In *Caldicellulosiruptor* sp. F32, a panel of glycoside hydrolases including one putative GH4 α-galactosidase (EC 3.2.1.22), one putative GH31 α-xylosidase (EC 3.2.1.177), five putative GH43 α-L-arabinofuranosidases (EC 3.2.1.55), one putative GH30 glucosylceramidase (EC 3.2.1.45), and one putative GH51 α-L-arabinofuranosidase (EC 3.2.1.55), which are involved in removing side chains from the xylose backbone of hemicellulose, were identified. In the genome of *Caldicellulosiruptor* sp. F32, there are 6 putative carbohydrate esterases (acetylxylan esterase, EC 3.1.1.72) belonging to CE4, CE7, and CE9, most of them might shear the *O*-acetyl groups from positions 2 and/or 3 on the β-D-xylopyranosyl residues of acetyl xylan, particularly arabinoglucuronoxylan[31]. In addition to xylan-degrading enzymes, there are 2 putative GH26 β-mannanases (EC 3.2.1.78) and 7 putative GH2 β-mannosidases (EC 3.2.1.25) which are related to the degradation of the main chain and side chains of glucomannans. CBMs are usually considered as auxiliary domain in conjunction with other CAZyme catalytic modules[35]. The traditional recognized function of non-catalytic CBMs is to bind polysaccharides, bringing the biocatalyst into its substrates by targeting and proximity effects[26]. There are over 70 families of CBMs, and these CBMs are classified into three types based on their binding preference towards crystalline polysaccharides, soluble polysaccharides, and soluble mono-, di- or tri-saccharides,

respectively[36]. It is elusory that the deletion of CBMs from the multi-modular GH10 and GH11 xylanases from *Caldicellulosiruptor* sp. F32 resulted in opposite effect on thermostability of the truncated mutants. Through homology modeling and cross-linking analysis, it turned out to be that optimized intramolecular interaction between CBM and catalytic domain conveyed improved thermostability of GH10 xylanase[9].

Chitin degradation

Chitin, a linear polymer of β-1,4-*N*-acetylglucosamine (GlcNAC), is found in the outer skeleton of insects, fungi, yeasts, algae, crabs, shrimps, and lobsters, and in the internal structure of other invertebrates[37]. Chitinases (E.C 3.2.1.14) are a widespread group of glycosyl hydrolases found in bacteria, fungi, yeasts, plants, actinomycetes, and arthropods[38]. The enzymes act on the *N*-acetyl-β-1,4-glucosaminide linkages in chitin and randomly split chitin at internal sites to produce oligosaccharides such as chitotriose, and chitotetraose[39]. *Caldicellulosiruptor* sp. F32 contains a putative extracellular GH18 chitinase (Table 2), which shares the highest amino acid identity of 40% with chitinase *Lp*ChiA from *Laceyella putida* (BAO37115.1)[40]. Members from GH18 family usually show chitinase or endo-*N*-acetyl-β-1,4-glucosaminidase activity[41]. In the genome of *Caldicellulosiruptor* sp. F32, a GH20 putative *N*-acetylglucosaminidase (EC 3.2.1.30) was predicated, which might play a role in the cleavage of oligomeric products of chitinase, thereby generating monomers of GlcNAc[39].

Pectin degradation

Pectin is mainly constructed by three types of polysaccharides including homogalacturonan (HG, 1,4-linked α-D-galactopyranosyluronic acid, Gal*p*A), rhamnogalacturonan-I (RG-I,

289 alternative α-1,2-rhamnosyl and α-1,4-galacturonic acid), and RG-II (α-1,4-linked Gal*p*A)[42,43]. A
290 panel of CAZymes, distributing in GHs, CEs, and PLs, are required for complete deconstruction
291 of the complex pectin. Usually, GH28 endopolygalacturonases (EC 3.2.1.15) and
292 exopolygalacturonases (EC 3.2.1.67) can cleave α-1,4-linkages and α-1,2-linkages in pectin,
293 respectively. Till now, CAZymes from the GH1, GH2, GH3, GH5, GH30, GH35, GH43, GH53,
294 and GH54 families have been reported to be able to degrade side chains of pectin[43]. In the
295 genome of *Caldicellulosiruptor* sp. F32, though GH28 CAZyme was not found, a GH105
296 putative rhamnogalacturonyl hydrolase (EC 3.2.1.172) was considered to be involved in the
297 cleavage of rhamnogalacturonan. A homology of Cbes_1853, which has been proposed as
298 rhamnogalacturonan lyase in *C. bescii*, containing the modules of PL11 and CBM3, is also
299 present in the genome of strain F32[44]. Furthermore, two PL3 CAZymes annotated as pectate
300 lyase in *Caldicellulosiruptor* sp. F32 might cleave bonds linking the α-1,4-galacturonan residues
301 to HG backbone at the non-reducing end[43].

302 **Starch degradation**

303 α-Amylase (1,4-α-D-glucan-4 glucanohydrolase; EC 3.2.1.1) catalyzes the hydrolysis of the
304 α-1,4 glycosidic linkages of starch, thereby generating dextrins and oligosaccharides with low
305 polymerization[45]. It is universally distributed throughout the animal, plant and microbial
306 kingdoms[46]. In the genome of *Caldicellulosiruptor* sp. F32, five putative GH13 α-amylases were
307 predicted. Among them, a family 13 glycoside hydrolase, located at 329573 to 330500 (reverse),
308 contains an *N*-terminal signal peptide, whereas the other four do not have. Two GH13
309 α-Amylases were flanked with CBM48 and CBM34, respectively (Fig. 1). Both CBMs from

family 48 and 34 are starch or glycogen-binding module. A family 15 glycoside hydrolase, predicted as a secreted protein, was annotated as glucoamylase (EC 3.2.1.3), which is exo-acting amylase that releases glucose from the nonreducing end of starch and related oligosaccharides[47]. Pullulanase (EC 3.2.1.41) is a de-branching enzyme that cleaves the 1,6-α-glucosidic linkages in pullulan or other amylaceous polysaccharides[48]. In the genome of *Caldicellulosiruptor* sp. F32, two GH13 putative pullulanases were predicted. One contains the GH13 catalytic domain, a signal peptide, an *N*-terminal CBM41 and a *C*-terminal CBM2. The other GH13 putative pullulanase is also a secreted protein flanked by an *N*-terminal CBM48 (Fig. 1). Altogether, 8 starch-degrading enzymes consist of a complex CAZymes system for pullulan and starch degradation.

β-1,3-1,4-glucan degradation

β-1,3-1,4-glucan (lichenin), a liner polysaccharide containing up to 1,200 β-D-glucosyl residues, is an important structural and storage polysaccharide in the endosperm cell walls of cereals and lichens[49]. β-1,3-1,4-glucanases (lichenase, EC 3.2.1.73) can selectively hydrolyze β-1,4 glycosidic bonds in 3-*O*-substituted glucosyl residues[49]. A GH5 enzyme F32EG5 (GenBank accession number KC958503, Fig. 1) showed the highest amino acid identity of 53% with an endoglucanase CelCCA (YP_002505438) from *C. cellulolyticum* and was proved to be a lichenase in *Caldicellulosiruptor* sp. F32 (data unpublished). The analysis of hydrolysis products indicated that F32EG5 cleaved the β-1,4 linkage while the 4-*O*-substitued glucose residue linked to a 3-*O*-substited glucose residue, as well as the β-1,3 linkage, which is completely different from extensively studied GH16 lichenase, which catalyzes strict endo-hydrolysis of the

β-1,4-glycosidic linkage adjacent to a 3-*O*-substitued glucose residue in the mixed linked β-glucans[49], suggesting F32EG5 is a novel type lichenase with specific substrate specificity and glycosidic bond cleavage pattern. A family 16 glycoside hydrolase, consisting of a signal peptide, two *N*-terminal SLH domains, a GH 16 catalytic domain, and six *C*-terminal CBMs with predicted endo-β-1,3-glucanase (EC 3.2.1.39) activity, was found in the genome of strain F32 (Fig. 1). Endo-β-1,3-glucanases are hydrolase specific to *O*-glycoside bonds between 1,3-linked glucopyranose residues found in variety of β-1,3-glucan including laminarin, yeast β-1,3-1,6-glucan, and β-1,3-1,4-glucan[50].

Other possible lignocellulose-degrading related modules

In the genome of *Caldicellulosiruptor* sp. F32, there are 19 extracellular solute-binding proteins. Two dimensional electrophoresis experiments have shown that there are 23 extracellular proteins which were significantly upregulated on microcrystalline cellulose (Avicel PH-101) than glucose[51]. Among them, 4 proteins are extracellular solute-binding proteins, indicating the important role in cellulose degradation. Yokoyama et al. have reported that multidomain enzymes of genus *Caldicellulosiruptor* might be recruited to regions bound by noncatalytic plant cell wall-binding proteins (PWBPs) on the plant cell wall, and thus involved in bacterial survival and lignocellulosic biomass deconstruction[52]. Surface layer homology (SLH) domain protein is a distinguishing feature of genus *Caldicellulosiruptor* and may play a role in lignocellulose deconstruction[29]. The GH5 glucanase (Csac_0678) from *C. saccharolyticus* DSM8903 was investigated in details. SLH domain did not show any impact on Csac_0678 GH activity, but contributed to the binding to S-layer of cells[29]. In the genome of

352 *Caldicellulosiruptor* sp. F32, 7 SLH domain containing ORFs including two CAZymes with
353 GH5 and GH16 catalytic domain were annotated, respectively. SLH domain-containing protein is
354 likely to anchor the catalytic domain near cells in order to quickly absorb the oligosaccharides
355 produced by extracellular CAZymes.

356 **Thermostable β-glucosidase BlgA synergistically facilitates enzymatic degradation of Avicel**
357 **by endo-1,4-β-glucanase CelB**

358 Recombinant β-glucosidase BlgA contains one catalytic domain belong to GH family 1. CelB
359 is constituted of an *N*-terminal GH5 catalytic domain annotated as endoglucanase, three
360 *C*-terminal SLH domains, and a family 28 CBM between GH5 and SLH domains. Further
361 analysis indicated that CelB has an *N*-terminal signal peptide at residues 1-28, suggesting it is a
362 secreted enzyme. Orthologs of BlgA and CelB also exist in all genome sequences of eight
363 *Caldicellulosiruptor* species, indicating the important role of these two enzymes in cellulose
364 deconstruction.

365 These two thermostable cellulose-degrading enzymes were heterologously expressed in *E. coli*
366 BL21 (DE3) and purified by His-tag mediated affinity chromatography to near-homogenous,
367 respectively. BlgA exhibited the highest specific activity against *p*NPC (22.1 U/mg) and
368 cellobiose (20 U/mg), indicating it is a β-glucosidase. CelB showed hydrolysis activity with
369 CMC (17.1 U/mg) and *p*NPC (10.4 U/mg) as substrates (Table 3). Both of them have a
370 temperature optimum of 75 °C and a pH optimum of 5.6. The half-lives of recombinant BlgA
371 and CelB at 75 °C were 48 h and 34 h, respectively.

372 Various cello-oligosaccharides (C2-C6) were incubated with BlgA, after 12 hours, all

substrates were converted to glucose completely (Fig. 2A). It has been reported that glucose tolerant β-glucosidase has significant advantage in the process of saccharification of lignocellulosic materials[53]. The search for β-glucosidases insensitive to high concentration of glucose has received a lot of attention, but few of them focus on thermostable enzyme. In this study, the specific activity of β-glucosidase BlgA was gradually inhibited while increasing the concentration of glucose, and a K_i of 350 mM glucose for BlgA was determined, which is higher than the K_i value of 200 mM reported for β-glucosidase from *Scytalidium thermophilum*[54] (Fig. 2B). The K_i value of other β-glucosidase from *Aspergillus niger*, *Thermoanaerobacterium thermosaccharolyticum*, and *A. tubingensis* has been reported to be 543, 600, and 600 mM, respectively, which indicated these enzymes exhibited higher glucose-tolerant[55-57]. BlgA can efficiently eliminate the product-inhibitors of endoglucanases during lignocellulosic biomass hydrolysis.

The degradation capacity of CelB was investigated in a time course approach with Avicel as substrate (Fig. 2C), and most of the released products were determined to be glucose and cellobiose, which is consistent with the substrate specificity results that the endoglucanase CelB exhibited high activity for *p*NPCase. In the following experiment, in order to eliminate feedback inhibition of hydrolysis products for CelB, BlgA was added into the reaction mixture. The synergistic effect of these two thermostable enzymes was investigated. After 36 h incubation, the hydrolysis ability of CelB toward Avicel was improved by 65%, 110% and 154% in presence of purified BlgA at the ratio of 1:100, 1:20 and 1:5 (BlgA:CelB), respectively (Fig. 2D). BlgA is boosting cellulose-degrading ability of endo-glucanase CelB with the degree of synergy of 2.54

394 at the ratio of 1:5 (BlgA:CelB). Apparently, thermostable β-glucosidase BlgA synergistically
395 facilitated enzymatic degradation of Avicel by CelB via removing cellobiose which inhibits the
396 enzyme activity. Furthermore, possible accessible substrate subsites may be provided to BlgA
397 during the hydrolysis of Avicel by CelB, which enhanced the synchronous action of synergism[58].

398 *Caldicellulosiruptor* spp. can efficiently degrade lignocellulosic plant biomass without
399 pretreatment that facilitated by a diverse set of CAZymes. Unpretreated plant biomass, such as
400 poplar, napier, and Bermuda grasses, can be utilized by *C. becsii*[59]. When unpretreated wheat
401 straw was used as sole carbon source, strain F32 showed better growth than DSM 8903. The
402 secretome of strain F32 showed 0.17 U mg^{-1} of FPA, 0.47 U mg^{-1} of endo-glucanase, and 8.7 U
403 mg^{-1} of xylanase, which was higher than that of *C. saccharolyticus*[8]. The physiological and
404 biochemical feature of strain F32 indicated that a wealthy of CAZymes including cellulase and
405 hemicellulase, and plus synergism effect of these enzymes, such as BlgA and CelB, jointly
406 enabled the strain to deconstruct complex plant cell wall.

407

408 **Conclusion**

409 *Caldicellulosiruptor* sp. F32 holds in store of a thermophilic CAZymes library for its strong
410 lignocellulose degradation capacity. Most of them have convergently evolved with those from
411 other type strains of genus *Caldicellulosiruptor*. Meanwhile, some CAZymes might be
412 transferred from other organism such as genus *Clostridium* via horizontal gene transfer, enabling
413 the isolate to adapt itself to diverse environmental conditions. The thermostable β-glucosidase
414 BlgA and endo-glucanase synergistically hydrolyze cellulose, thereby eliminating feedback

inhibition of intermediate product. All the strategies adopted by *Caldicellulosiruptor* sp. F32 revealed how the microbe evolved an efficient system for utilization of carbohydrates. These results provide a basis for further development of improved enzyme mixtures aiming at industrial processing of plant biomass.

Acknowledgments

This work was supported by grants from the National Basic Research Program of China (No. 2011CB707404), the National Science Foundation of China (No. 31400060), and the Shandong Province Natural Science Funds for Distinguished Young Scholar (JQ201507).

References

1. S. Q. Ji, D. D. Meng, K. D. Zhang and L. F. Li, in *Thermophilic microorganisms*, ed. L. F. Li, Caister Academic Press, Norfolk, 2015, ch. 2, pp. 13-46.
2. S. E. Blumer-Schuette, R. J. Giannone, J. V. Zurawski, I. Ozdemir, Q. Ma, Y. B. Yin, Y. Xu, I. Kataeva, F. L. Poole, M. W. W. Adams, S. D. Hamilton-Brehm, J. G. Elkins, F. W. Larimer, M. L. Land, L. J. Hauser, R. W. Cottingham, R. L. Hettich and R. M. Kelly, *J. Bacteriol.*, 2012, **194**, 4015-4028.
3. H. J. G. van de Werken, M. R. A. Verhaart, A. L. VanFossen, K. Willquist, D. L. Lewis, J. D. Nichols, H. P. Goorissen, E. F. Mongodin, K. E. Nelson, E. W. J. van Niel, A. J. M. Stams, D. E. Ward, W. M. de Vos, J. van der Oost, R. M. Kelly and S. W. M. Kengen, *Appl. Environ. Microbiol.*, 2008, **74**, 6720-6729.
4. A. L. VanFossen, M. R. A. Verhaart, S. M. W. Kengen and R. M. Kelly, *Appl. Environ. Microbiol.*, 2009, **75**, 7718-7724.
5. R. Brunecky, M. Alahuhta, Q. Xu, B. S. Donohoe, M. F. Crowley, I. A. Kataeva, S. J. Yang, M. G. Resch, M. W. Adams, V. V. Lunin, M. E. Himmel and Y. J. Bomble, *Science*, 2013, **342**, 1513-1516.
6. S. E. Blumer-Schuette, S. D. Brown, K. B. Sander, E. A. Bayer, I. Kataeva, J. V. Zurawski, J. M. Conway, M. W. W. Adams and R. M. Kelly, *FEMS Microbiol. Rev.*, 2014, **38**, 393-448.
7. Y. Ying, D. D. Meng, X. H. Chen and F. L. Li, *Enzyme Microb. Technol.* , 2013, **53**, 194-199.
8. Y. Ying, D. D. Meng and F. L. Li, *Microbiology China*, 2014, **41**, 211-217.
9. D. D. Meng, Y. Ying, X. H. Chen, M. Lu, K. Ning, L. S. Wang and F. L. Li, *Appl. Environ. Microbiol.* , 2015, **81**, 2006-2014.
10. L. V. Andreou, *Method Enzymol.*, 2013, **529**, 143-151.
11. A. L. Delcher, K. A. Bratke, E. C. Powers and S. L. Salzberg, *Bioinformatics*, 2007, **23**, 673-679.

12. J. Besemer and M. Borodovsky, *Nucleic Acids Res*, 2005, **33**, W451-W454.
13. F. B. Guo, H. Y. Ou and C. T. Zhang, *Nucleic Acids Res*, 2003, **31**, 1780-1789.
14. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers, *Nucleic Acids Res*, 2011, **39**, D32-D37.
15. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi and M. Tanabe, *Nucleic Acids Res*, 2012, **40**, D109-D114.
16. Y. B. Yin, X. Z. Mao, J. C. Yang, X. Chen, F. L. Mao and Y. Xu, *Nucleic Acids Res*, 2012, **40**, W445-W451.
17. A. Marchler-Bauer, S. N. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. X. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. C. Zhang, N. G. Zhang, C. J. Zheng and S. H. Bryant, *Nucleic Acids Res*, 2011, **39**, D225-D229.
18. M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman and R. D. Finn, *Nucleic Acids Res*, 2012, **40**, D290-D301.
19. J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Res*, 1994, **22**, 4673-4680.
20. E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch, *Nucleic Acids Res*, 2003, **31**, 3784-3788.
21. T. N. Petersen, S. Brunak, G. von Heijne and H. Nielsen, *Nat. Methods*, 2011, **8**, 785-786.
22. M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248-254.
23. N. Hu, B. Yuan, J. Sun, S. A. Wang and F. L. Li, *Appl Microbiol Biot*, 2012, **95**, 1359-1368.
24. M. Iakiviak, R. I. Mackie and I. K. Cann, *Appl. Environ. Microbiol.*, 2011, **77**, 7541-7550.
25. A. Levasseur, E. Drula, V. Lombard, P. M. Coutinho and B. Henrissat, *Biotechnol. Biofuels*, 2013, **6**.
26. C. Herve, A. Rogowski, A. W. Blake, S. E. Marcus, H. J. Gilbert and J. P. Knox, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 15293-15298.
27. A. Lochner, R. J. Giannone, M. Rodriguez, M. B. Shah, J. R. Mielenz, M. Keller, G. Antranikian, D. E. Graham and R. L. Hettich, *Appl. Environ. Microbiol.*, 2011, **77**, 4042-4054.
28. Z. L. Yi, X. Y. Su, V. Revindran, R. I. Mackie and I. Cann, *Plos One*, 2013, **8**.
29. I. Ozdemir, S. E. Blumer-Schuette and R. M. Kelly, *Appl. Environ. Microbiol.*, 2012, **78**, 768-777.
30. B. C. Saha, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 279-291.
31. F. M. Girio, C. Fonseca, F. Carvalheiro, L. C. Duarte, S. Marques and R. Bogel-Lukasik, *Bioresour. Technol.*, 2010, **101**, 4775-4800.
32. T. Collins, C. Gerday and G. Feller, *FEMS Microbiol. Rev.*, 2005, **29**, 3-23.
33. K. Zhang, X. Chen, W. H. Schwarz and F. Li, *Appl. Environ. Microbiol.*, 2014, **80**, 2592-2601.
34. A. L. VanFossen, I. Ozdemir, S. L. Zelin and R. M. Kelly, *Biotechnol. Bioeng.*, 2011, **108**, 1559-1569.
35. D. Guillen, S. Sanchez and R. Rodriguez-Sanoja, *Appl Microbiol Biot*, 2010, **85**, 1241-1249.
36. A. B. Boraston, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Biochem. J.*, 2004, **382**, 769-781.
37. J. Rangasamy, N. Nitar, N. Hideaki, F. Tetsuya and T. Hiroshi, *Macromol. Symp.*, 2008, **264**, 163-167.
38. D. Bhattacharya, A. Nagpure and R. K. Gupta, *Crit. Rev. Biotechnol.*, 2007, **27**, 21-28.
39. A. S. Sahai and M. S. Manocha, *Fems Microbiol Rev*, 1993, **11**, 317-338.
40. H. Shibasaki, K. Uchimura, T. Miura, T. Kobayashi, R. Usami and K. Horikoshi, *Appl Microbiol Biot*, 2014, **98**, 7845-7853.
41. A. Durand, R. Hughes, A. Roussel, R. Flatman, B. Henrissat and N. Juge, *FEBS J.*, 2005, **272**, 1745-1755.

42. B. L. Ridley, M. A. O'Neill and D. A. Mohnen, *Phytochemistry*, 2001, **57**, 929-967.
43. L. M. Blackman, D. P. Cullerne and A. R. Hardham, *Bmc Genomics*, 2014, **15**.
44. P. Dam, I. Kataeva, S. J. Yang, F. F. Zhou, Y. B. Yin, W. C. Chou, F. L. Poole, J. Westpheling, R. Hettich, R. Giannone, D. L. Lewis, R. Kelly, H. J. Gilbert, B. Henrissat, Y. Xu and M. W. W. Adams, *Nucleic Acids Res*, 2011, **39**, 3240-3254.
45. R. Rodriguez-Sanoja, B. Ruiz, J. P. Guyot and S. Sanchez, *Appl. Environ. Microbiol.*, 2005, **71**, 297-302.
46. R. Gupta, P. Gigras, H. Mohapatra, V. K. Goswami and B. Chauhan, *Process Biochem.*, 2003, **38**, 1599-1616.
47. S. K. Gudi, C. Gurramkonda, G. Rather, M. G. S. Chandra, U. K. Mangamuri, S. Podha and Y. L. Choi, *J. Korean Soc. Appl. Biol. Chem.*, 2013, **56**, 427-433.
48. R. S. Singh, G. K. Saini and J. F. Kennedy, *Carbohyd. Polym.*, 2011, **83**, 672-675.
49. N. Planas, *Biochim. Biophys. Acta.*, 2000, **1543**, 361-382.
50. B. W. Zhu, J. G. Zhao, J. F. Yang, T. Mikiro, Z. S. Zhang and D. Y. Zhou, *Process Biochem.*, 2008, **43**, 1102-1106.
51. Y. Ying, Ph.D., University of Chinese Academy of Sciences, 2012.
52. H. Yokoyama, T. Yamashita, R. Morioka and H. Ohmori, *J. Bacteriol.*, 2014, **196**, 3784-3792.
53. J. J. Pei, Q. Pang, L. G. Zhao, S. Fan and H. Shi, *Biotechnol. Biofuels*, 2012, **5**.
54. F. F. Zanoelo, L. Polizeli Mde, H. F. Terenzi and J. A. Jorge, *FEMS Microbiol. Lett.*, 2004, **240**, 137-143.
55. T. R. Yan and C. L. Lin, *Biosci. Biotechnol. Biochem.*, 1997, **61**, 965-970.
56. C. H. Decker, J. Visser and P. Schreier, *Appl Microbiol Biot*, 2001, **55**, 157-163.
57. J. Pei, Q. Pang, L. Zhao, S. Fan and H. Shi, *Biotechnol. Biofuels*, 2012, **5**, 31.
58. K. Riedel, J. Ritter and K. Bronnenmeier, *FEMS Microbiol. Lett.*, 1997, **147**, 239-243.
59. S. J. Yang, I. Kataeva, S. D. Hamilton-Brehm, N. L. Engle, T. J. Tschaplinski, C. Doeppke, M. Davis, J. Westpheling and M. W. Adams, *Appl. Environ. Microbiol.*, 2009, **75**, 4762-4769.

Figure legends:

Fig. 1 Glycoside hydrolases (GHs) involved in deconstruction of polysaccharides in *Caldicellulosiruptor* sp. F32. Csac, *C. saccharolyticus*; Athe, *C. bescii*; COB47, *C. obsidiansis*; Cthe, *Clostridium thermocellum*; Cphy, *Clostridium phytofermentans*. The black circle at the *N*-terminus indicates signal peptide sequence. GH, glycoside hydrolase; CBM, carbohydrate-binding module; SLH, surface layer homology domain. Modular arrangement and length of amino acid sequence for each GH from *Caldicellulosiruptor* sp. F32 are exhibited on the right of the figure. The putative annotation of each GH and gene number of its homolog are exhibited on the left of the figure.

Fig. 2 Enzymatic characteristics of BlgA and CelB from *Caldicellulosiruptor* sp. F32. (A) Hydrolytic activity of BlgA against β-1,4-linked cello-oligosaccharides. G, glucose; C2, cellobiose; C3, cellotriose; C4, cellotetraose; C5, cellopentaose; C6, cellohexaose. (B) The effects of glucose on BlgA activity. (C) Time course analysis of Avicel-hydrolysis products of CelB. Glucose, square; cellobiose, circle. (D) The changes of reducing sugar production from Avicel by synergy of CelB and BlgA. Time course analysis of Avicel hydrolysis by 10 μg CelB alone (square), or synergy with 0.1 μg (circle), 0.5 μg (triangle), and 2 μg (diamond) BlgA, respectively. The degree of synergy (DOS) of BlgA and CelB at various ratios at 36 h hydrolysis is shown as inset.

530 Table 1 Analysis of ORFs from *Caldicellulosiruptor* sp. F32

ORFs		*Caldicellulosiruptor* sp. F32	*C. saccharolyticus* DSM 8903[2]
Total ORFs of strain		2285	2679
Hypothetical protein		358	781
Amylase		8	4
S-layer domain-containing protein		7	8
ABC transporters		22	25
CAZymes	Glycoside hydrolase	44	59
	Polysaccharide lyase	3	1
	Carbohydrate esterase	6	6
	Glycosyl transferase	21	30
	Carbohydrate binding module	25	17

531

532 Table 2 The modules of CAZymes from *Caldicellulosiruptor* sp. F32

CAZymes family	Substrate	Enzyme activity	EC number	number
Glycoside hydrolase (GH)				
GH1	Cellulose	β-glucosidase	3.2.1.21	1
GH2	Hemicellulose	β-mannosidase	3.2.1.25	7
		β-galactosidase	3.2.1.23	
		β-glucuronidase	3.2.1.31	
		α-L-arabinofuranosidase	3.2.1.55	
		Exo-β-glucosaminidase	3.2.1.165	
GH3	Cellulose	β-glucosidase	3.2.1.21	5
	Pectin (RGI)	Exo-β-1,4-glucosidase	3.2.1.74	
	Hemicellulose	Xylan β-1,4-xylosidase	3.2.1.37	
		Glucan β-1,3-glucosidase	3.2.1.58	
		α-L-arabinofuranosidase	3.2.1.55	
GH4	Hemicellulose	α-glucosidase	3.2.1.20	1
		α-galactosidase	3.2.1.22	1
GH5	Cellulose	Cellulase	3.2.1.4	2
		Lichenase	3.2.1.73	1
GH9	Cellulose	Endoglucanase	3.2.1.4	1
GH10	Xylan	Xylanase	3.2.1.8	2
GH11	Xylan	Xylanase	3.2.1.8	1
GH13	Starch	α-amylase	3.2.1.1	5
	Pullulan	Pullulanase	3.2.1.41	2
GH15	Starch	Glucoamylase	3.2.1.3	1
GH16	Hemicellulose	Endo-β-1,4-glucanase	3.2.1.151	1
	β-1,3-glucans	Glucan endo-β-1,3-glucosidase	3.2.1.39	
GH18	Chitin	Chitinase	3.2.1.14	1
		Endo-β-N-acetylglucosaminidase	3.2.1.96	
GH20	Hemicellulose	N-acetylglucosaminidase	3.2.1.30	1
GH26	Mannan	β-mannanase	3.2.1.78	2
GH29	Hemicellulose	α-L-fucosidase	3.2.1.51	1
GH30	Hemicellulose	Glucosylceramidase	3.2.1.45	1
GH31	Starch	α-glucosidase	3.2.1.20	1
	Hemicellulose	α-xylosidase	3.2.1.177	1
GH42	Hemicellulose	β-galactosidase	3.2.1.23	1
GH43	Hemicellulose	α-L-arabinofuranosidase	3.2.1.55	5
		β-xylosidase	3.2.1.37	
GH48	Cellulose	Endo-β-1,4-glucanase	3.2.1.4	1
GH51	Hemicellulose	α-L-arabinofuranosidase	3.2.1.55	1
GH105	Hemicellulose	Rhamnogalacturonyl hydrolase	3.2.1.172	1

Carbohydrate esterase (CE)				
CE4	Hemicellulose	Acetyl xylan esterase	3.1.1.72	4
	N-linked oligosaccharides	Peptidoglycan GlcNAc deacetylase	3.5.1.-	
	Chitin	Chitin deacetylase	3.5.1.41	
CE7	Hemicellulose	Acetyl xylan esterase	3.1.1.72	1
CE9	Hemicellulose	N-acetylglucosamine 6-phosphate deacetylase	3.5.1.25	1
Polysaccharide lyase (PL)				
PL3	Pectin (HG, RGI)	Pectate lyase	4.2.2.2	2
PL11	Pectin	Rhamnogalacturonan lyase	4.2.2.-	1
Carbohydrate binding module (CBM)				
CBM2	Cellulose, chitin, xylan	Cellulose-binding		1
CBM3	Cellulose, chitin	Cellulose-binding		4
CBM4	Glucan	Glucan-binding		5
CBM6	Glucan	Glucan-binding		3
CBM22	Xylan	Xylan-binding		3
CBM27	Mannan	Mannan-binding		2
CBM28	Glucan	Cellulose-binding		1
CBM34	Starch	Starch-binding		1
CBM36	Xylan	Xylan-binding		1
CBM41	Pullulan	Glucan-binding		1
CBM48	Glycogen	Glycogen-binding		2
CBM50	Peptidoglycan	Peptidoglycan-binding		1

533

534 Table 3 Enzymatic properties of BlgA and CelB

Protein	Molecular weight (kDa)	Theoretical *p*I	Optimal temp (°C)	Optimal pH	Specific activity[a]				$t_{1/2}$[b] (h)
					CMC	Avicel	*p*NPC	cellobiose	
BlgA	53.4	5.4	75	5.6	0	0	22.1±0.3	20±0.03	48
CelB	81.3	5.1	75	5.6	17.1±0.4	0.52±0.03	10.4±0.5	-	34

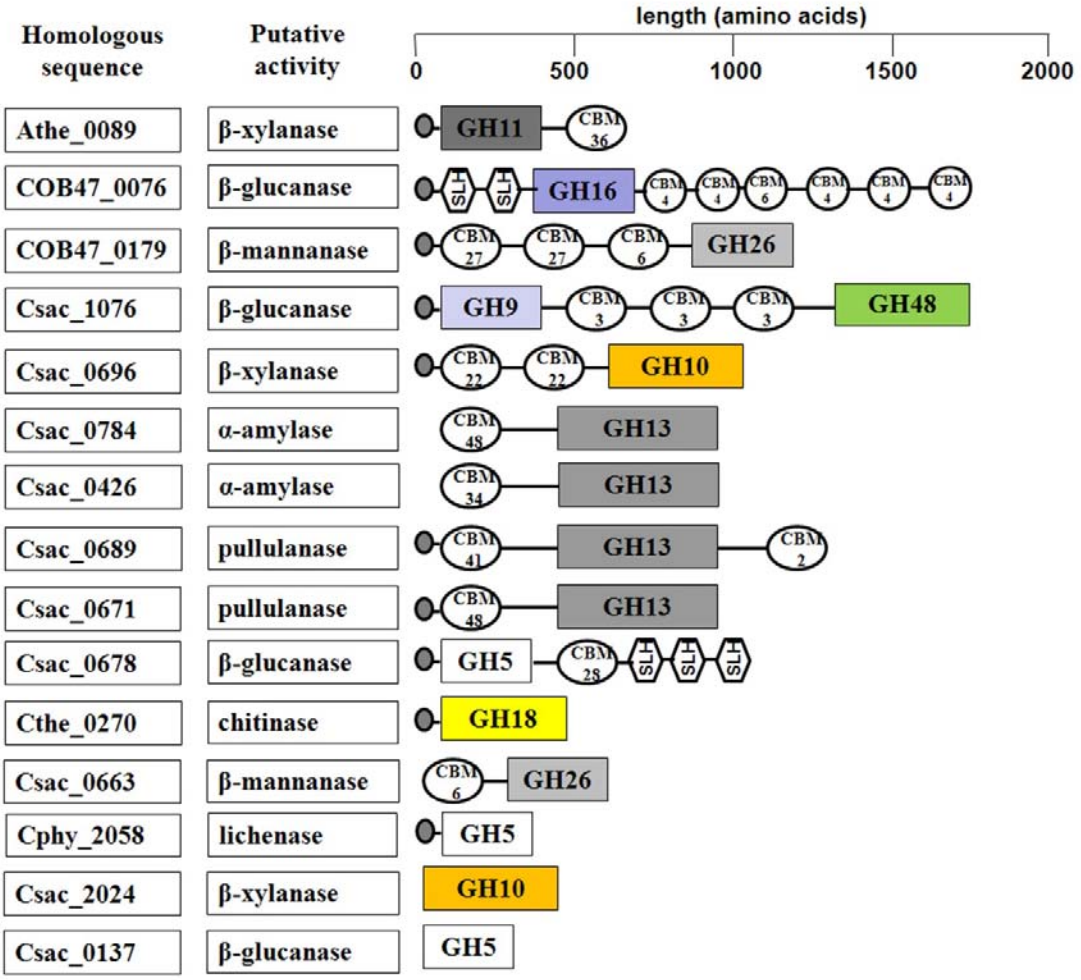
535 [a] The experiments were performed in triplicate, and data are reported as means ± S.D.

536 [b] $t_{1/2}$, half-life of enzyme at optimal temperature.

537 -, not detected.

538

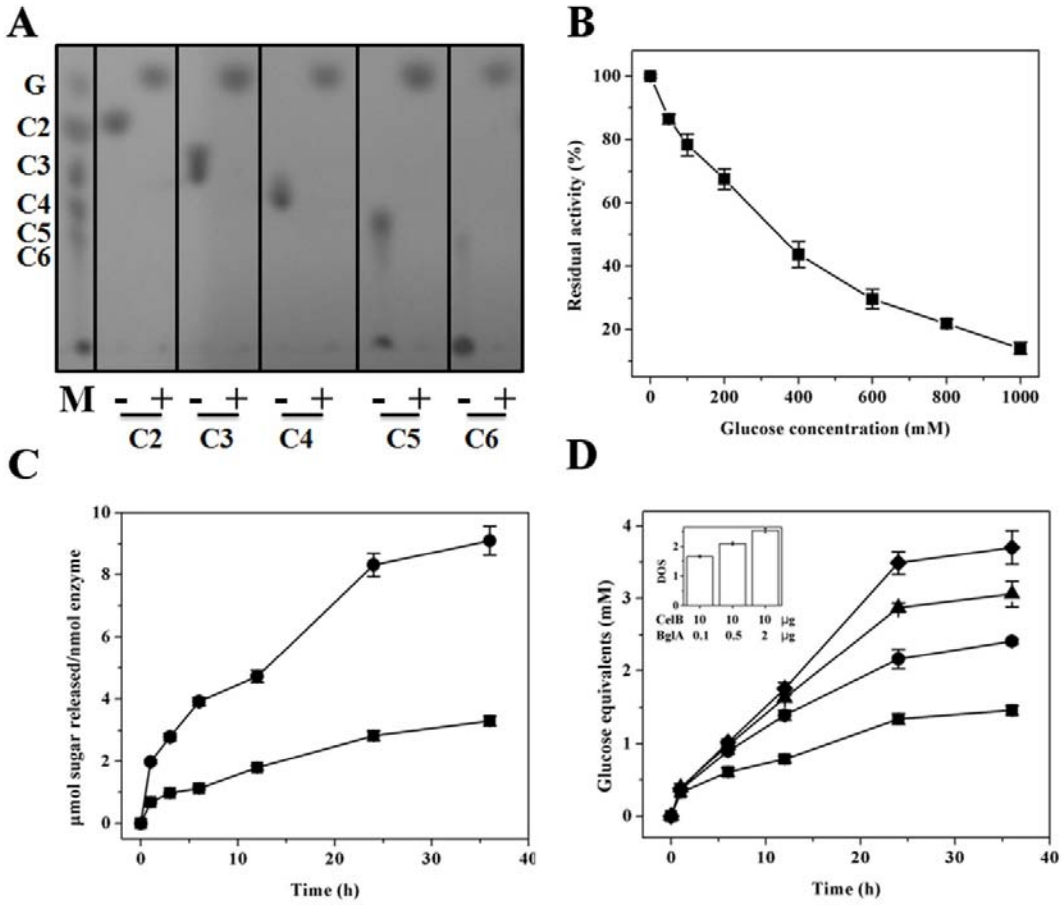
539 Fig. 1



540

541

542 Fig. 2



543