


RESEARCH

Open Access



# Secretomic analyses of *Ruminiclostridium papyrosolvens* reveal its enzymatic basis for lignocellulose degradation

Zhenxing Ren<sup>1,2†</sup>, Wuxin You<sup>3†</sup>, Shasha Wu<sup>1</sup>, Ansgar Poetsch<sup>3,4\*</sup> and Chenggang Xu<sup>1\*</sup> 

## Abstract

**Background:** Efficient biotechnological conversion of lignocellulosic biomass to valuable products, such as transportation biofuels, is ecologically attractive, yet requires substantially improved mechanistic understanding and optimization to become economically feasible. Cellulolytic clostridia, such as *Ruminiclostridium papyrosolvens* (previously *Clostridium papyrosolvens*), produce a wide variety of carbohydrate-active enzymes (CAZymes) including extracellular multienzyme complexes—cellulosomes with different specificities for enhanced cellulosic biomass degradation. Identification of the secretory components, especially CAZymes, during bacterial growth on lignocellulose and their influence on bacterial catalytic capabilities provide insight into construction of potent cellulase systems of cell factories tuned or optimized for the targeted substrate by matching the type and abundance of enzymes and corresponding transporters.

**Results:** In this study, we firstly predicted a total of 174 putative CAZymes from the genome of *R. papyrosolvens*, including 74 cellulosomal components. To explore profile of secreted proteins involved in lignocellulose degradation, we compared the secretomes of *R. papyrosolvens* grown on different substrates using label-free quantitative proteomics. CAZymes, extracellular solute-binding proteins (SBPs) of transport systems and proteins involved in spore formation were enriched in the secretome of corn stover for lignocellulose degradation. Furthermore, compared with free CAZymes, complex CAZymes (cellulosomal components) had larger fluctuations in variety and abundance of enzymes among four carbon sources. In particular, cellulosomal proteins encoded by the *cip-cel* operon and the *xyl-doc* gene cluster had the highest abundance with corn stover as substrate. Analysis of differential expression of CAZymes revealed a substrate-dependent secretion pattern of CAZymes, which was consistent with their catalytic activity from each secretome determined on different cellulosic substrates. The results suggest that the expression of CAZymes is regulated by the type of substrate in the growth medium.

**Conclusions:** In the present study, our results demonstrated the complexity of the lignocellulose degradation systems of *R. papyrosolvens* and showed the potency of its biomass degradation activity. Differential proteomic analyses and activity assays of CAZymes secreted by *R. papyrosolvens* suggested a distinct environment-sensing strategy for cellulose utilization in which *R. papyrosolvens* modulated the composition of the CAZymes, especially cellulosome, according to the degradation state of its natural substrate.

\*Correspondence: [ansgar.poetsch@rub.de](mailto:ansgar.poetsch@rub.de); [xucg@sxu.edu.cn](mailto:xucg@sxu.edu.cn)

†Zhenxing Ren and Wuxin You contributed equally to this work

<sup>1</sup> Key Laboratory of Chemical Biology and Molecular Engineering of Ministry of Education, Institute of Biotechnology, Shanxi University, Taiyuan 030006, Shanxi, China

<sup>3</sup> Department of Plant Biochemistry, Ruhr University Bochum, 44801 Bochum, Germany

Full list of author information is available at the end of the article



**Keywords:** *Ruminiclostridium papyrosolvens*, Secretome, Carbohydrate-active enzymes (CAZymes), Cellulosome, ATP-binding cassette (ABC) transporter

## Background

Lignocellulosic biomass is the most abundant biopolymer on earth, yet its recalcitrance to hydrolysis has severely hampered its exploitation for renewable energy and materials [1–3]. In nature, direct hydrolysis of lignocellulose is carried out exclusively by enzymes secreted by microorganisms. Enzymes involved in the degradation of these polysaccharides are designated carbohydrate-active enzymes (CAZymes) and classified into five distinct groups according to their activities and structural features: glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), glycosyl transferases (GTs) and enzymes with auxiliary activities (AAs), which often display a modular structure with non-catalytic carbohydrate-binding modules (CBMs). The CAZymes classification system has been integrated and meticulously updated in the CAZy database (<http://www.cazy.org>) [4, 5].

Cellulolytic clostridia, which are ubiquitous in cellulose anaerobic environments, represent a major class for efficient biological degradation of cellulosic biomass [6, 7]. Their cellulolytic machinery relies both on cellulosome complexes and on non-cellulosomal free CAZymes [3]. Cellulosome complexes are extracellular multi-enzyme machineries produced by numerous anaerobic and cellulolytic microorganisms, which consist of a non-catalytic multi-functional integrating subunit (called scaffoldin), responsible for organizing the various catalytic subunits into the complex [8, 9]. The integration is accomplished by the interaction of two complementary module classes, i.e., a cohesin module on the scaffoldin and a dockerin module on each enzymatic subunit [10]. These specific characteristics allow the cellulosome to degrade cellulosic substrates effectively. The host cells and their substrate degradation machineries [11–13] are being exploited in the production of cellulosic biofuels by a variety of approaches, notably consolidated bioprocessing (CBP; [14]).

As an important model for mesophilic anaerobic cellulolytic bacteria, for *Ruminiclostridium cellulolyticum* (previously *Clostridium cellulolyticum*) cellulolytic machinery expression and regulation of metabolism on cellulose and its derivatives have been widely investigated by employing transcriptomics [15, 16] and proteomics [17, 18]. It is known that the expression of CAZymes, including proteins of the cellulosome, is regulated by the type of substrate [15, 17]. However, *Ruminiclostridium papyrosolvens*, essentially the most derived species

among known mesophilic cellulolytic clostridia [16], has not yet been studied systematically because of a previous lack of genetic information and transformation method. Recently, genome sequences of two strains (DSM 2782 and C7) of *R. papyrosolvens* have been published [19, 20] and its transformation method has been successfully developed [21], which opens the door to thorough research of *R. papyrosolvens*.

In the present study, to achieve global insight into the cellulolytic machinery of *R. papyrosolvens* DSM 2782, we cultivated *R. papyrosolvens* on four different sources of carbon: glucose, cellobiose, microcrystalline cellulose and corn stover. Thereupon, extracellular proteomes (secretomes) were quantitatively compared by label-free liquid chromatography–tandem mass spectrometry (LC–MS/MS), and their catalytic activities for degradation of different substrates were examined. The results demonstrated that *R. papyrosolvens* modulates composition and abundance of its extracellular enzymes and transporters according to the growth substrate.

## Results

### Genomic features of cellulose degradation for *R.*

#### *papyrosolvens*

The draft genome of *R. papyrosolvens* DSM 2782 consists of 31 contigs with a GC content of 37.0% and a total length of 4,915,287 bp. It encodes 4039 proteins, 57 tRNAs and 18 rRNAs; 100-mL cultures grown on various carbon sources were harvested when the concentration of extracellular proteins reached the maximum (Additional file 1: Figure S1). After centrifuged (12,000g, 4 °C, 30 min), the supernatants were aspirated and filtered through a 0.22- $\mu$ m PES membrane (Jinteng, Tianjin, China). The residues after centrifugation of the cultures were washed and centrifuged for three times with 5 mL EDTA buffer (50 mM Tris–HCl, 5 mM EDTA, pH8.0), with eluates collected to obtain the proteins binding to the cellulose materials. Cell-free supernatants and the filtered eluates were pooled together and concentrated 100-fold using an ultrafiltration device containing a 10-kDa-cutoff membrane (Millipore, Germany). Protein concentrations were measured by the BCA assay (Sangon Biotech, Shanghai, China). The isolated samples were boiled for 5 min at 100 °C and loaded onto 12.5% SDS-PAGE. Proteins were visualized with a Coomassie Brilliant Blue (CBB-G250) stain as described by Dyballa and Metzger [52] (Additional file 2: Table S1; GenBank Accession Number NZ\_ACXX00000000.2;

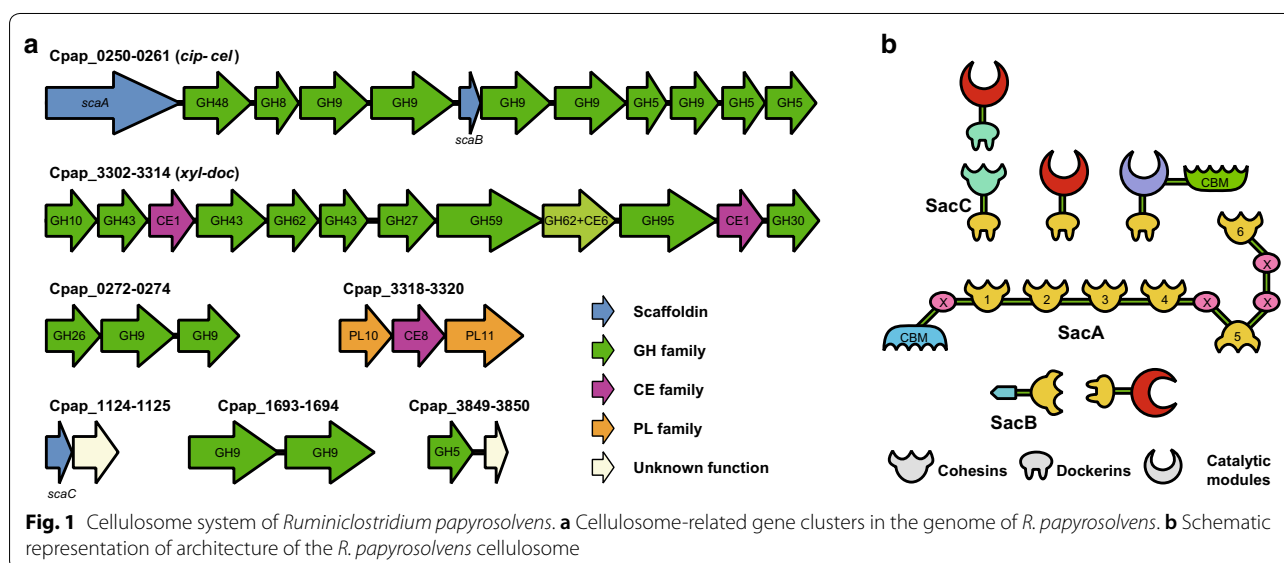
[19]). CAZymes were predicted by HMMER3.0 (<http://hmmer.org/>) [22] according to dbCAN database [23] definition: a total of 174 CAZyme genes were annotated for *R. papyrosolvans* DSM 2782 genome, including 106 GHs, 3 PLs, 34 CEs and 64 CBM-harboring proteins, in which more than 65% of genes are orthologous to *R. cellulolyticum* (Additional file 3: Table S2). Furthermore, among these CAZyme genes, there are 74 putative cellulosomal subunit-encoding genes, including 71 dockerin-encoding genes and 3 cohesin-encoding genes, which amounts to more than the 65 cellulosomal subunits of *R. cellulolyticum* [15, 17]. However, among them 57 orthologous cellulosomal genes are shared by *R. papyrosolvans* and *R. cellulolyticum* (Additional file 3: Table S2). Thus, it is suggested that *R. papyrosolvans* has evolved a very sophisticated cellulolytic system, which has remarkable orthologous relationships with *R. cellulolyticum* [24].

Cellulosomal genes from mesophilic clostridia tend to physically cluster along the chromosome [8], and *R. papyrosolvans* is no exception. Among the 74 cellulosomal genes in total, we identified seven clusters (Fig. 1a): (i) the “*cip-cel*” gene cluster of 12 genes (Cpap\_0250-0261) that encodes the major cellulosome components, including two cohesin-harboring scaffoldins, respectively, named ScaA and ScaB. SacA is composed of six cohesin domains of type I numbered from 1 to 6 from the N- to the C-terminus, in addition to a N-terminal cellulose-binding module (CBM) and four X modules, separated by short linker sequences, while SacB only harbors a cohesin domain in its C-terminus (Fig. 1b); (ii) the “*xyl-doc*” cluster of 12 genes (Cpap\_3302-3314) encoding exclusively secreted dockerin-containing proteins, which are probably involved in hemicellulose degradation

and herein named the *xyl-doc* gene cluster [17]; (iii) a couple of genes arranging another cohesin–dockerin interaction (named type II to distinguish from primary cohesin–dockerin interaction of type I, Cpap\_1124-1125), in which Cpap\_1124 encodes a type II cohesin domain along with the type I dockerin (named *scaC*), while Cpap\_1125 encodes a type II dockerin, resulting that it is assembled into Cpap\_1124-encoded type II cohesin, and then attached to the primary scaffoldin by mediation of type I dockerin of Cpap\_1124 (Fig. 1b); (iv) other small clusters (two or three genes) encoding cellulosomal enzymes (Cpap\_0272-0274, Cpap\_1693-1694, Cpap\_3318-3320 and Cpap\_3849-3850), in which two clusters, (Cpap\_0272-0274 and Cpap\_3318-3320), respectively, are located downstream of the *cip-cel* and *xyl-doc* clusters.

### Structure of secretome profiles in *R. papyrosolvans*

To identify the components of the cellulose degradation in *R. papyrosolvans*, we started by characterizing the populations of extracellular proteins in *R. papyrosolvans* cultures under a variety of carbon sources using proteomics with label-free quantitation. The carbohydrate substrates tested included cellulose and its derivatives glucose and cellobiose, and corn stover, the natural plant-derived lignocellulose. Growth was assessed by monitoring protein concentration of fermentation supernatants (Additional file 1: Figure S1). Proteins secreted to fermentation supernatants were concentrated by ultrafiltration when their concentrations reached the maximum and then analyzed by SDS-PAGE (Additional file 4: Figure S2). Total extracellular proteins isolated from each of the four growth conditions in three biological replicates were analyzed by



label-free LC–MS/MS. Proteins were identified using *R. papyrosolvans* protein sequence data from NCBI. In total, 1151 Protein sequences were found in the secretome (i.e., around 28% of the total 4172 proteins encoded in *R. papyrosolvans* genome).

Proteomes were obtained for three biological replicates under each carbon source. The results of PCA analysis indicated that the respective three biological replicates always closely cluster in the first and second components. However, there are much bigger distances between samples from different substrates, suggesting that the secretomes were significantly distinguished by their substrates (Additional file 5: Figure S3). In this study, a reliably expressed protein was defined as being identified in at least two biological replicates for one certain carbon source. Based on this principle, 912 proteins in total were found to be expressed in our experiments, of which 218 proteins were predicted to harbor signal peptides using SignalP5.0 [25] (Additional file 6: Table S3). Among them, 188, 170, 187 and 191 proteins were, respectively, identified under glucose, cellobiose, cellulose and corn stover; 149 proteins were shared among all four conditions being the core components of secretome, whereas the 22 exclusive proteins are most likely linked to the respective substrate (Fig. 2a, Additional file 7: Table S4).

To probe the functional implication of secretomes, 218 proteins with signal peptides were examined for association with functional categories as defined by Cluster of Orthologous Group (COG [[ftp.ncbi.nih.gov/pub/COG/COG2014/static/lists/homeCOGs.html](http://ftp.ncbi.nih.gov/pub/COG/COG2014/static/lists/homeCOGs.html)]; Additional file 7: Table S4). A broad spectrum of COGs was found, with these proteins being enriched in COG categories such as G (carbohydrate transport and metabolism), X (Mobilome: prophages, transposons) and R (General function prediction only) ( $P$ -value < 0.05, hypergeometric test; Fig. 2b). Furthermore, expression of proteins in COG category of G with the highest degree of enrichment was compared among four secretomes (Fig. 2c). The box plot showed that expression of proteins in G category was significantly different between any two substrates except glucose and cellulose, and among them, expression between cellobiose and corn stover produced the highest difference (Fig. 2c). Thus, it is suggested that *R. papyrosolvans* secretes different proteins related to carbohydrate transport and metabolism according to its growth substrates. Interestingly, expression pattern of these proteins in glucose appears to be similar to cellulose.

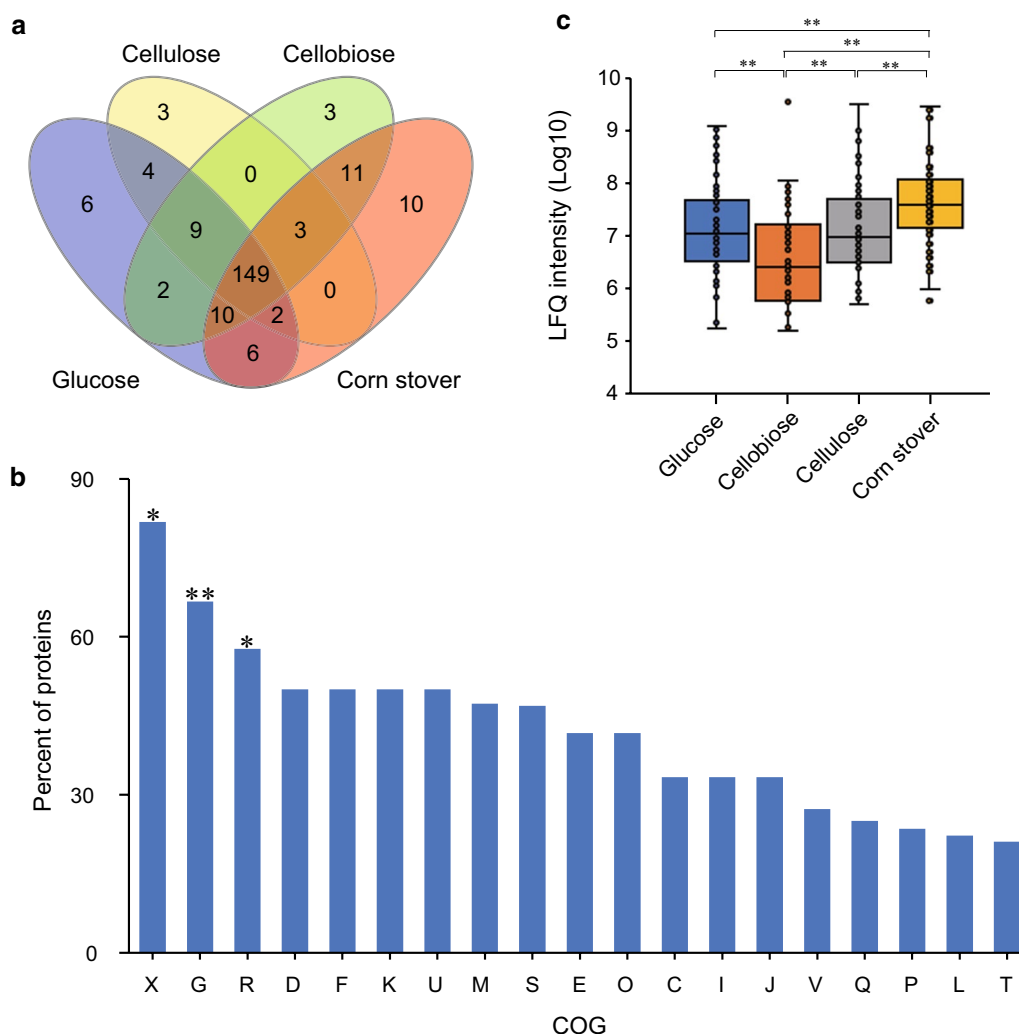
We then compared secretomic profiles between cellobiose and corn stover, which are, respectively, considered as easily and difficultly consumed carbon sources. It is observed that there are more high-abundant, up-regulated proteins under corn stover in comparison with

cellobiose (Fig. 3a). To determine the clustering function of these up-regulated proteins under corn stover, we applied a network analysis using STRING [26]. The result showed that up-regulated proteins displayed functionally distinct clusters. Compared with cellobiose, *R. papyrosolvans* consuming corn stover had increased expression of CAZymes including cellosomal components for lignocellulose degradation, extracellular SBPs of ATP-binding cassette (ABC) transporters and proteins involved in spore formation (Fig. 3b). We concluded that this proteomic evidence was consistent with corn stover exhibiting a complex structure composed of cellulose, hemicellulose, lignin and other soluble sugars. For utilization of corn stover, cells need to secrete an arsenal of degradative CAZymes and ABC transporters. Meanwhile, corn stover can be regarded as adverse environmental condition compared with cellobiose, triggering expression of proteins involved in sporulation in *R. papyrosolvans*. The similar phenomenon had been reported in *R. thermocellum* and *R. cellulolyticum*. Attachment to cellulose fibers could trigger sporulation in *R. thermocellum* [27], while Spo0A mutant of *R. cellulolyticum* abolished the sporulation ability and increased dramatically cellulose catabolism [28].

#### Expression pattern of CAZymes

To further probe the links among the substrate-specific proteins being implicated in plant cell wall degradation, we firstly counted the types and numbers of expressed CAZymes under the different substrates. In total, 116 of 174 CAZymes encoded by *R. papyrosolvans* genome were retrieved in its secretome. These expressed CAZymes were mainly distributed in COG categories of G, R and M. In particular, more than half (70) of CAZymes belonged to G category, which accounted for 51% of all proteins in this category. Furthermore, these secreted CAZymes included 67 cellosomal subunits and 49 non-cellosomal CAZymes, respectively, accounting for 91% and 49% of their genome-predicted proteins. Moreover, the abundance of non-cellosomal CAZymes and cellosomal subunits, respectively, accounted for 2.4% and 10.1% of the total abundance of secretome.

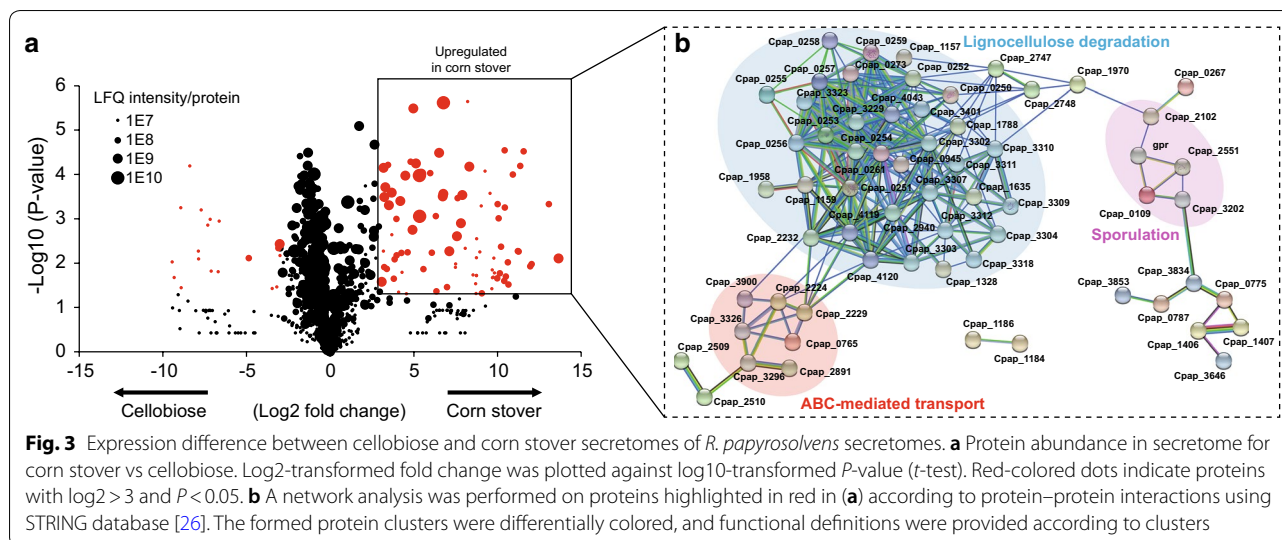
Out of the in total 116 expressed CAZymes, 101, 89, 101 and 111 CAZymes were, respectively, identified under glucose, cellobiose, cellulose and corn stover, demonstrating that cells secreted the most CAZymes, cellosomal components and CBM-harboring proteins when grown on corn stover (Fig. 4a). Interestingly, *R. papyrosolvans* expressed more CAZymes, cellosomal components and CBM-harboring proteins under glucose than cellobiose. Meanwhile, results of growth curves indicated that *R. papyrosolvans* preferred cellobiose over glucose (Additional file 1: Figure S1). These results on *R.*



**Fig. 2** Label-free LC-MS/MS analysis of *R. papyrosolvans* secretomes. **a** Venn diagram of total protein numbers for *R. papyrosolvans* secretomes in glucose, cellobiose, cellulose, and corn stover media. **b** Functional profile of proteins identified in secretomes. Percentages of such proteins in each COG term were shown in columns (X, mobilome: prophages, transposons; G, carbohydrate transport and metabolism; R, general function prediction only; D, cell cycle control, cell division, chromosome partitioning; F, nucleotide transport and metabolism; K, transcription; U, intracellular trafficking, secretion, and vesicular transport; M, cell wall/membrane/envelope biogenesis; S, function unknown; E, amino acid transport and metabolism; O, posttranslational modification, protein turnover, chaperones; C, energy production and conversion; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; V, defense mechanisms; Q, secondary metabolites biosynthesis, transport and catabolism; P, inorganic ion transport and metabolism; L, replication, recombination and repair; T, signal transduction mechanisms). Proteins identified were enriched in G, X and R of COG categories (\* $P < 0.05$ , \*\* $P < 0.01$ , hypergeometric test). **c** Box plot of differential expression of COG category of G in four secretomes under four carbon sources (\* $P < 0.05$ , \*\* $P < 0.01$ ,  $t$ -test)

*papyrosolvans* are completely consistent with our previous findings on *R. cellulolyticum*, suggesting activation of cellulase expression by non-preferred carbon source (i.e., glucose) and inhibition by a preferred substrate (i.e., cellobiose), which can be explained by the carbon catabolite repression (CCR) mechanism [15]. Furthermore, qualitative and quantitative differences of CAZymes among four carbon sources were mainly from cellulosomal components. For example, the number of cellulosomal

components increased from 51 in cellobiose to 66 in corn stover. CAZymes compared between cellulose and corn stover comprised nearly equal numbers of GH, CE and GT family enzymes (Fig. 4a). On the other hand, the expression of free CAZymes had no difference among four substrates (except between cellobiose and cellulose) with nearly equal average abundances, but expression of cellulosomal CAZymes was significantly different between any two substrates and the average abundances

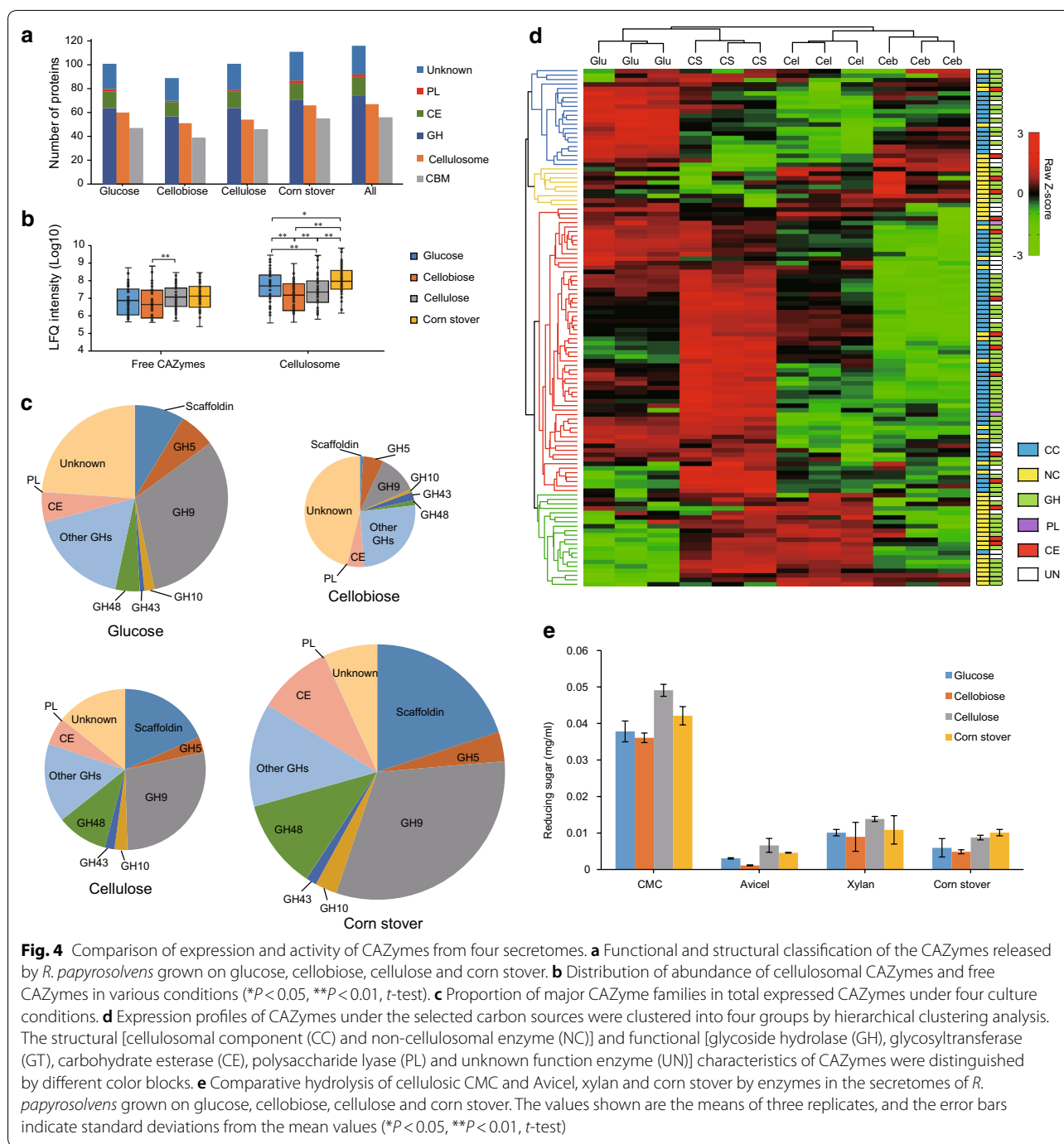


of cellulosomal subunits had the same tendency of changes with its number among four carbon sources, i.e., the cellulosome had the highest relative abundance under corn stover and the lowest expression abundance under cellobiose (Fig. 4b). This indicated that additional cellulosomal components with high expression level are required for degradation of the more structurally complex substrate lignocellulose.

It has been proved that scaffoldin, GH48 and GH9 family enzymes are the most abundant components of cellulosome, and GH5, GH10, GH43, CE and PL family enzymes are responsible for degradation of hemicellulose [9, 17]. Next, we investigated the proportion of these major CAZymes in total CAZymes under various carbon sources (Fig. 4c). The results indicated the largest portion of CAZymes in secretome was from corn stover, but the smallest was from cellobiose among four carbon sources. On the other hand, the proportion of key cellulosomal subunits (scaffoldin, GH48 and GH9 family proteins) and CE family proteins significantly increased with increasing proportion of CAZymes in secretome. For example, the proportion of scaffoldin increased from 1% in cellobiose to 20% in corn stover, while that of proteins with unknown function decreased from 46 to 7%. However, the proportion of GH5, GH10 and GH43 family proteins had no difference under four culture conditions. Thus, we proposed that these major cellulosomal subunits, especially scaffoldin, GH48 and GH9 family, were the core components of degradation system of lignocellulose, and CE and GH5, GH10 and GH43 family proteins also play important roles in hydrolysis of plant wall polysaccharides.

Based on their substrate-dependent expression patterns, the 116 CAZymes were clustered into four different

groups (Fig. 4d, Additional file 8: Table S5). Group 1 (blue) included 21 CAZymes that showed the highest relative abundance under glucose, which belong to GH families and proteins of unknown function harboring a CBM or dockerin domain. Most of this group of CAZymes were also cellulosomal components, of which 15 proteins harbor the dockerin domain. Group 2 (orange) included 9 CAZymes that showed higher relative abundance under cellobiose and cellulose than glucose and corn stover. CAZymes of this group were free CAZymes. Group 3 (red) was the largest group including 64 CAZymes that showed the highest relative abundance under corn stover and the lowest relative abundance under cellobiose. CAZymes of this group were mainly cellulosomal components, of which 50 proteins harbor dockerin or cohesin domains. Surprisingly, most of the cellulosomal subunits encoded by the *cip-cel* and *xyl-doc* gene clusters belonged to this group, suggesting that cellulosomal CAZymes, especially the key subunits encoded by the two large gene clusters, were the primary degraders of lignocellulose, such as corn stover. Group 4 (green) included 22 proteins that showed higher relative abundance under cellulose and corn stover than soluble sugars glucose and cellobiose. Proteins of this group were mostly free CAZymes. Thus, we found clear distinction on expression pattern between free CAZymes and cellulosomal components, suggesting that they could be controlled by different and likely independent mechanisms. Furthermore, the expressions of all the 116 CAZymes demonstrated a negative correlation between cellobiose and corn stover, with a high correlation coefficient ( $R=0.67$ ) (Additional file 9: Figure S4). Thus, it is suggested that the more difficult the substrate to be utilized, the more types and abundances of CAZymes are secreted by the bacterium.



To test the correlation between expression of CAZymes and their catalytic activity, we examined these four isolated extracellular samples for their degradation of four polysaccharide substrates: carboxymethyl cellulose (CMC), microcrystalline cellulose (Avicel), xylan and corn stover by the measurement of released total reducing sugars using the DNS method [29] (Fig. 4e). The results indicated that samples isolated from cellulose and

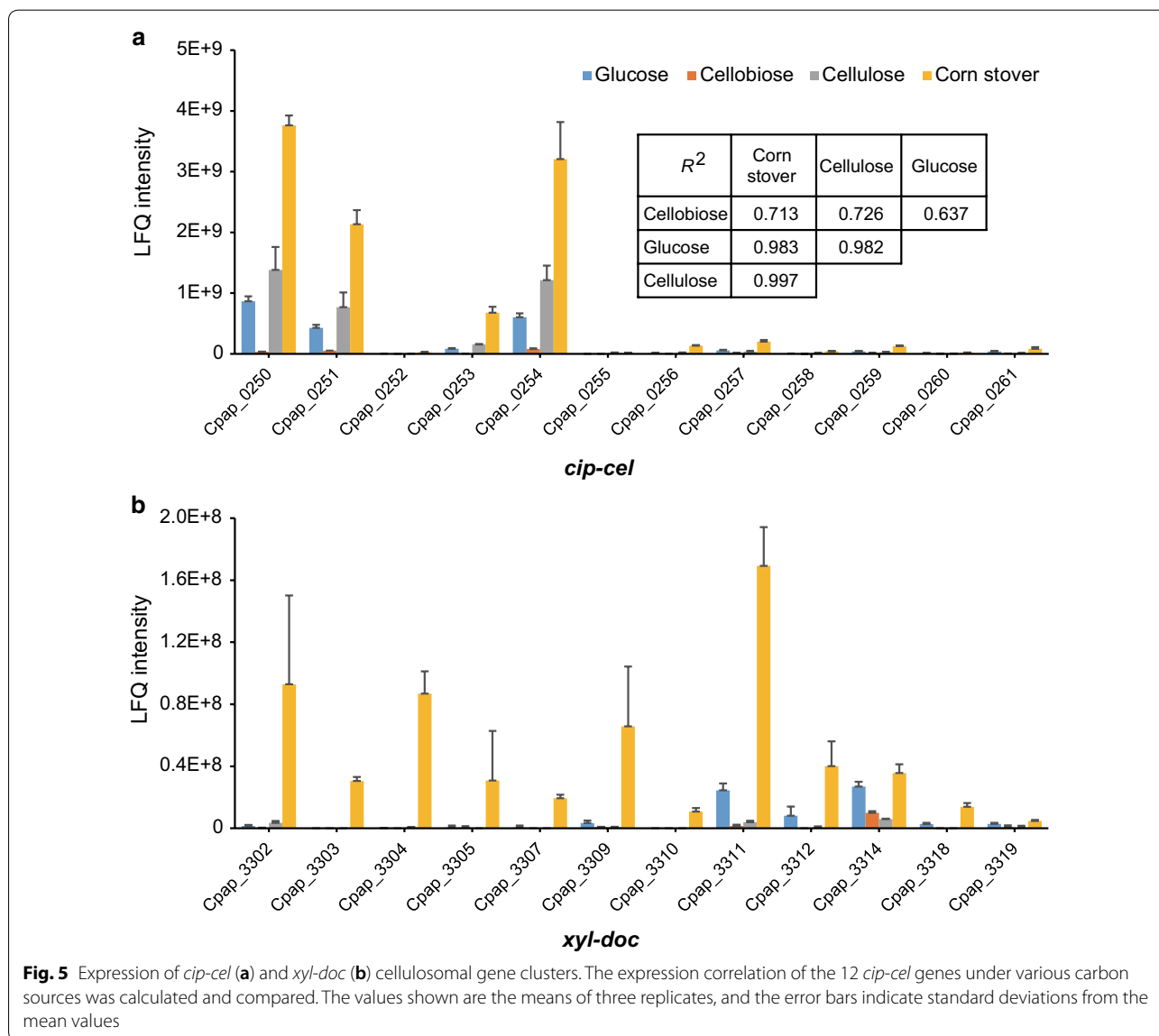
corn stover were more efficient on all polysaccharide substrates, respectively, while samples from cellobiose had the lowest activities. Samples from cellulose were most efficient to hydrolyze CMC and Avicel, and the most active samples on corn stover were samples isolated from corn stover. Thus, the catalytic activity of CAZymes has a strong link to supplied substrate, in which cells produce the optimal formulation of CAZymes.

### Expression of key CAZymes and ABC transporters

We first compared expression of cellulosomal subunits encoded by both of the *cip-cel* and *xyl-doc* clusters among different carbon sources (Fig. 5). The proteomic data reveal that expression of the *cip-cel* cluster appears to be negatively correlated with availability or preference of carbon sources due to the order of the average expression level of *cip-cel* (corn stover > cellulose > glucose > cellobiose). On the other hand, 12 proteins of *cip-cel* exhibit highly uneven abundance under all the four carbon sources tested, where the observed relative abundance of the first (Cpap\_0250, encoding the scaffoldin ScaA), second (Cpap\_0251, encoding an exoglucanase of GH48) and fifth (Cpap\_0254, encoding an endoglucanase of GH9) genes was far higher than the other genes in the

cluster (Fig. 5a). Interestingly, the ratio among 12 proteins of *cip-cel* was independent of the carbon sources, as demonstrated by the high correlations of abundance of *cip-cel* proteins among four carbon sources ( $R^2$  ranged from 0.637 to 0.997) (Fig. 5a). These results of *cip-cel* in *R. papyrosolvens* are in remarkably agreement with those in *R. cellulolyticum* in which expression difference of *cip-cel* was caused by the mechanism of selective RNA processing and stabilization (SRPS) in post-transcriptional level [15, 16], suggesting that the *cip-cel* cluster is not only conserved in protein-coding sequences, but also in regulation mechanism between *R. papyrosolvens* and *R. cellulolyticum*.

Like the *cip-cel* cluster, the *xyl-doc* cluster had the highest relative abundance under corn stover with highly





uneven abundance but was hardly expressed under other three carbon sources, except that the last five genes were expressed under glucose (Fig. 5b). It is fully consistent with our previous study on analysis of the promoter activity upstream of *xyl-doc* [21] and other group's study on transcription analysis of *xyl-doc* from *R. cellulolyticum* [30]. In short, the *cip-cel* and *xyl-doc* clusters from *C. papyrosolvans* were expressed in a manner very similar to that from *R. cellulolyticum*, implying that they may be regulated by many mechanisms that had been proved in *R. cellulolyticum*, such as CCR [15, 31], two-component system (TCS) [15, 30, 32] and SRPS [16].

In addition to CAZymes for degradation of lignocelluloses, bacteria are required to employ sugar transporters to transport lignocellulosic degradation products into cells (Fig. 3b). Thus, we further analyzed the expression of SBPs (extracellular subunits of ABC transporters) from secretome. In total 16 SBPs were found in our secretome, which was categorized into groups based on abundance trend similarity (Fig. 6a). It shows that some SBPs (Cpap\_0690, Cpap\_0704 and Cpap\_0906) were expressed at extremely low level among all substrates, while Cpap\_0701 had high relative abundance in all substrates. And others are expressed differentially among four carbon sources.

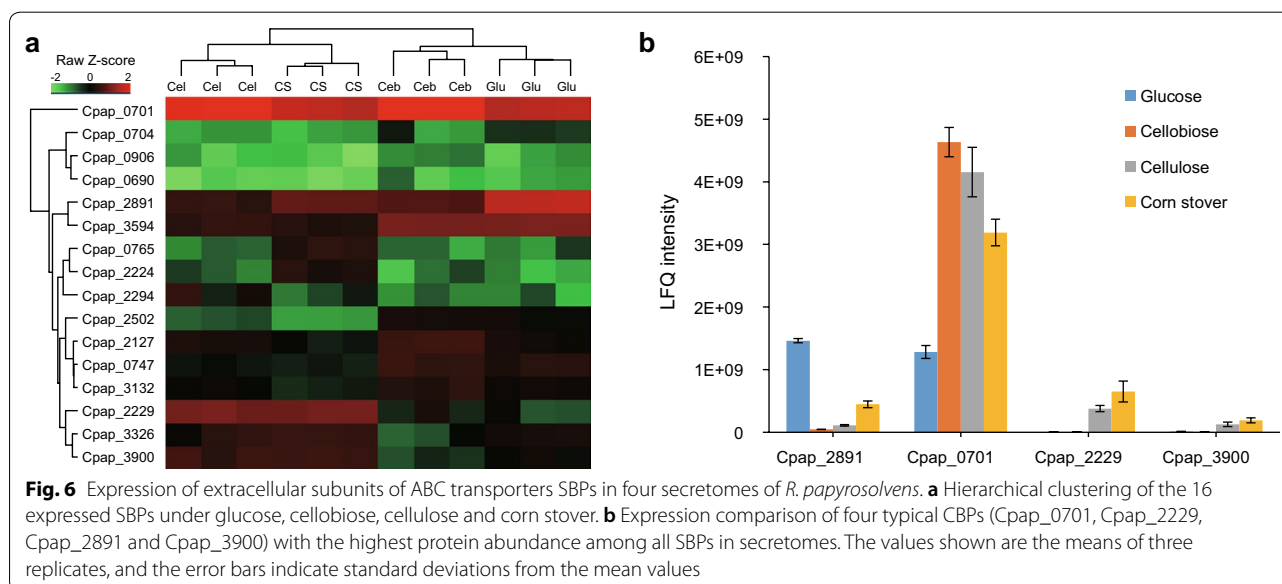
We further analyzed the expression pattern of four CBPs (Cpap\_0701, Cpap\_2229, Cpap\_2891 and Cpap\_3900) with the highest relative abundance (Fig. 6b). Cpap\_2891 had the highest relative abundance on glucose, moderate relative abundance on corn stover and low relative abundance on cellobiose and cellulose, suggesting that Cpap\_2891 was potentially in charge of

glucose transportation. Both Cpap\_2229 and Cpap\_3900 had high relative abundance on cellulose and corn stover, and almost no expression on glucose and cellobiose, suggesting that they play an important role in transportation of oligosaccharides.

It is worth noting that Cpap\_0701 accounted for the most mass of SBP elements, which made up more than 70% of all SBPs in secretome (Additional file 6: Table S3). It had much higher relative abundance on cellobiose, cellulose and corn stover than glucose (Fig. 6b), with same expression pattern as its ortholog CauA (Ccel\_2112, sequence 94% identity) from *R. cellulolyticum*. CauA-harboring ABC transporter in *R. cellulolyticum* is provably devoted to cellobiose and cellobioses uptake and regulated by its upstream two-component system [15, 32]. This suggests that the ABC transporter harboring Cpap\_0701 also facilitates an influx of cellobiose and cellobioses in *R. papyrosolvans*.

## Discussion

This study explores the complexity of the plant cell wall degradation system of *R. papyrosolvans*. Whole-genome analysis of *R. papyrosolvans* revealed a 174-CAZyme repertoire of 106 GHs, 34 CEs, 3 PLs and 64 CBMs, including 74 cellulosomal components, suggesting diversity and substrate adaptation in enzymatic activity. The most abundant GH families were GH5, GH9 and GH43, which constituted over 50% of the enzymatic domains identified. Our results were slightly different from the previous CAZyme prediction in *R. papyrosolvans* [24], in which there are 127 CAZymes including 103 GHs, 19 CEs, 5 PLs, 67 CBMs and 70 cellulosomal components. This



may be due to the different analytical methods and versions of genome annotation. However, both of CAZyme predictions in *R. papyrosolvans* revealed that CAZymes of *R. papyrosolvans* represented a notable increase compared to CAZymes observed in *R. cellulolyticum* [19, 24]. In addition to 115 orthologs of *R. cellulolyticum*, *R. papyrosolvans* evolved more particular enzymes to degrade unusual polysaccharides to adapt to the complex environment, such as GH31, GH39, GH109 and GT and CE families. The differences in numbers may be attributed to the size of the genomes, which is 4.92 Mb for *R. papyrosolvans* and 4.07 Mb for *R. cellulolyticum* [24].

Cellulosome architectures vary greatly among the cellulosome-producing bacterial species, but two global types of architectures of cellulosome systems have been observed, namely simple and complex [8]. Simple cellulosomes have so far been observed in mesophilic clostridial species, such as *Ruminiclostridium cellulovorans* [33], *R. cellulolyticum* and *R. josui* [34]. The simple cellulosome architecture includes a single scaffoldin protein, encoded by *cip-cel* cluster, comprising the primary scaffoldin gene followed downstream by a series of genes encoding for various dockerin-bearing enzymes. On the other hand, complex cellulosome systems contain multiple scaffoldin proteins. The major scaffoldin genes are clustered in the genome in a gene cluster. Complex cellulosomes have been observed in *Ruminiclostridium thermocellum* [35], *Bacteroides cellulosolvans* [36], *Acetivibrio cellulolyticus* [37] and *Ruminococcus flavefaciens* [38].

In addition to the *cip-cel* cluster, it was found that *R. papyrosolvans* harbors another two-gene cluster (Cpap\_1124-1125), encoding type II of scaffoldin protein and dockerin, distinguished from primary cohesin–dockerin interaction of *cip-cel* (Fig. 1b). Thus, *R. papyrosolvans* does not only possess the classic simple cellulosomes as mesophilic clostridia, but also has started to evolve the type II of cohesin–dockerin interaction-forming complex cellulosomes. Its architecture is constructed by two types of cohesin–dockerin interaction of scaffoldin and enzymes, which are, respectively, encoded in two-gene clusters comprising a scaffoldin gene followed downstream by enzyme genes (Fig. 1a). It is very different from the complex cellulosomes from *R. thermocellum*, in which type II of cohesin–dockerin interaction happens between two scaffoldins, but not between scaffoldins and enzymes, as they do in cellulosomes from *R. papyrosolvans*. The diversity of CAZymes and cellulosome architectures suggests that the various individual mesophilic clostridial species have evolved several specific strategies for carbohydrate degradation, some similar to, but others distinct from those of their intimate relatives.

To identify the particular proteins that were actually produced under specific substrate conditions including

soluble glucose and cellobiose and insoluble cellulose and corn stover, the extracellular secretomes were analyzed by label-free quantification (LFQ) proteomic method in MaxQuant, allowing inter- and intra-experiment comparison of relative protein abundances. However, there is a challenge with isolation of secretomes from insoluble substrates. Genome of *R. papyrosolvans* encodes 64 CBM-harboring proteins including cellulosomal scaffoldin (Additional file 3: Table S2), which potentially bind to insoluble polysaccharides via their CBM. Thus, cellulases harboring CBM might be underestimated in the secretomes, if they are not recovered from their insoluble substrates prior to quantitative proteomics. To improve recovery of these proteins, two alternative strategies may be adopted: one is that samples are collected after exhaustion of insoluble substrates, as binding proteins would be released then. However, such nutrient-depletion stress should affect the secreted proteome—in the worst case intracellular proteins are released due to cell death and confound results. This has been described for secretomes from bacteria at stationary growth phase (e.g., Indrelid et al. [39]). Thus, another approach is concentrating culture supernatants from insoluble substrates by ideally quantitative washes/extraction. The latter procedure has been successfully adopted in many cellulolytic microorganisms [40–43]. On the other hand, studying structure of a family 3 CBM from the cellulosomal scaffoldin subunit of *R. thermocellum* revealed that it harbors a calcium binding loop, whose interaction with  $\text{Ca}^{2+}$  modulates the mechanostability of CBM [44, 45]. Thus, in order to ensure the completeness of our secretomes, we washed the residual cellulose materials using EDTA buffer to chelate  $\text{Ca}^{2+}$ , thus releasing cellulose-binding proteins as much as possible. We estimated the elution effect of EDTA buffer for cellulose-binding proteins. The results indicated that EDTA buffer was able to elute protein from cellulose effectively, except very few individual proteins (Additional file 10: Figure S5). According to quantitative image analysis of the stained gel, this approach recovered 99% of the secreted proteins. Therefore, the method employed here avoided secretome interference from dying cells while providing quantitative recovery for the majority of insoluble substrate-bound proteins. Admittedly, incomplete removal of individual proteins from insoluble substrates could adversely affect their quantitative comparison with soluble substrates.

In this study, in total 917 proteins were identified in four different sets of secretomes isolated from *R. papyrosolvans* and the LFQ intensity ranged from  $4.65 \times 10^5$  to  $4.56 \times 10^9$ . Compared to proteomic data of *R. cellulolyticum* [17] and *R. thermocellum* [46], our experiments detected more cellulosomal proteins in more complex samples in which cellulosomes were not purified and

enriched. The LFQ protein data revealed that many extracellular proteins, and especially CAZymes, were differentially expressed among various carbon sources. As expected, *R. papyrosolvans* like many cellulolytic bacteria, such as *R. cellulolyticum* [15, 17] and *R. thermocellum* [46], expressed 27% more CAZymes on recalcitrant substrate corn stover than on its preferred sugar cellobiose. However, *R. papyrosolvans* expressed more CAZymes on glucose than cellobiose, being consistent with our previous transcriptomic data for *R. cellulolyticum* [15]. Whereas *R. acetobutylicum* [47] and *R. cellulovorans* [48], the mesophilic phylogenetic relatives of *R. papyrosolvans*, prefer glucose; in the latter, cellulases were transcriptionally repressed under glucose, but derepressed upon glucose exhaustion [48]. Thus, cellulolytic bacteria seem to activate expression of CAZyme by the non-preferred carbon sources and inhibit expression in the preferred substrates, a fact that can be explained by the carbon catabolic repression mechanism (CCR) [15]. Furthermore, cellobiose as the preferred sugar of cellulolytic bacteria, such as *R. papyrosolvans*, *R. cellulolyticum* and *R. thermocellum*, has two advantages: One is avoiding direct diet competition of cellulolytic organisms (who are often in minority, e.g., in the rumen only ~10% of the bacteria are cellulolytic [49]) with non-cellulolytic bacteria in carbon source. For most heterotrophic bacteria studied to date, glucose is the preferred (or primary) carbon source [50]. The other is that uptake of cellobiose or cellodextrins into the cell is more energy efficient than glucose. It is because the breakdown of cellobiose and cellodextrins into glucose-1-phosphate by the intracellular cellobiose/cellodextrin phosphorylase does not require the extra ATP [51].

## Conclusions

In this study, secretomes of *R. papyrosolvans* under various carbon sources (glucose, cellobiose, cellulose and corn stover) were investigated and compared by label-free LC-MS/MS quantification method; 912 proteins in total were found in four secretomes which, respectively, harbored a different functional profile of proteins. Compared with cellobiose, secretome under corn stover had increased expression of CAZymes for lignocellulose degradation, extracellular SBP from ATP-binding cassette (ABC) transporters and proteins involved in spore formation. Surprisingly, the number and abundance of free CAZymes were almost unchanged among four carbon sources, whereas the abundance of CAZymes from cellosome complexes, particularly those encoded by the *cip-cel* operon and the *xyl-doc* gene cluster, was significantly increased along with their numbers as cells were grown on corn stover. These features do not only delineate the secretome of lignocellulose degradation in *R.*

*papyrosolvans*, but also provide some targets for development of highly efficient biomass degradation systems by employing cellulolytic clostridia and their enzyme system.

## Materials and methods

### Strains and culture conditions

*Ruminiclostridium papyrosolvans* DSM 2782 was purchased from the Leibniz Institute DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). *R. papyrosolvans* was cultured anaerobically at 35 °C in 250-mL flasks with 100 mL working volume of modified DCB-1 medium supplemented with 2.0 g/L of glucose, cellobiose or 5.0 g/L of cellulose (Avicel PH101; Sigma-Aldrich) or corn stover, which was obtained from Taigu County, Shanxi Province, China, and milled using a beater pulverizer to pass through a mesh with a diameter of 2 mm. A 1% (vol/vol) inoculum of culture preadapted to various substrates in vials was used for inoculation. Cellular growth was measured based on the increase in extracellular proteins in the culture using the bicinchoninic acid (BCA) assay (Sangon Biotech, Shanghai, China). All cultivations were performed in triplicate.

### Isolation of extracellular proteins

100-mL cultures grown on various carbon sources were harvested when the concentration of extracellular proteins reached the maximum (Additional file 1: Figure S1). After centrifuged (12,000g, 4 °C, 30 min), the supernatants were aspirated and filtered through a 0.22- $\mu$ m PES membrane (Jinteng, Tianjin, China). The residues after centrifugation of the cultures were washed and centrifuged for three times with 5-mL EDTA buffer (50 mM Tris-HCl, 5 mM EDTA, pH8.0), with eluates collected to obtain the proteins binding to the cellulose materials. Cell-free supernatants and the filtered eluates were pooled together and concentrated 100-fold using an ultrafiltration device containing a 10-kDa-cutoff membrane (Millipore, Germany). Protein concentrations were measured by the BCA assay (Sangon Biotech, Shanghai, China). The isolated samples were boiled for 5 min at 100 °C and loaded onto 12.5% SDS-PAGE. Proteins were visualized with a Coomassie Brilliant Blue (CBB-G250) stain as described by Dyballa and Metzger [52].

To estimate the elution effect of EDTA buffer for cellulose-binding proteins, 1 mL isolated extracellular proteins from glucose culture (2 mg/mL) were incubated with 0.1 g cellulose (Avicel PH101) for 24 h at 35 °C in vitro and then washed the cellulose 3 times with 1 mL EDTA buffer. Finally, the washing solutions and residual proteins binding to cellulose were analyzed by SDS-PAGE and their amount calculated by gray scanning.

## Label-free quantitative LC–MS/MS analysis

### Proteolysis

For the proteome assay, samples (50 µg per lane) were allowed to run 2 cm beyond the stacking gel of SDS-PAGE. Protein bands were excised from the gels, cut into small cubes (ca.  $1 \times 1 \text{ mm}^3$ ) and destained according to Schluesener and colleagues [53]. Gel pieces were dried by incubation with 100% acetonitrile for 10 min at room temperature and then incubated with 50 mM DTT in 25 mM  $\text{NH}_4\text{HCO}_3$  (30 min at 60 °C) to reduce proteins disulfide bonds. The gel pieces were dried again with acetonitrile, and proteins were alkylated treating the gel pieces with 50 mM iodoacetamide in 25 mM  $\text{NH}_4\text{HCO}_3$  (1 h at room temperature in darkness). And then gel pieces were dried in a SpeedVac; trypsin (Sequencing grade modified; Promega, Madison, USA) solution (12.5 ng/mL in 25 mM ammonium bicarbonate, pH 8.6) was added until gel pieces were immersed completely in digestion solution (~200 µL). Protein digestion was performed overnight at 37 °C with a tempered shaker. After digestion and peptide extraction from gel pieces, the samples were centrifuged, and supernatants were transferred to new 1.5-mL tubes. The recovered peptides were dried using a SpeedVac and stored at room temperature. Prior to MS analysis, peptides were resuspended in 20 µL of 0.1% formic acid. Each measurement was performed with 8 µL of sample.

### Mass spectrometry analysis

An UPLC HSS T3 column (Waters, Milford, MA, USA) and an UPLC Symmetry C18 trapping column (Waters, Milford, MA, USA) for LC as well as a PicoTip Emitter (SilicaTip, 10 mm i.d., New Objective, Woburn, MA, USA) were used in combination with the nanoACQUITY gradient UPLC pump system (Waters, Milford, MA, USA) coupled to a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The peptides were eluted with a 105-min gradient of 2% to 85% acetonitrile with 0.1% formic acid at a flow rate of 400 nL/min (0–5 min: 2%; 5–10 min: 2–5%; 10–71 min: 5–30%; 72–77 min: 85%; 77–105 min: 2%). The LTQ Orbitrap Elite was operated via instrument method files of Xcalibur (Rev. 2.1.0) in positive ion mode. The linear ion trap and Orbitrap were operated in parallel, i.e., during a full MS scan on the Orbitrap in the range of 150–2000  $m/z$  at a resolution of 240,000 MS/MS spectra of the 20 most intense precursors were detected in the ion trap using the rapid scan mode. The relative collision energy for collision-induced dissociation (CID) was set to 35%. Dynamic exclusion was enabled with a repeat count of 1- and 45-s exclusion duration window. Singly charged and ions of unknown charge state were rejected from MS/MS.

### Protein identification

Protein identification was performed by Andromeda search engine [54] embedded in MaxQuant searching against the complete proteome database (GCF\_000175795.2) of *R. papyrosolvans* DSM2782 according to genome annotation in NCBI database (<https://www.ncbi.nlm.nih.gov/>). Mass tolerance for centroid match was set to 8 ppm; the mass tolerance for fragment ions was set to 0.4 Da. The PSM false discovery rate (FDR) and protein FDR were determined with MaxQuant, and the  $q$ -value was set to 1% [55]. For protein quantification, the “label-free quantification” function in MaxQuant was used.

### Functional enrichment analysis

The statistical significance of the enrichment of proteins from the secretomes in each COG category was calculated as follows: For example, let “ $N$ ” be the total number of proteins predicted in all COG categories, “ $n$ ” be the number of proteins predicted in a certain COG category, “ $M$ ” be the total number of proteins identified in secretomes, and “ $m$ ” be the number of secretomic proteins assigned to this COG category. The  $P$ -value was estimated for enrichment of proteins identified in a COG category based on the hypergeometric test:

$$P = \sum_{i=m}^{\min(n,M)} \frac{C(M, i)C(N - M, n - i)}{C(N, n)}$$

in which  $C(x, y)$  is the combinational number of choosing  $y$  items out of  $x$  items. Enrichment of COG-slim terms with  $P$ -value  $\leq 0.05$  was considered as statistically significant.

### Enzyme activity measurement

Activity assays of CAZymes were performed by incubating 0.1 mg/mL isolated secretome samples in a total volume of 500 µL assay mixture containing 1% (wt/vol) of substrates (CMC, Avicel, oat spelt xylan, or milled corn stover) in MES buffer (50 mM MES, 5 mM  $\text{CaCl}_2$ , pH6.0) at 50 °C for 8 h. The released sugar concentration was estimated by dinitrosalicylic acid (DNS) method [29] using glucose as standard. The absorbance was measured at 540 nm. All experiments were performed in triplicate.

### Additional files

**Additional file 1: Figure S1.** Growth curves of *Ruminiclostridium papyrosolvans* on glucose, cellobiose, cellulose and corn stover. Cell growth was monitored by determining the amount of proteins in supernatant of fermentation broths as described in Materials and Methods. The symbols indicate the means of three experiments, and the error bars indicate the standard deviations. The arrows show the time of sampling for MS analysis.

**Additional file 2: Table S1.** General features of the complete genome of *Ruminiclostridium papyrosolvans* DSM2782.

**Additional file 3: Table S2.** List of the putative CAZymes in the genome of *R. papyrosolvans*.

**Additional file 4: Figure S2.** SDS-PAGE analysis of the secretomes isolated from the four growth conditions (glucose, cellobiose, cellulose and corn stover).

**Additional file 5: Figure S3.** Principle component analysis (PCA) of secretomes including three biological replicates under four carbon sources (glucose, cellobiose, cellulose and corn stover).

**Additional file 6: Table S3.** The LFQ intensity of proteins in secretomes under various carbon substrates in *R. papyrosolvans*.

**Additional file 7: Table S4.** List of signal-peptide-harboring proteins identified in each carbon source.

**Additional file 8: Table S5.** Relative expression level (Z-score) of expressed CAZymes under four carbon sources (glucose, cellobiose, cellulose and corn stover) with three biological replicates.

**Additional file 9: Figure S4.** Correlation of the expression level (Z-score) of 116 expressed CAZymes between cellobiose and corn stover.

**Additional file 10: Figure S5.** Elution effect of EDTA buffer for cellulose-binding proteins. Lanes 1 and 2 secreted proteins before and after incubated with cellulose; lanes 3–5 fractions washed from cellulose; lane 6 residual proteins binding to cellulose after 3 times of washing.

#### Abbreviations

CAZyme: carbohydrate-active enzymes; GH: glycoside hydrolase; PL: polysaccharide lyases; CE: carbohydrate esterase; GT: glycosyl transferase; AA: enzyme with auxiliary activities; CBM: carbohydrate-binding module; CBP: consolidated bioprocessing; LC–MS/MS: liquid chromatography–tandem mass spectrometry; PCA: principal component analysis; COG: cluster of orthologous group; SBP: solute-binding protein; ABC: ATP-binding cassette; CCR: carbon catabolite repression; CMC: carboxymethyl cellulose; SRPS: selective RNA processing and stabilization; TCS: two-component system; LFQ: label-free quantification.

#### Acknowledgements

Not applicable.

#### Authors' contributions

ZR and WY participated in the design of the experiments, performed some of the experiments, analyzed the proteomics data and contributed to manuscript writing. SW performed some of the experiments and contributed to manuscript writing. AP participated in design of the experiments, provided the platform and method of label-free LC–MS/MS quantification and contributed to manuscript writing. CX supervised the research and participated in the design of the experiments, analyzed the proteomics data and was a major contributor in writing the manuscript. All authors have contributed to scientific discussion. All authors read and approved the final manuscript.

#### Funding

This work was supported by Grant 31571282 and 31871252 from National Natural Science Foundation of China and Grant 2018XYSDYY-01 from Research and Development Base on Solid Waste Comprehensive Utilization of Shanxi University in Changzhi.

#### Availability of data and materials

The datasets generated during the current study are available in the Proteomics IDentifications database (PRIDE) under Accession Number PXD013253.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup> Key Laboratory of Chemical Biology and Molecular Engineering of Ministry of Education, Institute of Biotechnology, Shanxi University, Taiyuan 030006, Shanxi, China. <sup>2</sup> Institute of Applied Chemistry, Shanxi University, Taiyuan 030006, Shanxi, China. <sup>3</sup> Department of Plant Biochemistry, Ruhr University Bochum, 44801 Bochum, Germany. <sup>4</sup> School of Biomedical and Healthcare Sciences, University of Plymouth, Plymouth PL48AA, UK.

Received: 25 June 2019 Accepted: 5 July 2019

Published online: 15 July 2019

#### References

- Delmer DP. Cellulose biosynthesis: exciting times for a difficult field of study. *Annu Rev Plant Physiol Plant Mol Biol.* 1999;50:245–76.
- McDonald JE, Houghton JN, Rooks DJ, Allison HE, McCarthy AJ. The microbial ecology of anaerobic cellulose degradation in municipal waste landfill sites: evidence of a role for fibrobacters. *Environ Microbiol.* 2012;14:1077–87.
- Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science.* 2007;315:804–7.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res.* 2014;42:D490–5.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZY): an expert resource for glycogenomics. *Nucleic Acids Res.* 2009;37:D233–8.
- Demain AL, Newcomb M, Wu JH. Cellulase, clostridia, and ethanol. *Microbiol Mol Biol Rev.* 2005;69:124–54.
- Dash S, Ng CY, Maranas CD. Metabolic modeling of clostridia: current developments and applications. *FEMS Microbiol Lett.* 2016;363:fnw004.
- Bayer EA, Lamed R, White BA, Flint HJ. From cellulosomes to cellulosomes. *Chem Rec.* 2008;8:364–77.
- Artzi L, Bayer EA, Morais S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat Rev Microbiol.* 2017;15:83–95.
- Bayer EA, Shimon LJ, Shoham Y, Lamed R. Cellulosomes—structure and ultrastructure. *J Struct Biol.* 1998;124:221–34.
- Xin F, Dong W, Zhang W, Ma J, Jiang M. Biobutanol production from crystalline cellulose through consolidated bioprocessing. *Trends Biotechnol.* 2019;37:167–80.
- Singh N, Mathur AS, Gupta RP, Barrow CJ, Tuli D, Puri M. Enhanced cellulolytic ethanol production via consolidated bioprocessing by *Clostridium thermocellum* ATCC 31924. *Bioresour Technol.* 2018;250:860–7.
- Xu C, Qin Y, Li Y, Ji Y, Huang J, Song H, Xu J. Factors influencing cellulosome activity in consolidated bioprocessing of cellulosic ethanol. *Bioresour Technol.* 2010;101:9560–9.
- Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. How biotech can transform biofuels. *Nat Biotechnol.* 2008;26:169–72.
- Xu C, Huang R, Teng L, Wang D, Hemme CL, Borovok I, He Q, Lamed R, Bayer EA, Zhou J, Xu J. Structure and regulation of the cellulose degradome in *Clostridium cellulolyticum*. *Biotechnol Biofuels.* 2013;6:73.
- Xu C, Huang R, Teng L, Jing X, Hu J, Cui G, Wang Y, Cui Q, Xu J. Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization. *Nat Commun.* 2015;6:6900.
- Blouzard JC, Coutinho PM, Fierobe HP, Henrissat B, Lignon S, Tardif C, Pages S, de Philip P. Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics.* 2010;10:541–54.
- Fendri I, Tardif C, Fierobe HP, Lignon S, Valette O, Pages S, Perret S. The cellulosomes from *Clostridium cellulolyticum*: identification of new components and synergies between complexes. *FEBS J.* 2009;276:3076–86.
- Hemme CL, Mouttaki H, Lee YJ, Zhang G, Goodwin L, Lucas S, Copeland A, Lapidus A, Glavina del Rio T, Tice H, et al. Sequencing of multiple clostridial genomes related to biomass conversion and biofuel production. *J Bacteriol.* 2010;192:6494–6.

20. Zepeda V, Dassa B, Borovok I, Lamed R, Bayer EA, Cate JH. Draft genome sequence of the cellulolytic bacterium *Clostridium papyrosolvens* C7 (ATCC 700395). *Genome Announc.* 2013;1(5):e00698-13.
21. Zou X, Ren Z, Wang N, Cheng Y, Jiang Y, Wang Y, Xu C. Function analysis of 5'-UTR of the cellulosomal xyl-doc cluster in *Clostridium papyrosolvens*. *Biotechnol Biofuels.* 2018;11:43.
22. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39:W29–37.
23. Huang L, Zhang H, Wu PZ, Entwistle S, Li XQ, Yohe T, Yi HD, Yang ZL, Yin YB. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* 2018;46:D516–21.
24. Dassa B, Borovok I, Lombard V, Henrissat B, Lamed R, Bayer EA, Morais S. Pan-cellulosomics of mesophilic clostridia: variations on a theme. *Microorganisms.* 2017;5(4):74.
25. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37:420–3.
26. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 2011;39:D561–8.
27. Wiegel J, Dykstra M. *Clostridium thermocellum*: adhesion and sporulation while adhered to cellulose and hemicellulose. *Appl Microbiol Biotechnol.* 1984;20:59–65.
28. Li Y, Xu T, Tschaplinski TJ, Engle NL, Yang Y, Graham DE, He Z, Zhou J. Improvement of cellulose catabolism in *Clostridium cellulolyticum* by sporulation abolishment and carbon alleviation. *Biotechnol Biofuels.* 2014;7:25.
29. Ghose TK. Measurement of cellulase activities. *Pure Appl Chem.* 1987;59:257–68.
30. Celik H, Blouzard JC, Voigt B, Becher D, Trotter V, Fierobe HP, Tardif C, Pages S, de Philip P. A two-component system (XydS/R) controls the expression of genes encoding CBM6-containing proteins in response to straw in *Clostridium cellulolyticum*. *PLoS ONE.* 2013;8:e56063.
31. Abdou L, Boileau C, de Philip P, Pages S, Fierobe HP, Tardif C. Transcriptional regulation of the *Clostridium cellulolyticum* cip-cel operon: a complex mechanism involving a catabolite-responsive element. *J Bacteriol.* 2008;190:1499–506.
32. Fosses A, Mate M, Franche N, Liu N, Denis Y, Borne R, de Philip P, Fierobe HP, Perret S. A seven-gene cluster in *Ruminiclostridium cellulolyticum* is essential for signalization, uptake and catabolism of the degradation products of cellulose hydrolysis. *Biotechnol Biofuels.* 2017;10:250.
33. Tamaru Y, Miyake H, Kuroda K, Ueda M, Doi RH. Comparative genomics of the mesophilic cellulosome-producing *Clostridium cellulovorans* and its application to biofuel production via consolidated bioprocessing. *Environ Technol.* 2010;31:889–903.
34. Kakiuchi M, Isui A, Suzuki K, Fujino T, Fujino E, Kimura T, Karita S, Sakka K, Ohmiya K. Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. *J Bacteriol.* 1998;180:4303–8.
35. Gilbert HJ. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Mol Microbiol.* 2007;63:1568–76.
36. Zhivin O, Dassa B, Morais S, Utturkar SM, Brown SD, Henrissat B, Lamed R, Bayer EA. Unique organization and unprecedented diversity of the *Bacteroides* (*Pseudobacteroides*) *cellulosolvens* cellulosome system. *Biotechnol Biofuels.* 2017;10:211.
37. Dassa B, Borovok I, Lamed R, Henrissat B, Coutinho P, Hemme CL, Huang Y, Zhou JZ, Bayer EA. Genome-wide analysis of *Acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. *BMC Genomics.* 2012;13(1):210.
38. Jindou S, Brulc JM, Levy-Assaraf M, Rincon MT, Flint HJ, Berg ME, Wilson MK, White BA, Bayer EA, Lamed R, Borovok I. Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*. *FEMS Microbiol Lett.* 2008;285:188–94.
39. Indrelić S, Mathiesen G, Jacobsen M, Lea T, Kleiveland CR. Computational and experimental analysis of the secretome of *Methylococcus capsulatus* (Bath). *PLoS ONE.* 2014;9:e114476.
40. Broecker J, Mechelke M, Baudrexl M, Mennerich D, Hornburg D, Mann M, Schwarz WH, Liebl W, Zverlov VV. The hemicellulose-degrading enzyme system of the thermophilic bacterium *Clostridium stercoarum*: comparative characterisation and addition of new hemicellulolytic glycoside hydrolases. *Biotechnol Biofuels.* 2018;11:229.
41. Ogunmolu FE, Kaur I, Pasari N, Gupta M, Yazdani SS. Quantitative multiplexed profiling of *Penicillium funiculosum* secretome grown on polymeric cellulase inducers and glucose. *J Proteom.* 2018;179:150–60.
42. Cai Y, Gong Y, Liu W, Hu Y, Chen L, Yan L, Zhou Y, Bian Y. Comparative secretomic analysis of lignocellulose degradation by *Lentinula edodes* grown on microcrystalline cellulose, lignosulfonate and glucose. *J Proteom.* 2017;163:92–101.
43. Rai R, Kaur B, Singh S, Di Falco M, Tsang A, Chadha BS. Evaluation of secretome of highly efficient lignocellulolytic *Penicillium* sp. Dal 5 isolated from rhizosphere of conifers. *Bioresour Technol.* 2016;216:958–67.
44. Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y, Steitz TA. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* 1996;15:5739–51.
45. Gunnoo M, Cazade PA, Orlowski A, Chwastyk M, Liu H, Ta DT, Cieplak M, Nash M, Thompson D. Steered molecular dynamics simulations reveal the role of Ca(2+) in regulating mechanostability of cellulose-binding proteins. *Phys Chem Chem Phys.* 2018;20:22674–80.
46. Yoav S, Barak Y, Shamshoum M, Borovok I, Lamed R, Dassa B, Hadar Y, Morag E, Bayer EA. How does cellulosome composition influence deconstruction of lignocellulosic substrates in *Clostridium* (*Ruminiclostridium*) *thermocellum* DSM 1313? *Biotechnol Biofuels.* 2017;10:222.
47. Servinsky MD, Kiel JT, Dupuy NF, Sund CJ. Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. *Microbiology.* 2010;156:3478–91.
48. Han SO, Yukawa H, Inui M, Doi RH. Regulation of expression of cellulosomal cellulase and hemicellulase genes in *Clostridium cellulovorans*. *J Bacteriol.* 2003;185:6067–75.
49. Russell JB, Muck RE, Weimer PJ. Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. *FEMS Microbiol Ecol.* 2009;67:183–97.
50. Bruckner R, Titgemeyer F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol Lett.* 2002;209:141–8.
51. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002;66:506–77.
52. Dyballa N, Metzger S. Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. *J Vis Exp.* 2009;30:e1431.
53. Schluesener D, Fischer F, Kruij J, Rogner M, Poetsch A. Mapping the membrane proteome of *Corynebacterium glutamicum*. *Proteomics.* 2005;5:1317–30.
54. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 2011;10:1794–805.
55. Kall L, Storey JD, MacCoss MJ, Noble WS. Posterior error probabilities and false discovery rates: two sides of the same coin. *J Proteome Res.* 2008;7:40–4.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.