

RESEARCH

Open Access



# Differential transcriptome analysis of *Sporocytophaga* sp. CX11 and identification of candidate genes involved in lignocellulose degradation

Jiwei Wang<sup>†</sup>, Ying Zhuang<sup>†</sup>, Xianghe Song, Xu Lin, Xiangyi Wang, Fan Yang\* and Xiaoyi Chen\*

## Abstract

Cellulose is the most abundant renewable bioresources on earth, and the biodegradation and utilization of cellulose would contribute to the sustainable development of global environment. *Sporocytophaga* species are common aerobic cellulose-degrading bacteria in soil, which can adhere to the surface of cellulose matrix and motile by gliding. In this study, a differential transcriptome analysis of *Sporocytophaga* sp. CX11 was performed and a total of 4,217 differentially expressed genes (DEGs) were identified. Gene Ontology enrichment results showed that there are three GO categories related to cellulose degradation function among the annotated DEGs. A total of 177 DEGs were identified as genes encoding carbohydrate-active enzymes (CAZymes), among which 54 significantly upregulated CAZymes were mainly cellulases, hemicellulases, pectinases, etc. 39 DEGs were screened to associate with gliding function. In order to explore unannotated genes potentially related to cellulose metabolism, cluster analysis was performed using the Short-Time Series Expression Miner algorithm (STEM). 281 unannotated genes were predicted to be associated with the initial-middle stage of cellulose degradation and 289 unannotated genes might function in the middle-last stage of cellulose degradation. *Sporocytophaga* sp. CX11 could produce extracellular endo-xylanase, endo-glucanase, FPase and β-glucosidase, respectively, according to different carbon source conditions. Altogether, this study provides valuable insights into the transcriptome information of *Sporocytophaga* sp. CX11, which would be useful to explore its application in biodegradation and utilization of cellulose resources.

**Keywords** Cellulose biodegradation, *Sporocytophaga*, Differential transcriptome analysis, Carbohydrate-active enzymes, Gliding motility

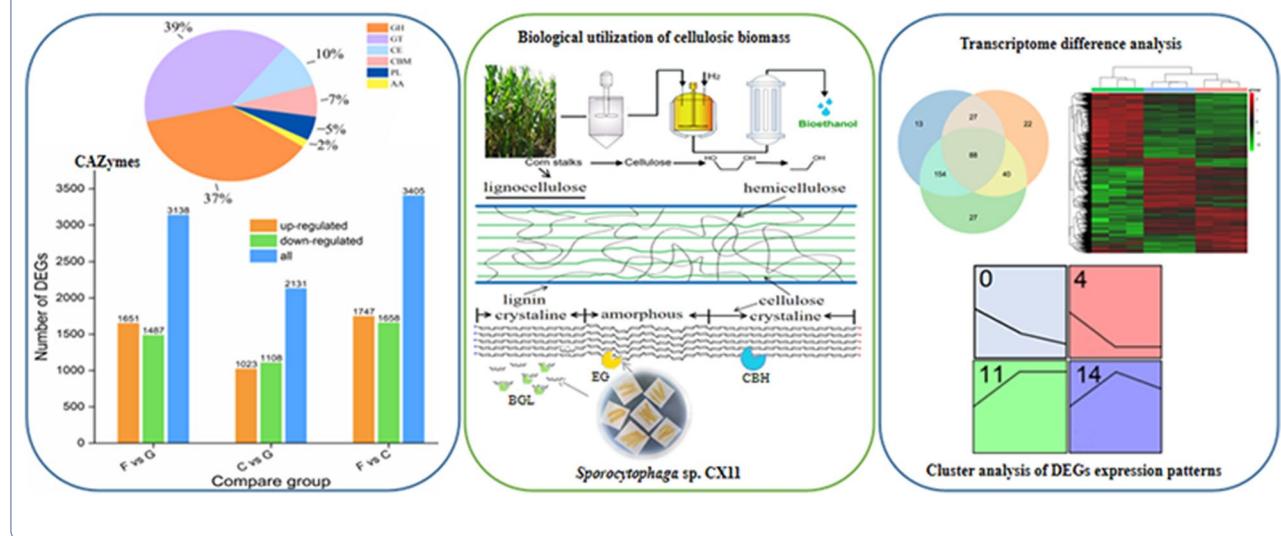
<sup>†</sup>Jiwei Wang and Ying Zhuang have contributed equally to this work

\*Correspondence:

Fan Yang  
yang\_fan@dlpu.edu.cn  
Xiaoyi Chen  
chen-xy@dlpu.edu.cn

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

## Graphical Abstract



## Introduction

Cellulose is the most abundant sustainable bioresource on earth, and can be used to generate biofuels such as ethanol (Lane et al. 2018). Compared with fossil energy, biofuels are renewable and will not cause greenhouse effect. The effective bioconversion and utilization of cellulose matrix is therefore the key to control environmental pollution and achieve both industrial and social sustainability.

As early as in 1919, *Sporocytophaga* species was isolated from soil and determined to be capable of efficiently degrading cellulose (Holt and Leadbetter 1967; Hutchinson and Clayton 1919). *Sporocytophaga*, together with *Cytophaga* (Stanier 1942) and *Sorangium* (Burkhard 1981), are kinds of gliding bacteria that are able to strongly adsorbed to the surface of insoluble cellulosic materials to obtain nutrients. The gliding behavior is closely related to their cellulose degradation ability. Strains of the genus *Sporocytophaga* can use cellulose, cellobiose or glucose as the only carbon source and the colonies grown on filter paper are yellow and gummy due to the production of extracellular slime. In 2014, Liu et al. determined the draft genome sequence of *Sporocytophaga* sp. PG-01 (Liu et al. 2014) and mined for genes encoding cellulases and hemicellulases that are closely related to cellulose degradation by the strain. Nonetheless, the studies on the comparative genomics and transcriptomics of *Sporocytophaga* strains have not been reported, and the mechanism of cellulose degradation of them remains to be clarified.

In previous work, a cellulose-degrading strain designated as CX11 has been isolated and preserved in our

lab. According to the polyphasic taxonomy results, strain CX11 should be a new species of the genus *Sporocytophaga* (data not published). In this study, we performed differential transcriptome analysis of *Sporocytophaga* sp. CX11 based on RNA-seq technology and mined genes encoding carbohydrate-active enzymes that play a key role in cellulose degradation of the strain. The experimental and sequencing data generated in this study are supposed to benefit the analysis of cellulose degradation process of strain CX11 and also lay the foundation for the clarification of cellulose degradation mechanism of the genus *Sporocytophaga*.

## Materials and methods

### Bacterial materials and growth conditions

*Sporocytophaga* sp. CX11 was isolated and preserved by our lab at Dalian Polytechnic University. The strain was inoculated on the filter paper double-layer solid medium and cultured at 30 °C for 3 days. Then the strain was transferred into liquid medium with filter paper, cellobiose and glucose as the only carbon source, respectively, and cultivated at 30 °C and 200 r/min until it reached the early stage of logarithmic phase.

The upper layer of the filter paper double-layer solid medium was Whatman No.1 filter paper and the lower layer was composed of 0.5 g L<sup>-1</sup> NaNO<sub>3</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub> 7H<sub>2</sub>O, 0.5 g L<sup>-1</sup> KCl, 1.0 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub> 3H<sub>2</sub>O, 0.01 g L<sup>-1</sup> FeSO<sub>4</sub> 3H<sub>2</sub>O, 20 g L<sup>-1</sup> agar. The composition of the liquid medium was: 5.0 g L<sup>-1</sup> carbon source, 0.5 g L<sup>-1</sup> NaNO<sub>3</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub> 7H<sub>2</sub>O, 0.5 g L<sup>-1</sup> KCl, 1.0 g L<sup>-1</sup> K<sub>2</sub>HPO<sub>4</sub> 3H<sub>2</sub>O and 0.01 g L<sup>-1</sup> FeSO<sub>4</sub> 3H<sub>2</sub>O.

### Total RNA extraction

The total RNA of CX11 cultured under different carbon sources was extracted, respectively, according to the instructions of SteadyPure Universal RNA Extraction Kit (Accurate Biology, China). The extracted RNA samples were checked for quality by agarose gel electrophoresis experiments and Aglient Bioanalyzer 2100 system (Agilent Technologies, USA).

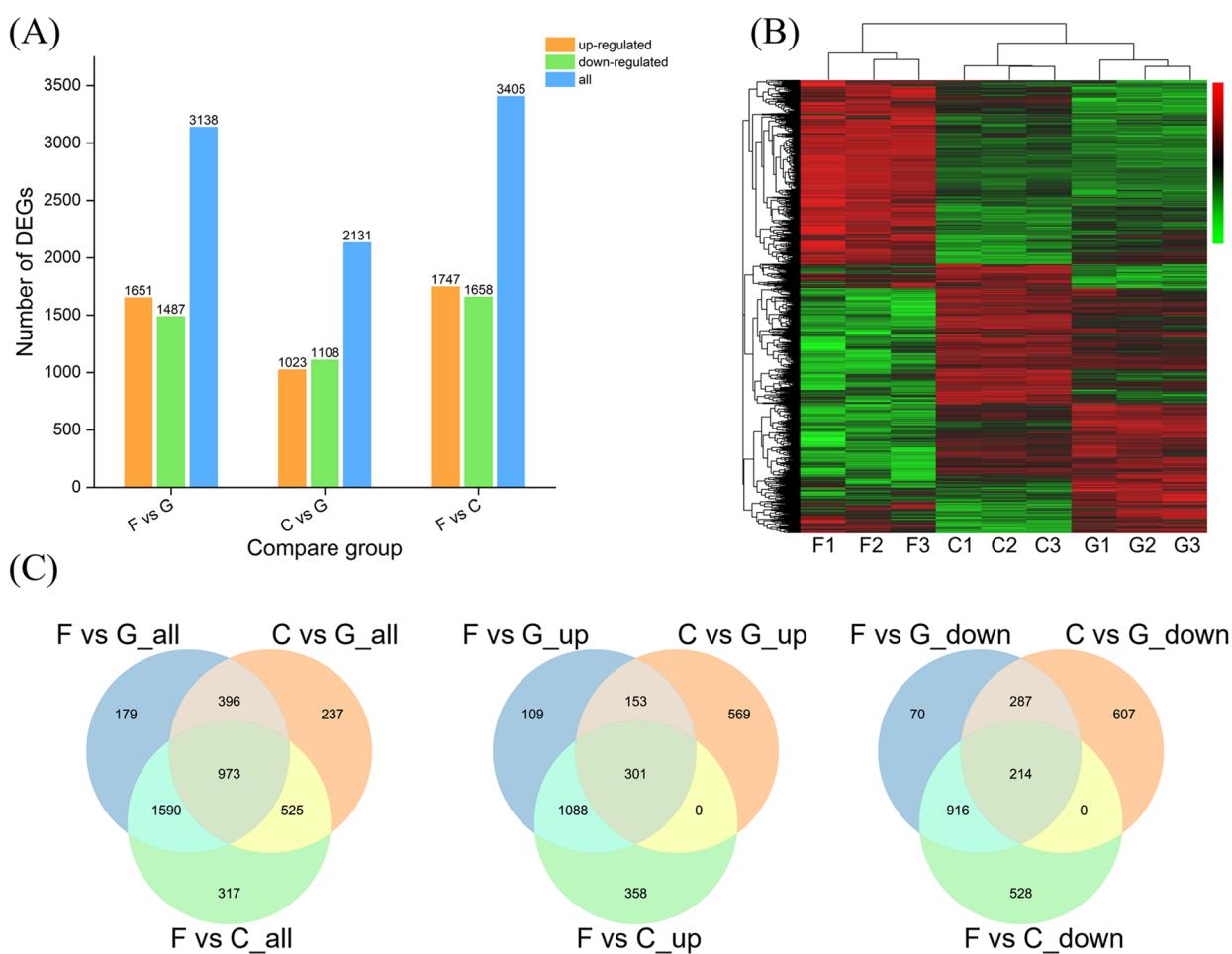
### cDNA library construction and RNA sequencing

RNA samples from cultures with three different carbon sources (nine samples) were submitted to Novogene Co., Ltd. (Beijing) for cDNA libraries construction and Illumina sequencing. Peak plots of sequencing results were converted to raw reads after identification by CASAVA 1.8. Clean reads were obtained by eliminating the low-quality reads (reads containing adapters, reads with quality values lower than 20 and reads containing more than 5% of unclear bases) (Macconnail et al. 2018). The clean

reads of nine cDNA libraries were localized to the reference genome of *Sporocytophaga* using Bowtie2 (Langmead and Salzberg 2012), and the localization results were counted to calculate the coverage of the reads on the reference genome. When the ratio was higher than 70%, it indicated that the reference genome was selected appropriately and the sequenced samples were free of exogenous species contamination (Yang et al. 2017).

### Bioinformatics analysis of RNA-seq data

After the sequencing data were processed, the expression levels of differentially expressed genes (DEGs) were calculated using the Fragments Per Kilobase of exon model per Million mapped (FPKM) (Zhao et al. 2021). Differential expression analysis was performed using DESeq package (Wang et al. 2010), and samples from three different carbon sources were compared pairwise. Differential genes were screened with  $p < 0.05$  and  $|\log_2 \text{FC}| > 0$



**Fig. 1** Analysis of DEGs. **A:** Statistical results of DEGs in different carbon source combinations. **B:** Heatmap of clustering of DEGs. **C:** Venn diagram of DEGs

as the threshold. The number of upregulated and down-regulated genes was subsequently counted according to fold change. Gene Ontology (GO) enrichment analysis of DEGs was implemented by GOseq (Young et al. 2010) to obtain GO categories. The GO function of DEGs was categorized with WEGO (<http://wego.genomics.org.cn>) (Ye et al. 2006). Hypergeometric test was applied to correct the *p*-value of the enrichment results, and those with *p*<0.05 were considered as significantly enriched GO categories for DEGs. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs was performed using KOBAS 2.0 software (Chen et al. 2011). Hypergeometric test was also applied to select metabolic pathways that were significantly enriched.

#### Prediction of gene function by short-time series expression miner algorithm

The functions of unannotated genes were predicted based on differential gene expression trends and functional annotations. The unannotated genes with FPKM mean values less than 50 and known functional genes clearly unrelated to cellulose degradation were removed from the three sets of samples, and the remaining genes were regarded as a gene set. The expression pattern analysis was performed using short-time series expression miner (STEM) algorithm (Ernst et al. 2005), and the genes with consistent patterns of expression change were clustered. The unannotated genes that were clustered into the same class with annotated genes were presumed to have similar functions with them.

#### Real-time quantitative PCR validation of candidate genes

Real-time quantitative PCR primers were designed by the online design tool (<https://sg.idtdna.com/>) and Primer Premier 5 (Additional file 1: Table S1). The extracted RNA was reverse transcribed to generate cDNA using the PrimeScript<sup>TM</sup> RT reagent Kit (Perfect Real Time). RT-qPCR was performed using the SYBR Green Polymerase Chain Reaction Kit. Each reaction was performed with a total volume of 20  $\mu$ L, with a pre-incubation program of 30 s at 95 °C, followed by 40 cycles of 10 s at 95 °C, 30 s at 60 °C and 15 s at 95 °C. The housekeeping gene 16S rRNA was used as a reference. PCR procedure for RT-qPCR was performed using standard methods, and the relative expression levels were calculated according to the standard  $2^{-\Delta\Delta Ct}$  method (Livak and Schmittgen 2001).

#### In situ enzymatic activity assays

To determine the lignocellulose-degrading capacity of strain CX11, enzymatic activity assays of fermentation broth were performed. The total cellulose-degrading activity, endo-1,4-beta-glucanase activity and xylanase

activity were tested using Whatman No.1 filter paper, sodium carboxymethyl cellulose (CMC-Na) and xylan (from beechwood) as substrates, respectively. The generated reducing sugars were measured by the method of 3,5-dinitrosalicylic acid (DNS) as described previously (Miller 1959). *p*-Nitrophenyl-β-D-glucopyranoside (*p*NPG) was used as substrate to test the beta-glucosidase activity as reported before (Fang et al. 2014). The fermentation broth was boiled for 30 min to inactivate any possible enzymatic activities, and was used as a negative control. One unit (U) of enzyme activity was defined as the amount of enzyme required to release 1  $\mu$ M of reducing sugar or *p*NP per minute, and was expressed as U mL<sup>-1</sup>.

#### Statistical analysis

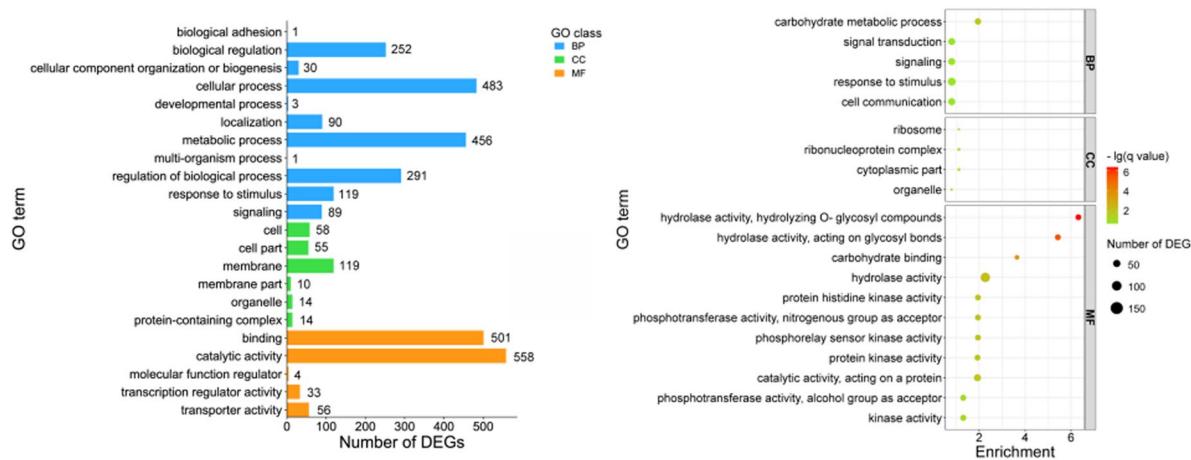
All the tests were performed in triplicate unless otherwise stated. Data were expressed as the mean±standard deviation of triplicate measurements.

### Results and discussion

#### Identification and annotation of differentially expressed genes (DEGs)

In the biodegradation of cellulosic substrates, the polysaccharide cellulose is originally hydrolyzed to cellobiose by endoglucanase and/or exoglucanase, which is then converted to glucose by β-glucosidase (Guerrero et al. 2020). It has been reported that the production of cellulases can be induced and regulated by fermentation conditions such as carbon sources, and certain enzymes are only synthesized with the existence of related substrates in the medium (Sukumaran et al. 2021). In this study, we adopted filter paper and cellobiose as the carbon sources, respectively, to induce the production of related cellulose-hydrolyzing enzymes. Glucose was chosen as the third carbon source to represent the situation that none of inducers exist in the medium. Besides, the chosen of the three carbon sources also simulated the process of cellulose degradation. Therefore, the differentially expressed genes under the three types of media should reflect the different roles of genes in cellulose degradation.

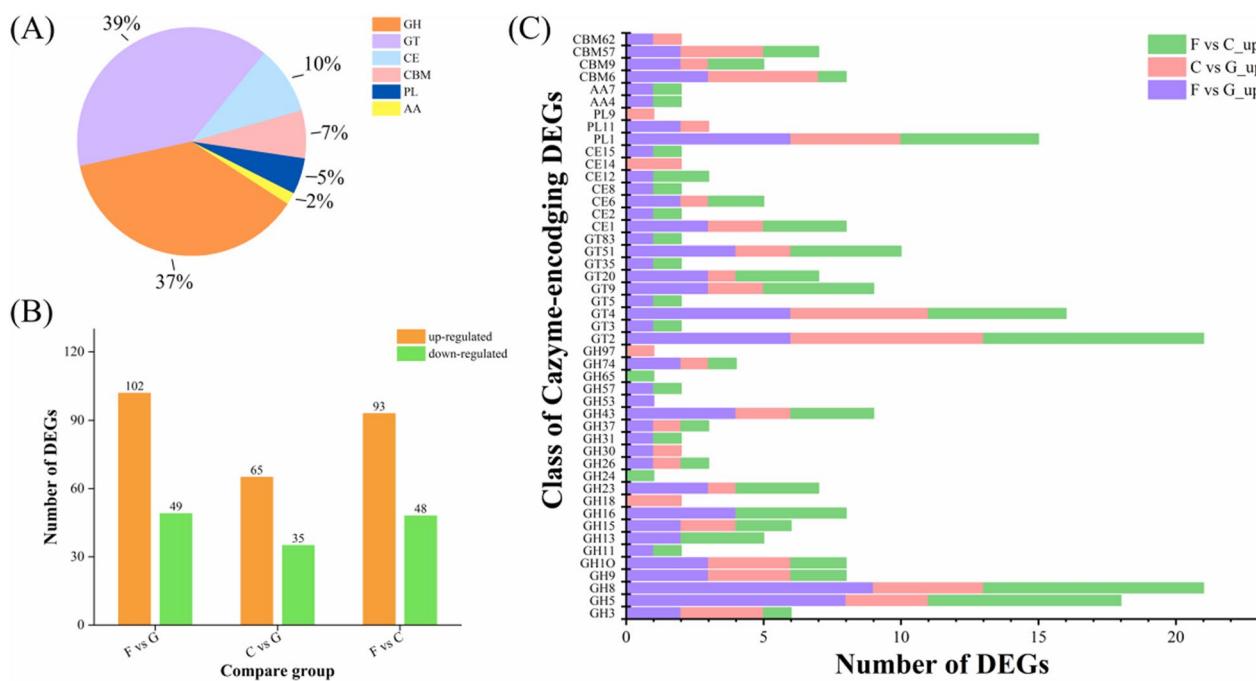
Differential expression genes analysis was performed on FPKM data for three comparative combinations of filter paper and glucose (F vs G), cellobiose and glucose (C vs G), and filter paper and cellobiose (F vs C). The carbon source with higher complexity in each combination was used as the comparison group, and the screening threshold was set to  $|\log_2 FC|>0$  and  $padj<0.05$ . A total of 4,217 DEGs were differentially expressed in the three combinations, of which a total of 3,138 DEGs were screened in the F vs G combination, 2,131 DEGs were screened in the C vs G combination, and 3,405 DEGs were screened in the



F vs C combination (\\* MERGEFORMAT Fig. 1A). The number of up/downregulated DEGs in each combination is displayed in Venn diagram plotting (\\* MERGEFORMAT Fig. 1C), which shows that 973 of the total 4,217 DEGs were identified in all the three combinations. Of these DEGs, 301 genes were upregulated and 214 genes were downregulated. The number of DEGs in the combination with filter paper as the comparison group was significantly greater than that in the combination with cellobiose as the comparison group. These results are consistent with that in previous studies that the

complexity of carbon source directly affects the number of DEGs (Gaskell et al. 2016).

The results of clustering analysis of DEGs are shown in \\* MERGEFORMAT Fig. 1B. Red color indicates high gene expression and green color indicates low gene expression. Genes in the similar color region were clustered and presumed to have similar functions or participate in regulating the same metabolic pathway. The expression pattern of the same gene in C and G groups is similar, while that in F groups is quite different, indicating that the complexity of carbon source can induce gene expression profile.



**Fig. 3** Statistical analysis of DEGs encodingCAZymes. **A:** Family classification of all DEGs. **B:** Statistics of up/downregulated CAZymes in each combination. **C:** Detailed classification of families with upregulated CAZymes in each combination

#### Gene Ontology enrichment analysis of DEGs

To further identify the possible biological functions of DEGs in cellulose degradation, GOseq software was used to perform GO enrichment analysis. A total of 1,589 DEGs were annotated into three major categories: biological process, cellular component and molecular function. In the biological process category, DEGs were mainly enriched on the functional categories such as biological process, metabolic process and cellular process. In the cellular component category, DEGs were mainly enriched on the functional categories such as cell part, membrane and membrane part. In the molecular function category, DEGs were mainly enriched on the functional categories such as catalytic activity, binding, etc. (\* MERGEFORMAT Fig. 2A).

In this experiment, the number of upregulated DEGs enriched on each functional category was significantly higher than that of downregulated DEGs (Additional file 1: Figures S1-S3). Therefore, the upregulated DEGs were selected for GO significance enrichment analysis (Zheng and Wang 2008). In the biological process category, only carbohydrate metabolic process was significantly enriched ( $p < 0.05$ ) in F vs G\_up combination (\* MERGEFORMAT Fig. 2B). In the cellular component category, no GO categories were significantly enriched in any of the three combinations. In the molecular function category, hydrolase activity acting on glycosyl bonds and O-glycosyl compounds were significantly enriched in

all the three combinations. Carbohydrate binding, catalytic activity acting on a protein, and kinase activity were significantly enriched in the F vs G\_up (\* MERGEFORMAT Fig. 2B) and C vs G\_up (\* MERGEFORMAT Fig. 2C).

Based on the Pfam database (Finn et al. 2014) comparison of DEGs enriched in the above categories, a total of three GO categories were found to be associated with cellulose degradation functions, which are carbohydrate metabolic processes, carbohydrate binding and O-glycosyl compound hydrolase activity.

#### CAZyme-encoding genes under different carbon sources mapped to the DEGs

A total of 177 DEGs were identified as genes encoding carbohydrate-active enzymes (CAZymes), and were classified into six major categories (Drula et al. 2022) according to their annotated functional properties (Fig. 3A). Among them, 66 DEGs were designated in the category of glycoside hydrolases (GHs) and distributed in 22 different families (Fig. 3C). GH8 and GH5 are the two predominant families, in which DEGs were mainly annotated as endoglucanases and reducing-end-xylose releasing exo-oligoxyylanases. DEGs in GH43 family were mainly annotated as xylosidases or arabinosidases, while GH9 and GH10 families primarily contain DEGs annotated as endoxylanases and endoglucanases.  $\beta$ -Glucosidases were chiefly distributed in family GH3. These enzymes are all

**Table 1** Summary of CAZyme-encoding genes with significantly increased expression in groups

Class	Gene ID	Enzyme description	CAZymes family	$\log_2$ fold change $\geq 1$ ; padj $\leq 0.01$		
				F vs G_up	C vs G_up	F vs C_up
Cellulases	CX11_GM000268	Beta-glucosidase	GH3	1.02	0.78	
	CX11_GM004302	Beta-glucosidase	GH3		1.89	
	CX11_GM002852	Endoglucanase	GH5	3.45	0.74	2.59
	CX11_GM004438	Endoglucanase	GH5	3.37		3.11
	CX11_GM002804	Endoglucanase	GH5	2.92		3.58
	CX11_GM004508	Endoglucanase	GH5	1.51	1.55	
	CX11_GM003220	Endoglucanase 3	GH5	2.9		2.62
	CX11_GM001350	Endoglucanase B	GH5		2.09	
	CX11_GM000434	Endoglucanase C	GH5	2.11		1.98
	CX11_GM003239	Endoglucanase 2	GH8	3.58	0.96	2.46
	CX11_GM000388	Endoglucanase 2	GH8	2.71		2.38
	CX11_GM001809	Endoglucanase 2	GH8	2.33	0.41	1.81
	CX11_GM005339	Endoglucanase 2	GH8	1.42	1.15	
	CX11_GM000618	Endoglucanase C	GH8	2.65		2.51
	CX11_GM005080	Endoglucanase A	GH9	1.09	0.48	0.5
	CX11_GM000056	Endoglucanase D	GH9	2.34		3.23
	CX11_GM000079	Beta-glucanase	GH16	1.33		1.18
	CX11_GM000626	Beta-glucanase	GH16	1.07		0.80
Hemicellulases	CX11_GM000042	Bifunctional acetylxyran esterase	CE1	2.62	1.44	1.07
	CX11_GM004263	Carbohydrate acetyl esterase/feruloyl esterase	CE1	1.18		1
	CX11_GM001962	Carbohydrate acetyl esterase/feruloyl esterase	CE6	2.15	1.23	0.81
	CX11_GM003964	Carbohydrate acetyl esterase/feruloyl esterase	CE6	2.97		1.84
	CX11_GM001963	Mannan endo-1,4-beta-mannosidase	GH5	2.92		2.51
	CX11_GM000038	Reducing-end-xylose releasing exo-oligoxylanase	GH8	2.91		2.43
	CX11_GM002090	Reducing-end-xylose releasing exo-oligoxylanase	GH8	2.74		2.27
	CX11_GM000628	Reducing-end-xylose releasing exo-oligoxylanase	GH8	2.60		1.68
	CX11_GM005472	Xyloglucan-specific endo-beta-1,4-glucanase	GH9	0.92	2.42	
	CX11_GM000081	Endo-1,4-beta-xylanase B	GH10	2.87	1.06	1.69
	CX11_GM003219	Endo-1,4-beta-xylanase B	GH10	2.02	0.93	0.97
	CX11_GM000047	Endo-1,4-beta-xylanase B	GH10	1.71	1.31	
	CX11_GM000039	Endo-1,4-beta-xylanase 3	GH11	1.38		1.04
	CX11_GM002851	Endo-1,4-beta-galactosidase	GH16	2.64		2.86
	CX11_GM004712	Mannan endo-1,4-beta-mannosidase	GH26	3.86	0.79	2.95
	CX11_GM000073	Glucuronoxylanase	GH30	1.08	0.71	
	CX11_GM004262	Arabinoxylan arabinofuranohydrolase	GH43	1.87		1.27
	CX11_GM000043	Xylosidase/arabinosidase	GH43		1.35	
	CX11_GM004261	Beta-xylosidase	GH43	1.48		1.34
	CX11_GM000048	Alpha-L-arabinofuranosidase	GH43	1.19	1.06	
	CX11_GM002260	Xyloglucanase	GH74	3.79		3.62
	CX11_GM000027	Xyloglucanase	GH74	1.85	1.57	
	CX11_GM000529	Mannosyl transferase	GT83	2.00		2.03

**Table 1** (continued)

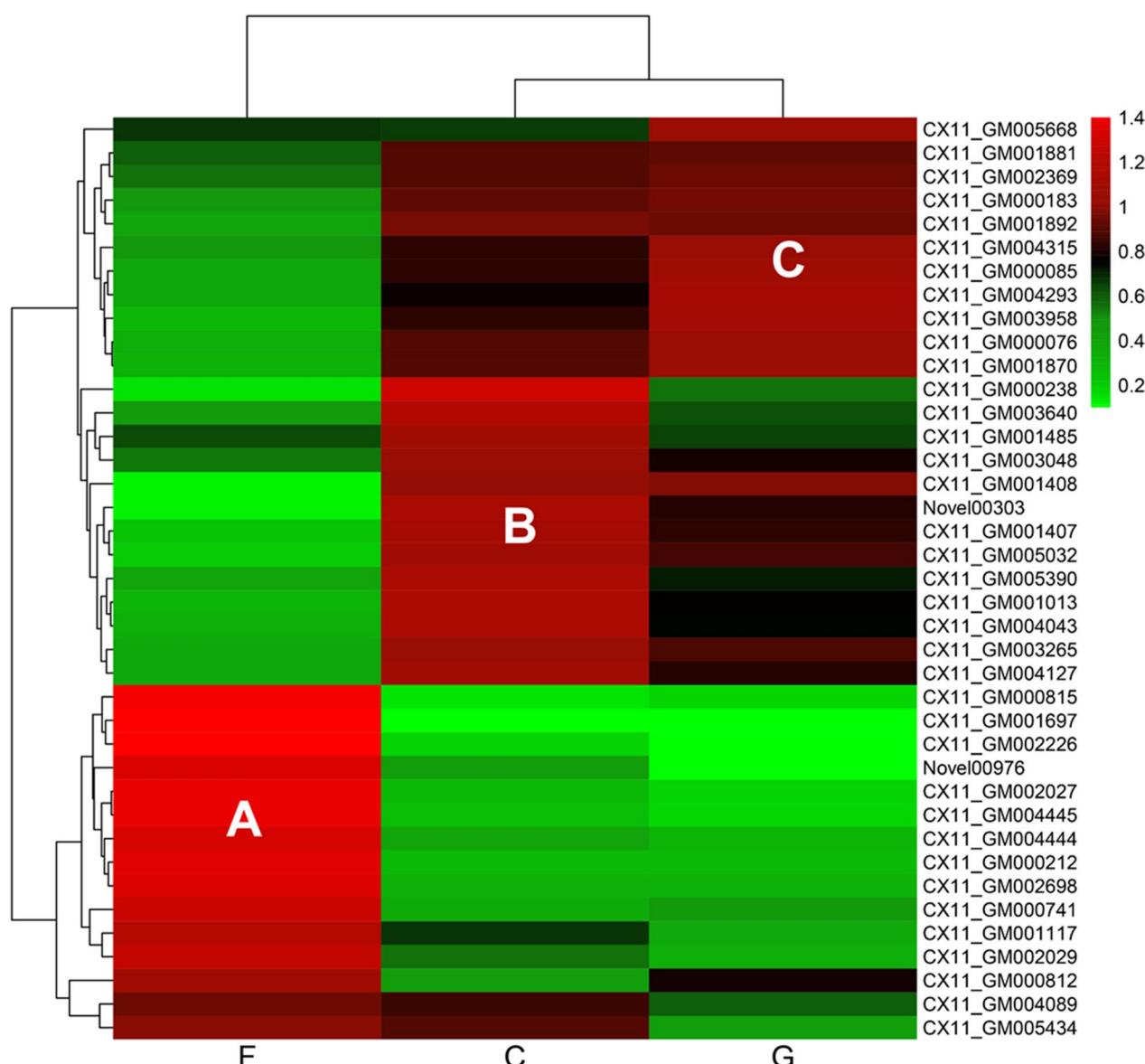
Class	Gene ID	Enzyme description	CAZymes family	$\log_2$ fold change $\geq 1$ ; $p\text{adj} \leq 0.01$		
				F vs G_up	C vs G_up	F vs C_up
Pectinases	CX11_GM001947	Pectinesterase 31	CE8	2.9		2.67
	CX11_GM003962	Rhamnogalacturonan acetylesterase	CE12	1.56		0.99
	CX11_GM000527	L-arabinose transferase	GT2	5.8	1.06	4.62
	CX11_GM004710	Pectate lyase E	PL1	2.85	2.14	0.6
	CX11_GM000632	Pectate trisaccharide-lyase	PL1	2.3	1.11	1.08
	CX11_GM000058	Pectate lyase C	PL1	3.23		1.57
	CX11_GM003503	Pectate lyase C	PL1	2.99	1.55	2.16
	CX11_GM003502	Pectate trisaccharide-lyase	PL1	1.96	0.68	1.18
	CX11_GM003504	Pectate trisaccharide-lyase	PL1	1.49		
	CX11_GM003956	Rhamnogalacturonan endolyase	PL11	1.58	1.22	
Chitinases	CX11_GM003955	Rhamnogalacturonan endolyase	PL11	1.07		
	CX11_GM002492	Chitinase	GH18		1.52	
	CX11_GM002954	Chitinase	GH18		1.02	

involved in the hydrolyzation of cellulose and hemicellulose. There were 17 DEGs designated in the category of carbohydrate esterases (CEs), and mainly distributed in family CE1 and CE6. Enzymes in these families involved acetylxyran esterases, carbohydrate acetyl esterase/feruloyl esterases, pectinesterases, etc., which usually catalyze the de-O or de-N-acylation of substituted saccharides such as acetylated xylan or pectin methyl esters (Biely 2012). 12 DEGs were annotated as carbohydrate-binding modules (CBMs) and distributed in four different families which are CBM6, CBM57, CBM9 and CBM62. CBMs are a sort of non-catalytic proteins that usually constitute large multi-module enzymes as a relatively independent module, and have been reported to promote enzymatic activity by increasing the effective enzyme concentration on the polysaccharide surface (Boraston et al. 2004).

In the F vs G combination, a total of 151 CAZymes were identified, of which 102 were upregulated and 49 were downregulated. A total of 100 CAZymes with 65 upregulated and 35 downregulated were identified in the C vs G combination. In the F vs C combination, a total of 141 CAZymes were identified, with 93 upregulated and 48 downregulated (\* MERGEFORMAT Fig. 3B). The statistical data showed that the CAZymes encoding genes with upregulated expression account for more than 65% in all the three combinations, indicating that these enzymes are induced with the existence of filter paper or cellobiose as substrates in the culture medium. A finer classification of the upregulated CAZymes in each combination was performed. As shown in \* MERGEFORMAT

Fig. 3C, the upregulated CAZymes were mainly distributed in the GH family, accounting for 48% (F vs G\_up), 43% (C vs G\_up) and 49% (F vs G\_up), respectively. By adjusting the screening threshold to  $|\log_2 \text{FC}| \geq 1$  and  $p \leq 0.01$ , 54 significantly upregulated carbohydrate-active enzymes were further screened, mainly including cellulases, hemicellulases, pectinases, etc. (Gouvêa et al. 2018) (\* MERGEFORMAT Table 1).

Most aerobic cellulolytic microorganisms degrade cellulose by secreting a series of free cellulases, while most anaerobic cellulolytic microorganisms perform the degradation process via cellulosome, a large multi-enzyme complex which binds to the outer surface of the microorganisms (Wilson 2009; Bayer et al. 2004). In both strategies, synergistic action of endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.91) and  $\beta$ -glucosidases (EC 3.2.1.21) is required to achieve effective degradation (Zhang et al. 2006). Recently, a few cellulolytic bacteria such as *Cytophaga hutchinsonii* and *Fibrobacter succinogenes*, have been proposed to degrade cellulose by a third strategy in which the bacteria neither secret free cellulases nor possess any cellulosomes (Suen et al. 2011; Zhao et al. 2020). The mechanism of this novel strategy still remains unclarified. Moreover, analysis of genomic sequence of *Cytophaga hutchinsonii* revealed no obvious homologues of known exoglucanases which are generally thought to be critical for the degradation of crystalline cellulose (Xie et al. 2007). As shown in \* MERGEFORMAT Table 1, endoglucanases (mainly distributed in GH5, GH8 and GH9 families) and  $\beta$ -glucosidases (mainly



**Fig. 4** Heat map of clustering of differentially expressed gliding function-related genes. (F/C/G: fermentation supernatant using filter paper, cellobiose and glucose as the sole carbon source, respectively.)

distributed in GH3 family) were synergistically upregulated. However, no exoglucanase or polysaccharide monooxygenase encoding genes (Taillefer et al. 2018) were identified in the genomic database of strain CX11 (data unpublished). Besides, the strain CX11 digests cellulose in a contact-dependent manner which is similar to that of *Cytophaga hutchinsonii*. Therefore, we speculated that *Sporocytophaga* sp. CX11 might also use the third novel strategy mentioned above to degrade cellulose (Zhu and McBride 2017).

#### Mining of DEGs related to gliding motility

*Sporocytophaga* sp. CX11 is closely related to *Cytophaga hutchinsonii* and exhibits many similarities in metabolism and lifestyle, such as the mode of contact-dependent degradation and rapid gliding motility over surfaces. It has been reported that *C. hutchinsonii* has orthologs for each of the core gliding motility genes of *Flavobacterium johnsoniae*, and the ability to glide over surfaces may allow cells to migrate along cellulose fibers in search of amorphous regions that are most amenable to enzymatic attack (Suen et al. 2011; Xie et al. 2007). *sprP*

**Table 2** Differentially expressed genes belong to 4 different expression patterns

Expression changing pattern	Gene ID	CAZy family	Annotation or putative information
Top 10 differential genes belong to difference changing pattern 0			
0	CX11_GM003764	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM004191	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM000840	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM003150	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM005552	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM002226	NA	Type IX secretion system membrane protein PorP/SprF
	CX11_GM004712	GH26	Endo-1,4-beta-mannosidase
	CX11_GM002852	GH5	Endoglucanase
	CX11_GM000081	GH10	Endo-1,4-beta-xylanase B
	CX11_GM000628	GH8	Reducing-end-xylose releasing exo-oligoxylanase
Top 10 differential genes belong to difference changing pattern 4			
4	CX11_GM004329	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM000033	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM000661	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM000293	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM004381	NA	Related to the initial and middle phase of cellulose degradation
	CX11_GM001697	NA	Type IX secretion system membrane protein PorP/SprF
	CX11_GM002804	GH5	Endoglucanase
	CX11_GM000815	NA	Gliding motility protein GldM
	CX11_GM004438	GH5	Endoglucanase
	CX11_GM003220	GH5	Endoglucanase 3
Top 10 differential genes belong to difference changing pattern 11			
11	CX11_GM003811	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM003111	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM005704	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM005293	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM000988	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM004267	GH5	Endo-1,4-beta-mannosidase
	CX11_GM001870	NA	Type IX secretion system membrane protein PorP/SprF
	CX11_GM001892	NA	Type IX secretion system membrane protein PorP/SprF
	CX11_GM003799	GH3	Beta-glucosidase
	CX11_GM003851	CBM9	Modules of ~170 residues found in xylanases and cellulase
Top 10 differential genes belong to difference changing pattern 14			
14	CX11_GM000997	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM000199	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM001026	NA	Related to the middle and last phase of cellulose degradation

**Table 2** (continued)

Expression changing pattern	Gene ID	CAZy family	Annotation or putative information
	CX11_GM002155	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM000995	NA	Related to the middle and last phase of cellulose degradation
	CX11_GM002492	GH18	Chitinase
	CX11_GM002954	GH18	Chitinase
	CX11_GM004043	NA	Gliding motility lipoprotein GldH
	CX11_GM003048	NA	Type IX secretion system membrane protein PorP/SprF
	CX11_GM005379	GH3	Beta-glucosidase

The significant patterns of 0, 4, 11 and 14 with more than 30 annotated cellulolytic genes were displayed, and the top 5 of relative abundance of putative cellulolytic genes and candidate cellulolytic genes were listed, respectively

deletion mutant of *C. hutchinsonii* was partially defective in gliding motility and failed to digest crystalline cellulose, suggesting a link between gliding motility and cellulose utilization (Zhu and McBride 2014). Therefore, DEGs related to gliding motility were also investigated in this study. The differential gene sequences were uploaded to the Pfam database for comparison, and 39 DEGs were screened to associate with gliding function (Additional file 1: Table S2). These genes encode three classes of gliding proteins, including Gld (Gliding motility) proteins, Spr (Surface protein) proteins and Sus (Starch utilization system) proteins. The Gld proteins are directly related to the gliding behavior of cells (McBride et al. 2009; Rhodes et al. 2011). The Spr proteins are a class of cell surface proteins that can transmit kinetic substances to support cells' gliding motility (Sato et al. 2021). The Sus proteins mainly participate in the binding and transporting of oligosaccharides, which are important in the digestion of extracellular carbohydrates (Anderson and Salyers 1989). The clustering of expression levels of differential gliding genes under different carbon sources is shown in \\* MERGEFORMAT Fig. 4. Under each carbon source condition, there are corresponding high expressed gliding genes, indicating that these gliding genes act at different stage of cellulose degradation.

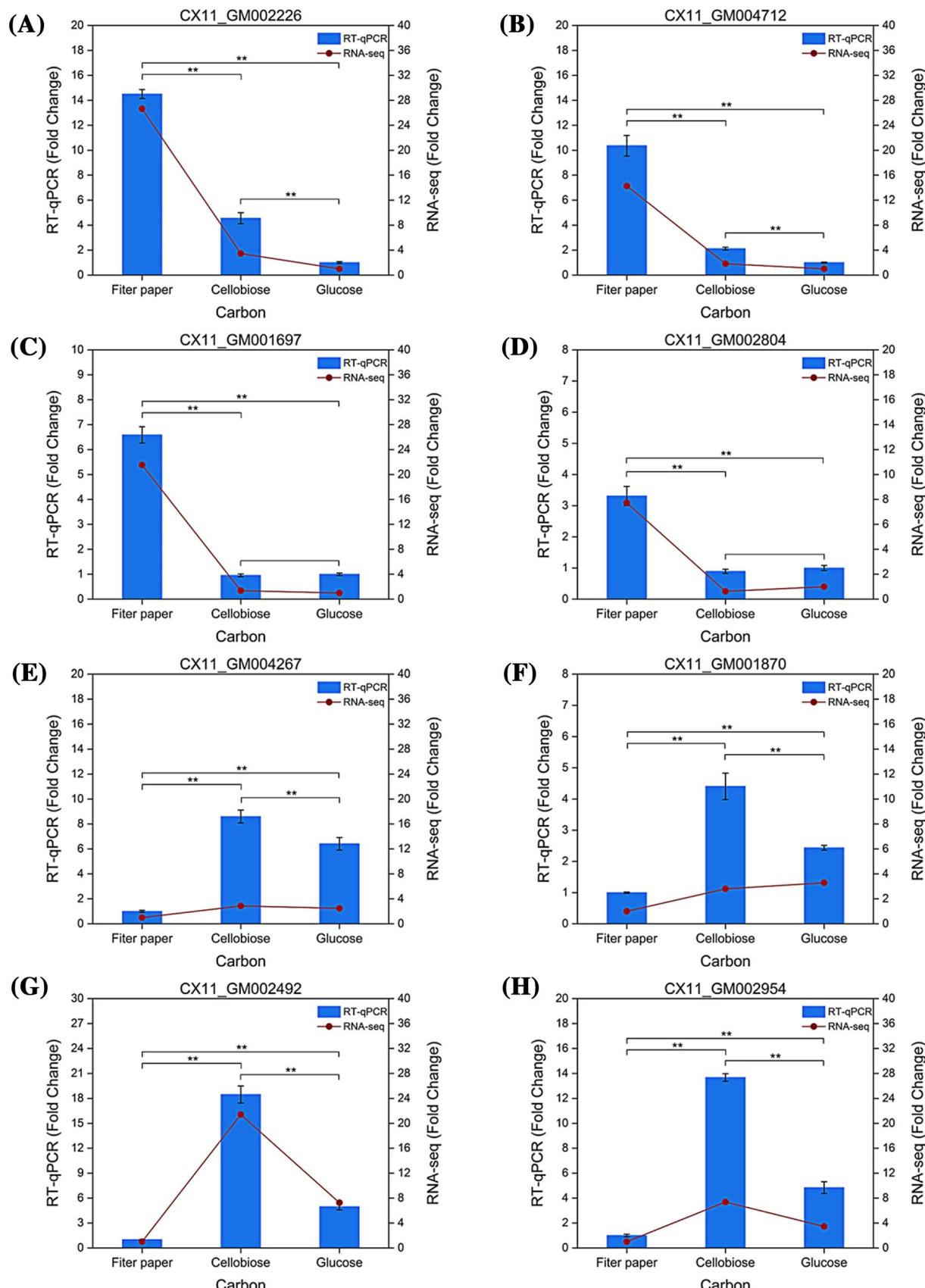
**Short-time series expression miner algorithm explores more cellulolytic enzymes from *Sporocytophaga* sp. CX11**  
Although functional annotations of DEGs have been performed through databases such as GO, KEGG and CAZy, there were still a large number of genes whose functions

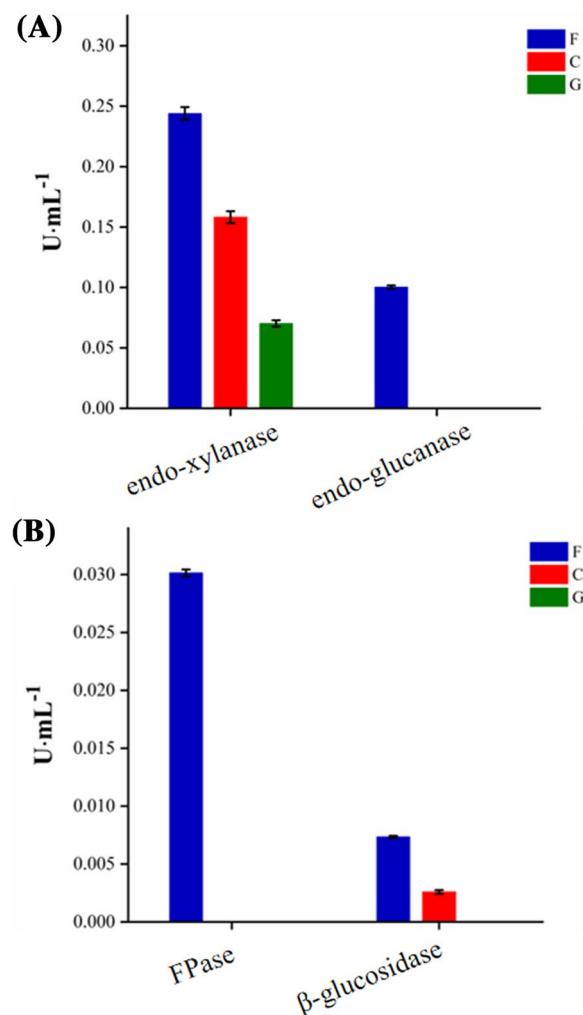
cannot be annotated. To further explore these unannotated genes potentially related to cellulose metabolism, cluster analysis was performed using the STEM algorithm based on different abundance variation patterns of genes, according the procedures proposed by (Yang et al. 2021). Based on the abundance variation patterns, functionally unannotated genes were clustered together with annotated cellulolytic genes and were thus assigned a certain function similar to those of annotated cellulolytic genes from the same pattern. In this study, a total of 16 patterns of gene expression changes were obtained (Additional file 1: Figure S4). Among them, four typical patterns (0, 4, 11 and 14), in which more than 30 genes with annotated functions, were selected for clustering (Table 2).

Pattern 0 and 4 are both downward patterns, with all genes in these patterns being expressed at relatively high levels under filter paper conditions. Most of the genes with known functions are annotated as endoglucanases, xylanases and gliding proteins. Therefore, the 281 unannotated genes in pattern 0 and 4 were presumed to be associated with the decomposition of polysaccharides such as cellulose and xylan, which usually own a high degree of polymerization. Pattern 11 and 14 are both upward patterns, and all genes in these patterns are expressed at relatively high levels under cellobiose or glucose conditions. Most of the annotated genes are β-glucosidases and some gliding proteins, suggesting that the 289 unannotated genes in pattern 11 and 14 might function mainly in the degradation of cellobiose or other cellulosic substrates with lower degree of polymerization

(See figure on next page.)

**Fig. 5** Comparison of RT-qPCR results with RNA-seq results. A and B are the top 2 DEGs in "0" pattern. C and D are the top 2 DEGs in "4" pattern. E and F are the top 2 DEGs in "11" pattern. G and H are the top 2 DEGs in "14" pattern. \*\*means  $p < 0.01$ , highly significant difference, and no \*means no significant difference

**Fig. 5** (See legend on previous page.)



**Fig. 6** Enzyme activities measured under different carbon source conditions. **A:** Endo-xylanase activity and endo-glucanase activity in fermentation broth under different carbon source conditions. **B:** FPase activity and β-glucosidase activity in fermentation broth under different carbon source conditions. (F/C/G: fermentation supernatant using filter paper, cellobiose and glucose as the sole carbon source, respectively.)

and basically existed in the hydrolysates of cellulosic polysaccharides.

#### Validation of the transcriptome data by RT-qPCR

RT-qPCR was used to further validate the expression level of genes identified in Illumina sequencing analysis. The top 2 functionally annotated genes in each typical pattern (highlighted in /\* MERGEFORMAT Table 2) were chosen as candidates for study (Additional file 1: Table S3). As demonstrated in Fig. 5, these 8 DEGs

showed a consistent pattern between Illumina sequencing results and RT-qPCR, indicating that the transcriptome data are reliable and accurate.

#### In situ enzymatic activity assays

As shown in Table 1, the 54 significantly upregulated carbohydrate-active enzymes were mainly designated as endoglucanases, xylanases or β-glucosidases. In order to assess the plausibility of the transcriptome analysis and to further identify the biological function of the strain CX11, enzymatic activity assays were performed. The strain was cultured under three different carbon source conditions, and extracellular enzymatic activities were determined.

When using filter paper as the sole carbon source, endo-xylanase, FPase, endo-glucanase and β-glucosidase activities were detected in the supernatant of fermentation broth, among which the endo-xylanase activity was the highest. When the strain was cultured with cellobiose as the sole carbon source, endo-xylanase and β-glucosidase activities were identified. Only endo-xylanase could be detected in the broth with glucose as the sole carbon source (Fig. 6). Endo-xylanase activity was detected under all the three different carbon source conditions, indicating that this kind of enzyme might be constitutively synthesized and secreted by strain CX11. While enzymes mainly acting on cellulose are most probably be induced by cellulosic substrates, which is consistent with the results mentioned above that most genes encoding cellulases were upregulated expressed with the existence of filter paper or cellobiose in the culture medium. All these results further confirmed that strain CX11 has a great potential application in the degradation of cellulosic substrates.

#### Conclusions

This is the first transcriptomic study of *Sporocytophaga* sp., and many DEGs associated with lignocellulose degradation were identified. These DEGs were involved in multiple GO functional category such as carbohydrate metabolic processes, carbohydrate binding and o-glycosyl compound hydrolase activity. The functional analysis and expression patterns of DEGs, as well as the prediction of functionally unannotated genes related to lignocellulose degradation, lay the foundation for further exploration of biological functions of strain CX11. The results of this study will also offer considerable potential for the application of candidate genes in the utilization of lignocellulose resources.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40643-023-00629-4>.

**Additional file 1: Figure S1.** Histogram of GO statistical results of differentially expressed genes in F vs G group. **Figure S2.** Histogram of GO statistical results of differentially expressed genes in C vs G group. **Figure S3.** Histogram of GO statistical results of differentially expressed genes in F vs C group. **Figure S4.** 16 expression patterns of differentially expressed genes. **Table S1.** Primers of RT-qPCR. **Table S2.** Summary of differentially expressed genes related to gliding function in groups. **Table S3.** Normalized relative expression of candidate genes under different carbon sources.

## Acknowledgements

This work was supported by the National Natural Sciences Foundation of China and Natural Science Foundation of Liaoning Province.

## Author contributions

JW performed parts of differential transcriptome analysis, enzymatic activity assays and wrote the manuscript; YZ and XS performed parts of differential transcriptome analysis and RT-qPCR experiment; XL and XW assisted with some experiment executing; FY and XC planned and supervised the work and revised the manuscript. All authors read and approved the final manuscript.

## Funding

Financial support provided by the National Natural Sciences Foundation of China (32072160, 31671796, 31801469), Natural Science Foundation of Liaoning Province (J2020041, 2020-MS-276), Key Research Projects of The Educational Department of Liaoning Province (LJKZZ20220061).

## Availability of data and materials

All data produced or analyzed for this study are included in the published article and its additional information files.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interest.

### Author details

<sup>1</sup>School of Biological Engineering, Dalian Polytechnic University, Ganjingziqu, Dalian 116034, People's Republic of China.

Received: 9 October 2022 Accepted: 12 January 2023

Published online: 30 January 2023

## References

- Anderson K, Salyers A (1989) Biochemical evidence that starch breakdown by *Bacteroides thetaiotomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J Bacteriol* 171(6):3192–3198. <https://doi.org/10.1128/jb.171.6.3192-3198.1989>
- Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58:521–554. <https://doi.org/10.1146/annurev.micro.57.030502.091022>
- Biely P (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnol Adv* 30(6):1575–1588. <https://doi.org/10.1016/j.biotechadv.2012.04.010>
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382:769–781. <https://doi.org/10.1042/BJ20040892>
- Burchard PR (1981) Gliding motility of prokaryotes: ultrastructure, physiology, and genetics. *Annu Rev Microbiol* 35(35):497–529. <https://doi.org/10.1146/annurev.mi.35.100181.002433>
- Chen X, Mao X, Huang J, Yang D et al (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 39:316–322. <https://doi.org/10.1093/nar/gkr483>
- Drula E, Garron M, Dogan S, Lombard V, Henrissat B, Terrapon N (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 50:D571–D577. <https://doi.org/10.1093/nar/gkab1045>
- Ernst J, Nau GJ, Bar-Joseph Z (2005) Clustering short time series gene expression data. *Bioinformatics* 21:i159–i168. <https://doi.org/10.1093/bioinformatics/bti022>
- Fang W, Song R, Zhang X, Zhang X, Zhang X, Wang X, Fang Z, Xiao Y (2014) Characterization of a novel beta-glucosidase from *Gongronella* sp. W5 and its application in the hydrolysis of soybean isoflavone glycosides. *J Agric Food Chem* 62(48):11688–11695. <https://doi.org/10.1021/jf502850z>
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* 42:D222–230. <https://doi.org/10.1093/nar/gkt1223>
- Gaskell J, Blanchette RA, Stewart PE, BonDurant SS, Adams M, Sabat G, Kersten P, Cullen D (2016) Transcriptome and secretome analyses of the wood decay fungus *Wolfiporia cocos* support alternative mechanisms of lignocellulose conversion. *Appl Environ Microbiol* 82(13):3979–3987. <https://doi.org/10.1128/AEM.00639-16>
- Gouvêa PD, Bernardi AV, Gerolamo LE, Emerson D, Riano-Pachón D, Uyemura SA, Dinamarco TM (2018) Transcriptome and secretome analysis of *Aspergillus fumigatus* in the presence of sugarcane bagasse. *BMC Genomics* 19(1):232. <https://doi.org/10.1186/s12864-018-4627-8>
- Guerrero EB, de Villegas RMD, Soria MA, Santangelo MP, Campos E, Talia PM (2020) Characterization of two GH5 endoglucanases from termite microbiome using synthetic metagenomics. *Appl Microbiol Biotechnol* 104(19):8351–8366. <https://doi.org/10.1007/s00253-020-10831-5>
- Holt SC, Leadbetter ER (1967) Fine structure of *Sporocytophaga myxococcoides*. *Arch Mikrobiol* 57(3):199–213. <https://doi.org/10.1007/BF00405947>
- Hutchinson HB, Clayton J (1919) On the decomposition of cellulose by an aerobic organism (*Spirochaeta cytophaga*, n. sp.). *J Agric Sci* 9(02):143. <https://doi.org/10.1017/s0021859600004755>
- Lane S, Dong J, Jin YS (2018) Value-added biotransformation of cellulosic sugars by engineered *Saccharomyces cerevisiae*. *Bioresour Technol* 260:380–394. <https://doi.org/10.1016/j.biortech.2018.04.013>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359. <https://doi.org/10.1038/nmeth.1923>
- Liu L, Gao P, Chen G, Wang L (2014) Draft genome sequence of cellulose-digesting bacterium *Sporocytophaga myxococcoides* PG-01. *Genome Announc* 2(6):e01154-e1214. <https://doi.org/10.1128/genomea.01154-14>
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25(4):402–408. <https://doi.org/10.1006/meth.2001.1262>
- Macconail LE, Burns RT, Anwesha N, Coleman HA, Slevin MK, Kristina G, Madelyn L, Kevin L, Mirna J, Mcneill MS (2018) Unique, dual-indexed sequencing adapters with UMLs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* 19(1):30. <https://doi.org/10.1186/s12864-017-4428-5>
- Mcbride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG, Gotsman E, Wang W, Xu J, Hunnicutt DW (2009) Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* 75(21):6864–6875. <https://doi.org/10.1128/AEM.01495-09>
- Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* 31(3):420–428. <https://doi.org/10.1021/ac60147a030>
- Rhodes RG, Samarasam MN, Groll EV, McBride MJ (2011) Mutations in *Flavobacterium johnsoniae* sprE result in defects in gliding motility and protein secretion. *J Bacteriol* 193(19):5322. <https://doi.org/10.1128/JB.05480-11>
- Sato K, Naya M, Hatano Y, Kondo Y, Sato M, Narita Y, Nagano K, Naito M, Nakayama K, Sato C (2021) Colony spreading of the gliding bacterium

- Flavobacterium johnsoniae* in the absence of the motility adhesin SprB. *Sci Rep* 11(1):967. <https://doi.org/10.1038/s41598-020-79762-5>
- Stanier RY (1942) The cytophaga group: a contribution to the biology of mycobacteria. *Bacteriol Rev* 6(3):143–196. <https://doi.org/10.1128/br.6.3.143-196.1942>
- Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D, Brumm PJ (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS ONE* 6(4):e18814. <https://doi.org/10.1371/journal.pone.0018814>
- Sukumar RK, Christopher M, Kooloth-Valappil P, Sreeja-Raju A, Mathew RM, Sankar M, Puthiyaramadam A, Adarsh VP, Aswathi A, Rebino V, Abraham A, Pandey A (2021) Addressing challenges in production of cellulases for biomass hydrolysis: targeted interventions into the genetics of cellulase producing fungi. *Bioresour Technol*. <https://doi.org/10.1016/j.biortech.2021.124746>
- Taillefer M, Arntzen M, Henrissat B, Pope PB, Larsbrink J (2018) Proteomic dissection of the cellulolytic machineries used by soil-dwelling *bacteroidetes*. *mSystems* 3:e00240. <https://doi.org/10.1128/mSystems.00240-18>
- Wang L, Feng Z, Wang X, Wang X, Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26(1):136–138. <https://doi.org/10.1093/bioinformatics/btp612>
- Wilson DB (2009) Evidence for a novel mechanism of microbial cellulose degradation. *Cellulose* 16(4):723–727. <https://doi.org/10.1007/s10570-009-9326-9>
- Xie G, Bruce DC, Challacombe JF, Chertkov O, Detter JC, Gilna P, Han CS, Lucas S, Misra M, Myers GL, Richardson P, Tapia R, Thayer N, Thompson LS, Bretton TS, Henrissat B, Wilson DB, McBride MJ (2007) Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl Environ Microbiol* 73(11):3536–3546. <https://doi.org/10.1128/AEM.00225-07>
- Yang Q, Guo B, Sun H, Zhang J, Liu S, Hexige S, Yu X, Wang X (2017) Identification of the key genes implicated in the transformation of OLP to OSCC using RNA-sequencing. *Oncol Rep* 37(4):2355–2365. <https://doi.org/10.3892/or.2017.5487>
- Yang M, Zhao J, Yuan Y, Chen X, Yang F, Li X (2021) Comparative metagenomic discovery of the dynamic cellulose-degrading process from a synergistic cellulolytic microbiota. *Cellulose* 28(4):2105–2123. <https://doi.org/10.1007/s10570-020-03671-z>
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:W293–297. <https://doi.org/10.1093/nar/gkl031>
- Young M, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11(2):R14. <https://doi.org/10.1186/gb-2010-11-2-r14>
- Zhang HP, Himmel ME, Mielenz JR (2006) Research review paper outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv* 24(5):452–481. <https://doi.org/10.1016/j.biotechadv.2006.03.003>
- Zhao D, Wang Y, Wang S, Zhang W, Qi Q, Lu X (2020) A disulfide oxidoreductase (CHU\_1165) is essential for cellulose degradation by affecting outer membrane proteins in cytophaga hutchinsonii. *Appl Environ Microbiol*. <https://doi.org/10.1128/AEM.02789-19>
- Zhao Y, Li M-C, Konaté MM, Chen L, Das B, Karlovich C (2021) TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *J Transl Med* 19(1):269. <https://doi.org/10.1186/s12967-021-02936-w>
- Zheng Q, Wang X (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res* 36(S2):W358–W363. <https://doi.org/10.1093/nar/gkn276>
- Zhu Y, McBride MJ (2014) Deletion of the *Cytophaga hutchinsonii* type IX secretion system gene sprP results in defects in gliding motility and cellulose utilization. *Appl Microbiol Biotechnol* 98(2):763–775. <https://doi.org/10.1007/s00253-013-5355-2>
- Zhu Y, McBride MJ (2017) The unusual cellulose utilization system of the aerobic soil bacterium *Cytophaga hutchinsonii*. *Appl Microbiol Biotechnol* 101(19):7113–7127. <https://doi.org/10.1007/s00253-017-8467-2>

## Publisher's Note

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

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)