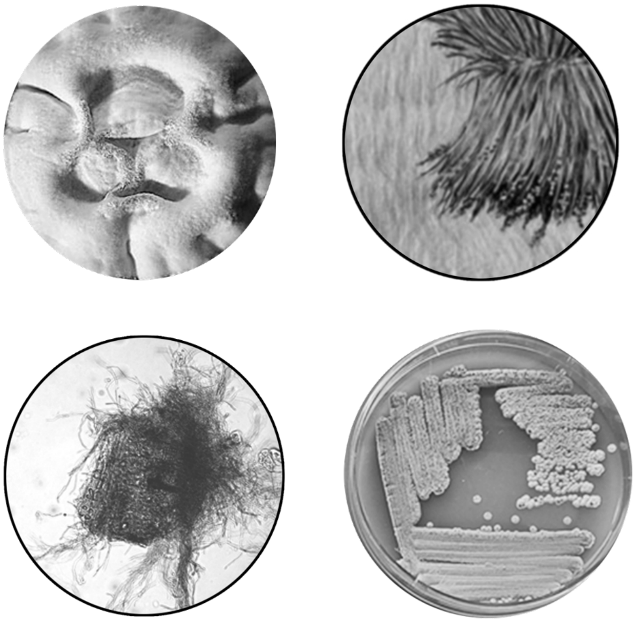
Unraveling modular architecture and domain engineering of carbohydrate-active enzymes: key insights for sustainable bio-based processes

Sidar, A.

Citation

Sidar, A. (2024, December 3). *Unraveling modular architecture and domain engineering of carbohydrate-active enzymes: key insights for sustainable bio-based processes*. Retrieved from https://hdl.handle.net/1887/4170491



Chapter 3

Highly Variable Domain Architecture in Carbohydrate-Active Enzymes Highlights *Streptomyces* as Promising Resource for Rice Straw Bioconversion

**Andika Sidar, Gerben P. Voshol, Ahmed El-Masoudi, Erik Vijgenboom,
Peter J. Punt**

Published in Bioresource Technology Reports, (2024).
DOI: https://doi.org/10.1016/j.biteb.2024.101775

Abstract

Streptomyces species secrete interesting carbohydrate-active enzymes (CAZymes) with unique domain organization, setting them apart from other soil-dwelling filamentous organisms like filamentous fungi. These *Streptomyces* CAZymes have potential for lignocellulose degradation, enhancing their application in biotechnological processes like biofuel production. However, the efficacy of *Streptomyces* enzymes varies based on the strain and enzymes produced. To explore enzymes involved in lignocellulosic rice straw degradation, eight *Streptomyces* strains were screened for hydrolytic enzyme activity. Three strains C34, MBT40, and MBT70 exhibited significant (ligno)cellulolytic activity, efficiently utilizing rice straw as a carbon source in submerged fermentation. Proteome analysis revealed a diverse array of secreted CAZymes in each strain, illustrating a highly variable hydrolytic enzyme repertoire. Although enzymes with similar functions were identified, the domain organization of the most abundant CAZymes differed significantly among the strains, highlighting the competency of different enzyme cocktails among *Streptomyces* strains in rice straw degradation.

Keywords: *Streptomyces, c*arbohydrate-active enzymes, enzyme domain organization, lignocellulose degradation, rice straw, proteome

3.1 Introduction

Lignocellulose, including rice straw, is a complex structure composed mainly of cellulose (40–60%), hemicellulose (20–40%), and lignin (10–24%) (Putro et al., 2016; Ramos et al., 2023). Rice straw represents one of the highly available lignocellulose sources worldwide as a byproduct of the rice harvesting (Mujtaba et al., 2023). In 2021, global rice production accounted for up to 787 million tons from 165 million hectares of harvested area (FAOSTAT, 2023) and resulted in around 500 million tons of rice straw per year (Van Hung et al., 2020). Meanwhile, the majority of the rice straw biomass is still not utilized optimally and ends up in the waste streams (Van Hung et al., 2020).

Recognized as a promising renewable carbohydrate resource, rice straw holds potential for producing biofuels and various biochemical products (Ramos et al., 2023; Satlewal et al., 2018). However, the complex structure of lignocellulose polymers arises from the intricate organization of cellulose, hemicellulose, and lignin, creating physical barriers that limit cellulase accessibility (Chang and Holtzapple, 2000; Satlewal et al., 2018; Zhu et al., 2008). Therefore, research has explored various strategies, including the development of enzymatic cocktails to enhance the efficient degradation of lignocellulosic biomass (Contreras et al., 2020).

The efficiency of lignocellulose degradation crucially depends on the performance of hydrolytic enzymatic cocktails produced by microorganisms (Gavande et al., 2021; Pedraza-Zapata et al., 2017). Microorganisms especially from the group of actinobacteria, firmicutes, and proteobacteria, has been demonstrated to play a pivotal role in the rapid degradation of hemicellulose and cellulose in rice straw biomass (Gavande et al., 2021; Wang et al., 2016). This complex degradation process involves various classes of carbohydrate active-enzymes (CAZymes), including glycoside hydrolase (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), and auxiliary activities (AA) (Drula et al., 2022). Among the CAZyme classes, GHs represent the largest class with 183 CAZyme families (Drula et al., 2022; Gavande et al., 2023). Particularly, GH classified as cellulases and hemicellulases play a major role in the cellulolytic system of lignocellulose degradation (Gavande et al., 2021).

GH enzymes are generally modular enzymes, consisting of a catalytic domain associated with one or multiple accessory domains that do not have catalytic activity (Berlemont et al., 2022; Sidar et al., 2020). The carbohydrate binding modules (CBM) are well-known representatives of the accessory domains with a recognized function to bind polysaccharides and to bring the enzyme in closer proximity toward its substrate (Gilbert et al., 2013). Other accessory domains were reported as well, but their function awaits elucidation, including fibronectin type three domain (Fn3), several types of Immunoglobulin-like domains, as well as domains of unknown function, so called DUF (Han et al., 2013; Lv et al., 2023). In the commonly found enzyme domain organization, the accessory domain is present at the C and/or N-terminus of the catalytic module, either as a single or as multiple copies of the same modules or in a combination of different accessory domains (Berlemont et al., 2022). Variation in the domain organization of glycoside hydrolases influences the hydrolysis performance of the enzymes (Hu et al., 2021; Sidar et al., 2023).

Lignocellulose degrading microorganisms are known to produce abundant CAZymes with diverse domain organization consisting of various accessory domains (Sidar et al., 2020; Talamantes et al., 2016). *Streptomyces* species represent a group of actinomycete bacteria producing a wide range of metabolites and CAZymes (Malik et al., 2020). Notably, *Streptomyces* strains have demonstrated significant cellulolytic activity and efficiency in degrading lignocellulosic substrates, such as rice straw (Feng et al., 2021; Xu and Yang, 2010). Moreover, research highlights *Streptomyces'* ability to secrete specific CAZymes including cellulase, xylanase, as well as laccase, when utilizing rice straw as substrate (Niladevi et al., 2007; Zhang et al., 2016) However, it is important to note that specific details regarding the genome mining of CAZyme families along with the modular-domain architecture, as well as proteome data set from the best-performing *Streptomyces* strains in relation to their secretion of a diverse array of CAZyme cocktails, specifically for rice straw degradation, have not yet been reported.

In the research presented here, several *Streptomyces* strains were explored for their lignocellulose-degrading enzymes based on genome mining and enzymatic activity assays on rice straw. Proteomic analysis was then performed specifically on the three best-performing strains in rice straw degradation. This analysis aimed to explore variations in CAZymes and their domain architecture, providing insight into the potential CAZymes involved in lignocellulosic rice straw degradation. Moreover, this highlights that the different enzyme cocktails among *Streptomyces* strains are competent for the degradation of lignocellulosic rice straw.

3.2 Materials and Methods

3.2.1 CAZyme specific genome analysis of *Streptomyces* strain

The genomic sequence data of eight *Streptomyces* strains (Table 1) available in the strain collection from Molecular BioTechnology (MBT), Institute of Biology Leiden was mined for their genes expressing CAZymes. The genes encoding CAZymes were predicted using dbCAN V12 program (Zheng et al., 2023). Furthermore, the complete domain organization of CAZymes particularly for the class of glycoside hydrolase enzyme correlated with hemicellulases and cellulases was identified using a server program called Interproscan (https://www.ebi.ac.uk/interpro/search/sequence/, accessed on 19 May 2023). Pairwise sequence alignment between GH representative sequences from each strain was computed by Clustal2.1 (https://www.ebi.ac.uk/Tools/msa/muscle/, accessed on 4 September 2023) to interpret the percent identity matrix (see Supplementary material). The representative sequence required for Clustal analysis was obtained from the GH sequence showing the lowest E-value against the CAZyme family HMMs (Hidden Markov Models) in CAZyme annotation via dbCAN V12 program (Zheng et al., 2023) when compared to the other sequences (data not shown). In the percent identity matrix (see Supplementary material), the highest similarity of a pairwise sequence was selected as the basis for GH sequence comparison which further be visualized in heatmap. For the CAZymes identification of proteome, protein annotation from in house genomic data was used as database to identify CAZymes in the proteome of three selected *Streptomyces* strains grown on rice straw.

Table 1. List of *Streptomyces* strains used in this study

Strains	Accession no.	References
Streptomyces C34	AZSD00000000.1	(Busarakam et al., 2014)
Streptomyces MBT21	KJ413756	(Zhu et al., 2014)
Streptomyces MBT28	LARV00000000.1	(Wu et al., 2015)
Streptomyces MBT39	KJ413774.1	(Zhu et al., 2014)
Streptomyces MBT40	KJ413775.1	(Zhu et al., 2014)
Streptomyces MBT63	JAIZ00000000	(Girard et al., 2014)
Streptomyces MBT66	JAIY00000000	(Girard et al., 2014)
Streptomyces MBT70	JAENRW000000000.1	(Westhoff et al., 2021)

3.2.2 Strains and culture conditions

The *Streptomyces* strains used in this study were obtained from the MBT strain collection. Prior to cultivation, fresh spores of the *Streptomyces* strains were generated by inoculating the strains from preserved stocks on agar plates using Soya Flour Mannitol (SFM) medium and incubation at 30°C for 7 days. The spores were harvested by pouring 10 mL of sterile Milli-Q water on the SFM-plate and gently scratching the colonies using a sterile cotton stick. The liquid containing the spores was transferred into a 15 mL sterile screw-cap tube and spores were pelleted by centrifugation for 15 min at 3500 ×*g*. Supernatant was discarded and the pellet was resuspended in 1 mL of 30 % glycerol solution using a vortex. The spores were stored at −20°C until further use.

3.2.3 Rice straw preparation and pretreatment

Rice straw was obtained from local farmers in Sleman district, Yogyakarta, Indonesia. Rice straw was washed, oven-dried (40°C, 24 h) and packed in airtight vacuum sealer plastic bags to keep it dry as well as to avoid deterioration. Furthermore, alkaline-based pretreatment was carried out to make the polysaccharides in complex lignocellulose materials more accessible to the enzymes as described by Valles et al. (2021). It has been reported that alkaline pretreatment is a more effective delignification process, increasing the release of the residual carbohydrates for the enzymes to attack (Kim and Han, 2012; Kim et al., 2016). Hundred gram of rice straw was cut into small pieces, washed with distilled water and eventually soaked in 2% NaOH in a 1:10 ratio (1 g rice straw fibers/10 mL 2% NaOH) for 3 h at 55°C. Subsequently, the pretreated rice straw was separated from the NaOH solution using miracloth filtration (Merck, 475,855) with the mesh size 22–25 μm and then washed 2 times with distilled water to neutralize the pH. Drying was done overnight in the incubator at 60°C until it reached approximately a constant weight. Lastly, the pretreated rice straw was grinded to reach a further reduced particle size by a blender machine and then it was sieved to obtain pretreated rice straw particle with approximately 0.4 mm length in size.

3.2.4 Submerged fermentation of rice straw

The fermentation was carried out by inoculating 10^6 spores/mL of each *Streptomyces* strains in 25 mL of liquid medium containing 250 mg of pre-treated rice straw, using a 100 mL Erlenmeyer baffled flask. Liquid minimal medium consists of the following compounds (g/L): K_2HPO_4 (0.5); $MgSO_4.7H_20$ (0.2); $(NH_4)_2Cl$ (1.0) and 2 mL of trace elements solution as described in Kieser et al. (2000). Cultivation was conducted in a shaker incubator (New Brunswick scientific, Innova 44®) at 28°C, 160 rpm, for 10 days. Crude enzyme samples, in total 40 mL, were collected at 24-hour intervals and stored as aliquots at −20°C for the evaluation of hydrolytic activity. The rice straw particles and the *Streptomyces* mycelium were removed by centrifugation. The inoculations were done in duplicate. The negative control was a shake flask prepared as stated above but without adding spores.

3.2.5 Enzymatic analysis

For qualitative cellulolytic enzyme determination, spores from eight *Streptomyces* strains were spotted on AZCL-cellulose plates, containing 1% agarose with 0.05% of AZCL-cellulose compounds (Megazyme), supplemented with 1% pretreated rice straw. The plates were subsequently incubated for 3 days at 30°C. Cellulase-positive strains were identified by the formation of a blue halo around the colony.

To evaluate lignocellulose degradation in the cultivation samples from *Streptomyces* grown on rice straw containing medium, overall carbohydrate hydrolytic enzyme activity was determined using reducing sugar release according to the 3.5-dinitrosalicylic acid (DNS) method (Miller, 1959). Spent medium samples from submerged fermentation were collected and centrifuged to separate the supernatant from residual rice straw and mycelium suspension. Subsequently, a 50 µL aliquot containing 25 µL of supernatant and 25 µL of 0.05 M NaAc-buffer (pH 5) was mixed with 50 µL of DNS reagent. The detailed DNS reagent comprises 1% 3.5-dinitrosalicylic acid, 0.05% sodium sulfite, and 1 % sodium hydroxide. The samples were incubated at 100°C for 10 min to allow the reaction. After incubation the reaction was stopped with 30 µL of sodium-potassium tetrahydrate solution (40% w/v). Absorbance was measured at 540 nm (TECAN, Spark 10 M) and plotted against a standard curve to quantify the presence of reducing sugar released (mM). The measurements were done in duplicate. Samples of supernatant from the control shake flask stated in Section 2.4. were used as a negative control.

Further enzyme activity analysis was carried out by incubating the cell-free supernatant samples with fresh pretreated rice straw. Crude samples from a 7-days shake flask cultivation (160 RPM, 28°C) were centrifuged at 14000 rpm for 10 min at 4°C (Eppendorf, Centrifuge 5415 R). Subsequently, supernatant was filtered through sterile PVDF Syringe Filter 0.2 µM (Fisherbrand™) and concentrated 5 times using a spin filter (Vivaspin, 10,000 MWCO, VS0102). In a 1.5 mL Eppendorf tube, 100 µL of concentrated spent medium was mixed with 400 µL NaAc buffer (0.05 M, pH 5.0) containing 5 mg of pre-treated rice straw, and then incubated at 30°C for 48 h. Subsequently, the DNS assay was conducted. The sample incubation with the fresh rice straw was done in triplicate.

3.2.6 Microscopic observation

Microscopic observation of mycelial growth was performed by light microscopy Axioplan 2 microscope (Carl Zeiss) equipped with an imaging system using AxioCam camera and AxioVision 4.8 software. For staining, 250 µL crude liquid culture from submerged fermentation was mixed with 750 µL staining solution containing 0.4% trypan blue (Merck, CI23850) dissolved in acetic acid, a fixative solution for mycelium as previously described by Wilkes et al. (2020). An aliquot of the mixing solution was placed in a clean dry microscope slide and covered with a cover slip for microscopic visualization with a 400× of total magnification.

3.2.7 Proteomic analysis

Proteomic analysis was carried out for the best performing strains in lignocellulose rice straw degradation, being strains C34, MBT40, and MBT70. Samples for proteomic analysis were prepared from 15 mL spent medium of cultures grown for 7 days. The spent medium was centrifuged at 14,000 rpm for 10 min at 4°C. The supernatant was collected and filtered through sterile PVDF Syringe Filter 0.2 µM (Fisherbrand™) to remove remaining mycelium and small particles of rice straw. Further ultrafiltration was also conducted using a spin column with 3 kDa molecular cut-off membrane (Amicon® Ultra-15 Centrifugal Filter Units, MWCO 3 kDa) to wash out the small oligosaccharide degradation products of rice straw. The proteins including the CAZymes larger than 3 kDa remain on top of the ultrafiltration device. Phosphate buffer (20 mM, 1 mL in total) was used to collect all protein from the ultrafiltration device. Protein content was measured using the Bradford calorimetric assay (Bio-Rad 5000006) using bovine serum albumin (BSA) as standard (Bio-Rad). The protocol was conducted according to the manufacturer's instruction manual. MS/MS analysis was carried out by Alphalyse (Alphalyse A/S, Odense, Denmark). The tandem mass spectrometry (MS/MS) data converted into mgf format were analyzed using X! Tandem version 17-02-01-4 (Craig and Beavis, 2004).

3.3 Results and discussion

3.3.1 CAZyme specific genome analysis of *Streptomyces* strain

To identify novel cellulolytic enzymes with potential to degrade largely underutilized rice straw-based waste streams, as a first step, the genome sequences from a number of *Streptomyces* strains from the MBT collection were analyzed for the presence of annotated glycoside hydrolase encoding genes (data not showed). Subsequently, eight strains were analyzed in more detail for the complete repertoire of genes encoding CAZymes, as described in method sections. A very broad distribution of CAZymes classes with diverse families was annotated from the genome of these eight *Streptomyces* strains, dominated by GH enzymes comprising of 81 families, followed by polysaccharide lyase (PL), glycosyl transferase (GT), carboxyl esterase (CE), and auxiliary activity (AA) consisting of 17, 15, 11, and eight families respectively (see Supplementary material).

Among the annotated GHs families, 41 families were predicted to belong to the group of β-glucosidases, cellulases, and hemicellulases (see Supplementary material) on the basis of CAZymes database (www.cazy.org, accessed on 1st of August 2023) (Drula et al., 2022). Within these families, three GHs including GH1, GH3, and GH4 were predicted as β-glucosidases (see Supplementary material). Furthermore, GH6, GH9 and GH48 were predicted as cellulases (see Supplementary material). In total 29 families of GH represent hemicellulases with a widespread activity including xylanase, galactosidase, mannosidase, glucuronidase, arabinofuranosidase, xyloglucanase, mannanase, rhamnosidase, and xylosidase (see Supplementary material), as well as four families consisting of GH5, GH12, GH51, GH74 are predicted as both cellulases and hemicellulases (see Supplementary material). In addition, a wide variety of enzyme domain organizations was revealed throughout these (hemi)cellulose degrading protein families (see Supplementary material).

This research was focused on mining the (hemi)cellulolytic GH families and their domain organization because cellulose and hemicellulose are the main components of rice straw polysaccharides. The β-glucosidase was also included in the genome mining since this enzyme promotes complete degradation of cellulose by cleaving the cellulose degradation product such as cellobiose and cellooligosaccharides into glucose, the desired end product in the cellulose degradation (Bai et al., 2013). All the strains in this research contain numerous (hemi)cellulolytic GH enzymes with various domain organizations, of which some are present in more than one strain, while many of them appeared to be strain specific (Fig. 1A & B). Overall, highly similar GH domain organizations were observed in two of the strains, MBT28 and MBT39, indicating that most of the GH sequences from these two strains are true orthologues (see Supplementary material & Fig. 1A).

In comparing the different enzyme classes, the GH family 1, 3, and 4 β-glucosidases with their specific domain structures are present in all eight strains (Fig. 1A). However, for almost all other selected families (see Supplementary material) a large variation of domain organizations was observed and different strains presented very different domain structures for the same enzyme family. An example of a GH family with most different domain organizations across the strains is GH2 (see Supplementary material). In addition, enzymes with more than one catalytic domain were also observed in the genomes, including GH62 coupled with both GH5 and GH10 (see Supplementary material).

No	Domain organizations	Strains
1	GH26_CBM11	C34
2	GH51_CBM2	
3	GH5 CBM10 CBM10 CBM10	MBT21
4	CBM6 GH5	
5	CBM32_GH8	
6	GH30_CBM13_CBM13_CBM13	
7	CBM13_CBM13_GH74	
8	GH10_CBM13_GH62	
9	GH30_Ricin-B-Lectin	MBT28
10	GH2 LamininG3 LacZ Bgal-small-N	MBT39
11	PA14_GH2	
12	GH27_CBM35	
13	CBM13_GH35	
14	GH39	
15	GH78	
16	ThuA_CBM6_GH99	
17	GH106	
18	GH110	
19	GH110_CBM13	
20	GH141_CBM32_CBM32	
21	GH2 CBM42	MBT40
22	PA14_GH2_LamininG3	
23	GH3_ExopC	
24	GH5_FN3_F5/8TypeC	
25	CBM4_CelDN_GH9_CBM2	
26	GH43 CBM61	
27	Big2 GH43	
28	GH93_CBM13_CBM13	
29	GH5_GH62	
30	GH6_CBM2	MBT63
31	CelDN GH9 CBM2	MBT66
32	GH12 RicinBLectin	
33	GH30 CBM13 CBM13	
34	AglN_GH31_DUF5110_CBM32	
35	GH2_IgGlcNase	MBT70
36	GH27_CBM51	
37	GH31 DUF5110	
38	GH43 CBM6	
39	GH43_Ricin-B-Lectin	
40	GH51_FN3_CBM2	
41	GH59_CBM13	
42	GH62	
43	GH10_GH62	

Fig. 1. The occurrences of domain organizations in GH (hemi)cellulases within eight *Streptomyces* strains. (1A) Group of domain organizations that are observed in more than a single strains. The percentage identity of representative sequences (see Section 2.1 and Table S1 for further details) is given as a heatmap chart. The darker the red color, the higher the sequence identity level. (1B) Group of domain organizations that are strain specific. For both 1A & 1B, font highlighted with blue color represents the gene encoding a protein detected in the proteome, while "P" sign in the rectangle indicates in which strain the protein was detected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

The in-silico genome mining was conducted to facilitate the identification of CAZyme-variability within the various strains. This genome mining revealed the abundancy of annotated CAZymes within the *Streptomyces* genomes, where strains C34, MBT39, MBT40, and MBT70 carry a richer variation of domain organizations than the other strains (see Supplementary material). In addition, in strain C34 more than 50% of the GH (hemi)cellulases are associated with a CBM (see Supplementary material). Meanwhile, strain MBT63 showed the lowest number of GH families (see Supplementary material).

Streptomyces, as a diverse genus of bacteria, encompasses numerous strains, each possessing unique genetic makeup and metabolic capabilities (Otani et al., 2022). Inherent differences in the genetic repertoire of these strains corresponds to the variation in CAZyme encoding genes (Gong et al., 2020; Song et al., 2016). Furthermore, previous research has highlighted substantial variation in the domain architecture of GH-cellulases from *Streptomyces* genera, particularly featuring a variety of accessory domains represented by multiple CBM families (Sidar et al., 2020). These structural variations are considered to affect the performance of cellulolytic cocktails for degrading different lignocellulose biomass sources. Consequently, further in vitro analysis is imperative to pinpoint strains demonstrating significant degradation of rice straw polysaccharides. Subsequently, the eight *Streptomyces* strains were subjected to further evaluate for their cellulolytic activity when cultivating them using rice straw as carbon source.

3.3.2 Evaluation of *Streptomyces* strains grown on rice straw

The eight genome analyzed *Streptomyces* strains were further evaluated qualitatively for their capacity in processing cellulose, the main component of rice straw. These strains were grown on solid minimal medium containing rice straw supplemented with AZCL-Cellulose as indicator substrate (Fig. 2). The cellulase activity is visualized by the formation of a blue halo around the spot of mycelial growth.

In this qualitative evaluation, the most significant capacity for degrading cellulose was observed for C34 followed by MBT40 and then MBT70, based on the size of the blue halo formed around the mycelial growth. Smaller and less intense blue halos were detected for MBT66 and MBT39, indicating a lower cellulase activity for these strains, while strains MBT21, MBT28 and MBT63 did not show any indication of cellulase activity. Even though not all the strains demonstrated strong cellulase activity, most of them were able to grow on the rice straw medium, indicated by colony formation. This can be due to the different carbon source preference of these strains (Gavande et al., 2021; Pedraza-Zapata et al., 2017). In rice straw, the carbon source could also be obtained from hemicellulose activity instead of cellulase, which are not detected by AZCL-cellulose. However, no or hardly any growth was observed for both MBT21 and MBT63, suggesting inefficient capability of these two strains to use rice straw as carbon source.

Fig. 2. The cellulases activity of the eight *Streptomyces* strains screened on AZCL-cellulose plates supplemented with rice straw as carbon source. The presence of a blue halo around the colony indicates the degradation of AZCL-cellulose substrate. Relative colony growth is indicated by +, ++, +++, ++++, and NG means no growth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the next experiment the strains were analyzed for their growth and enzyme production in the liquid minimal medium supplemented with rice straw as the sole carbon source. The production of reducing (oligo) sugars was monitored by the quantitative DNS assay to confirm rice straw polysaccharides degradation by the strains. Different strains showed the differences in the amount of reducing sugars released (Fig. 3). The highest concentration of reducing sugars was observed from strain MBT40, followed by C34 and MBT70 during 10 consecutive days of activity measurements (Fig. 3). This is similar to the results obtained in the AZCL-cellulose plate assay experiments. Relatively low concentration of reducing sugars was detected for the other strains, even for MBT39, which is a strain carrying many genes encoding CAZymes in the genome (see Supplementary material). This could suggest that only a particular subset of enzymes is required for efficient rice straw degradation and/or not all the enzymes encoded in the genome are produced in response to the presence of rice straw.

The results of submerged growth experiments showed that the strains released sugars from rice straw to support their growth. Obviously, some of these released sugars will be consumed and therefore, the amount of sugar detected in this analysis (Fig. 3) underestimates the total reduced sugar production. To exclude any influence of sugar consumption of the strains themselves on the interpretation of these results, the spent medium from day-7 of the submerged fermentations was also analyzed for the activity toward fresh rice straw.

In this experiment, spent medium was collected and filtered to remove all *Streptomyces* biomass and rice straw particles followed by concentration. Subsequently, these samples were incubated with fresh rice straw for 48 h and analyzed for the release of reducing sugar equivalents, as mentioned in the methods Section 2.5. Similar as shown in Fig. 3, a clearly different amount of the released reducing sugars for the eight strains was observed. Especially strain MBT40 showed a lower level of reducing sugar compared to C34 and MBT70 (Fig. 4). The variations in sugar consumption among microorganisms can be attributed to the specific sugars (oligomers) released from polysaccharides in the culture medium (Beisel and Afroz, 2016). Given that rice straw contains complex polysaccharides, several types of (oligo) sugars, including glucose, xylose, galactose, rhamnose, mannose, arabinose, and glucuronic acid, are released (Khantibongse and Ratanatamskul, 2023). The differences in released sugars may contribute to the observed variations in sugar consumption by different strains. For example, the presence of more sugar monomers compared to oligomers could influence the strain-specific sugar consumption pattern.

The distinctive subsets of CAZymes produced by these *Streptomyces* strains can contribute to variations in the types and amounts of released reducing (oligo/mono-) sugars during the fermentation process. These differences in sugar profiles, in turn, could lead to distinct sugar consumption patterns by different strains, explaining the difference in reducing sugar amounts released by MBT40 compared with C34 and MBT70 (Fig. 4). Despite the influence of sugar consumption, strains C34, MBT40, and MBT70 consistently emerged as the top performers in their capacity to release carbohydrates from rice straw compared to the other strains (Fig. 2, Fig. 3, Fig. 4).

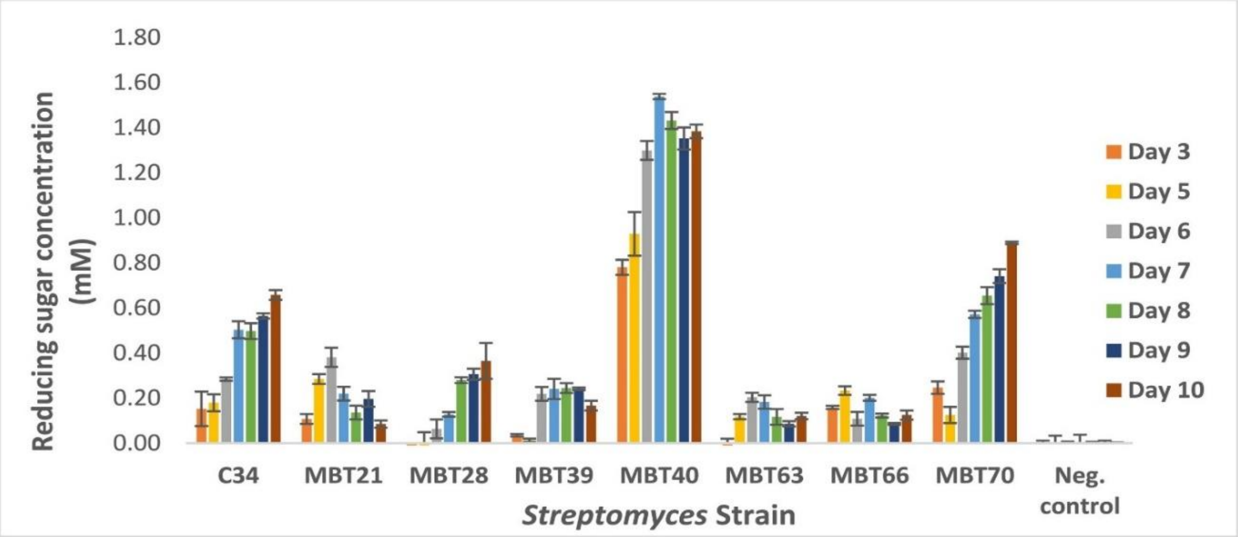


Fig. 3. Reducing (oligo) sugars released from rice straw submerged fermentation by the eight individual strains over 10 days. Error bars represent standard deviation of each experimental point (n = 2). Negative control is the incubated medium without inoculation.

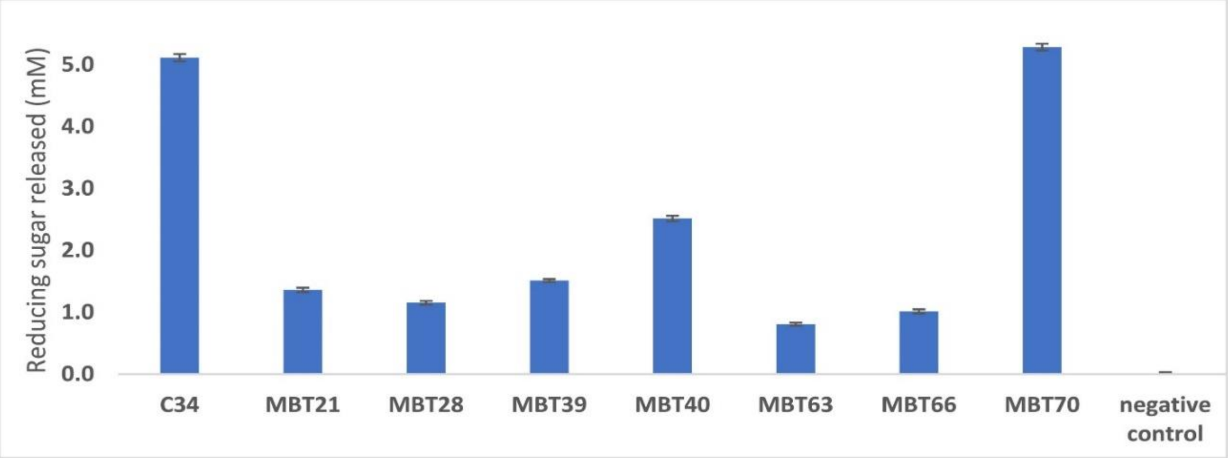


Fig. 4. Enzyme activity of cell-free spent medium on fresh rice straw. The spent medium was collected from day-7 of submerged fermentation and filter-concentrated (see methods Section 2.5), as well as incubated on fresh rice straw for 48 h.

In addition to measuring sugar release, microscopic observation was carried out using samples from the liquid cultures containing strains and rice straw particles to analyze the growth behavior of three best-performing strains, C34, MBT40, and MBT70, during submerged fermentation. All three strains showed a close intertwined interaction with the rice straw particles (Fig. 5).

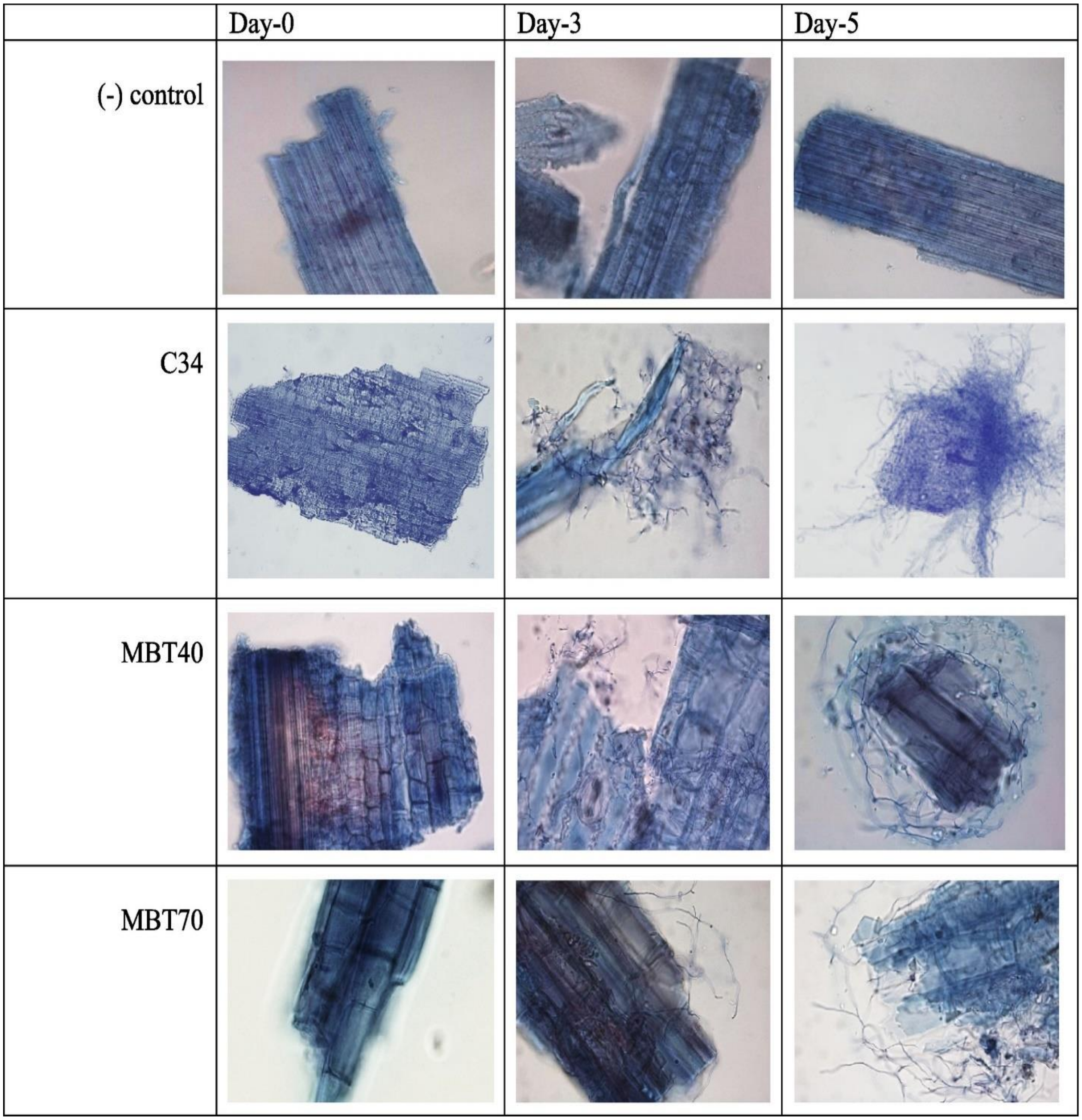


Fig. 5. Microscopic visualization of *Streptomyces* strains grown on liquid minimal medium containing rice straw.

This interaction provides an indication that the attachment between the *Streptomyces* mycelium and the rice straw as substrate may be important for the induction of CAZymes gene expression. This is in line with the reported observation that hydrolytic enzyme production and bacterial growth more likely occurred when the bacterial cell attached to the polysaccharide substrate particles (Guessous et al., 2023), which could be mediated by the different carbohydrate binding domains present in many of the genome-encoded CAZymes in these strains. However, the correlation between growth behavior and CAZyme production in rice straw degradation needs further study.

It has been reported that the substrate-induced CAZymes gene expression, together with different regulatory mechanisms among strains, can significantly influence the production of extracellular proteins (Buccitelli and Selbach, 2020; Daly et al., 2017; Kim et al., 2020). Gene expression across microbial species including *Streptomyces* is tightly controlled by complex regulatory systems at transcriptional and translational level (Kim et al., 2020). Moreover, variations in substrate dependent CAZyme gene expression has been documented, resulting in distinct compositions of enzyme cocktails when the same substrate is utilized within different strains (Song et al., 2016). To further explore if the differences in CAZyme encoding gene content also resulted in a different secreted CAZyme cocktail, a shot-gun proteomic analysis was employed to identify CAZymes corresponding to the detected rice straw degrading activities.

3.3.3 CAZymes identified in the proteome of selected *Streptomyces* strains during cultivation with rice straw as carbon source

Proteomic analysis was applied to reveal any differences in the enzyme cocktails secreted by three best-performing strains, C34, MBT40, MBT70 grown on rice straw medium. From this analysis various classes of CAZymes, including GH, CE, GT, PL, and AA were identified (Table 2). Analysis of the proteome data revealed a significant variation in the most abundantly secreted CAZymes in particular in relation to the domain organization of the detected proteins (highlighted in Fig. 1).

Table 2. CAZymes identified in the proteome of *Streptomyces* strains grown on rice straw medium. (3A) Strain C34, (3B) Strain MBT40, (3C) Strain MBT70. CAZymes predicted active on rice straw are highlighted in blue color. Different classes of CAZymes are separated by an empty row.

(2A).

Accession no[a]	Protein annotation number and description[b]	Peptides		CAZy Family HMM	
		Counts	Coverages	CAZy Class	CBM module
WP_029383598.1	c34_28115 Endoglucanase	2	5%	GH5	CBM2
WP_029381598.1	c34_16665 Endo-1 4-beta-xylanase C	2	4%	GH11	NA
WP_047122023.1	c34_27700 Beta-N-acetylglucosaminidase	2	24%	GH20	CBM32
WP_029385015.1	c34_12585 Family 20 glycosylhydrolase	1	4%	GH20	NA
WP_029386516.1	c34_01230 Beta-galactosidase	3	10%	GH35	NA
WP_029381561.1	c34_27035 Beta-galactosidase bgaB	1	4%	GH42	NA
WP_029386369.1	c34_04085 Exoglucanase	2	4%	GH48	CBM2
WP_242513948.1	c34_04180 Alpha L-arabinofuranosidase	1	6%	GH51	CBM2
WP_029387535.1	c34_00430 Alpha-L-arabinofuranosidase	9	15%	GH62	CBM13
WP_029383255.1	c34_34520 Alpha-1,4 polygalactosaminidase	1	10%	GH114	NA
WP_029382565.1	c34_09670 Glycoside hydrolase	1	5%	GH154	NA
WP_047122325.1	c34_33665 1,3 glucanase	2	5%	GH158	CBM32
WP_029387770.1	c34_20040 Dehydrogenase	1	8%	GH177	NA
WP_029381608.1	c34_11325 Putative oxidoreductase	1	4%	GH179	NA
WP_029385600.1	c34_18530 Laccase	1	4%	AA1	NA
WP_029381143.1	c34_08950 Esterase	2	10%	CE1	NA
WP_029386426.1	c34_02715 Feruloyl esterase family	1	5%	CE1	NA
WP_029380989.1	c34_19235 Polysaccharide deacetylase	2	10%	CE4	NA
WP_078648412.1	c34_17715 Glycosyltransferase family 2	1	6%	GT2	NA
WP_029387625.1	c34_32580 Glycosyltransferase family 4	2	13%	GT4	NA
WP_029381443.1	c34_21620 Glycosyltransferase	1	5%	GT4	NA
WP_029382497.1	c34_06605 Glycosyltransferase	3	5%	GT83	NA
WP_029386222.1	c34_09110 Argininosuccinate lyase	1	4%	PL1	NA
WP_242513963.1	c34_32985 Xanthan lyase	2	5%	PL8	NA

(2B).

Accession no[a]	Protein annotation number and protein description[b]	Peptides		CAZy Family HMM	
		Counts	Coverages	CAZy Class	CBM module
WP_103515247.1	MBT40_34705 Beta-glucoronidase	2	4%	GH2.	NA
WP_306610744.1	MBT40_29175 Endoglucanase	3	6%	GH5	NA
WP_203182926.1	MBT40_14530 Endo-1,4-beta-xylanase A	18	20%	GH10	CBM13
WP_147983992.1	MBT40_05160 Glycoside hydrolase family 18	2	7%	GH18	CBM4
WP_147975503.1	MBT40_29185 Exo-beta-D-glucosaminidase	3	6%	GH39	CBM35
WP_250293383.1	MBT40_09635 Endo-1,5-alpha-L-arabinosidase	1	6%	GH43	NA
WP_306105137.1	MBT40_37915 Glycoside hydrolase family 64	1	5%	GH64	NA
WP_203183683.1	MBT40_09590 cellulase 1,4 beta glucanase	2	6%	GH74	NA
WP_103516255.1	MBT40_17405 O-GlcNAcase	2	5%	GH84	CBM32
WP_306103708.1	MBT40_29325 Glycosyl hydrolase	1	5%	GH87	NA
WP_164286151.1	MBT40_19810 Alpha-1,2-Mannosidase	1	4%	GH92	CBM13
WP_103514531.1	MBT40_00035 Oxidoreductase	1	5%	GH109	NA
WP_187327307.1	MBT40_23630 Glycosyltransferase	1	7%	GT2	NA
WP_203182111.1	MBT40_11150 Glycosyltransferase family 2 protein	1	5%	GT2	NA
WP_103760377.1	MBT40_16445 Alpha-(1- 6)-mannopyranosyltransferase	3	12%	GT87	NA
WP_014158013.1	MBT40_30455 Pectate lyase	1	7%	PL9	NA

(2C.)

Accession no[a]	Protein annotation number and description[b]	Peptides		CAZy Family	
		Counts	Coverages	CAZy Class	CBM module
WP_189757469.1	MBT70_07695 6-phospho-beta-glucosidase	1	3%	GH4	NA
WP_189753610.1	MBT70_33795 Cellulase	3	7%	GH6	CBM2
WP_189754737.1	MBT70_27995 Endo-1,4-beta-xylanase B	8	27%	GH11	NA
WP_189755491.1	MBT70_02555 1,3 Beta-glucanase	1	5%	GH16	CBM13
MBK3525195.1	MBT70_26545 Glycoside hydrolase	1	3%	GH31	NA
WP_189756546.1	MBT70_17725 Beta-galactosidase	3	7%	GH35	NA
MBK3520186.1	MBT70_00535 Extracellular Arabinofuranosidase	3	14%	GH43	CBM13
MBK3526596.1	MBT70_33785 Exoglucanase B	4	4%	GH48	CBM2
GGR84518.1	MBT70_13630 Glycoside hydrolase family 65	2	4%	GH65	NA
MBK3526595.1	MBT70_33780 Xyloglucanase	1	3%	GH74	CBM2
MBK3522212.1	MBT70_11085 GH domain containing protein	1	3%	GH87	NA
WP_189757217.1	MBT70_15640 Glucose--fructose oxidoreductase	1	7%	GH109	NA
MBK3522970.1	MBT70_15025 Endo-1,4-beta-xylanase A	1	5%	GH10; GH62 (multi enzyme domain)	NA
WP_189751528.1	MBT70_13790 Acetyl xylan esterase	1	8%	CE1	NA
WP_189750933.1	MBT70_16675 Acetyl xylan esterase	1	6%	CE1	NA
WP_189755189.1	MBT70_08025 Esterase	1	4%	CE1	NA
WP_189751898.1	MBT70_36110 Esterase	1	7%	CE3	NA
WP_189753254.1	MBT70_37170 Esterase family	1	4%	CE3	NA
MBK3525472.1	MBT70_28000 Polysaccharide deacetylase	7	19%	CE4	NA
WP_189750944.1	MBT70_10835 Polysaccharide deacetylase	2	16%	CE4	NA
MBK3521892.1	MBT70_09395 Bifunctional xylanase/deacetylase	1	5%	CE4	NA
WP_200692102.1	MBT70_01950 Carbohydrate Esterase family10	2	7%	CE10	NA
MBK3521524.1	MBT70_07525 Carboxylesterase	1	6%	CE10	NA
MBK3522093.1	MBT70_10475 Prolyl endopeptidase	2	6%	CE10	NA
WP_189754058.1	MBT70_09340 Alpha/beta hydrolase	2	5%	CE10	NA
WP_189753385.1	MBT70_30455 Rhamnogalacturonan acetylesterase	1	5%	CE12	NA
MBK3521472.1	MBT70_07265 Glycosyltransferase family 1	1	8%	GT1	NA
MBK3525759.1	MBT70_29475 Glycosyltransferase family 2	2	5%	GT2	NA
WP_189757869.1	MBT70_31170 Alpha-D-kanosaminyltransferase	1	4%	GT4	NA
WP_189751861.1	MBT70_26875 Glucosyl-3-phosphoglycerate synthase	1	4%	GT81	NA

[a]Accession no is a protein reference from NCBI database.

[b]Protein annotation number is the ID of an in-house protein annotation from the genome that was used for the protein identification

Among all CAZymes, GHs were the most dominant enzyme class identified in the proteome (Table 2 & Fig. 1). In strain C34 two GH cellulases, GH5_CBM2 endoglucanase and CBM2_Big7_GH48 exoglucanase, were identified both with a CBM2 (Table 2A, Fig. 1), while in MBT40 the GH5 endoglucanase has an FN3 domain, instead of CBM2 domain (Table 2B, Fig. 1). In strain MBT70 the GH5 was not detected but GH6 and GH48 cellulases carrying CBM2 were found in the proteome (Table 2C). Moreover, GH4 annotated as a β-glucosidase was exclusively detected in MBT70. This variation of secreted GH cellulases may reflect the differences in cellulolytic activity in the strains. Furthermore, the proteomic data showed that in each of the three strain hits occurred for GH hemicellulases belonging to arabinofuranosidase families of GH62 and GH51 for C34, as well as GH43 for both MBT40 and MBT70. Moreover, xylanase families including GH10 and GH11 were identified for MBT40 and MBT70 respectively (Table 2).

In addition to GH cellulases and hemicellulases, other enzyme classes were also detected, such as carboxyl esterase, polysaccharides lyases, and auxiliary enzyme which were also reported to promote lignocellulose degradation (Nevalainen et al., 2023). Notably, several carboxyl esterase were identified in the proteome of MBT70 including five acetyl xylan esterases from carboxyl esterase family CE1 & CE3, three polysaccharide deacetylases from family CE4 as well as a rhamnogalacturonan acetylesterase from family CE12 (Table 2C). Esterase family CE1 and CE4 proteins were also detected in C34, but not in MBT40. Acetyl xylan esterase was reported to facilitate degradation of xylan by removing 2- or 3-*O*-acetyl group of xylan (Biely et al., 2003), and rhamnogalacturonan acetylesterase plays a role in the deacetylation of the polysaccharide backbone of lignocellulose, making this enzyme an essential prerequisite for the subsequent action of GHs (Mølgaard et al., 2000). Moreover, the feruloyl esterase detected for C34 could play a role to cleave ester linkage between ferulic acid and sugars in order to facilitate cellulose degradation (Duan et al., 2021). Laccases, as detected in strain C34 are multicopper oxidases which oxidize a variety of phenolic substrates including phenolic lignin (Janusz et al., 2020).

In relation to the auxiliary domains associated with the proteins detected in the proteomes, the CBM2 and CBM13 were found to be attached to the GH cellulase and/or hemicellulases. It was reported that CBM2 was essentially found to bind the surfaces of crystalline polysaccharides such as polycrystalline cellulose (Boraston et al., 2004; Gilbert et al., 2013), insoluble xylan (Wu et al., 2022) and crystalline chitin (Nakamura et al., 2008). In contrast to CBM2, the CBM13 was specifically found to bind the xylan-polysaccharide chain and cannot bind to microcrystalline cellulose (Boraston et al., 2000; J. Liu et al., 2023). The occurrences of these CBMs, especially CBM2 may suggest more efficient degradation of polysaccharides in C34 and MBT70.

In dissecting the proteomic landscape of three top-performing *Streptomyces* strains during rice straw degradation, our results unveiled distinct enzyme cocktails (Table 3). Notably, strain C34 exhibited a spectrum of GH-(hemi)cellulases, encompassing

GH5(CBM2), GH11, GH35, GH42, GH48(CBM2), GH51(CBM2), GH62(CBM13), GH154, alongside the carboxyl esterases such as CE1, CE4, and the presence of laccase (AA1). In contrast, MBT40 displayed distinctive GH-(hemi)cellulase repertoire, featuring GH2, GH5, GH10(CBM13), GH43, GH74, GH92(CBM13), although only GH5 shared with C34. However, GH5 presented with the accessory domain CBM2 in C34. Noteworthy, MBT40 especially detected Polysaccharide Lyase, PL9 (Table 3). On the other hand, Strain MBT70 showcased a wider array of GH-(hemi)cellulases, encompassing GH4, GH6(CBM2), GH10, GH11, GH31, GH35, GH43(CBM13), GH48(CBM2), GH62, GH74(CBM2), complemented by variation of carboxyl esterase families, including CE1, CE3, CE4, CE12. Interestingly, MBT70 presented two types of CBM domains, including CBM2, and CBM13 (Table 3).

This significant variation in enzyme composition among the strains emphasizes the competency of different enzyme compositions in lignocellulose rice straw degradation. Moreover, this underscore the functional diversity encoded within *Streptomyces* genomes toward bioconversion of complex lignocellulosic rice straw (Satlewal et al., 2018). This diversity holds immense potential for tailoring enzymatic cocktails to specific biotechnological applications, such as biofuel production and biochemical synthesis from agricultural residue like rice straw.

Table 3. Overview of the polysaccharide degradation capacity of *Streptomyces* strains.

	Genome mining		Polysaccharide degradation capacity			Proteomic analysis	
Strains	Total different domain organizations of GHs	GH domain organization containing CBM	Solid medium	Liquid cultures	In vitro analysis	CAZymes predicted to participate in rice straw degradation	Predicted Activity
C34	**45**	**25 (56%)**	++++	+++	+++++	GH5(CBM2), GH11, GH35, GH42, GH48(CBM2), GH51(CBM2), GH62(CBM13), GH154.	Cellulases/ Hemicellulases
						CE1, CE4, AA1	Feruloyl esterase, laccase
MBT21	31	14 (45%)	NA	+	++	ND	
MBT28	37	16 (43%)	NA	+	++	ND	
MBT39	55	24 (44%)	+	+	++	ND	
MBT40	**53**	**24 (46%)**	+++	+++++	+++	GH2, GH5, GH10(CBM13), GH43, GH74, GH92(CBM13).	Cellulase/ Hemicellulases
						PL9	Pectate Lyase
MBT63	15	6 (40%)	NA	+	+	ND	
MBT66	26	9 (35%)	+	+	+	ND	
MBT70	**58**	**25 (43%)**	++	+++	+++++	GH4, GH6(CBM2), GH10, GH11, GH31, GH35, GH43(CBM13), GH48(CBM2), GH62, GH74(CBM2)	Cellulases/ Hemicellulase
						CE1, CE3, CE4, CE12	Acetyl xylan esterase, Rhamnogalacturonan acetylesterase

3.4 Conclusion

This study systematically analyzed the eight *Streptomyces* genomes, revealing a large number of genes encoding CAZymes from different families and with different domain organizations. In enzymatic activity screens, among these eight strains, C34, MBT40 and MBT70, came out as best performers when grown on rice straw as carbon source. To gain deeper insights for the enzymatic activity in rice straw fermentation, in vitro enzyme activity analyses and proteome investigations were conducted for these top-performing strains. Combining these experimental approaches, demonstrated that *Streptomyces* strains have both a highly variable gene repertoire for CAZymes and produce different cellulolytic enzyme cocktails using rice straw as substrate. The majority of detected enzymes in the proteome of these top-three strains were GH-(hemi)cellulases, showcasing significant variations in GH family and domain organizations among the strains. Understanding the specific enzyme cocktails produced by *Streptomyces* strains toward rice straw degradation opens avenues for fine-tuning bioprocesses, emphasizing the significance of tailored approaches for optimizing bioprocesses such as biofuel production and other value-added products from lignocellulose biomass.

Appendix A. Supplementary data

Supplementary data to this chapter can be found online at this following link.
https://doi. org/10.1016/j.biteb.2024.101775.

CRediT authorship contribution statement

Andika Sidar: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Gerben P. Voshol:** Software, Resources, Methodology. **Ahmed El-Masoudi:** Investigation, Formal analysis. **Erik Vijgenboom:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization. **Peter J. Punt:** Writing – review & editing, Validation, Supervision, Resources, Conceptualization.

Acknowledgement

We are gratefully thankful to Prof. Gilles van Wezel for providing the in-house collection of *Streptomyces* strains from Molecular BioTechnology (MBT) Institute of Biology Leiden, as well as to Dr. V.J. Carrion Bravo for transferring the in-house genomic data of these MBT *Streptomyces* strains used in this study. Moreover, very thank you to Dr. Joost Willemse and Gerda E.M. Lamers for the kindly microscopic guidance and assistance. Also, the sincerely thank to Prof. Arthur F.J. Ram for the insightful discussion.

Funding

This work was funded by a scholarship of the Indonesia Endowment Fund for Education (LPDP) from the Ministry of Finance, Indonesia (20160422026103) to AS, and by a grant of the Dutch National Organization for Scientific Research NWO, in the framework of an ERA-IB project FilaZyme (053.80.721/EIB.14.021) to EV, GPV, and PJP.