3 Biotech (2020) 10:160
https://doi.org/10.1007/s13205-020-2148-z

GENOME REPORTS

Genome analysis of cellulose and hemicellulose degrading *Micromonospora* sp. CP22

Sye Jinn Chen[1] · Ming Quan Lam[1] · Suganthi Thevarajoo[1] · Fazilah Abd Manan[1] · Adibah Yahya[1] · Chun Shiong Chong[1]

Received: 2 December 2019 / Accepted: 16 February 2020 / Published online: 5 March 2020

Abstract

In this study, a bacterial strain CP22 with ability to produce cellulase, xylanase and mannanase was isolated from the oil palm compost. Based on the 16S rRNA gene analysis, the strain was affiliated to genus *Micromonospora*. To further investigate genes that are related to cellulose and hemicellulose degradation, the genome of strain CP22 was sequenced, annotated and analyzed. The de novo assembled genome of strain CP22 featured a size of 5,856,203 bp with G + C content of 70.84%. Detailed genome analysis on lignocellulose degradation revealed a total of 60 genes consisting of 47 glycoside hydrolase domains and 16 carbohydrate esterase domains predicted to be involved in cellulolytic and hemicellulolytic deconstruction. Particularly, 20 genes encode for cellulases (8 endoglucanases, 3 exoglucanases and 9 β-glucosidases) and 40 genes encode for hemicellulases (15 endo-1,4-β-xylanase, 3 β-xylosidase, 3 α-arabinofuranosidase, 10 acetyl xylan esterase, 6 polysaccharide deacetylase, 1 β-mannanase, 1 β-mannosidase and 1 α-galactosidase). Thirty-two genes encoding carbohydrate-binding modules (CBM) from six different families (CBM2, CBM4, CBM6, CBM9, CBM13 and CBM22) were present in the genome of strain CP22. These CBMs were found in 27 cellulolytic and hemicellulolytic genes, indicating their potential role in enhancing the substrate-binding capability of the enzymes. CBM2 and CBM13 are the major CBMs present in cellulases and hemicellulases (xylanases and mannanases), respectively. Moreover, a GH10 xylanase was found to contain 3 CBMs (1 CBM9 and 2 CBM22) and these CBMs were reported to bind specifically to xylan. This genome-based analysis could facilitate the exploration of this strain for lignocellulosic biomass degradation.

Keywords Genome · Glycosyl hydrolase · Lignocellulose · *Micromonospora*

Introduction

The genus *Micromonospora* is an aerobic and Gram-positive actinobacteria that belongs to the Micromonosporaceae family (Hirsch and Valdés 2010). The members of this genus are widely distributed in diverse habitats such as soil, compost, freshwater, root nodules and sediment from lakes or deep sea (Parte 2018). The *Micromonospora* spp. have long been recognized as prolific producers of bioactive compounds (Boumehira et al. 2016). Throughout the years, the genome analyses of members from this genus had been extensively carried out to explore their potential biotechnology applications. These studies were mainly focused on genes that are related to the production of antimicrobial and antitumor compounds. Genome of *Micromonospora* sp. HK10 was sequenced and revealed the presence of polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS) and siderophore gene clusters that are potential to be applied in biomedicine production (Talukdar et al. 2016). Moreover, *Micromonospora yangpuensis* DSM 45577 ^{T} was found to possess the YPM A gene, which encodes for enediynes, a potent natural product against cancer cell lines (Yan et al. 2017). *Micromonospora* sp. DSW705 was also reported to

Genome sequence accession number: The whole-genome shotgun project of strain CP22 has been deposited at DDBJ/ENA/GenBank under the accession number VCJO00000000. The 16S rRNA gene of the strain CP22 is available at DDBJ/ENA/GenBank under accession MK934350.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-020-2148-z) contains supplementary material, which is available to authorized users.

✉ Chun Shiong Chong
cschong@utm.my

Extended author information available on the last page of the article

produce antitumor cyclic depsipeptides rakicidins A and B based on its genome analyses (Komaki et al. 2016).

The ability of *Micromonospora* spp. to depolymerize the lignocellulosic biomass is relatively less studied compared to other genera from actinobacteria, especially from the genomic aspect. The genome of the other actinobacterial strains such as *Streptomyces* have been sequenced extensively and reported to possess lignocellulolytic genes such as cellulases (GH5, GH6, GH9, GH12, GH44 and G48), xylanases (GH10 and GH11), mannanase (GH26, GH38 and GH76) and ligninase (AA10) (Feng et al. 2015; Pinheiro et al. 2017). Additionally, strains from the genus *Thermobifida* such as *Thermobifida fusca* and *Thermobifida cellulosilytica* were well studied at the genomic level (Zhao et al. 2015; Tóth et al. 2017). The recombinant *xynA* from *T. cellulosilytica* was confirmed with its competent ability to degrade xylan (Zhao et al. 2015). *Micromonospora* spp. were also found to have potential in lignocellulose degradation. Several strains from this genus were reported to exhibit cellulase and β-glucosidase activities previously (Gallagher et al. 1996; de Menezes et al. 2008). However, limited studies on the genes related to lignocellulose degradation for genus *Micromonospora* have been reported up to date.

Here, we present the genome analysis of *Micromonospora* sp. strain CP22, which was isolated from the oil palm empty fruit bunch (EFB) compost. The strain CP22 was grown on R2A agar (HiMedia) supplemented with 1% (w/v) of carboxymethyl cellulose (CMC) (Merck), beechwood xylan (Apollo Scientific) or locust bean gum from# *Ceratonia siliqua* seeds (Sigma) for screening of cellulolytic, xylanolytic and mannanolytic activities, respectively. Clear zones were observed in all agar plates, indicating this strain could secrete extracellular cellulase, xylanase and mannanase (Teather and Wood 1982).

To study the bacterium, the bacterial genomic DNA was extracted by Quick DNA Microprep Plus kit (Zymo Research) and purified by DNA Clean and Concentrator™-25 (Zymo Research) by following the manufacturer's protocol. Nanodrop™ 1000 spectrophotometer and Qubit® 3.0 fluorometer were used to validate the DNA quality ($A_{260/280}$ ratio) and concentration, respectively. The 16S rRNA gene of strain CP22 was amplified by polymerase chain reaction (PCR) using the universal primers 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1525R (5′-AAGGAGGTGWTCCARCC-3′) (Lam et al. 2018; Lane 1991). The PCR product was purified using Wizard® SV Gel and PCR Clean-Up System (Promega) and the purified PCR product was sent to Apical Scientific Pte. Ltd., Seri Kembangan, Selangor, Malaysia for Sanger sequencing. The 16S rRNA gene sequence of strain CP22 was obtained and searched in EzBioCloud database (Yoon et al. 2017). Based on the 16S rRNA gene analysis, strain CP22 was identified as a member of the genus *Micromonospora*. The similarities between strain CP22 and other *Micromonospora* species were in the range of 97.61% to 99.44% (S. Table 1). The 16S rRNA gene sequence of strain CP22 showed its highest similarity of 99.44% to *Micromonospora sediminimaris* CGMCC 4.3550 ^{T}, an isolate from deep sea sediment (Dai et al. 2010).

Furthermore, the quantitative enzyme assays of endoglucanase, exoglucanase, β-glucosidase, xylanase and mannanase were performed (Miller 1959; Wood and Bhat 1988). An equal volume of crude enzyme solution from a 2-day-old culture was added into 50 mM sodium phosphate buffer (pH 7.0) amended with 0.5 mL of 1% (w/v) CMC, Avicel® (Sigma), beechwood xylan, locust bean gum or 0.25 mL of 10 mM of p-nitrophenyl-beta-D-glucopyranoside (pNPG) (Merck) as the substrate, respectively. The mixture was then incubated at 30 °C for 60 min. The absorbance was detected by using UV-spectrophotometer (Thermo Fisher Scientific) at 540 nm (for reducing sugar assay using DNS) or 430 nm (for detecting p-nitrophenol (pNP) released from pNPG). One unit of enzyme activity (U/mL) was defined as the amount of enzyme that released 1 µmol of the respective product per min under assay conditions. The results showed that the enzyme activities were detected in all the assays conducted (Fig. 1), including all necessary cellulases (endoglucanase 0.08 U/mL; exoglucanase 0.056 U/mL; β-glucosidase 0.042 U/mL), xylanases (0.397 U/mL) and mannanases (0.403 U/mL). A similar finding was also reported from the enzyme assays conducted in *Micromonospora chalcea* NCIMB 12879 ^{T}. The crude enzyme activities of strain NCIMB 12879 ^{T} were present (endoglucanase 0.013 U/mL; β-glucosidase 0.004 U/mL) after culturing in a cellulose-containing media for 2 and 4 days, respectively (Gallagher et al. 1996).

To further investigate the genes related to cellulose and hemicellulose degradation, the genome of strain CP22 was

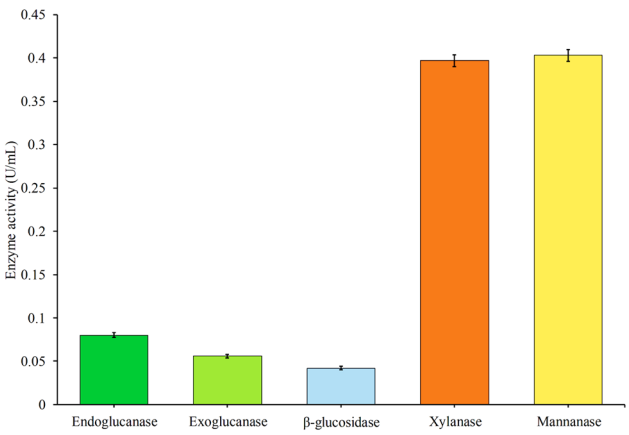


Fig. 1 Cellulolytic, xylanolytic and mannanolytic enzyme activities in cell-free supernatant of strain CP22. The experiment was performed in triplicate and standard deviations are indicated as error bar

sequenced and analyzed. In addition, the genome data were also used to validate the taxonomic affiliation of strain CP22. The whole genome sequencing was performed on an Illumina HiSeq 2500 sequencing platform (Illumina, California, USA) with 2 × 150-bp paired-end reads. After quality trimming and adapter removal, the sequences were assembled de novo using SPAdes software version 3.11.1 (Bankevich et al. 2012). The final de novo genome of *Micromonospora* sp. CP22 (NCBI accession: VCJO00000000.1) was assembled in 57 contigs with size of 5,856,203 bp (Table 1). The N_{50} is 297,587 bp and the longest contig assembled contains 764,068 bp (Table 1). The genome size of strain CP22 is smaller than its closely related strain: *Micromonospora maris* AB-18-032 ^{T} (6,407,450 bp) and *Micromonospora sediminimaris* CGMCC 4.3550 ^{T} (6,673,976 bp) (S. Figure 1). The G + C content of strain CP22 is 70.8%, which is similar to *Micromonospora maris* AB-18-032 ^{T} (70.9%) and *Micromonospora sediminimaris* CGMCC 4.3550 ^{T} (71.0%) (S. Fig. 1).

The assembled genome was analyzed using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.8 (Tatusova et al. 2016). Prodigal and GeneMarkS 2 + with best-placed reference protein set were used as the annotation method for gene identification. For tRNA and rRNA prediction, tRNAS-can-SE-1.23 (Lowe and Chan 2016) and RNAmmer server (Lagesen et al. 2007) were used, respectively. Based on the NCBI PGAP, the genome contains a total of 5034 protein coding genes, 3 rrn operons and 53 rRNA genes. A total of 3751 genes were annotated into 25 COG functional groups by using Clusters of Orthologous Groups (COGs) functional assignment using WebMGA server with RPS-BLAST version 2.2.15 based on NCBI COG database (Wu et al. 2011). The major COG categories of strain CP22 were general function prediction (R-13.81%), transcription (K-11.41%), amino acid transport and metabolism (E- 9.54%), carbohydrate transport and metabolism (G-9.03%) and liquid transport and metabolism (I-7.25%). Specifically, there were 339 genes (9.03%) encoded in the genome of strain CP22, which were categorized under the functional group of carbohydrate transport and metabolism (G), in which part of them were related to lignocellulose degradation (Fig. 2).

Table 1 Genome features and genome sequencing information of *Micromonospora* sp. CP22 (NCBI accession: VCJO00000000.1)

Attributes	Value
Genome feature	
Genome size (bp)	5,856,203
DNA coding region (bp)	5,266,789
DNA G + C content (%)	70.84
Number of contigs	57
Total genes	5272
Protein coding genes	5034
rRNA genes	3
tRNA genes	53
Genes assigned to COG	3751
MIGS 4.0 data	
Submitted_to_insde	GenBank
Investigation_type	Bacteria_archea
Project_name	*Micromonospora* sp. CP22 genome sequencing
Geo_loc_name	Johor, Malaysia
Collection_date	March 2018
Env_biome	Agricultural
Env_feature	Compost
Env_material	Compost
Biotic_relationship	Free living
Trophic_level	Chemoheterotroph
Rel_to_oxygen	Aerobic
Genome assembly data	
Seq_method	Illumina HiSeq 2500
Assembly	SPAdes version 3.11.1
Finishing_strategy	High-quality draft

To further validate the taxonomy affiliation of strain CP22, the 16S rRNA gene phylogenetic and phylogenomic trees between strain CP22 and other closely related species of *Micromonospora* were reconstructed. For 16S rRNA phylogenetic analysis, the full length 16S rRNA sequences of strain CP22 (NCBI accession: MK934350) retrieved from the genome sequence of strain CP22 were aligned using ClustalW with other related members (Thompson et al. 1994). For the phylogenomic analysis, REALPHY version 1.12 online server was used to align the genomic sequences of strain CP22 and other *Micromonospora* members (Bertels et al. 2014). Both phylogenetic and phylogenomic tree were reconstructed using three methods (neighbor joining, maximum likelihood and maximum parsimony) based on 1000 bootstrap replicates in MEGA X software (Huelsenbeck and Crandall 1997; Kumar et al. 2018; Saitou and Nei 1987; Swofford 1998). Distances were calculated using Kimura 2-parameter and the sequence of *Streptomyces coelicolor* A3(2) was used as the outgroup. In Fig. 3, strain CP22 is clustered within the clade comprising the *Micromonospora* species, indicating that strain CP22 belongs to the genus *Micromonospora*. In addition, based on both phylogenetic (Fig. 3a) and phylogenomic trees (Fig. 3b), strain CP22 showed its closest relationship with *Micromonospora sediminimaris* CGMCC 4.3550 ^{T}.

For lignocellulose degrading genes mining, the putative genes encoded for CAZymes from strain CP22 were identified using dbCAN 2 meta server (Zhang et al. 2018). This was followed by a further screening on the identified CAZymes based on the principle that the CAZyme domains were recognized by at least two of the three tools (DIAMONDS, HMMER and Hotpep) (Lam et al. 2020).

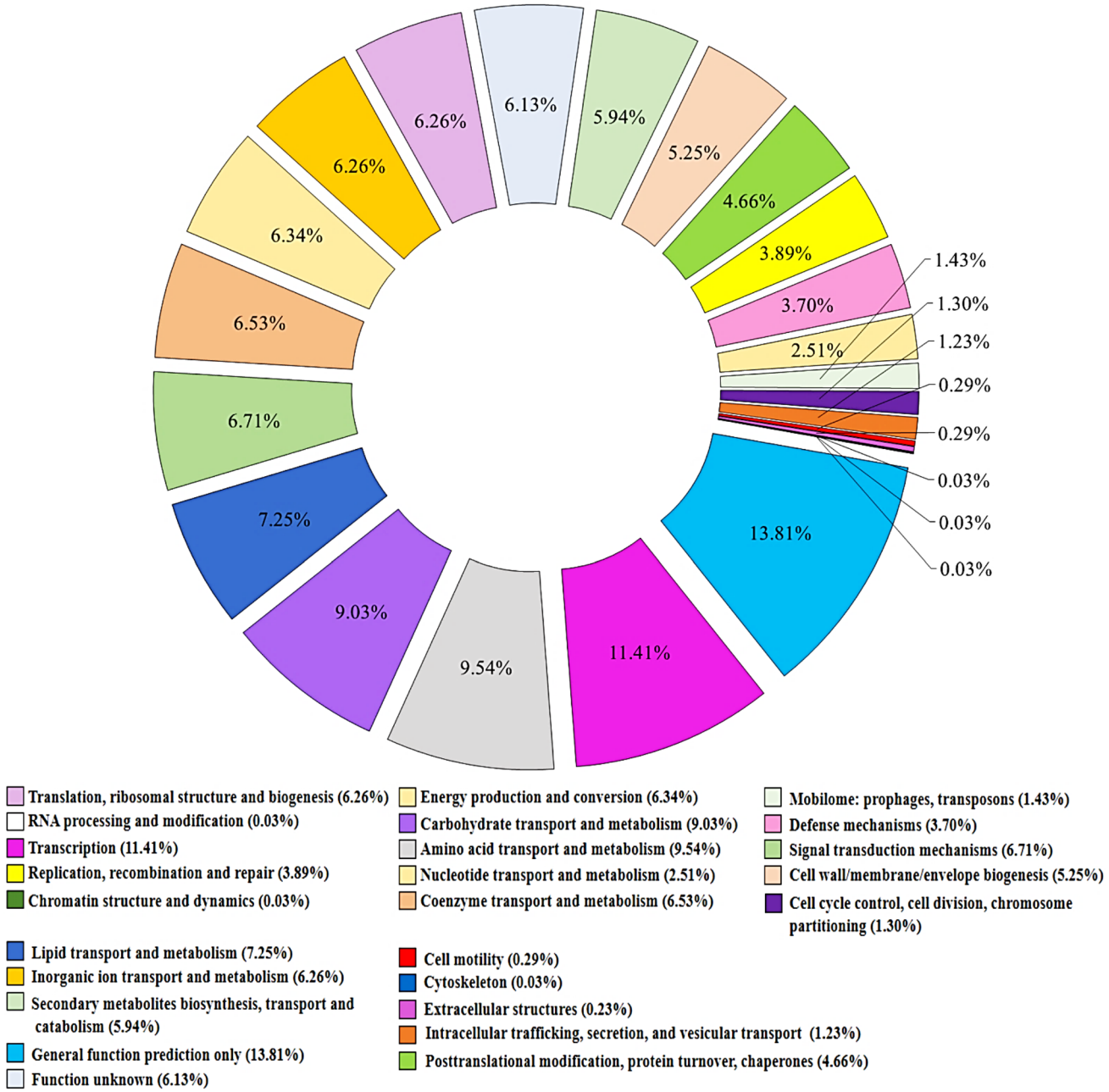
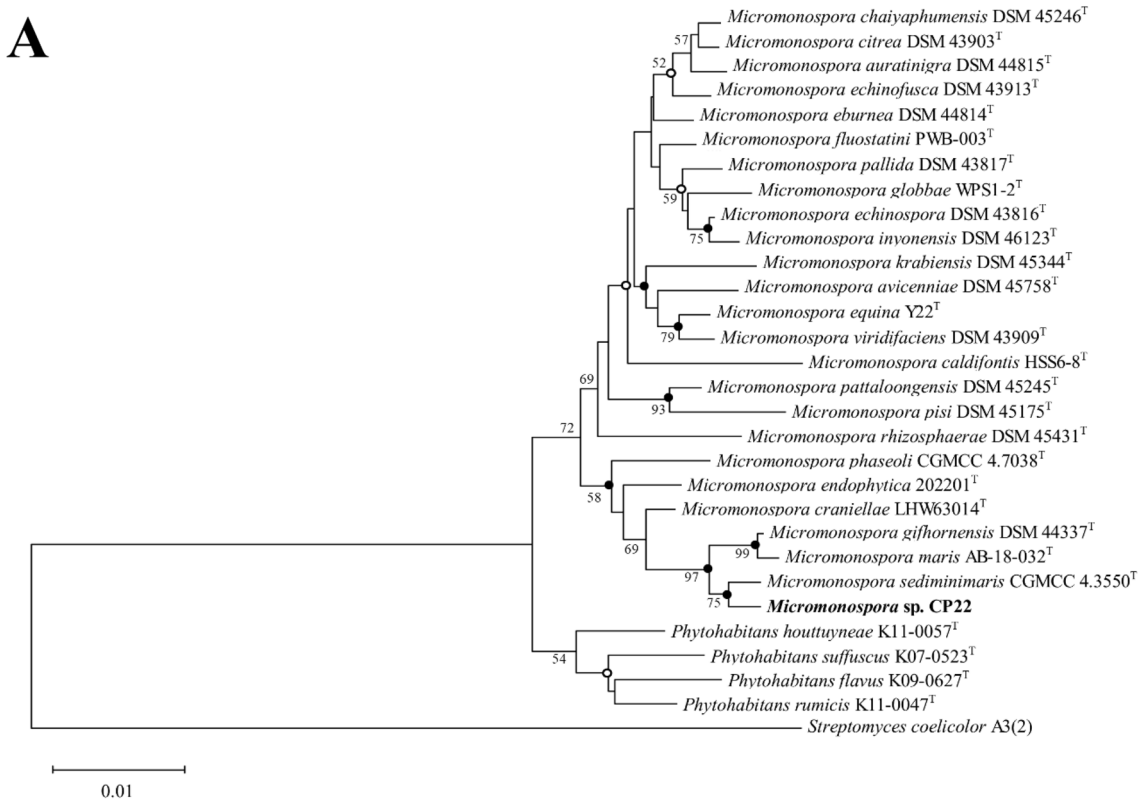


Fig. 2 COG functional categories of *Micromonospora* sp. CP22

The CAZymes were further screened to identify the genes responsible for cellulose and hemicellulose degradation (Lombard et al. 2013). In the genome of strain CP22, a total of 60 genes consisting of 47 glycoside hydrolase domains (GH1, GH2, GH3, GH5, GH6, GH8, GH9, GH10, GH11, GH12, GH27, GH30, GH39, GH43, GH48 and GH51), 16 carbohydrate esterase domains (CE1, CE3, CE4 and CE7) and 32 carbohydrate-binding proteins (CBM2, CBM4, CBM6, CBM9, CBM13 and CBM22) were predicted to be involved in cellulolytic and hemicellulolytic deconstruction

Fig. 3 **a** Neighbor-joining tree derived from 16S rRNA gene sequence data of *Micromonospora* sp. CP22 and its relatives. **b** The neighbor-joining phylogenomic tree derived from genome data of *Micromonospora* sp. CP22 and other strains which are publicly available online. *Streptomyces coelicolor* A3(2) was used as the outgroup. Filled circles indicate the corresponding nodes were recovered in the trees generated with all three methods (neighbor joining, maximum likelihood and maximum parsimony). Open circles indicate that the corresponding nodes were recovered in the trees generated with either two methods. Bar, 0.01 nucleotide substitutions per site


A
Micromonospora chaiyaphumensis DSM 45246T
Micromonospora citrea DSM 43903T
Micromonospora auratinigra DSM 44815T
Micromonospora echinofusca DSM 43913T
Micromonospora eburnea DSM 44814T
Micromonospora fluostatini PWB-003T
Micromonospora pallida DSM 43817T
Micromonospora globbae WPS1-2T
Micromonospora echinospora DSM 43816T
Micromonospora inyonensis DSM 46123T
Micromonospora krabiensis DSM 45344T
Micromonospora avicenniae DSM 45758T
Micromonospora equina Y22T
Micromonospora viridifaciens DSM 43909T
Micromonospora caldifontis HSS6-8T
Micromonospora pattaloongensis DSM 45245T
Micromonospora pisi DSM 45175T
Micromonospora rhizosphaerae DSM 45431T
Micromonospora phaseoli CGMCC 4.7038T
Micromonospora endophytica 202201T
Micromonospora craniellae LHW63014T
Micromonospora gifhornensis DSM 44337T
Micromonospora maris AB-18-032T
Micromonospora sediminimaris CGMCC 4.3550T
Micromonospora sp. CP22
Phytohabitans houttuyneae K11-0057T
Phytohabitans suffuscus K07-0523T
Phytohabitans flavus K09-0627T
Phytohabitans rumicis K11-0047T
Streptomyces coelicolor A3(2)
0.01

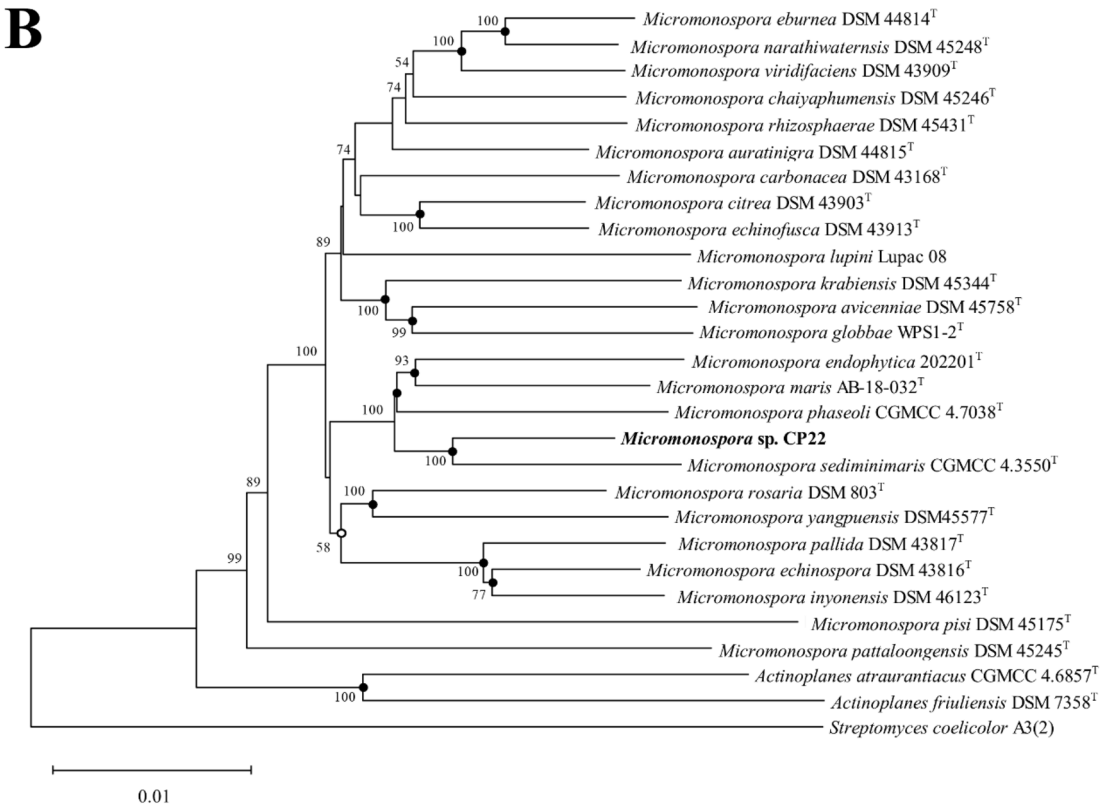

B
Micromonospora eburnea DSM 44814T
Micromonospora narathiwaternsis DSM 45248T
Micromonospora viridifaciens DSM 43909T
Micromonospora chaiyaphumensis DSM 45246T
Micromonospora rhizosphaerae DSM 45431T
Micromonospora auratinigra DSM 44815T
Micromonospora carbonacea DSM 43168T
Micromonospora citrea DSM 43903T
Micromonospora echinofusca DSM 43913T
Micromonospora lupini Lupac 08
Micromonospora krabiensis DSM 45344T
Micromonospora avicenniae DSM 45758T
Micromonospora globbae WPS1-2T
Micromonospora endophytica 202201T
Micromonospora maris AB-18-032T
Micromonospora phaseoli CGMCC 4.7038T
Micromonospora sp. CP22
Micromonospora sediminimaris CGMCC 4.3550T
Micromonospora rosaria DSM 803T
Micromonospora yangpuensis DSM45577T
Micromonospora pallida DSM 43817T
Micromonospora echinospora DSM 43816T
Micromonospora inyonensis DSM 46123T
Micromonospora pisi DSM 45175T
Micromonospora pattaloongensis DSM 45245T
Actinoplanes atraurantiacus CGMCC 4.6857T
Actinoplanes friuliensis DSM 7358T
Streptomyces coelicolor A3(2)
0.01

Table 2 List of potential cellulolytic and hemicellulolytic genes in the genome of strain CP22

Category	Predicted function	CAZyme families	Protein ID
Cellulase	Endoglucanase	GH5	FF096_08465
		GH5 + CBM2	FF096_17730
			FF096_23015
			FF096_24405
		GH6 + CBM2	FF096_13080
			FF096_03085
		GH9 + CBM2	FF096_06450
		GH12 + CBM2	FF096_22605
	Exoglucanase	GH6 + CBM2	FF096_21250
		GH9 + CBM2 + CBM4	FF096_13010
		GH48	FF096_14285
	β-Glucosidase	GH1	FF096_06480
			FF096_01150
			FF096_14290
			FF096_19310
		GH3	FF096_18490
			FF096_19365
			FF096_09040
			FF096_08385
			FF096_13220
Xylanase	Endo-1,4-β-xylanase	GH5 + GH43	FF096_09840
		GH5 + GH43 + CBM6	FF096_09835
		CBM6 + GH8 + CBM6	FF096_26150
		GH10	FF096_22465
			FF096_06350
			FF096_10160
			FF096_13595
		GH10 + CBM2	FF096_25100
		GH10 + CBM13	FF096_08470
		GH10 + CBM22 + CBM22 + CBM9	FF096_21655
		GH11	FF096_18125
		GH11 + CBM13	FF096_02220
		GH30 + CBM2	FF096_16355
		GH43	FF096_10275
			FF096_18160
	β-Xylosidase	GH39	FF096_21465
		GH43	FF096_06345
			FF096_05325
	α-Arabinofuranosidase	GH43 + CBM 6 + CBM 22	FF096_10270
		GH51	FF096_09215
		GH51 + CBM2	FF096_09800
	Acetyl xylan esterase	CE1	FF096_24510
			FF096_22890
			FF096_23180

Table 2 (continued)

Category	Predicted function	CAZyme families	Protein ID
		CE1 + CBM2	FF096_24445
			FF096_18200
			FF096_20900
			FF096_10225
		CE1 + CBM13	FF096_25110
		CE3 + CBM13	FF096_09300
		CE7	FF096_21435
	Polysaccharide deacetylase	CE4	FF096_06355
			FF096_06035
			FF096_11395
			FF096_00990
		CE4 + CBM13	FF096_06125
		CE4 + GT2	FF096_03080
Mannanase	β-Mannanase	GH5 + CBM2	FF096_06260
	β-Mannosidase	GH2	FF096_08210
	α-galactosidase	GH27 + CBM13	FF096_09830

(Table 2). Based on the annotation, all three hydrolytic cellulases including endoglucanase (GH5, GH6, GH9 and GH12), exoglucanase (GH6, GH9 and GH48) and β-glucosidase (GH1 and GH3) were found in strain CP22 genome, which were reported earlier to work cooperatively on cellulose degradation (Berlemont and Martiny 2013; Lombard et al. 2013). For hemicellulases, the related genes were annotated as endo-1,4-β-xylanase (GH5, GH8, GH10, GH11, GH30 and GH43), β-xylosidase (GH39 and GH43), α-arabinofuranosidase (GH43 and GH51), acetyl xylan esterase (CE1, CE3 and CE7) and polysaccharide deactylase (CE4), respectively (Shallom and Shoham 2003; Lombard et al. 2013). Moreover, a total of three genes (β-mannanase (GH5), β-mannosidase (GH2) and α-galactosidase (GH27)) were predicted for mannan (second major component in hemicellulose) degradation (Malgas et al. 2015; Lombard et al. 2013). The presence of cellulose and hemicellulose degrading genes in the genome of strain CP22 synchronized well with the crude enzyme analyses, wherein the endoglucanase, exoglucanase, β-glucosidase, xylanases and mannanases were expressed and active in degrading CMC, avicel, pNPG, xylan and locust bean gum, respectively (Fig. 1).

Besides that, 27 annotated genes related to cellulose and hemicellulose deconstruction were found to be coupled with carbohydrate-binding modules (CBMs) (Table 2). The family 2 CBM (CBM2) is the major CBMs found in the genome of strain CP22 (Table 2). CBM2 was reported to have the ability to bind cellulose or xylan (Lombard et al. 2013). A total of seven out of the eight annotated endoglucanases from strain CP22 were found to have CBM2 domains (Table 2). This CBM could facilitate the enzymatic reaction for cellulose degradation. Other than CBM2, there are five CBM13 present in the genes encode for xylananses, while one CBM13 is found in a mannanase (Table 2). CBM13 is found to have a variety of sugar-binding specificities and distributed in many types of CAZymes including xylanases and α-galactosidases (Fujimoto 2013). Researchers have reported that the presence of CBM13 in the enzyme could be advantageous. As CBM13 may contain more than one sugar-binding site with different sugar-binding preferences (type α, β or γ), this could permit the enzyme to bind different xylan chains in a complex xylan cluster at the same time (Fujimoto et al. 2002). On the other hand, there is one endo-1,4-β-xylanase (GH10) that contains multiple CBMs with two CBM22 and one CBM9 (Table 2). Both types of CBM were previously reported to have the ability to bind specifically to xylan (Lombard et al. 2013).

To investigate the unique genomic properties of strain CP22, particularly in the aspect of cellulose and hemicellulose degradation, the related genes of strain CP22 were compared to its closest related strain, *Micromonospora sediminimaris* CGMCC 4.3550 ^{T} (Fig. 4). A total of 63 cellulolytic or hemicellulolytic CAZYme domains were found in the genome of strain CP22, while there are 61 in the genome of *Micromonospora sediminimaris* CGMCC 4.3550 ^{T}. Strain CP22 contains a higher number of CE1 and CE4 genes compared to *Micromonospora sediminimaris* CGMCC 4.3550 ^{T} (Strain CP22: 8 CE1s, 6 CE4s-; *M. sediminimaris* CGMCC 4.3550 ^{T}: 7 CE1s, 3 CE4s). CE1 and CE4 are annotated to acetyl xylan esterase and polysaccharide deacetylase,

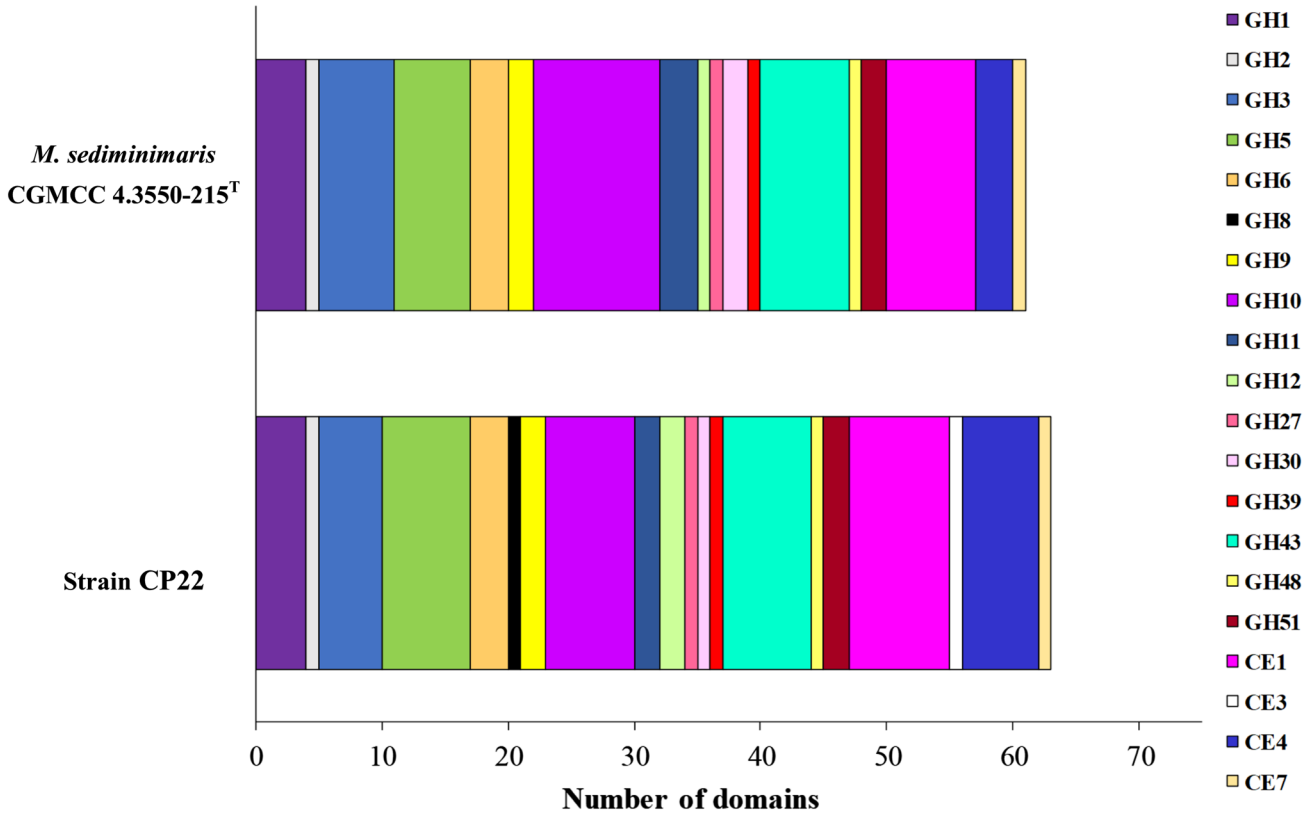


Fig. 4 The number of CAZyme domains containing enzymes involved in the cellulose and hemicellulose deconstruction in the genome of strain CP22 and *Micromonospora sediminimaris* CGMCC 4.3550 T

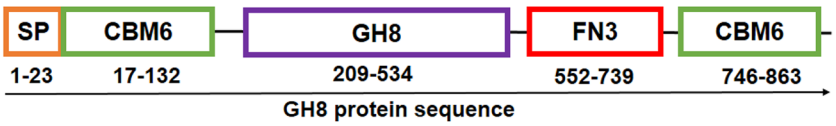


Fig. 5 The domain organization of GH8 protein sequence (Protein ID: FF096_26150) from the genome of strain CP22. From left to right: *SP* signal peptide; *CBM6* carbohydrate-binding module family 6; *GH8* glycoside hydrolase family 8; *FN3* fibronectin type 3

respectively, in which they are the key enzymes attributed to deacetylation of hemicellulose. The higher number of CE1 and CE4 genes found in strain CP22 might improve the rate of deacetylation and thus increase the biomass accessibility for enzymatic action of GHs (Biely 2012). On the other hand, the number of GH10 enzymes in *Micromonospora sediminimaris* CGMCC 4.3550 T was significantly higher than that in strain CP22 (strain CP22: 7, *M. sediminimaris* CGMCC 4.3550 T-10). GH10 enzymes mainly take part in hemicellulose degradation as endo-1,4-β-xylanase or endo-1,3-β-xylanase (Lombard et al. 2013). The strain CP22 possesses additional domains from GH8 and CE3 families, which were not found in the genome of *Micromonospora sediminimaris* CGMCC 4.3550 T (Fig. 4).

The GH8 protein sequence of strain CP22 (Protein ID: FF096_26150) was further investigated. Interestingly, no protein sequences from other *Micromonospora* species were found to be similar to GH8 protein of strain CP22 based on NCBI non-redundant database. The GH8 protein sequence of strain CP22 shared a relatively low similarity (57.46–77.61%) to the top ten most similar protein sequences of strains from different families or phyla including *Catellatospora*, *Kitasatospora*, *Demequima*, *Streptomyces*, *Deinococcus*, *Chloroflexi* and *Sphingomonadaceae*. In addition, the protein sequence of GH8 was analyzed by using InterProScan server 77.0 (Mitchell et al. 2018). The result revealed that the protein sequence consists of a signal peptide, a GH8 domain, a fibronectin type 3 (FN3) domain and

two CBM6 domains (Fig. 5). This indicates that the GH8 protein of strain CP22 could be an extracellular enzyme. FN3 domains was reported to have the ability to promote enzyme activity and stability in the hydrolysis of cellulosic or xylanosic substrate (Han et al. 2019; Kataeva et al. 2002). The GH8 in this study was annotated as xylanase with low similarity to other bacterial GH8, indicating the enzyme could be further studied for its functionality (Juturu and Wu 2014).

In conclusion, both genomic and experimental evidences demonstrated that *Micromonospora* sp. CP22 has the ability to degrade cellulose and hemicellulose. The strain CP22 possesses not only all necessary cellulases, but also a decent amount of hemicellulases that are active against xylan and mannan. The strain CP22 could become a potential industrial strain, which can be applied in processes such as pre-treatment and saccharification of the lignocellulose biomass. Functional characterization of the recombinant enzymes and the optimization of enzyme production are required in future to further study the potential of this cellulolytic and hemicellulolytic strain.

Acknowledgements This work was financially sponsored by Universiti Teknologi Malaysia under Research University Grant (GUP 2018 Tier 1) with project number 20H43. The authors would like to acknowledge palm oil mill, Johor for the permission for sample collection. Sye Jinn Chen acknowledges UTM Zamalah Scholarship from Universiti Teknologi Malaysia. Ming Quan Lam acknowledges Khazanah Watan Postgraduate (PhD) scholarship (scholar ID: 40852) from Yayasan Khazanah.

Author contribution statement SJC: carried out the experiment, analyzed the data and drafted the manuscript. MQL: designed the experiment, assisted in bioinformatics analyses and edited the manuscript. CSC: conceived of the presented idea, designed the experiment and provided expertise. AY, FAM and ST: provided expertise and edited the manuscript. All authors read, edited and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477

Berlemont R, Martiny AC (2013) Phylogenetic distribution of potential cellulases in bacteria. Appl Environ Microbiol 79(5):1545–1554

Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E (2014) Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol Biol Evol 31(5):1077–1088

Biely P (2012) Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnol Adv 30(6):1575–1588

Boumehira AZ, El-Enshasy HA, Hacene H, Elsayed EA, Aziz R, Park EY (2016) Recent progress on the development of antibiotics from the genus *Micromonospora*. Biotechnol Bioproc E 21(2):199–223

Dai HQ, Wang J, Xin YH, Pei G, Tang SK, Ren B, Ward A, Ruan JS, Li WJ, Zhang LX (2010) *Verrucosispora sediminis* sp. Nov., a cyclodipeptide producing actinomycete from deep-sea sediment. Int J Syst Evol Microbiol 60(8):1807–1812

de Menezes AB, Lockhart RJ, Cox MJ, Allison HE, McCarthy AJ (2008) Cellulose degradation by Micromonosporas recovered from freshwater lakes and classification of these actinomycetes by DNA gyrase B gene sequencing. Appl Environ Microbiol 74(22):7080–7084

Feng H, Sun Y, Zhi Y, Mao L, Luo Y, Wei X, Zhou P (2015) Lignocellulose degradation by the isolate of *Streptomyces griseorubens* JSD-1. Funct Integr Genomics 15(2):163–173

Fujimoto Z (2013) Structure and function of carbohydrate-binding module families 13 and 42 of glycoside hydrolases, comprising a β-trefoil fold. Biosci Biotechnol Biochem 77(7):1363–1371

Fujimoto Z, Kuno A, Kaneko S, Kobayashi H, Kusakabe I, Mizuno H (2002) Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. J Mol Biol 316(1):65–78

Gallagher J, Winters A, Barron N, McHale L, McHale AP (1996) Production of cellulase and β-glucosidase activity during growth of the actinomycete *Micromonospora chalcae* on cellulose-containing media. Biotechnol Lett 18(5):537–540

Han Z, Yang J, Shang-guan F (2019) Molecular and biochemical characterization of a bimodular xylanase fro*m Marinifilaceae* bacterium strain SPP2. Front Microbiol 10:1507

Hirsch AM, Valdés M (2010) *Micromonospora*: an important microbe for biomedicine and potentially for biocontrol and biofuels. Soil Biol Biochem 42(4):536–542

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Evol Syst 28(1):437–466

Juturu V, Wu JC (2014) Microbial exo-xylanases: a mini review. Appl Biochem Biotech 174(1):81–92

Kataeva IA, Seidel RD, Shah A, West LT, Li XL, Ljungdahl LG (2002) The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. Appl Environ Microbiol 68(9):4292–4300

Komaki H, Ichikawa N, Hosoyama A, Hamada M, Harunari E, Ishikawa A, Igarashi Y (2016) Draft genome sequence of *Micromonospora* sp. DSW705 and distribution of biosynthetic gene clusters for depsipeptides bearing 4-amino-2, 4-pentadienoate in actinomycetes. Stand Genomic Sci 11(1):84–93

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549

Lam MQ, Mut NN, Thevarajoo S, Chen SJ, Selvaratnam C, Hussin H, Jamaluddin H, Chong CS (2018) Characterization of detergent compatible protease from halophilic *Virgibacillus* sp. CD6. 3 Biotech 8(2):104

Lam MQ, Oates NC, Suganthi T, Tokiman L, Goh KM, MQ-Mason SJ, Bruce NC, Chong CS (2020) Genomic analysis of a lignocellulose degrading strain from the underexplored genus *Meridianimaribacter*. Genomics 112(1):952–960

Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in bacterial systematics. Wiley, Chichester, pp 125–175

Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35(9):3100–3108

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42(D1):D490–D495

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:54–57

Malgas S, van Dyk JS, Pletschke BI (2015) A review of the enzymatic hydrolysis of mannans and synergistic interactions between β-mannanase, β-mannosidase and α-galactosidase. World J Microb Biot 31(8):1167–1175

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31(3):426–428

Mitchell AL et al (2018) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res 47:D351–D360

Parte AC (2018) LPSN–List of Prokaryotic names with Standing in Nomenclature (bacterio. net), 20 years on. Int J Syst Evol Microbiol 68(6):1825–1829

Pinheiro GL, de Azevedo-Martins AC, Albano RM, de Souza W, Frases S (2017) Comprehensive analysis of the cellulolytic system reveals its potential for deconstruction of lignocellulosic biomass in a novel *Streptomyces* sp. Appl Microbiol Biotechnol 101(1):301–319

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Shallom D, Shoham Y (2003) Microbial hemicellulases. Curr Opin Microbiol 6(3):219–228

Swofford DL (1998) Phylogenetic analysis using parsimony. Illinois Natural History Survey, Champaign

Talukdar M, Das D, Bora C, Bora TC, Boruah HP, Singh AK (2016) Complete genome sequencing and comparative analyses of broad-spectrum antimicrobial-producing *Micromonospora* sp. HK10. Gene 594(1):97–107

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624

Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43(4):777–780

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680

Tóth Á, Baka E, Luzics S, Bata-Vidács I, Nagy I, Bálint B, Herczeg R, Olasz F, Wilk T, Nagy T, Kriszt B (2017) Plant polysaccharide degrading enzyme system of *Thermobifida cellulosilytica* TB100^T revealed by de novo genome project data. Acta Aliment 46(3):323–335

Wood TM, Bhat KM (1988) Methods for measuring cellulase activities. Methods Enzymol 160:87–112

Wu S, Zhu Z, Fu L, Niu B, Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics 12(1):444–452

Yan X, Chen JJ, Adhikari A, Yang D, Crnovcic I, Wang N, Chang CY, Rader C, Shen B (2017) Genome mining of *Micromonospora yangpuensis* DSM 45577 as a producer of an anthraquinone-fused enediyne. Org Lett 19(22):6192–6195

Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67(5):1613–1617

Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46(W1):W95–W101

Zhao L, Geng J, Guo Y, Liao X, Liu X, Wu R, Zheng Z, Zhang R (2015) Expression of the *Thermobifida fusca* xylanase Xyn11A in Pichia pastoris and its characterization. BMC Biotechnol 15(1):18–29

Affiliations

Sye Jinn Chen[1] · Ming Quan Lam[1] · Suganthi Thevarajoo[1] · Fazilah Abd Manan[1] · Adibah Yahya[1] · Chun Shiong Chong[1]

Sye Jinn Chen
syejinn0420@gmail.com

Ming Quan Lam
mqlam2@gmail.com

Suganthi Thevarajoo
suganthitheva@gmail.com

Fazilah Abd Manan
m-fazilah@utm.my

Adibah Yahya
adibah@fbb.utm.my

[1] Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia