Genome sequencing of strain *Cellulosimicrobium* sp. TH-20 with ginseng biotransformation ability

Fei Zheng^{1,4} · Wei Zhang² · Xiaodan Chu³ · Yulin Dai¹ · Jing Li¹ · Huanxi Zhao¹ · Liankui Wen⁴ · Hao Yue¹ · Shanshan Yu¹

Received: 25 April 2017 / Accepted: 6 June 2017 / Published online: 11 July 2017
© Springer-Verlag GmbH Germany 2017

Abstract Biotransformation for increasing the pharmaceutical effect of ginsenosides is getting more and more attractions. Strain *Cellulosimicrobium* sp. TH-20 isolated from ginseng soil samples was identified to produce enzymes contributing to its excellent biotransformation activity against ginsenosides, the main active components of ginseng. Based on phylogenetic tree and homology analysis, the strain can be designated as *Cellulosimicrobium* sp. Genome sequencing was performed using the Illumina Miseq to explore the functional genes involved in ginsenoside transformation. The draft genome of *Cellulosimicrobium* sp. TH-20 encoded 3450 open reading frames, 51 tRNA, and 9 rRNA. All ORFs were annotated using NCBI BLAST with non-redundant proteins, Gene Ontology, Cluster of Orthologous Gene, and Kyoto Encyclopedia of Genes and Genomes databases. A total of 11 genes were selected based on the functional annotation analysis. These genes are relevant to ginsenoside biotransformation, including 6 for beta-glucosidase, 1 for alpha-*N*-arabinofuranosidase, 1 for alpha-1,6-glucosidase, 1 for endo-1,4-beta-xylanase, 1 for alpha-L-

arabinofuranosidase, and 1 for beta-galactosidase. These glycosidases were predicted to catalyze the hydrolysis of sugar moieties attached to the aglycon of ginsenosides and led to the transformation of PPD-type and PPT-type ginsenosides.

Keywords *Cellulosimicrobium* sp. TH-20 · Genome sequencing · Biotransformation · Ginsenosides

Introduction

Ginsenosides are the main active components of ginseng with anti-fatigue, anti-allergenic, anti-oxidant, anti-inflammatory, and anti-cancer activities (Ru et al. 2015). More than 180 ginsenosides have been identified from ginseng. Ginsenosides can be divided into two groups consisting of protopanaxadiol (PPD) and protopanaxatriol (PPT) group. The PPD and PPT ginsenosides consist of a nonsugar component (aglycon) with a dammarane skeleton and a sugar component comprised of 1–4 molecules (glycon), including D-glucose, L-arabinopyranoside, L-arabinofuranoside, D-xylose, and/or L-rhamnose. Furthermore, differences in structures, including sugar types, numbers, and attachment positions, provide diversity in the pharmaceutical activities of ginseng. Recent research indicated that deglycosylated ginsenosides (ginsenosides Rh1, Rh2, F2, Rg3, C-K, etc.) are more pharmaceutically active than glycosylated ginsenosides (Rb1, Rb2, Rc, Rd, Rg1, etc.). Glycosidases have been considered to be useful tools to catalyze the hydrolysis of the sugar moieties of ginsenosides to produce deglycosylated ginsenosides. Hou et al. (2015) reported that Strain CNU 120806 represented a high degree of beta-glucosidase activity to convert ginsenosides Rb and Rd to gypenoside LXXV and Compound K, respectively (Hou et al. 2015). Hyun et al. (2012) observed that beta-D-xylosidase

✉ Hao Yue
yuehao001@aliyun.com

✉ Shanshan Yu
yushanshan001@aliyun.com

¹ Jilin Ginseng Academy, Changchun University of Chinese Medicine, Changchun 130117, China

² The Affiliated Hospital of Changchun University of Chinese Medicine, Changchun 130117, China

³ Life Science Research Center, Beihua University, Jilin 132011, China

⁴ College of Food Science and Engineering, Jilin Agricultural University, Changchun 130118, China

from *Bifidobacterium breve* K-110 can catalyze the hydrolysis of ginsenoside Ra1 to Rb2 (Hyun et al. 2012). Shin et al. (2013) reported the production of compound K from major ginsenosides by the combined use of alpha-L-arabinofuranosidase and beta-galactosidase from *Caldicellulosiruptor saccharolyticus* and beta-glucosidase from *Sulfolobus acidocaldarius* (Shin et al. 2013). A limited number of glycosidases available for ginsenoside biotransformation are identified. Thus, identifying new glycosidases for the production of ginsenosides is imperative.

The genus *Cellulosimicrobium* has attracted attention as rich sources of glycosidases, and *Cellulosimicrobium*-derived glycosidases are reported useful for hydrolyzing glycoside substances. Many glycosidases, including endo-beta-1,4-glucanase, endo-beta-1,4-xylanase, endo-beta-1,4-mannanase, and endo-beta-1,3-glucanase, have been identified and characterized from *Cellulosimicrobium*. *Cellulosimicrobium cellulans* strain E4-5 isolated from seaside soil was identified to hydrolyze (1 → 3)-linked beta-D-glucan (Fu et al. 2015). An endo-1,3-beta-glucanase was purified from *C. cellulans* DK-1 (Tanabe and Oda 2011). *C. cellulans* strain F16 that can remove the C-7 xylosyl group from 7-xylosyltaxanes was isolated from the root soil of an old *Taxus yunnanensis* tree (Hou et al. 2012). Furthermore, emerging evidence has demonstrated that microorganisms from *Cellulosimicrobium* are potential to be applied in the production of glycoside bioactive moleculars in the pharmaceutical industry. Yuan et al. (2015) reported that beta-glucosidase from strain *C. cellulans* sp. 21 represents transformation activity for producing ginsenosides, the main active components of ginseng (Yuan et al. 2015). Hence, research on the genomic information of *Cellulosimicrobium* may discover novel genes of glycosidases with important application in pharmaceutical field.

Whole-genome sequencing is the most comprehensive method for analyzing the genome. Whole-genome sequencing for *Cellulosimicrobium* can provide a large volume of gene data associated with its important functions. *Cellulosimicrobium* within the family *Promicromonosporaceae* was proposed by the reclassification of *C. cellulans* based on its phylogenetic and chemotaxonomic evidences (Schumann et al. 2001). With only three published species, *C. cellulans*, *Cellulosimicrobium funkie*, *terreum terreum*, and 31 16S rRNA gene sequences, *Cellulosimicrobium* remains underrepresented in present NCBI reference databases. The type strains of *C. cellulans* and *C. terreum* were isolated from soils, whereas that of *C. funkie* was isolated from the human blood. Currently, there are only four sequenced completed genomes for *Cellulosimicrobium*, including *Cellulosimicrobium* sp. strain MM (Sharma et al. 2014), *C. cellulans* LMG 16121 (Sharma et al. 2016), *C. cellulans* J36 (Sharma et al. 2016), and *C. funkie* U11 (Karthik et al. 2016).

In this study, *Cellulosimicrobium* sp. TH-20, which was isolated from ginseng soil samples, was identified to

display its effective activity to transform ginsenosides through HPLC analysis. A complete genome sequencing of *Cellulosimicrobium* sp. TH-20 was performed using the Illumina Miseq system to select the genes involved in ginsenoside transformation. The gene functions were annotated through BlastP searches in NCBI Nr, Gene Ontology (GO), Gene Ontology, Cluster of Orthologous Gene (COG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

Materials and methods

Materials

Ginseng root extract was purchased from ZeLang Health-tech Co., Ltd., China. Reference ginsenosides, including ginsenoside Rb1, Re, Rg1, Rd, GypXVII, Rg2, and PPT, were purchased from Jilin University (Changchun, China). Acetonitrile and methanol of HPLC grade were from Fisher (Waltham, MA, USA). HPLC-grade formic acid (96%) was from Tedia (Fairfield, OH, USA). Other reagents were of analytical purity.

Identification of *Cellulosimicrobium* sp. TH-20

The isolated *Cellulosimicrobium* sp. TH-20 was identified through phylogenetic analysis. The 16S rRNA gene of *Cellulosimicrobium* sp. TH-20 was sequenced. The 16S rRNA gene sequences of related taxa were obtained from GenBank (National Center for Biotechnology Information; Bethesda, MD, USA), and a phylogenetic tree was constructed by the neighbor-joining method using the MEGA 3.1 program (Tamura et al. 2007). A bootstrap analysis with 1000 replicates was conducted to obtain confidence levels for the branches.

Biotransformation of ginsenosides by *Cellulosimicrobium* sp. TH-20

Biotransformation of ginsenosides

The biotransformation procedure was carried out using biotransformation medium [0.3% (w/v) NaNO₃, 0.1% (w/v) K₂HPO₃·3H₂O, 0.05% (w/v) MgSO₄, and 0.01% (w/v) FeSO₄·7H₂O] with 125 mg L⁻¹ Rb1, Re, or Rg1 as the carbon source in a shaking incubator (170 rpm) at 30 °C. All media were sterilized at 121 °C for 20 min, and the initial pH was 7.0. An equal volume of *n*-butanol saturated with water was added to stop the reaction. After centrifugation (4 °C, 10,000 rpm), the reactant present in the *n*-butanol fraction was evaporated in a water bath at 60 °C. Samples were redissolved by methanol and diluted to 1.0 mL for further analysis.

HPLC analysis

Chromatographic separation was performed on an Agilent 1200 series HPLC system (Agilent Technologies, USA) equipped with a vacuum degasser, a binary pump, an autosampler, a thermostatically controlled column apartment, and connected to a diode array detector. Samples were loaded onto a C18 reverse-phase chromatographic column (5.0 cm × 3.0 mm, 2.7 μm; Supelco, USA). The mobile phase consisted of water (A) and acetonitrile (B). The gradient elution was programmed as follows: 0–2.5 min, 19% (B); 2.5–5 min, 19–30% (B); 5–11 min, 30–33% (B); 11–20 min, 33–45% (B); and 20–25 min, 45–65% (B). The detection wavelength was 203 nm. The column temperature was 35 °C. The flow rate was 0.2 mL/min, and the injection volume was 5 μL.

DNA extraction, whole-genome sequencing, and assembly

The extraction of genomic DNA from *Cellulosimicrobium* sp. TH-20 was carried out using a bacterial DNA kit (Omega) according to the manufacturer's instructions (Zhou et al. 2016). The genome was sequenced by a whole-genome shotgun sequencing using paired-end sequencing with MiSeq Illumina and long sequencing with PacBio RS (Pacific Biosciences) at Shanghai Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China). A paired-end library (400 bp) was constructed and sequenced, giving approximately 412.30x-fold genome coverage. An approximately 8–10 kb insert PacBio library was constructed and sequenced on an eight single-molecule real-time cell, yielding 81,365 reads (average length, 8215.7 bp). After quality control, these reads were assembled using a novel assembly strategy (Koren et al. 2012).

Genome annotation

Genes were predicted using Glimmer 3.02 (Delcher et al. 2007) and annotated through BlastP (BLAST 2.2.28+) searches in NCBI non-redundant (Nr) database, String (<http://string-db.org/>), GO with Blast2GO (V2.8.0) software, COG, and KEGG (Kanehisa et al. 2004) databases. rRNAs and tRNAs were predicted using Barrnap 0.4.2 and tRNAscan-SE v1.3.1 (Lowe and Eddy 1997), respectively.

Results and discussion

Characterization of *Cellulosimicrobium* sp. TH-20

Cellulosimicrobium sp. TH-20 was identified by phylogenetic analyses. The analyses based upon the 16S rRNA gene sequence of *Cellulosimicrobium* sp. TH-20 and

indicated that the gene sequence was actually grouped with *Cellulosimicrobium* sp. (Fig. 1). Moreover, the highest degree of 16S rRNA gene sequence identity was with *Cellulosimicrobium* sp. CH6 (99%). Hence, TH-20 can be designated as *Cellulosimicrobium* sp. based on the phylogenetic tree and homology analysis.

Biotransformation of ginsenosides by *Cellulosimicrobium* sp. TH-20

To screen effective microorganisms transforming ginsenosides, the ginseng soil samples were collected and cultured on ginseng agar medium plates. More than 20 microorganisms were confirmed to possess transformation ability. We have previously reported the *Aspergillus niger* strain TH-10a, which can transform ginsenoside Rb1 to Rd (Feng et al. 2015).

The transformation ability of *Cellulosimicrobium* sp. TH-20 toward ginsenosides was investigated by HPLC. PPD-type ginsenoside (Rb1) and PPT-type ginsenosides (Re and Rg1) were used in our study. As shown in Fig. 2b, c, Rb1 was almost totally transformed, and 1 and 2 were increased after 96 h of fermentation. By comparison with ginsenoside standards, 1 and 2 with retention time at 8.485 and 9.984 were identified as Rd and GypXVII, respectively. As shown in Fig. 2d, e, Re was completely converted into 3 with retention time at 7.231, and 3 was identified as Rg2 by comparison to the reference standard after 96 h of fermentation. As shown in Fig. 2f, g, Rg1 was significantly transformed, and only one final product 4 was obtained after 144 h of fermentation. In addition, 4 with retention time at 17.126 was assigned as PPT by comparison to the reference standard. Hence, *Cellulosimicrobium* sp. TH-20 can transform ginsenoside Rd, GypXVII, Rg2, and PPT from ginsenoside Rb1, Re, and Rg1.

Ginsenoside transformation by microorganisms and enzymatic deglycosylation were found to be very useful because of high substrate specificity, lower byproducts, and high production yields. Many microorganisms have been explored to transform ginsenosides. Microorganisms isolated from soil-cultivated ginseng, including *Absidia coerulea* (Chen et al. 2007), *Acremonium strictum* (Chen et al. 2008), *Rhizopus stolonifer* (Dong et al. 2003), and *Paecilomyces bainier* fungi (Zhou et al. 2008) and *Burkholderia pyrrocinia* (Kim et al. 2005), *Caulobacter leidyia* (Cheng et al. 2006), *Intrasporangium* sp. (Cheng et al. 2007), and *Microbacterium* sp. (Cheng et al. 2008) bacteria, have been used for the transformation of ginsenosides. Compared with previous reports, ginsenoside transformation by *Cellulosimicrobium* sp. TH-20 presented unique advantages as it can both transform PPD-type (Rb1) and PPT-type (Re and Rg1) ginsenosides. Hence, further research was performed to explore the features of

Fig. 1 Phylogenetic trees based on the 16S rRNA gene sequences, showing the phylogenetic relationships of the *Cellulosimicrobium* sp. TH-20

Cellulosimicrobium sp. TH-20 relevant to its ginsenoside-transforming ability.

Genome features

The genome sequencing of *Cellulosimicrobium* sp. TH-20 was performed by Illumina Miseq to select the genes involved in ginsenoside transformation. The draft genome encoded 3450 open reading frames (ORFs). Of these ORFs, 2114 were genes with predicted functions, 1336 were genes of unknown functions, and 60 were RNA-coding genes, including 51 tRNAs and 9 rRNAs. The average GC content in gene region was 74.7%, which is consistent with that of previously reported *Cellulosimicrobium* genomes. This whole-genome shotgun project has been deposited at GenBank under the accession no. CP020857.1. The *Cellulosimicrobium* sp. TH-20 genome is graphically represented in Fig. 3, whereas its genomic features are summarized in Table 1.

Functional annotation

The functional annotations of genes were performed using NCBI BLAST with Cluster of GO, COG, and KEGG. GO analysis was performed to specify all the annotated nodes comprising of GO functional groups. GO assignments were used to classify the functions of the predicted ORFs. The GO mapping provided the ontology of defined terms representing gene product properties, which were grouped into three main domains consisting of biological process, molecular function, and cellular component. In our study, a total number of 34 GO terms were enriched for the annotated ORFs. Based on GO annotation, maximum ORFs were categorized under biological processes (1855), followed by molecular functions (936) and cellular

components (811). Under the biological processes category, the highest number of ORFs was associated with metabolic process (493) and cellular process (419), followed by single-organism (340). In the molecular function category, catalytic activity (434) represented the most abundant term, followed by binding (315). Under the cellular component category, the highest number of ORFs was associated with cell (235) and cell part (203), followed by membrane (152). Other greater number of sequences was classified under the GO categories viz., localization, establishment of localization, membrane par, biological regulation, and regulation of biological process. Extremely low numbers of genes were classified in terms of membrane-enclosed lumen, extracellular region, and multi-organism process (Fig. 4).

For the analysis of phylogenetically widespread domain families, ORFs were compared against the string database for COG annotation. 636 ORFs were distributed functionally into at least 21 protein families (Fig. 5). Of these ORFs, the cluster for amino acid transport and metabolism represented the largest group (68), followed by inorganic ion transport and metabolism (58), function unknown (49) and signal transduction mechanisms (44), replication, recombination and repair (43), energy production and conversion (43), and general function prediction only (42). The least represented groups included intracellular trafficking, secretion, and vesicular transport (5), followed by cell motility (4) and RNA processing and modification (1). Chromatin structure and dynamics, extracellular structures, nuclear structure, and cytoskeleton had no gene distribution.

To further understand the biological significance of genes, pathway-based analysis was conducted. The KEGG pathway database contains a systematic analysis of inner-cell metabolic pathways and functions of gene products,

Fig. 2 HPLC analysis of the biotransformation process during different periods by *Cellulosimicrobium* sp. TH-20. Standards of ginsenosides (**a**). Transformation of ginsenoside Rb1 (**b**, **c**). Transformation of ginsenoside Re (**d**, **e**). Transformation of ginsenoside Rg1 (**f**, **g**)

which aid in studying the complex biological behavior of genes. In our study, a total of 749 ORFs were assigned to 147 KEGG pathways. The mapped ORFs represented the metabolic pathways of major biomolecules, such as carbon, pyrimidine, amino and nucleotide sugar, glycine, serine, threonine, and others. The mapped ORFs also represented the genes involved in the biosynthesis of secondary metabolites and amino acids, metabolic pathways, two-component system, ABC transporters, and others. The pathways with most representation by the ORFs were metabolic pathways (213), followed by the biosynthesis of secondary metabolites (108) and ABC transporters (63).

Features of *Cellulosimicrobium* sp. TH-20 genome relevant to its ginsenoside biotransformation ability

Cellulosimicrobium sp. TH-20 was identified to exhibit effective activity to transform PPD-type and PPT-type ginsenosides. The PPD and PPT ginsenosides consisted of a nonsugar component (aglycon) with a dammarane skeleton and a sugar component comprised of 1–4 molecules (glycon), including D-glucose, L-arabinopyranoside, L-arabinofuranoside, D-xylose, and/or L-rhamnose. As a result, beta-glucosidase (EC 3.2.1.21), beta-xylosidase (EC 3.2.1.37), alpha-L-arabinofuranosidase (EC 3.2.1.55), and alpha-L-rhamnosidase (EC 3.2.1.40) can be used to transform ginsenosides.

As shown in Fig. 6, the chemical structures of ginsenoside Rb1, Re, Rg1, and their hydrolysis products (Rd, GypXVII, Rg2, and PPT) were displayed. Ginsenoside Rb1 is PPD-type ginsenoside with two sugar moieties attached to the β -OH at C-3 and C-20 in the aglycon PPD, respectively. Rd was formed by cleavage of β -D-glucopyranose linkage of 20-O- β -D-glucopyranosyl-(1,6)- β -D-glucopyranose from Rb1. In addition, GypXVII was created by hydrolysis of β -D-glucopyranose linkage of 3-O- β -D-glucopyranosyl-(1,2)- β -D-glucopyranose from Rb1. Ginsenoside Re and Rg1 are PPT-type ginsenosides with two sugar or one sugar moieties linked to the α -OH at C-6 and one sugar moieties attached to β -OH at C-20 in the aglycon PPT. Rg2 was formed by cleavage of β -D-glucopyranose linkage of 20-O- β -D-glucopyranosyl-(1,6)- β -D-glucopyranose from Re. PPT was formed by cleavage of β -D-glucopyranose linkage of 20-O- β -D-glucopyranosyl-(1,6)- β -D-glucopyranose and 6-O- β -D-glucopyranosyl-(1 $\rightarrow$ 2)- β -D-glucopyranose from Rg1. Hence, beta-glucosidases with various substrate specificities from *Cellulosimicrobium* sp. TH-20 may contribute to different biotransformation of ginsenosides.

Glycoside hydrolases (GH) (EC 3.2.1.) were a widespread group of enzymes that hydrolyze the glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. GHs are

Fig. 3 Circular graph of *Cellulosimicrobium* sp. TH-20. From the outer to the innermost circle, *circle 1* presents the physical coordinates, in 0.5 Mb per interval. *Circles 2 and 3* depict the ORFs on the positive and negative strands. *Different colors* indicate

different COG functions of ORFs. *Circles 4* depict tRNA and rRNA. *Circles 5* depict GC content plot. *Red* indicates the GC content more than average of GC content, while *blue* indicates lower than average of GC content. *Circle 6* depicts the GC-skew (cyan >0, yellow <0)

divided into 135 families at the carbohydrate-active enzyme classification server, CAZY (<http://www.cazy.org/>). Beta-D-Glucosidases belong to the GH families 1, 3, 9, and 30. Beta-D-xylosidases belong to 3, 30, 39, 43, 52, and 54. Alpha-L-arabinofuranosidases belong to 3, 43, 51, 54, and 62. Alpha-L-rhamnosidases belong to 78 and 106.

Glycosidases with ginsenoside-transforming activities were mainly from GH1 and GH3 families. The glycosidases from *Sulfolobus solfataricus*, *Sphingomonas* sp. 2F2, *Thermus thermophilus*, *Paenibacillus mucilaginosus*, *Sphingopyxis alaskensis*, *Pyrococcus furiosus*, *Arthrobacter chlorophenolicus*, and *S. acidocaldarius* were reported to transform ginsenosides and grouped into GH1 (Oh et al.

Table 1 Genome statistics

Gene number	3450
Gene total length (bp)	3,893,364
Gene average length (bp)	1128
Gene density	0.808 genes per kb
GC content in gene region (%)	74.7
Gene/genome (%)	91.3
Intergenic region length (bp)	372,182
GC content in intergenic region (%)	73.4
Intergenic length/genome (%)	8.72

Fig. 4 Classification of protein-coding genes based on gene ontology into three main categories: biological processes, molecular functions and cellular components. The y-axis on the right indicates the number of genes in a category. The y-axis on the left indicates percentage of a specific category of genes in the main category

Fig. 5 Clusters of orthologous groups (COG) functional classification

2014; Noh et al. 2009; Noh and Oh 2009; Liang et al. 2011; Shin et al. 2014). Moreover, *Mucilaginibacter* sp. strain, *Pseudonocardia* sp. Gsoil 1536, *Sanguibacter keddiei*, *Terrabacter ginsenosidimutans*, *Microbacterium esteraromaticum*, *Flavobacterium johnsoniae*, *Actinosynnema mirum*, *Penicillium aculeatum*, *Aspergillus niger*, and *Cladosporium fulvum* were confirmed to produce glycosidases with ginsenoside-transforming activities and classified into

GH3 (Shin et al. 2009; Wei et al. 2011; An et al. 2010; Ruan et al. 2009; Cui et al. 2013; Quan et al. 2012).

In our study, genes relevant to glycoside hydrolases in the genome of *Cellulosimicrobium* sp. TH-20 were analyzed by blasting the genome sequence into the CAZy database (Table 2). Eleven ORFs were predicted to be related to ginsenoside transformation. These genes included 6 for beta-glucosidase, 1 for alpha-N-arabinofuranosidase, 1 for alpha-

Fig. 6 Molecular structures of ginsenoside Rb1, Re, Rg1, Rd, GypXVII, Rg2, and PPT

Table 2 Genes relevant to glycoside hydrolases in the genome of *Cellulosimicrobium* sp. TH-20

Family	Genes count	Genes list
GH1	2	orf00755_1, orf00598_1
GH3	4	orf00846_1, orf01813_1, orf02505_1, orf03101_1
GH43	1	orf01151_1
GH51	1	orf03423_1
GH13	1	orf00963_1
GH10	1	orf00991_1
GH42	1	orf00759_1

1,6-glucosidase, 1 for endo-1,4-beta-xylanase, 1 for alpha-L-arabinofuranosidase, and 1 for beta-galactosidase. These glycosidases were predicted to catalyze the hydrolysis of sugar moieties attached to the aglycon of ginsenosides and led to the transformation of PPD-type and PPT-type ginsenosides.

These 11 genes that are relevant to ginsenoside bio-transformation from *Cellulosimicrobium* sp. TH-20 were grouped into seven different glycosyl hydrolase families. Six ORFs identified to be beta-glucosidases were classified into GH1 and GH3. A 519 bp ORF coding for alpha-N-arabinofuranosidase belonged to glycoside hydrolase family 43. The sequence analysis of this ORF revealed maximum similarity to alpha-N-arabinofuranosidase from *C. cellulans* with 95% identity. Moreover, gene coding for alpha-L-arabinofuranosidase that could catalyze the hydrolysis of terminal non-reducing alpha-L-arabinofuranoside residues in alpha-L-arabinosides was predicted. This ORF was grouped into glycoside hydrolase family 51. Furthermore, gene coding for alpha-1,6-glucosidase, endo-1,4-beta-xylanase, and beta-galactosidase was predicted to be members of GH13, GH10, and GH42, respectively. Hence, the complete genome of *Cellulosimicrobium* sp. TH-20 described in this study provided a large volume of gene data associated with its important functions and served as a starting point for ginsenoside transformation using glycosidases from *Cellulosimicrobium* sp. TH-20.

Conclusion

In this study, the genome of the *Cellulosimicrobium* sp. TH-20 with ginsenoside-transforming activity was sequenced and analyzed. A total of 11 genes were predicted to be related to ginsenoside transformation and grouped into six different glycosyl hydrolase families. These results significantly enlarge the currently known gene repertoire of *Cellulosimicrobium*. These data will serve as a resource for the future functional studies of *Cellulosimicrobium*. The cloning and characterization of

11 genes are being studied for the production of glycoside bioactive molecules in the pharmaceutical industry.

Acknowledgements This research was supported by the National Natural Science Foundation of China (No. 3140040275).

Compliance with ethical standards

Conflict of interest The authors have declared no conflict of interest.

References

- An DS, Cui CH, Lee HG, Wang L, Kim SC, Lee ST, Jin F, Yu H, Chin YW, Lee HK, Im WT, Kim SG (2010) Identification and characterization of a novel *Terrabacter ginsenosidimutans* sp. nov. β -glucosidase that transforms ginsenoside Rb1 into the rare gypenosides XVII and LXXV. *Appl Environ Microbiol* 76:5827–5836
- Chen G, Yang M, Lu Z, Zhang J, Huang H, Liang Y, Guan S, Song Y, Wu L, Guo DA (2007) Microbial transformation of 20(S)-protopanaxatriol-type saponins by *Absidia coerulea*. *J Nat Prod* 70:1203–1206
- Chen GT, Yang M, Song Y, Lu ZQ, Zhang JQ, Huang HL, Wu LJ, Guo DA (2008) Microbial transformation of ginsenoside Rb(1) by *Acremonium strictum*. *Appl Microbiol Biotechnol* 77:1345–1350
- Cheng LQ, Kim MK, Lee JW, Lee YJ, Yang DC (2006) Conversion of major ginsenoside Rb1 to ginsenoside F2 by *Caulobacter leidyia*. *Biotechnol Lett* 28:1121–1127
- Cheng LQ, Na JR, Kim MK, Bang MH, Yang DC (2007) Microbial conversion of ginsenoside Rb1 to minor ginsenoside F2 and gypenoside XVII by *Intrasporangium* sp. GS603 isolated from soil. *J Microbiol Biotechnol* 17:1937–1943
- Cheng LQ, Na JR, Bang MH, Kim MK, Yang DC (2008) Conversion of major ginsenoside Rb1 to 20(S)-ginsenoside Rg3 by *Microbacterium* sp. GS514. *Phytochemistry* 69:218–224
- Cui CH, Kim SC, Im WT (2013) Characterization of the ginsenoside-transforming recombinant β -glucosidase from *Actinosynnema mirum* and bioconversion of major ginsenosides into minor ginsenosides. *Appl Microbiol Biotechnol* 97:649–659
- Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23:673–679
- Dong A, Ye M, Guo H, Zheng J, Guo D (2003) Microbial transformation of ginsenoside Rb1 by *Rhizopus stolonifer* and *Curvularia lunata*. *Biotechnol Lett* 25:339–344
- Feng L, Xu C, Li Z, Li J, Dai Y, Han H, Yu S, Liu S (2015) Microbial conversion of ginsenoside Rd from Rb1 by the fungus mutant *Aspergillus niger* strain TH-10a. *Prep Biochem Biotechnol* 46:336–441
- Fu Y, Cheng L, Meng Y, Li S, Zhao X, Du Y, Yin H (2015) *Cellulosimicrobium cellulans* strain E4-5 enzymatic hydrolysis of curdlan for production of (1 $\rightarrow$ 3)-linked β -D-glucan oligosaccharides. *Carbohydr Polym* 134:740–744
- Hou JG, Xue JJ, Sun MQ, Wang CY, Liu L, Zhang DL, Lee MR, Gu LJ, Wang CL, Wang YB, Zheng Y, Li W, Sung CK (2012) Highly selective microbial transformation of major ginsenoside Rb1 to gypenoside LXXV by *Esteya vermicola* CNU120806. *J Appl Microbiol* 113:807–821
- Hou JG, Xue JJ, Sun MQ, Wang CY, Liu L, Zhang DL, Lee MR, Gu LJ, Wang CL, Wang YB, Zheng Y, Li W, Sung CK (2015) Highly selective microbial transformation of major ginsenoside Rb1 to gypenoside LXXV by *Esteya vermicola* CNU120806. *J Appl Microbiol* 113:807–821

- Hyun YJ, Kim B, Kim DH (2012) Cloning and characterization of ginsenoside Ral-hydrolyzing beta-D-xylosidase from *Bifidobacterium breve* K-110. *J Microbiol Biotechnol* 22:535–575
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280
- Karthik C, Oves M, Thangabalu R, Sharma R, Santhosh SB, Indra Arulselvi P (2016) *Cellulosimicrobium funkei*-like enhances the grow of *Phaseolus vulgaris* by modulating oxidative damage under chromium (VI) toxicity. *J Adv Res* 7:839–850
- Kim MK, Lee JW, Lee KY, Yang DC (2005) Microbial conversion of major ginsenoside Rb1 to pharmaceutically active minor ginsenoside Rd. *J Microbiol* 43:456–462
- Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy Adam M (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* 30:693–700
- Liang WG, Liu QM, Sung BH (2011) Bioconversion of ginsenosides Rb1, Rb2, Rc and Rd by novel-glucosidase hydrolyzing outer 3-O glycoside from *Sphingomonas* sp. 2F2: cloning, expression, and enzyme characterization. *J Biotechnol* 156:125–133
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964
- Noh KH, Oh DK (2009) Production of the rare ginsenosides compound K, compound Y, and compound Mc by a thermostable β -glycosidase from *Sulfolobus acidocaldarius*. *Biol Pharm Bull* 32:1830–1835
- Noh KH, Son JW, Kim HJ, Oh DK (2009) Ginsenoside compound K production from ginseng root extract by a thermostable β -glycosidase from *Sulfolobus solfataricus*. *Biosci Biotechnol Biochem* 73:316–321
- Oh HJ, Shin KC, Oh DK (2014) Production of ginsenosides Rg1 and Rh1 by hydrolyzing the outer glycoside at the C-6 position in protopanaxatriol-type ginsenosides using β -glucosidase from *Pyrococcus furiosus*. *Biotech Lett* 36:113–119
- Quan LH, Jin Y, Wang C, Min JW, Kim YJ, Yang DC (2012) Enzymatic transformation of the major ginsenoside Rb2 to minor compound Y and compound K by a ginsenoside-hydrolyzing β -glycosidase from *Microbacterium esteraromaticum*. *J Ind Microbiol Biotechnol* 39:1557–1562
- Ru W, Wang D, Xu Y, He X, Sun YE, Qian L, Zhou X, Qin Y (2015) Chemical constituents and bioactivities of *Panax ginseng* (C.A. Mey.). *Drug Discov Ther* 9:23–32
- Ruan CC, Zhang H, Zhang LX, Liu Z, Sun GZ, Lei J, Qin YX, Zheng YN, Li X, Pan HY (2009) Biotransformation of ginsenoside Rf to Rh1 by recombinant β -glucosidase. *Molecules* 14:2043–2048
- Schumann P, Weiss N, Stackebrandt E (2001) Reclassification of *Cellulomonas cellulans* (Stackebrandt and Keddle 1986) as *Cellulosimicrobium cellulans* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 51:1007–1017
- Sharma A, Hira P, Shakarad M, Lal R (2014) Draft genome sequence of *Cellulosimicrobium* sp. strain MM, isolated from arsenic-rich microbial mats of a Himalayan hot spring. *Genome Announc* 2:e01020–e01034
- Sharma A, Gilbert JA, Lal R (2016) (Meta) genomic insights into the pathogenome of *Cellulosimicrobium cellulans*. *Sci Rep* 6:25527
- Shin KC, Seo MJ, Oh HJ, Oh DK (2009) A novel ginsenoside Rb1-hydrolyzing β -D-glucosidase from *Cladosporium fulvum*. *Process Biochem* 36:1287–1293
- Shin KC, Oh HJ, Kim BJ, Oh DK (2013) Complete conversion of major protopanaxadiol ginsenosides to compound K by the combined use of α -L-arabinofuranosidase and β -galactosidase from *Caldicellulosiruptor saccharolyticus* and β -glucosidase from *Sulfolobus acidocaldarius*. *J Biotechnol* 167:33–40
- Shin KC, Seo MJ, Oh HJ, Oh DK (2014) Highly selective hydrolysis for the outer glucose at the C-20 position in ginsenosides by β -glucosidase from *Thermus thermophilus* and its application to the production of ginsenoside F2 from gypenoside XVII. *Biotechnol Lett* 36:1287–1293
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Tanabe Y, Oda M (2011) Molecular characterization of endo-1,3- β -glucanase from *Cellulosimicrobium cellulans*: effects of carbohydrate-binding module on enzymatic function and stability. *Biochim Biophys Acta* 1814:1713–1722
- Wei Y, Zhao W, Zhang Q, Zhao Y, Zhang Y (2011) Purification and characterization of a novel and unique ginsenoside Rg1-hydrolyzing β -D-glucosidase from *Penicillium sclerotiorum*. *Acta Biochim Biophys Sin* 43:226–231
- Yuan Y, Hu YB, Hu CX (2015) Overexpression and characterization of a glycoside hydrolase family 1 enzyme from *Cellulosimicrobium cellulans* sp.21 and its application for minor ginsenosides production. *J Mol Catal B Enzym* 120:60–67
- Zhou W, Yan Q, Li JY, Zhang XC, Zhou P (2008) Biotransformation of *Panax noto ginseng* saponins into ginsenoside compound K production by *Paecilomyces bainier* sp. 229. *J Appl Microbiol* 104:699–706
- Zhou M, Liu Q, Xie Y, Dong B, Chen X (2016) Draft genome sequence of *Thermococcus* sp. EP1, a novel hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent on the East Pacific Rise. *Marine Genom* 26:9–11