Journal of Microbiology (2019) Vol. 57, No. 10, pp. 865–873
DOI 10.1007/s12275-019-9217-1

eISSN 1976-3794
pISSN 1225-8873

Antarctic tundra soil metagenome as useful natural resources of cold-active lignocelluolytic enzymes

Han Na Oh[1], Doyoung Park[1], Hoon Je Seong[1], Dockyu Kim[2*], and Woo Jun Sul[1*]

[1]*Department of Systems Biotechnology, Chung-Ang University, Anseong 17546, Republic of Korea*
[2]*Division of Polar Life Sciences, Korea Polar Research Institute, Incheon 21990, Republic of Korea*

(Received May 2, 2019 / Revised Jul 3, 2019 / Accepted Jul 4, 2019)

Lignocellulose composed of complex carbohydrates and aromatic heteropolymers is one of the principal materials for the production of renewable biofuels. Lignocellulose-degrading genes from cold-adapted bacteria have a potential to increase the productivity of biological treatment of lignocellulose biomass by providing a broad range of treatment temperatures. Antarctic soil metagenomes allow to access novel genes encoding for the cold-active lignocellulose-degrading enzymes, for biotechnological and industrial applications. Here, we investigated the metagenome targeting cold-adapted microbes in Antarctic organic matter-rich soil (KS 2-1) to mine lignolytic and celluloytic enzymes by performing single molecule, real-time metagenomic (SMRT) sequencing. In the assembled Antarctic metagenomic contigs with relative long reads, we found that 162 (1.42%) of total 11,436 genes were annotated as carbohydrate-active enzymes (CAZy). Actinobacteria, the dominant phylum in this soil's metagenome, possessed most of candidates of lignocellulose catabolic genes like glycoside hydrolase families (GH13, GH26, and GH5) and auxiliary activity families (AA7 and AA3). The predicted lignocellulose degradation pathways in Antarctic soil metagenome showed synergistic role of various CAZyme harboring bacterial genera including *Streptomyces*, *Streptosporangium*, and *Amycolatopsis*. From phylogenetic relationships with cellular and environmental enzymes, several genes having potential for participating in overall lignocellulose degradation were also found. The results indicated the presence of lignocellulose-degrading bacteria in Antarctic tundra soil and the potential benefits of the lignocelluolytic enzymes as candidates for cold-active enzymes which will be used for the future biofuel-production industry.

***Keywords*:** metagenomics, lignocellulose degradation, SMRT sequencing, CAZy, cold-active enzymes, Antarctica

*For correspondence. (D. Kim) E-mail: envimic@kopri.re.kr; Tel.: +82-32-760-5525; Fax: +82-32-760-5509 / (W.J. Sul) E-mail: sulwj@cau.ac.kr; Tel.: +82-31-670-4707; Fax: +82-31-670-3108.
Copyright ⓒ 2019, The Microbiological Society of Korea

Introduction

Lignocellulose is composed of cellulose, hemicellulose and lignin and accounts for 50% of the world biomasses (Goldstein, 1981). In the biofuel conversion process from lignocellulose, hydrolysis of polysaccharides to fermentable simple sugars and then fermentation of the sugars to ethanol are required (Sun and Cheng, 2002). However, due to the recalcitrant structure of lignocellulose, an usage as an alternative to other biofuel starting materials requires expensive manipulation such as chemical, physical, physicochemical, and biological conversion for removing hemicellulose and lignin before cellulose-derived glucose fermentation by yeast or bacteria. Especially for biological conversion, enzymatic hydrolysis are the expensive processes accounting for more than 60–65% of the total operating costs (Johnson, 2016). Over the past 50 years, various pretreatment methods for bioconversion of lignocellulose to biofuel have been developed and suggested (Rabemanolontsoa and Saka, 2016; Manisha and Yadav, 2017). Still, a large portion of the pretreatment process relies on the two pretreatment processes (Chemical or physical treatment and then biological treatment) because the biological conversion has the many drawbacks. These drawbacks are mainly because of the difficulty of culturing the lignocellulose-degrading filamentous fungi, its slow rates of biological conversion processing, the limited numbers of commercially available lignocellulose conversion enzymes (Maurya *et al.*, 2015), the lack of bacterial and archaeal extremozymes functioning under extreme environments such as higher temperatures andhighly acidic or basic pH conditions (Manisha and Yadav, 2017) and the hindrance due to by-products during the chemical and physical pretreatment processes. Thus, to improve the efficiency of the biological conversion process, discovering a wide variety of lignocellulose degradation enzymes is still demanded.

Here, we conducted metagenome sequencing for discovering the naturally derived cold-active enzymes which have the potential for degrading lignocellulose at extremely low temperatures in soil obtained from near King Sejong Station in Antarctica (a Korean research station operated by the Korea Polar Research Institute). Their optimal temperatures are closer to the global annual temperature than those of mesophilic or thermophilic enzymes (Struvay and Feller, 2012). Thus, they could overcome the limitation of unstable and inert enzyme activity owing to the narrow temperature range of the relevant enzymes during hydrolysis of lignocelluloses when supplied as a part of the enzyme consortium (Manisha and Yadav, 2017). Although many metagenomics studies have demonstrated the potential of lignocellulolytic enzymes, little is known about their capability to degrade lig-

nocellulose in low temperature environments like Antarctica (Pedersen *et al.*, 2011).

Here, we applied both Illumina Miseq and PacBio® RSII sequencing for systematic characterization of bacterial taxonomic profiling and lignocellulose degradation-related functional diversity. To investigate the potential for lignocellulose biodegradation in the cost-competitive production of bioethanol, our exploration was conducted with humic-like substances rich Antarctic tundra soil.

Materials and Methods

Description for sampling site: the Barton Peninsula

The Barton Peninsula is located in King George Island, the western Antarctic Peninsula. The annual average temperature of the Barton Peninsula is -1.8°C with an average temperature of 1.9°C in January. The annual relative humidity and precipitation are 89% and 437.6 mm, respectively. Two vascular plants (*Deschampsia antarctica* and *Colobanthus quitensis*), 33 moss species, and 68 lichen species have been reported in this maritime Antarctic region (Kim *et al.*, 2007, 2016), with lichens and mosses being expected to play the principal roles in vegetation succession. Total organic carbon content in the soil ranges from 0.06 to 2.97% with an average of 0.76%. It seems likely that organic material from seabirds is the strong influence on soil formation in this region (Lee *et al.*, 2004).

Soil sampling and DNA extraction

Jeonjaegyu Hill is located in the neighborhood of the Korean Antarctic Research Station (King Sejong) on the Barton Peninsula. A soil sampling site (designated KS 2; 62°13′31″S, 58°46′ 42″W; 97.3 m above sea level) on Jeonjaegyu Hill was surrounded by a dense moss carpet of *Chorisodontium aciphyllum* with *Sanionia uncinata*, *Polytrichum strictum*, and *Andreaea* sp. (Yu *et al.*, 2014). This site was selected because of the presence of a high amount of soil organic matter, most of which was presumably derived from *C. aciphyllum*. Samples of thawed soil (designated KS 2-1; total organic carbon, 16.3%; total nitrogen, 1.3%; pH 4.6; sand, 37.0%; silt, 33.2%; clay, 29.9%) were collected at depths of 0–10 cm during the summer season (2014 January; Fig. 1). The three sub-samples were combined together and homogenized in a plastic bag, which was stored at -80°C.

Fig. 1. The sampling site for organic matter-rich soils on Jeonjaegyu Hill, King George Island, the western Antarctic Peninsula (photographed in January, 2014).

Soil DNA was extracted from 0.25 g of KS 2-1 soil using PowerSoil® DNA Isolation Kit (MoBio Laboratories Inc.) according to the manufacturer's instructions. DNA concentration and quality were measured using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

Microbial community analysis by 16S rRNA gene sequencing

The extracted soil DNA was used as the template for PCR amplification. The primer set targeted the V4-V5 region of eubacterial 16S rRNA genes (Forward: CCA GCA GC[T,C] GCG GT[G,A] AN; Reverse: CCG TCA ATT CNT TT[G,A] AGT). The quality of PCR amplicon was checked by gel electrophoresis with Difco Noble agar (Becton, Dickinson and Company) and PCR cleanup proceeded using AMPure XP beads (Beckman Coulter). The product was incubated with 10 mM of Tris buffer (pH 8.5) and DNA barcoding was performed using Nextera® XT index kit v2 (Illumina). The quality of the PCR product purified with AMPure XP beads was checked using the aforementioned Nanodrop 2000 spectrophotometer and sequencing of the library was performed by Illumina Miseq platform at Macrogen.

The raw paired sequences were analyzed using QIIME. The sequences were grouped into operational taxonomic units (OTUs; sequences sharing 97% similarity) based on the UCLUST algorithm (Edgar, 2010). After the RDP classifier (Cole *et al.*, 2014) based taxonomic classification with default assignment confidence option, those clustered into the class Chloroplast and the family Mitochondria, which might be considered as originating from eukaryotic cells, were filtered out as described previously (Cole *et al.*, 2014; Oh *et al.*, 2017). Sequences were aligned using the PyNAST method (Caporaso *et al.*, 2009). Calculation of relative abundance of the representation of taxonomic groups in the clustered OTUs was performed by following scripts: *summarize_taxa.py* and *plot_taxa_summary.py*. The raw data were deposited in National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession number SRR5864037.

Metagenome shotgun sequencing and analysis

The metagenomic library was constructed from 0.25 g of KS 2-1 soil using four single molecules, real-time (SMRT) metagenomic sequences. The library was sequenced using the PacBio RS II system and P6/C4 chemistry (Pacific Biosciences) as described previously at DNA Link (Rhoads and Au, 2015). Assembly of the sequences was carried out using a hierarchical genome assembly process (HGAP v2.3) workflow (Chin *et al.*, 2013). The alignment of the short assembled contigs was based on longer contigs for error correction, and a seed library of 4 kb was used for further analysis. The assembled contigs were subjected to taxonomic classification, which was conducted using Burrows-Wheeler Aligner (BWA) mapping by Edge Perl scripts against NCBI RefSeq (O'Leary *et al.*, 2015; Li *et al.*, 2016). Functional gene annotation on the contigs whose sequence length exceeded 500 bp was performed by Prokka (Seemann, 2014). The relative abundance of the taxonomic bins was obtained by dividing the sum of the base

Fig. 2. Predominant taxonomic distribution of KS 2-1 soil at the phylum and class level. The abundance of taxonomic origin in KS 2-1 soil (A) 16S rRNA gene and (B) metagenome. The percent in this figure represents relative abundance of phylum in each result. Only taxa classified with an incidence >1% are shown.

pair to each taxon and normalized by the total number of base pairs in the sample. The relative gene abundance was calculated in terms of TPM (transcript per million) units for normalization. The accession number of metagenome sequences is SRR6040206 in NCBI SRA database.

Gene mining for lignocellulose degrading enzymes

Lignocellulose-degrading enzymes (e-value 1e-5) were predicted from the translated CDS (coding sequences) sequences against CAZymes in the dbCAN database (dbCAN-fam-HMMS.txt; Yin *et al.*, 2012) using the hmmscan tool (version HMMER 3.1b2) and the eggNOG database with the web-based eggNOG-mapper (Huerta-Cepas *et al.*, 2017). From these annotation results and SmartBLAST, several CDS (endoglucanases in GH12; alpha-amylases in GH13; multicopper oxidases in AA1) were selected as cold-active enzyme candidates in KS 2-1 soil metagenome. Phylogenetic analysis was performed to examine the structural and functional relatedness between the new cold-active enzyme and well-characterized psychrophilic, mesophilic, or thermophilic counterparts which were retrieved from the NCBI database. Following alignment of protein sequences by Muscle, the phylogenetic analysis was conducted by MEGA X using neighbor-joining methods (Kumar *et al.*, 2018).

Results and Discussion

Microbial diversity in KS 2-1 soil

The V4-V5 regions of 16S rRNA genes were PCR-amplified and sequenced to characterize the microbial community in the Antarctic organic matter-rich soil KS 2-1. A total of 153,386 paired-end reads were produced, merged, and clustered into 7,063 OTUs excluding those assigned to Chloroplast and Mitochondria. Most of the OTUs (99.98%) were assigned to bacterial origin, with a trace of archaea sequences. While taxonomy of OTUs from KS 2-1 belongs to 27 phyla, 167 orders, and 411 genera, the predominant phylum was Proteobacteria (24.3%), followed by Bacteroidetes (14.3%), Acidobacteria (13.5%), Chloroflexi (12.2%), Planctomycetes (10.0%), and Actinobacteria (7.4%) which collectively accounted for 80% of all the 16S rRNA gene sequences (Fig. 2A). The higher abundances of Proteobacteria were also observed in the dry soils with neutral to alkaline pH in the Ross Sea region located on the Antarctica (Aislabie *et al.*, 2008) and the soils with high primary productivity on Northern Victoria Land, Antarctica (Niederberger *et al.*, 2008).

The longer contigs obtained from KS 2-1 metagenome SMRT sequencing allowed to assign the definite bacterial genera by more accurate contig's taxonomic annotation and precisely predict lignocellulose degradation-related enzymes as below. A total of 35,919 reads (N50 length of 6,575 bp; mean read length of 3,437 bp) were produced using the SMRT cells and then assembled into 2,116 contigs with mean 34.3-fold coverage using the HGAP assembly (Table 1). Classifying contig's taxonomy using BWA mapping to NCBI RefSeq, a total of 10 phyla were identified, accounting for 79.2% of the con-

Table 1. Summary of metagenomic data obtained from KS 2-1 soil using PacBio® RS II and Illumina Miseq sequencing

Summary	Illumina Miseq
Total Bases	22,299,637 bp
Read Count	59,493
Average length	374.828
	PacBio RS II
Total Bases	123,452,758 bp
Read Count	35,919
Number of contigs	2,116
Mapped Bases (contig length* depth)	229,529,422.393
Mean coverage	34.3
N50 contig length	6,575
Max contig length	72,361
Number of CDS	11,436
Number of CAZy-annotated CDS	162

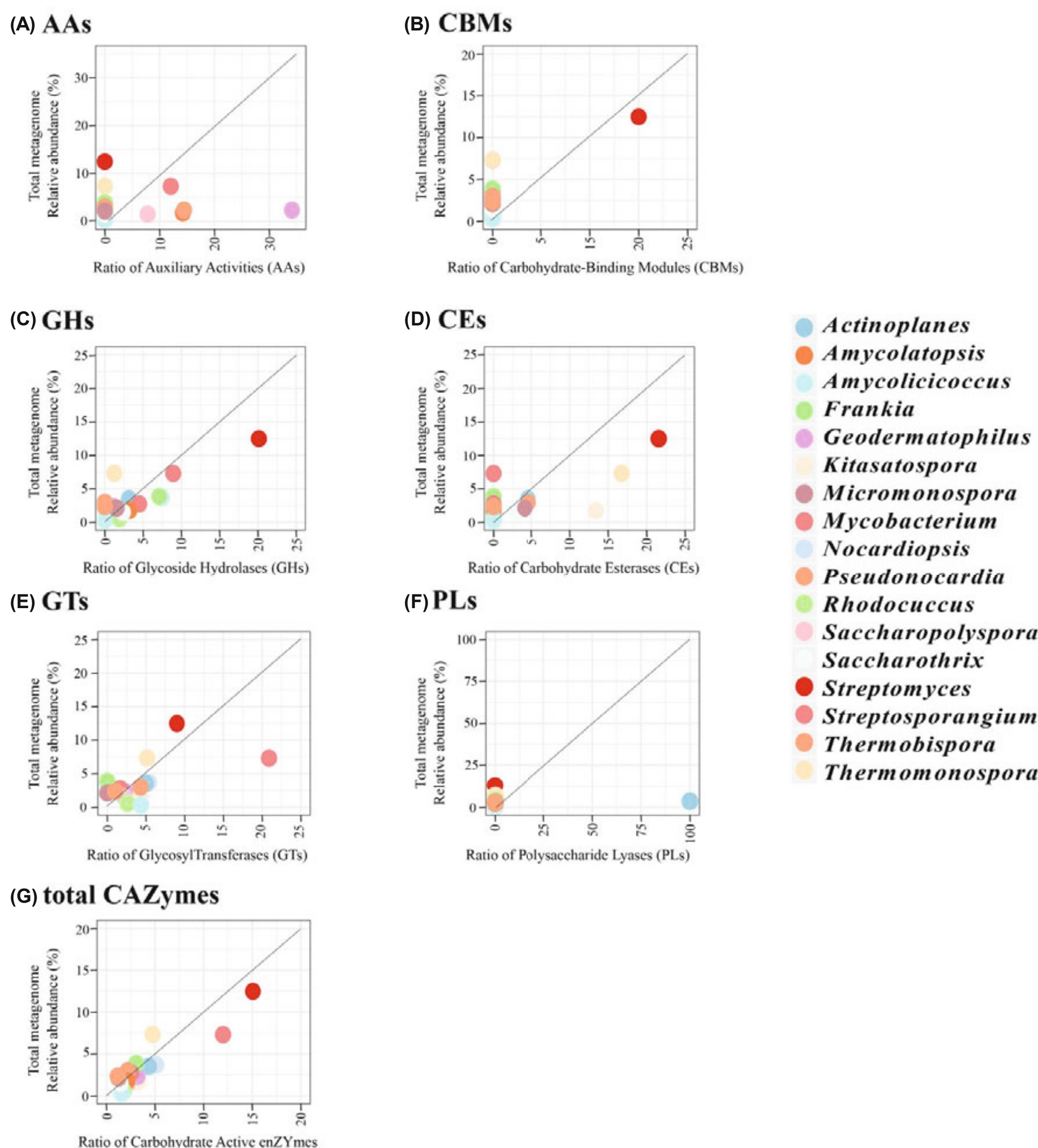

Fig. 3. Scatter plot of relative abundances of the total metagenome versus the proportion of (A) auxiliary activity, AA, (B) carbohydrate-binding module, CBM, (C) glycoside hydrolase, GH, (D) carbohydrate esterase, CE, (E) glycosyl transferase, GT, (F) polysaccharide lyase, PL, and (G) total CAZymes in the 17 genera with the greatest number of CAZyme genes. The x-axis values are relative proportion of genes in each CAZyme class from the 17 genera and the y-axis values are relative abundance of the 17 genera in total metagenome contigs. All genera have a relative abundance of >1.0% in the entire CAZymes in KS 2-1 soil metagenome data.

tigs and the rest of contigs were unidentified. In KS 2-1 metagenome, the predominant phylum was determined as Actinobacteria (73.4%), followed by Proteobacteria (4.1%; Fig. 2B). Although Actinobacteria and Proteobacteria are the most numerically abundant in both Arctic and Antarctic soils, in KS 2-1 soil Actinobacteria seem to better adapt and more actively metabolize (i.e., decomposition of soil organic matter) at low temperatures than Proteobacteria, exhibiting the more dominance of cold-adapted metabolic genes from the Actinobacteria (Makhalanyane *et al.*, 2016).

Abundance and distribution pattern of lignocellulose-degrading enzymes

To investigate lignocellulose degradation potentials in KS 2-1 soil, the genes in assembled contigs encoding CAZymes that catalyze carbohydrate-related substrates were analyzed. The CAZymes contains three main enzyme classes for lignocellulose degradation: Glycoside hydrolase (GH), carbohydrate-binding module (CBM), and auxiliary activity (AA). Specifically, AA, GH, and CBM contain enzymes degrading lignin,

degrading cellulose and hemicellulose, and promoting the association of the CAZymes and carbohydrate-related substrates, respectively. A total of 162 CDS (1.42%) were identified as CAZyme out of 11,436 CDS with cutoff E-value < 1e-5. Among these 162 CDS, 72 (44.4%), 3 (1.9%), and 8 (5.0%) CDS were assigned to GH, CBM, and AA classes, respectively. As GH members were detected even more frequently than those in AAs, the degradation efficiency and/or rate of cellulose and hemicellulose degradation in KS 2-1 is assumed to be much higher than that of lignin-related substrate degradation.

Seventeen genera were examined for the taxonomic distribution patterns of CAZymes in KS 2-1 (Fig. 3). Totally, the CAZymes were unevenly distributed among genera *Streptomyces* (15.1% in total CAZymes), *Streptosporangium* (12.0%) and *Nocardiopsis* (5.1%; Fig. 3A). Figure 3 showed distinct configuration in each CAZyme class: The genera including *Geodermatophilus*, *Thermobispora*, and *Amycolatopsis* had remarkably higher abundances in AA class than the others (Fig. 3A). Genus *Streptomyces* was the most significant in CBM class (Fig. 3B), and genera including *Streptomyces*, *Streptosporangium*, *Frankia*, and *Nocardiopsis* were dominant in GH class (Fig. 3C). Likewise, in other three enzyme classes, various genera had different pattern in each class. Genera including *Streptomyces*, *Thermomonospora*, and *Kitasatospora* had larger portion in carbohydrate esterase (CE) class, containing hydrolytic enzymes acting on ester bonds in carbohydrates, than the others (Fig. 3D) and genera including *Streptosporangium*, *Mycobacterium*, and *Actinoplanes* had higher abundances in glycosyl transferase (GT) class that has a role in formation of glycosidic bonds (Fig. 3E). The only one genus, *Actinoplanes*, had higher abundance in polysaccharide lyase (PL) class that cleaves glycosidic bonds non-hydrolytically and had one PL family, PL12. The scatter plots illustrating a correlation between the relative abundances of genera and each CAZyme class indicated that microbial consortium in KS 2-1 has a potential for lignocellulose degradation, with the highest abundance of CAZyme genes being assigned to Actinobacteria (73.9% in total CAZymes) including *Streptomyces* and *Streptosporangium* in KS 2-1. Recently, meta-transcriptomic and eco-physiological evidences showed that in addition to *Streptomyces* spp. strains, *Micromonospora*, and Actinobacterial strains were also identified as degraders of crystalline cellulose (Ventorino *et al.*, 2016) and even lignocellulosic biomass (Simmons *et al.*, 2014; Yeager *et al.*, 2017).

It has broadly been known that carbohydrates (i.e., cellulose and hemicellulose) and lignin are degraded through hydrolytic (cellulolytic) and oxidative (ligninolytic) systems, respectively. However, the differentiation between cellulolytic and lignolytic enzymes seems to be intricate. Thus, this novel AA class, including the families of redox enzymes (e.g., multicopper oxidases and Class II peroxidases) involved in lignin breakdown, was integrated in CAZy database (Levasseur *et al.*, 2013). To better profile the lignocellulolytic enzymes in the KS 2-1 metagenomic data, AA and GH classes were examined at each family level based on eggNOG database. In terms of ligninolytic AA, the enzymes in AA7 family (glucooligosaccharide oxidase) accounted for the greatest abundance ratio (3.0% in total CAZymes, Supplementary data Table S1) and followed by AA3 (GMC oxidoreductase) and AA4 (vanillyl alcohol oxidase). The functions of AA4 and AA7 associated with aromatic ring attack in lignin structures and was also confirmed in another metagenomics-based study for bacterial lignocellulose degradation in wheat straw, switch grass, and corn stover biomass samples (Jiménez *et al.*, 2016). Additionally, four CDS were assigned to three AA1

Table 2. The list of publicly avaiable enzymes used for comparative phylogenetic analysis with the cold-active enzyme candidate in KS 2-1 soil

Family	Enzyme	Organism	Opt. temp.	Reference
GH12; endoglucanase	Tm_Cel5A	*Thermotoga maritima*	80°C	Pereira *et al.* (2010)
	Eg1II	*Bacillus amyloliquefaciens*	50°C	Nurachman *et al.* (2010)
	EngB	*Clostridium cellulovarans*743B	45°C	Foong *et al.* (1991)
	EG5C	*Paenibacillus* sp. IHB B 3084	40°C	Dhar *et al.* (2015)
	celCCA	*Ruminoclostridium cellulolyticum*	28°C	Ducros *et al.* (1995)
	EGD	*Fibrobacter succinogenes* subsp. Succinogenes S85	35°C	Iyo and Forsberg (1999)
	CelG		25°C	
	Cel5G	*Pseudoalteromonas haloplanktis*	4°C	Garsoux *et al.* (2004)
GH13; α-amylase	chimeric amylase	*Bacillus amyloliquefaciens*	70°C	Kikani and Singh (2011)
	GH13	*Alicyclobacillus acidocaldarius*	75°C	Matzke *et al.* (1997)
	α-Amylase 1	*Thermotoga maritima* MSB8	70°C	Lim *et al.* (2003)
	α-Amylase 2	*Bacillus circulans*	48°C	Dey *et al.* (2002)
	α-Amylase 3	*Nesterenkonia* sp. strain F	45°C	Shafiei *et al.* (2010)
	α-Amylase 4	*Bacillus* sp. YX 1	40–50°C	Liu and Xu (2008)
	α-Amylase 5	*Microbacterium foliorum*	20°C	Kuddus (2014)
	GH13	*Aeromonas veronii*	10°C	Samie *et al.* (2012)
AA1; Laccase like multicopper-oxidase (LMCO)	cotA	*Bacillus subtilis* subsp. 168	75°C	Martins *et al.* (2002)
	CueO	*Escherichia coli* str. K-12 substr. MG1655	70°C	Kim *et al.* (2001)
	Laccase domain 1	*Azospirillum lipoferum*	70°C	Devasia and Nair (2016)
	CuOx	*Paenibacillus glucanolyticus* SLM1	40°C	Mathews *et al.* (2016)
	ScLac laccase	*Streptomyces cyaneus*	40°C	Niladevi *et al.* (2008)
	Laccase domain 2	*Streptomyces bikiniensis*	50–60°C	Devi *et al.* (2016)

family (multicopper oxidase) enzymes, containing laccase (AA1_1 subfamily) and laccase-like multicopper oxidase (LMCO, AA1_3 subfamily), and one dye-decolorizing peroxidase (DyP, not yet integrated to AA family) using eggNOG-mapper. These extracellular laccase and peroxidase enzymes from white-rot fungi and Actinobacteria are assumed to oxidize and then depolymerize lignin. The mono- and bi-aromatic compounds derived from lignin depolymerization are further degraded by bacteria. Recently, bacterial DyPs have been detected in various bacteria (Tian *et al.*, 2017) and shown to be active for oxidation of polymeric lignin in several soil bacteria. Bacterial LMCOs have been reported in various lignin-degrading bacterial genera (Wang *et al.*, 2016b; Granja-Travez *et al.*, 2018). However, the functional mechanism of LMCOs for lignin degradation is not yet fully characterized, mainly owing to their functional diversity and wide range of substrates in different bacterial strains.

Among the cellulose and hemicellulose degradation-related CAZymes, several GH families, GH13, GH26, GH65, GH15, and GH5, were frequently found in KS 2-1 metagenome data. In particular, endoglucanase, one of three main types of cellulases, in GH5 and α-amylase, catalyzing the hydrolysis of starch into sugars, in GH13 have been studied well in psychrophilic bacteria, deemed cold-active glucoside hydrolases (Struvay and Feller, 2012). A lignocellulose degradation pathway by KS 2-1 microbial consortium was predicted based on the known activities and taxonomic distribution pattern of various CAZymes in KS 2-1 (Fig. 4). Within the predicted pathway, cellulose was degraded by GH5, GH9, and GH12, and the cellulose degradation products were further degraded to glucose by GH3, GH4, and GH13. A main component of hemicellulose, xylan, was degraded by GH16, GH51, and GH3. The other components, mannan and galactan were degraded by specific CAZymes including GH26, GH59, and GH2. The presence of various cellulolytic and lignolytic enzymes in CAZyme database implied the potential of KS 2-1 soil as metagenomic resources for new and useful lignocellulose degradation enzymes (Fig. 4). To sum up, lignocellulose degradation is being processed by a synergistic response of diverse soil microbial communities. It is possibly assumed that a co-existing fungal-bacterial community within KS 2-1 soil is well adapted to and functionally active at low temperatures. Even though fungal group is one of the major microbial communities for decomposition of plant-derived organic compounds, diverse bacterial taxonomic groups (in this case, in phylum Actinobacteria) equipped with their specific CAZymes are maybe playing a more significant role in this organic matter-rich cold environment as in other published reports (Wang *et al.*, 2016a). Also, it seems that these cold-active enzymes could be more active at higher temperatures increased during short summer season in the Antarctic tundra, participating in all the cellulose, hemicellulose, and lignin degradation.

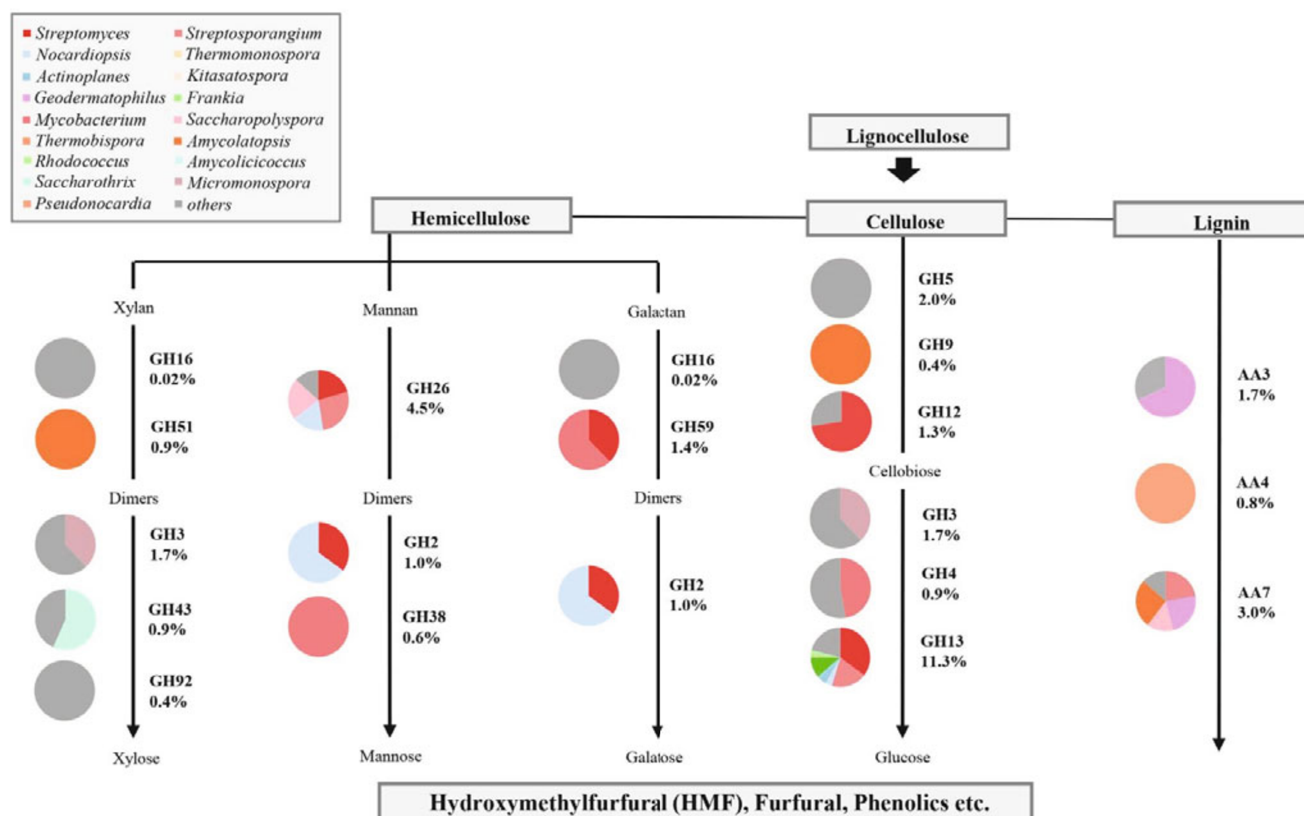

Fig. 4. Predicted lignocellulose degradation pathways in KS 2-1 soil with the relative abundance. The genera possessing lignocellulose degradation-related CAZymes with an incidence >1.0% in KS 2-1 soil data were shown. The percentages are relative abundance of each CAZyme family in the total CAZymes.

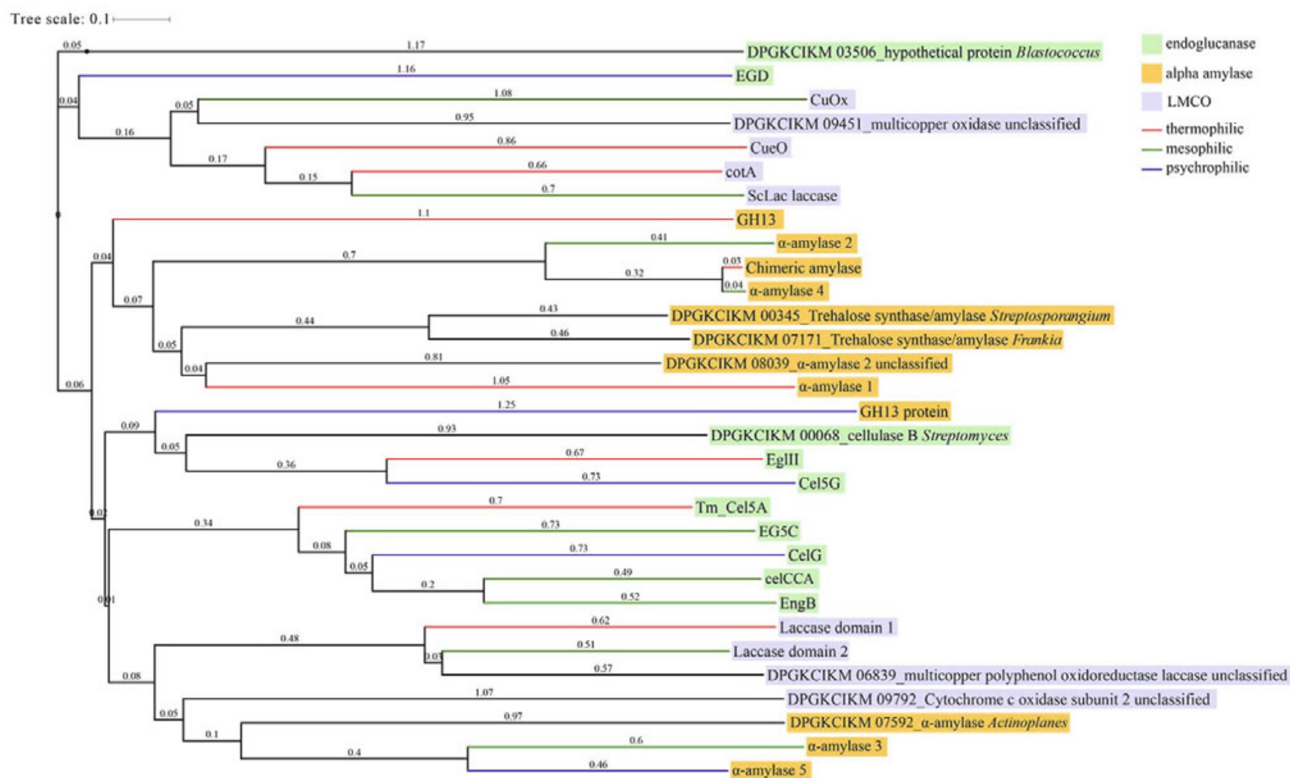

Fig. 2. Phylogenetic trees constructed using amino acid sequences corresponding to the cold-active gene candidates in KS 2-1 soil data with psychrophilic/mesophilic/thermophilic reference enzymes. The bacterial amino acid sequences having the same enzyme name with our cold-active gene candidates were retrieved from published data and clustered using MEGA X. GH12; endoglucanase, GH13; α-amylase and laccase like multicopper oxidase were selected as cold-active enzyme candidates based on the genes that played an important role in the degradation of cellulose, hemicellulose and lignin and showed relatively high abundances in KS 2-1 soil data.

Phylogenetic analyses of cold-active lignocellulolytic enzymes

To examine the phylogenetic similarities and differences between the cold-active enzyme candidates in KS 2-1 and other previously characterized psychrophilic/mesophillic/thermophillic enzymes, phylogenetic trees were constructed targeting endoglucanases in GH12, α-amylases in GH13, and LMCOs in AA1 (Table 2 and Fig. 5). Generally, the biological roles of laccases and LMCOs in nature are the construction and de-construction of polymeric lignin (Moghadam *et al.*, 2016). Cellulases are associated with the degradation of cellulose, lichenin, and cereal β-D-glucans, and α-amylases are involved in the hydrolysis of α(1→4) linkages in polysaccharides (McDonald *et al.*, 2008). The amino acid sequence of one endoglucanase candidate (DPGKCIKM 00068) in KS 2-1 closely clustered with those from several endoglucanase references were characterized to be thermophilic, mesophilic, or psychrophilic. The most closely clustered reference enzymes, being grouped into a subcluster, were eglII and Cel5G, thermophilic and psychrophilic endoglucanase, respectively (Garsoux *et al.*, 2004; Nurachman *et al.*, 2010). Another endoglucanase candidate, DPGKCIKM 03506, however, did not cluster with any other reference enzymes. A total of 18 amino acid sequences were found to be in GH13, of which four sequences were annotated to α-amylase and used to construct phylogenetic trees. Three candidates (DPGKCIKM 00345, 07171, and 08039) were found to be closely related to a thermophilic α-amylase from thermophile *Thermotoga maritima* MSB8 (Lim *et al.*, 2003) and the candidate DPGKCIKM 07592 was predicted to be similar to α-amylases from mesophilic *Nesterenkonia* sp. strain F and psychrophilic *Microbacterium foliorum*. On the other hand, two LMCO candidates, DPGKCIKM 09451 and 06839, grouped with the mesophilic *Paenibacillus glucanolyticus* SLM1 (Mathews *et al.*, 2016) and *Streptomyces bikiniensis* laccase domain 2 (Devi *et al.*, 2016), respectively. Like the candidate endoglucanase (DPGKCIKM 03506), DPGKCIKM 09792 did not cluster with any other LMCO reference enzymes. As shown in Fig. 5, the phylogenetic tree showed a result contrary to a general perception that a novel cold-active enzyme catalyzing lignocellulose degradation could be phylogenetically close to existing cold-active ones. Together, this phylogenetic analysis indicates that a wide range of cold-active lignocellulolytic CAZymes are present in Antarctic soil metagenomes including KS 2-1 and these cold-active enzymes are not stringently correlated to their natural environmental conditions.

Finally, from this study we suggest that cold-adapted bacteria are thought to be one of the main players in soil organic matter decomposition in natural cold environments, because they are both abundant and diverse, possessing versatile catabolic enzymes and pathways which are well adapted at low temperatures. The higher abundances of lignocellulolytic enzymes from phylum Actinobacteria in KS 2-1 metagenome

indicated that Actinobateria play an essential role in lignocellulose degradation in Antarctic soil. The organic matter-rich Antarctic soils are believed to be a reservoir of cold-active lignocellulolytic enzymes. Through this work, we provided a list of cold-active enzyme candidates in GH12, GH13, and AA1, capable of synergistic lignocellulose hydrolysis for the microbe-based biotechnological production of biofuel and value-added chemicals at relatively low temperatures.

Acknowledgements

This research was supported by a grant to the Korea Polar Research Institute (PE19090) and by the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (MOF) (No. 20180430).

References

Aislabie, J.M., Jordan, S., and Barker, G.M. 2008. Relation between soil classification and bacterial diversity in soils of the Ross Sea region, Antarctica. *Geoderma* **144**, 9–20.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., and Knight, R. 2009. Pynast: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., and Eichler, E.E. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. 2014. Ribosomal database project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633–642.

Devasia, S. and Nair, A.J. 2016. Screening of potent laccase producing organisms based on the oxidation pattern of different phenolic substrates. *Int. J. Curr. Microbiol. Appl. Sci.* **5**, 127–137.

Devi, P., Kandasamy, S., Chendrayan, K., and Uthandi, S. 2016. Laccase producing *Streptomyces bikiniensis* CSC 12 isolated from compost. *J. Microbiol. Biotechnol. Food Sci.* **6**, 794–798.

Dey, G., Palit, S., Banerjee, R., and Maiti, B. 2002. Purification and characterization of maltooligosaccharide-forming amylase from *Bacillus circulans* GRS 313. *J. Ind. Microbiol. Biotechnol.* **28**, 193–200.

Dhar, H., Kasana, R.C., Dutt, S., and Gulati, A. 2015. Cloning and expression of low temperature active endoglucanase EG5C from *Paenibacillus* sp. IHB B 3084. *Int. J. Biol. Macromol.* **81**, 259–266.

Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H.P., Belaich, J.P., Davies, G.J., and Haser, R. 1995. Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* **3**, 939–949.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

Foong, F., Hamamoto, T., Shoseyov, O., and Doi, R.H. 1991. Nucleotide sequence and characteristics of endoglucanase gene *engB* from *Clostridium cellulovorans. J. Gen. Microbiol.* **137**, 1729–1736.

Garsoux, G., Lamotte, J., Gerday, C., and Feller, G. 2004. Kinetic and structural optimization to catalysis at low temperatures in a psychrophilic cellulase from the antarctic bacterium *Pseudoalteromonas haloplanktis*. *Biochem. J.* **384**, 247–253.

Goldstein, I.S. 1981. Composition of biomass. *In* Goldstein, I.S. (ed.), Organic Chemicals from Biomass, pp. 9–19. CRC Press Inc., Boca Raton, FL, USA.

Granja-Travez, R.S., Wilkinson, R.C., Persinoti, G.F., Squina, F.M., Fülöp, V., and Bugg, T.D. 2018. Structural and functional characterisation of multi-copper oxidase CueO from lignin-degrading bacterium *Ochrobactrum* sp. reveal its activity towards lignin model compounds and lignosulfonate. *FEBS J.* **285**, 1684–1700.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122.

Iyo, A.H. and Forsberg, C.W. 1999. A cold-active glucanase from the ruminal *Bacteriumfibrobacter succinogenes* S85. *Appl. Environ. Microbiol.* **65**, 995–998.

Jiménez, D.J., de Lima Brossi, M.J., Schückel, J., Kračun, S.K., Willats, W.G.T., and van Elsas, J.D. 2016. Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Appl. Microbiol. Biotechnol.* **100**, 10463–10477.

Johnson, E. 2016. Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuel Bioprod. Biorefin.* **10**, 164–174.

Kikani, B. and Singh, S. 2011. Single step purification and characterization of a thermostable and calcium independent α-amylase from *Bacillus amyloliquifaciens* TSWK1-1 isolated from Tulsi Shyam hot spring reservoir, Gujarat (India). *Int. J. Biol. Macromol.* **48**, 676–681.

Kim, J.H., Ahn, I.Y., Lee, K.S., Chung, H., and Choi, H.G. 2007. Vegetation of Barton peninsula in the neighbourhood of King Sejong Station (King George Island, maritime Antarctic). *Polar. Biol.* **30**, 903–916.

Kim, S.C., Kim, J.S., Hong, B.R., Hong, S.G., Kim, J.H., and Lee, K.S. 2016. Assembly processes of moss and lichen community with snow melting at the coastal region of the Barton peninsula, maritime Antarctic. *J. Ecol. Environ.* **40**, 8.

Kim, C., Lorenz, W.W., Hoopes, J.T., and Dean, J.F. 2001. Oxidation of phenolate siderophores by the multicopper oxidase encoded by the *Escherichia coli yacK* gene. *J. Bacteriol.* **183**, 4866–4875.

Kuddus, M. 2014. Bio-statistical approach for optimization of cold-active α-amylase production by novel psychrotolerant *M. foliorum* GA2 in solid state fermentation. *Biocatal. Agric. Biotechnol.* **3**, 175–181.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. Mega X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549.

Lee, Y.I., Lim, H.S., and Yoon, H.I. 2004. Geochemistry of soils of king george island, South Shetland Islands, West Antarctica: Implications for pedogenesis in cold polar regions. *Geochim. Cosmochim. Acta.* **68**, 4319–4333.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., and Henrissat, B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**, 41.

Li, P.E., Lo, C.C., Anderson, J.J., Davenport, K.W., Bishop-Lilly, K.A., Xu, Y., Ahmed, S., Feng, S., Mokashi, V.P., and Chain, P.S. 2016. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Res.* **45**, 67–80.

Lim, W.J., Park, S.R., An, C.L., Lee, J.Y., Hong, S.Y., Shin, E.C., Kim, E.J., Kim, J.O., Kim, H., and Yun, H.D. 2003. Cloning and characterization of a thermostable intracellular α-amylase gene from the hyperthermophilic bacterium *Thermotoga maritima* MSB8. *Res. Microbiol.* **154**, 681–687.

Liu, X.D. and Xu, Y. 2008. A novel raw starch digesting α-amylase from a newly isolated *Bacillus* sp. YX-1: Purification and characterization. *Bioresour. Technol.* **99**, 4315–4320.

Makhalanyane, T.P., Van Goethem, M.W., and Cowan, D.A. 2016. Microbial diversity and functional capacity in polar soils. *Curr.*

Opin. Biotechnol. **38**, 159–166.

Manisha, and Yadav, S.K. 2017. Technological advances and applications of hydrolytic enzymes for valorization of lignocellulosic biomass. *Bioresour. Technol.* **245**, 1727–1739.

Martins, L.O., Soares, C.M., Pereira, M.M., Teixeira, M., Costa, T., Jones, G.H., and Henriques, A.O. 2002. Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the *Bacillus subtilis* endospore coat. *J. Biol. Chem.* **277**, 18849–18859.

Mathews, S.L., Smithson, C.E., and Grunden, A.M. 2016. Purification and characterization of a recombinant laccase-like multi-copper oxidase from *Paenibacillus glucanolyticus* SLM1. *J. Appl. Microbiol.* **121**, 1335–1345.

Matzke, J., Schwermann, B., and Bakker, E.P. 1997. Acidostable and acidophilic proteins: The example of the α-amylase from *Alicyclobacillus acidocaldarius*. *Comp. Biochem. Physiol.* **118**, 475–479.

Maurya, D.P., Singla, A., and Negi, S. 2015. An overview of key pretreatment processes for biological conversion of lignocellulosic biomass to bioethanol. *3 Biotech.* **5**, 597–609.

McDonald, A.G., Boyce, S., and Tipton, K.F. 2008. ExplorEnz: The primary source of the IUBMB enzyme list. *Nucleic Acids Res.* **37**, D593–D597.

Moghadam, M.S., Albersmeier, A., Winkler, A., Cimmino, L., Rise, K., Hohmann-Marriott, M.F., Kalinowski, J., Rückert, C., Wentzel, A., and Lale, R. 2016. Isolation and genome sequencing of four Arctic marine *Psychrobacter* strains exhibiting multicopper oxidase activity. *BMC Genomics* **17**, 117.

Niederberger, T.D., McDonald, I.R., Hacker, A.L., Soo, R.M., Barrett, J.E., Wall, D.H., and Cary, S.C. 2008. Microbial community composition in soils of Northern Victoria Land, Antarctica. *Environ. Microbiol.* **10**, 1713–1724.

Niladevi, K.N., Jacob, N., and Prema, P. 2008. Evidence for a halotolerant-alkaline laccase in *Streptomyces psammoticus*: Purification and characterization. *Process Biochem.* **43**, 654–660.

Nurachman, Z., Kurniasih, S.D., Puspitawati, F., Hadi, S., Radjasa, O.K., and Natalia, D. 2010. Cloning of the endoglucanase gene from a *Bacillus amyloliquefaciens* PSM3.1 in *Escherichia coli* revealed catalytic triad residues thr-his-glu. *Am. J. Biochem. Biotechnol.* **6**, 268–274.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., and Ako-Adjei, D., *et al.* 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745.

Oh, H.N., Lee, T.K., Park, J.W., No, J.H., Kim, D., and Sul, W.J. 2017. Metagenomic SMRT sequencing-based exploration of novel lignocellulose-degrading capability in wood detritus from *Torreya nucifera* in Bija Forest on Jeju island. *J. Microbiol. Biotechnol.* **27**, 1670–1680.

Pedersen, M., Johansen, K.S., and Meyer, A.S. 2011. Low temperature lignocellulose pretreatment: effects and interactions of pretreatment pH are critical for maximizing enzymatic monosaccharide yields from wheat straw. *Biotechnol. Biofuels* **4**, 11.

Pereira, J.H., Chen, Z., McAndrew, R.P., Sapra, R., Chhabra, S.R., Sale, K.L., Simmons, B.A., and Adams, P.D. 2010. Biochemical characterization and crystal structure of endoglucanase Cel5a from the hyperthermophilic *Thermotoga maritima*. *J. Struct. Biol.* **172**, 372–379.

Rabemanolontsoa, H. and Saka, S. 2016. Various pretreatments of lignocellulosics. *Bioresour. Technol.* **199**, 83–91.

Rhoads, A. and Au, K.F. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289.

Samie, N., Noghabi, K.A., Gharegozloo, Z., Zahiri, H.S., Ahmadian, G., Sharafi, H., Behrozi, R., and Vali, H. 2012. Psychrophilic α-amylase from *Aeromonas veronii* NS07 isolated from farm soils. *Process Biochem.* **47**, 1381–1387.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.

Shafiei, M., Ziaee, A.A., and Amoozegar, M.A. 2010. Purification and biochemical characterization of a novel SDS and surfactant stable, raw starch digesting, and halophilic α-amylase from a moderately halophilic bacterium, *Nesterenkonia* sp. strain F. *Process Biochem.* **45**, 694–699.

Simmons, C.W., Reddy, A.P., D'haeseleer, P., Khudyakov, J., Billis, K., Pati, A., Simmons, B.A., Singer, S.W., Thelen, M.P., and VanderGheynst, J.S. 2014. Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnol. Biofuels* **7**, 495.

Struvay, C. and Feller, G. 2012. Optimization to low temperature activity in psychrophilic enzymes. *Int. J. Mol. Sci.* **13**, 11643–11665.

Sun, Y. and Cheng, J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* **83**, 1–11.

Tian, M., Du, D., Zhou, W., Zeng, X., and Cheng, G. 2017. Phenol degradation and genotypic analysis of dioxygenase genes in bacteria isolated from sediments. *Braz. J. Microbiol.* **48**, 305–313.

Ventorino, V., Ionata, E., Birolo, L., Montella, S., Marcolongo, L., de Chiaro, A., Espresso, F., Faraco, V., and Pepe, O. 2016. Lignocellulose-adapted endo-cellulase producing streptomyces strains for bioconversion of cellulose-based materials. *Front. Microbiol.* **7**, 2061.

Wang, C., Dong, D., Wang, H., Müller, K., Qin, Y., Wang, H., and Wu, W. 2016a. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* **9**, 22.

Wang, L., Nie, Y., Tang, Y.Q., Song, X.M., Cao, K., Sun, L.Z., Wang, Z.J., and Wu, X.L. 2016b. Diverse bacteria with lignin degrading potentials isolated from two ranks of coal. *Front. Microbiol.* **7**, 1428.

Yeager, C.M., Gallegos-Graves, V., Dunbar, J., Hesse, C.N., Daligault, H., and Kuske, C.R. 2017. Polysaccharide degradation capability of *Actinomycetales* soil isolates from a Semiarid Grassland of the Colorado Plateau. *Appl. Environ. Microbiol.* **83**, pii: e03020-16.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. 2012. Dbcan: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451.

Yu, N.H., Kim, J.A., Jeong, M.H., Cheong, Y.H., Hong, S.G., Jung, J.S., Koh, Y.J., and Hur, J.S. 2014. Diversity of endophytic fungi associated with bryophyte in the maritime Antarctic (King George Island). *Polar Biol.* **37**, 27–36.