Genome Sequencing and Genome-Wide Identification of Carbohydrate-Active Enzymes (CAZymes) in the White Rot Fungus *Flammulina fennae*

Chang-Soo Lee¹, Won-Sik Kong², and Young-Jin Park¹*

¹Department of Biomedical Chemistry, Research Institute for Biomedical & Health Science, College of Biomedical and Health Science, Konkuk University, Chungju 27478, Republic of Korea
²Mushroom Research Division, National Institute of Horticultural and Herbal Science, Rural Development Administration, Eumseong 27709, Republic of Korea

Received: August 21, 2018 / Revised: August 24, 2018 / Accepted: August 30, 2018

Introduction

*Flammulina fennae* (*Physalacriaceae, Flammulina*) was first described by Bas in 1983 [1], but little is known about its biological and molecular characteristics. The biological and molecular properties of *F. fennae* were studied by Ripková [2]. According to previous reports, *F. fennae* showed an average SQ value (ratio of length and width of spores) of up to 1.75 and a nearly unbranched hymenial layer of the pileipellis [1, 2]. Additionally, Ripková et al. [2] demonstrated that the phylogenetic classification of *F. fennae* based on internal transcribed spacer (ITS) sequence was consistent with other morphological classifications. The study showed that *F. fennae* can be distinguished from other *Flammulina* species, based on a combination of morphological and molecular characteristics such as the SQ value, proportion of a certain type of terminal cells of ixohyphidia, and ITS sequence. Pérez-Butrón and Fernández-Vicente [3] considered that the habitat of *Flammulina* taxa is also important for identifying *Flammulina* species. They demonstrated that *F. fennae* is terrestrial or grows on roots, gregarious or scattered, and sometimes in clusters.

Basidiomycetes can efficiently degrade lignocellulosic biomass derived from plants because they contain vari-
ous lignocellulolytic enzymes [4]. Thus, basidiomycetes are frequently found in various natural environments such as woods, grasses, and crop wastes. Generally, wood-rotting fungi can be divided into two major groups, including brown rot and white rot fungi (90% of wood-rotting basidiomycetes), which degrade lignin (non-carbohydrate complex) and polysaccharides [5]. The ability of wood-decaying fungus to degrade and modify these lignocellulosic biomass is conferred by their carbohydrate-active enzymes (CAZymes). CAZymes are divided into glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), glycosyltransferases (GTs), auxiliary activities (AA), and carbohydrate-binding modules (CBMs), which increase the activity of carbohydrate enzymes [6–8]. These CAZymes are classified into several families, based on their catalytic activities towards substrates and amino acid sequence similarity [7] (http://www.cazy.org/).

CAZymes have received attention because of their great potential for various applications, such as those in the industrial and biotechnological fields to produce animal feed, food, paper, textile, and bioenergy [5]. Whole-genome sequencing of various fungi revealed a vast array of genes associated with lignocellulolytic enzymes [7, 9]. Understanding the genes associated with the biomass-degrading machinery in various fungi and their mechanisms is important for using these genes in various applications. Here, we first report the genome sequence and CAZyme repertoire of the white-rot fungus *F. fennae*. Determining the genome information and CAZyme repertoire is useful for understanding this fungus and facilitate its applications in the biotechnological and industrial fields.

**Materials and Methods**

**Genome Sequencing of *F. fennae***

*Flammulina fennae* KACC46185 was obtained from the National Agrobiodiversity Center (Rural Development Administration, Republic of Korea) and grown at 26°C on MCM agar (0.2% peptone, 2% glucose, 0.2% yeast extract, 0.05% MgSO₄, 0.046% KH₂PO₄, 0.1% K₂HPO₄, and 1.5% agar) for 2 weeks. Genomic DNA was extracted from *F. fennae* as described by Park et al. [10]. Next-generation sequencing of *F. fennae* was performed using the HiSeq 2000 platform according to the manufacturer’s protocol (Illumina, Inc.). The quality of the sequencing data was evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and further processed by Trimmomatic (version 0.32) [11] to detect poor-quality reads and conduct adapter trimming. The resulting short-reads were used for genome assembly with Velvet Optimiser [12]. *Ab initio* gene modeling was conducted by using the AUGUSTUS tool [13] trained in *L. bicolor*. The predicted genes sequences were compared to those in the National Center for Biotechnology Information (NCBI) non-redundant protein database by using both DIAMOND [14] and BLASTP (v 2.2.31) software. Furthermore, amino acid sequences of the predicted genes were compared with those in the protein family database (Pfam 31.0, http://pfam.xfam.org) by using Pfam-scan software [15]. tRNAscan-SE (version 2.0) software [16] was used to identify tRNAs in the *F. fennae* genome.

**Inferring Orthology**

Predicted genes in *F. fennae* were clustered into orthologous groups using OrthoFinder (version 2.2.1) software [17] by all-versus-all protein comparison with the following fungal species; *Aspergillus nidulans* FGSC-A4 [18], *Botrytis cinerea* B05.10 [19], *Agaricus bisporus* var. *bisporus* H97 [20], *Coprinopsis cinerea* okayama7#130 [21], *Cordyceps militaris* CM01 [22], *Cryptococcus neoformans* var. grubii H99 [23], *Flammulina velutipes* KACC42780 [10], *Laccaria bicolor* S238NH82 [24], *Lentinula edodes* [25], *Neurospora crassa* OR74A [26], *Phanerochaete chrysosporium* RP78 [27], *Saccharomyces cerevisiae* S288C [28], *Schizophyllum commune* H4-8 [29], *Trichoderma reesei* QM6a [30], and *Ustilago maydis* 521 [31].

**CAZyme Annotation and Signal Sequence Prediction**

CAZyme genes in *F. fennae* and other fungal species genomes were identified using HMMER 3.0 package software (http://hmmer.org/) with the dbCANCAZyme database (http://cslb.bmb.uga.edu/dbCAN/) [32] and obtained from CAZy database (http://www.cazy.org/) and DOE Joint Genome Institute (JGI Fungi Portal database; https://genome.jgi.doe.gov/programs/fungi/index.jsf). Additionally, prediction of signal peptides of the CAZyme genes was conducted using the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/) [33].
Results and Discussion

General Features of F. fennae Genome

Short reads (total 44,907,180 reads, 100-bp paired-end) derived from the Hiseq 2000 platform (Illumina, Inc., USA) were processed using the Trimmomatic tool for quality control and adapter trimming. The resulting short reads (41,905,226 reads, >Q30) were analyzed for de novo assembly, using the Velvet assembly tool with a kmer-size search range of 17–31. The resulting assembly (31 kmer) consisted of 11,953 sequence contigs with a total length of 32,423,623 bp (39% GC content) and N50 length of 52,210 bp. A total of 11,591 gene structures with an average gene length of 1,973 bp was predicted by ab initio gene prediction (Table 1).

Additionally, the average exon and intron lengths were 230.53 and 68.93 nucleotides, respectively. Of the 11,591 predicted genes, 86.9% (10,076) had sequence similarity (0.001 > e-value) to genes of documented proteins in NCBI-NR. A total of 9,678 (83.4%) of the predicted proteins shared sequence similarity to documented fungal sequences (Table S1). The total number of genes and genome size of F. fennae were comparable to those of its nearest sequenced species, F. velutipes [10]. The average exon size was also similar to those of other basidiomycetes.

Data Access

Sequence reads were deposited in the Sequence Read Archive (SRA) at NCBI under the following accession number: SRP151704.

Table 1. Flammulina fennae genome sequencing statistics.

<table>
<thead>
<tr>
<th>Data Access</th>
<th>Sequence reads were deposited in the Sequence Read Archive (SRA) at NCBI under the following accession number: SRP151704.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results and Discussion</td>
<td>General Features of F. fennae Genome</td>
</tr>
<tr>
<td>Short reads (total 44,907,180 reads, 100-bp paired-end) derived from the Hiseq 2000 platform (Illumina, Inc., USA) were processed using the Trimmomatic tool for quality control and adapter trimming. The resulting short reads (41,905,226 reads, &gt;Q30) were analyzed for de novo assembly, using the Velvet assembly tool with a kmer-size search range of 17–31. The resulting assembly (31 kmer) consisted of 11,953 sequence contigs with a total length of 32,423,623 bp (39% GC content) and N50 length of 52,210 bp. A total of 11,591 gene structures with an average gene length of 1,973 bp was predicted by ab initio gene prediction (Table 1). Additionally, the average exon and intron lengths were 230.53 and 68.93 nucleotides, respectively. Of the 11,591 predicted genes, 86.9% (10,076) had sequence similarity (0.001 &gt; e-value) to genes of documented proteins in NCBI-NR. A total of 9,678 (83.4%) of the predicted proteins shared sequence similarity to documented fungal sequences (Table S1). The total number of genes and genome size of F. fennae were comparable to those of its nearest sequenced species, F. velutipes [10]. The average exon size was also similar to those of other basidiomycetes.</td>
<td></td>
</tr>
<tr>
<td>Table 2. Genome comparison of Flammulina fennae and other basidiomycetes.</td>
<td></td>
</tr>
<tr>
<td>Fungal species</td>
<td>F. fennae</td>
</tr>
<tr>
<td>Strain</td>
<td>KACC46185</td>
</tr>
<tr>
<td>Genome assembly (Mb)</td>
<td>32.4</td>
</tr>
<tr>
<td>Number of protein-coding genes</td>
<td>11,591</td>
</tr>
<tr>
<td>GC contents (%)</td>
<td>39</td>
</tr>
<tr>
<td>Average gene length (bp)</td>
<td>1,980</td>
</tr>
<tr>
<td>Average exon size (bp)</td>
<td>230.53</td>
</tr>
<tr>
<td>Average intron size (bp)</td>
<td>68.93</td>
</tr>
</tbody>
</table>
mycetes with similar genome sizes (Table 2). Additionally, 256 tRNAs were identified by tRNAscan-SE [16] (Table S2) and 6,701 genes and 2,473 genes were annotated as functional proteins and multi-domain protein families, respectively, by Pfam 31.0 database searching (Table S3). As shown in Fig. 1A, *F. fennae* was classified into one group with *F. velutipes* by ortholog-based clustering analysis. Cluster analysis with other sequenced fungal species identified 6,715 (57.1%) groups containing at least one *F. fennae* protein (Table S4). Analysis of these clusters suggested that 10,667 genes (92%) of *F. fennae* proteins had orthologs amongst the Dikarya and were thus conserved in basidiomycetes and ascomycetes (Fig. 1B and Table S4). Among the set of homologous genes, 606 single-copy orthologs were detected. *Flammulina fennae* contained 23 species-specific genes, of which 16 were paralogous.

**CAZymes in *F. fennae* and Other Fungal Species**

Annotation of the predicted amino acid sequences of *F. fennae* genes against the dbCANCAZyme database revealed a series of genes associated with degradation and modification of carbohydrate complexes and lignin. The genome sequence of *F. fennae* contained 513 CAZymes, including 85 GTs, 220 GHs, 57 CEs, 45 CBMs, 20 PLs, and 86 AAs (Fig. 2, Table S5, and Table S6). In the present study, CAZymes of 15 other fungal species were annotated using the dbCAN software (http://csbl.bmb.uga.edu/dbCAN/) [32] or obtained from the CAZy database (http://www.cazy.org/) [7] and JGI Fungi Portal database (https://genome.jgi.doe.gov/programs/fungi/index.jsf).

**Glycosyltransferases (GTs)**

GTs (EC 2.4.-.-) are enzymes that catalyze formation
of the glycosidic linkage to form a glycoside and are involved in the biosynthesis of oligosaccharides, polysaccharides, and glycoconjugates [34, 35]. In the present study, 98 GTs were identified by amino acid comparison against the three different databases dbCAN, NCBI-NR, and pfam. Among them, 85, 66, and 46 genes predicted to encode GTs were identified by dbCANdbCAN, NCBI-NR, and pfam, respectively, in \textit{F. fennae} genome sequence (Fig. 3A, Table S1, and Table S3). Among the 98 GTs, 32 genes predicted to encode GT were commonly identified by all databases, while 18, 10, and 3 genes were uniquely identified by the dbCAN, NCBI-NR, and pfam databases searches, respectively (Fig. 3A and Table S7). In the CAZy database (http://www.cazy.org/), approximately half of the total number of GT families were GT2 and GT4. Indeed, complete genome sequences of various organisms, including archaea, bacteria, and eukaryotes, revealed a large number of GTs with approximately 1–2% of the gene products in the CAZy database (http://www.cazy.org/). Our results also revealed that the GT2 family was prominently present in the \textit{F. fennae} genome with 11 GT2 members (Fig. 4A and Table S7). It was previously reported that some of GT2 proteins act as chitin synthases [34]. Similarly, 3 genes predicted to encode GT2 family members were annotated as chitin synthases (EC 2.4.1.16) involved in fungal cell wall biosynthesis based on NCBI fungal genome databases searches (Table S1 and Table S7). To date, 105 GT families including more than 410,000 classified GT sequences have been annotated, of which more than 126,000 sequences from archaea, bacteria, eukaryota, and viruses were classified into the GT2 family in the CAZy database (http://www.cazy.org/) [7]. Genome-wide comparisons also revealed that GT2 is a major family in 12 fungal genomes, including 8 basidiomycetes and 4 ascomycetes species (Fig. 5A and Table S8). Most GTs are membrane proteins in the endoplasmic reticulum and Golgi apparatus. Therefore, they have a signal-anchor domain (16–20 amino acids), which acts as both a transmembrane region and uncleavable signal peptide [36, 37]. It has been reported that the difference between signal peptides and signal anchors is the length of the hydrophobic domain [38, 39]. Among the 98 GTs, 37 GTs were predicted as transmembrane proteins, of which 1 GT contained the signal peptide. Additionally, 6 GTs were further predicted to contain signal peptides within their amino acid sequences (Table S9). These results suggest that the predicted signal peptide sites in 7 genes are uncleavable and likely act as signal-anchor domains.

Although glycosyltransferases were classified into families based on amino acid sequence similarities [40, 41], functional prediction of a putative GT based on sequence homology is controversial. This is because although their amino acid sequence similarity is high,
their actual activities may differ. Breton et al. [34] reported that polyspecific GT families, including GT2 and GT4, showed sequence similarities within a small portion of the catalytic domain in amino acid sequence. However, sequence similarities have been observed in the entire catalytic domain of monospecific families.

Therefore, a new approach for identifying GTs is required because of the difficulty in identifying and classifying GTs based on sequence similarity. Recently, Mukai et al. [42] reported a computational method for determining the transmembrane region of Golgi-localized signal-anchor-type GTs. Further structural and mutational analyses are needed to precisely evaluate these enzyme characteristics.

**Glycoside hydrolases (GHs)**

GHs are common enzymes that catalyze the hydrolysis of glycosidic bonds of carbohydrates, such as cellulose, hemicellulose, and starch, the most abundant biomass in nature [43, 44]. A total of 220 GHs classified into 52 families were predicted in the *F. fennae* genome based on dbCAN database searching (Fig. 4F and Table S6). GH prediction revealed that GH16 was prominently present with 32 genes, while 17 families consisted of only one gene (Fig. 4F and Table S6). Furthermore, 134 and 162 GHs were identified by BLASTP (NCBI-NR) and protein family database (Pfam 31.0) searching, respectively (Table S1, Table S3, and Table S10) and 84 genes predicted to encode GHs were commonly identified among three different databases (Fig. 3B and Table S10). Among them, 39, 16, and 39 GHs were uniquely identified by dbCAN database, BLASTP (NCBI-NR), and protein family (Pfam 31.0 database) searches, respectively (Fig. 3B and Table S10). In genome comparisons, GH5, GH16, and GH18 families were also prominently present in other fungal species, except for in *S. cerevisiae* (Fig. 5C and Table S8).

It has been reported that most GH16 family enzymes, including agarase (EC 3.2.1.81), endo-beta-1,3-glucanase (EC 3.2.1.39), endo-beta-1,3,1,4-glucanase (EC 3.2.1.6), endo-beta-galactosidase (EC 3.2.1.103), lichenase (EC 3.2.1.73), xyloglucan xyloglucosyltransferase (EC 2.4.1.207), and kappa-carrageenanase (EC 3.2.1.83), contain the conserved motif EXDX(X)E in their amino acid sequences [45, 46]. It has also been revealed that both glutamic acid (E) residues at the first and last position of

http://dx.doi.org/10.4014/mbl.1808.08012
the motif are important for GH16 catalytic activity [47]. Similarly, our results revealed the EXDX(X)E motif within the amino acid sequences of all GH16 families in *F. fennae* (Fig. S1A).

Many GHs are secreted or targeted to other cellular locations such as the periplasmic space or Golgi body and thus possess a signal sequence. Our results also showed that 115 and 20 of the 220 GHs in *F. fennae* possessed signal peptides and were membrane proteins, respectively (Table S9). Approximately 33% of GHs have no signal sequence, suggesting that they are present inside the cell [33, 48]. Although further studies are needed, our results indicate that GHs without a signal sequence of *F. fennae* are located intracellularly. GHs are essential for the degradation of polysaccharides, such as cellulose and xylan from plants and chitin, which are important carbon and nitrogen sources in ecosystems [7]. Polysaccharides are degraded to short oligosaccharides by the endo-mode of action (endo-cellulase) and exo-mode of action (exo-cellulase) of GHs. β-Glucosidases (EC 3.2.1.21) convert cellobiose into glucose as members of GH families (GH1 and GH3). Therefore, most enzymes active towards polysaccharides belong to GH families (http://www.cazy.org/) [7, 49]. In nature, fungi play an important role in degrading cellulose, xylan, and chitin and therefore have potential for biotechnological applications such as in the food industry, animal feed, waste treatment, and other chemical industries [50–52]. CAZyme annotation revealed that genes associated with cellulase (GH5, -6, -7, -9, and -12), xylanase (GH10, -11, and -30), chitinases (GH18 and -85), and β-glucosidases (GH1 and -3) were present in the *F. fennae* genome (Fig. 4F, Table S6, and Table S10). These results suggest that *F. fennae* with more than 200 genes predicted to encode GHs have strong potential for various applications.

**Polysaccharide lyases (PLs)**

Polysaccharides are essential cellular components of all living organisms [53]. PLs are a group of enzymes (EC 4.2.2.-) that cleave polysaccharides in a β-elimination mechanism [54]. To date, PLs have been classified into 28 families based on recognizable sequence homologies in the CAZy database (http://www.cazy.org/). Our results revealed 20 PLs classified into 8 families in the *F. fennae* genome based on dbCAN database searching.
(Fig. 4B and Table S6). Among them, the PL1 and PL3 families were prominently present, each with 5 PLs in the *F. fennae* genome (Fig. 4B and Table S6). Additionally, 10 and 22 genes predicted to encode PLs were identified by BLASTP (NCBI-NR) and protein family database (Pfam 31.0) searching, respectively (Table S1 and Table S3), and 8 PLs were commonly identified in all three databases, including the dbCAN, NCBI-NR, and pfam databases (Fig. 3C and Table S11). Each 5, 2, and 10 of the 30 PLs were uniquely identified by dbCAN, NCBI-NR, and pfam databases searches, respectively (Fig. 3C and Table S11). Our results showed that several other PLs, including PL1, PL3, and PL14 families, were prominently present in some fungal species, although there is no PL14 family in ascomycetes (Fig. 5D and Table S8). Previous studies reported that PL11 and PL15 were only found in ascomycetes and basidiomycetes, respectively [9]. Our results also showed only *A. nidulans* and *C. cinerea* possessed PL families 11 and 15, respectively, and these families were not detected in other fungal species (Fig. 5D and Table S8). Furthermore, our results revealed that the PL14 family was prominently present in some basidiomycetes and is specific to the Basidiomycota (Fig. 5D and Table S8). The contents of pectate or pectin are lower than cellulose and hemicellulose compared to in plants [55, 56]. Pectate or

Fig. 5. Distribution of CAZymes in *F. fennae* and other fungal species. (A) GT families, (B) CBM families, (C) GH families, (D) PL families, (E) AA families, (F) CE families.
pectin lyases degrade smooth regions of polysaccharides such as polygalacturonan [36]. In the CAZy database, most pectin and pectate lyases are belonging to the PL1, -2, -3, -9, -10, and -22 families (http://www.cazy.org/). Among them, fungal-specific pectate lyases (EC 4.2.2.2 and EC 4.2.2.9) belong to PL1, PL3, and PL9 (http://www.cazy.org/). Similarly, the PL2, -10, and -22 families were not found in all fungal species evaluated in this study (Fig. 5D and Table S8).

Signal peptide prediction revealed 17 of 30 PLs predicted to contain signal sequences in their amino acid sequences. These results suggest that these PLs are secreted, except for one PL which was predicted to contain a transmembrane domain in its amino acid sequence (Table S9).

Although, in-depth studies are required for biochemical characterization of these enzymes, particularly in basidiomycetes, there is a great potential for using these enzymes in biotechnological applications because of the diverse ecological roles of basidiomycetes. In previous reports, S. commune was shown to contain abundant putative pectin-degrading lyases to enable high-levels of pectinase production [29, 57]. Moreover, S. commune produces high levels of polygalacturonase compared to that by A. niger in wheat bran [57]. Flammulina fennae showed a similar distribution of PLs as S. commune including PL families 1, 3, and 9, suggesting that this fungus is useful for biotechnological applications.

**Carbohydrate-binding modules (CBMs)**

CBMs, which are carbohydrate enzymes with carbohydrate-binding activities, increase the activity of carbohydrate enzymes [6, 8]. CBMs are commonly associated with several CAZymes such as GHs, PLs, and GTs [58]. Shoseyov et al. [8] reported that removal of the CBM from the scaffolding in cellulose, a non-covalent multi-protein complex possessing catalytic subunits, decreased the enzyme activity of the complex.

CBMs have been classified into 80 families based on amino acid sequence similarity in the CAZy database (http://www.cazy.org/). Our results revealed that 45 CBMs classified into 16 families were predicted in the F. fennae genome based on dbCAN database searching (Fig. 4D and Table S6). CBM family 1 was prominently present with 15 genes in the F. fennae genome (Fig. 4D and Table S6). Moreover, 15 and 20 CBMs were further identified by BLASTP (NCBI-NR) and protein family database (Pfam 31.0) searching, respectively (Table S1 and S3) and 2 CBMs were commonly identified in the three different databases (Fig. 3D and Table S12). In genome-wide comparison, the distribution of the CBM1, -13, and -50 families was similar to that in other some fungal species, although C. cinerea contained the largest number of CBM1 family members in its genome. According to a previous report by Zhao et al. [9], ascomycetes contained more CBM18 compared to in basidiomycetes. Our results also showed that ascomycetes contain more CBM18 family members than other basidiomycetes and have no CBM5 and CBM12 family members in their genomes (Fig. 5B and Table S8). Fernandez-Fueyo et al. [59] previously reported that the C. cinerea genome contains most of the CBM1 family members in its genome. This is also consistent with our results showing the largest number of CBM1 family members in C. cinerea identified by dbCAN database searching (Fig. 5B and Table S8).

CBMs are considered essential modules in cellobiohydrolases belonging to the GH6 and GH7 families [60]. Our results also revealed that GH6 and GH7 have CBM1 in their genes (Table S6). Our results also revealed that several CAZymes, including 17 GHs, 4 CEs, and 1 AA, contained CBMs in their genes, suggesting that these CAZymes require CBMs to degrade their substrates (Table S6).

**Carbohydrate esterases (CEs)**

CEs catalyze O-de- or N-deacylation to remove esters of substituted saccharides and are widely used in industrial and biotechnological processes [61–63]. CEs have been classified into 15 families in the CAZy database (http://www.cazy.org/). CEs act on various substrates, including acetic ester (acetyl esterases, EC 3.1.1.6), chitin (chitin deacetylases, EC 3.5.1.41), peptidoglycan (poly-N-acetylglucosamine deacetylases, EC 3.5.1.104), feruloyl-polysaccharide (feruloyl esterases, EC 3.1.1.73), pectin (pectinesterase, EC 3.1.1.11), and xylan (acetylxylanesterases, EC 3.1.1.72), among others [64].

In this study, 57 CEs classified into 10 families were predicted in the F. fennae genome by dbCAN database searching (Fig. 4C and Table S6). CE families 1 and 4 were prominently present, each showing 16 genes in the F. fennae genome (Fig. 4C). Moreover, 24 and 44 CEs...
were further identified by BLASTP (NCBI-NR) and protein family database (Pfam 31.0) searching, respectively (Table S1 and S3), and 13 CEs were commonly identified in the three different databases (Fig. 3E and Table S13). Among the total predicted 78 CEs, 24, 5, and 15 CEs were uniquely identified by dbCAN, NCBI-NR, and pfam databases searches, respectively (Fig. 3E and Table S13). Our results revealed that the total number of CEs in *F. fennae* was similar with those in some basidiomycetes and CE families 1, 4, and 16 were highly present in some basidiomycetes such as *F. velutipes*, *C. cinerea*, *L. edodes*, *S. commune*, and *U. maydis* (Fig. 5F and Table S8). Although many CE10 family members were identified in *F. fennae* (data not shown), they are not described in this study, as this family also acts on non-carbohydrate substrates (http://www.cazy.org/).

Among the 57 CEs, approximately half were predicted to possess signal sequences within their amino acid sequences, suggesting that these CEs are secreted (Table S9).

Among the CE families, the CE1, CE4, CE5, and CE7 families have been characterized as containing the SHD (Ser-His-Asp) catalytic triad and GXSXG (Gly-Xaa-Ser-Xaa-Gly) conserved motif [65]. Some these CE families were found to possess the GXSXG conserved motif in their amino acid sequences (Fig. S1B and Table S6). It has been reported that CE families remove acylated moieties from polysaccharides, enabling GHs to access polysaccharides which facilitates saccharification of these polymers [66]. The vast array of genes associated with CE families in the *F. fennae* suggests that this fungus is useful for biotechnological and industrial applications such as in food and biofuel production.

**Auxiliary activities (AAs)**

Recently, the GH61 and CBM33 families were reclassified into a new category, AA families, because they were found to be lytic polysaccharide monoxygenases [67]. AA families including lytic polysaccharide monoxygenases are involved in the depolymerization of lignin (non-carbohydrate structural component) in plants [5]. AAs are classified into 15 families, including 8 families of ligninolytic enzymes and 3 families of lytic polysaccharide monoxygenases, based on amino acid sequence similarities (http://www.cazy.org/).

CAZyme prediction of *F. fennae* revealed that 86 AAs classified into 11 families were identified by dbCAN database searching (Fig. 4E and Table S6). *Flammulina fennae* contained a vast array of genes predicted to encode AA3 (GMC oxidoreductase), AA7 (glucooligosaccharide oxidase), and AA9 (lytic polysaccharide monoxygenase; GH61) family members in its genome (Fig. 4E and Table S6). Additionally, 52 and 101 AAs were identified by BLASTP (NCBI-NR) and protein family database (Pfam 31.0) searching, respectively, and 33 AAs were commonly identified in the three databases (Fig. 3F and Table S14).

Basidiomycetes can degrade lignocellulosic biomass derived from plants such as hardwoods, softwoods, and forest waste [4]. These fungi are also known as wood-rotting basidiomycetes, which are largely divided into brown rot and white rot fungi. Particularly, white rot fungi account for more than 90% of wood-rotting basidiomycetes and can efficiently degrade both lignin and polysaccharides [5]. Thus, characterization of the genes involved in the lignocellulosic biomass degradation is important for understanding this fungus for appropriate application. A vast array of genes associated with lignocellulosic biomass degradation have been identified in several sequenced basidiomycetes. It has been reported that *C. neoformans* (xerophilic mold-like basidiomycete) contains a very limited pattern of genes coding for polysaccharide degradation enzymes [5]. Additionally, *U. maydis* (biotrophic plant pathogenic fungus) was found to contain a minimal set of AAs for plant defense responses [31]. In *L. bicolor*, the most abundant CAZymes are involved in modification of polysaccharide backbones, such as plant cell wall degradation [5]. Although other fungal species contained relatively low numbers of AAs in their genomes, the repertoire of genes predicted to encode AAs in *F. fennae* were similar to those in other white rot fungus, such as *L. edodes* (white rot), *F. velutipes* (white rot), and *S. commune* (white rot-like) (Fig. 5D and Table S8).

It was previously reported that some AA families, including AA1 and AA3, possess conserved motifs. The AA1 family (laccase; EC 1.10.3.2) contains copper binding motifs, His-Xaa-His-Gly, His-Xaa-Xaa-His-Xaa-His, and His-Cys-His-Xaa(3)-His-Xaa(4)-Met/Leu/Phe, within its amino acid sequence [68]. Similarly, 3 AA1 families in *F. fennae* contain these conserved motifs in amino acid sequences, suggesting the genes predicted to encode
AA1 families act on lignin (Fig. S1C and Table S6). Moreover, the AA3 family (glucose-methanol-choline oxidoreductase) contains the conserved motif Gly-Xaa-Gly-Xaa-Gly-Xaa-Gly-Xaa[18]. Glu, which interacts with flavin adenine dinucleotide cofactor [69–71]. Among the 25 AA3 family members in F. fennae, 17 contain the conserved motifs, and thus are involved in the oxidative degradation of lignocellulosic biomass (Fig. S1D and Table S6).

The method for producing bioethanol from plants is well-understood; however, a major limitation in bioethanol production is the difficulty in decomposing the large amount of woody matter (particularly lignin) present in plant cell walls. Several reports have suggested methods for overcoming this problem by using enzymes produced by microorganisms for recalcitrant lignin matrix degradation [72, 73]. In general, wood degradation by white rot fungus leads to the generation of highly reactive lignin radicals, starting with the depolymerization of lignin, which leads to a chain reaction of further degradation of the neighboring wood polymer [74, 75]. From this perspective, the various AA families of F. fennae have strong potential for use in the production of various biomaterials and bioenergy.

**Acknowledgement**

This research was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2016R1D1A1B03930808).

**Conflict of Interest**

The authors have no financial conflicts of interest to declare.

**References**

Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biol.* **12**: R116.


http://dx.doi.org/10.4014/mbl.1808.08012


