Metagenomic Analysis of Novel Lignocellulose-Degrading Enzymes from Higher Termite Guts Inhabiting Microbes

Nimchua, Thidarat*, Taksawan Thongaram, Tanaporn Uengwetwanit, Somchai Pongpattanakitshote, and Lily Eurwilaichitr

Enzyme Technology Laboratory, Bioresources Technology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand Science Park, Pathumthani 12120, Thailand

Received: August 17, 2011 / Revised: December 4, 2011 / Accepted: December 5, 2011

A metagenomic fosmid library was constructed from genomic DNA isolated from the microbial community residing in hindguts of a wood-feeding higher termite (Microcerotermes sp.) collected in Thailand. The library was screened for clones expressing lignocellulolytic activities. Fourteen independent active clones (2 cellulases and 12 xylanases) were obtained by functional screening at pH 10.0. Analysis of shotgun-cloning and pyrosequencing data revealed six ORFs, which shared less than 59% identity and 73% similarity of their amino acid sequences with known cellulases and xylanases. Conserved domain analysis of these ORFs revealed a cellulase belonging to the glycoside hydrolase family 5, whereas the other five xylanases showed significant identity to diverse families including families 8, 10, and 11. Interestingly, one fosmid clone was isolated carrying three contiguous xylanase genes that may comprise a xylanosome operon. The enzymes with the highest activities at alkaline pH from the initial activity screening were characterized biochemically. These enzymes showed a broad range of enzyme activities from pH 5.0 to 10.0, with pH optimal of 8.0 retaining more than 70% of their respective activities at pH 9.0. The optimal temperatures of these enzymes ranged from 50°C to 55°C. This study provides evidence for the diversity and function of lignocellulose-degrading enzymes in the termite gut microbial community, which could be of potential use for industrial processes such as pulp biobleaching and denim biostoning.

Keywords: Alkaline cellulase, alkaline xylanase, metagenomics, termite gut, uncultured bacteria

Cellulose and xylan are the principal structural components of plant and represent the major sources of renewable biomass in the biosphere. Cellulose is the most abundant biopolymer on Earth and can be degraded by cellulases, a group of enzymes comprising endo-β-glucanases (E.C. 3.2.1.4), exo-β-glucanases including cellobiohydrolases (E.C. 3.2.1.74), and β-glucosidases (E.C. 3.2.1.91) [23]. Xylan is the second most abundant biopolymer, which can be hydrolyzed to xylo-oligosaccharides and xylose residues by endo-1,4-β-D-xylanase (E.C. 3.2.1.8) [29] and β-xylosidase (E.C. 3.2.1.37) [31]. Cellulases and xylanases are used in many industries. In the pulp and paper industry, the pulping and bleaching processes are usually conducted under alkaline conditions [12, 37]. Likewise, alkaline conditions are preferred for the biopolishing process during cotton garment preparation, since fabric strength is maintained and backstaining is less frequent [1].

Termites play pivotal roles in the decomposition of lignocellulose, with the aid of their associated microbial symbionts [18]. The guts of higher termites (Family Termitidae) harbor a great variety of intestinal microbiota including bacteria, archaea, and eukaryotes [28], but typically lacking eukaryotic flagellated protists [36]. The prokaryote termite gut flora converts plant polysaccharides, including celluloses and hemicelluloses, into energy sources that can be utilized by the insect host. The hindgut of Termitidae is highly compartmentalized and comprises five main sections including the first proctodeal segment (P1). The contents of the P1 segment are strongly alkaline (pH>9.0) [4], and thus the P1 segment is a likely habitat for alkaliphilic microorganisms capable of degrading lignocellulosic substances at high pH. Many genes encoding lignocellulose-degrading enzymes have been characterized from microorganisms of termite guts, but most of them were isolated from pure cultured microbes [30, 39]. 16S rRNA phylogenetic profiling of higher termite gut samples has revealed highly diverse families of bacteria including several species deemed unculturable [6, 18, 36]. Therefore,
the uncultured bacteria in the termite gut are a potential source for the discovery of novel genes encoding lignocellulose-hydrolyzing enzymes.

Culture-independent strategies including metagenomic approaches have been developed to isolate microbial DNA directly from several environmental samples, providing a comprehensive insight into microbial diversity, including uncultured microbes. Novel genes originating from uncultured microbes encoding enzymes with unique biochemical and biophysical characteristics have been discovered by culture-independent approaches [8]. Several metagenome-derived genes encoding cellulases or xylanases have been identified in metagenomic libraries generated from various environmental samples including termite guts [34–36].

Despite many cellulase and xylanase genes having been isolated and characterized from different environmental genomic libraries [13, 17, 22], the biotechnological potential of novel lignocellulose-degrading genes isolated from microbial communities in termite gut is still far from being fully explored. Thus, we sought novel cellulase and xylanase genes from termite gut microorganisms. Direct cloning and activity-based screening of genomic DNA isolated from this environmental sample were performed, and several novel genes were isolated and analyzed. Novel cellulases and xylanases active in alkaline conditions were discovered with potential for biocatalysis applications.

**Materials and Methods**

**Samples, Bacterial Strains, Plasmids, and Culture Conditions**

Termitidae higher termites, *Microcerotermes* sp., were collected from a colony nest in Pathumthani Province, Thailand. *Escherichia coli* strains EP1300T1 and TOP10 were used for DNA cloning. The vector pC11FOS was purchased from Epicentre (USA) and the cloning vector pZErO-2 was purchased from Invitrogen (USA). *E. coli* was grown at 37°C in lysogeny broth (LB) or on agar [1.5% (w/v)] plates supplemented with appropriate antibiotics.

**Isolation of Total Genomic DNA and Construction of Metagenomic Fosmid Library**

Intestinal tracts of the termites were carefully dissected and the P1 segment was isolated with sterile fine-tipped forceps. Genomic DNA was extracted from a pool of 100 gut sections using an ISOPLANT II kit (Wako Nippon Gene, Japan) according to the manufacturer’s instruction. Subsequently, the extracted gDNA was purified with a DNeasy Tissue kit (QIAGEN, USA) according to the manufacturer’s instructions. This purified DNA was then used for preparation of a metagenomic DNA fosmid library. The metagenomic fosmid library was constructed using the CopyControl fosmid library production kit (Epicentre, USA) following the manufacturer’s instructions [7].

**Screening for Cellulase and Xylanase Activities**

The transformants were spread on LB plates supplemented with 12.5 µg/ml of chloramphenicol and 0.001% (w/v) arabinose and incubated overnight at 37°C. Each plate was then overlaid with 0.7% (w/v) agar supplemented with ENZhance cell permeabilizing reagent (pH 10.0) and suitable substrates (0.05% AZCL-xylan or 0.05% AZCL-HE cellulose purchased from Megazyme, Ireland) [7]. The reaction was further incubated at 37°C for 2 h. Blue color released from AZCL-substrate around colonies indicated positive clones possessing putative target enzymes. In order to differentiate the activity-positive clones, recombinant fosmids were extracted with a GeneJET Plasmid Miniprep kit (Fermentas, USA) and subsequently digested with NotI (Fermentas, USA). DNA patterns were visualized by agarose gel electrophoresis and ethidium bromide staining. Representative fosmids with different banding patterns (CL1, Xyn12, and Xyn9) were selected for further analysis.

**Subcloning of Cellulase- and Xylanase-Positive Clones**

Fosmid DNAs from clones CL1, Xyn12, and Xyn9 were partially digested with BspI1431i (Fermentas, USA). The DNA fragments larger than 1 kb were purified using a Wizard DNA Clean-Up system (Promega, USA) and ligated into the pZErO-2 cloning vector digested with BamHI. The shotgun library was transformed into *E. coli* TOP10 cells by electroporation at 25 µF, 600 Ω, and 2.0 kV using a Gene Pulser apparatus (BioRad, USA) and then incubated at 37°C for 1 h. The shotgun library was then screened for cellulase and xylanase activities on LB supplemented with 50 µg/ml kanamycin (LBkam) using ENZhance reagent and 0.05% AZCL-HE cellulose or AZCL-xylan [7]. Clones exhibited blue-color zones, indicative of corresponding cellulases and xylanases, were selected for DNA sequencing.

**DNA Sequencing Analysis**

DNA sequences of all fosmid-positive clones were obtained by pyrosequencing (Genome Institute, BIOTEC, Thailand), whereas sequencing of plasmids containing DNA subcloned from the metagenomic library clones was performed by standard dideoxy automated procedures (1st BASE, Malaysia). Subsequently, the DNA sequences obtained from both pyrosequencing and subcloning approaches were analyzed with several bioinformatic programs. Candidate cellulases and xylanases were identified by homology to related sequences in the database using BLASTx and BLASTp, available from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/BLAST). Multiple alignments of enzymes with homologous proteins were performed using the Align-X program (InfoMax, USA). Phylogenetic trees were constructed by the neighbor-joining method (MEGA 4.0 Software) with 1,000 bootstrap repetitions. The nucleotide sequences reported here have been submitted to the NCBI/GenBank database under accession numbers GU721061, GU721062, GU721063, and GU721064.

**Enzyme Assays**

*E. coli* transformants harboring activity-positive plasmids from the shotgun library were grown in 200 ml of LBkam at 37°C overnight. The cells were collected, resuspended in 20 ml of 100 mM potassium phosphate, pH 7.0, and lysed by sonication. Crude cell lysates were centrifuged and the clear supernatants were harvested as crude protein extracts for subsequent enzyme activity assays. Cellulolytic and xylanolytic activities were determined by measuring the reducing sugar liberated from carboxymethyl cellulose (CMC) and birchwood xylan, respectively, using the dinitrosalicylic acid (DNS) method [25]. One unit (U) was defined as the quantity of enzyme required
to liberate 1 µmole of glucose or xylose equivalent per minute at 50°C.

Effects of pH and Temperature on Enzyme Activities
To determine the optimum pH of the selected enzymes, cellulase or xylanase activity was measured at the pH range of 4.0 to 12.0 under standard conditions at 50°C. The buffers used were 100 mM sodium acetate (pH 4.0–6.0), 100 mM potassium phosphate (pH 6.0–8.0), 100 mM Tris-HCl (pH 8.0–9.0), and 100 mM glycine-NaOH (pH 9.0–11.0). The optimal temperature for enzyme activity was determined by subjecting reaction mixtures to different temperatures ranging from 25°C to 75°C at optimal pH. The relative activity was calculated by comparing the activity from each treatment with that of the maximal activity measured for each enzyme, which was considered as 100%.

RESULTS AND DISCUSSION
Construction of a Metagenomic Library from Termite Guts and Screening for Lignocellulolytic Enzymes
Total DNA from 100 termite gut samples were extracted with a final yield of approximately 2.6 µg of high-quality purified DNA. A fosmid library was then constructed from the purified genomic DNA with sizes ranging from 20 to 40 kb in length. The library consisted of approximately 88,000 clones. Restriction analysis with NotI of 30 randomly picked clones displayed unique patterns, indicating a high complexity and representative coverage of the termite gut biota (data not shown). The insert DNA fragments of these clones varied from 20 to 31 kb, with an average size of 25 kb, indicating the size of the library was approximately 2.2 Gb.

In order to discover fosmids harboring cellulase- and xylanase-encoding genes, the fosmid library was screened for cellulolytic and xylanolytic activities by colorimetric assay. Five colonies were found to exhibit cellulase activity, whereas 24 demonstrated xylanase activity as shown by blue-color zones around the colonies on agar plates containing suitable substrates at pH 10.0 (data not shown).

To differentiate the cellulase- and xylanase-positive clones, NotI restriction analysis was performed revealing two distinct band patterns among the cellulase-expressing clones and 12 patterns among the xylanase-expressing clones, suggesting that some of the clones recovered were repetitive with others. Fosmids representative of each of the NotI restriction patterns were pooled and subjected to pyrosequencing. After contigs assembly, five complete ORFs including one cellulase (c0001) and four xylanase (x1088, x1098.1, x1098.2, and x1098.3) genes were obtained. To complement the pyrosequencing, a shotgun library of the three fosmids showing the highest cellulase (CL1) and

<table>
<thead>
<tr>
<th>Fosmid clone</th>
<th>Predicted ORF</th>
<th>Protein length (aa)</th>
<th>Most homologous protein</th>
<th>Organism</th>
<th>Identity/similarity (%)</th>
<th>GH/conserved motif (range)/catalytic residues</th>
<th>GenBank Accession No.</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL1</td>
<td>c0001</td>
<td>404</td>
<td>GH5 (ABW39333) from metagenomes of termite gut</td>
<td>Uncultured bacterium</td>
<td>58/71</td>
<td>Cellulase/GH5 (82-372)/E205 and E338</td>
<td>GU721061</td>
<td>Warnecke et al. [36]</td>
</tr>
<tr>
<td>Xyn5</td>
<td>x1088</td>
<td>651</td>
<td>Endo-1,4-β-xylanase (AAS85781) from metagenomes of adult lepidopteran moth</td>
<td>Uncultured bacterium</td>
<td>50/65</td>
<td>Xylanase/GH8 (8-405)/E44 and D276</td>
<td>GU721062</td>
<td>Brennan et al. [5]</td>
</tr>
<tr>
<td>Xyn12</td>
<td>x0012</td>
<td>273</td>
<td>Endo-1,4-β-xylanase (AAS85784) from metagenomes of adult lepidopteran moth</td>
<td>Uncultured bacterium</td>
<td>56/73</td>
<td>Xylanase/GH11 (42-250)/E142 and E244</td>
<td>GU721063</td>
<td>Brennan et al. [5]</td>
</tr>
<tr>
<td>Xyn9</td>
<td>x1098.1</td>
<td>700</td>
<td>GH10 (ABN52146) C. thermocellum ATCC27405</td>
<td>40/58</td>
<td>Xylanase/GH10 (38-378)/E184 and E306</td>
<td>GU721064</td>
<td>Unpublished</td>
<td></td>
</tr>
<tr>
<td>Xyn9</td>
<td>x1098.2</td>
<td>664</td>
<td>Xylanase (AAS85784) from metagenomes of adult lepidopteran moth</td>
<td>Uncultured bacterium</td>
<td>49/66</td>
<td>Xylanase/GH11 (55-260)/E147 and E253</td>
<td>GU721064</td>
<td>Brennan et al. [5]</td>
</tr>
<tr>
<td>Xyn9</td>
<td>x1098.3</td>
<td>529</td>
<td>Xylanase (AAS85783) from metagenomes of caterpillars</td>
<td>Uncultured bacterium</td>
<td>59/73</td>
<td>Xylanase/GH11 (32-244)/E145 and E242</td>
<td>GU721064</td>
<td>Brennan et al. [5]</td>
</tr>
</tbody>
</table>
xylanase (Xyn12 and Xyn9) activities from colorimetric assay was constructed and screened in order to identify genes encoding the target enzymes. One plasmid with cellulase activity was isolated that had the same cellulase-encoding gene as c0001 found by pyrosequencing. Two plasmids with xylanase-encoding genes were isolated, one which was the same as x1098.3 found by pyrosequencing and another gene designated x0012. Together, one putative cellulase and five xylanase genes were obtained from both cloning approaches.

Amino acid sequence analysis of the six complete ORFs revealed relatively low sequence identity (40% to 59% pairwise amino acid sequence identities) to known cellulases and xylanases in the NCBI database (Table 1). These genes were most closely related to those from uncultured microbes and anaerobic bacteria (Table 1). Moreover, five of the six putative proteins identified were most closely related to enzymes identified from metagenomes of herbivorous insect guts including adult (X1088, X1098.2, and X0012) and caterpillar (X1098.3) of lepidopteran moths (Rotschildia lebæu) [5] and previously reported Nasutitermes sp. termites (C0001) [36]. In addition to uncleared bacteria, all glycosyl hydrolases (GH) found in this study were also related to groups of enzymes produced by anaerobic bacteria inhabited in a closed ecosystem such as cattle rumen (F. succinogenes) [33] and insect gut (Clostridium spp.) [5, 36]. The similarity of the genes isolated from the Microcerotermes sp. gut to genes from other anaerobic habitats is consistent with earlier reports that termite guts exclusively sustain an intrinsic anaerobic habitat analogous to the cattle rumen, widely known as a reservoir for anaerobic microbes [6, 15].

The six protein sequences were compared with known cellulases and xylanases for phylogenetic classification (Fig. 1). The termite gut microbial proteins belong to various GH families; C0001, X1088, and X1098.1 were affiliated with families 5, 8, and 10, respectively, whereas X0012, X1098.2, and X1098.3 were closely related to family 11. Multiple protein sequence alignments for these six ORFs showed signature motifs and catalytic residues of cellulase and xylanases specifically found in known GH families (Fig. 2) [5, 11, 21]. Moreover, the novel xylanases (X1088, X1098.1, X1098.2, and X1098.3) each possess catalytic modules with C-terminal domains, including

**Fig. 1.** Unrooted amino acid phylogenetic tree. Only bootstrap values higher than 50% are shown. The scale bar indicates 0.2 changes per amino acid site.
serine-rich domains, carbohydrate binding domain (CBM), and domains of unknown function (Fig. 3). It is widely recognized that modular glycosyl hydrolases can harbor non-catalytic motifs such as CBM that promote the association of the enzyme with the substrate [16], thermostabilizing domains that mediate enzyme thermostability [32], S-
layer-like domains that promote binding to the cell envelop [24], and dockerin domains whose function is to assist binding of the enzyme to cohesion domains within the scaffolding protein [3]. The C-terminal regions of X1098.1 and X1098.2 possess family CBM4 domains. This binding domain has been found frequently in bacterial endoglucanases and also in several other glycosyl hydrolases [27], including those from ruminal bacteria such as xylanases from \textit{F. succinogenes} [26] and cellulases from \textit{C. thermocellum} [20]. Two of the predicted amino acid sequences (X1088 and X1098.3) contained a C-terminal domain that showed no detectable homology to previously identified domains by BLAST analysis (Fig. 3). The lack of similarity between the C-terminal domain of X1088 and X1098.3 to other protein sequences in the NCBI database suggests that this xylanase has unique characteristics not previously attributed to known modular xylanases.

Three xylanase genes (x1098.1, x1098.2, and x1098.3) were identified in the \textit{Microcerotermes} sp. metagenome, which are contiguous, and thus may constitute an operon (Fig. 3). Operons encoding cellulases and xylanases have been discovered in anaerobic bacteria such as \textit{C. thermocellum} and \textit{Streptomyces olivaceovirides}. The encoded enzymes form protein complexes to promote synergistic degradation of cellulose and hemicellulose biomass [19, 23]. Cellulosomes are complexes of cellulolytic enzymes with high activity.

**Fig. 4.** Effect of pH on cellulase or xylanase activities of X0012 (A), X1098.3 (B), and C0001(C).
The enzymatic assay was performed at pHs ranging from 4.0 to 11.0 under standard assay conditions. The buffers used were 100 mM sodium acetate (pH 4.0–6.0) (◆), 100 mM potassium phosphate (pH 6.0–8.0) (□), 100 mM Tris-HCl (pH 8.0–9.0) (▲), and 100 mM glycine-NaOH (pH 9.0–11.0) (●). Effect of temperature on activities of X0012 (D), X1098.3 (E), and C0001 (F). The optimal temperature for enzyme activity was carried out by assaying enzymatic reactions at temperatures ranging from 25°C to 75°C for 10 min at the optimal pH. Carboxymethyl cellulose (CMC) and birchwood xylan were used as substrates for cellulase and xylanase, respectively.
against crystalline cellulose and have been characterized in several anaerobic bacteria and fungi [2]. Xylanosomes are multienzyme complexes that play an important role in the degradation of hemicelluloses, mainly xylan [19]. The C-terminal regions of the X1098.1, X1098.2, and X1098.3 xylanases contain serine repeat linker domains, which may serve as dockerins. Dockerins are non-catalytic flexible linker sequences rich in hydroxyamino acids separating catalytic domains that are found in polysaccharide-hydrolyzing enzymes organized into multienzyme complexes [14].

Characterization of the Cloned Cellulases and Xylanases with the Cytoplasmic Extracts of the Active Subclones

The cellulase C0001 and the xylanase enzymes X1098.3 and X0012, showing the highest cellulase and xylanase activities from the previous colorimetric assay, were investigated for their ability to degrade cellulose or hemicellulose at defined ranges of pH and temperature using CMC or birchwood xylan as substrates. These enzymes showed optimal activity at pH 8.0 (Fig. 4A–C). Moreover, X0012 retained more than 90% of its maximal activity at pH 9.0, whereas more than 70% of maximal activity remained at the same pH for C0001 and X1098.3 (Fig. 4A–C). The optimal temperature for C0001 was 50°C, whereas X0012 and X1098.3 were optimal at 55°C (Fig. 4D–F). It is well known that midgut contents of certain insect larvae, particularly among the Lepidoptera, Coleoptera, and Diptera, can be markedly alkaline [9, 10, 39]. Likewise, the intestinal pH of higher termite guts in family Termitidae is strongly alkaline (pH>9.0) in the P1 segment of the hindgut [4]. Therefore, it is quite reasonable to assume that termite guts would be a resource for alkaliphilic lignocelluloses-degrading enzymes. In this study, it is evident that the cellulose and xylanase enzymes exhibited optimal activities at alkaline pH. Nonetheless, they were active over a broad pH range (5.0–10.0) under standard assay conditions (Fig. 4A–C).

This work has demonstrated the identification of novel genes encoding lignocellulose-degrading enzymes obtained from termite gut using a metagenomic approach. According to phylogenetic inference, diverse groups of glycosyl hydrolases can be identified from this source, indicating the high diversity of this environment. Specifically, we have found that multiple glycosyl hydrolytic enzymes from different families were closely related to those of ruminal anaerobic bacteria. As different GH genes were discovered either as a single gene or as multiple genes within an operon, the corresponding proteins derived from genomes of the gut bacteria presumably acted synergistically to hydrolyze complex plant polysaccharides. This study provides evidence that the metagenomic approach is proved to be a potential technique to explore a variety of target enzymes from extreme environmental samples including animal intestinal tracts. Further study on the overexpression of the isolated enzymes in a suitable expression system and examination of their biochemical characteristics will be undertaken. It is expected that comprehensive experiments will provide important data for future application of these novel cellulase and xylanases for promising biotechnological processes.

Acknowledgments

The authors would like to thank Dr. Philip Shaw and Dr. Piyunarn Harnpicharnchai for critical editing of the manuscript and Dr. Sithichoke Tangphatsornruang for assistance with pyrosequencing analysis. This study was supported by a post-doctoral fellowship program and a research grant (P-00-20084) from the National Center for Genetic Engineering and Biotechnology (BIOTEC, Pathumthani, Thailand).

References