Acta Biochim Biophys Sin, 2015, 47(9), 741–748
doi: 10.1093/abbs/gmv070
Short Communication

Short Communication

Endogenous cellulolytic enzyme systems in the longhorn beetle *Mesosa myops* (Insecta: Coleoptera) studied by transcriptomic analysis

Jie Liu[1,†], Keqing Song[1,2,†], Huajing Teng[2,3], Bin Zhang[1,2], Wenzhu Li[1], Huaijun Xue[1,*], and Xingke Yang[1,*]

[1]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China, [2]University of Chinese Academy of Sciences, Beijing 100049, China, and [3]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

[†]These authors contributed equally to this work.
*Correspondence address. Tel: +86-10-64807237; Fax: +86-10-64807099; E-mail: yangxk@ioz.ac.cn (X.Y.)/xue@ioz.ac.cn (H.X.)

Received 12 November 2014; Accepted 22 April 2015

Abstract

The Cerambycidae (longhorn beetle) is a large family of Coleoptera with xylophagous feeding habits. Cellulose digestion plays an important role in these wood-feeding insects. In this study, transcriptomic technology was used to obtain one glycoside hydrolase family 45 (GH45) cellulase and seven GH5 cellulases from *Mesosa myops*, a typical longhorn beetle. Analyses of expression dynamics and evolutionary relationships provided a complete description of the cellulolytic system. The expression dynamics related to individual development indicated that endogenous GH45 and GH5 cellulases dominate cellulose digestion in *M. myops*. Evolutionary analyses suggested that GH45 cellulase gene is a general gene in the Coleoptera Suborder Polyphaga. Evolutionary analyses also indicated that the GH5 cellulase group in Lamiinae longhorn beetles is closely associated with wood feeding. This study demonstrated that there is a complex endogenous cellulolytic system in *M. myops* that is dominated by cellulases belonging to two glycoside hydrolase families.

Key words: cellulase, transcriptome, longhorn beetle, wood-feeding habit

Introduction

Cellulose is polymer comprising β(1,4)-linked D-glucose units. It is a stable compound and a key structural component of plants. Cellulose usually combines with lignin and other amorphous materials (e.g. hemicellulose, pectin) which makes hydrolysis difficult [1]. Most plant-feeding animals do not utilize cellulose with the exception of those with specialized feeding habits such as wood-feeding insects and ruminants [2]. In the previous studies, plant-feeding insects such as termites and longhorn beetles were found to have the ability of utilizing cellulose as their carbon source [3].

The enzyme cellulase plays a central role in the utilization of cellulose. The degradation of cellulose requires three kinds of related enzymes including endo-β-1,4-glucanase (EC 3.2.1.4), exo-β-1,4-glucanase (EC 3.2.1.91 and EC 3.2.1.74), and β-glucosidases (EC 3.2.1.21) [4]. Endo-β-1,4-glucanase, also called endocellullase, is a relatively common enzyme in animals [4] and it is the focus of this study. Exo-β-1,4-glucanase, also called exocellulase, is rarely reported in insects [4,5]. β-Glucosidase, also called cellobiase, is a widespread enzyme in all the current living animals, which means that it does not limit the use of cellulose by insects [4,6].

Many wood-feeding insects such as longhorn beetles and termites can utilize lignocellulose because of cellulases in their digestive tract [4]. Biological sources of cellulases can contribute to cellulose digestion in insects. Termites are a focal point of such research due to their special biological characteristics and economic significance. Many cellulase genes have been identified from termites and their hindgut endosymbionts [7]. A dual cellulolytic system from termites depends on both endogenous and exogenous cellulases [8,9].

The endogenous cellulolytic system is generated by the cellulase genes of the termite genome, and the cellulases of the exogenous cellulolytic system are encoded by digestive track endosymbionts [4]. Several cellulases have also been identified from beetle species including longhorn beetles (Cerambycidae) and leaf beetles (Chrysomelidae) [10,11].

Longhorn beetles, Cerambycidae, represent a very large family in the Coleoptera with a characteristic larval wood-feeding habit. The larvae of some longhorn beetles are serious pests that cause economic damage when born into tree trunks [12]. The digestion mechanisms of cellulose in longhorn beetles are not as well studied as in termites and only limited cellulase genes have been cloned or identified from longhorn beetle species.

Our understanding of the ability of animals to utilize cellulose has evolved over time. Early studies of termites revealed that cellulose digestion in insects depends on cellulases generated by intestinal symbionts [13,14]. Many endogenous cellulase genes in termites, beetles, and other insects have recently been reported. Endogenous cellulase encoded by the termite genome has proven the existence of endogenous cellulolytic systems in insects [15,16]. The co-existence of endogenous and exogenous cellulases occurs in the termite *Coptotermes formosanus*, indicating that some species have a dual cellulolytic system [8,9,17,18]. There are two types of cellulase expressed in different parts of the digestion tract. The exogenous cellulolytic system functions in the hindgut while the endogenous system works in the salivary gland and the midgut [17]. The hindgut in many termite species is significantly enlarged with a specialized shape and it typically houses a large number of endosymbionts including protists, fungi, and bacteria [19]. These hindgut endosymbionts provide exogenous cellulases for the host termites.

Endosymbionts are also found in the guts of longhorn beetles [20]. However, the digestive tract of longhorn beetle larvae differs from that of termites. The longhorn beetle hindgut has a small volume and simple structure compared with the termite hindgut [4,21], which suggests that the digestive tract of longhorn beetles cannot accommodate as many endosymbionts as termites. The exogenous cellulolytic system of longhorn beetles therefore cannot contribute as much to cellulose digestion as it can in termites.

Cellulase gene expression is a direct method for evaluating the cellulolytic system in longhorn beetles. Progress in RNA sequencing technology has enabled high throughput analyses of gene expression and such analyses have been accomplished in number of beetle species [22,23]. The genome data obtained on the red flour beetle *Tribolium castaneum* [24] also contribute to such research.

In this study, cellulases expressed in the longhorn beetle *Mesosa myops* were investigated through sequencing the transcriptome of larvae, pupae, and adults. A total of eight endocellulase genes from *M. myops* were identified and the sequences and expression levels were determined. The results indicated that *M. myops* contains a complex endogenous cellulolytic system.

Materials and Methods

Insects

Mesosa myops larvae were collected from *Salix babylonica* trees in Beijing (GPS: N40.278586, E116.046411), China. *Mesosa myops* is not a rare or protected species in China. Infested branches containing larvae were taken back to the laboratory and reared within the tree branches at 25°C with stable relative humidity in complete darkness. A sample of the tree branches was dissected at 5-day intervals to observe *M. myops* development. Larvae were transferred to a worm tube after feeding was completed. Three developmental stages of the life cycle were selected in this study: fourth instar larva (5 days following final ecdysis), pupa (7 days after pupation), and adult (4 days after eclosion). Adults were fed with leaves of *S. babylonica* for several days before the experiment. Six individuals were included in each group.

Laboratory methods about transcriptome

Sequencing of transcriptomes was performed using the Illumina Hiseq 2000 platform following standard protocols as provided by the manufacturer (Illumina, San Diego, USA). Briefly, total RNA was first isolated with an Rneasy Mini kit (Qiagen, Hilden, Germany) and then the mRNA was purified using Sera-mag Magnetic Oligo (dT) (Thermofisher, Wisconsin, USA). Purified mRNA was fragmented into 100–400 bp pieces with a divalent cations protocol to avoid priming bias in the subsequent cDNA synthesizing step. Then, double-stranded cDNA was synthesized using the SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, Carlsbad, USA) with random hexamer (N6) primers (Illumina). After the process of end-repair and phosphorylation (using T4 DNA polymerase, Klenow DNA polymerase, and T4 PNK), the 3′ ends of the cDNA fragments were adenylated with DNA polymerase Klenow Exo-(3′–5′ exo minus) (Illumina). Then, the Illumina Paired-end adapters were linked to the ends of these 3′-adenylated cDNA fragments. The cDNA fragment ranges within 200 ± 25 bp in length were extracted using 2% agarose gel in TAE-buffer. Then, 15 cycles of polymerase chain reaction (PCR) were performed to amplify the cDNA template using PCR Primer PE 1.0 and PCR Primer PE 2.0 (Illumina). The cDNA library was constructed using purified cDNA fragments (length ranges within 200 ± 25 bp). The quality of the library was validated using an Agilent 2100 Bioanalyzer with Agilent DNA 1000 chip kit (Agilent Technologies, Santa Clara, USA). Then, the library was sequenced on a flow cell using Illumina Hiseq 2000 platform with standard workflow comprising steps of hybridization, isothermal amplification, linearization, blocking, sequencing primer hybridization, and sequencing. The sequencing data were filtered and low quality bases were removed. Finally, the pure data were used for the subsequent analyses.

Transcriptome data processing

All of the raw data from the three developmental stages were assembled together with the software Trinity [25] using default parameters with the exception that minimum length of contigs was set to 200, and the parameter maximum_reads_per_graph was set to 200,000. The sequences homology was first searched against an NCBI non-redundant (nr) protein database using BLASTx with an *E*-value cut-off as 10^{-3} (minimum length of 20 amino acids). And then the sequences retrieving no BLASTx hits were searched again using BLASTn against an NCBI nr nucleotide database with cut-off of *E*-value as 10^{-10}. Functional annotation by gene ontology terms (GO, www.geneontology.org), metabolic pathways (Kyoto Encyclopedia of Genes and Genomes, KEGG) were performed using the BLAST2GO software suite v2.3.1. Expression of the unigenes was evaluated using the RPKM (Reads Per Kilobase of exon model per Million mapped reads) [26] value. The cellulase genes obtained in the transcriptome were then extracted.

Molecular evolution of the cellulases

Phylogeny analyses were used to infer the biological sources of the cellulase genes. Amino acid sequences of the recorded cellulase genes in Carbohydrate-Active enZYmes Database (CAZY, http://www.cazy.org/) [27] were downloaded and combined into one data matrix

with the cellulase obtained from the present research. But patent sequences, repeated sequences, and sequences with lengths too short or too long were excluded in the analyses. The related cellulase genes of another longhorn beetle, *Anoplophora glabripennis* [28], were also included in the subsequent analyses. The data matrixes are provided in **Supplementary data**. Amino acid sequences were aligned with software MUSCLE [29] (gap open = −2.9; gap extension = −0.5) and then used to construct phylogenetic trees. Data files were prepared using MEGA [30] and Bioedit [31] software. Tree building with neighbor joining (NJ) criteria was performed using MEGA and tree building with maximum likelihood (ML) criteria was performed with Garli software [32]. The evaluation of the final NJ and ML trees was performed using the bootstrap [33] method (1000 replications).

GH45 and GH5 cellulases data used for phylogenetic tree building

All the recorded GH45 members in CAZY are cellulases and a total of 218 GH45 cellulase genes were included in this analysis. All of the genes are from the CAZY database; the amino acid sequence data were downloaded from GenBank and their accession IDs and sequences are listed in **Supplementary data**. Some sequences have relatively short lengths. GH5 is a large enzyme family including cellulases and other enzyme types. GH5 has more entries in the CAZY database (4353 records) than GH45. Representative cellulases were extracted for tree building to infer the evolutionary relationships of the newly discovered GH5 cellulase genes. A total of 80 GH5 cellulase sequences were included in the final tree-building analyses. Accession IDs and sequence data matrixes of GH45 and GH5 cellulases used in tree building are given in **Supplementary data**. Sequences with fewer than 200 amino acid residues were excluded. One representative sequence was selected from each series of very similar or identical sequences for the final tree building.

Results

Information of transcriptome

The transcriptome of the three developmental stages was separately sequenced and produced 14.7 Gbp clean data (larvae: 4.62 Gbp; pupa: 4.5 Gbp; and adult: 5.58 Gbp). All the data were combined to produce 68,180 unigenes. Transcriptome information is summarized in Table 1.

Cellulase genes in *M. myops*

A total of eight cellulase genes from the entire transcriptome of *M. myops* were found, including six complete and two incomplete cDNA sequences (Table 2). One cellulase (CGH45A, complete cDNA sequence, 235 aa long) belongs to the GH45 and the remaining seven enzymes belong to the GH5. The lengths of the GH5 cellulases range from 317 to 327 aa predicted from the six complete sequences. All of the identified cellulase genes were submitted to GenBank and Accession IDs are listed in Table 2.

Expression dynamics of the cellulases genes

Expression levels varied significantly among the identified cellulase genes. The expression level of CGH45A, the only GH45 cellulase gene, was the greatest determined by the RPKM values. The RPKM value of CGH45A in larvae was 3913.4, indicating that this gene produced nearly 0.4% of the total mRNA, a relatively high expression level for a single gene. The expression level of CGH45A in larvae is relatively high among all of the unigenes (**Fig. 1**).

Expressions of GH5 cellulase genes in larvae were lower than that of the GH45 gene. However, the absolute expression intensities of several GH5 cellulase genes are greater than most of the other genes identified (**Fig. 1**). Expressions of the four GH5 cellulases (CGH5D, CGH5B, CGH5A, and CGH5C) were relatively high (RPKM > 500) in the larval stage (**Fig. 1**) compared with other stages. The high expression of the cellulase genes meets the requirement for cellulose digestion in the larvae.

Most cellulase genes show a declining trend of expression along with the insect development. There are two major expression patterns of the cellulase genes in this study (**Fig. 2**). Six cellulase genes follow pattern A (**Fig. 2A**) in which expression levels are down-regulated during development. Cellulases with pattern A have the highest expression in larvae and the lowest expression in adults. The remaining two cellulase genes follow pattern B (**Fig. 2B**), with low expression in larvae but high expression in adults.

Evolutionary analyses of *M. myops* cellulases

Understanding the evolution of cellulase genes is an important part in the study of cellulolytic systems. Phylogenetic analyses with available cellulases (including the cellulases yielded in this study and representatives of available cellulases obtained from public databases) were performed.

There are 216 GH45 cellulase gene records collected in CAZY (http://www.cazy.org/), a specific database for glycoside hydrolase. Among these, 194 were obtained from Eukaryota, including insects, fungi, nematodes, plants, and others. The GH45 cellulase genes (CGH45A) obtained in this study clustered with the GH45 cellulase genes found in other Coleoptera. The available GH45 cellulase members in the Coleoptera formed a well-supported linage in the phylogenetic tree (**Fig. 3**, GH45 tree). This relationship suggests that the GH45 cellulase may be universal in Coleoptera. Because of limited data

Table 1. Annotation summary for *M. myops* transcriptome assembly

Number of unigenes	68,180
Pure data (total)	14.7G
Average length of unigenes	1084 bp
Unigenes with BLASTP Hits (threshold of 10^{-5})	43,264 (63.5%)
Average length of CDS with blast hits	228 aa[a]
Number of larval expressing unigenes	33,026
Number of pupal expressing unigenes	34,082
Number of adult expressing unigenes	53,052
Number of unigenes with GO assignments	27,678 (40.6%)
Number of unigenes with KEGG assignments	17,730 (26.0%)
Number of unigenes with Pfam assignments	11,301 (16.6%)

[a]Amino acid.

Table 2. Cellulase genes identified in transcriptome of *M. myops*

Cellulase[a]	Length (bp)	Amino acid length	Organism	Accession ID
CGH5A	1182	327	*Ditylenchus destructor*	KP875552
CGH5B	1102	321	*Ditylenchus destructor*	KP875553
CGH45A	811	235	*Apriona germari*	KP875551
CGH5C	1394	324	*Rotylenchulus reniformis*	KP875554
CGH5D	1069	324	*Ditylenchus destructor*	KP875556
CGH5E	985	323	*Ditylenchus destructor*	KP875555
CGH5F	1013	317	*Ditylenchus destructor*	KP875557
CGH5G	212	70[b]	*Rotylenchulus reniformis*	KP875558

[a]Each cellulase was named by the Glycosyl hydrolase family it belongs.
[b]The sequence is not complete.

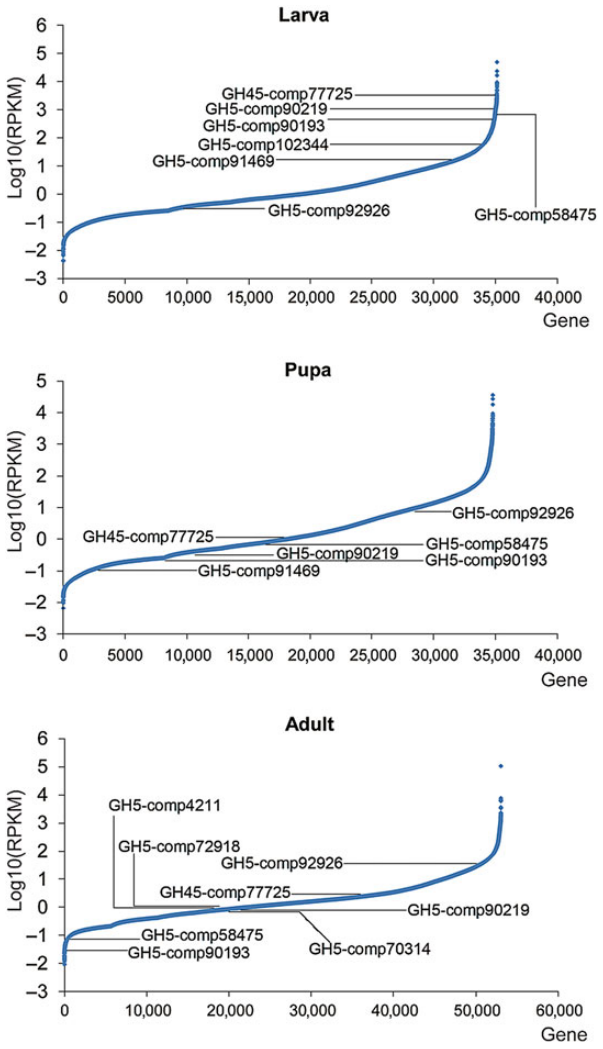

Figure 1. Relative expression intensity of the cellulase genes of *M. myops* All of the recorded unigenes are shown according to its expression intensities (with the format of logarithm of RPKM value) and the cellulases are marked to show their relative level. Horizontal axis: category axis for all the genes found in transcriptome; vertical axis: logarithm of RPKM value.

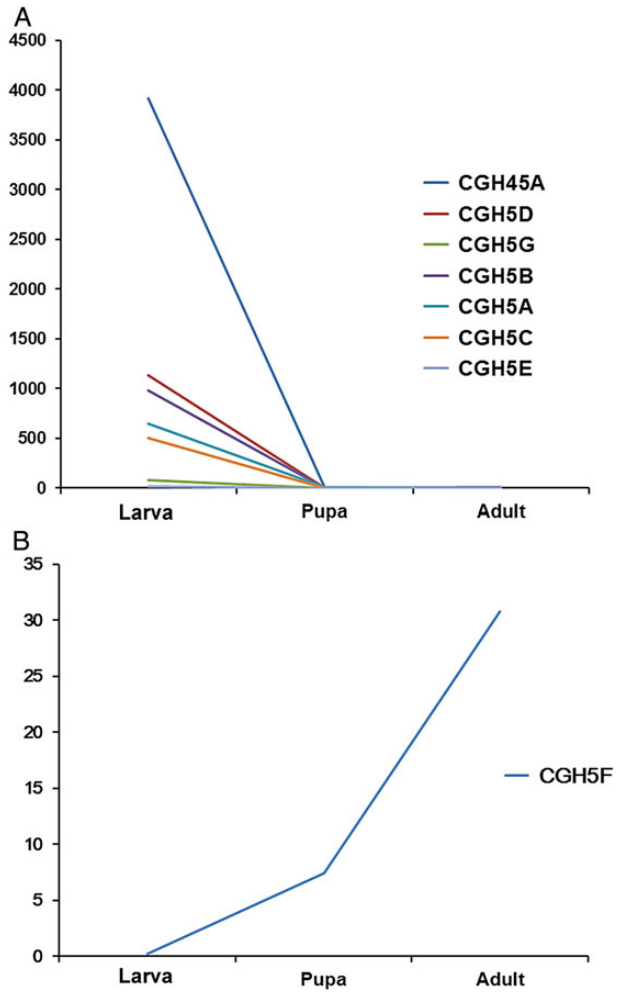

Figure 2. Cellulase gene expression dynamics in the three main developmental stages of *M. myops* (A) The up-regulated cellulases during development. (B) The down-regulated cellulases during development.

available, further studies with additional species are needed to confirm the potential universal presence of GH45 cellulases in Coleoptera.

Most of the GH45 cellulases were found in fungi. However, phylogeny analyses revealed polyphyletic linages of fungal cellulase. The Coleoptera GH45 cellulase linage was clustered with the linage of yeast (belong to Fungi) cellulases. This result indicated the complexity of GH45 cellulase origination in fungi. The cellulases of the symbiotic protists of termites formed a well-supported branch as well as the Mollusca cellulases and the nematode cellulases. These results suggested that GH45 cellulases are widespread in different animal groups, indicating the generic importance of this enzyme for a diversity of wood-feeding animals. No endogenous GH45 cellulase was found in termite genomes. However, large amount of exogenous GH45 cellulase occurs in termite symbiotic protists. The presence of GH45 cellulase is a good indicator of the wood-feeding habit and cellulose utilization in animals, especially in Coleoptera. Almost all of the insects, GH45 cellulases recorded in CAZY have been identified from species in the Coleoptera superfamilies Chrysomeloidea and Curculionoidea (excepting one Collembola species). More detailed studies covering a wider range of Coleoptera groups will be needed to determine if the GH45 cellulase is a characteristic cellulase in Coleoptera.

The ML tree of the GH5 cellulases indicates that the GH5 cellulase tree is more complex than the GH45 tree (**Figs.** 3 and 4). Most of the GH5 cellulases were identified from bacteria and fungi. However, the fungal GH5 cellulases formed series branches that were mixed with other branches (**Fig.** 4). All of the arthropod GH5 cellulases are derived from insects, specifically from longhorn beetles (CZAY and GenBank). The seven GH5 cellulases identified in this study together with the GH5 cellulases obtained from other longhorn beetle species [5,10,28,34] were included in the tree building. The phylogeny relationship (**Fig.** 4) supports the hypothesis that GH5 cellulases are endogenous in longhorn beetles. The beetle GH5 cellulases form a well-supported branch (bootstrap support = 100) and this result shows that longhorn beetles contain a series of actively expressed GH5 cellulase genes. The beetle branch, together with two well-supported branches formed by nematode and some bacterial cellulases, generates a large and well-supported branch (**Fig.** 4).

Discussion

Based on the expression analysis and evolutionary analysis, we reach the conclusion that there are two cellulose digestion systems in longhorn

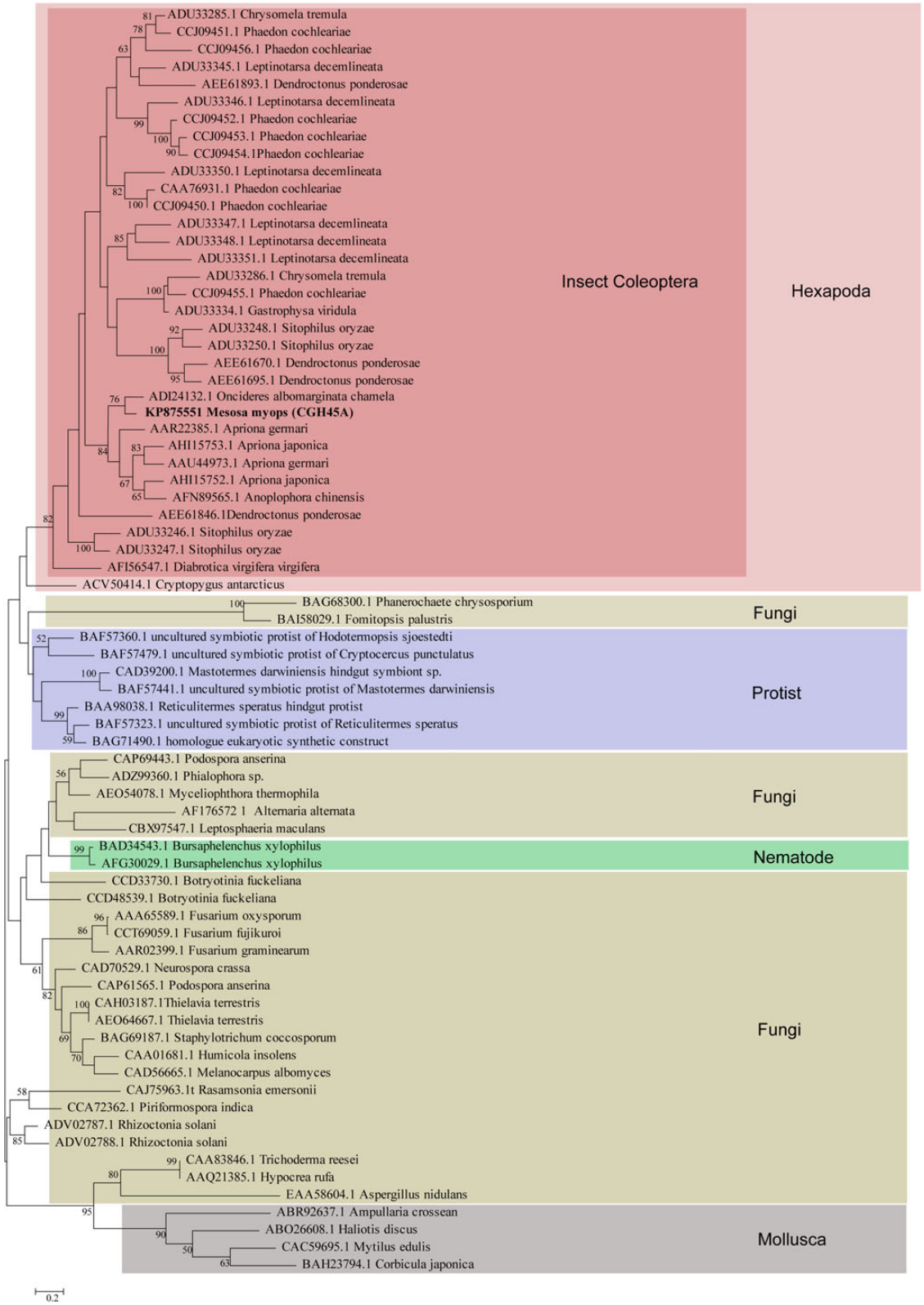

Figure 3. Maximum likelihood tree of GH45 cellulose genes Each cellulase is marked with the Accession ID and the name of organism. The data matrix is provided in appendix as **Supplementary data**. The tree is evaluated with bootstrap method and the support value is marked on the branch when it exceeds 50.

beetles. One system is represented by GH45 cellulase, which is widespread and characteristic to Coleoptera. The other system is represented by the GH5 cellulases. According to the expression characteristics and evolutionary relationships, we believe that these cellulases are products of endogenous genes. The expressions of most cellulase genes parallel to demands, especially in the larvae, for cellulose digestion in *M. myops*.

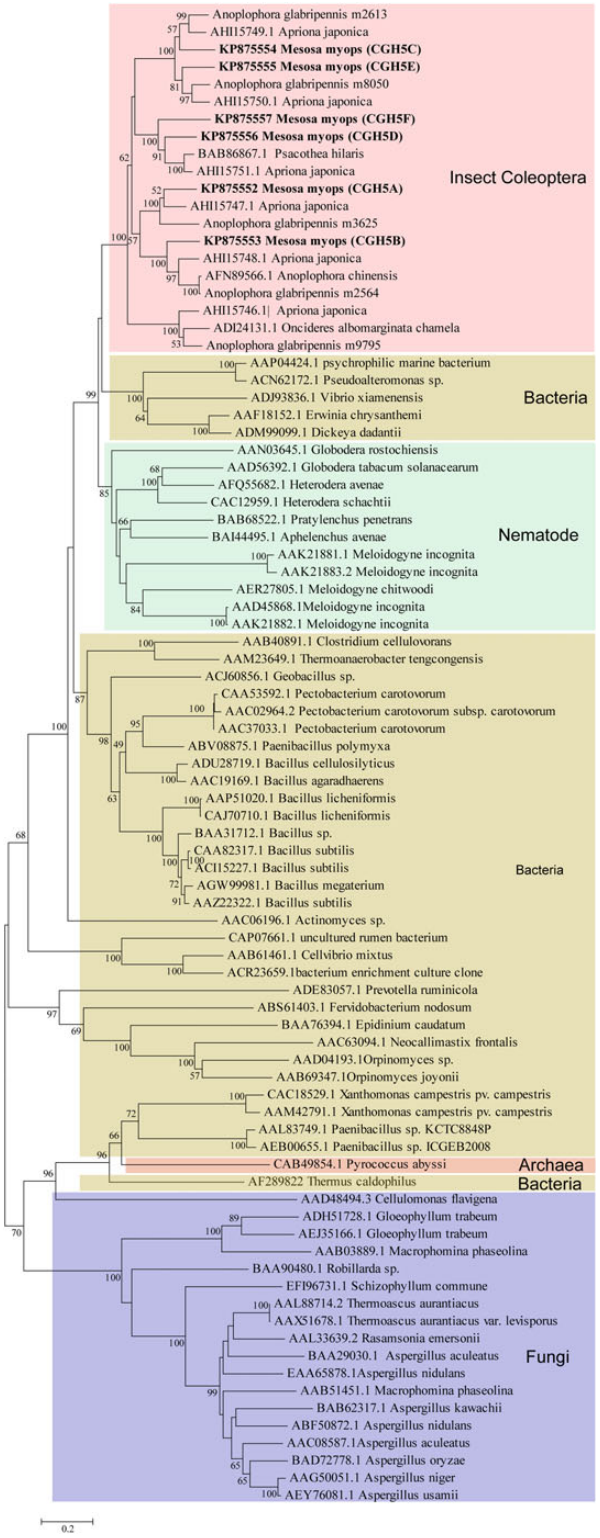

Figure 4. Maximum likelihood tree of GH5 cellulose genes Each cellulase is marked with the Accession ID and the name of organism. The data matrix is provided in appendix as **Supplementary data**. The tree is evaluated with bootstrap method and the support value is marked on the branch when it exceeds 50.

The origin of GH45 cellulase genes in Coleoptera is unclear. GH45 cellulase genes are common in some coleopteran linages, such as Chrysomelinae and Curculionidae. More studies on a wider taxonomic range of beetles will be needed to determine if GH45 cellulase is common in the Polyphaga. Phylogeny analyses across different species suggest that GH45 genes may be generated by endogenous genes. CAZY has recorded more than 40 GH45 cellulases from 32 Coleoptera species (in addition to the GH45 cellulase identified in this study). These GH45 cellulases form a monophyletic branch in the phylogenetic trees, indicating that these genes have a common ancestor gene which might render Coleoptera the ability to use cellulose for growth and metabolism. Horizontal gene transfer is one possible explanation for gene origination in organisms [35]. However, the probability is quite low that the horizontal transfer of one gene happens independently in most of the Coleoptera species. The widespread existence of GH45 cellulase gene in two Coleoptera superfamilies (suborder Polyphaga) suggested that it might have originated in the Polyphaga. An alternative explanation is that GH45 genes originated from common ancestor of Polyphaga.

CAZY currently has 4230 records of GH5. These include the activities of different enzymes, such as β-mannosidase (EC 3.2.1.25), chitosanase (EC 3.2.1.132), endo-β-1,4-glucanase (endocellulase, EC 3.2.1.4), and others. GH5 cellulase genes are widespread in different taxonomic groups and have been recorded in most currently living groups, including Archaea, Bacteria, and Eukaryota. GH5 cellulases from longhorn beetles form a monophyletic branch in the phylogenetic tree, indicating that the GH5 cellulases share a common ancestor. Several Lamiinae longhorn beetles contain relatively large numbers of GH5 cellulase copies [28,36]. This situation is special and has seldom been reported in insects.

Cellulases have a major influence on the evolution of plant-feeding insect species [37] and the present study supports this conclusion. However, only three GH5 cellulases were found previously in Coleoptera species (Cerambycidae) prior to this study. The expansion of GH5 cellulases in longhorn beetles, including *M. myops*, is therefore a rather unique phenomenon in insects. Recently, GH5 cellulases have also been identified, using transcriptomic methods, in the Cerambycid species *A. glabripennis* [28] and *Apriona japonica* [36]. It is well known that longhorn beetle larvae can ingest wood materials and therefore cellulose is the main carbon source for their metabolism and growth. Cellulose digestion is essential for longhorn beetle larvae compared with other plant-feeding animals because longhorn beetles utilize cellulose as their main carbon source. Cellulase helps to dissolve the cell walls, which improves the digestive efficiency [7]. The GH5 cellulase genes enable insects to utilize cellulose and enable longhorn beetle to adapt to a wood-feeding habit. The following questions are to be answered by future research. How has wood feeding evolved in other beetle species? How do the GH5 cellulases specifically function in the digestion tract of longhorn beetles? Why have longhorn beetles evolved numerous GH5 cellulase genes?

Most plant-feeding animals do not use cellulose as their main carbon source and the wood-feeding habit is clearly a special case among herbivores. One reason involves the stability of the cellulose molecule and the metabolic difficulty of releasing glucose from its molecular chains. Many plant-feeding animals could use limited amounts of cellulose with the aid of symbionts such as those found in ruminants [2]. However, most animals cannot use cellulose as a main food because they are unable to process adequate amounts of cellulose required to meet carbohydrate requirements. Details on how wood-feeding species such as termites and longhorn beetle larvae obtain sufficient carbohydrates remain inadequately understood.

Cellulase activity is a key factor for effective cellulose digestion. In most longhorn beetles, the larva is the main life stage utilizing cellulose

for nutrition. The expression of cellulase must be adequate to meet the requirements for cellulose digestion in larval *M. myops*. Transcription is the first step of gene expression and genes with greater production of mRNA usually have more final enzyme product. Furthermore, expression dynamics of the digestive enzymes should be correlated with the physiological requirements determined by the feeding habits. *Mesosa myops* larvae feed only on wood and the larva is the main feeding stage. *Mesosa myops* adults have relatively limited feeding requirements and their nutritional needs are met by leaf feeding. Larvae of longhorn beetles must depend on cellulases for conversion of cellulose into carbohydrates. From the habit characteristics, expression dynamics information, and expression levels, we conclude that the cellulases with Pattern A (**Fig. 2A**) are the dominant enzymes responsible for cellulose digestion (CGH45A, CGH5D, CGH5B, CGH5A, and CGH5C) in larvae of *M. myops* and the cellulases that follow Pattern B (CGH5F) are the dominant cellulases in adults.

The digestion of cellulose in some termites such as *C. formosanus* and *Reticulitermes flavipes* depends on hindgut enzymes produced by symbiotes [18,38,39]. At the same time, endogenous cellulases of the GH9 family are expressed in the foregut and midgut. The termites appear to have a dual cellulolytic system formed by both endogenous and exogenous cellulases. Similarly, the longhorn beetle *M. myops* also has dual cellulolytic system, but the detailed situation differs from that of the termites. Both of the two cellulolytic systems are endogenous in *M. myops* but they are evolutionarily distinctive. The first system has GH45 cellulase which might be common in Coleoptera. The second system includes a group of GH5 cellulases. In termites, the endogenous cellulase genes are mainly expressed in the salivary gland and the midgut, while the exogenous cellulolytic system exists in the hindgut. In longhorn beetles, the action mode of GH45 and GH5 cellulases and their spatial distribution in the digestive tract are still not clear.

There are at least two ways for endogenous cellulase to evolve. These are horizontal gene transfer and natural selection on advantageous mutations. It is difficult to identify a starting point for cellulase genes in insect linages such as Coleoptera and Isoptera based on our current knowledge. Irrespective of their origin, cellulase genes might be widespread endogenous genes in Coleoptera and might have contributed to the extreme diversity of Coleoptera, especially for the groups feeding on plant material.

This study raises several issues that need to be addressed by future study. What is the origin of GH45 cellulase? What is its role in the feeding habit differentiation among beetles? From the evolutionary and physiological perspectives, what are the roles of GH5 cellulases in the development of longhorn beetles wood-feeding habits? Information on the genomes and transcriptomes of a variety of other Coleoptera species will help to answer these questions.

Several cellulase genes are active in the wood-feeding longhorn beetle *M. myops*. We completed the characterization of one GH45 cellulase and seven GH5 cellulases from the transcriptome of *M. myops*. A dual endogenous cellulolytic system which includes a GH45 cellulase and a group of GH5 cellulases was characterized. The information about the active GH5 cellulase group could be valuable for the understanding of the wood-feeding habit in other longhorn beetle species. Phylogenetic and evolutionary analyses of the two kinds of cellulase indicated that both of the two systems are endogenous. The high expression levels and expression dynamics of these cellulase genes suggest that GH45 and GH5 cellulases are functional digestive enzymes that dominate cellulose digestion in *M. myops*. GH45 cellulases are likely to be common cellulases in Coleoptera. GH5 cellulases appear to be characteristic enzymes in longhorn beetles and they may play a role in the evolution of the wood-feeding habit in Coleoptera. The results of this study provide an example of the complexity of cellulose digestion in wood-feeding insects.

Supplementary Data

Supplementary data is available at *ABBS* online.

Funding

This work was supported by the grants from the Key Innovation-Projects from CAS (No. KSCX2-EW-G-4) and the National Science Fund for Fostering Talents in Basic Research (Special Subjects in Animal Taxonomy J1210002).

References

1. Tomme P, Warren RJ, Gilkes NR. Cellulose hydrolysis by bacteria and fungi. *Adv Microb Physiol* 1995, 37: 1–81.
2. Teunissen MJ, Smits AM, Op den Camp HM, Huis in't Veld JJ, Vogels GD. Fermentation of cellulose and production of cellulolytic and xylanolytic enzymes by anaerobic fungi from ruminant and non-ruminant herbivores. *Arch Microbiol* 1991, 156: 290–296.
3. Foley B. Heterotrophic nutrition in animals. In: Knox B, Ladiges P, Evans B, Saint R (ed). *Biology*, 2nd edn. Roseville: McGraw-Hill, 2001, 486–515.
4. Watanabe H, Tokuda G. Cellulolytic systems in insects. *Annu Rev Entomol* 2010, 55: 609–632.
5. Chang CJ, Wu CP, Lu SC, Chao AL, Ho TH, Yu SM, Chao YC. A novel exo-cellulase from white spotted longhorn beetle (*Anoplophora malasiaca*). *Insect Biochem Mol Biol* 2012, 42: 629–636.
6. Cairns JK, Esen A. β-Glucosidases. *Cell Mol Life Sci* 2010, 67: 3389–3405.
7. Calderón-Cortés N, Quesada M, Watanabe H, Cano-Camacho H, Oyama K. Endogenous plant cell wall digestion: a key mechanism in insect evolution. *Annu Rev Ecol Evol* 2012, 43: 45–71.
8. Veivers PC, Musca AM, O'Brien RW, Slaytor M. Digestive enzymes of the salivary glands and gut of *Mastotermes darwiniensis*. *Insect Biochem* 1982, 12: 35–40.
9. Veivers PC, O'Brien RW, Slaytor M. Selective defaunation of *Mastotermes darwiniensis* and its effect on cellulose and starch metabolism. *Insect Biochem* 1983, 13: 95–101.
10. Sugimura M, Watanabe H, Lo N, Saito H. Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, *Psacothea hilaris*. *Eur J Biochem* 2003, 270: 3455–3460.
11. Pauchet Y, Wilkinson P, Chauhan R, ffrench-Constant RH. Diversity of beetle genes encoding novel plant cell wall degrading enzymes. *PLoS ONE* 2010, 5: e15635.
12. MacLeod A, Evans HF, Baker RA. An analysis of pest risk from an Asian longhorn beetle (*Anoplophora glabripennis*) to hardwood trees in the European community. *Crop Protect* 2002, 21: 635–645.
13. Slaytor M, Veivers PC, Lo N. Aerobic and anaerobic metabolism in the higher termite *Nasutitermes walkeri* (Hill). *Insect Biochem Mol Biol* 1997, 27: 291–303.
14. Slaytor M. Energy metabolism in the termite and its gut microbiota. In: Abe T, Bignell DE, Higashi M (eds). *Termites: Evolution, Society, Symbioses, Ecology*. Dordrecht: Kluwer, 2000, 307–332.
15. Yokoe Y. Cellulase activity in the termite, *Leucotermes speratus*, with new evidence in support of a cellulase produced by the termite itself. *Sci Papers Coll Gen Educ Univ Tokyo* 1964, 14: 115–120.
16. Yokoe Y, Yasumasu I. The distribution of cellulase in invertebrates. *Comp Biochem Physiol* 1964, 13: 323–338.
17. Nakashima K, Watanabe H, Saitoh H, Tokuda G, Azuma JI. Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. *Insect Biochem Mol Biol* 2002, 32: 777–784.

18. Tokuda G, Lo N, Watanabe H, Arakawa G, Matsumoto T, Noda H. Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. *Mol Ecol* 2004, 13: 3219–3228.
19. Schmitt-Wagner D, Friedrich MW, Wagner B, Brune A. Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). *Appl Environ Microb* 2003, 69: 6018–6024.
20. Schloss PD, Delalibera I Jr, Handelsman JO, Raffa KF. Bacteria associated with the guts of two wood-boring beetles: *Anoplophora glabripennis* and *Saperda vestita* (Cerambycidae). *Environ Entomol* 2006, 35: 625–629.
21. Watanabe H, Tokuda G. Animal cellulases. *Cell Mol Life Sci* 2001, 58: 1167–1178.
22. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. *Insect Biochem Mol Biol* 2009, 39: 403–413.
23. Keeling CI, Henderson H, Li M, Yuen M, Clark EL, Fraser JD, Huber DPW, *et al*. Transcriptome and full-length cDNA resources for the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major insect pest of pine forests. *Insect Biochem Mol Biol* 2012, 42: 525–536.
24. Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Bucher G. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 2008, 452: 949–955.
25. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, *et al*. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 2013, 8: 1494–1512.
26. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Method* 2008, 5: 621–628.
27. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014, 42: 490–495.
28. Scully ED, Hoover K, Carlson JE, Tien M, Geib SM. Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle. *BMC Genomics* 2013, 14: 850.
29. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, 32: 1792–1797.
30. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007, 24: 1596–1599.
31. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999, 41: 95–98.
32. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation. The University of Texas at Austin, 2006.
33. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 1993, 42: 182.
34. Calderón-Cortés N, Watanabe H, Cano-Camacho H, Zavala-Páramo G, Quesada M. cDNA cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle *Oncideres albomarginata* Chamela (Cerambycidae). *Insect Mol Biol* 2010, 19: 323–336.
35. Hotopp JC. Horizontal gene transfer between bacteria and animals. *Trends Genet* 2011, 27: 157–163.
36. Pauchet Y, Kirsch R, Giraud S, Vogel H, Heckel DG. Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle *Apriona japonica*. *Insect Biochem Mol Biol* 2014, 49: 1–13.
37. Michael MM, Jones CG, Bernays EA. The evolution of cellulose digestion in insects (and discussion). *Phil Trans R Soc Lond B* 1991, 333: 281–288.
38. Inoue T, Moriya S, Ohkuma M, Kudo T. Molecular cloning and characterization of a cellulase gene from a symbiotic protist of the lower termite, *Coptotermes formosanus*. *Gene* 2005, 349: 67–75.
39. Zhou X, Smith JA, Oi FM, Kohler PG, Bennett GW, Scharf ME. Correlation of cellulose genes expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipens*. *Gene* 2007, 395: 29–39.