CMLS, Cell. Mol. Life Sci. 59 (2002) 1983–1992
1420-682X/02/111983-10

CMLS Cellular and Molecular Life Sciences

Research Article

New endo-β-1,4-glucanases from the parabasalian symbionts, *Pseudotrichonympha grassii* and *Holomastigotoides mirabile* of *Coptotermes* termites

H. Watanabe *,[a], K. Nakashima[b], H. Saito[a] and M. Slaytor[c]

[a] National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8634 (Japan), Fax +81 298 38 6108, e-mail: hinabe@affrc.go.jp
[b] Division of Environmental Science and Technology, Graduate School of Agricultural Science, Kyoto University, Sakyo-ku, Kyoto 606-8502 (Japan)
[c] Department of Biochemistry, The University of Sydney, Sydney, New South Wales 2006 (Australia)

Received 18 September 2002; accepted 20 September 2002

Abstract. An endo-β-1,4-glucanase (EG) was purified from the hindgut of an Australian mound-building termite, *Coptotermes lacteus*. The hindgut extract had a peak separate from those for extracts obtained from the salivary glands and the midgut based on sephacryl S-200 gel chromatography, and also demonstrated an origin different from the endogenous EGs of the termite itself. The recovery was further purified by SDS-PAGE, and its N-terminal amino acid sequence analyzed. This showed high homology to EGs from glycoside hydrolase family (GHF) 7. PCR-based cloning methods were applied to the hindgut contents of *C. lacteus* and individual protozoan symbionts from *C. formosanus*. cDNAs encoding putative EGs homologous to GHF7 members were then identified. The functionality of one of the putative proteins was confirmed by its expression in *Escherichia coli*.

Key words. Parabasal; cellulase; GHF7; Formosan termite; cloning.

Although termites are well known as xylophagous animals, until recent reports of endogenous and protozoan cellulose genes [1–3], little had been known about cellulolytic enzymes apart from chromatographic separations of cellulolytic activities [4, 5]. Termites (order Isoptera), which comprise seven families, are divided into two categories, one of protozoa-independent species (family Termitidae) and the other of protozoa-dependent species (families Mastotermitidae, Kalotermitidae, Termopsidae, Hodotermitidae, Rhinotermitidae, and Serritermitidae) [6]. The first group do not generally harbor cellulolytic protozoa in their guts, and thus to account for the deviating distribution of cellulase activity detected in the midgut of this group, termite researchers have assumed that cellulolytic enzymes must be produced in the midgut tissue where no cellulolytic symbionts have been identified [5, 7–9]. Still, the endogenous production of cellulolytic enzymes had not been generally accepted until the recent finding of a gene encoding an endo-β-1,4-glucanase (EG; EC 3.2.1.4; a type of cellulolytic enzyme which often represents cellulase) of glycoside hydrolase family (GHF) 9 from the protozoan-independent termite, *Nasutitermes takasagoensis* [2, 4].

In the second group, composed of protozoa-dependent species, the original proposal for protozoan cellulose digestion was made by Cleveland [10], and a combined cellulose digestion of endogenous and protozoan cellulases in termites was proposed because of the simultaneous distribution of cellulolytic activities both in secreting organs (the salivary glands and midgut, which do not have

* Corresponding author.

cellulolytic symbionts), and a non-secreting organ (the hindgut) harboring protozoan fauna [7, 11]. The endogenous and protozoan enzymes have been considered to share responsibility for cellulose digestion [8, 11–13]. Despite this historical background, only the role of hindgut protozoan fauna has been emphasized, but gene level demonstration had been done only for the endogenous EGs (GHF9) [1] until the recent finding of a GHF45 cellulase from hindgut hypermastigote protozoa of *Reticulitermes speratus* [3]. Nakashima and Azuma [14] purified a cellulase to be placed in GHF7 from the termite, *Coptotermes formosanus*, and attributed the salivary glands as the producer of this enzyme, but we recently renounced this and reattributed it to the hindgut protozoan fauna based on identification and relocation of each cellulase among different digestive organs (the salivary glands, foregut, midgut, and hindgut) [15]. Yamaoka and Nagatani [11] demonstrated the presence of two different enzymes in the gut of the protozoa-inhabited termite, *R. speratus*: one from the salivary glands was active against carboxymethylcellulose (Cx-cellulase; endo-β-1,4-glucanase), and the other from the hindgut protozoan fauna was active against filter paper (C1-cellulase; cellobiohydrolase), which were assumed to digest crystalline cellulose synergistically. Despite the identification of cellulolytic activities in the termite digestive systems, the exact contribution of each species of protozoa in cellulose digestion largely remains to be clarified. The report of cellulase genes from the hindgut protozoa of *R. speratus* has partially addressed the nature of the cellulolytic enzymes produced by the symbiotic protozoa [3], however, it does not provide a general explanation for the composition of protozoan enzymes in other protozoa-inhabited termite species.

In the present study, we purified possible protozoan EGs from hindgut extracts of the Australian mound-building termite, *C. lacteus*, analyzed their amino acid sequences, and elucidated the major protozoan cellulolytic component. Following the isolation of EGs, we demonstrated expression of EG-coding genes in the hindgut contents (containing protozoan fauna) of *C. lacteus* and, in addition, those from the parabasalian symbionts of the worldwide pest species, *C. formosanus,* which shares a common hindgut protozoan composition of the genera *Pseudotrichonympha, Holomastigotoides*, and *Spirotrichonympha* with other *Coptotermes* species [16–18], employing PCR-based cloning techniques.

Materials and methods

Termites

C. lacteus workers were collected from a termite mound in Wyong, N.S.W., Australia, and used for protein purifications. Some *C. lacteus* workers were preserved in absolute acetone for subsequent cDNA cloning [19] because of limited supply and access to the collection site. *C. formosanus* workers were collected on Iriomote Island in Okinawa Prefecture, Japan, and used for cDNA cloning in fresh form. Termites for fresh use were kept with their nest materials and tissue paper or sawdust for food for several days until they were used for the experiments.

Detection of EG activity

Enzymatic samples (crude extracts or fractions of gel filtration, 10 µl) were incubated in 200 µl of 2% (w/v) sodium carboxymethylcellulose [CMC; standard molecular weight 250,000, degree of carboxymethyl substitution 0.7 (w/v); Aldrich] in sodium acetate buffer (0.1 M, pH 5.5) at 37 °C for 5 min. Reducing sugars were detected with tetrazolium blue reagent (Sigma) as glucose equivalents [20].

Gel filtration and SDS-PAGE

The hindgut, salivary glands, and the midgut including the esophagus and proventriculus from 200 workers of *C. lacteus* were homogenized separately in 500 µl of ammonium acetate buffer (0.3 M, pH 5.0). Each homogenate was centrifuged for 5 min at 9,000 *g* and the supernatant was applied to a gel filtration column (HIPREP 16/60 Sephacryl S-200 High Resolution; Amersham Biosciences). Samples were eluted with ammonium acetate buffer at a flow rate of 0.5 ml/min, and the peak fractions and its shoulders were stored separately. Shoulder fractions were applied to the same column and eluted under the same conditions. The recovered peak fraction was added to the first recovery. The stored recovery was denatured with SDS-PAGE sample buffer and SDS-PAGE was performed (0.75%, 1 mm thick, 100 mm running length) according to the conventional method [21].

N-terminal amino acid analysis

Proteins on SDS-PAGE gels were transferred to a PVDF membrane using a semi-dry blotting cell (Bio-Rad) according to the method of Hirano and Watanabe [22]. Protein bands on the membrane were cut out and applied to an N-terminal amino-acid sequencer (Beckman).

cDNA cloning

mRNA was extracted from the hindgut contents using the QuickPrep micro mRNA extraction kit (Amersham Biosciences). First-strand cDNA was constructed with oligo-dT primer with (for 3′-RACE) or without an anchor sequence and Superscript II reverse transcriptase (Invitrogen) according to the manufacturer's recommended conditions. PCR amplification of the first-strand cDNA was performed with a set of degenerate primers designed from the N-terminal sequence of the purified protein with recombinant Taq polymerase (Takara, Japan) according to the manufacturer's recommended buffer and substrate

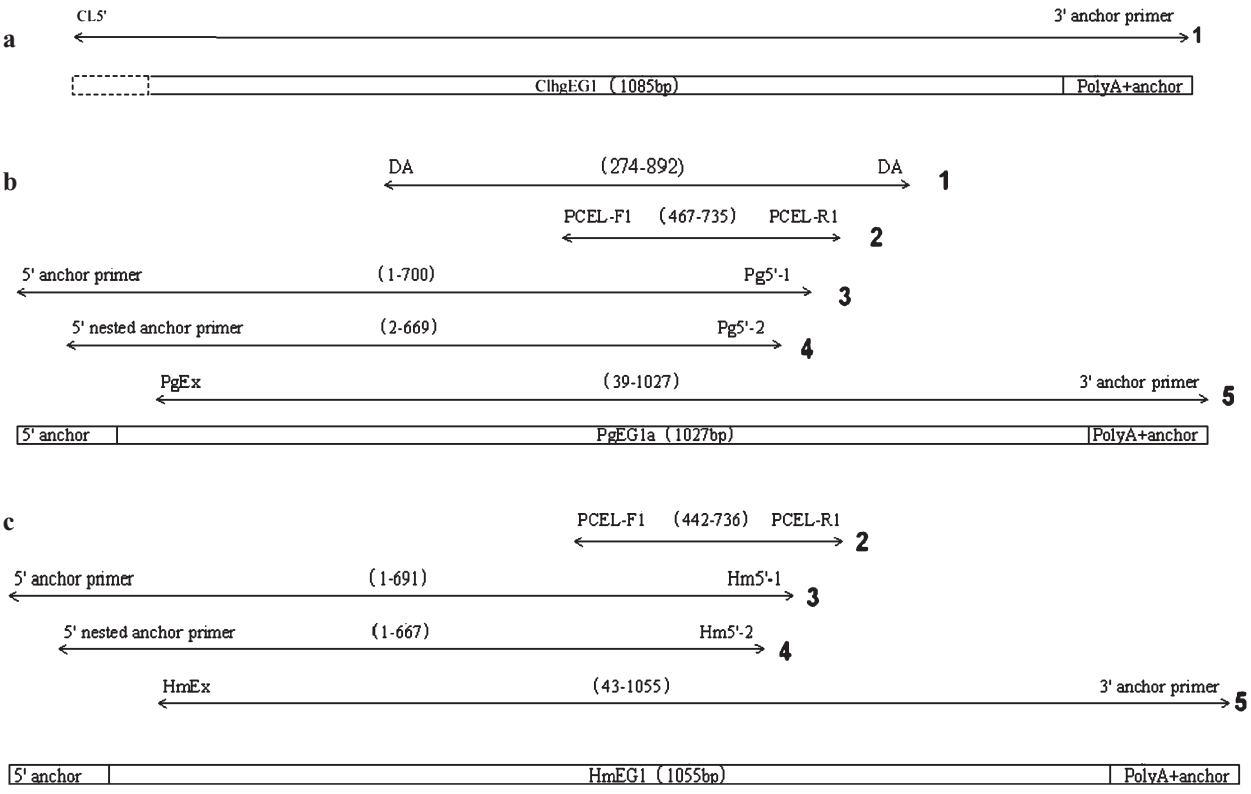

Figure 1. Cloning strategies for endo-β-1,4-glucanase cDNAs from the hindgut of the termites, *Coptotermes lacteus* [CLhgEG1 (*a*)] and *C. formosanus* [PgEG1a (*b*) and HmEG1 (*c*)]. The ranges amplified by PCR are designated by the arrow-headed lines and the nucleotide numbers (5′ and 3′ terminals in parentheses) including the primer sequences except the 5′ end of *CLhgEG1* (*a*). Numbers in bold characters indicate the orders in PCR amplifications of each strategy. Amplification **2** was performed by single-cell RT-PCR, while the others amplifications were performed on mRNA extracted from the hindgut contents by RT-PCR. Other homologues (CLhgEG2, PgEG1b-PgEG1h, PgEG2, PgEG3, HmEG2, and HmEG3) were also cloned by the same strategy for each host species. Details of the experimental conditions are described in the text.

conditions. For identification of 5′- and 3′-flanking regions, rapid amplification of cDNA ends (RACE) was employed with gene-specific primers. For cloning cDNA of EG from the hindgut of *C. lacteus*, a degenerate primer was redesigned from its N-terminal amino acid sequence, and the 3′-RACE method was directly applied to amplify the cDNA [23]. All RACE amplifications were performed using the SMART RACE cDNA amplification kit (Clontech). Nested PCR was performed for RACE of the 5′-flanking regions. The cloning strategy employed is summarized in figure 1.

Identification of protozoan species expressing EG genes

To identify EG-expressing protozoan species, RT-PCR based on individual protozoan cells (single-cell RT-PCR) was employed. The hindgut contents of the workers were suspended in a 90% concentration of solution U [24], and each protozoan cell was separated using a pipette with a 200-µl tip under a microscope. Subsequently, each protozoan cell was ejected into 100 µl of acetone in a 0.5-ml PCR tube to be dehydrated. The acetone was dried by heating the tube on a heating block (100 °C) for 1 min. Ten microliters of the first-strand mixture [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 5 mM $MgCl_2$, 1 µM of antisense gene-specific primer, 0.25 µM each of dNTPs, 0.5 µl of reverse transcriptase (Superscript II, Invitrogen) and 0.25 µl of human placenta ribonuclease inhibitor (Toyobo, Japan)] was added to the dried specimen and incubated for 1 h at 42 °C. The reaction was stopped by heating at 70 °C for 5 min, and following this the tube was placed on ice. Supplemental reagent solution was added to make 50 µl of PCR reaction mixture [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 15 mM MgCl2 and, finally, 2.5 units of recombinant Taq polymerase (Takara)]. The PCR reaction was performed using the hot-start technique. The single-cell RT-PCR was used to identify the origin of particular mRNAs but was inadequate for cloning full-length cDNA because of the small recovery of cDNA from the reverse transcription step.

DNA sequencing

For all PCR amplifications, the products were cloned into a plasmid cloning vector using the pMOSBlue blunt-end cloning kit (Amersham Biosciences), or pGEM-T plasmid cloning vector (Promega) with JM109 bacterial host.

Table 1. Primers used for the experiments.

Name	Design	Corresponding sequence (position), side
DS	5′ CAY CCN AAR TTY ACN TGG CAR 3′	'HPKFTWQ' (3–9), sense
DA	5′ RTT RTC RTT NGT RCA NCC RC 3′	'CGCTNDN' (14–20), antisense
PCEL-F1	5′ TGG ATA TGG TTA CTG TGA TGC GAA CTG CGT 3′ (C) (C) (C) (G)	 PgEG1a (438–467), sense HmEG1 (442–471), sense
PCEL-R1	5' GAT CGT CTT CCC ACC CTG CAC ATA CAA ACG 3′	PgEG1a (73–706), antisense HmEG1 (736–707), antisense
Hm5'-1	5′ CCG TTA GCG AMC CTC CTG ACC CCA C 3′ (C)	HmEG1 (698–674), antisense
Hm5'-2	5′ CGA MCC TCC TGA CCC CAC AAA STG 3′ (C) (C)	HmEG1 (691–667), antisense
HhEx	5′ CGA GAA GCA TCC GAA GTT TGT G 3′	HmEG1 (4–64), sense
Pg5'-1	5′ CCC ATT CAA GTT ACC AGA AGC ATC C 3′ (A) (G)	PgEG1a (700–675), antisense
Pg5'-2	5′ CAA GTT ACC AGA AGC ATC CGC ACG G 3′ (G)	PgEG1a (693–669), antisense
PgEx	5′ CGA GAA GCA TCC GAA CTT TAC 3′	PgEG1a (3–59), sense
CL5′	5′ GAR AAR CAY CCN AAR TTY ACN TGG CA 3′	EKHPKFTWQ (1–9), sense

The degenerate primers (DS and DA) were designed from the N-terminal amino acid sequence of protozoan endo-β-1,4-glucanase purified from the hindgut of *Coptotermes lacteus*. The letters in parentheses indicate actual bases finally confirmed in the corresponding sequences. Abbreviations for positions with degenerate codon positions are: Y for C+T, R for A+G, M for A+C, S for C+G, and N for A+T+G+C. DDBJ/GenBank accession numbers for PgEG1a and HmEG1 are AB071001 and AB071011, respectively.

The constructed plasmids were extracted from the host bacterial cells using conventional alkaline mini-prep and sequenced with an ABI377 or ABI3700 automated DNA sequencer using BigDye Terminator techniques with appropriate primers (designed for vector sequences or the same primers as used for the PCR amplifications). All gene-handling procedures were performed according to textbook methods [25] unless otherwise noted.

Expression of protozoan EG cDNA in *Escherichia coli*

The region coding from the N-terminus to the stop codon (i.e., excluding the putative signal peptide) of one protozoan EG cDNA was amplified by PCR using a sense primer (HmEx; table 1) and an anchor primer (joined at the 3′ end) following the method for 3′-RACE, and then ligated in a *Lac Z* coding frame at the *Eco*RV site of the pMOSBlue plasmid vector. The construct was transformed by pMOSBlue host cells (Amersham Biosciences). Expression and activity detection were performed using the method described previously [2].

Sequence analysis

Homology searches of the determined sequences were performed using the BLAST program [26] through the NCBI server at http://www. ncbi.nlm.nih.gov/BLAST/. Alignment of the retrieved, mature-form homologues was performed using CLUSTAL X (Version 1.8) for Windows (National Center for Biotechnology Information) with Gonnet 250 pairwise parameters [27]. Phylogenic trees were constructed using TREE-PUZZLE tree reconstruction software, with quartet-sampling and neighbor-joining of 1000 puzzling steps using the WAG substitution model [28].

Results

Purification of the protozoan EG and N-terminal sequencing

Hindgut, midgut, and salivary gland extracts of *C. lacteus* were applied separately to the gel filtration column. EG activity of the hindgut formed a peak at 69.6 ml, while the others formed peaks at 74.4 ml (salivary glands), or 72.0 and 74.4 ml (midgut), as seen in figure 2. The hindgut peak fraction from each chromatography was stored and applied to SDS-PAGE. Proteins on the SDS-PAGE gel were transblotted on PVDF membrane where one band of EG was separated from contaminations. N-terminal amino acid sequencing was conducted for the EG band, to give the amino acid sequence EKHPKFTWQEVTK ©G©TNDN (©, blank positions assumed to be Cys residues, which are undetectable using conventional amino acid sequencers), which showed homology to other EGs in GHF7 following BLAST X searches.

Cloning of protozoan EG cDNAs

Degenerate primers (DS for sense side and DA for the antisense side) (table 1) were designed from the N-terminal amino acid sequence of the purified hindgut EG of *C.*

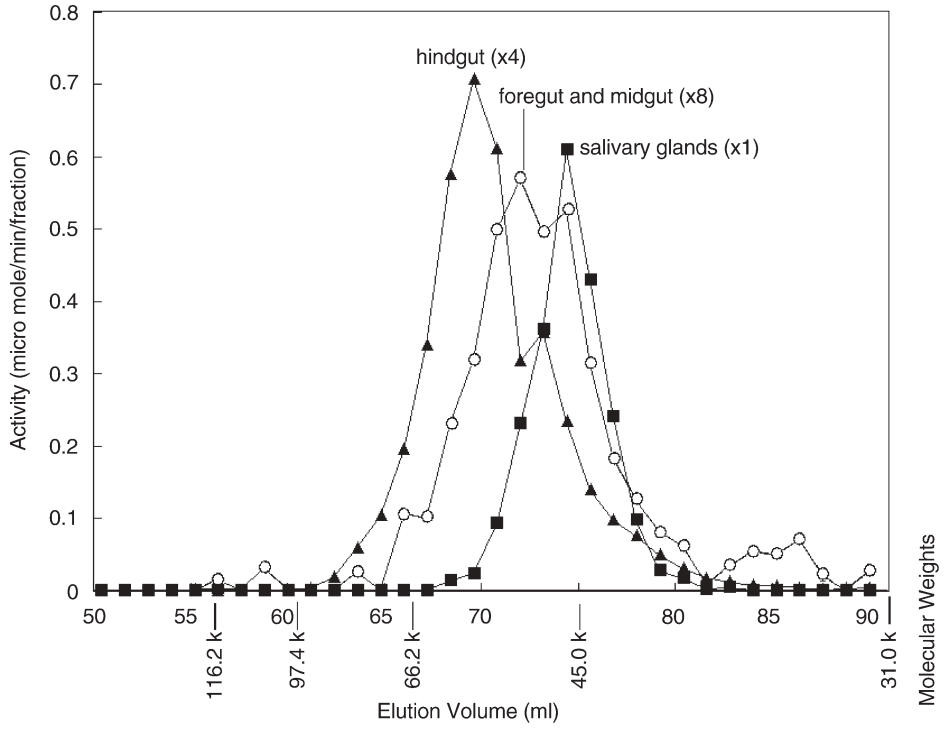

Figure 2. Elution profiles of endo-β-1,4-glucanase activities from the salivary gland (solid squares), the foregut and midgut (open circles), and the hindgut (solid triangles) extracts of the termite, *C. lacteus* on Sephacryl S-200. Enzymatic activity of the hindgut extract and that of the foregut and midgut extract were magnified to four and eight times the original values, respectively. Details of the experimental conditions are given in the text.

lacteus. First-strand cDNAs were synthesized with oligo-dT primers with or without an anchor sequence from mRNA extracted from the hindgut contents from *C. lacteus,* which had been dehydrated and preserved in absolute acetone, and from that of freshly collected *C. formosanus*. The first-strand cDNAs were applied for PCR amplifications to isolate target cDNA. The overall PCR strategies for cloning are summarized in figure 1. From the hindgut contents of *C. lacteus*, two cDNAs (CLhgEG1 GenBank accession AB089800 and CLhgEG2 AB089801, 1031 and 1085 bp, respectively) were obtained (fig. 1a). From *C. formosanus*, although the N-terminal coding region of the intended cDNA was not amplified by the designed degenerate primers, a partial sequence of protozoan EG cDNA [corresponding to bases 274–892 of PgEG1a with the antisense primer sequence (DA, 20 bp, table 1) at both ends] was obtained (fig. 1b). From this partial sequence, a sense primer (PCEL-F1) and an antisense primer (PCEL-R1) were designed (table 1). Using these primers, two different groups of partial cDNAs were identified from the protozoa, *Pseudotrichonympha grassii* and *Holomastigotoides mirabile* (PgEGs and HmEGs, respectively) by single-cell RT-PCR (fig. 1a, b). The amplification of HmEGs from *H. mirabile* was apparently due to the unexpected similarity of PCEL-F1 and PCEL-R1 in the corresponding sequences (table 1). From these partial sequences, antisense primers for 5′-RACE (Pg5′-1 and Hm5′-1) and for nested amplifications (Pg5′-2 and Hm5′-2) were designed (table 1). By the 5′-RACEs and the nested PCR amplifications which followed, 5′ regions of PgEGs and HmEGs were obtained (fig. 1b, c). The obtained DNA sequences of the 5′ regions encoded a putative N-terminal amino acid sequence homologous to that of the purified EG from the hindgut of *C. lacteus*. We designed sense primers (PgEx and HmEx) from the DNA sequences of the putative N-terminal coding regions (table 1) and conducted 3′ RACEs with these. Ten different cDNAs were identified for PgEGs and three for HmEGs. These sequences were aligned with the 5′-RACE sequences. As a result, PgEG1a (GenBank accession AB071001), HmEG1 (AB071011), HmEG2 (AB071012), and HmEG3 (AB071013) were identified with their region spanning from the putative start codons to downstream of the stop codons, while for the other cDNAs (PgEG1b–PgEG1h, AB071002–AB071010), the 5′-RACE did not reach either to the putative start codons or to the following signal-peptide-coding regions.

Origin of hindgut EGs

The single-cell RT-PCR method elucidated the origin of the cloned EG cDNAs in additions to clarifying their nucleotide sequences. Each protozoan cell from the hindgut of *C. formosanus* was carefully separated with a

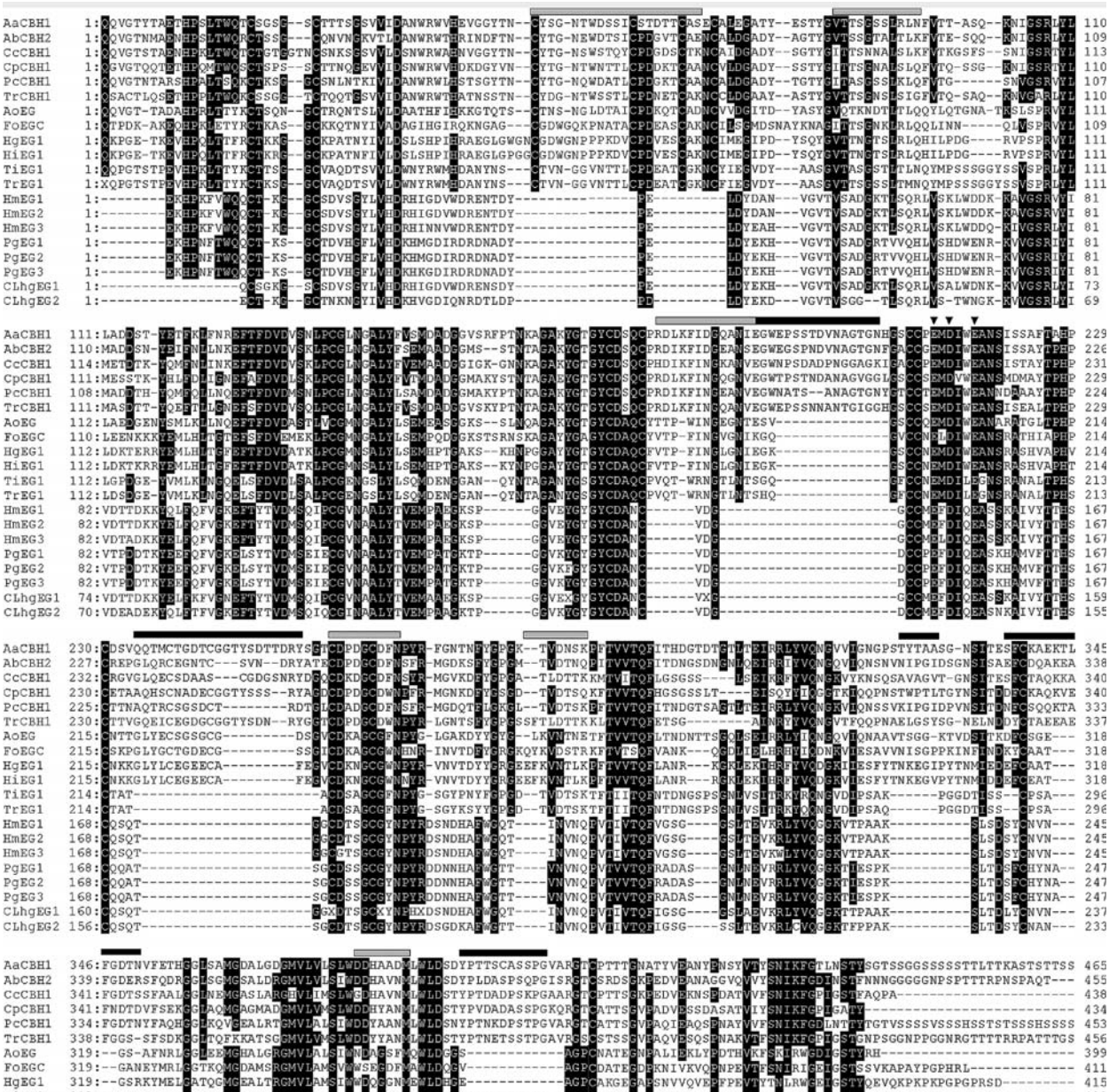

Figure 3. Multiple alignments of amino acid sequences of protozoan endo-β-1,4-glucanases and fungal cellulases from GHF7. Shaded characters indicate the coincidence of the majority of the amino acids in each column. The alignment methods are described in the text. CBH and EG in the abbreviated names refer to cellobiohydrases (EC 3.2.1.91) and endo-β-1,4-glucanases (EC 3.2.1.4), respectively. Closed wedges indicate important catalytic residues for GHF7 members corresponding to Glu212, Asp214, and Glu217 of TrCBH1. The GenBank accession numbers are O59843 (AaCBH1 from *Aspergillus aculeatus*), Q92400 (AbCBH2 from *Agaricus bisporus*), Q00328 (CcCBH1 from *Cochliobolus carbonum*), Q00548 (CpCBH1 from *Cryphonectria parasitica*), P13860 (PcCBH1 from *Phanerochaete chrysosporium*), P00725 [TrCBH1 from *Trichoderma reesei* (= *Hypocrea jecorina*)], BAA22589 (AoEG from *Aspergillus oryzae*), P46237 (FoEGC from *Fusarium oxysporum*), Q12622 (HgEG1 from *Humicola grisea*), P56680 (HiEG1 from *H. insolens*), Q12714 (TlEG1 from *T. longibrachiatum*), P07981 (TrEG1 from *Trichoderma reesei*, BAB64553, BAB64561, BAB64562 (PgEG1, PgEG2, PgEG3, respectively, from the hindgut protozoa, *Pseudotrichonympha grassii* of the termite, *C. formosanus*), BAB64563, BAB64564, BAB64565 (HmEG1, HmEG2, HmEG3, respectively, from the hindgut protozoa, *Holomastigotoides mirabile* of *C. formosanus*), BAB089800, BAB089801 (CLhgEG1, CLhgEG2, respectively, from the hindgut contents of the termite, *C. lacteus*).

pipette while keeping record of the species with a digital camera. RT-PCR was performed for each single protozoan cell, priming the first-strand cDNA with PCEL-F1, and amplifying with PCEL-F1 and PCEL-R1 (table 1, fig. 1b, c). The PCR products were then cloned and sequenced. All PgEG and HmEG homologues were amplified from *P. grassii* and *H. mirabile*, respectively. No cross-contamination or exceptions occurred, as verified by our recordings of our careful preparation of the protozoa. No amplification was obtained using background Solution U as a control sample. We could not obtain fresh samples of *C. lacteus* to perform the single-cell protozoan RT-PCR.

Sequence comparison of protozoan EG cDNAs

Cloned cDNAs from *P. grassii* and *H. mirabile* encoded PgEG1 (328 amino acids on PgEG1a), HmEG1, HmEG2, and HmEG3 (326 amino acids on each), and part of PgEG1 (from the putative N terminal to the putative C terminal, 318 amino acids on PgEG1b–PgEG1h). The putative amino-acid sequences encoded on the isolated cDNAs showed over 99% identity in nucleotides within species except for CLhgEG1 and CLhgEG2, and 61–62% between protozoan species. The putative proteins encoded by each cDNA had 1 amino acid substitution between each other, except for HmEG3, which had 13 substitutions against HmEG1 and 14 against HmEG2. The cDNAs obtained from the hindgut content of *C. lacteus* (CLhgEG1 and CLhgEG2) encoded 307 and 303 amino acids, respectively, corresponding from downstream of the N-terminal region used to design the degenerate primer (CL5′; table 1, fig. 1a) to downstream of the putative C terminal and stop codon. CLhgEG1 showed 76.9%, 88.6–89.9%, and 69.1–69.9% identity at the amino acid level to CLhgEG2, HmEGs, and PgEGs, respectively, while CLhgEG2 showed 68.9 % and 77.8–79.8% identity at the amino-acid level to PgEGs and HmEGs, respectively. By BLAST X homology searches, PgEGs, HmEGs, and CLhgEGs showed high homology to EGs and endo-β-1,4-glucanasecellobiohydrases (CBH; E.C. 3.2.1.91) both from GHF7 (fig. 3). The PgEGs, HmEGs and CLhgEGs encoded only putative catalytic domains. No other domains homologous to linkers or cellulose-binding domains of multi-domain cellulases were encoded. Specifically, the PgEGs, HmEGs and CL hindgut EGs showed 32–39% identity in amino acid sequence to TrEG1 (=Cel7B), and 37–43% identity to TrCBH1 (= Cel7A) from *Trichoderma reesei*, whose crystal structures have been firmly elucidated [29–31]. The protozoan and hindgut EGs lacked all of the loop forming regions and many of the shift loop-forming regions covering the catalytic clefts of TrCBH1 and TrEG1, respectively [29–31] (fig. 3). Phylogenetic analysis of protozoan EGs and CLhgEGs with other members of GHF7 showed that protozoan EGs, fungal EGs, and CBHs form distinctive clades (fig. 4).

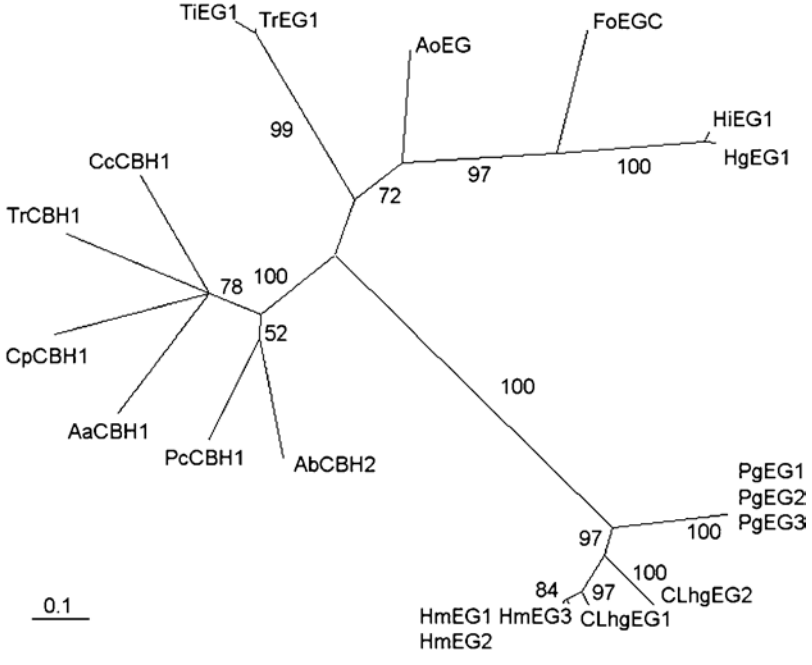

Figure 4. Phylogenetic relationships between protozoan endo-β-1,4-glucanases and GHF 7 members. The tree was constructed by the maximum-likelihood method with 282 aligned amino acid positions, corresponding to the putative catalytic domains of 12 fungal genes (6 endo-β-1,4-glucanase and 6 cellobiohydrolase genes) in GenBank and putative protozoan endo-β-1,4-glucanases. Numbers at the internal branches indicate the percent occurrence of a group in 1000 quartet puzzling steps. The scale bar indicates the number of substitutions per position as given by TREE-PUZZLE [28]. The GenBank (protein) accession numbers are given in the legend to figure 3.

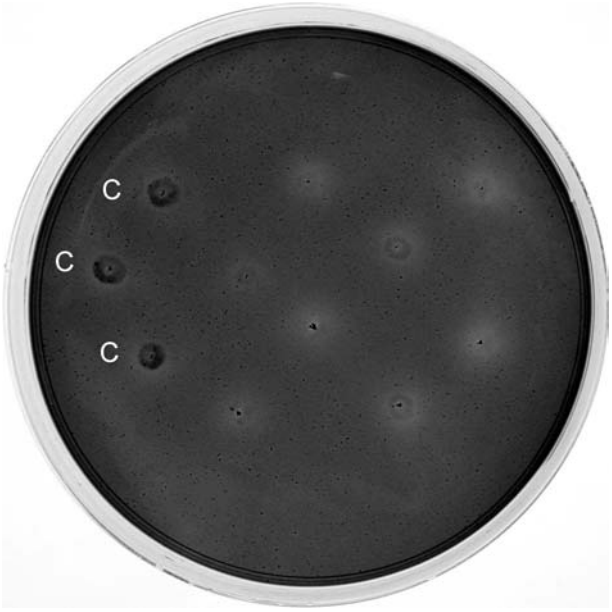

Figure 5. Expression of endo-β-1,4-glucanase in *Escherichia coli*. *PgEG1* cDNA from the hindgut protozoa, *Pseudotrichonympha grassii* of the termite *C. formosanus* was ligated into pMOSBlue and transformed into *E. coli*. The control (C) was transformed non-inserted pMOSBlue. The halo seen in the image on the right was made by the enzymatic degradation of CMC in agarose. The experimental conditions are described in the text.

Expression of the protozoan EG

To confirm the functionality of the protozoan EGs, one of the obtained EG cDNAs (HmEG3) was sub-cloned to a pMOSBlue cloning vector and the expression of a functional EG was confirmed by the formation of halos on an activity-stained CMC-LB plate (fig. 5). The control, transformed pMOSBlue without insert, did not create any recognizable halos (fig. 5).

Discussion

There have been several descriptions of *Holomastigotoides* sp. in *Coptotermes* [17, 18, 32]. In *C. formosanus,* '*Holomastigotoides hartmanni*' was originally identified as a solo *Holomastigotoides* species by Koizumi [18]. Recently, phylogenic analyses of the hypermastigotes in *C. formosanus* were performed by Ohkuma et al. [33], in which *Holomastigotoides* sp. was described as *H. mirabile*. Although the name *H. hartmanni* is still supported by many researchers, we have followed Ohkuma's description in order to avoid possible confusion on DNA databases.

The gel filtration using Sephacryl S-200 distinguished EGs from the salivary glands, the midgut, and the hindgut of *C. lacteus*. The simultaneous presence of endogenous and protozoan cellulase activities was first proposed by Yamaoka and Nagatani [11] based on their investigations of the salivary glands and hindgut of *R. speratus* from the family Rhinotermitidae of which *Coptotermes* species are members. Following this report, endogenous and protozoan activities were also identified in the digestive system of *C. lacteus* and chromatographically separated using a Superose 6B molecular sieving column [7]. Our present results coincide with these earlier studies.

In addition to the peak at the same fraction as that from the salivary gland extract, the *C. lacteus* midgut extract had a second active peak at a higher molecular weight, which was assumed to be an aggregate form of the protein present in the first peak (fig. 2). In termites, all known endogenous EG genes encode proteins consisting of 448 amino acids and their expression in the salivary glands and/or midgut has been established [1, 2, 4, 15].

Homologues to the termite endogenous EG genes have also been found in cockroaches, including omnivorous species, and a crayfish, which are evolutionarily cross-related to termites and to insects, respectively [4, 34].

The esophagus and proventriculus (foregut) included in the midgut extract are unlikely to be origins of EG activity, since no secretory cells or stable symbionts have ever been reported in these organs in insects as far as we know, and the foregut is generally considered a non-secreting organ [35]. Since the salivary glands are connected to the esophagus through the salivary canals, enzymes produced by both the salivary glands and midgut would contribute to the activity from the extract of foregut and midgut [15].

The purified hindgut EG and CLhgEGs from *C. lacteus* showed high homology to PgEGs and HmEGs, consisting of 315 and 318 amino acids, respectively. Therefore, the hindgut EGs of *C. lacteus* are expected to have similar molecular weights, and, thus, would be lower in molecular weight than the termite endogenous EGs with 432 amino acids without signal peptides [1, 2, 15], although the hindgut EG of *C. lacteus* was eluted earlier than those from its salivary glands and midgut from Sephacryl S-200 (fig. 2). This might again be explained by molecular clustering due to the experimental conditions.

The result demonstrated that *H. mirabile* and *P. grassii* (both from Hypermastigida) of *C. formosanus* have genes encoding GHF7 members, as is expected for the hindgut fauna of *C. lacteus*, which, like other *Coptotermes* species, harbors *Holomastigotoides* and *Pseudotrichonympha* [16, 17]. Possible prokaryote origins (externally attached spirochaeta, secondary bacterial or archeal intracellular symbionts) of the hindgut cellulases are unlikely since our experimental methodology recovered poly-(A)-tailed mRNA specific to eukaryotes. The negative result for the RT-PCR amplification of the background Solution U, which possibly contained those microorganisms except protozoa, also supports this conclusion. In another protozoon, *Trichonympha sphaerica*

(Hypermastigida) from the hindgut of *Zootermopsis* sp., cellulose digestion in the protozoon was concluded as being independent of its endosymbiotic bacteria [36]. The fact that no GHF7 members have ever been reported outside the Fungi further suggests that prokaryote origins for these genes are unlikely. The phylogenetic analysis of GHF7 members also supports the protozoan origin of PgEGs, HmEGs, and CLhgEGs, since a unique clade for them was recovered (fig. 4). We could not find GHF45 members or their cDNA in the hindguts of either *C. lacteus* or *C. formosanus*. This might be explained by the difference in protozoan fauna between *Coptotermes* and *Reticulitermes* termites. Many of the isolated GHF45 cDNAs were unidentified in terms of their origin, but some of them were attributed to *Teranympha mirabilis* and *Trichonympha agili* of *R. speratus* [3], which are absent in the hindgut of *Coptotermes* [16, 17].

GHF 7 consists of both EGs and CBHs [37] (http:afmb. cnrsmrs.fr/~pedro/CAZY/ghf. html). The three-dimensional structure of GHF7 has been elucidated for CBH1 and EG1 from *T. reesei* [29–31]. CBH1 from *T. reesei* has a β sandwich structure consisting of two large anti-parallel β sheets opposite to each other [29–31]. Our comparison of amino-acid sequences suggests that this overall structure of CBH1 seems to be conserved among GHF7 members including the protozoan EGs (fig. 3). CBH1 and EG1 from *T. reesei* are retention-type cellulases having two glutamate residues, one acting as a nucleophile and the other as a Bronsted acid/base catalyst, and one aspartic acid residue, which acts as a nucleophile supporter at the catalytic center (corresponding to Glu212, Glu217 and Asp214 of TrCBH1, respectively; fig. 3) [29–31]. These three residues are conserved in all protozoan (hindgut) EGs cloned in the present study and in the other GHF7 members compared with them (Fig. 3). These structural similarities and homologies of the protozoan EGs to other GHF7 members reported previously suggest that these putative proteins are functional and, furthermore, the functionality of one protozoan EG (*HmEG3*) was experimentally confirmed (fig. 5). The high homology between the N terminals of the putative proteins coded on the protozoan EG cDNAs from *C. formosanus* and the purified protozoan EG from *C. lacteus* also supports this conclusion.

The catalytic domain of CBH1 from *T. reesei* forms a tunnel-like structure comprised of two anti-parallel β sheets with loop-forming residues stitching the sheets together at both sides, and thus access to substrate cellulose chains is limited to the terminals [29]. However, in EG1 from *T. reesei*, less than half of the loop-forming amino acid residues are conserved in the sequence [30]. The stitching function of the missing residues is substituted by shift-loop-forming regions which form a more open structure at the catalytic center, thus allowing this enzyme to randomly access the middle of cellulose chains [30, 38]. This open versus closed structural difference in the catalytic centers determines the enzyme affinity to either crystalline or amorphous cellulose, respectively, since in crystalline cellulose, the only parts accessible to the enzymes are at the ends of the chains, while amorphous cellulose, like CMC, allows access to the middle of the chains if a cellulase has an open-form catalytic center [39]. In the termite protozoan EGs examined here, many of the amino acid residues corresponding to both the loop-forming and shift-loop-forming regions of CBH1/EG1 are absent or substituted (fig. 3), although there should be some residues which hold two β sheets together to form a catalytic cleft. Therefore, these protozoan enzymes would be randomly accessible to amorphous-form cellulose chains, as EGs usually are. This is also supported by the activities, against CMC as a specific substrate for EGs [40], of purified hindgut cellulase of *C. lacteus* and HmEG3 (a homologue of HmEG1, HmEG2, PgEGs, and CLhgEGs) expressed in *E. coli*. We previously demonstrated that the detectable amount of endogenous EGs stored in the midgut of *C. formosanus* does not flow into the hindgut and the crude extracts of the midgut and the hindgut hydrolyze crystalline cellulose independently [14]. Here, we conclude that the hindgut protozoa of *Coptotermes* termites produce EGs ('Cx-cellulase' [11]) by themselves.

Acknowledgements. This work was supported by the Promotion of Basic Research Activities for Innovative Biosciences Fund from the Bio-oriented Technology Research Advancement Institution (BRAIN; Omiya, Saitama, 331-8537 Japan; www.brain.go.jp) and by the Pioneer Research Project Fund (No. PRPF-0022) from the Ministry of Agriculture, Forestry and Fisheries of Japan. The research on *C. lacteus* was conducted with financial support by the Japan Science and Technology Corporation (JST), which allowed H. W. to stay at the Department of Biochemistry of the University of Sydney as a JST fellowship researcher from April 1998 to March 1999.

1 Watanabe H., Noda H., Tokuda G. and Lo N. (1998) A cellulase gene of termite origin. Nature **394:** 330–331
2 Tokuda G., Lo N., Watanabe H., Slaytor M., Matsumoto T. and Noda H. (1999) Metazoan cellulase genes from termites: intron/exon structures and sites of expression. Biochim. Biophys. Acta **1447:** 14–159
3 Ohtoko K., Ohkuma M., Moriya S., Inoue T., Usami R. and Kudo T. (2000) Diverse genes of cellulase homologues of glycosyl hydrolase family 45 from the symbiotic protists in the hindgut of the termite *Reticulitermes speratus*. Extremophiles **4:** 343–349
4 Watanabe H. and Tokuda G. (2001) Animal cellulases. Cell. Mol. Life Sci. **58:** 1167–1178
5 Slaytor M. (2000) Energy metabolism in the termite and its gut microbiota. In: Termites: Evolution, Society, Symbioses, Ecology, pp. 307–332, Abe T., Bignell D. E. and Higashi M. (eds.) Kluwer, Dordrecht
6 Krishna K. (1969) Introduction. In: Biology of Termites, vol. 1, pp. 1–13, Krishna K. and Weesner F. M. (eds) Academic Press, New York
7 O'Brien G. W., Veivers P. C., McEwen S. E., Slaytor M. and O'Brien R. W. (1979) The origin and distribution of cellulase in

the termites, *Nasutitermes exitiosus* and *Coptotermes lacteus*. Insect Biochem. **9:** 619–625
8 Breznak J. A. and Brune A. (1994) Role of microorganisms in the digestion of lignocellulose by termites. Annu. Rev. Entomol. **39:** 453–487
9 Slaytor M. (1992) Cellulose digestion in termites and cockroaches: what role do symbionts play? Comp. Biochem. Physiol. **103B:** 775–784
10 Cleveland L. R. (1924) The physiological and symbiotic relationships between the intestinal protozoa of termites and their host, with special reference to *Reticulitermes flavipes* Kollar. Biol. Bull. Mar. Biol. Lab. **46:** 117–227
11 Yamaoka I. and Nagatani Y. (1975) Cellulose digestion system in the termite, *Reticulitermes speratus* (Kolbe). I. Producing sites and physiological significance of two kinds of cellulase in the worker. Zool. Mag. **84:** 23–29
12 O'Brien R. W. and Slaytor M. (1982) Role of microorganisms in the metabolism of termites. Aust. J. Biol. Sci. **35:** 239–262
13 Hogan M. E., Schulz M. W., Slaytor M., Czolij R. T. and O'Brien R. W. (1988) Components of termite and protozoal cellulases from the lower termite, *Coptotermes lacteus* Frogatt. Insect Biochem. **18:** 45–51
14 Nakashima K. and Azuma J. (2000) Distribution and properties of endo-beta-1,4-glucanase from a lower termite, *Coptotermes formosanus* (Shiraki). Biosci. Biotechnol. Biochem. **64:** 1500–1506
15 Nakashima K., Watanabe H., Saitoh H., Tokuda G. and Azuma J. I. (2002) Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. Insect Biochem. Mol. Biol. **32:** 777–784
16 Kitade O. and Matsumoto T. (1998) Characteristics of the symbiotic flagellate composition within the termite family rhinotermitidae (Isoptera). Symbiosis **25:** 271–278
17 Yamin M. A. (1979) Flagellates of the orders Trichomonadida Kirby, *Oxymonadida grasse* and *Hypermastigida grassi* & *foa* reported from lower termites (Isoptera families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, and Serritermitidae) and from the wood-feeding roach *Cryptocercus* (Dictyoptera: Cryptocercidae). Sociobiology **4:** 3–119
18 Koizumi M. (1921) Studies on the intestinal protozoa found in the termites of Japan. Parasitology **13:** 235–310
19 Fukatsu T. (1999) Acetone preservation: a practical technique for molecular analysis. Mol. Ecol. **8:** 1935–1945
20 Jue C. K. and Lipke P. N. (1985) Determination of reducing sugars in the nanomole range with tetrazolium blue. J. Biochem. Biophys. Methods **11:** 109–115
21 Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680–685
22 Hirano H. and Watanabe T. (1990) Microsequencing of proteins electrotransferred onto immobilizing matrices from polyacrylamide gel electrophoresis: application to an insoluble protein. Electrophoresis **11:** 573–580
23 Tokuda G., Saito H. and Watanabe H. (2002) A digestive β-glucosidase from the salivary glands of the termite, *Neotermes koshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5′- and 3′-RACE amplifications with degenerate primers. Insect Biochem. Mol. Biol. **32:** 1681– 1689
24 Trager W. (1934) The cultivation of a cellulose-digesting flagellate *Trichomonas termopsidis* and of certain other termite protozoa. Biol. Bull. **66:** 182–190
25 Ausubel F. M., Brent R., Kingston R. E., Moore D. D., Seidman J. G., Smith J. A. et al. (1989) Current Protocols in Molecular Biology, Wiley, New York
26 Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402
27 Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990) Basic local alignment search tool. J. Mol. Biol. **215:** 403–410
28 Strimmer K. and Von Haeseler A. (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:** 964–969
29 Divne C., Stahlberg J., Reinikainen T., Ruohonen L., Pettersson G., Knowles J. K., et al. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. Science **265:** 524–528
30 Kleywegt G. J., Zou J. Y., Divne C., Davies G. J., Sinning I., Stahlberg J. et al. (1997) The crystal structure of the catalytic core domain of endoglucanase I from *Trichoderma reesei* at 3.6 Å resolution, and a comparison with related enzymes. J. Mol. Biol. **272:** 383–397
31 Divne C., Stahlberg J., Teeri T. T. and Jones T. A. (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. J. Mol. Biol. **275:** 309–325
32 Grassi G. B. (1917) Flagelati viventi nei termiti. Att. Nazil. Lincei. Mem. **12:** 331–394
33 Ohkuma M., Ohtoko K., Iida T., Tokura M., Moriya S., Usami R. et al. (2000) Phylogenetic identification of hypermastigotes, *Pseudotrichonympha*, *Spirotrichonympha*, *Holomastigotoides*, and parabasalian symbionts in the hindgut of termites. J. Eukaryot. Microbiol. **47:** 249–259
34 Lo N., Tokuda G., Watanabe H., Rose H., Slaytor M., Maekawa K. et al. (2000) Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. Curr. Biol. **10:** 801–804
35 Chapman R. F. (1998) The Insects. Cambridge University Press, Cambridge
36 Yamin M. A. (1980) Cellulose metabolism by the flagellate *Trichonympha* from a termite is independent of endosymbiotic bacteria. Science **211:** 58–59
37 Henrissat B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:** 309–316
38 Rouvinen J., Bergfors T., Teeri T., Knowles J. K. and Jones T. A. (1990) Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249:** 380–386
39 Teeri T. T. (1997) Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. Trends Biotech. **15:** 160–167
40 Klesov A. A. (1991) Biochemistry and enzymology of cellulose hydrolysis. Biokhimiya **55:** 1295–1318