Review

Biofpr
Biofuels Bioproducts & Biorefining

Termite digestomes as sources for novel lignocellulases

Michael E. Scharf, University of Florida,Gainesville, FL
Aurélien Tartar, Nova Southeastern University, Fort Lauderdale, FL

Received July 22, 2008; revised version received September 2, 2008; accepted September 8, 2008
Published online in Wiley InterScience (www.interscience.wiley.com); DOI: 10.1002/bbb.107;
***Biofuels, Bioprod. Bioref*. (2008)**

Abstract: For most animals, lignocellulose is a nutritionally poor food source that is highly resistant to enteric degradation. Termites, however, have the unique ability to digest lignocellulose with high efficiency, often using it as a sole food source. Another interesting aspect of termite biology is their symbiotic associations with prokaryotic and eukaryotic gut symbionts. Termite symbionts contribute to lignocellulose digestion efficiency, but by no means are they responsible for 100% of lignocellulose digestion in the termite gut. The termite *digestome* can be defined as the pool of genes, both termite and symbiont, that contribute to lignocellulose depolymerization and digestion, as well as simple sugar fermentation, nutrient transport, and nutrient assimilation. A central goal of termite *digestomics* research is to define/understand the relative contributions of termite and symbiont gene products to collaborative lignocellulose digestion. While efficient microbial cellulases have already been identified and are presently being used in industrial applications, efficient and inexpensive pre-treatments for lignin and hemicellulose depolymerization are not yet well developed. In this respect, termite digestomics has already offered significant insights, and can continue to identify relevant enzymes, as well as reveal how to optimally combine and utilize these enzymes for maximum synergy. The topics covered in this review are as follows: lignocellulose structure with emphasis on its potential for depolymerization by termite and gut endosymbiont-derived digestive enzymes; termite biology and ecology from the perspectives of termite nutrition, gut physiology, and lignocellulose digestion; and trends identified through recent termite digestomics research. © 2008 Society of Chemical Industry and John Wiley & Sons, Ltd

Keywords: biomass; lignase; hemicellulase; cellulase; xylanases; bioethanol

Introduction

Lignocellulose is a sustainable global resource with a great deal of relevance to renewable energy production.[1,2] It is a naturally occurring complex of plant-derived materials that includes the β-1,4-linked sugar polymers cellulose and hemicellulose, and the phenolic polymer lignin.[3] In plants, lignocellulose provides key structural support for cell walls. Because it is plant-derived, lignocellulose is the most abundant and widespread bioenergy feedstock available on Earth. However, a major limitation in plant biomass utilization as a renewable energy source is the inefficiency of industrial lignocellulose depolymerization.[1,2,4] This inefficiency increases energy inputs,

Correspondence to: Michael E. Scharf, Entomology and Nematology Department, PO Box 110620, University of Florida, Gainesville, FL 32611-0620, USA. E-mail: mescharf@ufl.edu

SCI

reduces product yields, drives production costs higher, encourages political scepticism, and ultimately limits acceptance of cellulose-based renewable bioenergy.[1,2]

With respect to the problem of lignocellulose recalcitrance, it is germane that a number of invertebrate animals, and to some extent, their symbiotic gut fauna, have evolved specialized enzymes that cooperate in lignocellulose processing.[5–9] In particular, endogenous lignocellulases encoded in marine and terrestrial invertebrate genomes can often confer high degrees of digestion capabilities to these organisms.[9–11] When endogenous insect lignocellulases work synergistically with symbiont-derived enzymes, this can confer extremely high efficiency in lignocellulose processing.[12,13] Termites (order Isoptera) are one of the most well recognized examples of an organism that subsists on lignocellulose; and thus, lignocellulase enzymes from termites and their gut symbionts have many potential bioenergy applications that warrant careful consideration.

Here, we define a new term, *termite digestome,* as the pool of genes, both termite and symbiont, that contribute to lignocellulose depolymerization and digestion, as well as simple sugar fermentation, nutrient transport, and nutrient assimilation. Due to recent advances in genomics and metagenomics, the mining of termite and symbiont genomes, or *digestomics*, is now an extremely approachable and fruitful area of science. Here, we use the more encompassing term *digestomics* rather than *metagenomics*, which is more concerned with sampling microbial genes from environments such as the termite hindgut lumen. This review focuses on recent findings of digestomics research from termites and their mutualistic gut fauna.

Lignocellulose: structure, degradation and depolymerization

Lignocellulose can serve as a starting material for a number of industrial biorefinery processes, namely pyrolysis, gasification and hydrolysis.[1] All three processes currently rely on acid and/or heat pre-treatment (= energy-intensive); however, energy and cost inputs into the latter process of hydrolysis can be greatly reduced by implementation of recombinant lignocellulases. It is logical to assume that enzymes derived from insects that feed on lignocellulose diets, such as termites, are relevant for this purpose. The search for termite and symbiont lignocellulases requires an understanding of industrial lignocellulose uses, lignocellulose structure, and lignocellulose depolymerization. These topics are overviewed in the following section.

Lignocellulose structure

Lignocellulose is a general term referring to a natural complex of the three biopolymers *cellulose*, *hemicellulose* and *lignin*. The different categories of monomeric lignocellulose building materials are shown in Fig. 1. Cellulose is composed of rigid, high molecular weight β-1,4-linked polymers of glucose that are held together in bundles by hemicellulose.[1,2,5] Hemicellulose is composed of shorter β-1,4-linked polymers of mixed sugars. Mannose is usually the dominant sugar present in hemicelluloses of softwoods fed upon by termites, with lesser amounts of xylose, galactose, rhamnose, arabinose, glucuronic acid, mannuronic acid and galacturonic acid.[14] Lignin is not a carbohydrate, but a 3-dimensional polymer of phenolic compounds that are linked to each other and to hemicellulose by ester bonds.[4] Lignin is composed of the three 'mono-lignol' monomers *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol combined in different ratios depending on the plant species.

Another noteworthy aspect of hemicellulose is its high degree of esterification with monomers and dimers of phenolic acids, which are analogous to the mono-lignols noted above.[4,14] Figure 2 provides an example of one such phenolic acid ester; in this case, feruloylated arabinose. Phenolic acid esters are derived mostly from the mono-lignols *p*-coumaryl and coniferyl alcohol (i.e., coumaric acid and ferulic acid; see Fig. 1).

The three individual lignocellulose components cellulose, hemicellulose and lignin, respectively, compose approximately 40, 25 and 20% of lignocellulose.[1] The remaining 15% consists of minor components that include proteins, terpenic oils, fatty acids, fatty acid esters, and inorganic components that include nitrogen, phosphorous and potassium.[1]

Lignocellulose degradation and depolymerization

The biologically mediated degradation of lignocellulose into fermentable or otherwise utilizable sugars is a complex process that requires a diversity of enzymes. As a first step in the process, depolymerization of hydrophobic lignin polymers is extremely important to enable hemicellulose

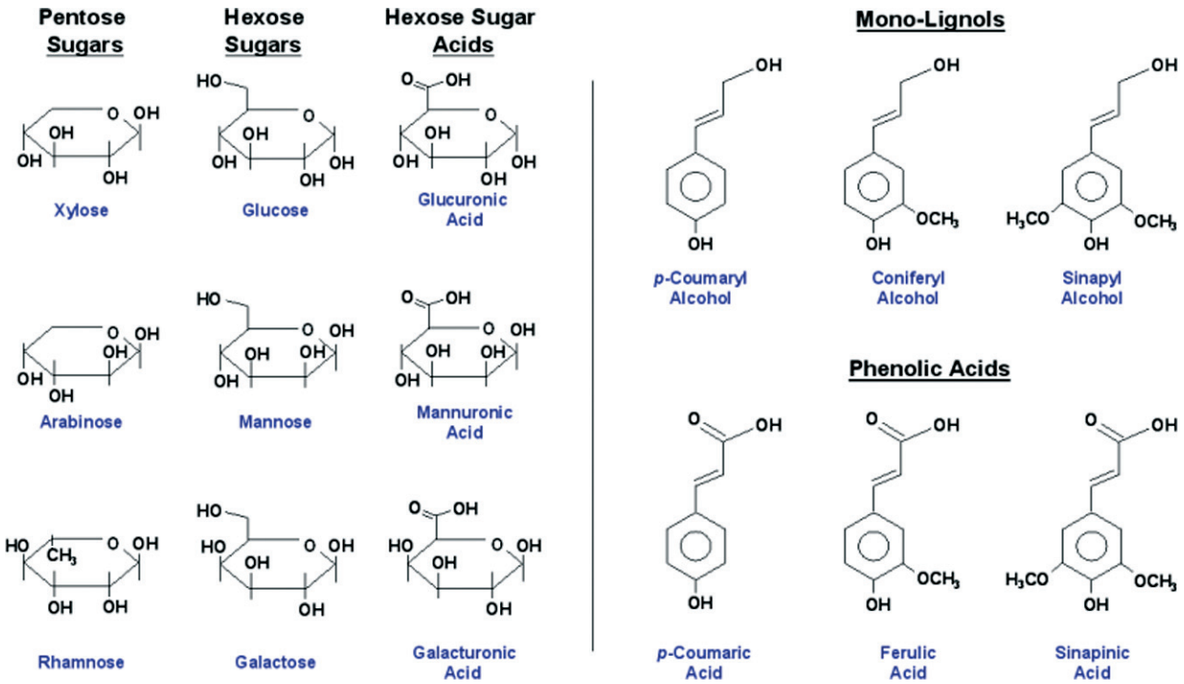


Figure 1. Monomeric building blocks of lignocellulose: carbohydrates (left) and lignin phenolics (right).

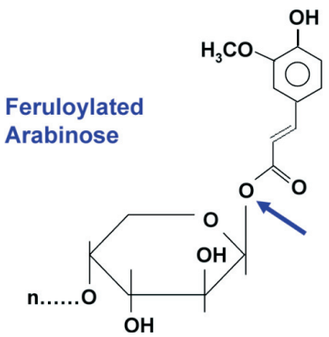


Figure 2. Structure of feruloylated arabinose, a conjugate of the phenolic acid ester ferulic acid with arabinose, as occurs in hemicellulose. Arrow indicates the critical carboxyl-ester bond. Carboxyl-esterases that hydrolyze this ester bond are considered critical with respect to hemicellulose degradation.

degradation. Because lignin is not a carbohydrate, carbohydrolases such as cellulases and hemicellulases theoretically should play no roles in lignin degradation /depolymerization. It is also important to note that lignin breakdown requires oxygen,[5,15] which is supported by evidence that termite guts are not completely anaerobic environments.[16] Indeed, lignin degradation/modification has been documented in termite guts (see later).[5,15,17–22] Relevant enzymes involved in lignin oxidation include laccases and peroxidases, such as the highly effective forms that are known to exist in lignin degrading fungi.[4,8,23,24] Other seemingly relevant enzymes could include enzymes classically considered in xenobiotic metabolism or defense, such as alcohol dehydrogenase, catalase, superoxide dismutase, cytochrome P450, epoxide hydrolase, reductase, glutathione-S-transferase, esterase, *etc*.[25–27] However, for any of these enzymes to act upon lignin in the termite gut, they would likely need to be secreted into the gut lumen. This may preclude cytochrome P450s, which are not secreted and typically are membrane-bound.

The second step in the overall lignocellulose digestion process, hemicellulose degradation, is critically important for making cellulose accessible for depolymerization.[14] Complete biodegradation of hemicellulose requires the combined activity of endo- and exo-β-1,4-xylanases, β-xylosidases, α-arabinofuranosidases, α-uronidases, and esterases such as acetylxylan esterase, ferulic acid esterase and *p*-coumaric acid esterase.[14] Specific conversions accomplished by each of these enzymes are as follows: (i) endo-xylanases hydrolyze

the β-1,4-xylose linkages in the xylan backbone; (ii) exo-xylanases hydrolyze reduced β-1,4-xylan linkages releasing xylobiose; (iii) β-xylosidases act on xylobiose to liberate xylose and other short chain oligosaccharides; (iv) α-arabinofuranosidases hydrolyze terminal non-reducing α-arabinofuranose from arabinoxylans; (v) α-uronidases release α-glucuronic, -mannuronic and -galacturonic acids; and (vi) esterases hydrolyze phenolic ester bonds (e.g., Koseki *et al.*[28]), namely those associated with acetyl xylans, ferulic acid xylans and *p*-coumaric xylans.

As a third step, cellulose depolymerization requires the action of three primary enzymes that include endo-β-1,4-glucanases, exo-β-1,4-glucanases, and β-glucosidases.[2,5] Specific conversions accomplished by each of these enzymes are as follows: (i) endoglucanases hydrolyze β-1,4-glycosyl linkages in the primary cellulose backbone releasing glucose, cellobiose, cellotriose or other longer oligomers; (ii) exoglucanases or 'cellobiohydrolases' target the terminal regions of polymeric chains to liberate either glucose or cellobiose; and finally (iii) β-glucosidases, then, hydrolyze cellobiose and cellotriose to liberate glucose monomers. As noted by Breznak & Brune[5] complete or nearly complete hydrolysis of cellulose typically requires synergistic collaboration by each of these three types of cellulases.

Clearly, biological degradation and depolymerization of lignocellulose are highly complex processes that involve a wide array of enzymes. Because of the complexity of these enzyme systems, it has generally been considered impractical to attempt to characterize lignocellulose degradation from termites (and other insects) by a one-dimensional approach involving enzyme biochemistry alone.[5] In this respect, the powerful 'reverse-genetics' approach offered by digestomics and metagenomics can literally allow for all relevant enzymes to be identified from a single, carefully executed sequencing project. Subsequently, digestive biochemistry and physiology can readily be deduced from comprehensive sets of translated gene sequence and then specifically characterized through hypothesis-driven, functional genomics research.

Termite biology, ecology and digestive physiology

Historically, termites have been considered important insects for reasons relating equally to ecology[29] and pest status.[30] For example, termites play important ecological roles in the decomposition of lignocellulosic plant materials and global carbon cycling; however, they also cause substantial economic damage to human structures and commodities on a global scale. More recently, with new emphasis being placed on renewable bioenergy, termites are considered as one of the most important bioreactors on the planet.[8]

Based mostly on the presence or absence of cellulolytic protozoa in the hindgut, termites are separated into the two subgroups of lower and higher termites, respectively.[31] Termite cellulose digestion was long thought to rely solely on microbial gut symbionts.[5,31,32] Evidence that began to shift this paradigm came from biochemical studies that identified cellulase activity from symbiont-free salivary gland and foregut extracts.[33] Subsequently, endoglucanase-encoding genes were discovered from a termite genome, verifying that termites do indeed produce their own 'endogenous' cellulases.[34] More recently it has been shown that symbiotic- and termite-derived cellulase gene transcripts are both present in termites,[12,35–39] supporting the contention that cellulases of both endosymbiont and termite origin are important to termite cellulose digestion; a fact that has mostly been overlooked[40,41] in recent termite gut metagenomics research. With respect to lignin and hemicellulose digestion, the relative roles of endogenous and symbiotic genes and proteins remain undefined; however, it is clear that cellulose, hemicellulose, and lignin are all significantly degraded as they pass through the termite gut (e.g., Breznak *et al.,* Shuji *et al.,* Katsumata *et al.* [5,21,22] and references therein).

The termite gut (Fig. 3), particularly the anaerobic fermentation chamber of the hindgut, is a rich source of microbial diversity. Micro-organisms from diverse taxa, including bacteria/spirochetes and protists are present in the termite gut.[5] In lower termites such as *R. flavipes*, eukaryotic symbionts are considered to contribute significantly to cellulose degradation while bacteria are considered important to acetogenesis, methanogenesis, nitrogen economy, O_2 physiology, and monosaccharide fermentation. In higher termites, many cellulolytic bacteria have been identified. Spirochetes, which are very difficult to culture, are found in large numbers in the termite hindgut. Spirochetes play roles in acetogenesis and nitrogen fixation in the lower termite hindgut,[42,43] and also occur as cytoplasmic symbionts of hindgut protists.[45,46]

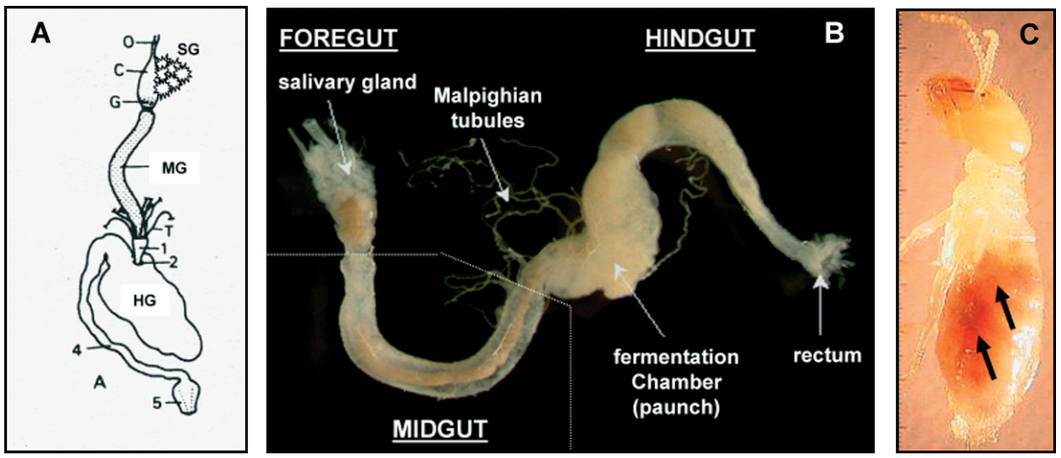


Figure 3. Morphology and features of the *R. flavipes* gut. (A & B) Drawing and photo showing different regions of the termite gut, O (oesophagus); C (crop); G (gastric cecum); MG (midgut); T (Malpighian tubules); HG (hindgut). Shown in (C) is the external morphology of a worker termite, in which arrows denote the highly visible fermentation chamber of the hindgut. Drawing in A is modified from Wood & Johnson.[36] Gut photo by JA Smith. Termite photo by ME Scharf.

Worker termites compose the majority of individuals in termite colonies and also perform the overwhelming majority of feeding and lignocellulose digestion. The worker termite gut is composed of three main regions: foregut, midgut and hindgut.[5,31] The foregut region consists of the oesophagus, crop and attached salivary glands. The salivary glands secrete endogenous endoglucanases and other relevant enzymes into the digestive tract. The midgut is a slender tubular region that secretes a peritrophic membrane around food materials, and presumably is a location where some lignocellulose degradation/depolymerization occurs. The midgut of higher termites is also known to secrete endoglucanases.[44] Analogous to the vertebrate kidney, malpighian tubules connect at the junction of the midgut and hindgut, and participate in waste excretion and nitrogen recycling. The hindgut consists of a fermentation chamber or 'paunch' that is generally anaerobic, but does possess a micro-oxic zone around its periphery.[16,42] The hindgut houses the majority of gut symbionts, and is the location where most cellulose degradation, as well as fermentation occurs. In terms of where nutrient assimilation takes place, there are no clear answers at the present time.

Termite digestive genomics research

To date, there have been three examples of digestive genomics research in termites. This work includes symbiont metagenomic sequencing the Japanese termite *Reticulitermes speratus*,[40] termite/symbiont digestome sequencing in the US termite *Reticulitermes flavipes*,[35,36,39,47] and symbiont metagenome sequencing in an undescribed *Nasutitermes* species from Costa Rica.[41]

Reticulitermes speratus metagenomics

R. speratus is a lower termite, meaning they possess cellulolytic protists in their digestive tracts, and that workers are temporally arrested juveniles (e.g., Zhou *et al.*[48]). A first example of termite digestive genomics research was reported by Todaka *et al.*[40] who sequenced 910 clones from an 'environmental' cDNA library prepared from pooled hindgut symbiont samples of *R. speratus*. These 910 sequences clustered into 580 tentative genes that were annotated based on homology to known sequences in public databases. Cellulases containing motifs from ten glycosyl hydrolase (GH) families (3, 5, 7, 8, 10, 11, 26, 43, 45 and 62) were identified, with GHF7 being the best represented (Table 1). No lignase-coding genes were identified. Of the nine GHF7 members identified, three encoded cellobiohydrolases/exoglucanases and six encoded endoglucanases. Additionally, the GHF7 findings were verified by N-terminal sequencing of dominant ~45 kDa proteins from symbiont protein extracts. These proteomic results revealed sequence motifs present

in GHF7 gene sequences that were identified from library sequencing, leading the authors to conclude that GHF7 carbohydrolases are major expressed proteins by the symbiotic protistan system of *R. speratus*, and that they are likely active against crystalline cellulose.

In a broader context, with ten GHFs (glycosyl hydrolase families) being represented, these *R. speratus* results provide a glimpse into the diversity of hindgut symbiotic protist cellulases, with GHF7 being the most diverse. A number of xylanolytic enzymes were also identified in this study, but did not receive additional attention. Thus, protistan symbionts of *R. speratus* clearly confer strong cellulolytic and some xylanolytic capabilities. However, no apparent lignase-coding genes were identified from these studies, supporting the hypothesis that lignases are termite-derived, or at least not derived from gut endosymbionts.

Reticulitermes flavipes digestomics

A second digestive genomics example from lower termites comes from the targeted sequencing of a 'polyphenic' cDNA library[35,36] prepared from whole *R. flavipes* termites of multiple castes.[49] The goal of this work was not specifically to identify lignocellulases, but to use cDNA arrays to identify differentially expressed genes among different caste phenotypes. Because the library included symbiont genes, cDNA arrays revealed both endogenous and symbiont-derived cellulase genes specifically expressed in the worker caste.[35,36] Several dozen ESTs were identified encoding cellulolytic enzymes. These ESTs were assembled into four contiguous partial open reading frame (ORF) sequences. Full-length sequences for these genes were subsequently obtained by aligning contiguous expressed sequence tag (EST) sequences, primer walking, and 5' rapid amplification of cDNA ends (RACE).[39] Most likely, the abundance of these four cellulases and the ability to identify them using an array-based approach underscores their overall abundance and relative importance to cellulose depolymerization.

Two of the *R. flavipes* cellulase genes have significant translated homology to endoglucanases, while two share significant homology with exoglucanases and xylanases. One of the endoglucanases (named *Cell-1*), a glycosyl hydrolase family 9 (GHF9) member, has significant homology to cellulases identified from termite, cockroach, and bivalve genomes.[39] Alternatively, the other three cellulases (named *Cell-2*, *-3* and *-4*) are all GHF7 members and have strong homology to endoglucanases (*Cell-2*) and exoglucanases (*Cell-3* and *-4*) identified previously from termite gut symbionts.[12,40,50,51] The endogenous and symbiotic origins for these four genes were further validated by defaunation studies in which UV light was used to eliminate termite hindgut symbionts.[38] Following UV treatment, no reduction in the endogenous *Cell-1* transcript levels occurred; however, transcript abundance for the symbiotic *Cell-2*, *Cell-3* and *Cell-4* genes all declined significantly, confirming their symbiotic origins. Additionally, knowledge of these genes and corresponding proteins enabled a novel line of research into prototype cellulase inhibitors and RNA interference for termite control.[52,53]

Based on sequencing results and examinations of cellulase activity across the gut, it was concluded that the four genes, each having different amino acid compositions, likely play collaborative or synergistic roles in cellulose digestion.[39,52] The existence of such a collaborative system, in which host and symbiont enzyme systems complement each other, is in contrast to previous conclusions of a 'dual' system in which symbionts work independently of the host.[12,40,54]

In an effort to better understand collaborative lignocellulose digestion, more comprehensive sequencing efforts have now been completed in *R. flavipes*.[47] For this purpose, over 11 000 ESTs were sequenced from the termite gut transcriptome and the gut symbiont meta-transcriptome (Genbank Accession numbers FL634956-FL640828 and FL641015-FL645753).[47] The combined *R. flavipes* sequencing efforts have identified GHF members from 28 families (Table 1), distinct host/symbiont profiles, and several endogenous termite lignase candidates.[47] The putative termite lignases identified include laccases, peroxidases, alcohol dehydrogenases, glutathione peroxidases, glutathione transferases, esterases/carboxylesterases, and epoxide hydrolases.[47]

These findings, identifying apparent *R. flavipes*-derived lignase genes, support previous findings that the *R. flavipes* gut can metabolize the phenylpropanoids benzoate and cinnamate as rapidly as cellulose.[15] Our identification of candidate lignase genes also agrees with earlier biochemical findings showing various degrees of modification of lignin, mono-lignol, and/or phenolic compounds in the

guts of diverse termites, including *R. flavipes*.[5,17–22] Finally, numerous genes encoding antioxidant enzymes (e.g., superoxide dismutase, catalase, alcohol dehydrogenase, glutathione transferase) were also identified from the *R. flavipes* gut digestome.[47] Because lignin degradation is known to result in free radical generation,[5] the expression of diverse antioxidant enzymes in the *R. flavipes* gut provides further evidence of lignin degradation capabilities.

Nasutitermes metagenomics

A third digestive genomics study was recently reported by Warnecke *et al.*[41] As stated by the authors, this metagenomic study 'illustrates how complex a 1-µl environment can be' and that bacterial symbionts of higher termites can reveal previously unknown enzymes and illuminate strategies for conversion of cellulose to biofuels.[41]

This study represents the most robust digestive genomics effort in a termite, and indeed, one of the most comprehensive metagenomic screens conducted in any organism to date. The termite studied in this research was a termite from the genus *Nasutitermes* collected from arboreal nests in Costa Rica; it is an undescribed species that is phylogenetically close to *N. ephratae* and *N. corniger*. Higher termites, such as the *Nasutitermes,* lack protozoan symbionts in their digestive tracts, but instead secrete their own endogenous cellulases and host cellulolytic symbiotic bacteria.[37,44,55–58]

The *Nasutitermes* study used a combination of novel pyrosequencing and traditional Sanger sequencing on whole-metagenome shotgun libraries. The libraries were prepared from symbiont genomic DNA samples isolated from the anterior 'P3' region of the hindgut lumen. This research also included a phylogenetic survey based on 16S rDNA sequence, proteome sequencing, and functional expression of recombinant enzymes. The 16S sequence survey revealed an unexpectedly broad diversity of bacteria that included 12 phyla and 216 phylotypes. The dominant taxa were from the genus *Treponema* and phylum Fibrobacteres, suggesting that these two taxa were the source of the majority of genomic DNA that was analyzed. It is not clear at this time if all termites harbour such bacterial symbiont diversity.

The *Nasutitermes* metagenomic sequencing can be summarized as follows. Two genomic shotgun libraries, one small insert library (2–4 kb inserts) and one large fosmid library (~32 kb inserts), were sequenced by conventional Sanger protocols. In combination, these two libraries produced 105 084 reads containing 71.38 Mb of sequence. A second approach involved isolation of 15 fosmid library clones for full sequencing using novel pyrosequencing technology. Pyrosequencing yielded 3 million reads of ~100 bp, or 30 million base pairs of DNA. A third and final approach involved pooling and shearing of the fosmid DNA, subcloning it into pUC vector, and end-sequencing 7680 clones; this latter procedure assisted with the assembly of pyrosequencing contigs into fosmid fragments (7 complete and 8 partial).

From this sequence pool, no potential lignases were identified, further supporting the hypothesis that lignin degradation and/or depolymerization are not conferred by hindgut endosymbiota. This sequence dataset, however, did reveal over 200 genes or gene fragments involved in cellulose and/or hemicellulose hydrolysis from 45 glycosyl hydrolase families (GHFs) (Table 1). Interestingly, the protistan symbiont GH family 7 noted above for *Reticulitermes* was absent from the *Nasutitermes* dataset. Also interestingly, GH family 9, the family represented mostly by endogenous animal endoglucanases, had 9 representatives in the *Nasutitermes* bacterial symbiont dataset; they are apparently Fibrobacter-derived. Approximately half of the GHFs identified from the *Nasutitermes* hindgut metagenome were from at least ten distinct families; the most prevalent of these groups were as follows: GHF3 (69 members), GHF94 (68), GHF5 (56), GHF 23 (52), GHF13 (48), GHF10 (46), GHF31 (26), GHF2 (23), GHF1 (22). However, hindgut proteomic analyses revealed that only a fraction of this diversity may be translated and secreted into the gut lumen, suggesting that many of these genes may be repressed or weakly expressed, and therefore may have limited significance in cellulose digestion.

Many other noteworthy findings were also revealed by the *Nasutitermes* metagenomic dataset, such as identification of 31 putative proteins with carbohydrate binding functions, 159 iron-only hydrogenases, 14–37 variants of all but one component (formate dehydrogenase) of the Ljungdahl CO_2-reductive acetogenesis pathway, novel ferredoxin oxidoreductases and membrane proteins, a rich diversity

of genes encoding nitrogen fixation homologs from many different categories, and finally, about 1500 genes encoding apparent chemotaxis and chemosensation functions. This latter category of genes is predicted to be involved in gut re-colonization after molting (via trophallactic transfer of symbionts from donors to recipients), and/or after regular peristaltic mixing and redistribution of gut contents.

Emerging trends from termite digestomics

Although termite gut digestomics is a relatively new area of research, a cursory comparison of existing genomic datasets reveals some important trends. First, the protistan symbiont glycosyl hydrolase family (GHF) 7 identified in both *R. speratus* and *R. flavipes* was conspicuously absent from the *Nasutitermes* hindgut bacterial metagenome dataset. This provides convincing evidence supporting that GHF7 is truly a termite protistan-affiliated GHF. [12,40,50,51] Second, eight total GH families were identified from the two *Reticulitermes* species that are not represented in the *Nasutitermes* hindgut metagenome data set (families 7, 30, 37, 47, 62, 70, 76 and 85; Table 1). There are 34 GH families identified thus far that are represented in *Nasutitermes* and absent in *Reticulitermes* (Table 1). The most noteworthy of these *Nasutitermes*-specific GHFs are the GHF94 members, of which 68 distinct genes or gene fragments were identified (Table 1). These GHF94 members are classified as cellobiose and cellodextrin phosphorylases; they presumably act intracellularly on substrates after transport into bacterial cytoplasm.[41]

It is also interesting that GHF9, the family composed mostly of endogenous animal endoglucanases, including many from *Reticulitermes* and a number of other higher and lower termites, had nine apparent fibrobacter-derived members in the *Nasutitermes* hindgut metagenome. The identification of bacterial GHF9 members in *Nasutitermes* bacterial symbionts is in agreement with recent analyses suggesting ancient origins for GHF9 endoglucanases.[10] With time and further functional analyses, the significance of these various asymmetrically distributed GH families among higher and lower termites will undoubtedly be more clearly resolved. These kinds of comparative genomics investigations will be an important direction for future termite digestomics research.

Also, as discussed above, lignin must first be degraded and/or at least partially depolymerized before cellulose and hemicellulose digestion can occur. In this respect, it bears re-emphasizing that no lignases have been identified from any of the *R. flavipes*, *R. speratus*, or *Nasutitermes* symbiont datasets. This paucity of lignases within symbiont sequence pools is contrasted by recent sequences obtained from a *R. flavipes* gut tissue library, which reveal a number of likely termite-derived lignases; in particular, laccases and peroxidases.[47] Both laccases and peroxidases from fungi play well-documented roles in lignin degradation;[4,8,23] and some are even produced by fungal cultivars of higher termites.[24] Because termite guts are not completely anoxic, but actually contain well-defined micro-oxic zones,[16] this supports that significant lignin degradation can take place in the termite gut,[15,21] and suggests legitimate roles for recently identified lignase gene candidates in the *Reticulitermes* gut.[47]

Thus, from termite digestomics research, a clearer picture of collaborative lignocellulose digestion is now emerging (Fig. 4). Perhaps most importantly, from *R. flavipes* sequencing work[47] and recent unpublished functional studies, a trend is emerging suggesting collaboration among termite-derived genes expressed in the salivary gland/foregut and midgut, and symbiont genes expressed in the hindgut. Specifically, in relation to individual gut regions, there is now evidence implicating: (i) lignases, GHF1 β-glucosidases, GHF9 endoglucanases, and GHF43 β-xylosidases in the foregut/salivary gland; (ii) apparent feruloyl esterases in the midgut; and (iii) a rich diversity of at least 17 symbiont-derived GHFs in the hindgut (i.e., GHF 2, 3, 5, 7, 10, 11, 16, 20, 26, 30, 42, 45, 47, 53, 77 and 92). Of the various symbiont GHFs, family 7 exoglucanases are undeniably the most diverse (see Fig. 4 and Todaka *et al.* and Tartar *et al.*[40,47]). Together, these enzyme systems provide the basis for a dual or collaborative system that, without question, must be far more efficient than host-derived enzymes working alone.[12,39,54]

Conclusions

An underlying goal of termite digestomics research is to define collaborative lignocellulose digestion; i.e., to define how termite and symbiont systems complement one another to achieve efficient lignocellulose digestion. The GHF9 and

Table 1. Summary of glycosyl hydrolase families (GHF) identified to date from different symbiont metagenomes / transcriptomes and termite genomes. Numbers indicate the number of genes identified within each GHF. Solid circles (•) denote undeterminable numbers. Dashes (–) indicate missing GHF members. *R. flavipes* sequences are deposited in the Genbank database under accession numbers FL634956-FL640828 and FL641015-FL645753.

GHF No.	*Nasutitermes* (symbiont) [1*]	*R. speratus* (symbiont) [2]	*R. flavipes* (symbiont) [3*]	*R. flavipes* (endogenous) [4*]
GHF 1	22	–	–	2
GHF 2	23	–	1	1
GHF 3	69	1	10	2
GHF 4	14	–	–	–
GHF 5	56	3	11	2
GHF 7	–	•	35	10
GHF 8	5	9	1	1
GHF 9	9	–	–	2
GHF 10	46	•	1	1
GHF 11	14	•	10	5
GHF 13	48	–	–	10
GHF 16	1	–	3	3
GHF 18	17	–	5	8
GHF 20	15	–	3	4
GHF 23	52	–	–	–
GHF 25	1	–	–	–
GHF 26	15	•	5	1
GHF 27	4	–	–	1
GHF 28	6	–	–	–
GHF 30	–	–	1	3
GHF 31	26	–	–	–
GHF 35	3	–	–	–
GHF 36	5	–	–	–
GHF 37	–	–	–	1
GHF 38	11	–	–	2
GHF 39	3	–	–	–
GHF 42	24	–	1	–
GHF 43	16	•	–	1
GHF 44	6	–	–	–
GHF 45	4	2	4	1
GHF 47	–	–	1	–
GHF 51	18	–	–	–
GHF 52	3	–	–	–
GHF 53	12	–	1	1
GHF 57	17	–	–	–
GHF 58	1	–	–	–
GHF 62	–	•	–	–
GHF 65	6	–	–	–

Table 1. Continued

GHF 67	10	–	–	–
GHF 70	–	–	–	1
GHF 74	7	–	–	–
GHF 76	–	–	–	1
GHF 77	14	–	1	–
GHF 85	–	–	–	1
GHF 88	9	–	–	–
GHF 91	1	–	–	–
GHF 92	2	–	1	–
GHF 94	68	–	–	–
GHF 95	12	–	–	–
GHF 98	1	–	–	–
GHF 103	3	–	–	–
GHF 106	2	–	–	–
GHF 109	3	–	–	–

[1] Warnecke *et al*.[43] (*Nasutitermes* symbiont metagenomic sequencing).

[2] Todaka *et al*.[42] (*R. speratus* symbiont cDNA library).

[3] Tartar *et al*.[47] (*R. flavipes* symbiont cDNA library).

[4] Tartar *et al*.[47] (*R. flavipes* normalized gut cDNA library).

* GHF identities were assigned using the CAZy (carbohydrate-active enzymes) database and nomenclature system (http://www.cazy.org/).

7 families, in which the termite genome contributes GHF9 endoglucanases and protistan symbionts contribute GHF7 exoglucanases, are one example of apparent enzymatic collaboration. Another example is with lignin degradation, which is apparently accomplished with enzymes derived from the termite genome. Attaining functional knowledge on such enzyme collaboration will lead to two important outcomes. First, it will lead to identification of industrially relevant enzymes; and second, it will reveal how to combine recombinant enzymes for maximum efficiency in industrial lignocellulose processing. For example, while suitable and efficient cellulases have been identified and are already used in industrial applications, efficient and inexpensive pre-treatments for lignin and hemicellulose depolymerization are not yet well developed. In this respect, termite digestomics has already offered significant insights. Also, as demonstrated by the identification of potential new lignases from a termite genome,[47] there is now strong rationale for a termite genome project, or other transcriptomic sequencing projects that specifically target symbiont-free termite gut or salivary gland tissues. From the perspective of defining collaborative lignocellulose digestion, continued transcriptomic and metagenomic sequencing from *both* symbionts and their termite hosts will also be extremely important. Developing such a holistic understanding of termite digestomics will not only provide a quantum leap in our understanding of the fascinating phenomenon of termite/symbiont co-evolution, it will also meet the ever-important need for refinement in industrial lignocellulose processing.

Acknowledgements

We thank our collaborator Drion Boucias for helpful comments, discussions, and support. Research by the authors was supported by a University of Florida – IFAS Innovation Grant, and in part through the Consortium for Plant Biotechnology Research, Inc. by DOE Prime Agreement No. DE-FG36-02GO12026 (this support does not constitute an endorsement by DOE or by the Consortium for Plant Biotechnology Research, Inc. of the views expressed in this publication). *R. flavipes* sequences as part of reference [47] are deposited in Genbank under accession numbers FL634956-FL640828 and FL641015-FL645753.

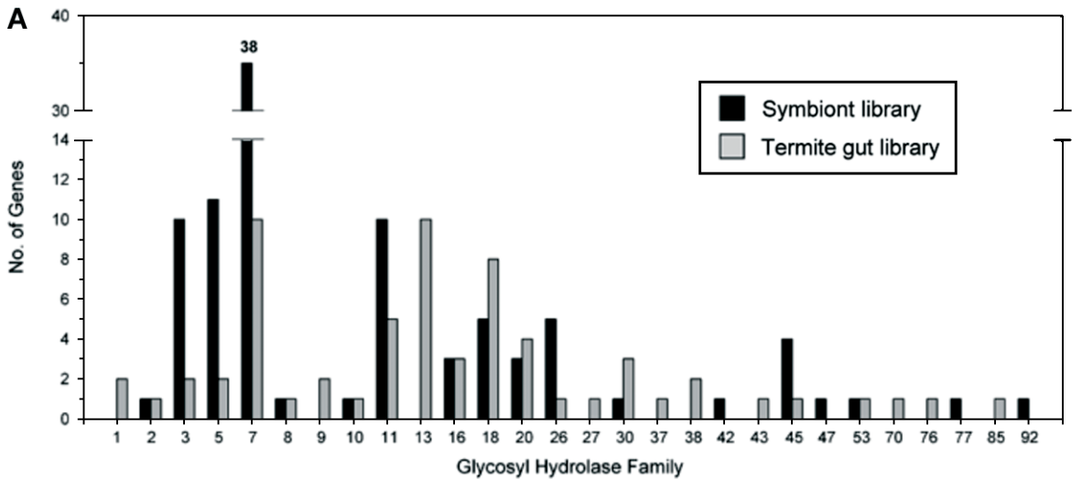


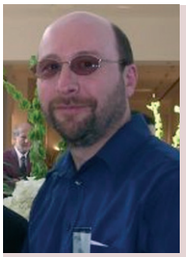
Figure 4. Evidence of host-symbiont collaboration in cellulose and hemicellulose digestion by a lower termite. (A) Asymmetric distribution of GH families among *R. flavipes* endogenous (termite; gray bars) and symbiotic (black bars) sequence pools.[47] GHFs represented in both data sets are assumed to be symbiont derived.[47] (B) GHF members represented only in the endogenous termite sequence pool. (C) Presumed symbiont GHF members represented in both the symbiont and host sequence pools. GHF7 is by far, the most diverse GHF represented in termite protozoan symbiont data sets.[42,47]

References

1. Lange JP, Lignocellulose conversion: an introduction to chemistry, process and economics. *Biofuels Bioprod Bioref* **1**:39–48 (2007).
2. Ragauskas A, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA *et al.*, The path forward for biofuels and biomaterials. *Science* **311**:484–489 (2006).
3. Ljungdahl LG and Erickson KE, Ecology of microbial cellulase degradation. *Adv Micro Ecol* **8**:237–299 (1985).
4. Anderson WF and Akin DE, Structural and chemical properties of grass lignocelluloses related to conversion for biofuels. *J Ind Microbiol Biotechnol* **35**:355–366 (2008).
5. Breznak JA and Brune A, Role of microorganisms in the digestion of lignocellulose by termites. *Annu Rev Entomol* **39**:453–487 (1995).
6. Lee SJ, Kim SR, Yoon HJ, Kim I, Lee KS, Je YH *et al.,* cDNA cloning, expression and enzyme activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. *Comp Biochem Physiol B* **139**:107–116 (2004).
7. Lee SJ, Lee KS, Kim SR, Gui ZZ, Kim YS, Yoon HJ *et al.,* A novel cellulase gene from the mulberry longicorn beetle, *Apriona germari*: gene structure, expression and enzymatic activity. *Comp Biochem Physiol* **B140**:551–560 (2005).
8. Ohkuma M, Termite symbiotic systems: efficient bio-recycling of lignocellulose. *Appl Microbiol Biotechnol* **61**:1–9 (2003).
9. Watanabe H and Tokuda G, Animal cellulases. *Cell Mol Life Sci* **58**: 1167–1178 (2001).
10. Davison A and Blaxter M, Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol* **22**:1273–1284 (2005).

11. Lo N, Watanabe H and Sugimura M, Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. *Proc R Soc Lond B (suppl)* **270p**:S69–S72 (2003).
12. Nakashima KI, Watanabe H, Saitoh H, Tokuda G and Azuma JI, Dual cellulose digesting system of the wood feeding termite, *Coptotermes formosanus*. *Insect Biochem Mol Biol* **32**:777–784 (2002).
13. Richard FJ, Mora P, Errard C and Rouland C, Digestive capacities of leaf-cutting ants and the contribution of their fungal cultivar to the degradation of plant material. *Comp Biochem Physiol B* **175**:297–303 (2005).
14. Saha BC, Hemicellulose conversion. *J Indust Microbiol Biotechnol* **30**:279–291(2003).
15. Brune A, Miambi E and Breznak JA, Roles of oxygen and the intestinal microflora in the metabolism of lignin-derived phenylpropanoids and other monoaromatic compounds by termites. *Appl Env Microbiol* **61**:2688–2695 (1995).
16. Brune A, Emerson D and Breznak JA, Termite gut microflora as an oxygen sink: microelectrode determinations of oxygen and pH gradients in guts of lower and higher termites. *Appl Env Microbiol* **61**:2681–2687 (1995).
17. Butler JHA and Buckerfield JC, Digestion of lignin by termites. *Soil Biol Biochem* **11**:507–511 (1979).
18. Cookson LJ, ^{14}C-lignin degradation by three Australian termite species. *Wood Sci Technol* **21**:11–25 (1987).
19. Cookson LJ, The site and mechanism of ^{14}C-lignin degradation by *Nasutitermes exitosus. J Insect Physiol* **34**:409–414 (1988).
20. Esenther GR and Kirk TK, Catabolism of aspen sapwood in *Reticulitermes flavipes*. *Ann Entomol Soc Am* **67**:989–991 (1974).
21. Shuji I, Ueshima K, Tanaka H and Enoki A, Degradation of wood components by subterranean termite, *Coptotermes formosanus*. *Mokuzai Gakkaishi* **41**:580–586 (1995).
22. Katsumata KS, Jin Z, Hori K and Iiyama K, Structural changes in lignin of tropical woods during digestion by termite, *Cryptotermes brevis*. *J Wood Sci* **53**:419–426 (2007).
23. Erikkson KEL, Blanchette RA and Ander P, *Microbial and Enzymatic Degradation of Wood and Wood Components*. Springer, New York (1990).
24. Taprab Y, Johjima T, Maeda Y, Moriya S, Trakulnaleamsai S, Noparatnaraporn N *et al.,* Symbiotic fungi produce laccases potentially involved in phenol degradation in fungus combs of fungus-growing termites in Thailand. *Appl Env Microbiol* **71**:7696–7704 (2005).
25. Feyereisen R, Insect cytochrome P450, in *Comprehensive Molecular Insect Science [vol. 4]: Biochemistry and Molecular Biology*, ed by Gilbert LI, Iatrou K and Gill SS. Elsevier, Amsterdam, pp. 1–77 (2005).
26. Matsumura F, *Toxicology of Insecticides*. Plenum Press, New York (1986).
27. Siegfried BD and Scharf ME, Mechanisms of organophosphate resistance in insects, in *Biochemical Sites of Insecticide Action and Resistance*, ed by Ishaaya I. Springer-Verlag, New York, pp. 269–272 (2001).
28. Koseki T, Takahashi K, Handa T, Yamane Y, Fushinobu S and Hashizume K, N-linked oligosaccharides of *Aspergillus awamori* feruloyl esterase are important for thermostability and catalysis *Biosci Biotechnol Biochem* **70**:2476–2480 (2006).
29. Sugimoto A, Bignell DE, and MacDonald JA, Global impact of termites on the carbon cycle and atmospheric trace gasses, in *Termites: Evolution, Sociality, Symbioses, Ecology*, ed by Abe T, Bignell DE and Higashi M. Kluwer, the Netherlands, pp. 409–435 (2000).
30. Su NY and Scheffrahn RH, Termites as pests of buildings, pp. in *Termites: Evolution, Sociality, Symbioses, Ecology*, ed by Abe T, Bignell DE and Higashi M. Kluwer, the Netherlands, pp. 437–453 (2000).
31. Wood TG and Johnson RA, The biology, physiology, and ecology of termites, in *Economic Impact and Control of Social Insects*, ed by Vinson SB and Johnson RA. Praeger, New York , pp. 1–68 (1986)
32. Cleveland LR, Symbiosis between termites and their intestinal protozoa. *Proc Natl Acad Sci USA* **9**:424–428 (1923).
33. Slaytor M, Cellulose digestion in termites and cockroaches: what role do symbionts play? *Comp Biochem Physiol* **103B**:775–784 (1993).
34. Watanabe H, Noda H and Lo N, A cellulase gene of termite origin. *Nature* **394**:330–331 (1998).
35. Scharf ME, Scharf WD, Pittendrigh BR and Bennett GW, Caste- and development-associated gene expression in a lower termite. *Genome Biol* **4**:R62 (2003).
36. Scharf ME, Wu-Scharf D, Zhou X, Pittendrigh BR and Bennett GW, Gene expression profiles among immature and adult reproductive castes of the termite *R. flavipes*. *Insect Mol Biol* **14**:31–44 (2005).
37. Tokuda G and Watanabe H, Hidden cellulases in termites: revision of an old hypothesis. *Biol Lett* **3**:336–339 (2007).
38. Wheeler MM, Zhou X, Scharf ME and Oi FM, Molecular and biochemical markers for monitoring dynamic shifts in cellulolytic protozoa in *Reticulitermes flavipes*. *Insect Biochem Mol Biol* **37**:1366–1374 (2007).
39. Zhou X, Smith JA, Oi FM, Koehler PG, Bennett GW and Scharf ME, Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *R. flavipes*. *Gene* **395**:29–39 (2007).
40. Todaka N, Moriya S, Saita K, Hondo T, Kiuchi I, Hirotoshi T, *et al.,* Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *FEMS Microbiol Ecol* **59**:592–599 (2007).
41. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT *et al.,* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560–565 (2007).
42. Leadbetter JR, Schmidt TM, Graber JR and Breznak JA, Acetogenesis from H_2 plus CO_2 by spirochetes from termite guts. *Science* **283**: 686–689 (1999).
43. Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR and Breznak JA, Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292**:2495–2498 (2001).
44. Tokuda G, Lo N, Watanabe H, Arakawa G, Matsumoto T and Noda H, Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. *Mol Ecol* **13**:3219–3228 (2004).
45. Stingl U, Radek R, Yang H and Brune A. "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl Environ Microbiol* **71**:1473–1479 (2005).
46. Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T *et al*., Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci USA* **105**:5555–5560 (2008).

47. Tartar A, Zhou X, Boucias DG and Scharf ME, Digestome analysis reveals evidence of collaborative digestion and lignin degradation in the termite *Reticulitermes flavipes*. In Preparation (2008).
48. Zhou X, Oi FM and Scharf ME, Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc Natl Acad Sci USA* **103**:4499–4504 (2006).
49. Wu-Scharf D, Scharf ME, Pittendrigh BR and Bennett GW, Expressed sequence tags from a polyphenic *Reticulitermes flavipes* cDNA library. *Sociobiology* **41**:479–490 (2002).
50. Nakashima KI, Watanabe H and Azuma JI, Cellulase genes from the parabasalian symbiont *Pseudotrichonympha grassii* in the hindgut of the wood-feeding termite *Coptotermes formosanus*. *Cell Mol Life Sci* **59**:1554–1560 (2002).
51. Watanabe H, Nakashima KI, Saito H and Slaytor M, New endo-beta-1,4-glucanases from the parabasalian symbionts, *Pseudotrichonympha grassii* and *Holomastigotoides mirable* of *Coptotermes termites*. *Cell Mol Life Sci* **59**:1983–1992 (2002).
52. Zhou X, Wheeler MM, Oi FM and Scharf ME, Inhibition of termite cellulases by carbohydrate-based cellulase inhibitors: evidence from *in vitro* biochemistry and *in vivo* feeding studies. *Pestic Biochem Physiol* **90**:31–41 (2008).
53. Zhou X, Wheeler MM, Oi FM and Scharf ME, RNA interference in the termite *Reticulitermes flavipes* by ingestion of double-stranded RNA. *Insect Biochem Mol Biol* **38:**805–815 (2008).
54. Tokuda G, Watanabe H and Lo N, Does correlation of cellulase gene expression and cellulolytic activity in the gut of termite suggest synergistic collaboration of cellulases? *Gene* **401**:131–134 (2007).
55. Khademi S, Guarino LA, Watanabe H, Tokuda G and Mayer EF, Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Biol Crystallog* **D58**:653–659 (2002).
56. Tokuda G, Watanabe H, Matsumoto T and Noda H, Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis*: distribution of cellulases and properties of endo-beta-1,4-glucanase. *Zool Sci* **14**: 83–93 (1997).
57. Tokuda G, Saito H and Watanabe H, A digestive beta-glucosidase from the salivary glands of the termite, *Neotermes koshunensis*: distribution characterization and isolation of its precursor cDNA by 5'- and 3'-RACE amplifications with degenerate primers. *Insect Biochem Mol Biol* **32**:1681–1689 (2002).
58. Tokuda G, Lo N, Watanabe H, Slaytor M, Matsumoto T and Noda H, Metazoan cellulase genes from termites: intron/exon structures and sites of expression. *Biochim Biophys Acta* **1447**:146–159 (1999).



M.E. Scharf

Dr Michael E. Scharf is an Assistant Professor of Insect Molecular Toxicology at the University of Florida in Gainesville, FL, focusing on development of next-generation insecticides, termite genomics, and functional genomics of the termite gut. Dr Scharf received his BS, MS and PhD degrees from Purdue University in 1991, 1993 and 1997, respectively. He has been at the University of Florida since 2004.

Aurélien Tartar

Dr Aurélien Tartar is an Assistant Professor of Biology at Nova Southeastern University in Fort Lauderdale, FL. His research interests are centered on insect-associated microbes, including both symbiotic and pathogenic microorganisms. Dr Tartar received his MS degree from AgroParisTech in 1998 and his PhD degree from the University of Florida in 2004. He has been at Nova Southeastern University since 2008.

