Minireview

Termites in the woodwork
Samuel Chaffron and Christian von Mering

Address: Institute of Molecular Biology and Swiss Institute of Bioinformatics, University of Zurich, Winterthurerstrasse, 8057 Zurich, Switzerland.

Correspondence: Christian von Mering. Email: mering@molbio.uzh.ch

Abstract

Termites eat and digest wood, but how do they do it? Combining advanced genomics and proteomics techniques, researchers have now shown that microbes found in the termites' hindguts possess just the right tools.
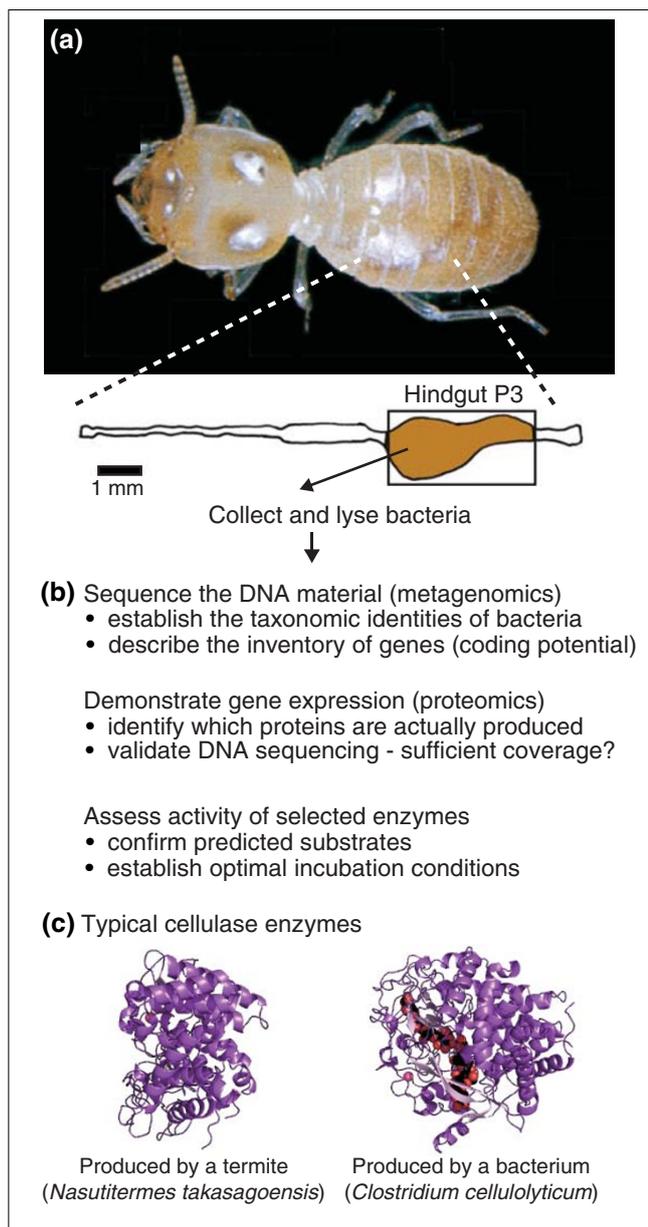
Most animals, from insects to mammals, carry complex communities of microbes in their digestive tracts. In the case of wood-eating termites, these gut microbes are particularly important: they are thought to provide most of the capabilities needed for efficient digestion of wood, which is otherwise a largely inaccessible food source. They also help to compensate for the paucity of some nutrients in wood, for example by fixing atmospheric nitrogen, and they synthesize essential amino acids and other compounds for their hosts [1,2].

Despite their importance, relatively little is known about gut microbes in termites. This is partly because gut microbes are often difficult to grow in pure culture (as is the case for most microbes sampled from natural environments). Furthermore, a single termite can harbor a very complex assemblage of hundreds of different microbial lineages, whose members may vary widely in terms of abundance and growth rates. Without access to cultivated strains, researchers have to rely on so-called 'cultivation-independent' molecular techniques to analyze such communities. A clever combination of these techniques has now been applied to a section of the termite hindgut, aiming to identify molecular tools used by the microbes in this compartment to degrade wood [3]. Here, we review the procedures and results of this study, and discuss insights into the biological system as well as implications for the generation of biofuels.

A comprehensive inventory

As recently as 2004, biologists had rather limited experimental options for taking stock of uncultured microbes in their natural environments. They could analyze selected phylogenetic marker genes to assess taxonomic identity (using *in situ* hybridization or PCR-based sequencing), or they could use expression cloning to screen for genes encoding a specific activity of interest. Another possibility was to clone and sequence individual DNA fragments isolated from the community, in the hope of finding phylogenetic marker genes and important functional genes together on the same fragment: this latter approach can help to map lifestyles to a given lineage [4,5]. However, none of these strategies simultaneously provides a global inventory of both the taxonomic and the functional properties of a microbial community.

To overcome this limitation, researchers have since begun to apply genomics (and proteomics) technologies in high-throughput mode, analyzing entire microbial assemblages without first cloning individual strains [6-9]. These exciting new research approaches ('environmental genomics', 'metagenomics' and 'metaproteomics') put the possibility of a molecular description of an entire microbial community within reach for the first time. For the termite gut ecosystem, Warnecke and colleagues [3] have now attempted just that, in a formidable *tour de force*. They even went a step further by complementing their work with a preliminary

(a)

Hindgut P3

1 mm

Collect and lyse bacteria

(b) Sequence the DNA material (metagenomics)
• establish the taxonomic identities of bacteria
• describe the inventory of genes (coding potential)

Demonstrate gene expression (proteomics)
• identify which proteins are actually produced
• validate DNA sequencing - sufficient coverage?

Assess activity of selected enzymes
• confirm predicted substrates
• establish optimal incubation conditions

(c) Typical cellulase enzymes

Produced by a termite
(*Nasutitermes takasagoensis*)

Produced by a bacterium
(*Clostridium cellulolyticum*)

Figure 1
Exploring the termite hindgut. **(a)** Photograph of a worker termite from the genus Nasutitermes. **(b)** The gut contents from the third proctodeal segment (P3) were sampled, and analyzed using a variety of techniques. **(c)** Three-dimensional structures of two typical cellulase enzymes (left, PDB1ksd; right, PDB1f9d). Photograph: CSIRO.

biochemical analysis of some of the enzymes they discovered.

The team began by sampling the luminal contents of the P3 hindgut segment, pooling the material from 165 adult worker termites. This is the largest of the gut compartments, yet still contains only about 1 μl of material (Figure 1). From this material, the authors purified the genomic DNA,

fragmented, subcloned and sequenced it. They generated about 70 megabases (Mb) of raw shotgun sequence and also selected several fosmid inserts to be sequenced separately for more detailed inspection. Warnecke *et al.* [3] mainly used classical capillary sequencing; today, this technology is being rapidly surpassed and the next-generation sequencing technologies will increase the scope of such studies by orders of magnitude [10]. As is the case for most metagenomics projects, the shotgun reads could not be assembled into complete genomes. In fact, relatively little assembly was possible at all - the longest assembled contig encompasses a mere 14.7 kb - owing to the complexity of the microbial community.

The metagenomics sequencing effort was complemented by a more focused strategy to sample a single phylogenetic marker gene (using PCR amplification and cloning of 16S ribosomal RNA genes). These 16S sequences were combined with similar sequences from the shotgun approach and analyzed in order to ask the question: which phyla and how many species are present in the termite gut?

As previously reported, members of the bacterial phyla Spirochaetes and Fibrobacteres dominated the community. Notably, Warnecke *et al.* [3] did not detect any archaeal sequences, nor did they find much eukaryotic material (there was apparently very little contaminating DNA from the host, if any). They discovered 216 distinct 'phylotypes' of bacteria (that is, groups of 16S sequences with at least 99% sequence identity) and estimated from the redundancy in these phylotypes that what they had found represented about 70-90% of the total diversity. This is roughly similar to the diversity of the human gut microbial flora [11].

Apart from a phylogenetic characterization, the authors carried out a quantitative analysis of functional genes in the sample. They focused on certain categories of interest: how many genes would encode enzymes known to degrade cellulose, xylan or lignin? Would there be evidence for nitrogen fixation? To find out, the authors grouped the predicted genes into families and orthologous groups, annotated them, and compared the abundance of each gene family to the respective occurrences of these genes in other environments, such as soil [7], seawater [6] or the human gut [12].

First and foremost, they found a large number of glycoside hydrolases; that is, enzymes that can degrade polysaccharides. The authors classified these genes according to known sequence families and predicted substrates, and attempted to assign them to the most likely source organism. Forty-five distinct groups were detected, and composition-based analysis predicted *Treponema* (a genus of Spirochaetes) as the most likely origin for the majority of these enzymes. In addition, a number of gene families known to associate with glycoside hydrolase

domains were found, including carbohydrate-binding domains and other functional domains. In total, hundreds of new enzymes were described, many of which significantly extend our knowledge of the various enzyme families. Remarkably, no enzymes were found for the degradation of lignin, a major constituent of wood that is partly responsible for its strength. Some enzymes capable of lignin degradation have previously been described, but none of these was found among the sequences retrieved here. Of course, as yet undescribed enzymes could do the task, or this activity could be located in a different compartment of the termite gut. The latter might well be the case, as many of the enzymes known to degrade lignin require molecular oxygen and the P3 segment is largely anoxic.

As expected, several other functional processes known (or suspected) to be carried out by the gut microbes were represented among the sequences. These include nitrogen fixation, chemotaxis and chemosensation, as well as carbon fixation from carbon dioxide via the Wood-Ljungdahl pathway [13].

Metaproteomics and activity assays

The detection of an open reading frame alone does not suffice to show that the protein is actually made, nor does it readily indicate when and where the gene is expressed. To assess the more abundant proteins at least, mass spectrometry is a promising tool, provided that the community is not too complex and it has been sampled deeply enough at the nucleotide level [9].

Warnecke and co-workers [3] have focused on a particular subset of the proteome (the secreted extracellular proteins) by analyzing centrifuged and clarified P3 luminal fluid using mass spectrometry. Although they were able to detect only a relatively small fraction of the expected proteins, they confirmed for the first time that bacterial glycosidases are indeed produced in the termite gut. What is more, they actually demonstrated activity for a number of these enzymes. More than 40 of the glycosidase genes were individually cloned, expressed heterologously and tested on acid-solubilized and microcrystalline cellulose. Although this is unlikely to match the situation in which these genes work in vivo, it shows convincingly that termite guts harbor secreted functional glycosidase enzymes.

Who encodes what?

The most pressing question in any metagenomics analysis is to what extent the molecular functions identified can be assigned to particular microbial lineages. This information is still almost entirely lacking for all but the simplest microbial communities, but it is crucially important for any deeper understanding of the ecology of these assemblages. The problem remains largely unsolved: in the current study [3],

compositional analysis of the DNA provided classification for only 9% of the contigs beyond the phylum level, leading to uncertainties; for example, none of the *nifH* nitrogen-fixation genes could be assigned. Even where it does work, compositional analysis is probably not very reliable, as microbial genomes can harbor large stretches of recently acquired genetic material, which may not yet have equilibrated with the host genome. For individual genes of interest, clever use of coupled PCR reactions has recently shown a way to reliably map genes to their host genomes [14], but for a global solution we will probably have to wait for single-cell sequencing [15].

One of the most intriguing results of this study actually concerns a class of proteins to which no molecular function can be assigned so far. Warnecke *et al.* [3] identified a number of previously uncharacterized protein families that were strongly enriched compared with other metagenomes, and that were in some cases even quite specific to the termite gut microbes. This is exciting because the degradation of lignocellulose in most cases requires not just individual enzymes operating in isolation, but large macromolecular complexes that guide and coordinate the process. These complexes have been termed 'cellulosomes' and are (partially) known for a number of microbial species [16]. Scaffold proteins and accessory proteins may, however, be different from lineage to lineage, and this could mean that a number of unknown cellulosome-like proteins are contained in the specifically enriched proteins discovered in this study.

As an aside, we hope that the success of the gene-based approaches illustrated here and elsewhere will not deter those who seek to characterize individual microbial lineages more thoroughly. Isolating and growing microbes in pure culture remains an art, and one that continues to produce ground-breaking insights [17-19]. It provides unequivocal anchors for taxonomists and for functional studies, and allows access to the slow-growing, rare community members that can contribute essential functions. Comparative genomicists depend on a continued input of high-quality, well annotated genome sequences to sort out phylogeny and to understand the effects of horizontal gene transfers and other evolutionary phenomena. It is to be hoped that those who produce isolates and complete genome sequences will continue to be given appropriate credit for their work.

Wood as a source of biofuel

Can the results of this study help us make better use of wood as a fuel? Humans have used wood as an energy source for thousands of years, mostly for domestic heating and cooking. But it has also been used to generate power, for example in steam engines and occasionally by converting it to fuel for use in combustion engines (Figure 2). Conversion of wood into a biofuel, such as ethanol, is again a hot topic [20,21] because of its potential for at least partially replacing
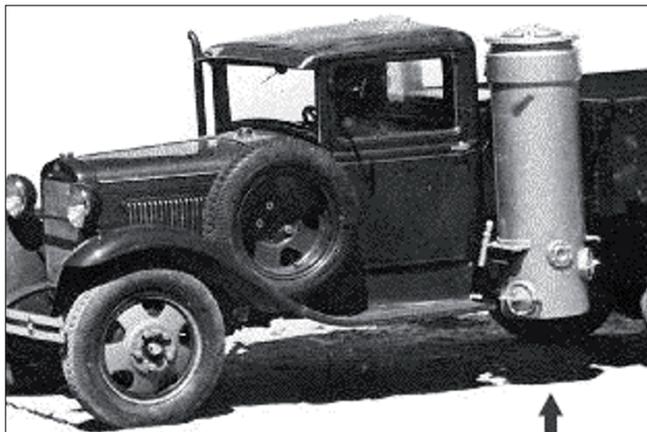
Figure 2
Making fuel from wood. The photograph, taken in 1951, shows a Russian automobile fitted with a 'wood gasifier' (arrow). Similar vehicles were relatively widespread in Europe in the 1940s and 50s, and achieved conversion efficiencies of roughly 3 kg of wood consumed per power-output equivalent to 1 liter of gasoline. Modern biotechnological approaches, using enzymes like the ones found in termite guts, are still struggling to surpass that efficiency [20]. But they do offer a much more convenient and clean fuel product, ethanol.

fossil fuels in transportation and thereby lowering greenhouse gas emissions.

Unlike some first-generation biofuels derived from just a small, energy-rich part of the plant (such as the seeds), which have been criticized on environmental grounds [22], wood-based biofuels use almost the whole plant. Trees in particular seem suitable for biofuel production, as they can be grown on marginal soils with very little water or fertilizer and do not compete with food crops.

Today, wood conversion is being attempted on the industrial scale using biotechnology. Cellulases and hemicellulases are already being used in this process and these enzymes can be further optimized. Many bigger challenges remain: how best to deal with the lignin, how best to pre-treat the wood and how to more efficiently release all sugars for fermentation. As termites achieve all of that in a volume of 1 µl, and at ambient temperatures, it seems that we have a lot to learn from them. It would be very satisfying if basic research into termite physiology could ultimately end up helping us to make better, environmentally friendly fuels.

References
1. Breznak JA, Brune A: **Role of microorganisms in the digestion of lignocellulose by termites.** *Annu Rev Entomol* 1994, **39**:453-487.
2. Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA: **Nitrogen fixation by symbiotic and free-living Spirochaetes.** *Science* 2001, **292**:2495-2498.
3. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, *et al.*: **Metagenomic and functional analysis of hindgut micro-biota of a wood-feeding higher termite.** *Nature* 2007, **450**:560-565.
4. Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, *et al.*: **Bacterial rhodopsin: evidence for a new type of phototrophy in the sea.** *Science* 2000, **289**:1902-1906.
5. Treusch AH, Leininger S, Kletzin A, Schuster SC, Klenk HP, Schleper C: **Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling.** *Environ Microbiol* 2005, **7**:1985-1995.
6. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, *et al.*: **Environmental genome shotgun sequencing of the Sargasso Sea.** *Science* 2004, **304**:66-74.
7. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, *et al.*: **Comparative metagenomics of microbial communities.** *Science* 2005, **308**:554-557.
8. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment.** *Nature* 2004, **428**:37-43.
9. Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, 2nd, Shah M, Hettich RL, Banfield JF: **Community proteomics of a natural microbial biofilm.** *Science* 2005, **308**:1915-1920.
10. Ryan D, Rahimi M, Lund J, Mehta R, Parviz BA: **Toward nanoscale genome sequencing.** *Trends Biotechnol* 2007, **25**:385-389.
11. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA: **Diversity of the human intestinal microbial flora.** *Science* 2005, **308**:1635-1638.
12. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE: **Metage-nomic analysis of the human distal gut microbiome.** *Science* 2006, **312**:1355-1359.
13. Graber JR, Breznak JA: **Physiology and nutrition of *Treponema primitia*, an H2/CO2-acetogenic Spirochaete from termite hindguts.** *Appl Environ Microbiol* 2004, **70**:1307-1314.
14. Ottesen EA, Hong JW, Quake SR, Leadbetter JR: **Microfluidic digital PCR enables multigene analysis of individual environmental bacteria.** *Science* 2006, **314**:1464-1467.
15. Lasken RS: **Single-cell genomic sequencing using multiple displacement amplification.** *Curr Opin Microbiol* 2007, **10**:510-516.
16. Doi RH, Kosugi A: **Cellulosomes: plant-cell-wall-degrading enzyme complexes.** *Nat Rev Microbiol* 2004, **2**:541-551.
17. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, *et al.*: **Genome streamlining in a cosmopolitan oceanic bacterium.** *Science* 2005, **309**:1242-1245.
18. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, *et al.*: **Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia.** *Science* 2007, **316**:1307-1312.
19. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M, *et al.*: **The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism.** *Proc Natl Acad Sci USA* 2003, **100**:12984-12988.
20. Wyman CE: **What is (and is not) vital to advancing cellulosic ethanol.** *Trends Biotechnol* 2007, **25**:153-157.
21. Hahn-Hagerdal B, Galbe M, Gorwa-Grauslund MF, Liden G, Zacchi G: **Bio-ethanol - the fuel of tomorrow from the residues of today.** *Trends Biotechnol* 2006, **24**:549-556.
22. Pimentel D, Patzek T, Cecil G: **Ethanol production: energy, economic, and environmental losses.** *Rev Environ Contam Toxicol* 2007, **189**:25-41.