# Gut anatomical properties and microbial functional assembly promote lignocellulose deconstruction and colony subsistence of a wood-feeding beetle

Javier A. Ceja-Navarro[1]*, Ulas Karaoz[1], Markus Bill[1], Zhao Hao[1], Richard A. White III[2], Abelardo Arellano[1], Leila Ramanculova[1], Timothy R. Filley[3], Timothy D. Berry[3,4], Mark E. Conrad[1], Meredith Blackwell[5,6], Carrie D. Nicora[2], Young-Mo Kim[2], Patrick N. Reardon[7], Mary S. Lipton[8], Joshua N. Adkins[6], Jennifer Pett-Ridge[9] and Eoin L. Brodie[1,10]*

## Abstract

Beneficial microbial associations enhance the fitness of most living organisms, and wood-feeding insects offer some of the most striking examples of this. *Odontotaenius disjunctus* is a wood-feeding beetle that possesses a digestive tract with four main compartments, each of which contains well-differentiated microbial populations, suggesting that anatomical properties and separation of these compartments may enhance energy extraction from woody biomass. Here, using integrated chemical analyses, we demonstrate that lignocellulose deconstruction and fermentation occur sequentially across compartments, and that selection for microbial groups and their metabolic pathways is facilitated by gut anatomical features. Metaproteogenomics showed that higher oxygen concentration in the midgut drives lignocellulose depolymerization, while a thicker gut wall in the anterior hindgut reduces oxygen diffusion and favours hydrogen accumulation, facilitating fermentation, homoacetogenesis and nitrogen fixation. We demonstrate that depolymerization continues in the posterior hindgut, and that the beetle excretes an energy- and nutrient-rich product on which its offspring subsist and develop. Our results show that the establishment of beneficial microbial partners within a host requires both the acquisition of the microorganisms and the formation of specific habitats within the host to promote key microbial metabolic functions. Together, gut anatomical properties and microbial functional assembly enable lignocellulose deconstruction and colony subsistence on an extremely nutrient-poor diet.

## Main

Over evolutionary time, arthropods have developed critical microbial associations, which enhance the host's metabolic capabilities and confer multiple ecological advantages[1,2,3]. In wood-feeding insects, digestive tract physicochemical gradients have evolved to enhance the kinetics and thermodynamic favourability of metabolic processes such as lignocellulose oxidation, hydrolysis, fermentation and acetogenesis[4,5,6,7]. While a number of studies have related these physicochemical parameters to metabolic products and microbial composition[8,9,10], the complex spatial arrangement of microbial metabolism required for woody debris transformation and host subsistence on a recalcitrant diet is poorly characterized. Understanding how

functional roles are partitioned across distinct compartments of the digestive tract and between genomes of gut microbiota is key to understanding the process by which beneficial microbiota are selected, enriched, distributed, and maintained in the digestive tract of animals. Significantly, in the case of social and subsocial insects, morphological gut differentiations and microbiome compartmentalization may have been selected not only for the benefit of the host, but also for the benefit of its brood, and colony continuity.

*Odontotaenius disjunctus*, the passalid beetle, is a wood-feeding insect that is considered one of the most highly subsocial arthropods[11], with pairs defending log tunnels, remaining with their offspring from egg to adulthood, and demonstrating parent–offspring cooperation for protection and caring of other offspring[12,13]. These beetles are usually found in second-stage decayed wood[14], where adults grind the lignocellulose substrate into small particles that are ingested and processed through a complex, morphologically differentiated digestive system[15]. Passalid feeding accelerates woody biomass decomposition in forested ecosystems, with each colony member consuming approximately 4.5 times its weight in woody material per day[16,17]. Digested woody material is released as frass, which is typically consumed by other members of the colony, particularly larvae, inoculating them with the adult's microbiome and providing nutrients required for survival and development[18].

The external and internal morphology of the adult *O. disjunctus* digestive tract is characterized by four morphologically differentiated regions (Fig. 1), each consisting of a complex landscape with varying cuticle thickness and unique topography[15]. These landscapes allow diverse microbial populations to establish, suggesting that such anatomical properties and separation of these compartments may be beneficial to the processes of energy extraction from woody biomass15. Isolation and analysis of *O. disjunctus*' fungal microbiota, particularly in the posterior hindgut region (PHG), has identified a large diversity of yeasts with xylose fermentation capabilities, suggestive of their role in host metabolism[19,20,21,22,23,24]. Previous research has demonstrated that passalid beetle digestive tracts harbour diverse bacterial and archaeal populations, with steep exterior-to-interior oxygen gradients allowing close coexistence of aerobic and anaerobic microorganisms[10]. Analyses of microbial composition and *nifH* gene expression suggested that important processes (for example, nitrogen fixation and methanogenesis) are segregated within the four primary digestive tract regions and that deterministic processes impact microbiome assembly[10]. On the basis of this previous work, and evidence that younger-stage passalids are dependent on the feeding activity of adults[18], we hypothesized that co-evolution of the insects' digestive tract and its microbiome has resulted in a functional segregation of microbial metabolism. The segregation probably facilitates energy extraction from the host's extremely recalcitrant diet, both for adult hosts and their offspring.

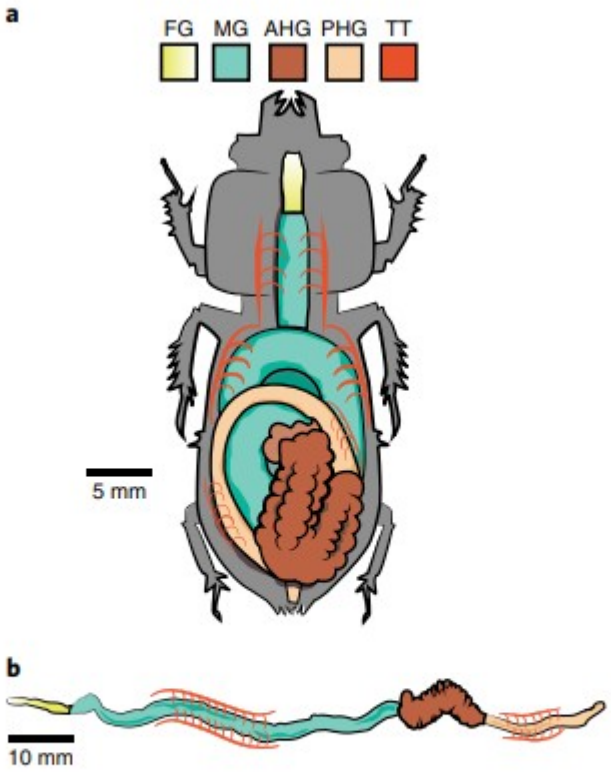


**Fig. 1 | Digestive tract of *O. disjunctus*.** **a**, Dorsal view of a dissected beetle showing the arrangement of its main four regions within the abdominal cavity. **b**, Depiction of the digestive tract and its four main regions when fully removed and extended. FG, foregut; MG, midgut; TT, tracheal tubes.

Here, we demonstrate how spatially segregated microbiome functionality facilitates lignocellulose deconstruction and transformation in the passalid digestive tract. Using spectroscopic, isotopic and analytical chemistry approaches, along with metagenome and metaproteome analyses, we reconstructed the pathways for sequential lignocellulose transformation, the subsequent fermentation of lignocellulose products, and the distribution of key microbial catalysts throughout the passalid digestive tract. Metabolomics of the gut lumen and frass confirmed the production of metabolites predicted by metagenomics and metaproteomics, and showed that excreted frass is rich in substrates that may contribute to overall colony nutrition. During its evolution, the passalid beetle has developed a digestive tract with specialized compartments for lignocellulose deconstruction and fermentation, and anatomical features that direct the flux of reducing equivalents ($H_2$) towards metabolic products that benefit both the host and its progeny.

Results

## Woody substrate transformation during passage through the passalid beetle digestive tract

To examine the transformation of oak wood following passage through the gut of *O. disjunctus*, we determined the transit time of woody substrates through the digestive tract of replicated adult beetles using fluorescent microspheres. We used different approaches to compare the relative chemical composition of wood substrate versus excretion products (frass). The transit time for woody substrate through the digestive tract was 10–13 h. Bulk C:N ratios declined from 167 to 49, demonstrating significant nitrogen enrichment in the frass (total carbon: $P = 0.01$; total nitrogen: $P = 2.1 \times 10^{-4}$) (Fig. 2a). Fourier transform infrared (FTIR) spectroscopy supported this observation, as spectral features corresponding to the amide II bond of proteins (N–H bending and C–N stretch at ~1,550 $cm^{-1}$) increased by approximately twofold in the frass ($P = 6.9 \times 10^{-5}$) (Fig. 2c). FTIR also showed a reduction of glycosidic linkages (~997 $cm^{-1}$) of over 9%, and an 86% reduction of hemicellulose ester side chains (~1,730 $cm^{-1}$) ($P = 6.2 \times 10^{-12}$), suggesting biological deconstruction of these materials. Similarly, a 45% reduction of aromatic compounds was observed ($P = 4.2 \times 10^{-7}$) (aromatic C=C stretch vibration at ~1,510 $cm^{-1}$; Fig. 2c). To confirm lignin transformation and investigate the potential mechanisms for its deconstruction, we used a thermochemolysis method described previously[25]. This demonstrated that an increase in both the guaicyl and syringyl lignin acid-to-aldehyde ratios occurred between the woody substrate diet and frass (Fig. 2b). These observations are diagnostic of lignin side-chain oxidation and ring hydroxylation as a process contributing to lignin depolymerization within the digestive tract.

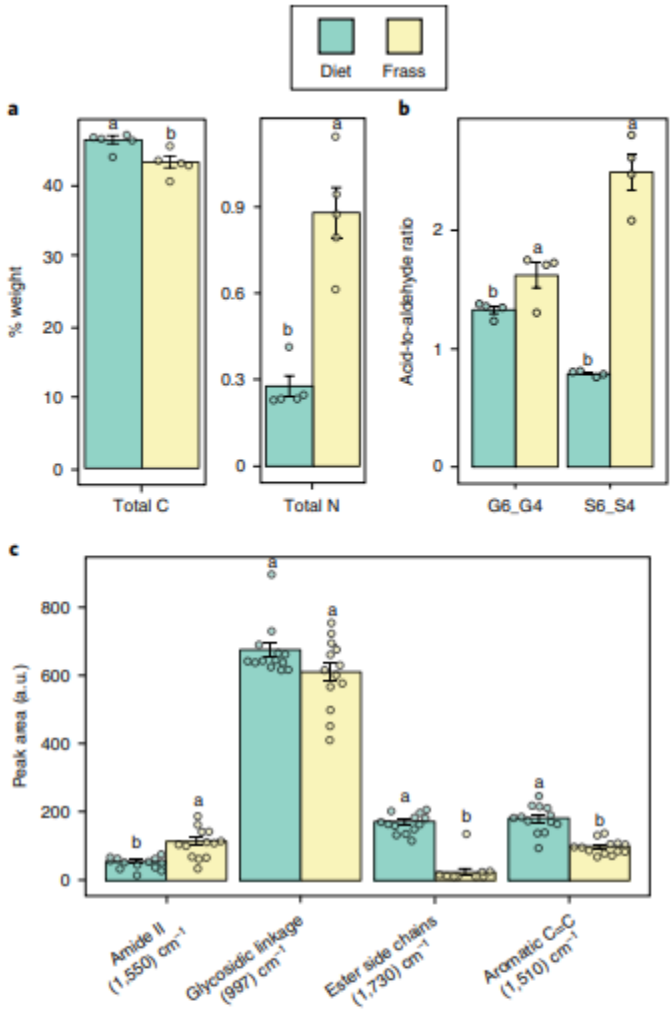


**Fig. 2 | Wood biomass is transformed as it passes through the digestive tract of *O. disjunctus*.** **a**, Bulk chemical analyses show significant differences in total carbon and total nitrogen content between the fed-wood and excreted frass ($n=5$). **b**, $^{13}$C-tetratmethylammonium hydroxide (TMAH) thermochemolysis demonstrates significant differences in acid-to-aldehyde ratios of guaicyl and syringyl lignin between the fed-wood and frass ($n=4$). These changes indicate that lignin is modified by side-chain oxidation and ring demethylation in the beetle gut. G4, 3,4-dimethoxybenzaldehyde; G6, 3,4-dimethoxybenzoic acid; S4, 3,4,5-trimethoxybenzaldehyde; S6, 3,4,5-trimethoxybenzoic acid. **c**, FTIR analyses of fed-wood and frass ($n=13$) show reductions in the chemical signatures of cellulose (glycosidic linkage), xylan (ester side chains) and aromatic compounds (aromatic carbon double bond), and an increase in nitrogen content (amide II bonds). Targeted absorption area peaks were normalized by the total number of detected peaks. Diet (green) = 2.5 g pulverized oak wood with 0.08 g agarose. Frass (yellow) = beetle's excretion products. Values with same letters are not significantly different ($P<0.05$). Bar heights and error bars represent means ± s.e. and each point denotes a biological replicate measurement. Statistical differences were assessed using one-way analysis of variance and the least significant difference test, with *P* values adjusted using the Bonferroni correction method.

Nuclear mangenetic resonance (NMR) analysis was carried out to identify the metabolic products of lignin, cellulose and xylan decomposition. To confirm that these components of wood were the primary contributors to metabolic by-products in the digestive tract, we compared the metabolites from beetles fed with oak wood with those from beetles fed with an artificial diet containing pure lignin, cellulose and xylan (Supplementary Fig. 1). Similar metabolites and concentrations were detected in the frass of insects feeding with oak wood and the artificial diet, confirming that lignin, cellulose and xylan transformations are key to metabolite production (Supplementary Fig. 2). The most abundant products identified in frass were the fermentation products acetate, formate, lactate and ethanol, in addition to xylose, aromatic monomers such as benzoate, and phenylpropionate. Knowing that lignin, cellulose and xylan are all transformed in the passalid hindgut, we next tested the production of signature compounds in the individual gut regions using gas chromatography–mass spectrometry (GC-MS) (Supplementary Fig. 3). Glucose, for example, was present at the highest concentrations in the midgut region, while concentrations of benzoic acid and hydrocinnamic acid (lignin derivatives) increased through the digestive tract. A number of free amino acids were detected, typically decreasing in concentration between the midgut and frass. 3,4-dihydroxybenzoic acid—another lignin derivative produced through oxidation—was lowest in the mostly anaerobic anterior hindgut (AHG) region (Supplementary Fig. 3).

Microbial depolymerization of lignocellulose is compartmentalized and enriched in more aerobic gut regions

To identify the probable microbial catalysts of woody substrate transformations and their spatial arrangement within the beetle digestive tract, we sequenced 16 metagenomic libraries from 4 independent beetle specimens, and extracted proteins and metabolites from 5 specimens, analysing each gut region. Metagenome reads were co-assembled across all beetles and gut regions. A total of 111,177 scaffolds larger than 1 kilobase pairs (kbp) were produced, with an average scaffold size of 4.1 kbp, an average N50 size of 6.8 kbp and a maximum scaffold size of 258 kbp (Supplementary Datasets 1 and 3). Open reading frames (ORFs) were predicted, yielding a total of 401,301 ORFs, and annotated against the Pfam database and the database for automated carbohydrate-active enzyme annotation (dbCAN) (Supplementary Datasets 2 and 3). Predicted proteins were used to predict peptide spectra for metaproteomic analysis. Mass spectrometry runs resulted in 829,099 raw spectra, 75,754 unique peptides and 31,749 unique proteins (1 peptide per protein). Of the 31,749 unique proteins, 9,752 were mapped to the metagenomes, and 3,634 were identified as relevant for the microbial processes of lignin, xylan and cellulose degradation, fermentation and nitrogen fixation, and methane and acetate production (Supplementary Table 1).

Bacterial genes coding for enzymes associated with lignin deconstruction processes[26], including multi-copper laccases, catalase-peroxidases,

manganese-dependent superoxide dismutases, DyP-type peroxidases, aromatic-ring-opening dioxygenases, feruloyl esterases, protochatecuate dioxygenases and vanillyl-alcohol dioxygenases, were detected with highest coverage at the MG and PHG regions (Fig. 3a and Supplementary Dataset 4). Genes coding for glycoside hydrolases (families with the prefix 'GH')[27] involved in the deconstruction of cellulose and xylan were detected across all regions of the beetle gut (Fig. 3b). These glycoside hydrolases included enzymes such as endocellulases (families GH5, 6, 8, 9, 44 and 48), endoxylanases (families GH5, 8, 10, 11, 30 and 43), cellobiohydrolases (families GH5, 6, 9 and 48), cellodextrinases (families GH1 and 3), β-glucosidases (families GH1, 3 and 116), β-xylosidases (families GH3, 30, 39, 43 and 116) and lytic polysaccharide monooxygenases (family AA10). With the exception of GH9, GH30 and GH39, all other glycoside hydrolases were more abundant at the MG and PHG (Fig. 3b and Supplementary Dataset 4)—the two regions with the highest density of connected tracheal tubes[15]. Xylose isomerases, required for the transformation of xylose, were identified at high coverage and protein expression in the MG, AHG and PHG (Fig. 4a). Families of glycoside hydrolases involved in the degradation of other wood polysaccharides, such as arabinan, mannan, glucomannan and galactomannan, were also detected, including α-L-arabinofuranosidase (GH3 and 43), β-mannosidases (GH1, 2 and 5) and β-mannanases (GH5 and 26), which were detected with highest coverage at the more aerobic gut regions —the MG and PHG (Supplementary Dataset 4). A total of 12,000 metagenome-predicted ORFs were identified as of arthropod origin; from these, 13 were associated with β-glucosidases (GH1 and 116), 7 with endoxylanases and β-xylosidases (GH30 and 116), and 2 with lytic polysaccharide monooxygenases from the family AA15 (Supplementary Dataset 4). Although not detected in the metaproteome, these host enzymes may also complement the activity of the larger arsenal of microbial enzymes, as observed in other arthropods such as *Thermobia domestica*[28].

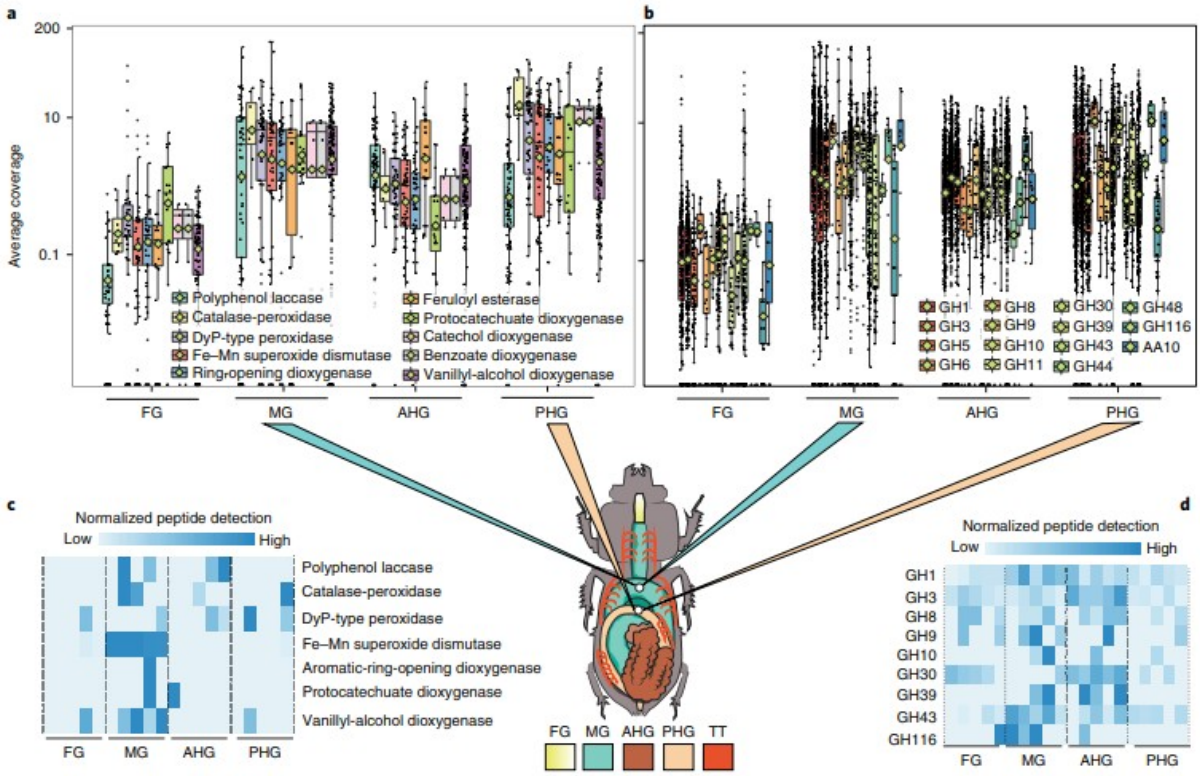


**Fig. 3 | Distribution of the microbial genetic potential for lignocellulose degradation and its expression through the digestive tract of *O. disjunctus*.** **a**, Genes coding for different enzymes involved in lignin degradation were detected through the four gut regions of four *O. disjunctus* specimens ($n=16$), with highest coverage levels detected at the MG and PHG. **b**, Glycoside hydrolases were detected in the metagenome, and their coverage distribution was similar to the ligninolytic enzymes. Coloured lines connect the plots to the areas in the beetle gut were the detected genes had higher average coverage. In each boxplot, a point represents a single gene per category and its detected coverage. Diamonds represent means, box boundaries represent the first and third quartiles of the distribution, and the median is represented as the horizontal line inside each box. Box plot whiskers span 1.5 times the interquartile range of the distribution, and outliers are denoted as large points outside the whiskers. Average coverage was calculated as the numbers of reads mapped to each assembled scaffold per sample and normalized to reads per kb per million reads mapped. **c**,**d**, A metaproteomic survey identified expression of 6 out of 10 metagenome-detected lignolytic enzymes, with highest expression in the MG region (**c**), and 9 out of 15 metagenome-detected glycoside hydrolases distributed through the beetle gut (**d**). Heat plots show average values of normalized protein expression per gut region. The colour gradient from white to blue indicates higher peptide detection across an average of five beetles per gut region. Statistical differences were evaluated by Kruskal–Wallis test. Pairwise comparisons were performed by two-sided Wilcox test with *P* values adjusted using the Benjamini–Hochberg method.

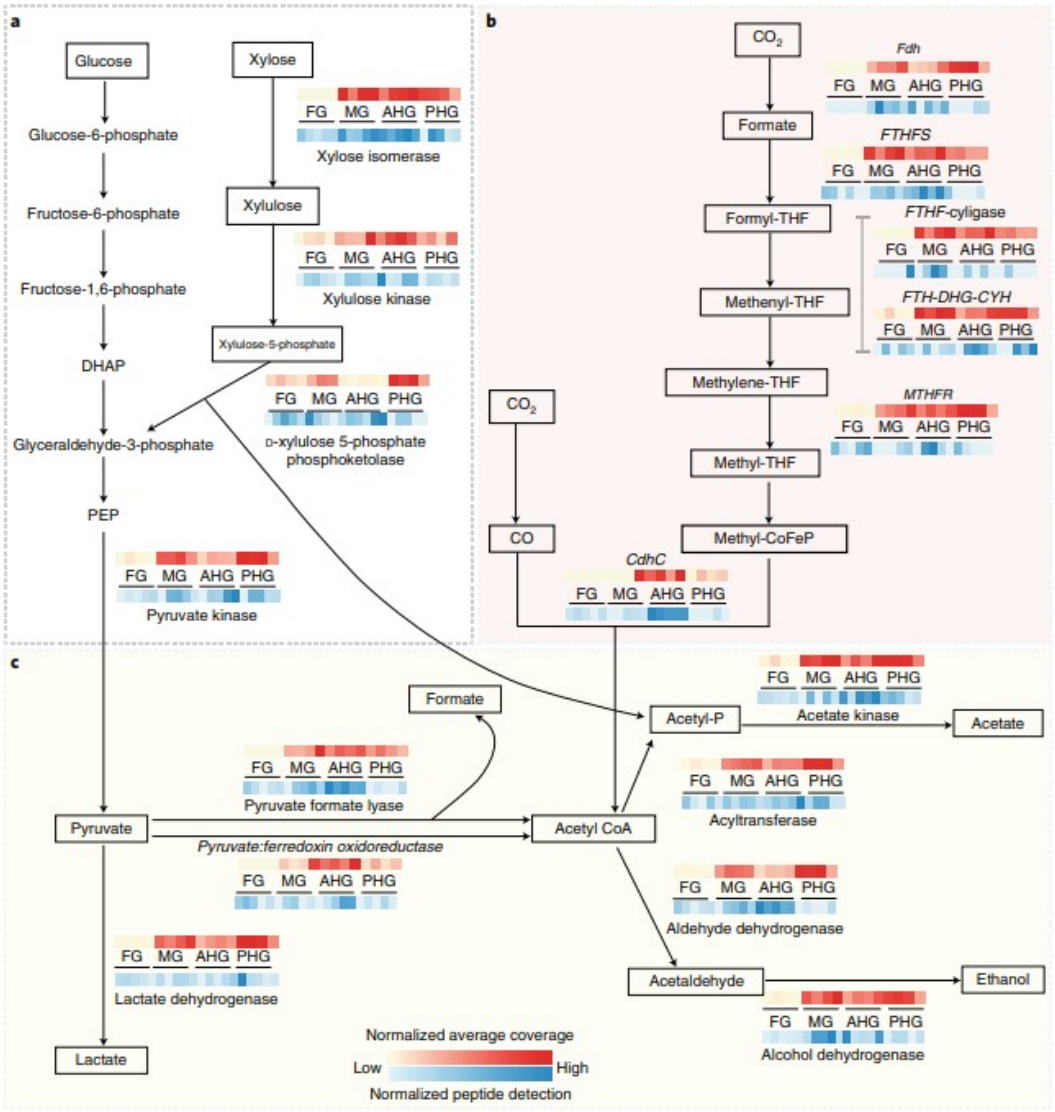


**Fig. 4 | Schematic representation of the coverage and expression of microbial metabolic pathway distribution through the beetle gut from the metagenomic and metaproteomic analyses.** For the metagenomic analyses, the cream-to-red colour gradient indicates increased coverage averaged by gut region across four individual beetles. For the metaproteomic analyses, the white-to-dark-blue colour gradient indicates increased counts of detected peptides averaged by gut region across five individual beetles. **a**, Glucose is converted into pyruvate via the EMP pathway, while xylose is converted into xylulose and xylulose-5-phosphate by xylose isomerases and xylulose kinases. The produced xylulose-5-phosphate feeds the EMP pathway and acetate production by the activity of the xylulose-5-phosphate phosphoketolase. DHAP, dihydroxyacetone phosphate; PEP, phosphoenolpyruvate. **b**, One of the pathways for the production of acetate in the beetle gut is the Wood–Ljungdahl pathway via both the methyl and carbonyl branches. CdhC, CO dehydrogenase/acetyl-CoA synthase; CoFeP, corrinoid sulfur protein; Fdh, formate dehydrogenase; FTHF-cyligase, formyltetrahydrofolate cyclo-ligase; FTHFS, formate-tetrahydrofolate ligase; MTHFR, methylenetetrahydrofolate reductase; TFH-DHG-CYH, tetrahydrofolate dehydrogenase/cyclohydrolase; THF, tetrahydrofolate. **c**, Metabolism of pyruvate results in the production of lactate, formate and acetyl-CoA by the action of enzymes whose highest levels of coverage and expression were detected mostly at the AHG. Acetyl-CoA is converted into acetate and ethanol by the activity of acetate kinase and aldehyde/alcohol dehydrogenases.

To identify key bacterial and archaeal players in lignocellulose deconstruction and fermentation processes across the gut regions, we reconstructed genomes using shotgun metagenomics. A total of 76 genome bins were obtained, with 15 bins probably originating from a single strain/population with genome completeness ranging from 15–98% (Supplementary Table 2). The phylogenetic affiliation of these 15 genomes and the microbial composition of the whole metagenome (Supplementary

Datasets 5 and 6) were determined through analysis of their bacterial/archaeal markers. Genome coverage within each gut section was calculated as a metric for relative abundance (Supplementary Dataset 4 and Supplementary Fig. 4a). These 15 genome bins were further annotated to quantify the distribution of genes coding for enzymes involved in lignocellulose deconstruction, nitrogen fixation, fermentation, acetogenesis and methanogenesis (Supplementary Fig. 4b).

In the midgut, reconstructed genomes were identified as *Turicibacter*-g13c1 and -g78c1, *Clostridiales*-g10c4, *Novosphigobium*-g64c4 and *Methanobrevibacter*-g73c4. Across the heterotrophs, genes coding for the enzymes required for cellulose and xylan deconstruction were detected (Supplementary Fig. 4b), including glycoside hydrolases from the GH5 (endocellulases, endoxylanases and cellobiohydrolases), GH8 (endocellulases and endoxylanases), GH10 (endoxylanases), GH43 (cellobiohydrolases and β-glucosidases), GH1 and GH3 (cellodextrinases, β-glucosidases and β-xylosidases) and AA10 (lytic polysaccharide monooxygenase; only in *Turicibacter*-g13c1) families, in addition to xylose isomerases. These genomes also contained polyphenol laccases and Fe–Mn superoxide dismutases (only in *Turicibacter*-g13c1 and -g78c1)—a possible indicator of their role in the deconstruction of lignin in the midgut region. *Novosphingobium*-g64c4 was also detected with high coverage in the PHG and contained glycoside hydrolases from families GH3 (cellodextrinases, β-glucosidases and β-xylosidases) and GH43 (xylanases and β-xylosidases), together with a wide array of putative ligninolytic enzymes, including polyphenol laccases, Fe–Mn superoxide dismutases, feruloyl esterases, ring-opening dioxygenases, and catechol and benzoate dioxygenases. All midgut genomes also contained genetic elements involved in nitrogenase formation (Supplementary Fig. 4b).

Fermentation, acetogenesis and methanogenesis are enriched in the gut region with the thickest gut wall

Metagenomic and metaproteomic analyses showed that some cellulose and xylan depolymerization probably continues in the AHG (~1.8 cm in length), and that glucose and xylose are primarily converted into pyruvate in this region. While glucose enters to the Embden–Meyerhof–Parnas (EMP) pathway for the production of pyruvate, xylose is transformed into glyceraldehyde-3-phosphate via xylose isomerase, xylulose kinase and xylulose-5-phosphate phosphoketolase, feeding the pentose phosphate to EMP pathway for pyruvate production (Fig. 4a). Xylose is also converted into acetylphosphate via xylulose kinase and xylulose-5-phosphate phosphoketolase for direct acetate production (Fig. 4a). Pyruvate and acetyl phosphate generated from glucose and xylose are further transformed in the AHG, where genes involved in pyruvate fermentation that mediate the production of acetyl-CoA are highly expressed (Fig. 4c). According to the reconstructed pathways, and as confirmed by NMR analysis (Supplementary Fig. 1), the fermentation of pyruvate in the beetle gut results in the production of substrates that can be

utilized by both the host and the associated microbiome, including lactate, formate, acetate and ethanol (Fig. 4c).

Key contributors to the continuous deconstruction of lignocellulose and the fermentation of xylose and glucose in the AHG were also identified by genome binning. Genomes *Parabacteroides*-g0c3 and -g38c4, *Erysipelothrix*-g12c1 and *Lactococcus*-g6c4 contained the genes encoding candidate glycoside hydrolases from the families GH5, GH9 (only for -g38c4), GH8 (only in *Lactococcus*-g6c4), GH43, xylose isomerases and some of the genes involved in lignin oxidation, such as multi-copper laccases, DyP-type peroxidases and Fe–Mn superoxide dismutases (in *Parabacteroides*-g0c3). These genomes also contained genetic potential for the utilization of pyruvate and acetyl-CoA for the production of lactate, formate, acetate and ethanol (Supplementary Fig. 4a,b).

A consensus finding of our metagenomic and metaproteomic analyses suggests that, in addition to direct fermentation, acetate—a critical energy source for the beetle host—is also produced via $CO_2$ reduction through both the carbonyl and methyl branches of homoacetogenesis (Fig. 4b). This process is connected with fermentation processes through the production of acetyl-CoA and $H_2$. Although genes coding for the enzymes of these pathways were detected throughout the beetle gut, CO dehydrogenase/acetyl-CoA synthase had the highest coverage and protein expression in the AHG, suggesting an important role of this region as the main area of fermentation and homoacetogenesis. Another relevant process occurring in the beetle gut that may represent competition for $H_2$, $CO_2$ and/or acetate utilization is methanogenesis. Incubations with living beetles showed methane production at a rate of over 14 $\mu mol\, h^{-1}$ $beetle^{-1}$ (Fig. 5a), and metagenome coverage data suggested that methanogenesis via $CO_2$ reduction was the predominant pathway (Fig. 5b). However, analysis of methane hydrogen and carbon isotopic fractionation indicated that methane was probably produced from both the acetoclastic and $CO_2$ reduction pathways ($\delta^2H$: −320‰; $\sigma = 6.6$; $\delta^{13}C$: −78‰; $\sigma = 3.1$). Proteomics data demonstrated that methanogenesis was most significant within the AHG (inset in Fig. 5b). The production of acetate through fermentation and homoacetogenesis is thermodynamically more favourable than methanogenesis under higher hydrogen partial pressures[29,30]. Using micrometre-scale electrodes, we observed that $H_2$ concentrations were higher in the AHG where maximal concentrations near the centre of the lumen reached 33–44 $\mu mol\, l^{-1}$, with significant axial gradients observed (Fig. 5c and inset). In addition to greater activity of $H_2$-producing pathways in this region, accumulation of hydrogen in the AHG is probably related to its anatomical properties that reduce $H_2$ diffusive loss, with the gut wall thickness (~250 μm) being at least two times greater in this region relative to other regions (PHG = 110 μm; foregut and midgut not thicker than the PHG)[15]. In the AHG, two genomes were found to contain genetic potential for the reduction of $CO_2$: *Methanobrevibacter*-g50c4 (a hydrogenotrophic

methanogen (Supplementary Fig. 4b)) and *Clostridiales*-g64c1 (containing genes for the production of acetate via the carbonyl branch of the Wood–Ljungdahl pathway (Supplementary Fig. 4b)).

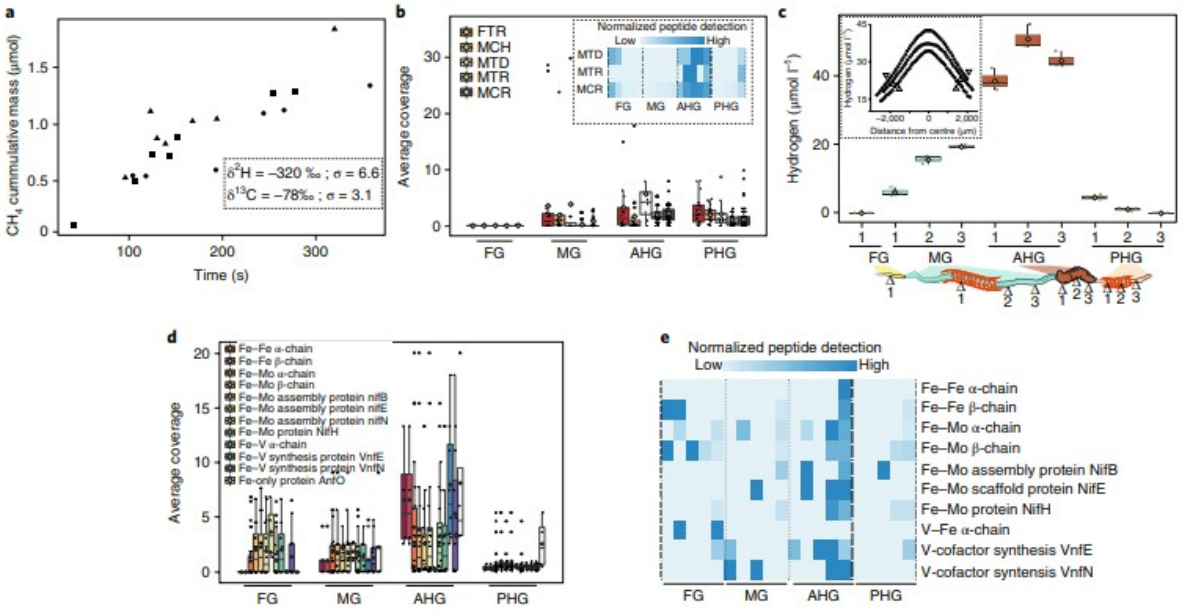


**Fig. 5 | Hydrogen and methane production, as well as nitrogen fixation, occur in the beetle's digestive tract. a**, Methane is produced at a rate of 14 µmol $h^{-1}$ $beetle^{-1}$. Carbon and hydrogen isotopic fractionation suggests a combination of acetoclastic and hydrogenotrophic sources ($n = 3$ beetles; each represented by either squares, triangles or circles). **b**, Average coverage distribution of assembled genes involved in the hydrogenotrophic production of methane ($n = 16$). Inset, proteome-detected enzymes involved in the hydrogenotrophic production of methane (the colour gradient from white to blue indicates peptide detection across each gut region of 5 beetles; $n = 20$). Metagenome and metaproteome analyses show that methane is primarily produced in the AHG by hydrogenotrophic methanogenesis. FTR, formylmethanofuran-tetrahydromethanopterin formyltransferase; MCH, cyclohydrolase; MCR, methyl-coenzyme M; MTD, methylenetetrahydromethanopterin dehydrogenase; MTR, tetrahydromethanopterin S-methyltransferase. **c**, Hydrogen concentrations were highest in the AHG. Axial measurements were performed at the centre of each gut region of three beetles. Inset, a well-defined radial profile was detected only at the AHG ($n = 3$). Arrows indicate the points of electrode contact with the gut wall. **d**, Distribution of coverage of assembled genes coding for nitrogenase elements ($n = 4$ beetles). Fe–Fe, Fe–Mo, Fe–V and iron-only nitrogenase elements were detected with the highest coverage in the AHG. **e**, Nitrogenase expression was highest in the AHG (the colour gradient from white to blue indicates higher peptide detection across each gut region of 5 beetles; $n = 20$). In the box plots in **b** and **d**, points represent single genes per category, while diamonds represent means. Average coverage was calculated as the numbers of reads mapped to each assembled scaffold per sample and normalized to reads per kb per million reads mapped. In the box plot in **c**, points represent biological replicates, and diamonds represent means. Box boundaries represent first and third quartiles of the distribution, while the horizontal line represents the median. Box plot whiskers span 1.5 times the interquartile range of the distribution. Outliers are denoted as large points outside the whiskers.

Nitrogen fixation is another process of importance in arthropods subsisting on diets with high C:N ratios. Metagenomic and metaproteomic analyses allowed the identification of genes coding for subunits of iron-only, Fe–Mo and vanadium nitrogenases (Fig. 5d,e), which were present among all of the reconstructed genomes of the AHG with the exception of *Methanobrevibacter*-g50c4. The highest coverage of genes and protein expression of nitrogenase subunits was also detected in the AHG, in agreement with our previous results for nitrogenase transcription[10], and probably contributing to elevated $H_2$ concentrations in this region.

Depolymerization continues through the beetle gut for energy extraction and the release of nutrient-enriched substrates into the colony

Coverage data suggested that pathways related to depolymerization of lignocellulose are enriched in the relatively more oxic midgut and the PHG

regions. Metaproteomic analysis confirmed the expression of different polyphenol oxidoreductase laccases, catalase-peroxidases, DyP-type peroxidases, Fe–Mn superoxide dismutases, aromatic-ring-opening dioxygenases and protocatechuate dioxygenases (Fig. 3c), again with expression being highest in the midgut and continuing through the AHG and PHG. Metaproteomic analysis also confirmed the expression of 68 metagenome-identified glycoside hydrolase families, 9 of which are involved in cellulose and xylan deconstruction (Fig. 3d). From both metagenomic coverage and metaproteomic data, it appears that although depolymerization of lignocellulose starts in the midgut, providing fermentable substrates to the AHG, it continues in the PHG. The genomes *Novosphingobium*-g64c4 (also abundant at the midgut) and *Lactococcus*-g71c2 were present with high coverage at the PHG. As mentioned previously, *Novosphingobium*-g64c4 encoded candidate β-glucosidases and xylosidases, as well as a broad array of genes coding for enzymes involved in lignin deconstruction. In contrast, *Lactococcus*-g71c2 contained all of the genetic elements encoding candidate enzymes for cellulose and xylan deconstruction, including xylose isomerases (Supplementary Fig. 4).

Our results demonstrate that the primary depolymerization region in the passalid digestive tract is the midgut, which feeds the fermentation vessel (the AHG), where acetogenesis and nitrogen fixation are key to hosting energy and nutrition. Depolymerization and some degree of fermentation continue at the PHG, where the residual material is packed and released in the form of frass. Our results also show that the produced frass is not a waste product, but a substrate that contains elevated concentrations of nitrogen, sugar monomers and aromatic monomers, together with products of fermentation such as acetate, ethanol, lactate and formate (Fig. 2 and Supplementary Figs. 1 and 3), thus providing a further energy and nutrient source for the beetle’s offspring.

## Discussion

Microbial associations are known to enhance the fitness of many living organisms. However, the establishment of beneficial microbial partners within a host requires more than the acquisition of the microorganisms. The host itself must provide numerous habitats developed through co-evolutionary processes[5,31] to sustain and promote the key microbial metabolic functions on which the arthropod and its whole colony rely. Several studies have shown the importance of host–microbe associations for energy extraction from lignocellulose in different arthropods and ruminants[32,33,34,35,36]. However, most have either focused on analysis of the whole gut (or pooled samples in the case of small arthropods) or considered only larger gut compartments (as in the case of some termites, beetles or ruminants), often overlooking the possible contributions of other digestive tract regions that are probably critical for biomass transformation and nutrient acquisition[36,37,38]. Here, we describe the interplay between gut anatomical properties and the distribution of microbial processes for the

transformation of lignocellulosic biomass through the four main anatomical regions of the digestive tract of an animal—the wood-feeding beetle *O. disjunctus*. Our results show that the anatomical development of the passalid digestive tract has crafted specific environments favouring microbial metabolic processes that enable subsistence on a recalcitrant diet (Fig. 6). For the effective use of woody biomass as a food and energy source, aerobic and anaerobic metabolism must co-occur, with spatial compartmentalization enhancing their relative efficiency while also allowing essential coupling to occur between these processes. We propose that the anatomical properties of the beetle enable this through the promotion of radial and axial redox gradients where oxidative processes can co-occur and feed fermentative processes. In the case of the passalid beetle, a gas-permeable midgut with oxygen supply from tracheal tubes[15,39] selects for microbial metabolic traits necessary for lignin oxidation, in tandem with cellulose and xylan depolymerization into simple sugars. The restricted diffusion of oxygen into (and hydrogen out of) the anterior hindgut promotes the establishment of bacteria that ferment these simple sugars, and directs the flow of electron equivalents away from substrates that do not contribute directly to host energy (methane) into substrates (such as acetate and fixed nitrogen) that are key to host energy and nutrition. Several studies analysing the behaviour and feeding habits of passalids[12,40] have shown that larvae, which are known to be unable to feed on hardwoods due to their underdeveloped mandibles[41,42], spend their time in the colony near the adults, feeding on pre-digested wood that seems to be the only type of diet that allows the larvae to survive in the colony and progress through developmental stages[13,18]. We show here that adult beetle gut microbiome processing of low-nutrient woody material results in an energy- and nutrient-rich product, on which their offspring subsist and develop (Fig. 6), demonstrating that the benefit of microbial associations extends across generations. By illuminating the mechanisms underlying interactions between host anatomy, the formation of distinct physicochemical gradients and the selection for microbial metabolic traits, we can begin to develop generalizable models for host–microbiome interactions and their importance in host adaptation to specific niches, as well as inform the design of processes for biomass-to-bioproduct production.

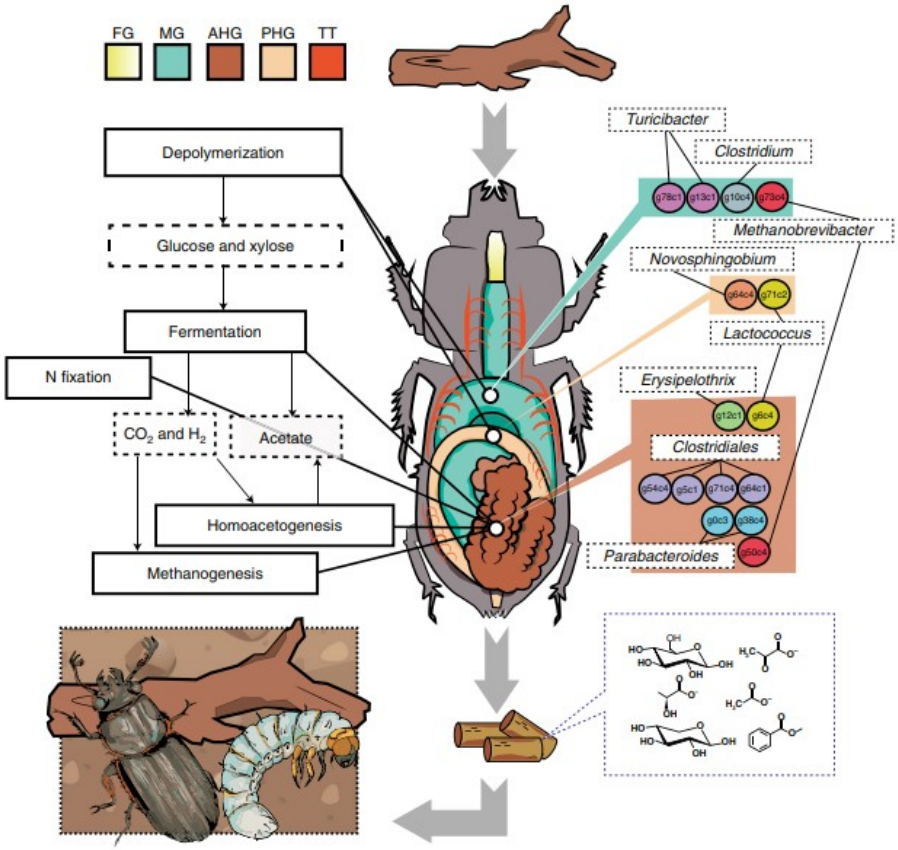


**Fig. 6 | Distribution of processes for the deconstruction and fermentation of lignocellulose in the digestive tract of *O. disjunctus*, and key microbial players.** Passalid beetles feed on decaying wood that is sequentially transformed in the four main regions of the insect's digestive tract. Wood polymers are primarily deconstructed in the MG and PHG. Depolymerization products are fermented in the AHG, where nitrogen fixation, methanogenesis and homoacetogenesis also occur. Each gut region has selected for different microbial groups that contain the genetic potential to perform the detected lignocellulose transformation processes. Through their feeding activity, insects release enriched frass that is required by their offspring to complete their development. Depicted chemical compounds include (from left to right) glucose, pyruvate, lactate, acetate, xylose and benzoate.

## Methods

### Colony collection and maintenance

Adult specimens of *O. disjunctus* were collected at the Burden Research Plantation of the Louisiana State University in Baton Rouge, Louisiana, USA (30° 40′ 59″ N, 91° 10′ 29″ W). Beetles were transported to the Lawrence Berkeley National Laboratory with pieces of pecan wood. Once received, the beetles were placed in sealed terrariums modified to allow the flow of air, and the pecan wood was replaced with water-soaked second-stage-decayed oak wood. These beetles were left to form a colony while feeding on the provided wood and producing frass that covered the surface of their containers (Supplementary Fig. 5). Containers were kept at room temperature and provided with a mist of water twice per week to provide humidity. New batches of beetles were shipped on pecan wood every time they were collected. They were left to starve for 24 h in individual containers, and their produced frass was removed continuously. After the starvation period, the beetles were introduced to our original colonies and left to acclimatize for at least one month before any experimental procedures were

performed. Only fully developed adult beetles (recognized by the black colour of their cuticle) were randomly selected (regardless of sex) from our colonies for each experiment. Given that our previous studies showed little variation in microbiome composition among our colony beetles[10], a minimum of three to six specimens were used for each assay.

### Transit time of wood diet

Four beetles were fed with a pellet of wood consisting of 2.5 g pulverized oak wood with 0.08 g agarose and fluorescent spheres (Fluospheres; Invitrogen) at a final concentration of 3,000 particles per mg of diet. Before feeding, the beetles were left to starve overnight and the excreted frass was removed every time it was produced to avoid re-ingestion. After the starving period, the beetles were fed pellets of the diet and the time of the first feeding event was recorded. The beetles were fed on the diet with fluorescent spheres for 12 h. After the 12 h feeding time, the diet was replaced with pellets of wood without the fluorescent spheres. Produced frass was collected every time excretion occurred from the beginning of the feeding experiment to two days after diet replacement. Collected frass samples were inspected for the presence of the spheres using a microscope at 40× magnification under ultraviolet light. Two time points were of special interest: the time after the first spheres appeared in the frass and the time at which all spheres were cleared out.

### Carbon and nitrogen concentrations in wood and frass

Five frass samples collected from 5 beetles were loaded in 8 mm × 5 mm tin capsules (Elemental Microanalysis) using 10–20 mg for nitrogen quantification and 2 mg for carbon. The capsules were then folded and loaded into a zero blank autosampler connected to an ECS 4010 Element Analyzer (Costech Analytical) coupled to a Delta V Plus isotope ratio mass spectrometer (Thermo Fisher Scientific). Based on the laboratory atropine standard associated with the runs, the analytical precisions on concentration yields were $4.83 \pm 0.09$ wt% (1 s; $n = 10$) for nitrogen and $70.14 \pm 1.29$ wt% (1 s; $n = 7$) for carbon.

### Detection of lignocellulosic materials in wood and frass

Attenuated total reflectance (ATR) FTIR spectroscopy was used for the detection of the changes in plant polymer content across the digestive tract of *O. disjunctus*. Five beetles were fed with a pellet of wood—2.5 g pulverized oak wood in 10 ml of 0.8% agarose in sterile water. Before feeding, the beetles were left to starve overnight and the their frass was removed every time it was produced to avoid re-ingestion. After the starving period, the beetles were provided with known amounts of the diet and left to feed for 12 h. These starving/feeding incubations were repeated for ten days to allow clearing of the gut content and account only for the diet passage through the insect’s digestive tract. Frass produced on the last day of incubation was stored for ATR analysis.

The infrared spectra of the frass samples were obtained using ATR FTIR spectroscopy. The frass samples were gently pressed down on a clean surface of the germanium crystal in an ATR configuration (Smart Orbit; Thermo Fisher Scientific). Infrared light beamed from the interferometer (Nexus 870; Nicolet) was focused onto the interface between the sample and the top surface of the crystal through the lower facet. The sample spectrum was recorded with a spectral resolution of 4 $cm^{-1}$ over the infrared range (400–4,000 $cm^{-1}$). The signature vibrational peaks of cellulose/xylan, lignin and nitrogen used in this study locate at ~997 $cm^{-1}$ for the glucosidic linkages and 1,730 $cm^{-1}$ for the ester groups in cellulose and xylan, respectively, ~1,510 $cm^{-1}$ for the C = C stretch in aromatic groups, and ~1,550 $cm^{-1}$ for N–H bending and C–N stretching for amide II, respectively. A multi-oscillator model was used to extract the vibrational strengths from the absorption peaks with a fitting coefficient of determination ($R^2$) greater than 0.999, and we retrieved the peak areas for each vibrational signature, which were proportional to the corresponding concentrations by Beer's law. All of the spectra were normalized to their total spectral weights after polynomial baseline correction.

For FTIR, bulk chemistry and $^{13}$C-tetratmethylammonium hydroxide, significant differences between treatments (diet versus frass) were determined using one-way analysis of variance and least significant differences with Bonferroni correction using the R software environment[43].

Metabolite analyses in the frass using NMR

Ten beetles were removed from their colonies and left overnight in separated containers while removing any pieces of their produced frass. After this starving incubation, the beetles were separated into 2 groups, where 5 individuals from each group were fed with the oak wood diet described above and the other 5 were fed with an artificial diet consisting of 10% lignin, 45% sigmacell cellulose and 45% beechwood xylan mixed in 0.8% agarose (Sigma). Individual beetles from each group were placed on a Petri dish with pieces of their corresponding diets, and kept in the dark while being monitored for the production of frass and its removal. Beetles were fed with their corresponding diets for 10 d, and the last pieces of the produced frass (3 pieces per beetle) were stored in 100 mM ammonium bicarbonate buffer (pH 8.0). Metabolites were extracted from the frass using a chloroform, methanol and water extraction procedure[44]. Homogenized passalid frass (150 µl) in ammonium bicarbonate buffer was combined with 600 µl methanol. The suspension was vortexed, combined with 450 µl chloroform and vortexed again to mix. A final 150 µl of $H_2O$ was added to the suspension and vortexed, after which the samples were placed at −20 °C for 30 min for phase separation. Samples were centrifuged at 21,000*g* for 2 min, after which the top layers containing polar metabolites were transferred to a microfuge tube and dried in a SpeedVac overnight. Dried samples were stored at −80 °C until they were used for analysis. Dried samples were dissolved in 250 µl 90% $H_2O$, 10% $D_2O$, 0.5 mM d6-4,4-dimethyl-4-

silapentane-1-sulfonic acid and 0.03% sodium azide, and transferred to 3 mm NMR tubes for analysis. Samples were analysed using an Agilent VNMRS 600 MHz (proton frequency) NMR spectrometer equipped with a cryogenically cooled HCN triple resonance probe. One-dimensional NOESY spectra were collected with a spectral window of 12 ppm, an acquisition time of 4 s, a recycle delay of 1 s, a 100 ms mixing time and 1,024 scans. The temperature was regulated at 25 °C. NMR data were zero-filled to twice their original size, apodized (0.5 Hz line broadening), Fourier transformed and baseline corrected using the Chenomx NMR Suite. Spectra were referenced to d6-4,4-dimethyl-4-silapentane-1-sulfonic acid at 0 ppm. NMR spectra were profiled using the Chenomx NMR Suite for identification of metabolites. Pearson correlation coefficients were calculated using the R software.

Carbon and hydrogen isotope ratios and $CH_4$ concentrations produced by living beetles

Three living specimens of *O. disjunctus* were placed in glass vials and sealed with a rubber stopper. Then, headspace aliquots were sampled at different time intervals. Concentrations and carbon isotope ratios were measured via direct injection into the inlet of a Tracegas pre-concentrator interfaced with a Micromass JA Series Isoprime isotope ratio mass spectrometer (TG-IRMS; Micromass). Briefly, the gas samples were injected into an evacuated inlet tube (~3 Torr) using a gas-tight syringe. The samples were transferred to a liquid nitrogen trap by flushing the inlet tube with ultra-pure helium. During the sample transfer, water, $CO_2$ and CO were removed using traps in series consisting of magnesium perchlorate, Carbosorb and Sofnocat Pt–Pd–SnO oxidation catalyst, then the $CH_4$ was oxidized to $CO_2$ at 1,000 °C in a ceramic tube loaded with nickel, copper and platinum wires. By heating the liquid nitrogen traps, the $CO_2$ was transferred and separated chromatographically from $N_2O$ on a PoraPLOT Q-fused silica capillary column (30 m × 0.32 mm), and the carbon isotope ratios were measured in the mass spectrometer. Carbon isotope ratios are reported in the conventional δ-notation relative to the Vienna PeeDee Belemnite (VPDB) scale. Repeated injections of laboratory standards associated with sample analysis yield $\delta^{13}C$ values of $-38.84 \pm 0.34‰$ (1σ; $n = 13$) for $CH_4$. Sample $CH_4$ concentrations were determined using the mass 44 ($CO_2$) area peak.

Hydrogen isotope ratios of $CH_4$ were determined using a gas chromatograph and an isotope ratio mass spectrometer interfaced with a pyrolysis reactor (GC-P-IRMS; Thermo Fisher Scientific). The gases were sampled using a gas-tight syringe and transferred to a stainless-steel loop mounted on a 6-port valve (Valco). Samples were transferred into the gas chromatograph by switching the 6-port valve. $CH_4$ was separated chromatographically on an HP Molesieve-fused silica capillary column (30 m × 0.32 mm). Following gas chromatography separation, the $CH_4$ was pyrolysed in an empty ceramic tube at 1,450 °C. Hydrogen isotope ratios were acquired in the isotope-ratio mass spectrometer. Hydrogen isotope ratios are reported in the conventional

δ-notation relative to Vienna Standard Mean Ocean Water (VSMOW) scales. Repeated injections yield $\delta^2H$ values of $-141.0 \pm 6.9‰$ (1σ; $n = 5$) for $CH_4$.

## Measurement of hydrogen concentration profiles

A set of three beetles was used to measure the concentration and profiles of hydrogen at each gut region using Clark-type hydrogen microelectrodes ($H_2$-25; Unisense). Before use, the electrodes were polarized overnight and calibrated in water saturated with 10 and 20% hydrogen gas in the CAL 300 calibration chamber (Unisense), as well as in degasified water, which represented our 0% $H_2$ calibration point. Calibration was carried out before performing the measurements in each beetle. The current was measured with a Unisense microsensor multimeter and recorded using SensorTracePRO software (Unisense). Before microelectrode measurements, 30 ml of low-melting-point agarose consisting of 1% agarose in insect Ringer's solution (111 mM NaCl, 3.3 mM KCl, 4.5 mM $CaCl_2$ and 2.8 mM $Na_2CO_3$) was cast into a microchamber. A freshly dissected gut was placed on this layer of agarose, fully extended and immediately covered with a second layer of molten agarose at 30 °C. Microelectrodes were positioned using a motorized micromanipulator (MXU2; PyroScience). Measurements were performed radially starting at the surface of the gut wall (0 µm) through the beetle gut until the tip completely penetrated the whole tissue. The progress of the tip was followed with a digital microscope (44032 Celestron) connected to a computer. All measurements were carried out at room temperature and over a time span of ~20 min from the dissection to the actual electrode measurements. Hydrogen profiles were measured at three different points of each gut region. To correct for differences in the length of each dissected gut, sampling points were chosen based on the selected sampling positions in the first beetle. These sampling positions were linearly measured from the beginning of the analysed gut region, and divided by its total length: (sampling point in cm)/(total length of the gut region in cm). The calculated correction factors were multiplied by the total length of the different gut regions to define sampling points in each beetle. A radial profile was only detected at the AHG, but measurements for the other regions are reported as the hydrogen concentration at the centre of a given gut region.

## DNA extraction and library preparation

Four passalid beetles that were part of a long-term colony feeding on oak wood were selected, and their digestive tracts were extracted and separated into their four morphologically differentiated sections. Each gut section was then cut open while submerged in a house-made solution of RNAlater. The lumen content of each section, as well as the material attached to the gut wall, was removed from the beetle tissue. The separated lumen content from each section was then kept in RNAlater at 4 °C overnight. RNAlater was removed before DNA extraction by centrifuging at 8,000*g* for 5 min. The extracted gut contents were then transferred to a Lysing Matrix E tube (MP Biomedicals) containing 500 µl of lysis solution (1% cetyltrimethylammonium

bromide in 0.5 M NaCl). The samples were bead-beaten in a FastPrep Instrument (MP Biomedicals) at 4 m $s^{-1}$ for 30 s. Lysozyme buffer (180 µl; 20 mM Tris-HCl (pH 8.0), 2 mM ethylenediaminetetraacetic acid and 1.2% Triton X) containing 10 mg $ml^{-1}$ of lysozyme was added to each tube and the mixture was incubated at 37 °C for 30 min. Some 5 µl of Proteinase K (Ambion) was added to the tubes, mixed by tapping and incubated at 56 °C for 30 min. After the Proteinase K lysis, 500 µl of phenol:chloroform:isoamyl alcohol (25:24:1) was added to each tube, and the samples were bead-beaten in the FastPrep instrument as before. The samples were centrifuged at 10,000*g* for 1 min at 4 °C and the supernatants were transferred to a MaXtract High Density Tube (Qiagen) containing 500 µl of chloroform:isoamyl alcohol (24:1). The samples were centrifuged at 10,000*g* for 1 min at 4 °C, and the supernatants were transferred to a microcentrifuge tube containing 1 ml of isopropanol and 1 µl of linear acrylamide (Ambion). The DNA/RNA mixture was precipitated by incubating for 10 min at room temperature, and centrifuged at 10,000*g* for 5 min at 4 °C, before the isopropanol was removed. The obtained pellet was washed with 70% ethanol and centrifuged at 10,000*g* for 1 min at 4 °C. The ethanol was completely removed and the pellet was dissolved in diethyl pyrocarbonate-treated water. The DNA was purified with the QIAamp DNA Micro kit (Qiagen) following the manufacturer's protocol, which was modified to treat the samples with RNase (Ambion) for 5 min before the washing steps. The DNA concentration was determined with the use of a Qubit fluorometer (Qiagen) and the DNA size was assessed by bioanalyzer and gel electrophoresis. The extracted DNA from each sample was diluted to yield a final mass of 100 ng for metagenomic library preparation at the QB3 Vincent J. Coates Genomics Sequencing Laboratory. In resume, DNA was sheared using Covaris and the library preparation was performed on the IntegenX Apollo 324 robot.

Metagenome sequencing quality control, trimming and filtering

Each sample was sequenced using an Illumina platform (HiSeq 2500 Rapid Run) to obtain paired ends of 150 base pairs (bp) with 400-bp inserts. The quality score profile of each sample was checked using FastQC. On the basis of this, each sequence was trimmed 10 bp from the 5′ end and 15 bp from the 3′ end to remove regions with sequencing biases. The mean Q scores for all trimmed sequences were >30. Low-complexity sequences were removed using the dust approach with the percentage of ambiguous bases set to <50% and the maximum allowed dust score set to 50 using PRINSEQ (http://sourceforge.net/projects/prinseq/files/standalone/). The orphan reads (that is, only one pair passing the filtering process) were removed using an in-house script.

Metagenome co-assembly and annotations

Trimmed and filtered sequences from all samples were concatenated into a single fastq file, giving a total of 30,787 megabase pairs of sequence from 16 samples. These sequences were co-assembled using an in-house iterative

metagenome assembly pipeline. In each iteration, the assemblies were performed using Ray Meta[45], the mapping of reads to the scaffolds was performed using Bowtie 2 (ref. [46]), and the post-processing of sequence alignments was performed using SAMtools[47] and custom R and python scripts. The iterative assembly consists of the following steps. (1) The first step is to determine the bins of coverage in the dataset with an initial assembly with a *k*-mer size that will cover the whole range of genome abundances, and hence coverages. Each coverage bin will correspond to one or more genomes. For this initial assembly, we set the *k*-mer size to roughly one-third of the mean read length. In practice, a few different trials of *k*-mer sizes might be warranted to obtain an accurate estimate of coverage bins. Ray Meta was run with the scaffolding option turned off, and the minimum contig length was set to 1 kilobase (kb). (2) Next, the coverage distribution plot (that is, the number of base pairs versus the *k*-mer coverage) of the contigs from the initial assembly is calculated and plotted using custom R scripts. The peaks of coverage bins in the coverage distribution plot are identified and each of the coverage bins is assembled in a separate iteration from the highest to lowest coverage bin using estimated optimized assembly parameters for each. The parameter set for each bin included the *k*-mer size, minimum-seed coverage depth and maximum-seed coverage depth. For each bin, whenever feasible, we chose a *k*-mer size giving a *k*-mer coverage between 30 and 50 under the constraint that $k \geq 31$. (3) Iterative assembly: each iteration consisted of the assembly of unmapped reads from the previous iteration, with the optimized parameters turning on the scaffolding option. The scaffolds from the current bin of coverage are accumulated into the final set of scaffolds. We checked all of the final scaffolds for the presence of chimeric assemblies using paired-end read and depth of coverage consistency with in-house scripts. Obtained scaffolds were in reads per kilobase of assembled contig per million mapped reads (RPKM), normalized to account for sequencing depth and scaffold length. The number of reads for each assembled contig in each sample was normalized to RPKM. Post-binning, to rule out the effect of host reads on normalization, we checked the contribution of potential host reads to the total coverage within each metagenome and across gut regions (Supplementary Table 3). RPKM-normalized coverage values were used as proxies for the abundance of each scaffold/contig in a sample. ORFs were predicted from the final set of scaffolds using Prodigal's[48] meta procedure (-p meta). Protein domain annotations were predicted using HMMER's hmmscan[49] against the PfamA-30 and dbCAN-V6 Hidden Markov Model protein domain databases using an e-value of $1 \times 10^{-5}$. Identified carbohydrate-active proteins were screened against the National Center for Biotechnology Information database using BLASTp with an e-value of $1 \times 10^{-5}$. Obtained BLAST alignments were imported to the MEGAN community edition (version 6.11.7) software for taxonomy parsing and separation of microbial and host carbohydrate-active enzymes. BLAST alignments of the host-identified enzymes were screened for gaps that could indicate erroneous protein predictions. Statistical

significance of the coverage distribution of identified genes was assessed using the Kruskal–Wallis test, and pairwise comparisons were carried out using the Wilcox test and the Benjamini–Hochber method for *P* value adjustment using R software.

## Sample preparation for metaproteomics and metabolomics

Beetle gut luminal content from the 4 main regions of the digestive tract of 5 beetles was placed into individual 2 ml snap-cap centrifuge tubes (Eppendorf) along with 400 µl of 100 mM ammonium bicarbonate (pH 8) and a 3 mm stainless-steel bead. The tissues were homogenized using a TissueLyser II system (Qiagen) at 30 Hz for 2 min on each side with ice-cold adaptors. On ice, a handheld pestle homogenizer (Fisher Scientific) was used to complete the homogenization. The sample was centrifuged at 21,000*g* for 10 min at 4 °C. The supernatant was removed into a methanol/chloroform-compatible 15 ml tube (Genesee Scientific), and methanol/chloroform extraction was done by adding ice-cold (−20 °C) cholorform:methanol mix (prepared 2:1 (*v/v*)) to the sample in a 5:1 ratio over the sample volume and vigorously vortexed. The sample was then placed on ice for 5 min and then vortexed for 10 s followed by centrifugation at 10,000*g* for 10 min at 4 °C. The upper, water-soluble metabolite phase was removed and placed in a labelled glass vial and dried completely for metabolomics analysis. Then, 1 ml of ice-cold methanol was added to the precipitated protein. The sample was again centrifuged at 10,000*g* for 10 min to pellet the protein, and the methanol was decanted off. The remaining protein was placed in a fume hood to dry. The precipitated protein pellet and pelleted debris were suspended in 8 M urea, reduced with 10 mM dithiothreitol and incubated for 30 min at 60 °C with constant shaking at 800 r.p.m. Denatured and reduced samples were diluted tenfold with 50 mM ammonium bicarbonate pH 7.8, and $CaCl_2$ was added to a final concentration of 1 mM before enzymatic digestion. Sequencing-grade modified trypsin was activated by incubating for 10 min at 37 °C. 5 µg of activated trypsin was then added to the samples, and they were digested at 37 °C for 3 h with constant shaking at 800 r.p.m. Reactions were quenched by rapid freezing in liquid nitrogen. Digested samples were desalted using a four-probe positive pressure Gilson GX-274 ASPEC system (Gilson) with Discovery C18 100 mg/1 ml solid-phase extraction tubes (Supelco) using the following protocol: 3 ml of methanol was added for conditioning, followed by 2 ml of 0.1% trifluoroacetic acid (TFA) in $H_2O$. The samples were then loaded onto each column, followed by 4 ml of 0.1% TFA in $H_2O$:acetonitrile (ACN) at a ratio of 95:5. Samples were eluted with 1 ml of 0.1% TFA in ACN:$H_2O$ at a ratio of 80:20. The samples were concentrated down to ~30 µl using a Speed Vac, and a final bicinchoninic acid assay was performed to determine the peptide concentration of each sample for normalization and subsequent liquid chromatography and mass spectrometry measurements.

## High-pH reversed-phase C18 fractionation

Aliquots of individual sample protein digests were subjected to high-pH reversed-phase liquid chromatography fractionation. All samples were diluted to a volume of 900 µl with 10 mM ammonium formate buffer (pH 10.0) and resolved on an XBridge C18 (250 mm × 4.6 mm; 5 µM) with 4.6 mm × 20 mm guard column (Waters). Separations were performed at 0.5 ml $min^{-1}$ using an Agilent 1100 series high-performance liquid chromatography system (Agilent Technologies) with mobile phases A (10 mM ammonium formate, pH 10.0) and B (10 mM ammonium formate, pH 10.0:acetonitrile at a ratio of 10:90). The gradient was adjusted from 100% A to 95% A over the first 10 min, 95% A to 65% A over minutes 10–70 and 65% A to 30% A over minutes 70–85, then maintained at 30% A over minutes 85–95, re-equilibrated with 100% A over minutes 95–105 and held at 100% A until minute 120. Fractions were collected every 1.75 min, starting at 10 min and ending at 115 min (60 fractions over the entire gradient). The plate was partially dried in a SpeedVac, then every 12th fraction was combined for a total of 12 samples (each with $n = 5$ fractions pooled) with 50% acetonitrile rinsing. The fractions were then completely dried down and 25 µl of 25 mM ammonium bicarbonate was added to each fraction for storage at −20 °C until liquid chromatograpgy tandem mass spectrometry analysis.

## Liquid chromatography tandem mass spectrometry

A Waters NanoAcquity liquid chromatographer (Waters) was coupled to an LTQ Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). Orbitrap spectra (automatic gain control: $1 \times 10^6$) were collected from 300–1,800 *m/z* at a resolution of 60,000 (full width at half maximum at 200 *m/z*), followed by data-dependent orbitrap higher-energy collisional dissociation top-10 tandem mass spectrometry (centroid mode) at a resolution of 7,500. Spectra were acquired with a collision voltage set to 32 V. The heated capillary temperature and spray voltage were 325 °C and 2.2 kV, respectively. Charge-state screening was enabled to reject singly charged ions, and the default charge state was set to 4. The mass spectrometer was outfitted with a custom electrospray ionization interface. Electrospray emitters were custom made using 360 µm o.d. × 20 µm i.d. chemically etched fused silica[50].

Data were acquired for 100 min after a 25-min delay from when the gradient started. The sample was first cleaned up using a solid-phase extraction column packed in house by slurry packing 3.6 µm Aeris WIDEPORE C18 (Phenomenex) into 5 cm × 360 µm o.d. × 150 µm i.d fused silica (Polymicro Technologies) using a 1 cm sol-gel frit on each end for media retention[51] for 20 min. The reversed-phase column was prepared in house by slurry packing 3.0 µm Jupiter C18 (Phenomenex) into 70 cm × 360 µm o.d. × 75 µm i.d fused silica (Polymicro Technologies) using a 1 cm sol-gel frit for media retention. Mobile phases consisted of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) operated at 300 nl $min^{-1}$. The 5 µl injection volume was

focused onto the analytical column, then eluted with a gradient profile (min: %B) of 0:0.1, 2:8, 20:12, 75:30, 97:45 and 100:95, and held for 10 min before going back to the starting conditions. The column was equilibrated for 52 min before the next sample was loaded onto the column.

### Metaproteomic analysis peptide identification and protein roll-up

Mass spectra from the resulting analyses were evaluated using the $MSGF^+$ software with the metagenomic assemblies mentioned above. Briefly, after conversion of the metagenomic assemblies into predicted ORFs (for example, predicted proteins via Prodigal), libraries were then created using the forward and reverse direction to allow determination of the false discovery rate (FDR). The reverse decoy database allowed for the detection of false hits, which in turn allowed for FDR calculation and appropriate filtering of the data to maximize the real peptide identifications while minimizing the spurious ones. $MSGF^+$ was then used to search the experimental mass spectra data against both the forward and reverse decoy databases in the SQL data management server (for example, data manager server). Cut-offs for the data included an $MSGF^+$ spectra probability of $1 \times 10^{10}$ (equivalent to a BLAST e value), a mass accuracy of ±5 ppm, a protein-level FDR of 5% and further requirement for one unique peptide per protein.

### GC-MS-based metabolomics analyses through the beetle gut

Water-soluble metabolites were co-extracted with the proteins as described above and analysed as previously reported[52]. Briefly, extracted metabolites were completely dried in a vacuum and subjected to chemical derivatization for GC-MS analysis. The collected data were processed using the MetaboliteDetector software (version 3.1)[53], and experimental metabolite spectra and retention indices were matched to entries in an in-house version of the Agilent Fiehn metabolomics library, as well as to the NIST 14 GC-MS library, or using spectra alone (denoted ‘NIST’). Metabolomics data were further validated for suspicious identifications on the detected metabolites.

### Genome extraction

The final set of quality-controlled scaffolds from the metagenome were binned into genome bins using Gaussian mixture model clustering using tetranucleotide frequencies and depth of coverage across 16 samples as features. The depth-of-coverage profiles for the scaffolds were estimated based on mapping of the reads from each sample to the scaffolds using Bowtie 2, and calculating the mean coverage for each scaffold from each of the samples normalized by the scaffold length. Gaussian mixture model clustering was performed using CONCOCT[54] using 3, 2 and 1 kb as the minimum contig lengths for clustering. The first clustering run with CONCOCT used a minimum contig length of 3 kb. Obtained bins were evaluated characterized against eukaryotic and prokaryotic protein markers using Phylosift[55], and those consisting of exclusively prokaryotic elements were screened for their content of prokaryotic essential single-copy genes. Nearly

complete bins containing 90–100 unique single-copy genes were extracted and separated from the scaffold's file, and the CONCOCT was run again at the other two contig-length levels (2 and 1 kb), extracting the nearly complete bins after each run. A total of 76 bins were obtained. with 15 probably originating from a single strain/population, and with completeness ranging from 15–98%. Proteins and their corresponding annotations were extracted from the metagenome based on the list of scaffolds corresponding to each genome (see Supplementary Dataset 3). Amino acid sequences from each genome were fed to the AmphoraNet webserver[56] and screened against the bacterial and archaeal protein-coding marker genes of Amphora's database for phylotyping. Coverage information for the scaffolds of each genome was extracted from the calculated coverage data RPKM normalized for each scaffold in the metagenome. Bin abundances in each replicated gut sample were calculated as the average RPKM coverage value over all of the scaffolds in a bin.

References

1. Shapira, M. Gut microbiotas and host evolution: scaling up symbiosis. Trends Ecol. Evol. 31, 539–549 (2016). 2. Beran, F. & Gershenzon, J. Microbes matter: herbivore gut endosymbionts play a role in breakdown of host plant toxins. Environ. Microbiol. 18, 1306–1307 (2016). 3. Ceja-Navarro, J. A. et al. Gut microbiota mediate cafeine detoxifcation in the primary insect pest of cofee. Nat. Commun. 6, 7618 (2015). 4. Warnecke, F. et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450, 560–565 (2007). 5. Schmitt-Wagner, D. & Brune, A. Hydrogen profles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (Cubitermes spp.). Appl. Environ. Microbiol. 65, 4490–4496 (1999). 6. Lemke, T., van Alen, T., Hackstein, J. H. P. & Brune, A. Cross-epithelial hydrogen transfer from the midgut compartment drives methanogenesis in the hindgut of cockroaches. Appl. Environ. Microbiol. 67, 4657–4661 (2001). 7. Pester, M. & Brune, A. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. ISME J. 1, 551–565 (2007). 8. Kohler, T., Dietrich, C., Schefrahn, R. H. & Brune, A. High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (Nasutitermes spp.). Appl. Environ. Microbiol. 78, 4691–4701 (2012). 9. Dietrich, C., Kohler, T. & Brune, A. Te cockroach origin of the termite gut microbiota: patterns in bacterial community structure refect major evolutionary events. Appl. Environ. Microbiol. 80, 2261–2269 (2014). 10. Ceja-Navarro, J. A. et al. Compartmentalized microbial composition, oxygen gradients and nitrogen fxation in the gut of Odontotaenius disjunctus. ISME J. 8, 6–18 (2014). 11. Schuster, J. C. & Schuster, L. B. Social behavior in passalid beetles (Coleoptera: Passalidae): cooperative brood care. Florida Entomol. 68, 266–272 (1985). 12. Wicknick, J. A. & Miskelly, S. A. Behavioral interactions between noncohabiting bess beetles, Odontotaenius disjunctus (Illiger) (Coleoptera: Passalidae). Coleopt.

Bull. 63, 108–116 (2009). 13. Schuster, J. C. & Schuster, L. B. in Te Evolution of Social Behavior in Insects and Arachnids (eds by Choe, J. C. & Crespi, B. J.) 260–269 (Cambridge Univ. Press, 1997). 14. Lindblad, I. Wood-inhabiting fungi on fallen logs of Norway spruce: relations to forest management and substrate quality. Nord. J. Bot. 18, 243–255 (1998). 15. Nardi, J. B. et al. Communities of microbes that inhabit the changing hindgut landscape of a subsocial beetle. Arthropod Struct. Dev. 35, 57–68 (2006). 16. Castillo, M. L. & Reyes-Castillo, P. in Tropical Biology and Conservation Management. Encyclopedia of Life Support Systems Vol. VII (eds Del Claro, K., Oliveira, P. S. & Rico-Gray, V.) 112–133 (2009). 17. Ulyshen, M. D. Wood decomposition as infuenced by invertebrates. Biol. Rev. Camb. Phil. Soc. 91, 70–85 (2014). 18. Pearse, A. S., Patterson, M. T., Rankin, J. S. & Wharton, G. W. Te ecology of Passalus cornutus Fabricius, a beetle which lives in rotting logs. Ecol. Monogr. 6, 455–490 (1936). 19. Urbina, H. & Blackwell, M. Multilocus phylogenetic study of the Schefersomyces yeast clade and characterization of the N-terminal region of xylose reductase gene. PLoS ONE. 7, e39128 (2012). 20. Lichtwardt, R. W., White, M. M., Cafaro, M. J. & Misra, J. K. Fungi associated with passalid beetles and their mites. Mycologia 91, 694–702 (1999). 21. Suh, S. O., Marshall, C. J., McHugh, J. V. & Blackwell, M. Wood ingestion by passalid beetles in the presence of xylose-fermenting gut yeasts. Mol. Ecol. 12, 3137–3145 (2003). 22. Zhang, N., Suh, S.-O. & Blackwell, M. Microorganisms in the gut of beetles: evidence from molecular cloning. J. Invertebr. Pathol. 84, 226–233 (2003). 23. Nguyen, N. H., Suh, S.-O., Marshall, C. J. & Blackwell, M. Morphological and ecological similarities: wood-boring beetles associated with novel xylosefermenting yeasts, Spathaspora passalidarum gen. sp. nov. and Candida jefriesii sp. nov. Mycol. Res. 110, 1232–1241 (2006). 24. Urbina, H., Schuster, J. & Blackwell, M. Te gut of Guatemalan passalid beetles: a habitat colonized by cellobiose- and xylose-fermenting yeasts. Fungal Ecol. 6, 339–355 (2013). 25. Geib, S. M. et al. Lignin degradation in wood-feeding insects. Proc. Natl Acad. Sci. USA 105, 12932–12937 (2008). 26. De Gonzalo, G., Colpa, D. I., Habib, M. H. M. & Fraaije, M. W. Bacterial enzymes involved in lignin degradation. J. Biotechnol. 236, 110–119 (2016). 27. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. Te carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495 (2014). 28. Sabbadin, F. et al. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. Nat. Commun. 9, 756 (2018). 29. Cord-Ruwisch, R., Seitz, H.-J. & Conrad, R. Te capacity of hydrogenotrophic anaerobic bacteria to compete for traces of hydrogen depends on the redox potential of the terminal electron acceptor. Arch. Microbiol. 149, 350–357 (1988). 30. Hoehler, T. M., Alperin, M. J., Albert, D. B. & Martens, C. S. Termodynamic control on hydrogen concentrations in anoxic sediments. Geochim. Cosmochim. Acta 62, 1745–1756 (1998). 31. Douglas, A. E. Multiorganismal insects: diversity and function of resident microorganisms. Annu. Rev. Entomol. 60, 17–34 (2014). 32. Tarmadi, D. et al. Te efects of various lignocelluloses and lignins on physiological responses

of a lower termite, Coptotermes formosanus. J. Wood Sci. 63, 464–472 (2017). 33. Zhou, J. et al. Diversity, roles, and biotechnological applications of symbiotic microorganisms in the gut of termite. Curr. Microbiol. https://doi.org/10.1007/ s00284-018-1502-4 (2018). 34. Brune, A. Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. 12, 168–180 (2014). 35. Shabat, S. K. et al. Specifc microbiome-dependent mechanisms underlie the energy harvest efciency of ruminants. ISME J. 10, 2958–2972 (2016). 36. Scully, E. D., Hoover, K., Carlson, J. E., Tien, M. & Geib, S. M. Midgut transcriptome profling of Anoplophora glabripennis, a lignocellulose degrading cerambycid beetle. BMC Genomics 14, 850 (2013). 37. Abdul Rahman, N. et al. A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. Microbiome 3, 5 (2015). 38. Zhou, M. et al. Assessment of microbiome changes afer rumen transfaunation: implications on improving feed efciency in beef cattle. Microbiome 6, 62 (2018). 39. Westneat, M. W. et al. Tracheal respiration in insects visualized with synchrotron X-ray imaging. Science 299, 558–560 (2003). 40. Jackson, H. B., Baum, K. A., Robert, T. & Cronin, J. T. Habitat-specifc movement and edge-mediated behavior of the saproxylic insect Odontotaenius disjunctus (Coleoptera: Passalidae). Environ. Entomol. 38, 1411–1422 (2009). 41. King, A. & Fashing, N. Infanticidal behavior in the subsocial beetle Odontotaenius disjunctus (Illiger) (Coleoptera: Passalidae). J. Insect Behav. 20, 527–536 (2007). 42. Krause, J. B. & Ryan, M. T. Te stages of development in the embryology of the horned passalus beetle, Popilius disjunctus Illiger. Ann. Entomol. Soc. Am. 46, 1–20 (1953). 43. R Core Development Team R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2013). 44. Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purifcation. Can. J. Biochem. Physiol. 37, 911–917 (1959). 45. Boisvert, S., Raymond, F., Godzaridis, É., Laviolette, F. & Corbeil, J. Ray Meta: scalable de novo metagenome assembly and profling. Genome. Biol. 13, R122 (2012). 46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). 47. Li, H. et al. Te sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). 48. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identifcation. BMC Bioinformatics 11, 119 (2010). 49. Eddy, S. R. A probabilistic model of local sequence alignment that simplifes statistical signifcance estimation. PLoS Comput. Biol. 4, e1000069 (2008). 50. Kelly, R. T. et al. Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. Anal. Chem. 78, 7796–7801 (2006). 51. Maiolica, A., Borsotti, D. & Rappsilber, J. Self-made frits for nanoscale columns in proteomics. Proteomics 5, 3847–3850 (2005). 52. Kim, Y.-M. et al. Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. Front. Microbiol. 6, 209 (2015). 53. Hiller, K. et al. Metabolitedetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. Anal. Chem. 81, 3429–3439

(2009). 54. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. Nat. Methods 11, 1144–1146 (2014). 55. Darling, A. E. et al. PhyloSif: phylogenetic analysis of genomes and metagenomes. PeerJ. 2, e243 (2014). 56. Kerepesi, C., Bánky, D. & Grolmusz, V. AmphoraNet: the webserver implementation of the AMPHORA2 metagenomic workfow suite. Gene 533, 538–540 (2014).

## Acknowledgements

This work was funded by the Department of Energy's Genomic Science Program (grant SCW1039). Part of this work was performed at Lawrence Berkeley National Laboratory and Lawrence Livermore National Laboratory under United States Department of Energy contract numbers DE-AC02-05CH11231 and DE-AC52-07NA27344, respectively. A portion of this research was also performed under an Environmental Molecular Sciences Laboratory Science Theme Project (awarded to E.L.B.), which is a Department of Energy Office of Science User Facility sponsored by the Office of Biological and Environmental Research and operated under contract DE-AC05-76RL01830 (EMSL). DNA sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303. We thank K. Burnum-Johnson for helpful discussion.