

# Comparative Genomic Analysis of the Endosymbionts of Herbivorous Insects Reveals Eco-Environmental Adaptations: Biotechnology Applications

Weibing Shi<sup>1,2</sup>, Shangxian Xie<sup>1,2,3</sup>, Xueyan Chen<sup>2,4</sup>, Su Sun<sup>1,2</sup>, Xin Zhou<sup>2</sup>, Lantao Liu<sup>1,2</sup>, Peng Gao<sup>1,2</sup>, Nikos C. Kyrpides<sup>5</sup>, En-Gyu No<sup>2</sup>, Joshua S. Yuan<sup>1,2\*</sup>

**1** Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas, United States of America, **2** Institute for Plant Genomics and Biotechnology, Texas A&M University, College Station, Texas, United States of America, **3** School of Life Sciences and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China, **4** Department of Veterinary Pathology, Texas A&M University, College Station, Texas, United States of America, **5** DOE Joint Genomes Institute, Walnut Creek, California, United States of America

## Abstract

Metagenome analysis of the gut symbionts of three different insects was conducted as a means of comparing taxonomic and metabolic diversity of gut microbiomes to diet and life history of the insect hosts. A second goal was the discovery of novel biocatalysts for biorefinery applications. Grasshopper and cutworm gut symbionts were sequenced and compared with the previously identified metagenome of termite gut microbiota. These insect hosts represent three different insect orders and specialize on different food types. The comparative analysis revealed dramatic differences among the three insect species in the abundance and taxonomic composition of the symbiont populations present in the gut. The composition and abundance of symbionts was correlated with their previously identified capacity to degrade and utilize the different types of food consumed by their hosts. The metabolic reconstruction revealed that the gut metabolome of cutworms and grasshoppers was more enriched for genes involved in carbohydrate metabolism and transport than wood-feeding termite, whereas the termite gut metabolome was enriched for glycosyl hydrolase (GH) enzymes relevant to lignocellulosic biomass degradation. Moreover, termite gut metabolome was more enriched with nitrogen fixation genes than those of grasshopper and cutworm gut, presumably due to the termite's adaptation to the high fiber and less nutritious food types. In order to evaluate and exploit the insect symbionts for biotechnology applications, we cloned and further characterized four biomass-degrading enzymes including one endoglucanase and one xylanase from both the grasshopper and cutworm gut symbionts. The results indicated that the grasshopper symbiont enzymes were generally more efficient in biomass degradation than the homologous enzymes from cutworm symbionts. Together, these results demonstrated a correlation between the composition and putative metabolic functionality of the gut microbiome and host diet, and suggested that this relationship could be exploited for the discovery of symbionts and biocatalysts useful for biorefinery applications.

**Citation:** Shi W, Xie S, Chen X, Sun S, Zhou X, et al. (2013) Comparative Genomic Analysis of the Endosymbionts of Herbivorous Insects Reveals Eco-Environmental Adaptations: Biotechnology Applications. *PLoS Genet* 9(1): e1003131. doi:10.1371/journal.pgen.1003131

**Editor:** Paul M. Richardson, Progentech, United States of America

**Received:** April 10, 2012; **Accepted:** October 15, 2012; **Published:** January 10, 2013

**Copyright:** © 2013 Shi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The research is funded by SouthCentral Sungrant and Texas Agrilife Bioenergy Research Initiative. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: syuan@tamu.edu

## Introduction

Insects represent one of the most diverse groups of organisms on the planet that can adapt to the extremely diverse eco-environments. In particular, herbivorous insects can exploit a wide range of the plant species as food sources [1]. Insect gut symbionts play an essential role in the insect adaptation to various food types and they have been shown to be important for lignocellulosic biomass degradation, nutrient production, compound detoxification, and environmental adaptation [2–7]. Disrupting insect gut symbionts can significantly reduce the fitness of insects and can even cause serious diseases such as CCD (Colony Collapse Disease) [8]. Moreover, insect gut symbionts also were shown to be maternally inheritable from generation to generation, which suggests the symbiotic microbiota is a dynamic

component of the competitive evolution between plants and herbivorous insects as well as a driving force for insect speciation [9,10]. For these reasons, insect gut symbionts have been the subject of extensive studies in recent years [10]. Previous studies highlighted several important features of some insect gut symbionts including their reduced genome size, convergent evolution, co-speciation, and complementary function with the host genome [11–15]. Recent studies also expanded our understanding of the roles of insect gut symbionts in non-conventional functions like nitrogen recycling, reproductive manipulation, pigment production and many other aspects related to insect fitness [16,17].

Despite the progress toward understanding insect-symbiont relationships, there is still much to be learned especially with regard to facultative symbionts. Moreover, limited research has

## Author Summary

The symbiotic gut microbiome of herbivorous insects is vital for their ability to utilize and specialize on plants with very different nutrient qualities. Moreover, the gut microbiome is a significant resource for the discovery of biocatalysts and microbes with applications to various biotechnologies. We compared the gut symbionts from three different insect species to examine whether there was a relationship between the diversity and metabolic capability of the symbionts and the diet of their hosts, with the goal of using such a relationship for the discovery of biocatalysts for biofuel applications. The study revealed that the metabolic capabilities of the insect gut symbionts correlated with insect adaptation to different food types and life histories at the levels of species, metabolic pathway, and individual gene. Moreover, we showed that the grasshopper cellulase and xylanase enzymes generally exhibited higher activities than those of cutworm, demonstrating differences in capabilities even at the protein level. Together, our findings confirmed our previous research and suggested that the grasshopper might be a good target for biocatalyst discovery due to their high gut cellulytic enzyme activities.

focused on comparing the gut symbionts from insect species that specialize on different food sources. For this reason, we systemically compared the gut enzyme activities and microbial diversity in several insect species relevant to biotechnology applications [2,3,18]. Previous studies comparing gut symbionts from woodbore (*Cerambycidae sp.*, (Coleoptera)), silkworm (*Bombyx mori* (Lepidoptera: Bombycidae)), and grasshopper (*Acrida cinerea* (Orthoptera: Acrididae)) suggested that the insect gut cellulytic enzyme activities were generally correlated with the lignocellulosic biomass composition in the food consumed [2]. Furthermore, the comparison of the microbial community structure of gut symbionts from woodbore, silkworm, grasshopper, and cutworm (*Agrotis sp.* (Lepidoptera: Noctuidae)) using DGGE (Denaturing Gradient Gel Electrophoresis) revealed significant differences in symbiotic community correlating with food adaptation [3]. Despite the progress, an in-depth understanding of the eco-evolutionary adaptation to food types requires metabolic and phylogenetic analysis that cannot be offered by traditional approaches like DGGE [18]. Most of the previous comparative studies of symbionts from different insect species were either carried out with DGGE or focused on one or few symbiotic species [19,20]. Compared to those conventional techniques, new platforms like metagenomics could help define the function of symbionts in the food adaptation of insects and promote discovery of biocatalysts for biotechnology applications [18].

From the deep sea to the human intestine system, metagenome analysis has emerged as a major approach to study the composition, function, and evolution of various microbiota [21]. Metagenome analysis and metabolic reconstruction of the termite gut symbiotic microbiota revealed potential functionality in these microbiomes that might be required for biomass degradation, nutrient synthesis and other functions essential to the insect [22–24]. Moreover, those studies also highlighted the potential for biotechnology application of insect gut symbionts, since many potential glycosyl hydrolases (GH) family enzymes have been identified from the termite gut [24]. Further studies revealed the potential complementary function between the host and symbionts enzymes for highly efficient biomass degradation [23]. Despite the progress, previous research mainly focused on the

metagenome sequencing of symbionts in single insect species or the same symbiont in different insect species [17,25–27]. Few studies have systematically compared the metagenomes of symbiotic microbiota from insect species with distinctly different diets, environmental adaptations, or life histories. This type of comparative metagenomics approach has the potential to substantially improve our understanding of the adaptive significance of insect gut symbionts for insect diet specialization as well as facilitates the discovery of novel biocatalysts for biorefinery applications.

In this study, we selected three insect species that are from different insect orders and have different diets and life histories characteristics: grasshopper (*Acrida cinerea* (Orthoptera)), cutworm (*Agrotis ipsilon*) (Lepidoptera) and termite, *Nasutitermes sp.* (Isoptera: Termitidae). The grasshopper is a polyphagous insect specializing on different plant leaves, mainly from the monocot grass species. Previous studies revealed that the grasshopper diet contains about 37.2% of forbs, 58% of grasses and sedges and 4.8% of others [28]. The cutworm is also a polyphagous, generalist that can adapt to a broad range of food sources including cabbage, asparagus, bean, and other crucifers [29]. In contrast, the termite is monophagous insect that specializes on lignocellulosic biomass as a food source. The three insects also differ in life cycle. The cutworm is a holometabolous insect that undergoes complete metamorphosis with a pupal stage [30], whereas the grasshopper and termite are hemimetabolous, having incomplete metamorphosis and juveniles with morphologies similar to adults [31].

Metagenome data from the gut symbiotic microbiota of grasshopper and cutworm were generated using Illumina Genome Analyzer, and these metagenome data were compared with the updated sequencing data from gut symbionts of the wood-feeding higher termite [24]. As one of the first comprehensive comparisons of insect gut symbiotic metagenome, the goal was to examine the relationships between the taxonomic and potential metabolic diversity of the insect gut microbiomes and the diets and life histories of their insect hosts at the community, metabolic pathway, and molecular levels. The analysis indicated that the composition of gut symbionts was correlated with their function in biomass degradation and nutrient biosynthesis. The metabolic reconstruction revealed the presence of specific pathways relevant to the utilization and transport of diverse carbohydrate sources in cutworm and grasshopper. The diversity, phylogenetic, metabolic, and functional analyses all supported the hypothesis that insects and their gut symbionts co-evolved with the food preferences of the insect toward optimal capacities in biomass degradation, macromolecule intake and utilization, complementary nutrient synthesis, and other aspects related to insect life style. In addition, we cloned 24 biomass degrading enzymes based on the predicted gene models and characterized four of them. Enzyme assays revealed that grasshopper cellulytic enzymes were generally more active than the cutworm cellulytic enzymes, which confirmed the presence of functional diversity at the protein. The enzyme characterization indicated that insect guts were useful resources for discovering novel biocatalysts for biorefinery applications.

## Results/Discussion

The metagenome sequencing results were summarized in Table 1. The sequence assembly rendered more than 20,000 of predicted gene models for the gut symbionts from grasshopper and cutworm, respectively. In order to analyze the composition-function relationship, we compared the grasshopper and cutworm gut microbiota with the updated termite gut microbiota sequences (JGI IMG Database GOLD ID: GM00013 and Sample ID:

**Table 1.** Summary of sequence data obtained from gut microbiomes of grasshopper and cutworm, respectively.

Parameters	Grasshopper	Cutworm
Total length of bases	14,036,933	11,308,910
Total length of coding bases	8,208,120	7,663,722
G+C content%	42.08	38.14
Total Scaffolds	39,301	35,554
Total CDSs	22,335	25,208
Average CDS length, bp	371	302
Archaea CDSs (% of total CDSs)	16 (0.17)	36 (0.31)
Bacteria CDSs (% of total CDSs)	2,420 (26.10)	7,720 (67.15)
Eukarya CDSs (% of total CDSs)	1,977 (21.32)	361 (3.14)
Plasmid CDSs (% of total CDSs)	22 (0.24)	54 (0.47)
Virus CDSs (% of total CDSs)	166 (1.79)	214 (1.86)
Unassigned CDSs (% of total CDSs)	4,672 (50.38)	3,112 (27.07)
CDS density,%	98.94	99.01
CDS with designed function	12209	14211
CDS connected to KEGG pathways	1105	900
CDS connected to KEGG Orthology (KO)	2077	1468
CDS with COGs	8954	11317
COG clusters	2301	1728
CDS with Pfam	10604	11420
Pseudogenes	0	0
rRNA	188	102
tRNA	77	104

doi:10.1371/journal.pgen.1003131.t001

GS000048), with respect to the phylogenetic diversity, microbial abundance, putative gene function, and metabolic capacity. As described above, the three host species are from distinct insect orders and have different diet specializations and life histories.

### The Microbial Species Distribution as Revealed by Gene-Coding Sequences Reflected the Function of Insect Gut Symbionts

Relative abundance of symbiotic microbial species in each insect gut was estimated based on the species distribution of the gene-coding sequences as annotated by the BLAST search. The cluster analysis of bacterial species distribution for the gut symbionts was shown in Figure 1. It should be pointed out that Figure 1 only represented a rough estimation of the microbial species distribution because of the genome size variations in different symbionts, which complicated the data interpretation. Nevertheless, the comparison of the relative abundance of the bacteria phyla in the microbiota from the three different insect species revealed that the microbiota composition was rather different from each other and these differences might be relevant to the functions they provided for their insect hosts. The dominant groups differed among the three insect species. For the cutworm, the phylum *Bacilli* was the dominant group (24.14%), followed by *Clostridia* (4%), *Erysipelotrichi* (3.64%) and  $\gamma$ -*proteobacteria* (1.43%) (Figure 1). For the grasshopper, the most common bacterial genes were from  $\gamma$ -*proteobacteria* (25.16%), followed by *Erysipelotrichi* (3.51%), *Clostridia* (1.27%), and *Bacilli*, (0.84%), respectively (Figure 1). For both species, the most abundant groups comprised

about 25% of the diversity, whereas the second most abundant groups comprised less than 5%.

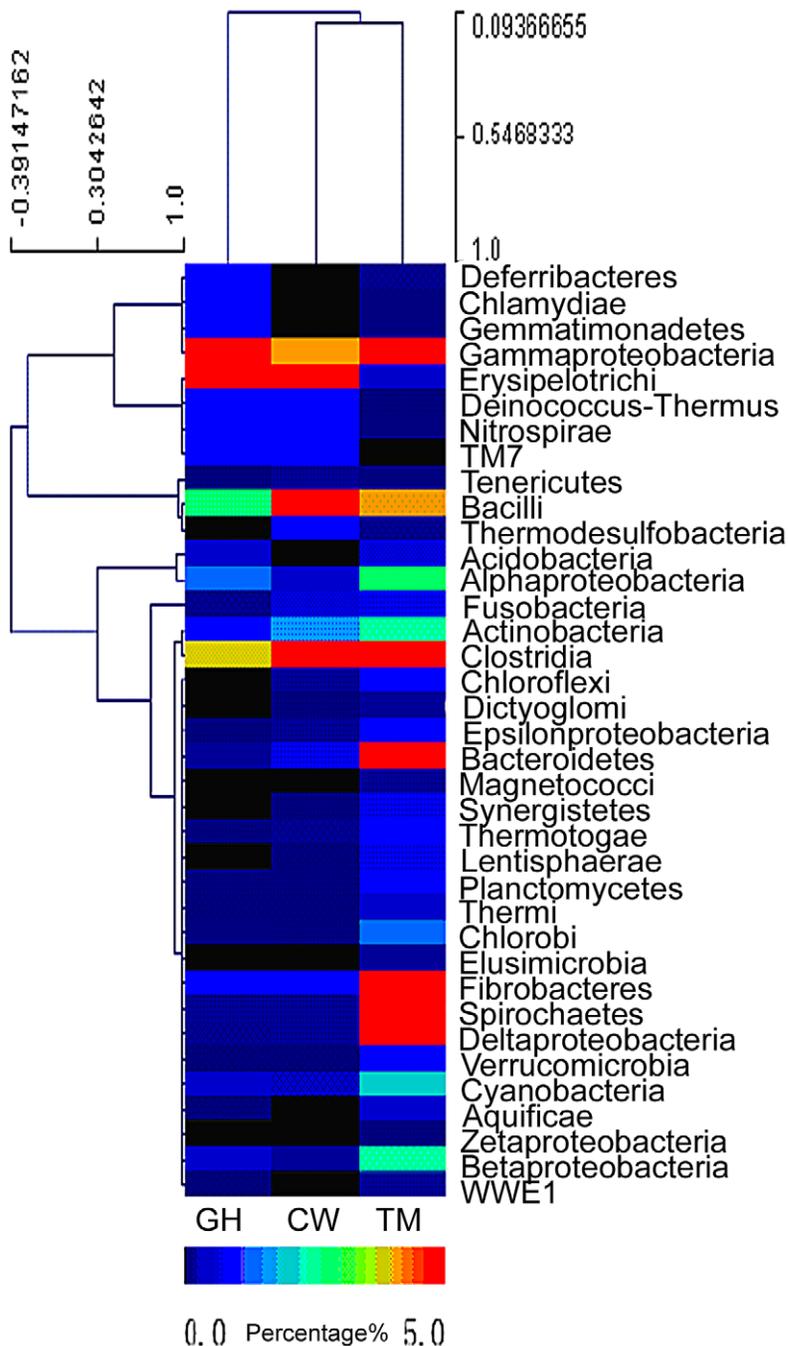
Even though the insects differed in microbial composition, there were some similarities that likely were related to function. Both *Clostridia* and *Bacilli* species have been shown to be the major groups of microbes responsible for biogas production and biomass conversion in microbial communities [32]. Many *Clostridia* species such as *C. thermocellum* and *C. ljungdahlii* are anaerobic *Firmicutes* known to have a robust capacity to use cellulose, hemicellulose, and other carbohydrate [33–35]. The presence of a large proportion of *Clostridia* was likely to be important for lignocellulosic biomass degradation [34,36]. However, the predominance of the  $\gamma$ -*proteobacteria* in grasshopper was unexpected, because  $\gamma$ -*proteobacteria* has not been shown previously to be involved in biomass utilization. However, recent work revealed that  $\gamma$ -*proteobacteria* might be important nutrient providers for host insects. For example,  $\gamma$ -*proteobacteria* as facultative or obligate endosymbionts were shown to play essential roles for insects like tsetse fly in the utilization of low nutrient food sources [37]. Similarly, the predominance of  $\gamma$ -*proteobacteria* in grasshoppers might be important for the utilization of the grasses, which characteristically have high fiber content.

Compared to the grasshopper and cutworm microbiomes, the microbial composition of the termite microbiome reflected its unique adaptation to utilization of woody species, where both the *Clostridia* and the *Spirochaetes* species were predominant (Figure 1) [24]. Additionally, the termite microbiome was composed of several major groups with more than 5% abundance. Morphologically diverse *spirochaetes* were consistently present in the hindgut of all termites [38], and was found as ectosymbionts attached to the surface of cellulose-digesting protists [39]. Overall, the microbial populations of the cutworm, grasshopper and wood-feeding termite gut systems appeared to consist of taxa with known capacities for degrading and utilizing the different types of foods on which their insect hosts specialize.

### Diversity of Insect Gut Microbiota as Evaluated by the 16S rRNA

In addition to gene-coding sequence-based analyses, we also implemented two types of phylogenetic analyses. First, two partial 16S rRNA clone libraries were established from the PCR amplified 16S rRNA sequences using 515F/1492R primers. Sanger sequencing was used to sequence individual 16S rRNA clones as summarized in Table S1. The phylogenetic analysis was presented in Figure 2. The second phylogenetic analysis was based on the annotation of the contigs derived from the metagenome sequence assembly. The assembled contigs were first aligned to the 16S rRNA genes from the recent release of RDP database using blastn. The analysis resulted in 188 and 102 contigs assigned to be 16S rRNA for cutworm and grasshopper, respectively (Table S1). The most similar partial or complete 16S rRNA sequences from the database were used for the multiple sequence alignment and phylogenetic analysis using Maximum likelihood method (RAxML). The analysis results were presented in Figure S1. The results from the two types of analysis generally were consistent; although the phylogenetic analysis based on the annotated contigs (Figure S1) provided a deeper coverage of microbial species and a better representation of uncultured species.

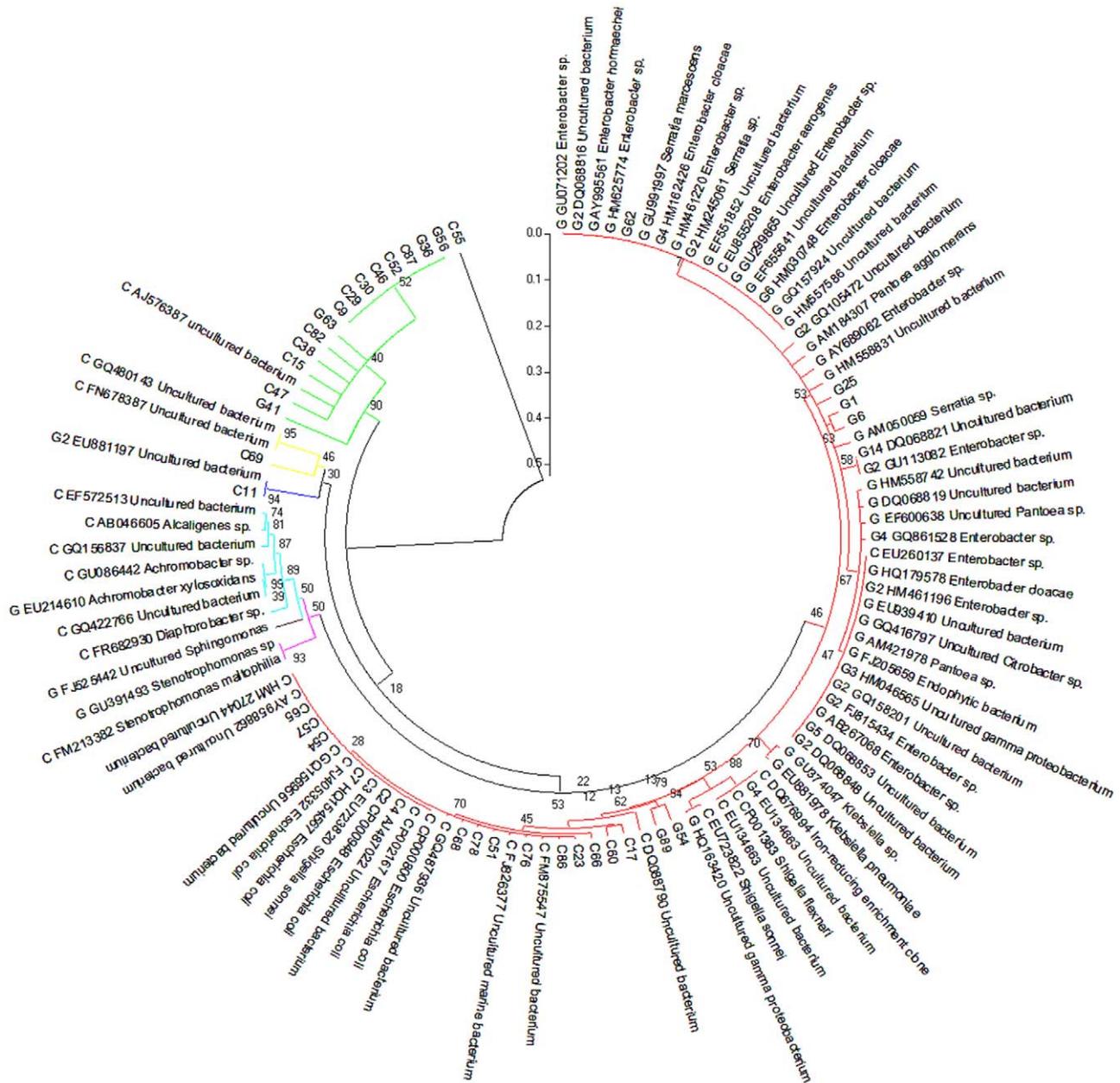
The phylogenetic analyses (Table S1, Figure 2, Figure S1) revealed three features. First, *proteobacteria* represented the most diverse group of the microbes in the microbiomes of both grasshopper and cutworm. Among the *proteobacteria*,  $\gamma$ -*proteobacteria* was the predominant taxa and the 16S rRNA sequences from cutworm and grasshopper formed two distinct clades,



**Figure 1. Abundance of bacterial phyla based on the predicted gene models in the gut microbiota of grasshopper (GH), cutworm (CW), and termite (TM), respectively.** The relative abundance ranged from 0–26%. Except for the three most abundant bacteria phyla, all other phyla are less than 5%. To better visualize, the heat map scale set from 0–5%. doi:10.1371/journal.pgen.1003131.g001

indicating the relatively independent evolution of the gut microbiome in the two species. The 16S rRNA-based phylogenetic analysis correlated well with the microbial abundance analysis using gene models (Figure 1). The studies confirmed the differences in abundance, phylogeny, and evolution of gut symbionts between cutworm and grasshopper. A second feature of the analyses was that the cutworm had more species of gut symbionts than grasshopper (188 vs. 102, Figure S1). We speculated that the greater diversity of symbionts in the cutworm gut as compared to that of the grasshopper might be relevant to its

being both more of a dietary generalist. A third feature was the discovery of large number of uncultured species or unknown species. Uncultured species referred to the species that cannot be cultured in standard medium, whereas unknown species referred to those lacking taxonomic information. Due to the deeper coverage of metagenomic sequencing compared to the PCR cloning library, Figure S1 showed almost 60% sequences were from uncultured or unknown species. The results highlighted our limited knowledge of the diversity of insect gut symbionts. It was proposed that the existence of many unculturable species might be related to the



**Figure 2. Composition of grasshopper (G) and cutworm (C) gut microbiomes as revealed by 16S analysis.** From a PCR-based library, 54 and 56 nearly complete sequences of the 16S rRNA V3–V9 region belonging to different bacterial species were obtained from the gut microbiomes of grasshopper and cutworm, respectively. These were used in a Maximum Likelihood analysis (RA×ML). Species identification was determined based on sequence similarity greater than 97% using the 16S rRNA sequences available in NCBI GenBank. Genbank accession numbers are given. The strains belonging to different group were indicated using different color, i.e. red ( $\gamma$ -proteobacteria/Enterobacteriales), magentas ( $\gamma$ -proteobacteria/Xanthomonadales), brown ( $\alpha$ -proteobacteria), cyan ( $\beta$ -proteobacteria), blue (Cyanobacteria), yellow (Bacteroidetes), and green (Firmicutes). doi:10.1371/journal.pgen.1003131.g002

significant reduced genome and limited metabolic capacity of some symbiotic microbes [40–43]. The phenomena indicated that the metabolic capacity of insect gut microbiota should be considered as a whole instead of based on individual species.

Another observation was that 14 and 10 16S rRNA sequences were assigned to *Acetobacter pasteurianus* (AP011163) for cutworm and grasshopper, respectively (Figure S1). *Acetobacter* strains belong to acetic acid bacteria (AAB), which are often found in various categories of fruits, flowers, and fermented foods [44] and some

insect guts [45]. *Acetobacter* might have originally been acquired from the food sources of cutworm and grasshopper and subsequently become a more permanent symbiont for the two species or might occur as a transient resident. *Acetobacter* can produce alcohol dehydrogenase (ADH), which could potentially contribute to lignin oxidation for lignin degradation/modification in termite guts [46,47]. Overall, the phylogenetic analysis indicated correlations between microbial composition and function and insect diet preference.

### Comparative Functional Analysis of Microbiome from Three Different Insect Orders

Metagenome sequencing provided more detailed functional comparisons of different gut symbionts using pathway analysis based on COGs (Clusters of Orthologous Groups) and KEGG (Kyoto Encyclopedia of Genes and Genomes) [48,49]. KEGG maps the genes within the biological pathways to derive potential functions [50], whereas COG analysis uses evolutionary relationships to group functionally relevant genes [51]. The annotation of the cutworm and grasshopper gut microbiomes yielded 11,317 and 8954 hits for the COG database as well as 900 and 1105 hits for the KEGG pathways, respectively.

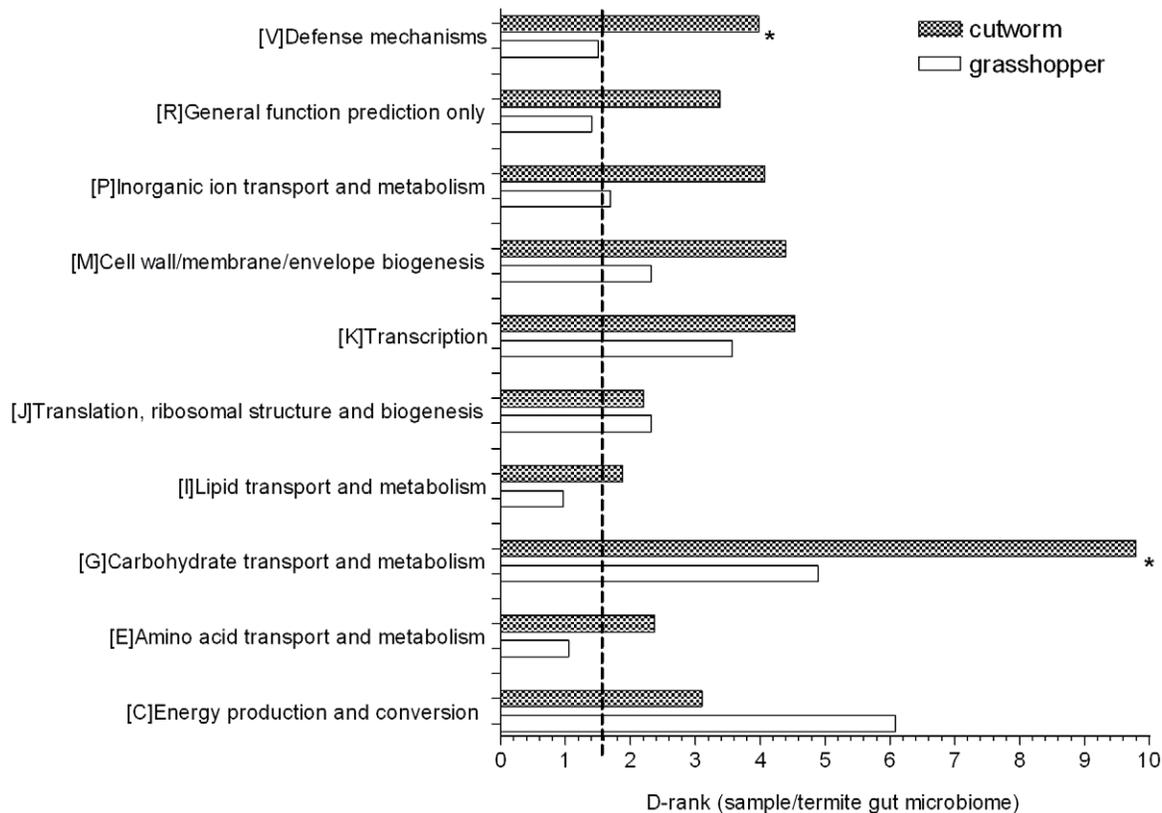
D-ranks analysis was used to evaluate the relative enrichment of COG and KEGG gene categories in the cutworm and grasshopper gut symbiotic metagenomes compared to the termite metagenome. The enrichment or under-representation of COG categories were as shown in Figure 3. Both cutworm and grasshopper gut symbionts were enriched in several metabolic pathways compared to termite gut symbionts. Cutworm gut symbionts were enriched with genes for carbohydrate transport and metabolism, and defense mechanisms ( $P < 0.05$ ) relative to grasshopper symbionts. The diversity in carbohydrate metabolism genes correlated well with the taxonomic diversity of the gut microbiomes (Figure S1) and were consistent with the hypothesis that the greater diversity in species composition and carbohydrate metabolism observed in the cutworm may be related to the broader diet preference and more complicated life histories of the cutworm compared to those of the grasshopper.

The ontology analysis based on KEGG revealed similar patterns as shown in Table S2, where flagella assembly in cell motility and type III secretion system ( $P < 0.05$ ) are more enriched in termite gut symbionts than those of cutworm and grasshopper, although it is unclear why this would be so. Overall, the metagenomic composition of genes in all categories reflected their potential function in adaptation to insect diet and life history. A more detailed functional relevance can be derived from examination of specific pathways.

### Metabolic Reconstruction of Symbionts from Three Insect Species at Pathway Level

Metabolic reconstruction provided comparison of potential biocatalyst functionality in four general COG categories and thus a means of relating the metabolic diversity and capability of the microbiome to the insect diet and life style.

**Plant polysaccharide degradation (Carbohydrate transport and metabolism).** Insect guts are believed to be dual systems where enzymes from both the host and symbiotic microorganisms work synergistically to degrade and utilize the cell wall components [23,24,52]. Highly efficient natural biocatalyst systems like insect guts are important resources to discover novel enzymes for biorefinery applications [24,53]. We carried out the domain identification for all gene models using global alignment of the Glycosyl Hydrolase (GH) catalytic domains, Carbohydrate Binding Modules (CBM), and glycosyl transferase (GT) domains as shown in Table S3. A total of 31, 40, and 52 different GH CAZY families (carbohydrate-active enzymes; <http://www.cazy.org>) were



**Figure 3. COG analysis reveals metabolic functions that are enriched or under-represented in grasshopper and cutworm gut.** Gene categories with D-Rank values greater than indicated by the dashed line are significantly enriched in the cutworm and grasshopper gut symbiotic metagenome as compared to that of termite ( $P < 0.05$ ); Asterisks indicate categories that are significantly different between grasshopper and cutworm gut microbiomes ( $P < 0.05$ ).

doi:10.1371/journal.pgen.1003131.g003

detected from the guts of the grasshopper, cutworm, and termite, respectively.

There was a clear correlation between the primary food source and the categories of enzymes predicted from the metagenomic analysis. The termite gut featured the most abundant putative cellulases and hemicellulases among the three insect species, correlating with the fact that termite is an extremely successful wood-degrading organism. There were 125 GH5 cellulases and 101 GH10 xylanase along with a number of GH8, 9, and 45 endoglucanases from termite gut symbionts. However, only GH5 and GH8 family cellulases existed in the grasshopper gut. The cutworm gut only had GH5 family cellulase (Table S3).

A striking feature of the cutworm and grasshopper biomes was the significant enrichment in GH1 family enzymes, where 181 and 34 gene models were assigned to GH1 from cutworm and grasshopper gut microbiomes, respectively. The GH1 family enzymes include a diverse group of enzymes such as  $\beta$ -glucosidases,  $\beta$ -galactosidases, 6-phospho- $\beta$ -galactosidases, myrosinases, and others [54]. Most of the GH1 family members attack  $\beta$ -glycosidic bonds between a pyranosyl glycon and an aglycon. Among these GH1 enzymes,  $\beta$ -glucosidases cleave non-reducing carbohydrates in oligosaccharides and hydrolyze cellobiose to glucose [54]. Other enzymes catalyze a broad spectrum of activities for carbohydrate usage.

Other than GH 1, many  $\beta$ -glucosidases in GH 3, 4, and 31 also were identified in the microbiomes of the three insect species. Other enzymes discovered from cutworm and grasshopper guts include GH 13 ( $\alpha$ -amylase), GH 18 (Chitinase), GH 23, GH 28 (endopolygalacturonase), GH 38 ( $\alpha$ -mannosidase), and GH 43 ( $\beta$ -xylosidase). There were seven different types of CBM domains identified from the termite gut microbiome and three types of CBM domains in the grasshopper gut microbiome (Table S3). CBM is a protein domain usually found in carbohydrate-degrading enzymes for binding specific plant structural polysaccharides [55,56]. In the metabolic reconstruction, we identified a number of plant polysaccharide degradation enzymes and relevant domains in grasshopper, cutworm, and termite gut microbiome (Figure 4A and Table S3). Overall, the distribution of the GH family enzymes and CBM domains predicted from the metagenomic analysis were consistent with differences among insect hosts in food specialization, indicating that the plant polysaccharide degradation capacity of the symbionts reflected diet specialization of the insect.

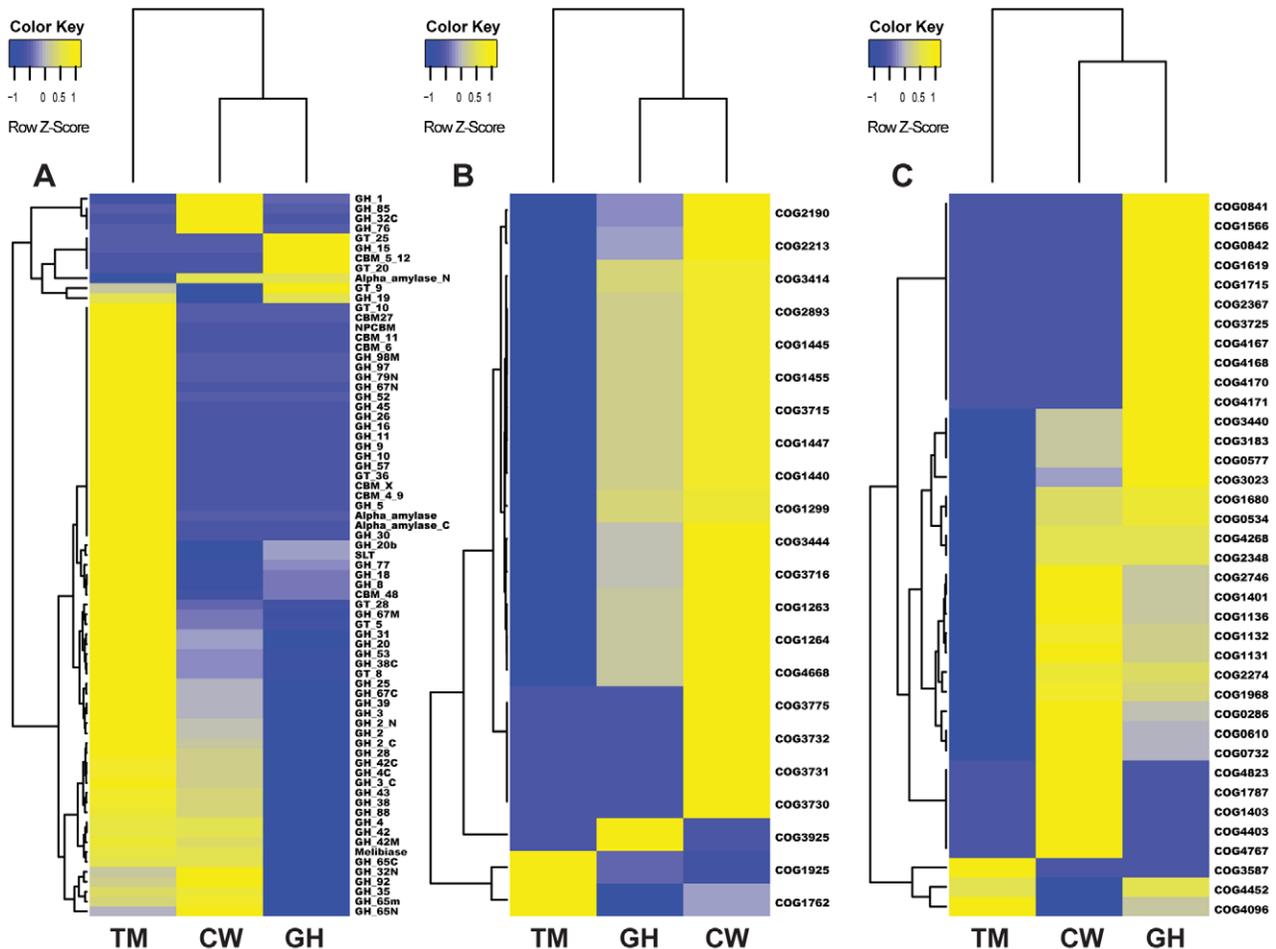
As expected, the termite microbiome was enriched in lignocellulosic biomass degrading enzymes including cellulase and hemicellulase. However, the cutworm microbiome was enriched with various GH family enzymes, in particular, GH1 enzymes involved in utilization of a variety of carbon sources. The grasshopper microbiome was intermediate having fewer lignocellulosic enzymes than the termite microbiome, but more CBM domains, cellulases and xylanases than the cutworm microbiome (Figure 4A and Table S3). The pattern might be important for the degradation of high fiber grass leaves. We cloned and characterized several cellulytic enzymes to both verify the function of the symbionts and exploit them for biofuel applications (see 5 below). Overall, the distribution of GH family enzymes in the microbiomes of the three insects generally reflected their adaptation to different food types.

**Phosphotransferase system (PTS) for sugar membrane transport.** Another group of proteins relevant to carbohydrate utilization was Phosphotransferase (PTS) for sugar transport across membrane. Comparative analysis of KEGG pathways revealed that both cutworm and grasshopper gut microbiomes were more enriched in PTS genes than the wood-feeding termite microbiome

( $P < 0.01$ ) (Table S2). Cluster analysis clearly indicated that the grasshopper gut microbiome had a profile similar to cutworm, but distinct from termite (Figure 4B). The PTS complex consisted of three catalytic components including Enzyme I, Enzyme II (membrane-bound sugar-specific permeases), and HPr (heat-stable, histidine-phosphorylatable protein) [57]. Enzyme II is the component important for carbohydrate transport across the bacterial membrane and was identified in all three species [58] (Figure 4A and Table S4). The COG analysis also revealed the prevalence of different Enzyme II components in the cutworm and grasshopper gut microbiomes. However, the termite gut symbionts seemed to lack most types of the Enzyme II systems as shown in Figure 4A and Table S4. The results highlighted the differences in carbohydrate transport and processing among the microbiomes of the three insect species. The polyphagous cutworm and grasshopper gut microbiomes were much more enriched and had a higher diversity of PTS components than the microbiome of the monophagous termite. The diversity of food types and carbohydrate substrates in the diets of polyphagous insects might contribute to the maintenance of PTS diversity in the microbiomes of these insects. Overall, the diversity in microbes, their carbohydrate transport, and carbohydrate utilization genes were correlated with the diversity of food types in the insect diet, consistent with the hypothesis that more complicated diets require more complicated carbohydrate transport and utilization systems at the species, metabolic capacity, and molecular pathway levels.

**Energy production, conversion, and nitrogen metabolism.** COG analysis also revealed that energy production, conversion and other relevant metabolic functions were enriched in the grasshopper and cutworm gut symbionts as compared to termite symbionts (Figure 3). The cluster analysis of COG category enrichment or under-representation was as shown in Table S5. Notably, cutworm microbiome was enriched with COG malmate/lactate dehydrogenases (COG0039), Isocitrate dehydrogenases (COG0538) and other TCA (Tricarboxylic acid) pathway components as compared to that of termite (Table S5). However, both the termite and grasshopper gut microbiomes were more enriched in nitrogen metabolism enzymes than the cutworm microbiome (Table S6). For instance, 22 nitrogenase homologues were identified in the termite gut microbiome and some nitrate reductases were identified only in the grasshopper gut microbiome (Table S6). Since termite and grasshopper rely on food (wood and grasses) with less protein content as compared to cutworm, and functional enzymes like nitrogenase for nitrogen fixation and nutrient synthesis might be important for supplementing low nitrogen in the diet.

**Detoxification and defense-relevant mechanisms.** As compared to the wood-feeding termite, another COG category enriched in both grasshopper and cutworm gut microbiome was the detoxification and defense-related proteins (Figure 3, Table S7). As shown in Figure 4C, grasshopper and cutworm gut microbiomes were enriched in several ABC transporter-related COGs, such as ABC-type multidrug transport system (COG1131), ATPase and permease components (COG1132), ABC-type antimicrobial peptide transport system (COG1136), and ABC-type bacteriocin/lantibiotic exporters (COG2274) (Table S7 and Figure 4C). The ATP binding cassette (ABC) transporters are important components for the uptake and efflux systems in different organism including bacteria, lower eukaryotes [59,60]. ABC transporters are known for their detoxification functions. For example, the ABC transporter-based detoxification pumps in bacteria include several major classes: the ABC super family [59], the major facilitator super family (MFS) [61], the small multidrug resistance (SMR) family [62], and the resistance-nodulation-cell



**Figure 4. Cluster analysis of genes in three metabolic pathways in the gut microbiomes of grasshopper (GH), cutworm (CW), and termite (TM).** A. biomass degradation enzymes in carbohydrate transport and metabolism; B. Phosphotransferase system; and C. Defense mechanism.

doi:10.1371/journal.pgen.1003131.g004

division (RND) family [63]. Genome sequencing has revealed that these ABC transporters are present in a broad range of microorganisms such as *Escherichia coli*, *Haemophilus influenzae*, *Mycoplasma genitalium*, *Bacillus subtilis*, *Methanococcus janneschii*, and *Synechocystis* PCC8603. [64–68]. The enrichment of detoxification and defense genes in cutworm and grasshopper may be related to their diverse food intake or more variable host environment.

### Verification of Sequence Assembly and Characterization of Enzymes for Biorefinery Applications

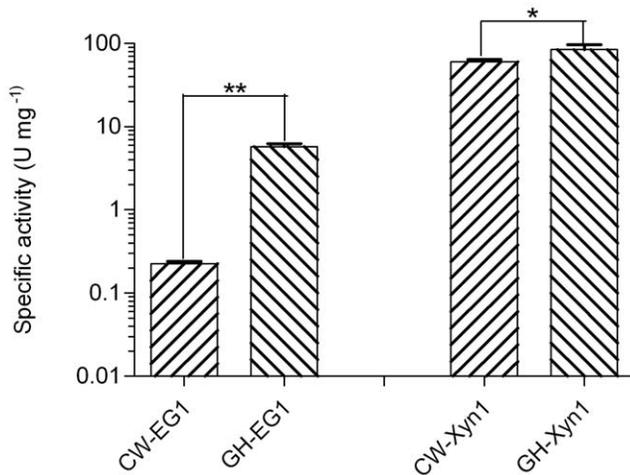
The ultimate goal of this research was to discover novel biocatalysts for biorefinery applications. We therefore cloned and characterized several enzymes for functional validation. A total of 24 ORFs of predicated plant polysaccharides degradation enzymes were PCR amplified using primers based on the assembled sequences (Figure S2). A total of 22 out of 24 ORFs amplified and the sequences of all of the amplicons were consistent with the assembled sequences (Figure S2). The results highlighted the reliability of the Illumina metagenomic sequencing and assembly to identify degradation enzymes. Our research represents one of the few metagenome sequencing efforts to rely mainly on the Illumina Genome Analyzer [69].

We further characterized an endoglucanase (CW-EG1 and GH-EG1) and a xylanase (CW-Xyn1 and GH-Xyn1) from both the

grasshopper and cutworm guts, respectively. The selected enzymes were expressed and purified by a His-trap nickel column, as indicated by SDS-PAGE (Figure S3). The enzyme performance under different temperature and pH conditions was as shown in Figure S4. All four of the enzymes exhibited activity, and the activities were significantly influenced by temperature and pH. Most enzymes had temperature optima at 60–70°C and pH optima at 7.0–9.0 (Figure S4). This pH range correlates with the fact that many insect gut systems have a slightly basic environment [70]. Considering that many traditional filamentous fungi enzymes had optimal activity in the weakly acidic pH range, the insect gut enzymes provided complementary capacity for biomass degradation.

We further compared the specific activity of the same category of enzymes from cutworm and grasshopper gut microbiome. Interestingly, for both cellulase and xylanase, the grasshopper gut enzymes were significantly higher than those of cutworm ( $P < 0.05$ , Figure 5). The result correlated with our previous analyses of gut content activities, even though the differences could also result from the choice of enzymes and other factors [2]. The adaptation to relatively higher temperature made the enzymes good candidates for some biomass conversion applications.

Together with many recent studies, our research indicated that insect gut symbionts are substantial resources for enzyme discovery



**Figure 5. Comparison of the specific activities of enzymes important for biomass deconstruction from grasshopper and cutworm gut microbiomes.** \*\*means  $P < 0.01$  and \*means  $P < 0.05$  in student t-test.

doi:10.1371/journal.pgen.1003131.g005

for biorefinery applications. The relationship between the diversity and potential functional capabilities of the gut microbiomes and insect food preference is particularly relevant improvements in biomass degradation, and thus should be explored for biotechnology applications [71–75]. Due to the technical limitations, we particularly focused on the bacterial symbionts in this study. Nevertheless, the fungal and protozoal symbionts in insect guts were also widely studied for their biomass degradation capacity. These eukaryote symbionts should be investigated for their roles in biomass deconstruction, food and life history adaptation in the follow-up studies.

## Materials and Methods

### Metagenomic DNA Extraction

Metagenome analysis requires comprehensive coverage of most multiple species in the sample [76]. To obtain sufficient high-quality DNA for sequencing with Illumina Genome Analyzer, approximately 2000 third to fifth instar grasshoppers and 50 fourth to fifth instar cutworms were dissected to extract genomic DNA from gut symbionts. A recently developed indirect DNA extraction method was modified for the insect gut metagenomic DNA extraction [77]. The extracted metagenomic DNA were quantified by a Nano Drop ND-1000 spectrophotometer and characterized by electrophoresis. Moreover, the quality of the DNA was verified by PCR amplification of conserved 16S rRNA for bacteria and conserved 18S rRNA for insect host contamination [29]. The results confirmed that the metagenomic DNA is free from host DNA contaminations, because the 18S rRNA did not amplified.

### Library Construction and Metagenome Sequencing

Metagenome sequencing of cutworm and grasshopper gut symbiotic microbiota was carried out using Illumina Genome Analyzer II (Illumina, Inc. CA, USA) with paired-end 76 base sequencing. Library construction was carried out following the manufacturer's recommendation using Illumina Paired-End Sequencing Kit (Cat. No. PE-102-1001). Briefly, 2 to 5 µg metagenomic DNA was sheared by nebulization to generate DNA fragments and the ends were repaired with Klenow,

followed by several steps to add the adapters. Adapter-ligated DNA fragments of length 300–350 bp were isolated from a 2% agarose gel using QIAquick Gel Extraction Kit. The fragments were then amplified by 11 cycles of PCR reaction to generate the DNA library at a concentration of 20–35 ng/µl. The median size of the library was evaluated using 2% agarose gel. The PHIX Control V2 Library was prepared by Illumina (Cat. No CT-901-2001) and used for sequencing. Approximately 5 pmol DNA libraries were subjected to cluster generation and sequenced by DNA core of Institute of Plant Genomics and Biotechnology. The images were processed using version 0.3 of the GAPipeline software supplied by Illumina.

### Sequence Assembly

After base-calling with GAPipeline software, the remaining 44,155,246 (cutworm) and 58,033,340 (grasshopper) reads (each is about 76 bases) were trimmed and assembled using Velvet version 0.7.55 (<http://www.ebi.ac.uk/~zerbino/velvet/>, European Bioinformatics Institute, EMBL-EBI). The resulted assembly consisted of 64,065 and 78,991 contigs for cutworm and grasshopper, respectively.

### Loading of Data into IMG/M and Function Annotation

The draft assembled contigs ( $\geq 100$  bp) were loaded into IMG/M (<http://www.jgi.doe.gov/m>) [78]. Before further analysis, the IMG/M system first carried out a gene model validation process, including editing overlapping CDSs, correcting start codons, and identifying missed genes and pseudogenes [78]. The predicted coding sequences (CDSs) and some functional RNAs were recorded with start/end coordinates in the contigs. The predicted genes were assigned to COGs (clusters of orthologous groups) based on RPS-BLAST (reverse position specific BLAST) and NCBI's Conserved Domain Database (CDD), using an e-value threshold of  $10^{-2}$  without low-complexity masking [79]. Genes were also probed against Pfam database using HMMER search (<http://hmmer.janelia.org/>) [80,81]. Protein-coding sequences were further annotated for molecular function and pathways using KEGG pathways. In addition, the metagenome sequences and gene models were binned to rank domain, phylum, and class using PhyloPythia [82].

### 16S rRNA Analysis

The phylogenetic analysis of 16S rRNA was carried out with two types of analyses. First, two clone libraries were prepared using PCR products amplified from cutworm and grasshopper gut metagenome DNA with one pair of primers broadly targeting the V3–V9 region of 16S rRNA. The primer sequences were 515F (5'-GTGCCAGCAGCCGCGTAATACCTTGTACGACTT-3') and 1492R (5'-GGTTACCTTGTACGACTT-3') [83]. 87 and 97 near complete 16S rRNA V3–V9 region sequences were obtained for cutworm and grasshopper gut microbiome, respectively. The 16S rRNAs was then used for phylogenetic analysis.

In addition to sequencing of the V3–V9 region, we also sought to reach a deep coverage of symbiotic species by analyzing the assembled metagenome sequences. 16S rRNA sequences were identified using BLASTN ( $E < 1 \times 10^{-5}$  and a sequence length hit  $> 50$  nt) search against the SSU rRNA genes from release 16.3.3 of the RDP database (<http://rdp.cmc.msu.edu/>) [84], and the European Ribosomal RNA database (<http://www.psb.ugent.be/rRNA/index.html>). Due to the high similarity, it is usually difficult to isolate the 16S rRNA genes from *de novo* assembly of metagenome data. A total of 96 and 53 partial and near complete 16S sequences were extracted from 188 and 102 assembled contigs for cutworm and grasshopper gut microbiomes, respectively. The

sequences were then aligned with the NAST aligner [85], and imported into an ARB database (<http://greengenes.lbl.gov>) [86]. The nearest aligned full length sequences were used for classification and phylogenetic tree construction using RAxML [87].

Phylogenetic analysis was carried out using the Minimum Evolution method with the sum of branch length = 5.0 [88]. The evolutionary distances were computed using the Maximum Composite Likelihood method with 1000 replicates of bootstrap tests [89].

### Comparative Metabolic Pathway Analysis

In order to compare the metabolic pathways for different microbiota, the coding sequences were analyzed with KEGG and COG (Clusters of orthologous groups). Both grasshopper and cutworm symbiotic metagenome and updated termite metagenome data (JGI IMG Database GOLD ID: GM00013 and Sample ID: GS0000048) [24] were compared. For KEGG analysis, all coding sequences were converted into KEGG orthologous (KO) groups, and the KEGG pathway annotation was extracted based on the latest release of KEGG version (Release 55.1, September 1, 2010). The COG assignment was based on RPS-BLAST and NCBI's Conserved Domain Database (CDD). Only 4.95%, 3.48%, and 6.41% of predicted genes were assigned to KEGG pathway for grasshopper, cutworm, and termite gut microbiome, respectively. 39.4%, 44.41%, and 53.56% of coding sequences were assigned to COG terms for grasshopper, cutworm and termite gut microbiome, respectively.

In order to further define the enrichment or under-representation of a KEGG pathway or a COG term in a certain microbiome, two metrics were used in this study. For the comparison of a protein family between a query metagenome and a reference metagenome, the *D*-scores were calculated using a binomial distribution. We calculated the *D*-score using  $(f1-f2)/\sqrt{pq * (1/n1+1/n2)}$ , where  $f1 = x1/n1$  = frequency of functional occurrence in query group,  $f2 = x2/n2$  = frequency of functional occurrence in reference group,  $p = (x1+x2)/(n1+n2)$  = probability of occurrence,  $q = 1-p$  = probability of non-occurrence. Specifically, *x1* was the number of a given function in query group, *x2* was the number of a given function in reference group, *n1* was total counts of all function occurrences in query group, and *n2* was total counts of all function occurrences in reference group. Further analysis involved *D*-rank, a normalization ranking for each pair wise comparison. *D*-rank was calculated by adding the *D*-scores of all protein families assigned to a certain functional category and then normalized by the square root of the number of total categories [90,91].

### Sequence Assembly Verification, Subcloning, Expression, and Activity Assay for Biocatalysts

In order to verify the quality of sequence assembly and discover novel biocatalysts, 24 predicted coding genes for carbohydrate degrading enzymes were amplified, among which 22 showed positive results. Among the 22, four were expressed and analyzed. The same batch of sequenced metagenomic DNAs were used as template for PCR amplification. The PCR mixture (50  $\mu$ l) contained 5  $\mu$ l of 10 $\times$  PCR buffer, 4  $\mu$ l of MgCl<sub>2</sub> (25 mM), 1  $\mu$ l of dNTP, 1  $\mu$ l of each primer (10 mM), 37  $\mu$ l of sterile Milli-Q water, 0.5  $\mu$ l of Taq polymerase (AmpliTaq Gold DNA Polymerase, Applied Biosystems, CA, USA), and 0.5  $\mu$ l of DNA templates. PCR were carried out under the following conditions: an initial denaturation at 94°C for 5 min; 35 cycles of denaturation at 94°C 30 s, annealing at 55°C 1 min, and extension at 72°C for 1.5 min. The final step of the PCR was an extension step at 72°C for 7 min, followed by cooling at 4°C. The PCR products were analyzed by gel electrophoresis. Two predicted endoglucanase genes and two xylanase genes were cloned and expressed as

described by Shi et al (2011) [29]. Briefly, the endoglucanase and xylanase genes were cloned into pET161 vector (Cat No. K160-01, Invitrogen, USA) with a 6 $\times$ His-tags. The enzyme expressions were induced in BL21 (DE3) cells with 0.5 mM IPTG at 25°C for 5 hours. The expressed enzymes were purified through a 5-ml nickel affinity column in AKTA FPLC system (GE healthcare, USA). Cellulase and xylanase activities were measured by the amount of reducing sugars released using dinitrosalicylic acid [92]. One unit was calculated as 1  $\mu$ mol reducing sugar released per minute using glucose as standard.

### Sequence Accession Numbers

This Whole Genome Shotgun project was deposited at DDBJ/EMBL/GenBank under the accession AKYZ00000000 and AKZA00000000 for grasshopper and cutworm, respectively. The version described in this paper is the first version, AKYZ01000000 and AKZA01000000. The Genbank ID for the four enzymes was as follows; cutworm EG1 is JX434086; grasshopper EG1 is JX434088; cutworm XYN1 is JX434089; and grasshopper XYN1 is KC155983.

### Supporting Information

**Figure S1** Phylogenetic tree of 16S rRNA annotated sequences. A. Grasshopper; B, cutworm. A total of 96 and 53 partial 16S rRNA sequences were extracted from cutworm and grasshopper gut microbiomes, respectively. The sequences were then aligned with the NAST aligner, and imported into an ARB database. The nearest aligned full length sequences were used for classification and phylogenetic tree construction using RAxML. Genbank accession numbers were presented in the figure.

(TIF)

**Figure S2** PCR amplification of cellulytic enzyme Open Reading Frames (ORFs) from the same metagenome DNA sample for sequencing library construction.

(TIF)

**Figure S3** Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of purified enzymes from cutworm and grasshopper microbiomes. M: Pertained marker (Invitrogen); 1: purified endoglucanase from grasshopper (GH-EG1); 2: purified endoglucanase from cutworm (CW-EG1); 3: purified xylanase from cutworm (CW-Xyn1); 4: purified xylanase from grasshopper (GH-Xyn1).

(TIF)

**Figure S4** The effect of temperature and pH conditions on enzyme activities (mean  $\pm$  SD) for the four enzymes cloned from cutworm and grasshopper microbiomes. A and B. One endoglucanases from grasshopper (GH-EG1) and one from cutworm (CW-EG1) gut microbiomes. C and D. one xylanase from grasshopper (GH-Xyn1) and one from cutworm (CW-Xyn1) gut microbiomes.

(TIF)

**Table S1** Summary of the 16S rRNA gene sequences identified from the PCR clone library of V3–V9 region for both the cutworm and grasshopper gut microbiome.

(PDF)

**Table S2** Enriched or under-represented KEGG pathway categories in grasshopper and cutworm gut microbiome as compared to those of termite gut.

(PDF)

**Table S3** Comparison of Glycosyl Hydrolase (GH), Carbohydrate Binding Modules (CBM), and Glycosyl Transferase (GT)

domain counts in grasshopper (GH), cutworm (CW), and termite (TM). GH stands for grasshopper, CW stands for cutworm, and TM stands for termite. (PDF)

**Table S4** Distribution of genes belonging to the phosphotransferase system (PTS) in the grasshopper (GH), cutworm (CW), and termite (TM). (PDF)

**Table S5** Comparison of grasshopper (G) and cutworm (C) gut microbiome with termite (T) gut microbiome showed the enrichment of energy production and conversion COGs. (PDF)

**Table S6** Comparison of grasshopper (GH) and cutworm (CW) gut microbiome with termite (TM) gut microbiome showed the enrichment for nitrogen metabolism KEGGs. (PDF)

## References

- Despres L, David JP, Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol Evol* 22: 298–307.
- Shi WB, Ding SY, Yuan JS (2011) Comparison of insect gut cellulase and xylanase activity across different insect species with distinct food sources. *Bioenergy Res* 4: 1–10.
- Shi WB, Uzuner U, Jesudhasan PR, Pillai SD, Yuan JY (2011) Comparative analysis of insect gut symbiotic composition and diversity as adaptation to different food type. *Biofuels* 2: 529–544.
- Dunbar HE, Wilson AC, Ferguson NR, Moran NA (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* 5: e96. doi:10.1371/journal.pbio.0050096
- Oliver KM, Degnan PH, Burke GR, Moran NA (2010) Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annu Rev Entomol* 55: 247–266.
- Ohkuma M (2003) Termite symbiotic systems: efficient bio-recycling of lignocellulose. *Appl Microbiol Biotechnol* 61: 1–9.
- Moran NA (2007) Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci U S A* 104 Suppl 1: 8627–8633.
- Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318: 283–287.
- Moran NA, Dunbar HE (2006) Sexual acquisition of beneficial symbionts in aphids. *Proc Natl Acad Sci U S A* 103: 12803–12806.
- Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42: 165–190.
- McCutcheon JP, McDonald BR, Moran NA (2009) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci U S A* 106: 15394–15399.
- McCutcheon JP, McDonald BR, Moran NA (2009) Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* 5: e1000565. doi:10.1371/journal.pgen.1000565
- Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA (2009) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc Natl Acad Sci U S A* 106: 9063–9068.
- Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 4: e188. doi:10.1371/journal.pbio.0040188
- Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, et al. (2006) The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science* 314: 267.
- Tsuchida T, Koga R, Horikawa M, Tsunoda T, Maoka T, et al. (2010) Symbiotic bacterium modifies aphid body color. *Science* 330: 1102–1104
- Sabree ZL, Kambhampati S, Moran NA (2009) Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont. *Proc Natl Acad Sci U S A* 106: 19521–19526.
- Shi WB, Syrenne R, Sun JZ, Yuan JS (2010) Molecular approaches to study the insect gut symbiotic microbiota at the 'omics' age. *Insect Sci* 17: 199–219.
- Dillon RJ, Webster G, Weightman AJ, Dillon VM, Blanford S, et al. (2008) Composition of Acridid gut bacterial communities as revealed by 16S rRNA gene analysis. *J Invertebr Pathol* 97: 265–272.
- Reeson AF, Jankovic T, Kasper ML, Rogers S, Austin AD (2003) Application of 16S rDNA-DGGE to examine the microbial ecology associated with a social wasp *Vespa germanica*. *Insect Mol Biol* 12: 85–91.
- Konopka A (2009) What is microbial community ecology? *Isme J* 3: 1223–1230.
- Tartar A, Wheeler M, Zhou X, Coy M, Boucias D, et al. (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuels* 2: 25.
- Scharf ME, Tartar A (2008) Termite digestomes as sources for novel lignocellulases. *Biofuel Bioprod Bior* 2: 540–552.
- Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450: 560–565.
- Oliver KM, Campos J, Moran NA, Hunter MS (2008) Population dynamics of defensive symbionts in aphids. *Proc Biol Sci* 275: 293–299.
- Degnan PH, Moran NA (2008) Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. *Mol Ecol* 17: 916–929.
- Hongoh Y (2010) Diversity and genomes of uncultured microbial symbionts in the termite gut. *Biosci Biotech Bioch* 74: 1145–1151.
- Joern A (1983) Host plant utilization by grasshoppers (Orthoptera: acrididae) from a sandhills prairie. *J Range manag* 36:793–797.
- Busching MK, Turpin FT (1997) Survival and development of black cutworm (*Agrotis ipsilon*) larvae on various species of crop plants and weeds. *Environ Entomol* 6: 63–65.
- Harris CR MJ, White GV (1962) The life history of the black cutworm, *Agrotis ipsilon* (Hufnagel), under controlled conditions. *Can Entomol* 94: 1183–1187.
- Moore JE (1998) Animal life cycles. Monterey, California, USA: Evan-Moor Educational Publisher.
- Schluter A, Bekel T, Diaz NN, Dondrup M, Eichenlaub R, et al. (2008) The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *J Biotechnol* 136: 77–90.
- Brown SD, Lamed R, Morag E, Borovok I, Shoham Y, et al. (2012) Draft genome sequences for *Clostridium thermocellum* wild-type strain YS and derived cellulose adhesion-defective mutant strain AD2. *J Bacteriol* 194: 3290–3291.
- Kopke M, Held C, Hujer S, Liesegang H, Wiezer A, et al. (2010) *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc Natl Acad Sci U S A* 107: 13087–13092.
- Mohammadi M, Younesi H, Najafpour G, Mohamed AR (2012) Sustainable ethanol fermentation from synthesis gas by *Clostridium ljungdahlii* in a continuous stirred tank bioreactor. *J Chem Technol Biot* 87: 837–843.
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66: 506–577.
- Snyder AK, Deberry JW, Runyen-Janecky L, Rio RVM (2010) Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *P Roy Soc B-Biol Sci* 277: 2389–2397.
- Breznak JA, Brune A (1994) Role of microorganisms in the digestion of lignocellulose by termites. *Annu Rev Entomol* 39: 453–487.
- Wenzel M, Radek R, Brugerolle G, Konig H (2003) Identification of the ectosymbiotic bacteria of *Mixotrichia paradoxa* involved in movement symbiosis. *Eur J Protistol* 39: 11–23.
- Vogel KJ, Moran NA (2011) Sources of variation in dietary requirements in an obligate nutritional symbiosis. *P Roy Soc B-Biol Sci* 278: 115–121.
- Sabree ZL, Degnan PH, Moran NA (2010) Chromosome stability and gene loss in cockroach endosymbionts. *Appl Environ Microbiol* 76: 4076–4079.
- Nikoh N, McCutcheon JP, Kudo T, Miyagishima S, Moran NA, et al. (2010) Bacterial genes in the aphid genome: absence of functional gene transfer from Buchnera to its host. *PLoS Genet* 6: e1000827. doi:10.1371/journal.pgen.1000827
- Hansen AK, Moran NA (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A* 108: 2849–2854.
- Sievers M, Swings J (2005) Family II. Acetobacteraceae. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, editors. *Bergey's manual of systematic bacteriology*, vol 2, *The Proteobacteria*. Springer, New York, pp 41–54

45. Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, et al. (2008) Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. *Science* 319: 777–782.
46. Brune A, Miambi E, Breznak JA (1995) Roles of oxygen and the intestinal microflora in the metabolism of lignin-derived phenylpropanoids and other monoaromatic compounds by termites. *Appl Environ Microbiol* 61: 2688–2695.
47. Butler JHA, Buckerfield JC (1979) Digestion of lignin by termites. *Soil Biol Biochem* 11: 507–513.
48. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32: D277–280.
49. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
50. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30.
51. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312: 1355–1359.
52. Watanabe H, Noda H, Tokuda G, Lo N (1998) A cellulase gene of termite origin. *Nature* 394: 330–331.
53. Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, et al. (2008) Genome of an endosymbiont coupling N-2 fixation to cellulolysis within protist cells in termite gut. *Science* 322: 1108–1109.
54. Hill AD, Reilly PJ (2008) Computational analysis of glycoside hydrolase family 1 specificities. *biopolymers* 89: 1021–1031.
55. Yin QY, Teng YG, Ding M, Zhao FK (2011) Site-directed mutagenesis of aromatic residues in the carbohydrate-binding module of *Bacillus endoglucanase* EGA decreases enzyme thermostability. *Biotechnol Lett* 33: 2209–2216.
56. Chou WY, Pai TW, Jiang TY, Chou WI, Tang CY, et al. (2011) Hydrophilic aromatic residue and in silico structure for carbohydrate binding module. *PLoS ONE* 6: e24814. doi:10.1371/journal.pone.0024814
57. Saier MH, Jr. (2001) The bacterial phosphotransferase system: structure, function, regulation and evolution. *J Mol Microbiol Biotechnol* 3: 325–327.
58. Deutscher J, Francke C, Postma PW (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* 70: 939–1031.
59. Higgins CF (1992) ABC transporters - from microorganisms to man. *Ann Rev Cell Biol* 8: 67–113.
60. Fath MJ, Kolter R (1993) ABC transporters - bacterial exporters. *Microbiol Rev* 57: 995–1017.
61. Marger MD, Saier MH (1993) A major superfamily of transmembrane facilitators that catalyze uniport, symport and antiport. *Trends in Biochem Sci* 18: 13–20.
62. Paulsen IT, Skurray RA, Tam R, Saier MH, Turner RJ, et al. (1996) The SMR family: A novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. *Mol Microbiol* 19: 1167–1175.
63. Paulsen IT, Brown MH, Skurray RA (1996) Proton-dependent multidrug efflux systems. *Microbiol Rev* 60: 575–608.
64. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496–512.
65. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270: 397–403.
66. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453–62.
67. Bult CJ, White O, Olsen GJ, Zhou LX, Fleischmann RD, et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273: 1058–1073.
68. Saier MH, Paulsen IT, Sliwinski MK, Pao SS, Skurray RA, et al. (1998) Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *Faseb J* 12: 265–274.
69. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawa H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331: 463–467.
70. P. J. Gullan PC (2000) *The insects: an outline of entomology* 2nd ed. Malden, MA, Oxford: Blackwell Science.
71. Zhou XG, Kovaleva ES, Wu-Scharf D, Campbell JH, Buchman GW, et al. (2010) Production and characterization of a recombinant beta-1,4-endoglucanase (Glycohydrolase family 9) from the termite *Reticulitermes flavipes*. *Arch Insect Biochem Physiol* 74: 147–162.
72. Sami AJ, Anwar MA, Rehman FU, Shakoori AR (2011) Digestive cellulose hydrolyzing enzyme activity (endo-beta-1, 4-D-glucanase) in the gut and salivary glands of blister beetle, *Mylabris pustulata*. *Pak J Zool* 43: 393–401.
73. Sugimura M, Watanabe H, Lo N, Saito H (2003) Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, *Psacothea hilaris*. *Eur J Biochem* 270: 3455–3460.
74. Arakawa G, Watanabe H, Yamasaki H, Maekawa H, Tokuda G (2009) Purification and molecular cloning of xylanases from the wood-feeding termite, *Coptotermes formosanus* shiraki. *Biosci Biotech Bioch* 73: 710–718.
75. Matteotti C, Haubruge E, Thonart P, Francis F, De Pauw E, et al. (2011) Characterization of a new beta-glucosidase/beta-xylosidase from the gut microbiota of the termite (*Reticulitermes santonensis*). *Fems Microbiol Lett* 314: 147–157.
76. Guazzaroni ME, Belouqi A, Golyshin PN, Ferrer M (2009) Metagenomics as a new technological tool to gain scientific knowledge. *World J Microbiol Biot* 25: 945–954.
77. Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, et al. (2005) Metagenomic gene discovery: past, present and future. *Trends Biotechnol* 23: 321–329.
78. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, et al. (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34: D344–348.
79. Marchler-Bauer A, Panchenko AR, Ariel N, Bryant SH (2002) Comparison of sequence and structure alignments for protein domains. *Proteins* 48: 439–446.
80. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32: D138–141.
81. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–288.
82. McHardy AC, Martin HG, Tsirigos A, Hugenholtz P, Rigoutsos I (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nat Methods* 4: 63–72.
83. Ghosh A, Dey N, Bera A, Tiwari A, Sathyaniranjan K, et al. (2010) Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. *Saline Systems* 6: 1.
84. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33: D294–296.
85. DeSantis TZ, Jr., Hugenholtz P, Keller K, Brodie EL, Larsen N, et al. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 34: W394–399.
86. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069–5072.
87. Stamatakis A, Ludwig T, Meier H (2005) RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456–463.
88. Rzhetsky A, Nei M (1992) A simple method for estimating and testing minimum evolution trees. *Mol Biol Evol* 9: 945–967.
89. Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* 101: 11030–11035.
90. Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, et al. (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* 36: D534–538.
91. Freedman D PR, Purves R (1998) *Statistics*. New York: W. W. Norton & Comp.
92. Miller GL (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal Chem* 31: 426–428.