environmental microbiology

Environmental Microbiology (2014) doi:10.1111/1462-2920.12526

Saccharide breakdown and fermentation by the honey bee gut microbiome

Fredrick J. Lee,[1] Douglas B. Rusch,[2] Frank J. Stewart,[3] Heather R. Mattila[4] and Irene L. G. Newton[1]*

[1]*Department of Biology and* [2]*Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN 47405, USA.*

[3]*School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA.*

[4]*Department of Biological Sciences, Wellesley College, Wellesley, MA 02481, USA.*

Summary

The honey bee, the world's most important agricultural pollinator, relies exclusively on plant-derived foods for nutrition. Nectar and pollen collected by honey bees are processed and matured within the nest through the activities of honey bee-derived microbes and enzymes. In order to better understand the contribution of the microbial community to food processing in the honey bee, we generated a metatranscriptome of the honey bee gut microbiome. The function of the microbial community in the honey bee, as revealed by metatranscriptome sequencing, resembles that of other animal guts and food-processing environments. We identified three major bacterial classes that are active in the gut (γ-Proteobacteria, Bacilli and Actinobacteria), all of which are predicted to participate in the breakdown of complex macromolecules (e.g. polysaccharides and polypeptides), the fermentation of component parts of these macromolecules, and the generation of various fermentation products, such as short-chain fatty acids and alcohol. The ability of the microbial community to metabolize these carbon-rich food sources was confirmed through the use of community-level physiological profiling. Collectively, these findings suggest that the gut microflora of the honey bee harbours bacterial members with unique roles, which ultimately can contribute to the processing of plant-derived food for colonies.

Received 10 March, 2014; revised 23 May, 2014; accepted 23 May, 2014. *For correspondence. E-mail irnewton@indiana.edu; Tel. 812 855 3883; Fax 812 855 6705.

Introduction

Many insects with specialist diets (e.g. nitrogen-limited, sap-based diets) live in symbiosis with microorganisms that supplement vital nutrients that are absent in the hosts' primary food (Douglas, 1998; Currie *et al*., 2003; Ohkuma, 2003; Thomas *et al*., 2009). Some of these insect-associated microbes (symbionts) provide their hosts with essential amino acids, while others provide them with vitamins, or aid in making nutrients more readily available for the host (Douglas, 2009). Many insects do not produce enzymes that digest complex and recalcitrant plant cellular materials. For these insects, such as termites, partnership with microbial symbionts that express cellulases or hemicellulases is crucial for obtaining energy and materials from a plant-based diet (Ohkuma, 2003; Newton *et al*., 2013).

The honey bee is an herbivorous insect that forages for food in the form of plant-produced nectar and pollen. Within their nest, worker bees process nectar and pollen into honey and bee bread respectively. Honey is the main source of carbohydrates for bees in a colony, while bee bread is their main source of essential amino acids, lipids, vitamins and minerals (Haydak, 1970; Herbert and Shimanuki, 1978; Brodschneider and Cralisheim, 2010). The mechanisms that underlie honey and bee bread production are not completely understood and probably depend in part on metabolic transformations by bee-associated microbial communities, both within worker host guts, as well as during maturation of honey and bee bread in wax comb and outside of the host (Anderson *et al*., 2011). Because the production of these food products involves inoculation of plant-derived resources with gut microflora from bees, gut community function and composition likely affects food production, and ultimately colony health. Although both maturation processes occur primarily outside of the honey bee body per se, each one requires worker-derived substances to create the final food product.

To make honey, nectar is collected from plants by foragers, which begin to reduce its water content through evaporation as they fly home (Nicolson and Human, 2008). The process of producing honey continues within the nest when enzymes (e.g. invertase) and microbes (from the honey bee crop, or foregut) are added to nectar

as it is handled by multiple workers to further reduce its water content through regurgitation and evaporation (Haydak, 1970; Olofsson and Vasquez, 2008). Honey, in its final form, is comprised primarily of a concentrated mix of sugars that can be stored in the nest for a long period of time (Doner, 1977).

The process by which bee bread matures is relatively poorly understood. When honey bees collect raw pollen from various plant sources and subsequently pack it into comb, workers add glandular secretions to the pollen before the cell is sealed with a drop of honey, creating what is thought to be an anaerobic environment (Herbert and Shimanuki, 1978; Gilliam, 1979a,b; Vasquez and Olofsson, 2009; DeGrandi-Hoffman *et al*., 2013). This mixture, called bee bread, matures for several weeks before it is consumed, resulting in a food that is usually more nutritious for honey bees than raw pollen alone (Beutler and Opfinger, 1949; Haydak and Vivino, 1950; Human and Nicholson, 2006; Ellis and Hayes, 2009). Pollen that is transformed into bee bread has reduced complex polysaccharides, a shift in amino acid profile and an increase in simple carbohydrates (Beutler and Opfinger, 1949; Herbert and Shimanuki, 1978). The final composition of bee bread is carbohydrate-rich (35–61% dry weight), lower in pH than raw pollen and contains essential amino acids in appropriate quantities required for honey bee development (Human and Nicholson, 2006; Brodschneider and Cralisheim, 2010). The exact mechanisms of bee bread maturation are unclear, but it is hypothesized that fermentation by anaerobic microbes is a major contributor to this process (Gilliam, 1979a; Vasquez and Olofsson, 2009). Indeed, lactic acid bacteria associated with the honey bee foregut are also found in bee bread, suggesting a link between these two microbial habitats (Vasquez and Olofsson, 2009). Thus, in addition to potentially aiding the enzymatic digestion of honey and bee bread in the midgut, it is likely that a worker's foregut microbial community is critical for the maturation of honey and bee bread *outside* of the host.

A recently published honey bee gut metagenome revealed the diversity of bacterial strains that exist within the honey bee gut and described potential functions for these strains (Engel *et al*., 2012). Polysaccharide-degrading enzymes as well as sugar-transport mechanisms were identified for bee-associated bacteria (Engel *et al*., 2012). In that study, bacterial isolates from the *Gilliamella* genus were found to have the potential to digest pectin, a polysaccharide found in pollen tubes. Interestingly, the ability to digest pectin was found to correlate with the evolutionary history of different *Gilliamella* isolates, suggesting niche adaptation through functional diversification (Engel *et al*., 2012).

Here, we aim to corroborate functions predicted by the honey bee gut metagenome (total DNA from the community) through characterization of the gut metatranscriptome (total RNA from the community). We utilized Illumina sequencing of cDNA generated from multiple honey bee guts to identify active microbial members and to infer community metabolic functions. Our analyses focused primarily on processes relating to carbohydrate metabolism, the dominant metabolic category identified from the transcriptome. Notably, we identified key microbial groups and fermentative pathways that are predicted to be active in the microbiome, including pathways that produce short-chain fatty acids (SCFAs), an important energetic product of fermentation. Additionally, we utilized community-level physiological profiling (CLPP) to identify microbial utilization of common carbon-based substrates (e.g. carbohydrates, amino acids) that are found in the honey bee diet, the consumption of which are predicted to occur based on our metatranscriptomic data. Finally, we statistically explored the use of a microbial RNA enrichment protocol, showing that subtractive removal of eukaryotic mRNA based on polyadenylation did not greatly alter our estimation of rank abundance in our samples and only increased our ability to deeply sample the host-associated microbial communities in them.

Results

Sequencing and assembly

Total RNA was extracted from three individual honey bee guts, from workers collected in Wellesley, MA. A total of 24 930 128 reads (1.25 Gb) from six libraries (range: 3 346 710–6 180 859 sequences per library) were generated from the honey bee gut metatranscriptome. These datasets, which represent composites of transcripts from the host *and* the microbial community, were pooled to generate a single, overall assembly, consisting of 25 170 contigs. A total of 19 875 contigs (or 79% of the total 25 170 contigs) were classified as belonging to the host honey bee, based on significant sequence homology to the *Apis mellifera* genome. Our analysis of metabolic pathways focuses solely on the remaining *non-host* fraction ($n = 5295$ contigs), with an N50 length of 5.5 Mb, an N50 value of 450 (i.e. half of the assembly was in 450 contigs) and an L50 of 6173 bp (i.e. the length of the shortest contig among those that sum to 50% of the assembly).

The honey bee gut metatranscriptome is dominated by γ-Proteobacteria, Bacilli and Actinobacteria

We assessed the taxonomic composition of the metatranscriptome-derived bee gut microbiome based on 16S rRNA reads. Because a significant fraction of reads did not map to the overall assembly (~ 9–25%, depending

Table 1. Total amount of sequencing and fraction of reads mapping to the entire metatranscriptome assembly before and after bacterial mRNA enrichment.

Library	Total num reads	High-quality, dereplicated reads	% Unmapped reads
Sample 1 – total RNA	4 430 459	1 765 311	8.81
Sample 1 – enriched RNA	3 426 500	1 682 487	23.39
Sample 2 – total RNA	3 758 829	1 582 842	9.84
Sample 2 – enriched RNA	3 346 710	1 710 925	20.85
Sample 3 – total RNA	6 180 859	2 607 509	14.49
Sample 3 – enriched RNA	3 786 772	1 626 544	25.51

The percentage of unmapped reads increases as the eukaryotic fraction is removed.

on the library, Table 1), we mined the reads for homology to a database of 16S rRNA sequences belonging to known members of the honey bee bacterial community, in addition to the ARB-SILVA database (Newton and Roeselers, 2012). We utilized 10 975 618 high-quality (q30/80) de-replicated reads for this analysis, 412 839 of which had significant homology to curated, 16S rRNA gene databases (~ 4% of our reads). Based on transcript fragment mapping (bit score > 50) from enriched libraries, we identified a total of 19 bacterial phyla in the bee gut, dominated by Proteobacteria, Firmicutes and Actinobacteria (Fig. 1A), recapitulating our previous work and those of others based on 16S rRNA gene sequences (Cox-Foster *et al*., 2007; Martinson *et al*., 2011; Mattila *et al*., 2012; Moran *et al*., 2012). The relative contribution of these bacterial groups to the metatranscriptomic libraries differed between the three enriched libraries; sample two contained many more reads homologous to the 16S rRNA from Lactobacilli (Firmicutes, Fig. 1A). The location to which short reads (100 bp) align along the length of the 16S rRNA can affect the accuracy of taxonomic designations (as in Dinsdale *et al*., 2008). To explore how this variability impacted our taxonomic analyses, we compared identifications based on short read alignments to distinct 16S rRNA hypervariable regions for Sample 1 (microbe-enriched). A total of 25 770, 29 675, 10 477 and 7539 reads mapped to the V1V2, V3V4, V5V6 or V7V9 regions of the 16S rRNA gene respectively (Table 2). At both the phylum and class levels, we saw a general concordance across the V1–V6 regions with regard to relative proportions of different bacterial groups (Table 2). However, reads mapping to the V7V9 regions was relatively enriched in sequences matching Firmicutes and relatively depleted in those matching Proteobacteria. The variability among regions increased further at lower taxonomic levels (order and family; data not shown). Based on these results, we report only phylum- or class-level taxonomic classifications for our other five libraries when mapping short reads to all regions of the 16S rRNA gene (Fig. 1).

Taxonomic composition was also inferred from the annotations of protein-coding genes identified on assembled contigs (referred to as *transcripts* throughout). Annotations provided by the Metagenomics Analysis Server (MG-RAST), based on homology to protein-coding genes, were used to taxonomically classify each contig (Meyer *et al*., 2008). The taxonomic composition (relative abundances of dominant bacterial members) derived from mining 16S rRNA reads broadly resembled that based on transcript annotation using MG-RAST. Importantly, the three most abundant bacterial classes based on 16S rRNA short reads (γ-Proteobacteria, Bacilli and Actinobacteria) contribute 91% of the protein-coding transcripts in the dataset (Fig. 1B, data pooled across all six libraries). The total number of *reads* mapping to the transcripts from the γ-Proteobacteria (402 795) was substantially higher than those mapping to Bacilli (282 949) and Actinobacteria (76 100).

In contrast to bacteria, archaea and microbial eukaryotes were *not* identified as prominent members of the honey bee gut microbiome. Only three contigs with homology to archaea were identified (average amino acid per cent identity: 64%). These transcripts, which included a trehalose-6-phosphate synthase, a glucose-1-phosphate cytidylyltransferase (EC 2.7.7.33) and an integral membrane protein, were not found in all libraries and ranged from 0–47 transcripts per million (TPM) (Supporting Information Table S1). In addition, archaea were rare among the short reads matching 16S rRNA genes. At 98% identity thresholds and using a bit score of 50, no reads matching archaeal 16S rRNA were identified, perhaps reflecting the relative paucity of sequences in the archaeal 16S rRNA gene databases. At 95% identity thresholds, 13 428 reads matched archaeal 16S rRNAs while 399 411 matched the bacterial 16S rRNAs. Similarly, when focusing on the unenriched total RNA libraries, microbial eukaryotes represented 1% of the actively transcribed non-rRNA *transcripts*, and based on 18S rRNA short read homology were found at much lower abundances than that of the dominant bacterial community in the metatranscriptome (42 762 reads with significant homology to microbial eukaryote 18S rRNA gene sequences). Although microbial eukaryotes did not contribute substantially to the gut metatranscriptome, these organisms, notably the ascomycete yeasts, have been reported previously in association with the honey bee and may play important roles in the processing of bee bread (Gilliam, 1979b).

Finally, taxonomic classifications were also performed by mapping our transcriptomic reads to the six taxonomic bins that were identified in the previously published metagenomic project: alpha-1, alpha-2, beta, bifido,

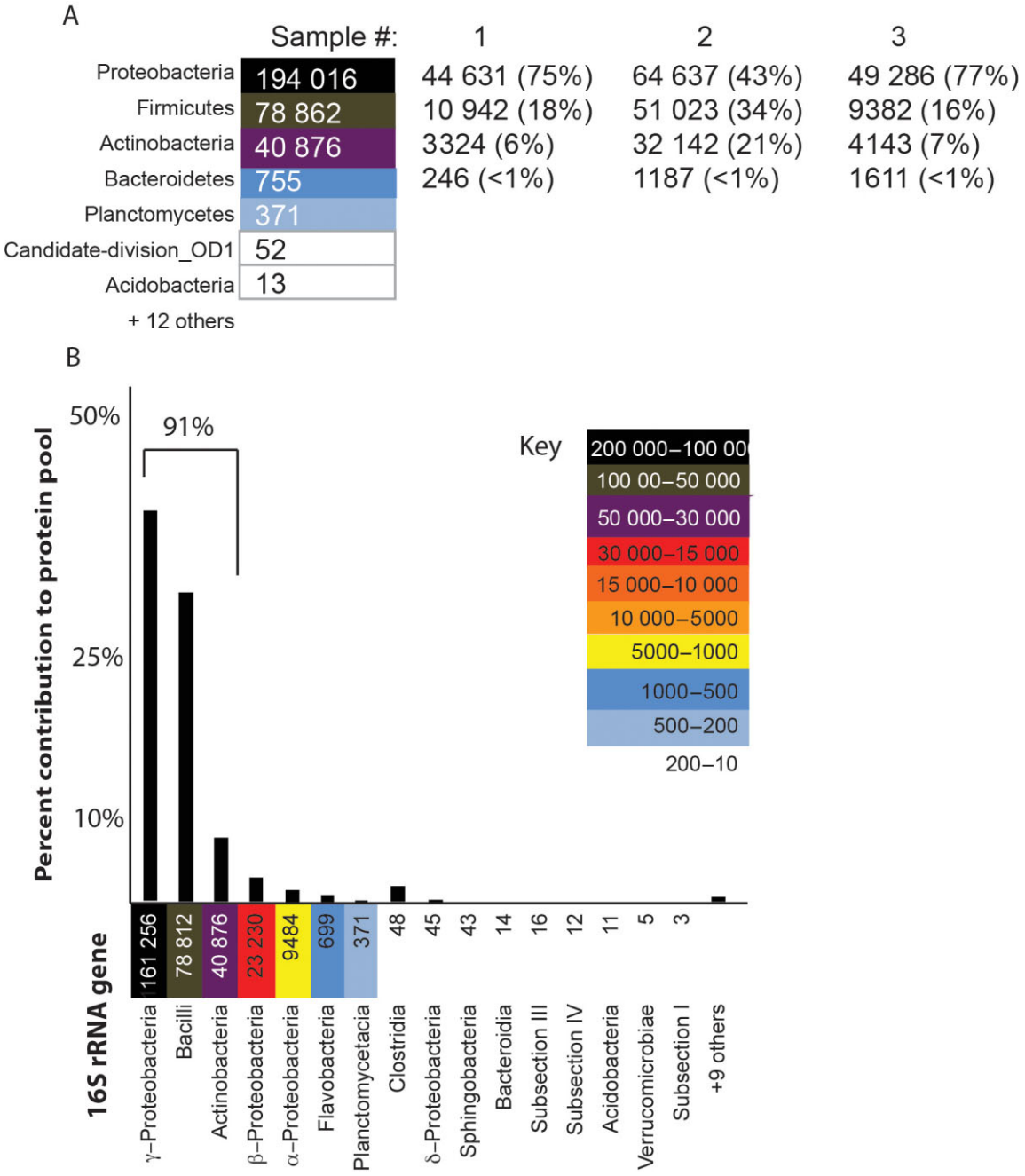

Fig. 1. The most active bacterial taxa based on 16S rRNA transcript abundance also contribute strongly to the functionally annotated mRNA transcripts.
A. The total number of transcripts with homology to bacterial 16S rRNA across three enriched metatranscriptomic libraries, each prepared from an individual honey bee digestive tract, indicate that the Proteobacteria, Firmicutes and Actinobacteria dominate the bee gut, although their relative percentages differ slightly across the three enriched samples (see sample 2). Only the top seven phyla are listed (12 others are present below this threshold).
B. The most abundant classes found in these three enriched libraries based on 16S rRNA read abundance contribute 91% of the total annotated, assembled mRNAs.

gamma and lacto (Engel *et al.*, 2012). These bins of genomic scaffolds differ in their total number of base pairs, and accordingly, the number reads from our transcriptome that map to each bin is clearly linked to bin size (Supporting Information Table S2). Additionally, these bins are substantially larger than most single bacterial genomes (the largest bin being almost 24 Mb in size), and therefore likely represent multiple strains within the honey

Table 2. Classification of high-quality reads from the honey bee gut transcriptome at the phylum and class levels using the 16S rRNA gene as a maker for a single honey bee sample (enriched RNA sample 1).

	V1V2	V3V4	V5V6	V7V9
Total reads mapping	25 770	29 675	10 477	7539
Phylum				
Proteobacteria	20 919 (81%)	23 125 (78%)	7976 (76%)	4992 (66%)
Firmicutes	3709 (14%)	4170 (14%)	2013 (19%)	1957 (26%)
Actinobacteria	1102 (4%)	2373 (8%)	478 (5%)	549 (7%)
Bacteroidetes	22	6	8	39
Acidobacteria	9	0	0	1
Deferribacteres	1	0	0	0
Gemmatimonadetes	1	0	0	0
Nitrospirae	1	0	0	0
SPAM	1	0	0	0
WPS-2	1	0	0	0
TM7	0	0	2	0
	V1V2	V3V4	V5V6	V7V9
Class				
γ-Proteobacteria	18 470 (72%)	20 104 (68%)	6878 (66%)	4130 (55%)
Bacilli	3700 (14%)	4167 (14%)	2013 (19%)	1955 (26%)
β-Proteobacteria	1722 (7%)	1942 (7%)	579 (6%)	606 (8%)
Actinobacteria	1102 (4%)	2373 (8%)	478 (5%)	549 (7%)
α-Proteobacteria	718 (3%)	1,078 (4%)	518 (5%)	256 (3%)
Flavobacteria	17	6	6	36
Clostridia	9	3	0	2
δ-Proteobacteria	7	0	1	0
Solibacteres	7	0	0	0
Bacteroides	5	0	2	1
Acidobacteria	2	0	0	1
SPAM	1	0	0	0
Deferribacteres	1	0	0	0
ε-Proteobacteria	1	1	0	0
Gemmatimonadetes	1	0	0	0
Nitrospira	1	0	0	0
Nostocophycideae	1	0	0	0
Synechococcophycideae	1	0	0	0
ζ-Proteobacteria	1	0	0	0
WPS	1	0	0	0
Sphingobacteria	0	0	0	2
TM7	0	0	2	0

The total number and the percentage of the reads classified at each level is shown if > 1% (in parentheses).

bee gut. Roughly three million reads were mapped to these 'core' honey bee taxa, although a great majority did not map (21 627 379, allowing for 20% divergence). These results can be explained in two ways: (i) these core taxa do not dominate transcriptionally *in vivo*, or more likely (ii) that representatives from the metagenome's 'core' taxa are divergent from those found in our metatranscriptomic dataset. That is, bees sampled for the metatranscriptome were hosts to these core groups of bacteria, but their members are genetically different from those in the published metagenome. Indeed, our assembled transcripts do match portions of the metagenomic contigs, although with a range of per cent identities (~ 60–100%, Supporting Information Fig. S1).

Carbon metabolism in the honey bee gut microbiome

The metatranscriptome of the bee gut microbiome is dominated by seven functional categories: protein metabolism, carbohydrate metabolism, RNA metabolism, respiration, membrane transport, stress response and virulence (Fig. 2A). Excluding functional categories associated with the core cellular processes of tRNA and ribosomal protein synthesis, which make up a large fraction (70%) of the protein metabolism category, the largest functional signature in the dataset is associated with carbohydrate metabolism (Fig. 2). As is true of the metatranscriptome overall, this functional category is dominated by transcripts matching genes of Bacilli, γ-Proteobacteria and Actinobacteria (Fig. 2). Outside of central intermediary metabolism (which includes glycolysis, the Entner–Doudoroff pathway, pentose phosphate and the citric acid cycle), the next most common metabolism identified in the dataset is the use of simple sugars (mono 214, and di- and oligosaccharides, 162 annotated transcripts), followed by fermentation (336 annotated transcripts) and sugar alcohol metabolism (119 annotated transcripts). We, therefore, focused

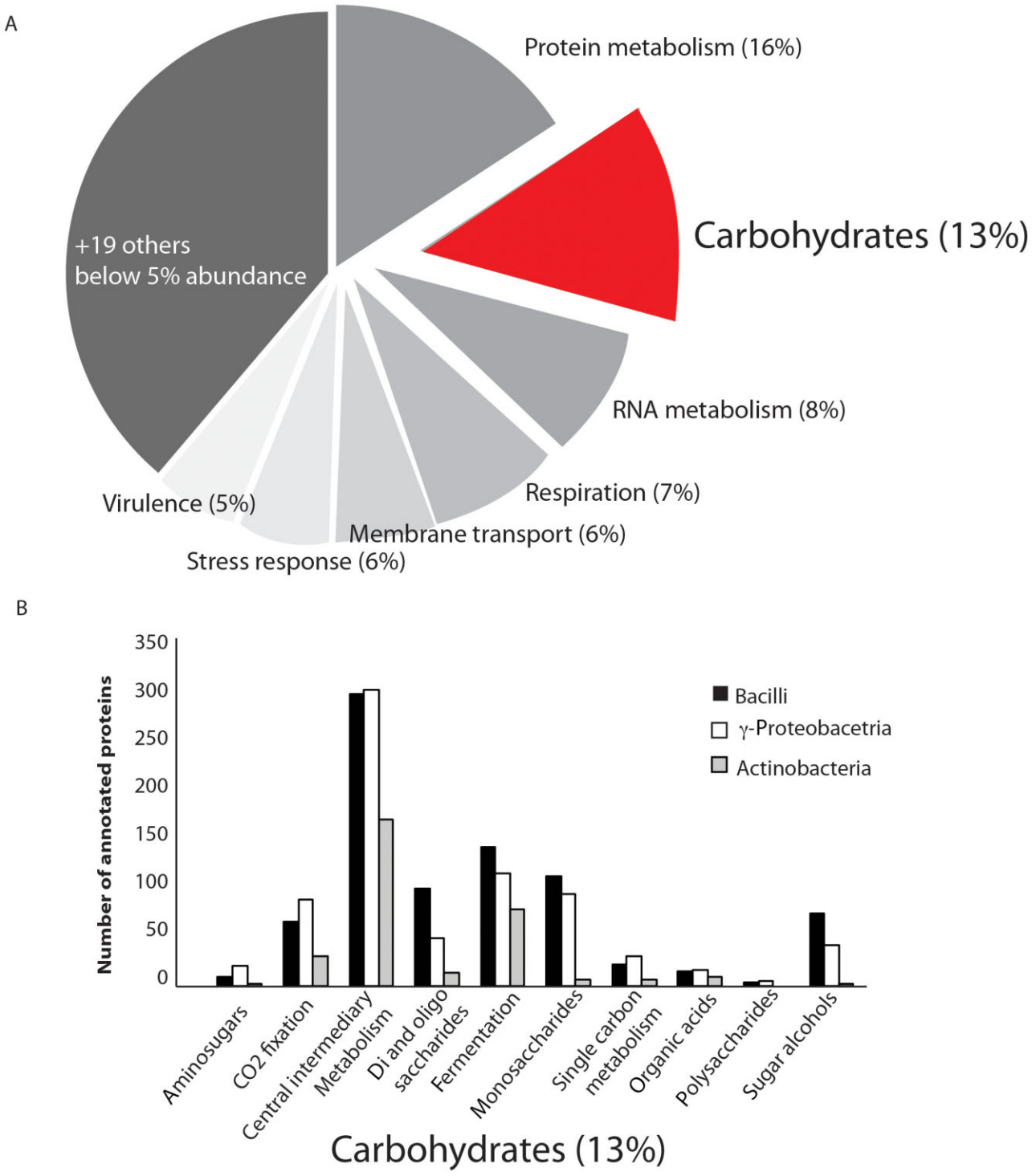

Fig. 2. The second most abundant functional class (based on MG-RAST subsystems hierarchical annotation) of transcripts is carbohydrate metabolism.
A. Thirteen per cent of the annotated transcripts are associated with the uptake, export and metabolism of carbohydrates.
B. The most abundant bacterial classes (γ-Proteobacteria, Bacilli and Actinobacteria) contribute to this category in varying proportions and with emphasis on different functions. For example, the Bacilli are most prominent in metabolizing di- and oligosaccharides, fermentation, monosaccharides, and sugar alcohols while the γ-Proteobacteria are enriched for transcripts involved in anaplerotic reactions and central intermediary metabolism. Importantly, like many functional annotation categories, overlap exists between categories with regard to membership of particular transcripts.

our analysis on the potential for the microbiota to digest macromolecules, take up sugars and ferment carbohydrates.

In order to generate a metabolic model for the microbial members in the gut, we utilized the published, assembled metagenome of the honey bee gut (Engel *et al*., 2012) in conjunction with our metatranscriptomic dataset. The assembled metagenome contigs were annotated via the same pipeline used for the metatranscriptome contigs (MG-RAST), and used to identify predominant carbohydrate utilization pathways. Interestingly, although the metagenome was generated from honey bees from a different geographical area (Arizona, USA), the functions predicted by the metagenome were largely in agreement with the metatranscriptomic data from our samples (Wellesley, MA), identifying a prominent role for fermentative pathways in the three major bacterial classes. This congruence suggests stability of the core honey bee microbiome, as has been observed previously (Martinson *et al*., 2011; Moran *et al*., 2012). Interestingly, however, we identified several carbon utilization pathways not identified in the metagenome (Fig. 3). Differences in the functional composition of the metatranscriptome and metagenome datasets may reflect variation due to methodology (RNA versus DNA sequencing), age and genetic backgrounds of the sampled bees, or microbiome composition as well as enrichment strategies used in the metagenomic project (Engel *et al*., 2012). Below, we describe key carbon utilization pathways identified in the metatranscriptome.

Saccharide and protein breakdown and uptake by the bacterial community in the honey bee

The metatranscriptome confirms a major role of the honey bee microbiome in the degradation of plant carbohydrates and polypeptides, indicating active transcription of genes encoding glycoside hydrolases to break down complex polysaccharides and peptidases for protein hydrolysis. Notably, we identified transcripts matching diverse peptidase-encoding genes from Bacilli, including genes for proline aminopeptidase (EC 3.4.11.5; TPM average 5.65), dipeptidase (EC 3.4.13.9; TPM average 3.55) as well as aminopeptidases C (EC 3.4.22.40; TPM average 7.83). In contrast, peptidase-encoding transcripts from the γ-Proteobacteria were represented by only a single oligopeptidase A with a relatively high expression level (EC 3.4.24.70; TPM average 32.33). The honey bee genome itself encodes several glycoside hydrolases, including some that may act on plant cell wall components, such as GH13 and GH16 families (Supporting Information Table S3). Interestingly, our transcriptome assembly did not include many of these genes; only two out of the five predicted CAZy enzymes were transcribed substantially in our honey bee gut samples. In contrast, we found evidence for the expression of several glycoside hydrolases by bacteria in the honey bee gut, the most highly transcribed of which came from the Bacilli (Table 3) and included those of the GH13 carbohydrate-active enzyme (CAZy) family (ECs 3.2.1.70 and 3.2.1.135, amylopullulanase). These glycoside hydrolases degrade alpha-1,4-glucosidic linkages between simple sugars but also have de-branching activity, hydrolysing the alpha-1- > 6 linkages in oligosaccharides (Ara *et al*., 1995). Substrates of the GH13 family include amylopectin and starch polymers commonly found within plant cells (Ball and Morell, 2003). A beta-glucosidase (EC 3.2.1.21), of the GH3 CAZy family, was also well represented in the metatranscriptome, matching genes of the Actinobacteria, γ-Proteobacteria and β-Proteobacteria (Table 3). This enzyme is responsible for the breakdown of diverse oligosaccharides found in cellulosic biomass. Enzymes of this class have unusually broad substrate specificities, being capable of binding to oligosaccharides with diverse sizes and residues, and could therefore contribute significantly in the breakdown of pollen. The metatranscriptome also contained sequences matching Clostridial and γ-Proteobacterial genes encoding alpha-mannosidase (EC 3.2.1.24, GH38 CAZy family), which is predicted to act on mannose containing substrates (Suits *et al*., 2010). Other GH families detected in the dataset (GH4, GH1, GH32) are predicted to act on the soluble disaccharides maltose, cellobiose and sucrose respectively (Table 3).

The metatranscriptome also encoded transporters and permeases for the uptake of key products of complex carbohydrate and protein degradation, including amino acids, peptides, di- and oligo-saccharides. Over 180 transcripts from the assembled dataset were classified as membrane transporters, including phosphotransferase systems (104 transcripts), ABC transporters (68 transcripts), permeases (19 transcripts) and symporters (9 transcripts). Interestingly, more than one third of these systems are specific to amino acids, not sugars (13 amino acid transporters, across all three bacterial classes compared with 21 sugar transporters) (Fig. 3; Table 4). These amino acid transporters include those taking up oligopeptides, dipeptides (e.g. putrescine and spermidine), urea and single amino acids (methionine, proline, alanine, glycine, isoleucine, valine, and leucine, arginine, serine, threonine, glutamate, aspartate and histidine). Based on both our metatranscriptomic data and the published metagenomic contigs (Engel *et al*., 2012), honey bee gut Bacilli are predicted to contain relatively few pathways for amino acid biosynthesis (the only complete metabolic pathway identified in the meta-omic data for Bacilli was for alanine anabolism). Bacilli likely take up

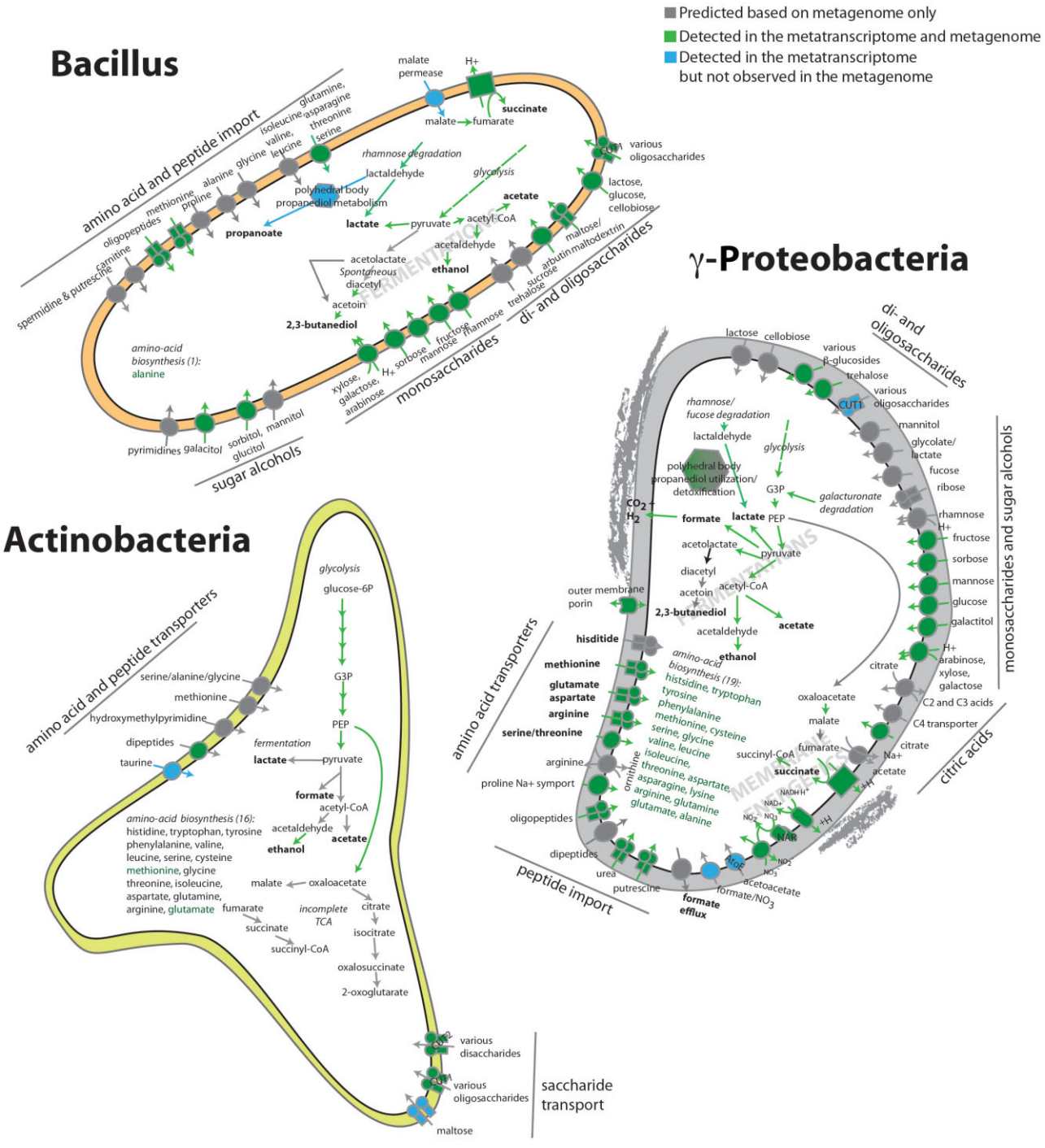

Fig. 3. Predicted uptake and metabolism of carbon and nitrogen sources by the three major classes of bacteria in the honey bee gut. These processes were found to be taxonomically associated with each of these three major classes, although we cannot definitively conclude that all processes occur within a single cell. Support for the presence of the metabolism in each bacterial class is shown and coloured based on source (the metagenome or the metatranscriptome). Every pathway depicted as transcribed is supported by a TPM of five in at least two of the six libraries.

amino acids from the environment of the honey bee gut (either derived from diet or from biosynthesis by other members of the community). In contrast, and again based on the combined datasets from the metagenome and metatranscriptome, the γ-Proteobacteria and the Actinobacteria are predicted to synthesize all essential amino acids, as well as other non-essential amino acids (Fig. 3, Table 4).

Table 3. Genes encoding carbohydrate active enzymes detected as transcribed in the honey bee gut based on their coverage (TPM values) and annotations from the metatranscriptome.

CAZy family	Putative function	Potential substrates	Phylum classification (no.)	Average TPM (min, max)
GH13	Amylase (EC 3.2.1.70); neopullulanase (EC 3.2.1.135);	Starch, amylopectin, glycogen	Bacilli (2)	34.64 (20.38, 50.91)
GH4	Maltose-6-phosphate glucosidase (EC 3.2.1.122)	Maltose	Bacilli (1)	24.98 (0, 49.96)
GH1	6-phospho-β-glucosidase/ β-galactosidase (EC 3.2.1.86)	Cellobiose	Bacilli (11); γ proteo (5); β proteo (1)	24.69 (0.45, 151.71)
GH32	Invertase; endo-inulinase	Sucrose	Bacilli (1); actino. (2)	17.12 (1.41, 45.81)
GH3	β-glucosidase/xylan 1,4-β-xylosidase (EC 3.2.1.21)	Cellulose, xylan oligosaccharides	Actino. (1); γ proteo (2); β proteo (1)	12.7 (3.97, 18.86)
GH38	α-mannosidase (EC 3.2.1.24)	Mannose- and xylose-containing glycans	Clostridia (1); γ proteo (1);	10.43 (10.87, 52.01)

All of these CAZy families were also identified in the published metagenomic analysis (Engel *et al.*, 2013). Data were sorted based on TPM value.

Diverse bacterial sugar transporters were detected in the honey bee gut. We detected transcripts matching Bacilli genes encoding transporters for 16 distinct saccharides, as well as transcripts matching the CUT1 family of oligosaccharide transporters, which are known for their broad substrate affinity. The γ-Proteobacteria-related transcript pool encoded at least nine distinct saccharide transporters, as well as two broad affinity transporters (Fig. 3). For the Actinobacteria, which were not as deeply sampled in this dataset, we detected transcripts encoding three sugar transporters, the CUT1 and CUT2 (disaccharide transporter) families and a maltose ABC transporter (Table 4).

Sugar fermentation and the production of SCFAs by the honey bee microbiota

The metatranscriptome data suggest that many of the sugars transported into the cytoplasm by bacteria in the bee gut are shunted through glycolysis or the pentose phosphate pathway, both of which were found as complete pathways in the metatranscriptome and metagenome (data not shown). These pathways generate the important metabolic intermediates phosphoenolpyruvate and pyruvate, as well as reducing equivalents (NADH, H^+) that require recycling in the honey bee gut, which is predicted to be an anaerobic habitat. These products likely play key roles in diverse fermentative pathways in the bee gut (described below).

Based on both metagenome and metatranscriptome evidence, acetate, ethanol, lactate, acetoin, propanoate and 2,3-butanediol are likely products of pyruvate fermentation by bee gut Bacilli (Fig. 3). Transformation of pyruvate to these products is evidenced by the detection of genes and transcripts encoding acetate kinase (EC 2.7.2.1), phosphate acetyltransferase (EC 2.3.1.8), lactate dehydrogenase (EC 1.1.2.4), alcohol dehydrogenase (EC 1.1.1.1), 2,3-butanediol dehydrogenase/ acetoin reductase (EC 1.1.1.76) and acetoin dehydrogenase (EC 2.3.1.190) (Fig. 3, Table 4). Short-chain fatty acids produced by the Bacilli (acetate, lactate) may then be utilized by the host or by other microbial community members. These SCFAs are predicted to readily diffuse through cell membranes or be taken up through specific transporters (e.g. AtoE) (Jenkins and Nunn, 1987; Matta *et al.*, 2007). Based on the metatranscriptome, the *Lactobacilli* are predicted to produce 2,3-butanediol, a carbon storage and anti-freeze compound, via diacetyl and acetoin in the pathway described for *Lactobacillus* (Crow, 1990). Interestingly, diacetyl and acetoin are both chemicals that produce butter flavour in fermented foods and are of interest to the dairy and wine industry. The production of 2,3-butanediol is stimulated by a slightly acidic pH and used by microbes to prevent intracellular acidification (Syu, 2001). If the honey bee gut is slightly acidic, as predicted based on the fermentative pathways identified herein, we would expect some of the pyruvate to be shunted to 2,3-butanediol production. The detection of Bacilli transcripts encoding propanediol dehydratase (EC 4.2.1.28) suggests that these bacteria are also involved in propanoate production in the bee gut (Table 4). Propanoate production via propanediol dehydratase is known to produce a highly toxic intermediate (propanal), which in many systems is sequestered within the so-called polyhedral bodies (Syu, 2001). We found evidence for the transcription of genes involved in both the production of propanal and the polyhedral bodies within the Bacilli. The metatranscriptome data support the hypothesis that the Bacilli experience an anaerobic environment in the honey bee gut and likely utilize fumarate reductase to generate an electrochemical gradient for ATP synthesis (as well as the acid succinate) (Table 4). No evidence for other mechanisms of energy generation by the Bacilli, aside from substrate-level

Table 4. Number of transcripts per million mapped reads (TPM) for bacterial transcripts important in different functional categories, identified to be associated with different classes, from three, microbially enriched honey bee metatranscriptomic libraries.

Bacterial class	Functional category	Annotation	*1*	*2*	*3*	Average
Actinobacteria	Transport of amino acids and peptides	Dipeptide transporter (TC 3.A.1.5.2)	0.00	78.06	0.00	26.02
	Transport of carbohydrates	Maltose/maltodextrin (Mal system)	0.00	26.78	0.00	8.93
		CUT1 family transporters, oligosaccharides (TC 3.A.1.1.-)	7.09	20.37	6.99	11.48
		CUT2 family transporters, monosaccharides (TC 3.A.1.2.-)	16.88	25.93	0.00	14.27
	Fermentation	Alcohol dehydrogenase (EC 1.1.1.1)	0.00	43.42	1.06	14.83
		Phosphoenolpyruvate carboxylase (EC 4.1.1.31)	0.00	60.54	0.00	20.18
γ-Proteobacteria	Transport of amino acids and peptides	Outer membrane porin 1b	355.44	63.08	2.78	140.43
		Putrescine ABC-type transporter (TC 3.A.1.11.2)	73.91	36.31	0.00	36.74
		Glutamate-aspartate ABC transport system (GltK, TC 3.A.1.3.4)	31.54	190.62	9.66	77.28
		Methionine ABC transport (TC 3.A.1.24.1)	25.45	3.22	3.23	10.63
		Oligopeptide transporter (TC 3.A.1.5.1)	216.60	133.28	56.96	135.62
		Urea ABC transporter	0.00	12.18	0.00	4.06
		Serine-Threonine transporter (DsdX)	0.00	161.33	0.00	56.78
		Arginine ABC transporter (Art type)	27.27	0.00	0.00	9.09
		Proline/Na + symport (TC 2.A.21.2.1)	33.72	0.00	0.00	11.24
	Transport of acids	Citrate transporter	271.10	193.25	73.74	179.36
		Formate/nitrite transporter	57.18	21.73	10.58	29.83
		Short-chain fatty acids/H + symporter (atoE, TC 2.A.73)	0.00	11.99	0.00	4.00
	Transport of carbohydrates	CUT1 family transporters, oligosaccharides (TC 3.A.1.1.-)	30.98	22.80	0.00	17.93
		Arabinose-proton symporter (AraE, TC)				
		Fructose PTS (TC 4.A.2.1.8)	58.45	446.99	3.35	169.60
		Glucose PTS (TC 4.A.1.1.1)	17.92	10.71	0.00	9.54
		Galactitol PTS (TC 4.A.5.1.1)	54.24	29.53	14.07	32.62
		β-glucoside PTS (TC 4.A.1)	46.15	12.92	0.00	19.69
		Man-family PTS (TC 4.A.6) – mannose, fructose, sorbose	89.35	110.79	0.00	66.71
		Trehalose PTS (TC 4.A.1.2.4)	0.00	14.22	0.00	4.74
	Fermentation					
	Propanediol metabolism	PduA (polyhedral body protein)	27.00	0.00	0.00	9.00
	Acetate	Phosphate acetyltransferase (EC 2.3.1.8)	87.66	157.87	0.00	81.84
		Acetate kinase (EC 2.7.2.1)	14.86	20.16	0.00	11.67
	Formate	Pyruvate formate-lyase (EC 2.3.1.54)	333.07	337.70	14.14	228.30
		Formate dehydrogenase (1.2.1.2)	19.83	92.84	0.82	37.83
	Succinate	Fumarate reductase (EC 1.3.99.1)	62.2	66.34	0	42.85
	Lactate	Lactate dehydrogenase (EC 1.1.1.28)	0.00	151.48	10.38	53.95
	Ethanol	Acetaldehyde dehydrogenase (EC 1.2.1.10/EC 1.1.1.1)	114.01	77.00	1.87	64.29
	2,3 Butanediol metabolism	Acetolactate synthase (EC 2.2.1.6)	66.35	76.53	0.00	47.63

Bacilli						
	Fermentation					
	Propanediol metabolism	Propanediol dehydratase (EC 4.2.1.28)	7.11	67.40	0.00	24.84
		PduJ (polyhedral body protein)	0.00	16.73	0.00	5.58
		PduN (polyhedral body protein)	17.61	6.22	0.22	8.02
		PduL (phosphate propanoyltransferase, EC 2.3.1.222)	0.00	16.73	0.00	5.58
		PduP (CoA-dependent propionaldehyde dehydrogenase, EC 1.2.1.87)	0.00	17.88	0.00	5.96
	Succinate	Fumarate reductase (EC 1.3.99.1)	0.00	47.67	0.00	15.89
	Acetate	Phosphate acetyltransferase (EC 2.3.1.8)	7.41	2.82	0.00	3.41
		Acetate kinase (EC 2.7.2.1)	0.00	28.93	0.00	9.64
		Pyruvate dehydrogenase (EC 1.2.1.51)	3.16	9.60	0.00	4.25
	Lactate	Lactate dehydrogenase (EC 1.1.1.28)	99.04	100.86	4.60	68.17
	Ethanol	Alcohol dehydrogenase (EC 1.1.1.1)	85.18	67.85	1.37	51.47
	2,3 butanediol metabolism	2,3-Butanediol dehydrogenase/acetoin reductase (EC 1.1.1.76)	6.36	24.32	4.81	11.83
		Acetoin dehydrogenase complex (EC 2.3.1.190)	0.00	43.61	0.00	14.54
	Butyrate	3-hydroxybutyryl-CoA dehydrogenase (1.1.1.157)	31.17	190.29	0.64	74.03
	Transport of amino acids and peptides	Amino acid ABC transport (PAAT, TC 3.A.1.3.-)	7.98	154.16	3.49	55.21
		Methionine ABC transport (Met system)	0.00	11.48	0.00	3.83
		Oligopeptide ABC transporter (Opt family, TC 3.A.1.5.1)	135.41	233.46	11.32	126.73
	Transport of short-chain fatty acids	Malate permease	13.79	14.01	0.00	9.27
	Transport of carbohydrates	Lactose/cellobiose PTS	41.20	216.20	0.00	85.80
		Fructose PTS (TC 4.A.2.1.8)	23.60	66.82	18.79	36.40
		CUT1 family transporters, oligosaccharides (TC 3.A.1.1.-)	32.25	72.85	0.00	35.03
		Arbutin PTS	13.10	3.33	0.82	5.75
		Arabinose/xylose proton symporter (TC 2.A.1.1.55)	359.39	0.00	0.00	119.80
		Maltose/maltodextrin ABC transport (TC 3.A.1.1.27)	49.43	12.13	3.28	21.61
		Rhamnose permease (2.A.1.1.74)	4.01	18.31	0	7.44
		Galactitol PTS (TC 4.A.5.1.1)	21.66	167.57	0.00	63.08
		Sorbitol/glucitol PTS system	0.00	18.60	0.00	6.20
		Man-family PTS (TC 4.A.6) – mannose, fructose, sorbose	167.91	465.91	80.32	238.05

phosphorylation, was found in our dataset or the metagenome (neither decarboxylation coupled to ion extrusion *nor* the efflux of organic acids).

Our data suggest that the γ-Proteobacteria also contributes strongly to fermentation within the honey bee gut. Like the Bacilli, these bacteria are predicted, based on the metatranscriptome, to produce lactate, acetate, ethanol and 2,3-butanediol. Instead of *generating* propanoate, the data suggested that the γ-Proteobacteria likely *utilize* this fatty acid, produced by the Bacilli. The metatranscriptome contained transcripts related to the γ-Proteobacteria and encoding polyhedral body structural proteins, thought to sequester the toxic intermediate involved in the utilization of propanoate, but the γ-Proteobacteria lacked the enzymes necessary for propanoate synthesis. This hypothesis is supported by the detection in the γ-Proteobacterial metagenome (but not the metatranscriptome) of two methylcitrate cycle enzymes (EC 4.2.1.99 and EC 4.2.1.117), which are involved in the consumption of propanoate (Engel *et al*., 2012). Unlike the Bacilli, and based on the metatranscriptomic data, the γ-Proteobacteria are predicted to generate formate through pyruvate-formate lyase (EC 2.3.1.54) and can presumably generate acetyl-coA without the reduction of NAD + . Transcripts encoding pyruvate-formate lyase are highly represented in the metatranscriptome data (Table 4), and this enzyme is expressed only under anaerobic conditions (as pyruvate dehydrogenase expression is repressed by NADH in the related γ-Proteobacterium, *Escherichia coli*) (Sawers and Suppmann, 1992). The γ- Proteobacteria metatranscriptome also encodes a formate dehydrogenase (EC 1.2.1.2, likely under redox control) that could, in conjunction with hydrogenase, convert the formate produced by pyruvate-formate lyase into molecular hydrogen (H_2) and carbon dioxide (CO_2). The production of these gases during anaerobic fermentation could enable interspecies hydrogen transfer in the honey bee gut. For example, the formate hydrogen lyase complex (FHL, made up of formate dehydrogenase and hydrogenase) could be coupled to energy conservation if the hydrogen produced by FHL activity is consumed by another community member (Dolfing *et al*., 2008). FHL is known to be controlled by the concentration of formate in the cell and by pH (Rossmann *et al*., 1991; Yoshida *et al*., 2005). As the formate dehydrogenase active site is periplasmic, the formate efflux transporter in the γ-Proteobacteria metagenome could provide the periplasmic formate necessary for FHL activity, although we did not find evidence of its expression (Fig. 3). Potentially, Clostridial acetogens could be involved in consuming the gases produced by FHL, generating acetate in the process *or* methanogenic archaea within the honey bee gut could be performing a similar role. The meta-omic data provide some evidence that both Clostridia and methanogenic archaea are present and may be participating in interspecies hydrogen transfer: a total of 46 transcripts with homology to the Clostridia were identified in the metatranscriptome, including transcripts matching acetate kinase (EC 2.7.2.1) (Supporting Information Table S1), while genes encoding archaeal 6-phospho-3-hexuloisomerase, part of the ribulose monophosphate pathway for formaldehyde fixation, were found in the metagenome (data not shown) (Engel *et al*., 2012).

Based on the metatranscriptome, the SCFAs lactate, formate and acetate are predicted to be produced by the Actinobacteria, as well as by the Bacilli and γ-Proteobacteria (Fig. 3). Additionally, the activity of phosphoenolpyruvate carboxylase (average TPM 20.18) was detected only in the Actinobacteria, suggesting that perhaps an incomplete TCA cycle could be generating intermediates for biosynthesis.

Confirmation of predicted utilization of carbon substrates through community-level metabolic profiling

Based on annotations of protein-coding genes and transcripts detected in this analysis, several sugars and amino acids are predicted to be utilized by the honey bee gut microbiome. We used Ecoplates™ to test the ability of the bee gut microbiome to metabolize diverse carbon substrates, which allowed us to confirm some of the metabolic predictions we made based on the metagenomic and metatranscriptomic data. Ecoplates contain a pre-aliquoted set of carbon substrates (in separate wells) that are combined with an indicator tetrazolium dye, whereby substrate utilization is detected through a concomitant colour/absorbance change due to reduction of the dye (see *Experimental procedure*). This community metabolic profiling was replicated using the gut microbiomes of 10 different bees (from Bloomington, Indiana) with plate incubations performed in an *anaerobic* environment. Indeed, incubation of plates with gut microbiomes at ambient oxygen partial pressure resulted in no colour change (data not shown). As confirmation of the fermentation of these substrates, we measured the resulting pH of each well after incubation (see *Experimental procedure*) and found an overall average drop from the original pH of 8 to 6.96 (min = pH 5, max = pH 8). We observed a high degree of variability in substrate utilization across the 10 gut communities (Fig. 4), reflective of both the stochasticity associated with sampling for the Ecoplate analysis (utilizing serial dilution) and potential differences in the abundance of functional groups among different bees. Statistically significant utilization patterns (green in Fig. 4) include the use of the amino acids asparagine, serine and glycyl-L-glutamate. As we found evidence of transporters for these amino acids and peptides in the

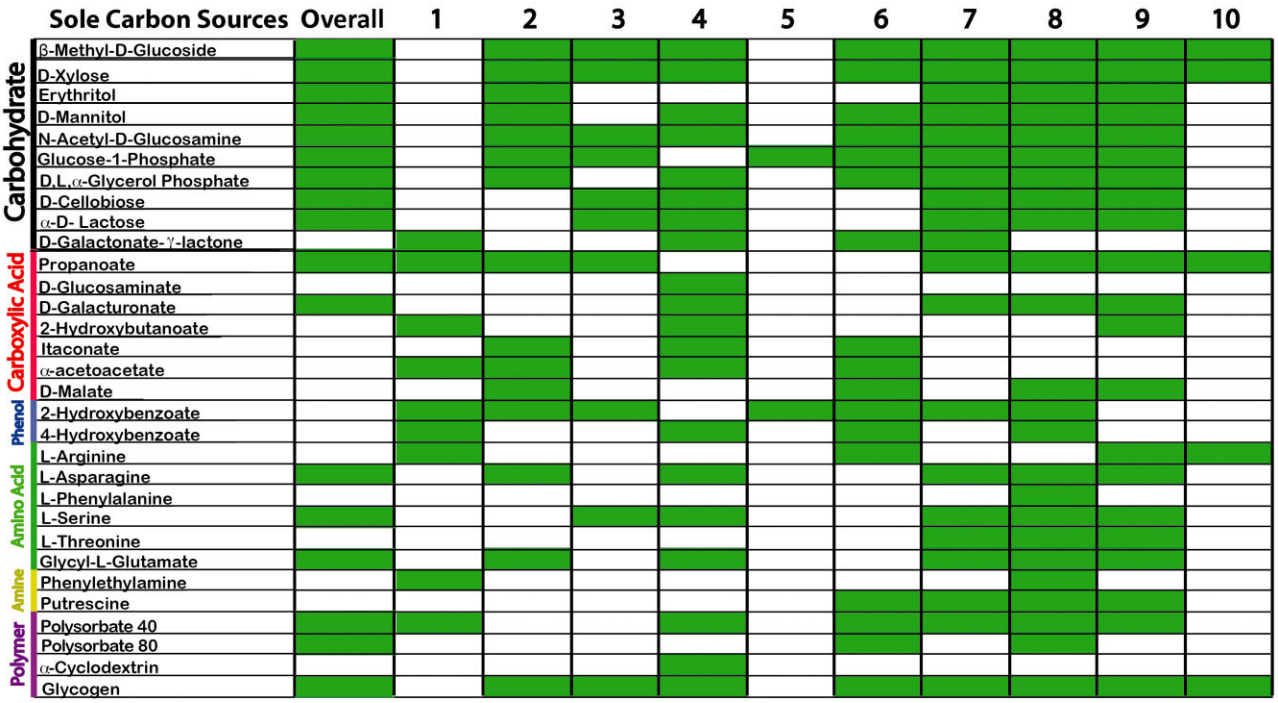

Fig. 4. Community metabolic profiling for 10 individual honey bee guts using 31 defined substrates in six categories (carbohydrate, carboxylic acid, phenol, amino acid, amine or polymer) with Biolog Ecoplates. To determine significant levels of metabolic utilization (as reflected by tetrazolium chloride (TTC) reduction measured spectrophotometrically), triplicates within each sample were compared with a sterile water-only substrate control. To identify statistically significant utilization across the entire dataset, sample profiles for each substrate were pooled across each of the 10 samples and compared with the control wells (with a Bonferroni correction). Samples that produced statistically significant redox reactions with TTC ($P < 0.05$), meaning they utilized the substrate, are coloured in green.

transcripts matching γ-Proteobacteria, variability in the use of this substrate among sampled microbiomes could reflect the differential abundance of this bacterial class across individual guts.

The gut community was able to utilize a large array of saccharides and diverse organic acids, including the carboxylic acids propanoate and galacturonate (Fig. 4). The uptake and utilization of propanoate have been discussed previously, and are likely performed by the γ-Proteobacteria. Galacturonate is a predicted degradation product of pectin, a polysaccharide found in plant cell walls. Glycogen and surfactant polymers (polysorbate 40 and 80) were also utilized by the gut community. Glycogen is a common carbon-storage compound that is used by animals, fungi and bacteria, and is similar in structure to amylopectin starch in plants (Ball and Morell, 2003). PS 80 is an ingredient in the media used to isolate *Lactobacilli* (de Man, Rogosa and Sharpe) and is thought to facilitate nutrient uptake (de Man *et al.*, 1960).

Nearly all sugars that were tested were utilized by the honey bee gut microbial community, with the exception of D-galactonate-γ-lactone, a cyclic intermediate in some galactose degradation pathways (Fig. 4). As lactonase (EC 3.1.1.25) was not found in either the metatranscriptome or the metagenome (Engel *et al.*, 2012), this result was not surprising. Evidence for the utilization of all but one of the sugars that were tested on the Ecoplates (xylose, mannitol, NAG, glucose-1-phosphate, cellobiose and lactose) was also provided by the combined meta-omic datasets. i-Erythritol was the only sugar compound that was utilized in the CLPP metabolic analysis for which we could not identify a specific biochemical mechanism for its metabolism in the meta-omic data. It is possible that some of the PTS sugar transporters, dehydrogenases and kinases that were identified in the meta-omic dataset could be promiscuous enough to metabolize this sugar.

The use of the MicrobEnrich™ kit does not systematically bias the data, but can result in loss of low-coverage transcripts

Depletion of eukaryotic RNAs prior to sequencing significantly increased the observed expression levels (number of reads mapping to assembled contigs) for many bacterial transcripts – a total of 2102 transcripts were enriched in mapped reads after the use of the MicrobEnrich™ kit (Supporting Information Fig. S2). On average, transcript abundance was increased threefold by enrichment; the mean and median fold-change values (before versus after

enrichment) were significantly greater than zero (mean = 3.24, median = 2.49; confidence interval = 3.01, 3.48). We were interested in identifying whether or not the use of MicrobEnrich would result in a bias in our transcriptome, for example by the preferential loss or enrichment of particular taxonomic categories or protein-coding genes. No single bacterial taxonomic class was lost in the enrichment process (Supporting Information Fig. S3). However, the enrichment protocol did lead to the loss of genes with low read coverage. Contigs with low TPMs (10–20) were lost in the *post*-enrichment libraries, likely during the extra time and manual handling of the RNA that is required for the protocol (Supporting Information Table S4). This category of transcripts would not have been detected if we had not sequenced both pre- and post-enrichment libraries, and included predicted metabolic functions such as propanediol utilization, β-glucosidases, PTS transport systems, urease and oligopeptide transporters, among others. However, genes enriched (or lost) in the MicrobEnrich™-treated dataset were not significantly associated with a specific functional or taxonomic category ($P > 0.05$, Kruskal–Wallis and χ^2 tests).

Discussion

Metatranscriptome sequencing revealed that major bacterial groups and fermentative processes are active in the honey bee gut. Bees forage on plant-derived nutrients (pollen and nectar), and it is important to consider how the chemical composition of these nutrients affects metabolic processing by the bee gut microbiome. Nectar is predominantly composed of sucrose and its component monosaccharides, fructose and glucose (Doner, 1977; Nicolson and Thornburg, 2007). Hydrolysis by the plant itself (Pate *et al*. 1985) results in the mix of disaccharides and monosaccharides that are observed in floral nectaries. In addition to these simple sugars, other trace monosaccharides are present in plant-derived nectar (mannose, arabinose, xylose), as well as the disaccharides maltose and melibiose, the oligosaccharide raffinose (on occasion) (Nicolson and Thornburg, 2007), and the sugar alcohol sorbitol (Haydak, 1970; Nicolson and Thornburg, 2007). However, not all of these sugars are equally utilized by the honey bee. Feeding experiments have determined that honey bee health declines when bee diets are restricted to certain carbohydrates, including rhamnose, fucose, mannose, sorbose, lactose, melibiose, dulcitol, erythritol and inositol (Haydak, 1970; Tan *et al*., 2007). Indeed, mannose is poisonous to honey bees, presumably due to an imbalance between high hexokinase activity and low mannose-6-phosphate isomerase activity, leading to an accumulation of mannose-6-phosphate and a decrease in ATP (Delafuente *et al*., 1986). However, the honey bee gut microbiome may facilitate the metabolism of this sugar (Engel *et al*., 2012), potentially detoxifying some nectars for honey bees (Adler, 2000).

Honey bees also derive much of their nutrition from pollen that is collected and matured in the nest to make 'bee bread', which is a combination of glandular secretions, nectar, and foraged pollen that is capped and allowed to mature for ~2 weeks before consumption (Haydak, 1970; Winston, 1987). Importantly, by adding contents of the honey bee foregut (crop), workers necessarily inoculate pollen with their own gut microflora. Pollen grains differ from other plant cells in that they have very thick, enzymatically recalcitrant cell walls that are composed of the poorly characterized compound sporopollenin, in addition to hemicellulose and lignin (Brooks and Shaw, 1968; Burton *et al*., 2010; Ariizumi and Toriyama, 2011). As a result, pollen is difficult to digest, and indeed honey bees cannot survive on raw pollen alone (Haydak, 1970). Several characteristics of bee bread suggest that the microbial community that is associated with the honey bee gut might assist in this process of pollen maturation. Specifically, bee bread is lower in pH than pollen, it contains fewer complex polysaccharides, and it undergoes a shift in amino acid profile (Gilliam, 1979a,b; Vasquez and Olofsson, 2009; DeGrandi-Hoffman *et al*., 2013). These changes do not preclude the further digestion of bee bread within the bee midgut.

Using metatranscriptomic (our data) and metagenomic (Engel *et al*., 2012) evidence from the honey bee gut microbiome and knowledge of the diet of the bee, we propose a conceptual model for a microbial food chain in the honey bee gut (Fig. 5). Bacterial members of the community that produce glycosidases and peptidases (the Bacilli, the γ- and β-Proteobacteria, the Actinobacteria, and the Clostridia) likely participate in the breakdown of plant polysaccharides and oligopeptides, initiating the processing of complex plant material. Shorter saccharides, peptides and amino acids produced by this extracellular digestion would be taken up and fermented by Bacilli, γ-Proteobacteria and Actinobacteria. The resultant products (organic acids, alcohols and gases) could then be consumed by methanogenic archaea (such as *Acidilobus* and *Methanocorpusculum*, both of which were detected in the metatranscriptome), as well as by Clostridia (such as *Clostridium*, *Ruminococcus* and *Anaerostipes* species detected in the metatranscriptome) (Supporting Information Table S1). This model assumes an anaerobic habitat within the honey bee gut, which is supported by our metatranscriptomic data (e.g. the high transcriptional activity of pyruvate-formate lyase and the presence of several fermentative pathways) and our inability to detect utilization of carbohydrates with Biolog Ecoplates under ambient oxygen pressure. That is not to

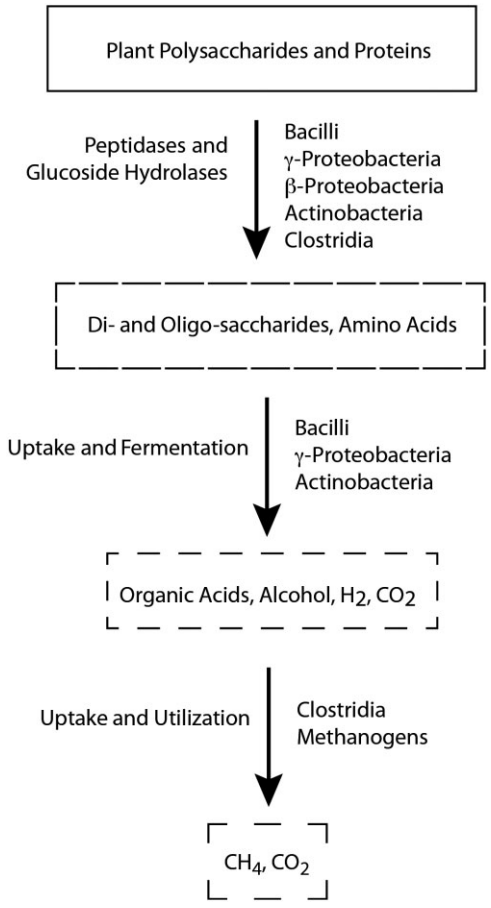

Fig. 5. Conceptual model of a microbial food chain in the honey bee gut based on metatranscriptomic (reported here) and metagenomic data (Engel *et al*., 2012). The degradation of plant-derived compounds, the fermentation of saccharides and amino acids, and the utilization of organic acids and gases are performed by different consortium members.

say that all of these bacteria are obligate anaerobes. Indeed, the production of the protective enzymes catalase (EC 1.11.1.6, with similarity to Bacilli, TPM range 0–8.69) and superoxide dismutase (with similarity to γ-Proteobacteria, TPM range 0–32.08) was found in the metatranscriptome, suggesting that these bacteria may withstand oxygen exposure.

The meta-omic data suggest potential roles for and syntrophies between the bacterial groups found in the honey bee gut. We predict the Bacilli are not able to make their own amino acids and instead rely on import of these substrates from the honey bee diet or from other microbes in the gut. Indeed, we found no evidence that the Bacilli contain or express TCA cycle enzymes, but they instead appear capable of reducing fumarate derived from malate (with the help of a malate permease; Fig. 3). This result is in line with previous observations for other *Lactobacilli* where incomplete or absent citric acid cycles are observed (Tammam *et al*., 2000; Chaillou *et al*., 2005; Diaz-Muniz *et al*., 2006). In contrast, however, both the Actinobacteria and the γ-Proteobacteria are predicted to be able to synthesize all essential amino acids (Fig. 3), and may therefore provide these amino acids to their host or to other microbial community members. Importantly, we could not detect the transcription of these amino acid biosynthetic genes in our dataset likely because of either (i) transcriptional regulation of these enzymatic pathways or (ii) relatively low sequencing depth for the Actinobacteria. Certain Actinobacteria in the honey bee gut may have an incomplete TCA cycle resembling that of obligate autotrophs, likely utilized for synthesis of essential intermediates with the consumption of reducing equivalents. Although the mechanism for transport of these intermediates from the Actinobacteria to the external environment has not been identified in sequencing projects so far, it is possible that these citric acids are secreted and then used by the γ-Proteobacteria, which harbour multiple citric acid importers (Fig. 3).

The stability of the core honey bee microbiome composition (based on 16S rRNA gene homologies) is well established, and these bacteria have been referred to as groups alpha-1, alpha-2 (within the α-Proteobacteria), beta (now *Snodgrassella alvi*, within the β-Proteobacteria), gamma-1 (*Gilliamella apicola* within the γ-Proteobacteria), gamma-2 (*Frischella perrara* within the γ-Proteobacteria), firm-4 and firm-5 (closely related to *Lactobacilli*), and bifido (related to *Bifidobacterium* within the Actinobacteria) (Martinson *et al*., 2011; Moran *et al*., 2012). Certain clades within this bacterial community are honey bee-specific and have new taxonomic designations (*Gilliamella*, *Snodgrassella*, *Frischella*) (Kwong and Moran, 2012; Engel *et al*., 2013). It is worth reemphasizing that functions predicted by our metatranscriptomic data largely corroborate the metagenomic data derived from geographically distant colonies (Engel *et al*., 2012). That being said, the metatranscriptome analysis revealed clear differences in the relative prevalence of bacterial classes and core honey bee specific groups (Fig. 1A, Supporting Information Table S2), as well as the expression of different bacterial genes among the three worker bees that were examined individually here (Table 4). In addition, community metabolic profiling analysis of 10 different worker bees suggested substantial, statistically significant differences in carbohydrate utilization between workers (Fig. 4). Although the factors that affect the composition and functional activity of the bee gut microbiome are not well understood, it is likely that the microbiome is significantly affected by the honey bee diet, which is correlated with age and caste membership (Haydak, 1970), and potentially social interactions experienced by workers. For example, we did not identify transcription of pectinases or pectate lyases in the transcriptome, although these

enzyme-coding genes are present in the metagenome, and both datasets were analysed for carbohydrate-active enzymes using the CAZy database (Engel *et al*., 2012). We did find that the community could utilize the derived sugar, D-galacturonate, through both the community physiological profiling assay and the expression of enzymes (UxaAC, KdgLK and KdgR, as well as the rhamnogalacturonide transporter RhiT). This discrepancy (between metagenomic and transcriptomic data) may be due to differences in the age of sampled bees, as older bees are expected to consume little in the way of pollen while younger adult bees are thought to eat a bee bread-rich diet (Hagedorn and Moeller, 1967; Haydak, 1970; Winston, 1987). Other potential explanations for the lack of a pectate lyase in our metatranscriptome include functional and composition variation among the microbiota of the sampled bees, or low sequencing depth. In the future, it will be of interest to examine the functional changes in the honey bee microbiome in response to variation in diet, season and age.

Our analysis of transcription in the honey bee gut highlights the importance of bacterial community members in the honey bee microbiome. The dominance of three major bacterial classes in the metatranscriptome data is supported by 16S rRNA gene amplicon studies (Mattila *et al*., 2012; Moran *et al*., 2012). However, our sequencing strategy revealed underexplored diversity. For example, the transcriptional activities of some of the members of the 'core' microbiome of the honey bee (consisting of the groups alpha-1, alpha-2, beta, gamma-1, gamma-2, firm-4, firm-5 and bifido) (Martinson *et al*., 2011) are rivalled in activity by other, potentially overlooked, bacterial groups that emerged in this dataset (e.g. see Clostridia in Fig. 1B, and Enterobacteriaceae and Flavobacteriaceae in Supporting Information Fig. S3). Previous to this work, the Clostridia have never been considered as significant contributors to the honey bee microbiome; culture-dependent studies have not been designed with this group in mind, and culture-independent studies have overlooked this group's potential contribution to the honey bee gut (Mattila *et al*., 2012; Vojvodic *et al*., 2013). In the metatranscriptome, 46 different genes belonging to the Clostridia were identified (128 contigs in the metagenome), including genes encoding glycoside hydrolases (Table 3, EC 3.2.1.24), as well as PTS systems and permeases for an array of sugars including galacticol, mannose, fructose, sorbose, lactate and glycerol. Clostridia may also participate in the production of butyrate, an important SCFA not produced by any of the 'core' members of the community (see Supporting Information Table 1). Additionally, these data propose a role for the methanogenic archaea in the honey bee, as fermentative metabolisms producing hydrogen and carbon dioxide are made more exergonic by the consumption of these gases. We suspect that deeper sampling of the bee gut community will continue to reveal more of its important functional members. Disentangling the relationships between these members and key environmental or life history variables (e.g. host diet, age structure) will be critical for understanding how gut microbiome composition and activity relate to overall honey bee health and nutrition.

Experimental procedure

Sampling, RNA extraction and microbial RNA enrichment

During summer 2011, adult workers were collected from a single, healthy hive of *Apis mellifera* located in a research apiary at Wellesley College in Wellesley, MA. Worker ages were not known, but we avoided collecting very young bees (slow and shiny-haired) or older bees (visible hair loss and wing wear) and collected only from the brood area of the colony. Thus, we presumed that workers were intermediately aged 'nurse' bees (Seeley, 1982) that consume bee bread and honey to produce food to feed to growing larvae (Winston, 1987). Hindgut dissections were performed in the field with aseptic technique and sterile equipment. Digestive tracts were removed and placed in sterile, screw cap tubes, which were immediately flash-frozen with liquid nitrogen. Samples were transported on dry ice and stored at −80°C until processed for RNA extraction. Total RNA was extracted from three samples (complete digestive tracts from three individuals) using a TRI reagent-based method coupled to lysis by grinding with a mortar and pestle under liquid nitrogen. To deplete the eukaryotic RNAs in each sample, we utilized the Ambion MicrobEnrich™ (Life Technologies, NY) kit on the total volume of each RNA sample, and followed the prescribed protocol. The MicrobEnrich™ kit utilizes a depletion technique whereby both polyadenylated mRNAs and eukaryotic ribosomal RNAs are removed from a mixture of RNAs using magnetic beads. We measured the integrity and quantity of RNA via the Agilent 2100 Bioanalyzer using an RNA Nano chip. Only samples with high-quality RNA (RNA integrity number: 9–10) were DNAse-treated, and they were then stored at −80°C until they were utilized for construction of cDNA libraries. We extracted ~ 30 μg of total RNA from each samples (34.7, 34.5 and 28.9 μg for samples 1, 2 and 3 respectively). After enrichment, a significant amount of RNA was still available for sequencing (3.0 μg, 4.1 μg and 2.3 μg for samples 1, 2 and 3 respectively).

cDNA synthesis and library preparation

Six cDNA libraries were constructed for sequencing. These libraries represent both a non-microbe enriched and a microbe-enriched sample (see above) from the complete digestive tracts for each of the three individuals (Table 1). Vacuum-concentrated RNA was added to elute, fragment and prime mix, and incubated at 94°C for 2 min. The Illumina TruSeq RNA Sample Preparation protocol was used without modification to prepare for multiplexed sequencing. Following purification with 1X Agencourt Ampure XP beads, double-

stranded cDNA samples were eluted in EB buffer and assayed on a D1K High Sensitivity Tape Station (Agilent) before cluster formation and sequencing on two lanes of a single flow cell run (100 × 100 bp paired end) on the Illumina Genome Analyzer IIx system.

Contig assemblies and data pre-processing

Quality controlled reads (q30 threshold) from all six samples were pooled and then assembled *de novo* using the transcriptome assembly program Trinity (Haas *et al.*, 2013). Trinity.pl was used (with parameters –seqType fq –JM 10G) on the merged, concatenated fastq files for F and R reads from all six libraries. We then used alignReads.pl to align the raw reads to the assembly and command-line utilities (bwa, -n 0.05 and samtools) to extract unmapped reads (Li and Durbin, 2009; Li *et al.*, 2009). Because the fasta headers produced by Trinity were incompatible with the bwa read aligner, we used an in-house perl script to remove the 'path = ' from each header. Bwa (Li and Durbin, 2009) was then used to make an IGV readable output of the assembly and to map reads from each library to the assembly (based on 95% ID). Honey bee host-derived contigs (19 875 contigs out of 25 170) were identified via BLASTN searches against the honey bee genome (Amel_4.0) and removed from the analysis (blast threshold: e value > 10^5, percent ID ≥ 99). Resulting 'non-bee' contigs were then used for functional gene annotation. The extent of coverage of bee-specific taxonomic groups identified in the previously published metagenome (Engel *et al.*, 2012) and the amount of sequencing effort expended on the host were determined by mapping reads (using bwa –n 0.20) to either the taxonomically identified scaffolds (JGI project ID 2498) or the honey bee genome (Amel_4.0). Additionally, we used blast to identify regions of homology between our assembled transcripts and the metagenomic contigs.

Functional annotation

Assembled metatranscriptome contigs (this study) and published metagenomic contigs from a prior honey bee gut microbiome analysis (JGI project ID 2498; (Engel *et al.*, 2012) were used as input to the MG-RAST (Meyer *et al.*, 2008). For taxonomic identification of protein-coding genes, we utilized the M5NR database with a 60% identity cut-off over a minimum alignment length of 50 amino acids, and $1e^{-5}$ confidence threshold. Putative functions of open reading frames were assigned by homology to proteins in the MG-RAST subsystems framework, a conglomeration of protein sequence databases including GO, JGI, KEGG, SEED and NCBI-nr databases using a 60% identity cut-off and minimum alignment length of 50 amino acids.

Library normalizations and TPM calculations

To control for both library size and depth of sequencing for different transcripts, and to identify the extent to which microbiome expression profiles differed between individual cDNA libraries, we utilized the program RSEM (Li and Dewey, 2011). We used our Trinity-assembled metatranscriptome as a reference (resem-prepare-reference –no-polyA) for calculating expression levels (and their modelled confidence intervals) for each of the six cDNA libraries (rsem-calculate-expression – calc-ci – paired-end). Within each library, expression values were sorted based on normalized values of fragments per kilobase per million mapped reads generated by RSEM. For between-library comparisons of expression, we used the normalized TPM counts and calculated confidence intervals (TPM_ci_upper and lower bound) to better assess statistical differences between libraries ($P < 0.05$). Fold change was calculated based on the average TPMs for each of the three libraries before and after enrichment [(average TPM_{after}/average TPM_{before}), where ratios less than 1 were replaced by the negative of the inverse].

16S rRNA gene transcript mining

We used the following pipeline to identify 16S rRNA transcripts. Non-assembled reads passing the quality filter (80% of bases higher than q30) were dereplicated using the FASTX toolkit and then queried via BLASTN against curated custom databases of near full-length bacterial and archaeal rRNA gene sequences (Newton and Roeselers, 2012) and the SILVA ARB-EUK database to identify prokaryotic and eukaryotic members respectively. Reads with significant matches above bit score 50 were classified according to the taxonomic annotation of the top blast hit. Because the 16S rRNA fragments identified in our analysis may cover either variable or conserved regions that differ in information content for taxonomic classification, we repeated our analyses using only taxonomically informative variable regions of the 16S rRNA gene, including the V1 + V2 region, V3 + V4, V5 + V6, and the V7 + V8 + V9 (V7V9 below) regions. These subregions of the alignment were generated using the MOTHUR software suite (summary.seqs in combination with screen.seqs on our aligned database).

Biolog Ecoplate analyses to confirm metatranscriptomic findings

To support the findings of the metatranscriptome analysis and to identify metabolic capabilities of the honey bee gut microbiome, we used CLPP with Ecoplates (Biolog, CA) to determine sole utilization of common carbon-based environmental substrates (31 substrates per plate). In the CLPP assay, a colorimetric change occurs as a result of the production of reducing equivalents, not necessarily coupled to growth, and the subsequent reduction of a tetrazolium dye (TTC), which absorbs at 590 nm. During summer 2013, 10 adult worker bees were collected from a hive located on a research plot at Indiana University in Bloomington, IN. Worker bees were collected in individual, sterile vials and brought into the laboratory. Hindgut dissections were performed using sterile technique, and individual bee gut samples were homogenized in sterile PBS (pH 8) with a sterile pestle. A 1:10 dilution was created from the initial homogenate and used for the CLPP analysis. Before incubation, initial absorbance at 590 nm was measured for each plate well. The 10 inoculated Ecoplates were then incubated under anaerobic conditions (using the GasPak EZ system,

BD, New Jersey) supplemented with CO_2 at 37°C for 5 days. To measure reduction of TTC (and utilization of substrate), absorbance readings were taken daily under anaerobic conditions. In order to support the hypothesis that fermentative processes resulted in the consumption of these substrates and the production of organic acids, after the last absorbance reading the pH of each well of two randomly selected plates was measured.

Statistical analyses

All statistical analyses were performed using the IBM SPSS software (v20). To determine substrates that were utilized by the honey bee gut community, we conducted *t*-Tests with a Bonferroni correction to compare the absorbance for each of the 31 carbon substrate across all 10 plates to that of the control wells (tetrazolium dye and buffer) after 4 days of incubation. We also utilized *t*-tests to confirm significant utilizations *within* a single plate, that is, per individual digestive tract, by comparing the absorbance readings for each substrate to the control wells (each in triplicate). Finally, Kruskal–Wallis tests were used to identify significant differences in functional category (MG-RAST based) abundances between datasets generated with and without microbial RNA enrichment.

Data deposition statement

Annotation data are available through the MG-RAST server (4519189.3) and raw reads deposited and made public through the NCBI SRA (ProjectID PRJNA253368).

Acknowledgements

We would like to thank Dr. Jake McKinlay for his thoughtful comments on the metabolic pathways proposed for these organisms, James Ford for his help in library generation and sequencing, and Newton laboratory members Melissa Horton, Randy Hicks, Zach Rokop and Kathy Sheehan for their assistance in annotation. We are grateful to two anonymous reviewers whose suggestions improved the manuscript. ILGN was supported by a generous start-up fund from Indiana University, and analysis of the data at Indiana University was supported by the National Center for Genome Analysis Support (NCGAS). HRM was supported by a Knafel Endowed Chair in the Natural Sciences from Wellesley College.

References

Adler, L.S. (2000) The ecological significance of toxic nectar. *Oikos* **91:** 409–420.

Anderson, K.E., Sheehan, T., Eckholm, B., Mott, B., and DeGrandi-Hoffman, G. (2011) An emerging paradigm of colony health: microbial balance of the honey bee and hive (*Apis mellifera*). *Insectes Soc* **58:** 431–444.

Ara, K., Saeki, K., Igarashi, K., Takaiwa, M., Uemura, T., Hagihara, H., *et al*. (1995) Purification and characterization of an alkaline amylopullulanase with both alpha-1,4 and alpha-1,6 hydrolytic activity from alkalophilic Bacillus sp. KSM-1378. *Biochim Biophys Acta* **1243:** 315–324.

Ariizumi, T., and Toriyama, K. (2011) Genetic regulation of sporopollenin synthesis and pollen exine development. *Annu Rev Plant Biol* **62:** 437–460.

Ball, S.G., and Morell, M.K. (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* **54:** 207–233.

Beutler, R., and Opfinger, E. (1949) Pollenernährung und Nosemabefall der Honigbiene (*Apis mellifica*) [Pollen feeding and Nosema infection of the honey bee (*Apis mellifica*)]. *Zeitschrift für Vergleichende Physiologie* **32:** 383–421.

Brodschneider, R., and Cralisheim, K. (2010) Nutrition and health in honey bees. *Apidologie* **41:** 278–294.

Brooks, J., and Shaw, G. (1968) Chemical structure of exine of pollen walls and a new function for carotenoids in nature. *Nature* **219:** 532–533.

Burton, R.A., Gidley, M.J., and Fincher, G.B. (2010) Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat Chem Biol* **6:** 724–732.

Chaillou, S., Champomier-Verges, M.C., Cornet, M., Crutz-Le Coq, A.M., Dudez, A.M., Martin, V., *et al*. (2005) The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. *Nat Biotechnol* **23:** 1527–1533.

Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., *et al*. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318:** 283–287.

Crow, V.L. (1990) Properties of 2,3-butanediol dehydrogenases from *Lactococcus lactis* subsp. lactis in Relation to citrate fermentation. *Appl Environ Microbiol* **56:** 1656–1665.

Currie, C.R., Wong, B., Stuart, A.E., Schultz, T.R., Rehner, S.A., Mueller, U.G., *et al*. (2003) Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science* **299:** 386–388.

DeGrandi-Hoffman, G., Eckholm, B.J., and Huang, M.H. (2013) A comparison of bee bread made by Africanized and European honey bees (Apis mellifera) and its effects on hemolymph protein titers. *Apidologie* **44:** 52–63.

Delafuente, M., Penas, P.F., and Sols, A. (1986) Mechanism of mannose toxicity. *Biochem Biophys Res Commun* **140:** 51–55.

Diaz-Muniz, I., Banavara, D.S., Budinich, M.F., Rankin, S.A., Dudley, E.G., and Steele, J.L. (2006) Lactobacillus casei metabolic potential to utilize citrate as an energy source in ripening cheese: a bioinformatics approach. *J Appl Microbiol* **101:** 872–882.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., *et al*. (2008) Functional metagenomic profiling of nine biomes. *Nature* **452:** 629–U628.

Dolfing, J., Jiang, B., Henstra, A.M., Stams, A.J., and Plugge, C.M. (2008) Syntrophic growth on formate: a new microbial niche in anoxic environments. *Appl Environ Microbiol* **74:** 6126–6131.

Doner, L. (1977) The sugars of honey – a review. *J Sci Food Agric* **28:** 443–456.

Douglas, A. (2009) The microbial dimension in insect nutritional ecology. *Funct Ecol* **23:** 38–47.

Douglas, A.E. (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. *Annu Rev Entomol* **43:** 17–37.

Ellis, A., and Hayes, G. (2009) An evaluation of fresh versus fermented diets for honey bees (*Apis mellifera*). *J Apic Res* **48:** 215–216.

Engel, P., Martinson, V.G., and Moran, N.A. (2012) Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci USA* **109:** 11002–11007.

Engel, P., Kwong, W., and Moran, N.A. (2013) *Frischella perrara* gen. nov., sp. nov., a gammaproteobacterium isolated from the gut of the honey bee, *Apis mellifera*. *Int J Syst Evol Microbiol* **63:** 3646–3651.

Gilliam, M. (1979a) Microbiology of pollen and bee bread – genus bacillus. *Apidologie* **10:** 269–274.

Gilliam, M. (1979b) Microbiology of pollen and bee bread – the yeasts. *Apidologie* **10:** 43–53.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., *et al*. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8:** 1494–1512.

Hagedorn, H., and Moeller, F. (1967) The rate of pollen consumption by newly emerged honey bees. *J Apic Res* **6:** 159–162.

Haydak, M., and Vivino, A. (1950) The changes in the thiamine, riboflavin, niacin, and pantothenic acid content in the food of female honeybees during growth with a note on the vitamin K activity of royal jelly and bee bread. *Ann Entomol Soc Am* **43:** 361–367.

Haydak, M.H. (1970) Honey bee nutrition. *Annu Rev Entomol* **15:** 143–157.

Herbert, E., and Shimanuki, H. (1978) Chemical composition and nutritive value of bee-collected and bee-stored pollen. *Apidologie* **9:** 33–40.

Human, H., and Nicholson, S. (2006) Nutritional content of fresh, bee-collected and stored pollen of *Aloe greatheadii* var. davyana (Asphodelaceae). *Phyotchemistry* **67:** 1486–1492.

Jenkins, L.S., and Nunn, W.D. (1987) Genetic and molecular characterization of the genes involved in short-chain fatty acid degradation in *Escherichia coli*: the ato system. *J Bacteriol* **169:** 42–52.

Kwong, W., and Moran, N.A. (2012) Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of Snodgrassella alvi gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and Gilliamella apicola gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order 'Enterobacteriales' of the *Gammaproteobacteria*. *Int J Syst Evol Microbiol* **63:** 2008–2018.

Li, B., and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323 DOI: 10.1186/1471-2105-12-323.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25:** 2078–2078.

de Man, J., Rogosa, M., and Sharpe, M. (1960) A medium for the cultivation of Lactobacilli. *J Applied Bacteriology* **23:** 130–135.

Martinson, V.G., Danforth, B.N., Minckley, R.L., Rueppell, O., Tingek, S., and Moran, N.A. (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* **20:** 619–628.

Matta, M.K., Lioliou, E.E., Panagiotidis, C.H., Kyriakidis, D.A., and Panagiotidis, C.A. (2007) Interactions of the antizyme AtoC with regulatory elements of the Escherichia coli atoDAEB operon. *J Bacteriol* **189:** 6324–6332.

Mattila, H.R., Rios, D., Walker-Sperling, V.E., Roeselers, G., and Newton, I.L.G. (2012) Characterization of the active microbiotas associated with honey bees reveals healthier and broader communities when colonies are genetically diverse. *PLoS ONE* **7:** DOI: 10.1371/journal.pone.0032962.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** DOI: 10.1186/1471-2105-9-386.

Moran, N.A., Hansen, A.K., Powell, E., and Sabree, Z.L. (2012) Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS ONE* **7:** e36393.

Newton, I.L.G., and Roeselers, G. (2012) The effect of training set on the classification of honey bee gut microbiota using the Naive Bayesian Classifier. *BMC Microbiol* **12:** DOI: 10.1186/1471-2180-12-221.

Newton, I.L.G., Sheehan, K., Lee, F.J., Horton, M.A., and Hicks, R.D. (2013) Invertebrate systems for hypothesis-driven microbiome research. *Microbiome Sci Med* **1:** 1–9.

Nicolson, S., and Thornburg, R. (eds) (2007) *Nectar Chemistry. Nectaries and Nectar*. Dordrecht, The Netherlands: Springer.

Nicolson, S.W., and Human, H. (2008) Bees get a head start on honey production. *Biol Lett* **4:** 299–301.

Ohkuma, M. (2003) Termite symbiotic systems: efficient bio-recycling of lignocellulose. *Appl Microbiol Biotechnol* **61:** 1–9.

Olofsson, T.C., and Vasquez, A. (2008) Detection and identification of a novel lactic acid bacterial flora within the honey stomach of the honeybee Apis mellifera. *Curr Microbiol* **57:** 356–363.

Pate, J.S., Peoples, M.B., Storer, P.J., and Atkins, C.A. (1985) The extrafloral nectaries of cowpea (Vigna unguiculata (L.) Walp.) II. Nectar composition, origin of nectar solutes, and nectary functioning. *Planta* **166:** 28–38.

Rossmann, R., Sawers, G., and Bock, A. (1991) Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. *Mol Microbiol* **5:** 2807–2814.

Sawers, G., and Suppmann, B. (1992) Anaerobic induction of pyruvate formate-lyase gene expression is mediated by the ArcA and FNR proteins. *J Bacteriol* **174:** 3474–3478.

Seeley, T.D. (1982) Adaptive significance of the age polyethism schedule in honeybee colonies. *Behav Ecol Sociobiol* **11:** 287–293.

Suits, M.D., Zhu, Y., Taylor, E.J., Walton, J., Zechel, D.L., Gilbert, H.J., *et al*. (2010) Structure and kinetic investigation of Streptococcus pyogenes family GH38 alpha-mannosidase. *PLoS ONE* **5:** e9006.

Syu, M.J. (2001) Biological production of 2,3-butanediol. *Appl Microbiol Biotechnol* **55:** 10–18.

Tammam, J.D., Williams, A.G., Noble, J., and Lloyd, D. (2000) Amino acid fermentation in non-starter Lactobacillus spp. isolated from Cheddar cheese. *Lett Appl Microbiol* **30:** 370–374.

Tan, K., Guo, Y.H., Nicolson, S.W., Radloff, S.E., Song, Q.S., and Hepburn, H.R. (2007) Honeybee (*Apis cerana*) foraging responses to the toxic honey of *Tripterygium hypoglaucum* (Celastraceae): changing threshold of nectar acceptability. *J Chem Ecol* **33:** 2209–2217.

Thomas, G.H., Zucker, J., Macdonald, S.J., Sorokin, A., Goryanin, I., and Douglas, A.E. (2009) A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Syst Biol* **3:** 24 DOI: 10.1186/1752-0509-3-24.

Vasquez, A., and Olofsson, T.C. (2009) The lactic acid bacteria involved in the production of bee pollen and bee bread. *J Apic Res* **48:** 189–195.

Vojvodic, S., Rehan, S.M., and Anderson, K.E. (2013) Microbial gut diversity of Africanized and European honey bee larval instars. *PLoS ONE* **8:** e72106.

Winston, M.L. (1987) *The Biology of the Honey Bee*. Cambridge, MA, USA: First Harvard University Press.

Yoshida, A., Nishimura, T., Kawaguchi, H., Inui, M., and Yukawa, H. (2005) Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing *Escherichia coli* strains. *Appl Environ Microbiol* **71:** 6762–6768.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Number of assembled contigs from the metatranscriptome, finding matches in the metagenomic data (Engel *et al*., 2012) at each percent identity. A bitscore threshold of 40 was used and all transcriptomic reads with hits above this threshold are plotted. The number of base pairs matching each group corroborates annotation based on MG-RAST in the primary manuscript: gamma = 2 409 080; lacto = 883 972; bifido = 114 347; alpha-1 = 279 914; beta = 194 979; alpha-2 = 79 722.

Fig. S2. A rank abundance blot of number of transcripts (i.e. assembled contigs) in the honey bee metatranscriptome and the fold change experienced after use of the MicrobEnrich protocol.

Fig. S3. Number of reads with matches to the 16S rRNA gene based on BLASTN (bitscore > 50) utilizing dereplicated, quality (q30/80) reads from the metatranscriptome. Number having matches to different bacterial phyla, classes and orders shown as a heat map.

Table S1. Contigs in the metatranscriptomic dataset containing genes with homology to archaea and Clostridiales and their average transcripts per million counts before (UE = unenriched) and after (E = enriched) use of the MicrobEnrich kit.

Table S2. The number of reads from the metatranscriptome that map to either the honey bee genome or honey bee-specific bacterial taxa identified in the metagenomic project (Engel *et al*., 2013) allowing a 20% edit distance. Total number of reads = 24 930 128.

Table S3. Genes encoding carbohydrate active enzymes in the *Apis mellifera* 4.0 genome, with CAZy annotation and transcriptional activity detected in the digestive tract.

Table S4. Descriptive statistics (95% CIs) for expression levels of bacterial mRNAs responding one of three ways to the MicrobEnrich protocol: an increase from zero, an increase or decrease from a detectable baseline, or the loss of that transcript.