Functional diversity within the simple gut microbiota of the honey bee

Philipp Engel[a,1], Vincent G. Martinson[b], and Nancy A. Moran[a,1]

[a]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520; and [b]Center for Insect Science, University of Arizona, Tucson, AZ 85721

Edited by Gene E. Robinson, University of Illinois at Urbana–Champaign, Urbana, IL, and approved May 21, 2012 (received for review February 27, 2012)

Animals living in social communities typically harbor a characteristic gut microbiota important for nutrition and pathogen defense. Accordingly, in the gut of the honey bee, *Apis mellifera*, a distinctive microbial community, composed of a taxonomically restricted set of species specific to social bees, has been identified. Despite the ecological and economical importance of honey bees and the increasing concern about population declines, the role of their gut symbionts for colony health and nutrition is unknown. Here, we sequenced the metagenome of the gut microbiota of honey bees. Unexpectedly, we found a remarkable degree of genetic diversity within the few bacterial species colonizing the bee gut. Comparative analysis of gene contents suggests that different species harbor distinct functional capabilities linked to host interaction, biofilm formation, and carbohydrate breakdown. Whereas the former two functions could be critical for pathogen defense and immunity, the latter one might assist nutrient utilization. In a γ-proteobacterial species, we identified genes encoding pectin-degrading enzymes likely involved in the breakdown of pollen walls. Experimental investigation showed that this activity is restricted to a subset of strains of this species providing evidence for niche specialization. Long-standing association of these gut symbionts with their hosts, favored by the eusocial lifestyle of honey bees, might have promoted the genetic and functional diversification of these bee-specific bacteria. Besides revealing insights into mutualistic functions governed by the microbiota of this important pollinator, our findings indicate that the honey bee can serve as a model for understanding more complex gut-associated microbial communities.

symbiosis | coevolution | pectin degradation | *Gilliamella* | *Snodgrassella*

As a pollinator, the honey bee, *Apis mellifera*, is a key species for agricultural production and contributes significantly to the human food supply (1, 2). Recent losses of *A. mellifera* and bumble bees (genus *Bombus*), and the potential association of these declines with various infectious agents, call for a better understanding of the bees' microbiota (3, 4). In many animals, the gut microbial community, in particular, confers functions related to nutrition and susceptibility to disease and thus might also play an important role in the health and resilience of honey bees.

Honey bees pool resources, divide labor, and communicate in highly structured social colonies (5). Sterile female worker bees predominate within colonies in which they initially clean cells, rear brood, and store food, then leave the hive, and forage for pollen and nectar (5). Diet and nutrition also shift as workers age (6, 7), but throughout the life cycle, pollen represents the only dietary source of fat and amino acids (8).

Independent studies of bacterial community profiles based on 16S rRNA sequences show that workers of *A. mellifera* and some *Bombus* species consistently harbor a distinctive gut microbiota not shared with solitary bees (9–11). This microbiota consists of eight distinct species or phylotypes (i.e., closely related strains with ≥97% sequence identity in 16S rRNA sequences, hereafter referred to as species): three Gram-positive species (two closely related Firmicutes within *Lactobacillus* and one within *Bifidobacterium*) and five Gram-negative species (one β-proteobacterium with the *Candidatus* name "*Snodgrassella alvi*," two closely related γ-proteobacteria, one with the *Candidatus* name "*Gilliamella apicola*," and two α-proteobacteria) (12). Recent deep sequencing of 16S rRNA amplicons from guts of individual worker bees has confirmed these results and has further shown that these species make up more than 99% of all bacterial sequences detected in the worker bees' intestines and that the β-proteobacterium (hereafter, *Snodgrassella* both γ-proteobacteria, and both *Lactobacilli* are present in every bee (13). The consistent occurrence of these particular species and their deep phylogenetic divergence from other known bacteria suggests coevolution and mutualistic interactions with their hosts. Indeed, in a European *Bombus* species, it was recently shown that the gut microbiota is transmitted via feces and protects against the intestinal trypanosomatid pathogen *Crithidia bombi* (14). In addition to defending against pathogens, the gut microbiota might be involved in nutritional processes such as breakdown and utilization of pollen grains or degradation of toxic compounds encountered in the environment (15, 16).

Analyses of 16S rRNA sequences are useful for characterizing species composition but provide little information about metabolic capabilities of these species. Although a 16S rRNA sequence can indicate a close phylogenetic relationship to organisms for which physiology has been characterized, even closely related species can display massive differences in functional gene content, due to the dynamic nature of bacterial genomes. For example, strains of *Escherichia coli* differ by <1% in 16S rRNA sequences but have very different gene sets and effects on host organisms. In extending beyond 16S rRNA analyses, a first picture of metabolic and functional capabilities can be revealed from genomic sequences, which enable reconstruction of pathways and functions with strong predictive value. Such information from genome sequences provides a rich set of candidate genes and processes for experimental tests of functions.

Here, we report analyses of the metagenomic sequence of the resident gut microbiota of *A. mellifera*. Our results confirm the simple composition of the microbiota on the species level, but we discovered extensive strain-level diversity within each of the bacterial species present. Annotation and pathway reconstruction based on the metagenomic gene content suggest that specific functions related to host interactions, digestion, and defense are associated with particular bacterial species or with strains within species. We investigated experimentally one such potential function, that of pectin degradation, an activity presumably involved in breakdown of

Author contributions: P.E., V.G.M., and N.A.M. designed research; P.E. and V.G.M. performed research; P.E. analyzed data; and P.E. and N.A.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The 16S rRNA gene sequences reported in this paper are deposited in the GenBank database (accession nos. JQ673232–JQ673261; and the metagenome deposited in the Integrated Microbial Genomes with Microbiome Samples (IMG/M), http://img.jgi.doe.gov/cgi-bin/m/main.cgi (project ID 2498).

[1]To whom correspondence may be addressed. E-mail: philipp.engel@yale.edu or nancy.moran@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1202970109/-/DCSupplemental.

pollen walls. Our experiments demonstrate that differences in gene content within species result in differences in functional capabilities that are expected to affect host ecology.

Results and Discussion

Metagenomic Sequencing and Strain Variation in Species of the Gut Microbiota of Honey Bees. De novo assembly of 81,343,096 paired-end reads resulted in 54,700 scaffolds of contigs with a total length of 76.6 Mb (*SI Appendix*, Table S1). Annotation of the gene content predicted 112,128 protein-coding sequences (CDSs). Analyses of the overall community profile indicated that the metagenome is dominated by sequences from α-, β-, γ-proteobacteria, *Lactobacillus* (Firmicutes), and *Bifidobacterium* (Actinobacteria) (Fig. 1 *A* and *B*). Reads corresponding to 16S rRNA genes verified that the eight typical species previously detected in honey bees (9, 11–13) were all present and dominated the gut community. Phylogenies inferred from eight conserved, single copy marker genes (Fig. 1*C* and *SI Appendix*, Fig. S1) further revealed that the taxonomic diversity is largely limited to six distinct phylogenetic lineages, referred to as Alpha-1, Alpha-2, Snodgrassella, Gamma, Firm, and Bifido. These distinct clusters reflect the eight dominant species with the two closely related Firmicutes (Firm-4 and Firm-5) and the two γ-proteobacteria (*Gilliamella* and Gamma-2) each being pooled into one group. Interestingly, the presence of several divergent copies of marker genes per phylogenetic cluster (Fig. 1*C* and *SI Appendix*, Fig. S1) suggested the existence of marked variation despite the limited number of major species in the honey bee gut. Notably, a few sequences showed closest homology to Bacteroidetes, probably from a species previously detected in both honey bee and *Bombus* guts (Fig. 1*B*, *SI Appendix*, Fig. S1*B*) (10, 17). Because of the very low representation of this lineage (indicated by small contigs and shallow read coverage) we did not include it in subsequent analyses.

To analyze gene contents specific to a given species or species group, we assigned scaffolds to the six different phylogenetic clusters shown in Fig. 1*C*, using a combination of homology- and nucleotide composition-based binning methods (*SI Appendix*, Table S1). We then assessed the completeness of the genomic content for each of these bins by determining the presence of a minimal set of essential (mostly single-copy) genes (18). Except for the Bifido cluster, we identified full-length homologs for ≥85% of these gene sets in all bins, and even genes scored as missing were usually present as gene fragments. Thus, each bin contained sequences representing nearly complete genomic contents for the corresponding species or species group (Fig. 1*D*), implying nearly complete genomic representation for the major species of the bee gut microbiota.

Inspection of these bins revealed that there were several related genomes within each, as indicated by the existence of several distinct copies of the genes in the minimal gene set (Fig. 1*D*). Particularly within Alpha-1, Gamma, and Firm, individual genes were present as numerous copies. The presence of multiple strains was also supported by (*i*) the total contig length obtained for these bins exceeding the size of typical bacterial genomes (*SI Appendix*, Table S1) and (*ii*) our phylogenetic marker gene analysis, indicating the presence of several divergent copies per phylogenetic

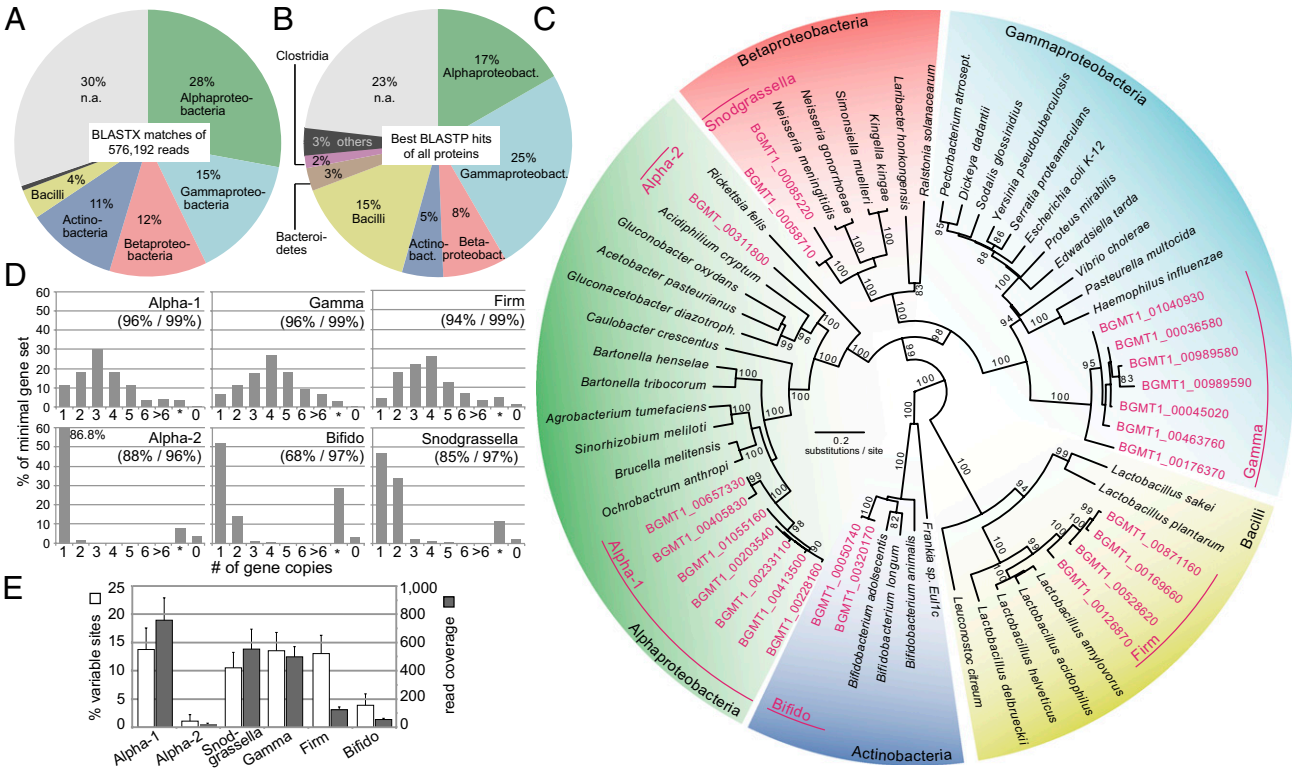

Fig. 1. Identification of bacterial species or species groups in the honey bee microbiome and analysis of their genetic diversity. Phylogenetic profile based on (*A*) classification of 576,192 reads mapping against 31 marker proteins with MetaPhyler (48) and (*B*) best BLASTP hit distribution of all 112,128 CDSs. n.a., reads or CDSs not assigned. (*C*) Maximum-likelihood protein phylogeny of UvrC. All eight phylogenies (see *SI Appendix*, Fig. S1 for other trees) revealed that most sequences from the honey bee microbiome (shown in pink) fall into the same six distinct clusters. These phylogenetic clusters are referred to as Alpha-1, Alpha-2, Snodgrassella, Gamma, Bifido, and Firm. We considered all closely related taxa with available genomes for this analysis. Bootstrap values >80 are shown. (*D*) Percentages of the minimal gene set present in each bin are depicted in parentheses (only full-length copies/including fragmented genes). Graphs show distribution of genes of the minimal gene set based on identified full-length copies per bin. Asterisks indicate fragmented genes. (*E*) Average percentage of variable sites and average read coverage for 27 ribosomal protein-encoding genes of each bin.

cluster (Fig. 1*C* and *SI Appendix*, Fig. S1). Pairwise protein identities of homologs of the minimal gene set in each bin mostly ranged from 80 to 100% (*SI Appendix*, Fig. S2*A*), providing further evidence for the presence of extensive strain variation. To quantify the degree of diversity, we remapped all reads against 27 ribosomal protein-encoding genes including one set of genes for each bin. Detection of single nucleotide polymorphisms (SNPs) revealed high percentages of variable sites in Alpha-1 (13.8% ± 3.8), Snodgrassella (10.5% ± 2.7), Gamma (13.5% ± 3.2), and Firm (13.1% ± 3.2). The lower percentages of variable sites detected for Bifido (3.9% ± 2.0) and Alpha-2 (1.1% ± 1.2) might reflect low representation in the metagenomic dataset (Fig. 1*E*). Despite having low numbers of gene copies for the minimal gene set, the Snodgrassella bin contained a high percentage of variable sites. This discrepancy probably reflects the poor assembly of sequences from this bin resulting in many fragmented gene copies.

Overall, these findings corroborate surveys based on 16S rRNA that indicated dominance of the bee gut microbiota by eight characteristic bacterial species. Further, they show that in a single honey bee colony, distinct strains of these species coexist. The low frequencies of many SNP variants, particularly in bins with a high number of variable sites (*SI Appendix*, Fig. S2*B*), provides additional evidence for the presence of many, closely related strains. Recent analysis of 16S rRNA sequences revealing multiple strains of *Snodgrassella* and *Gilliamella* even within individual bees (13) further support our findings. Maintenance of this diversity is likely promoted by the eusocial lifestyle of honey bees. First, colonies are founded by swarms consisting of a large number of workers and their queen, so the associated gut bacteria do not undergo a bottleneck by passing through a single individual founder. Second, trophallaxis and other interactions within the colony could facilitate recurrent colonization of individual bees. Third, intercolony interactions such as robbing of food resources in neighboring hives and mixing of colonies by beekeepers might further enhance bacterial diversity. The implications of diversity in the bee gut microbiota are unknown. Variants of a given species may colonize different microenvironmental niches in the gut or potentially, may compete among each other, even at host expense. Alternatively, as in other microbial gut communities (19, 20), this diversity might confer fitness benefits to bees when facing environmental challenges such as exposure to pathogens, varying food sources, or toxic compounds.

Gene Functions Enriched in the Metagenome of Honey Bees. The metagenome provides an initial picture of the functional capabilities of the bee gut microbiota. For an indication of what functions might be unusually prominent in the honey bee-associated community, compared with the gut microbiota of other animals, we compared the profile of Clusters of Orthologous Group (COG) functions to nine metagenomic datasets from the gut-associated microbial communities from five mammals and four insects (21–26). We found a total of 72 COGs to be significantly enriched in the honey bee dataset in at least eight of the nine comparisons (*SI Appendix*, Table S2). "Carbohydrate metabolism and transport" was the most abundant category among these functions (20%, *SI Appendix*, Fig. S3) and was predominant among those with the highest enrichment scores (Fig. 2). A majority of these functions represent components of various phosphotransferase systems (PTSs), which are responsible for the import of sugars from the environment (27) and were particularly prominent in the Firm and Gamma bins. The enrichment of such transporters could be an adaptation to the diet and gut environment of the honey bee. Bee gut bacteria are expected to encounter a broad spectrum of carbohydrate substrates, including various sugars present in nectar and honey or originating from pollen cell walls or from host glycans. Sugar uptake and metabolism involves a complex regulatory network (27), and many sugar transporter operons, such as the PTSs, encode their own regulatory genes (28), which could explain why a number of transcriptional regulator COGs belonged to the

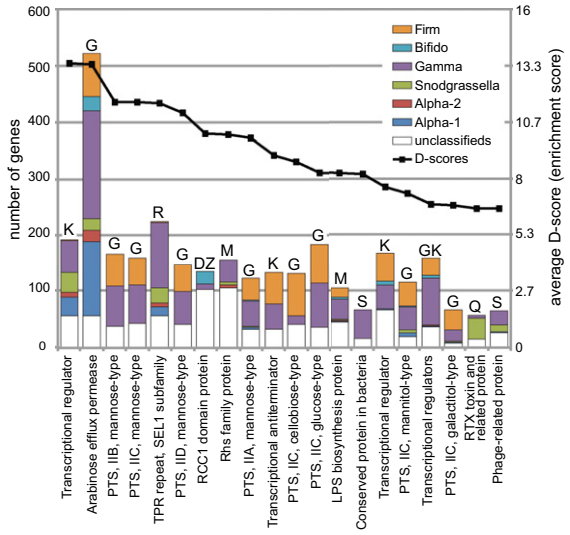

Fig. 2. Functions enriched in the honey bee microbiome relative to nine other gut microbiomes. These represent the 20 most enriched functions from a total of 72 COG functions that received significant enrichment scores in eight of nine comparisons (*SI Appendix*, Table S2). Average *D*-score represents the mean enrichment score over all nine comparisons. For each function, the total number of genes found in the metagenomic dataset and the distribution in the six bins is shown. Letters above graphs indicate COG category. "G" and "K" stand for "carbohydrate metabolism and transport" and "transcription," respectively.

most enriched functions (Fig. 2). Another carbohydrate-related function highly enriched in the honey bee microbiome and detected in all bins, was the family of "arabinose efflux permease." Proteins of this COG function belong to the major facilitator superfamily responsible for the import or export of a broad range of different substrates, including sugars, drugs, and peptides (29). We found a high diversity among these proteins in our metagenomic data with a marked number showing homology to drug resistance efflux pumps (*SI Appendix*, Fig. S4). Honey, in particular, is known for its broad antimicrobial activity resulting from high sugar concentration, low pH, and the presence of various antimicrobial compounds (30). The latter compounds are either produced in the bee itself or are molecules naturally present in nectar or pollen (31). Plants produce a plethora of defensive compounds, as protection against herbivores or phytopathogens (32). Similar to plant-associated microbes (33), bacterial species colonizing bee guts might have acquired an arsenal of such efflux systems to ensure resistance against the broad range of plant-derived antimicrobial compounds as well as the varied carbohydrates ingested by bees as part of their nectar and pollen diet. Exposure of honey bees to pesticides, like antibiotics applied to the hive or used in agriculture (16) or antimicrobial substances produced by the gut bacterial community itself could impose additional selective pressure favoring accumulation of these efflux pump functions.

Gene Functions Specific to Different Species or Species Groups. Different species localize to specific niches in the honey bee gut (12), suggesting that they fulfill distinct functions. To identify such species-specific functions, we determined which COG categories were enriched within each bin relative to other bins. We used normalization ranks of COG categories based on the relative abundance of all COG functions belonging to a given category (*SI Appendix*, Table S3).

Host Cell Interaction and Biofilm Formation. Previous analysis using fluorescence microscopy revealed that *Snodgrassella* and *Gilliamella* (the dominant component of our Gamma bin) form biofilm-

like layers on the epithelium of the longitudinal invaginations of the ileum, a segment of the hindgut connecting midgut and rectum (12; 34). *Snodgrassella* is in direct association with the host tissue followed by a thick layer of *Gilliamella.* Strikingly, the Snodgrassella and Gamma bins showed significant enrichment in the categories of "intracellular trafficking, secretion, and vesicular transport" and "cell motility" (*SI Appendix*, Table S3). In both bins, we detected genes for type IV pili/type II secretion. The Gamma bin was further found to encode flagellar and type VI secretion system genes (*SI Appendix*, Fig. S5). Genes encoding RTX proteins, a family of surface proteins with a broad spectrum of activities including host cell interaction (35), were additionally enriched in Snodgrassella. Together these gene functions are likely involved in the formation of the biofilm on the gut epithelial surface and in the intimate interaction with the host. Interestingly, the microbial community of *Bombus* species, shown to protect against a protozoan parasite, is dominated by *Gilliamella* and *Snodgrassella* (14), suggesting the possibility that the *Gilliamella*/*Snodgrassella* biofilm functions as a protective layer against parasite invasion. The mechanism of this protection has not yet been elucidated. Besides providing a physical barrier, defenses could be mediated by identified secretion systems or RTX proteins. In particular, the RTX proteins might display important determinants of this function, as they can act as bacteriocins or contribute to defense against environmental stressors by forming protective surface layers or by supporting biofilms via cell–cell adhesion (35). Alternatively, RTX proteins could play a role in establishing the intimate association of *Snodgrassella* with the epithelium after pupal eclosion and thus might affect developmental processes in the newly formed gut of adult honey bees (12).

Carbohydrate Transport. Sugar transport functions were enriched in the overall community of honey bees compared with other gut-associated microbiomes (Fig. 2). These carbohydrate-related functions were specifically abundant in the Gamma, Firm, and Bifido bins compared with the remaining bins (*SI Appendix*, Table S3). Relatives of all three species or species groups include typical gut commensals known to metabolize carbohydrates (36). The large number of different PTSs particularly present in the Gamma and Firm bins suggests that they metabolize a variety of sugars. The most abundant PTSs were transporters of the mannose family, known for their broad substrate specificity (37) (*SI Appendix*, Fig. S6). In contrast to nutritive sugars like sucrose, glucose, or fructose, various nectar- and pollen-derived sugars cannot be metabolized by bees or even are toxic (15, 38, 39). In particular, mannose and melibiose are known to be poisonous to newly emerged workers (39). By metabolizing such sugars, the gut microbiota could be critical for the detoxification of food components. Besides such a mutualistic role, the large variety of transporter systems could facilitate these bacteria to thrive in the gut environment of honey bees by broadening the spectrum of potential energy sources.

Carbohydrate Breakdown. Polysaccharides present in pollen walls or host glycans as well as certain nectar sugars have to be broken down before they can be used as energy sources. Accordingly, carbohydrate-degrading enzymes were found in the Gamma, Firm, and Bifido bins (*SI Appendix*, Table S3). Using the CAZy (carbohydrate-active enzyme) database (40), we classified all glycoside hydrolases (GHs) and polysaccharide lyases (PLs) of the honey bee microbiome (*SI Appendix*, Table S4 and Dataset S1). This analysis identified fructosidases to be highly abundant in the metagenome. These enzymes hydrolase the disaccharide sucrose into fructose and glucose and thus could support breakdown of one of the most common sugars in nectar (38).

We further detected a number of genes encoding enzymes that target plant cell wall polysaccharides (*SI Appendix*, Table S4). Whereas cellulases and hemicellulases were mostly absent, we found genes encoding pectin-degrading enzymes, including pectate lyases targeting the polygalacturonic acid (PGA) backbone of pectin and debranching enzymes (Fig. 3). Most of these enzymes as well as specific transporters and pathways to import and catabolize

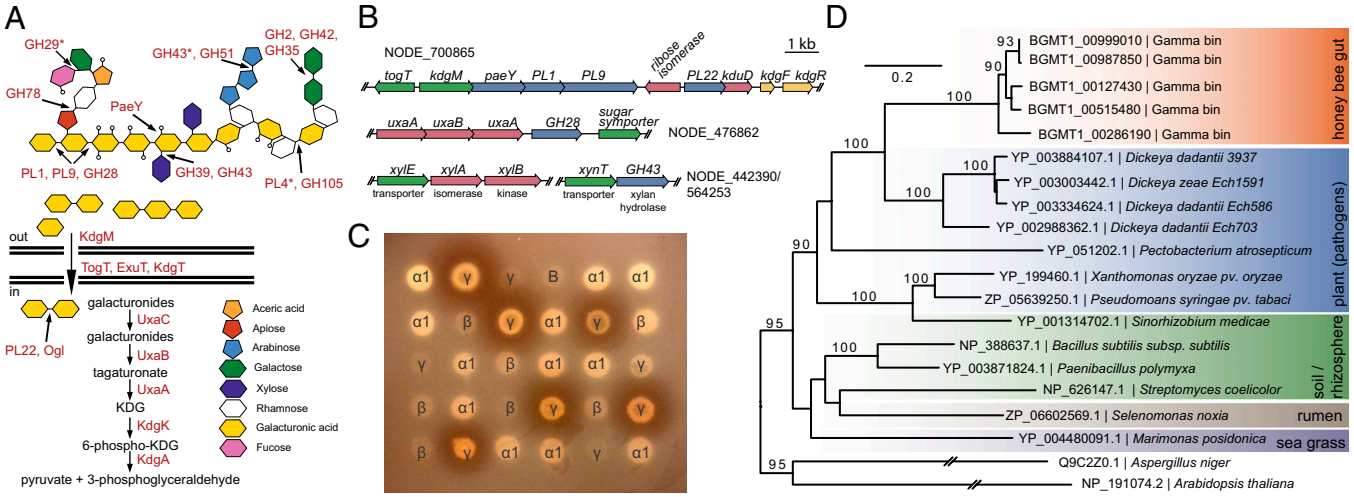

Fig. 3. Pectin degradation by the honey bee gut microbiota. (*A*) Proteins involved in pectin catabolism, for which respective genes were identified, are indicated in red. The illustrated pectin molecule shows the homo- and rhamnogalacturonate backbone with side chains (50) and cleavage sites for identified GHs and PLs. PaeY indicates a pectin acetyl esterase removing ester residues. Except for proteins marked with an asterisk, the encoding genes were present in the Gamma bin. Genes encoding the oligogalacturonide-specific porin KdgM, the inner-membrane transporters TogT, ExuT, and KdgT, and a pathway for galacturonide conversion were also identified. (*B*) Four different loci assigned to the Gamma bin encoding key proteins for pectin breakdown (scaffold numbers are indicated). The first two loci encode proteins involved in metabolism of PGA. The latter two might be involved in breakdown and metabolism of xylose side chains. Blue, green, magenta, and yellow indicate functions associated with glycoside cleavage, transport, cellular catabolism, and regulation, respectively. (*C*) PGA degradation is detected by the formation of clearance zones around bacterial lawns. The species to which each bacterial isolate belongs is indicated. γ, *Gilliamella*; β, *Snodgrassella*; α1, Alpha-1; and B, Bifido. (*D*) Maximum-likelihood protein phylogeny of metagenomic pectate lyases of the PL1 family and their closest homologs found in other bacteria as well as outgroups *Aspergillus niger* and *Arabidopsis thaliana*. Bootstrap values >80 and ecological niche of bacteria are indicated.

PGA derivatives were encoded on scaffolds of the Gamma bin (Fig. 3*A*). Pectin constitutes a major component of plant cell walls, and pectic polysaccharides are involved in the formation of different layers in pollen walls (41). They are particularly prominent in pollen tubes, where they control cell outgrowth during maturation (42). On the basis of current knowledge, pollen breakdown in honey bees seems to be a multifactorial process including the rupture or perforation of the exine layers by osmotic shock and/or pseudogermination (43). These processes may result in the exposure of the intine, a layer similar to primary plant cell walls (42), which subsequently could be digested by corresponding enzymes. Strikingly, histochemical studies previously revealed that pectin of ruptured pollen grains is digested in the honey bee midgut (44), a part of the alimentary tract in which the γ-proteobacterial species are present (12). Phylogenetic analysis indicates that the major pectate lyases (PL1) encoded in the metagenome are most closely related to corresponding enzymes of plant pathogens (Fig. 3*D*). Pectin-degrading enzymes of these bacteria are essential for weakening plant cell walls allowing invasion (45). Similarly, pectin digestion in the honey bee gut might facilitate pollen perforation resulting in the release of its nutrient-rich content. Alternatively, because pectin has been shown to be toxic for honey bees (15), its catabolic breakdown by bacteria could simply permit bees to avoid intoxication.

Experimental Tests of Pectinase Activities. Using an overlay agar containing PGA, we tested bacterial isolates from the bee gut for their ability to degrade pectin. Strikingly, only bacteria belonging to the Gamma bin, and specifically *Gilliamella* isolates, were able to degrade PGA (Fig. 3*C*). PCR screening confirmed that these strains encode the pectate lyases identified in the metagenome (*SI Appendix*, Fig. S7). Interestingly, some *Gilliamella* isolates did not degrade PGA in our experiment and there was an exact correspondence between ability to degrade PGA and presence of pectate lyase on the basis of our PCR screen (*SI Appendix*, Fig. S7). Although these isolates share about 99% identity in 16S rRNA sequences and thus belong to the same species, phylogenetic reconstruction showed that isolates with PGA degradation activity cluster independently from the ones without this capability (*SI Appendix*, Fig. S8). This corroborates our hypothesis that the genetic variation observed within different bacterial species (Fig. 1) might reflect divergent niche adaptation within the gut of honey bees.

Conclusions

In summary, our study provides insights into the functional capabilities of the distinctive microbiota of honey bees. Our results support the view that these bacteria likely play roles in nutrition and pathogen defense in honey bee colonies and provide a large set of candidate roles worthy of further experimentation. The bacteria from the honey bee gut have not been detected outside of bees, but, despite their restricted niche, the degree of diversity within the few members of this bacterial community is remarkable. Although our study considered only a pooled sample from one colony, 16S rRNA-based studies from single individuals from several colonies also show extensive strain variation (13). Coevolution among strains within the microbiota might have resulted in functional divergence, as exemplified by differences between *Gilliamella* strains in ability to degrade pectin. Our study provides a basis for future research aiming at a better understanding of the functional roles of these bacteria for colony health with respect to nutrition and pathogen defense.

Materials and Methods

Metagenome Sequencing. A bacterial DNA sample was prepared from the hindguts of 150 adult worker bees sampled on a single day from a single colony in Tucson AZ. Sequencing of a 300-bp library was conducted with a Illumina Genome Analyzer IIx at the Yale Center for Genome Analysis. The assembly was generated with Velvet v1.0.19 (46) based on ∼40 million quality-filtered read pairs and subjected to a gap-closing analysis. For annotation of the gene content, we used the Integrated Microbial Genomes with Microbiome samples (IMG/M) system (47). Details about DNA preparation and assembly are described in *SI Appendix, SI Materials and Methods*.

Taxonomic Profiling. We determined the community profile in the metagenomic dataset by two independent approaches. First, we taxonomically binned all reads mapping against a set of 31 phylogenetic marker proteins with the BLASTX version of the program MetaPhyler (48). Second, we used the "Phylogenetic Distribution" tool of IMG/M to classify CDSs on the basis of best BLASTP hits with an identity cutoff of 30% (47). Due to the lack of reference data and the deep-branching phylogenetic positions of most bee gut bacterial species, sequences could only be classified on the phylum/class level.

Phylogenetic Analyses. Using Geneious v5.3.6 (Biomatters), protein sequences were aligned with MUSCLE or CLUSTALW, overhanging ends trimmed off, and maximum-likelihood phylogenies inferred with PHYML (WAG model). For the phylogenetic marker gene analyses, all metagenomic CDSs belonging to the corresponding COG function were included, if aligning over >50% of the full-length sequence. We only analyzed phylogenetic marker COGs having a single-copy per genome and a length >300 aa, as defined by IMG/M.

Metagenomic Sequence Binning. To sort out sequences originating from host DNA, we queried all scaffolds against the genome of *A. mellifera* using BLASTN. We then assigned scaffolds to the six different phylogenetic clusters identified in Fig. 1*C* resulting in species or species group bins. We first used BLASTX to query scaffolds against genomes from species (*SI Appendix*, Table S5) related to these six clusters. A scaffold was assigned to a given bin, if >50% of all best BLASTX hits originated from relatives of the same phylogenetic cluster. Unassigned scaffolds ≥2 kb were further analyzed with the compositional-based method ClaMS (49) (for details see *SI Appendix, SI Materials and Methods*).

Minimal Gene Set and Copy Number Analysis. We used the minimal gene set described by ref. 18 mostly containing single-copy housekeeping genes. Protein sequences of each bin were blasted with BLASTP against a subset of this minimal gene set conserved in related species (*SI Appendix*, Table S5). We used relative alignment length, *E* value, and protein identity to identify homologs of this minimal gene set in each bin (for details see *SI Appendix, SI Materials and Methods*).

Analysis of Variable Sites. We selected one copy of the same 27 ribosomal protein-encoding genes from each bin and remapped all quality-filtered reads against them using CLC Genomics Workbench (CLC bio). The same software was then used to detect SNPs in the mapped sequence data (for details see *SI Appendix, SI Materials and Methods*).

COG Function Enrichment Analysis of the Honey Bee Microbiome. We used the "Function Comparison" tool of IMG/M (47) to compare the honey bee microbiome against nine other gut-associated microbiomes available on IMG/M including human, dog, panda, pig, wallaby, termite (21–26), southern pine beetle (IMG/M sample Id: 10345), Asian longhorned beetle (IMG/M sample Id: 10503), and European Sirex wasp (IMG/M sample Id: 391). The relative abundance of COG functions was calculated on the basis of normalized gene counts and expressed as *D*-scores i.e., the standard variation from the null hypothesis (relative gene counts in metagenome A = relative gene counts in metagenome B). For each comparison, an individual *P* value cutoff for significant *D*-scores was determined using a false discovery rate of 0.05.

COG Category Enrichment Analysis of Different Metagenome Bins. To compare the relative abundance of COG category functions between different bins, we applied the method used in the "Function Category Comparison" tool of IMG/M (47). The analysis assesses the statistical significance of the relative frequencies of genes assigned to different functional categories in terms of *D*-rank, representing a normalization ranking of each pairwise comparison. Each bin was compared with the sum of the other bins. *D*-ranks were calculated by adding the *D*-scores (see above) of all protein functions in a given category normalized by the square root of the number of these functions. The null hypothesis of equal probabilities of functional categories was rejected, if *D*-rank values were greater than 2.33 at $P < 0.01$.

Identification of Carbohydrate-Degrading Enzymes. Using HMMERv3 hmmscan, we queried all CDSs of the metagenome against Pfam hidden Markov models

(HMMs) of the PL and GH families present in the CAZy database (40). Proteins with relative hit alignment lengths >0.5 and *E* values $<10^{-3}$ were counted and named according to the CAZy nomenclature. For GH and PL families with no available Pfam HMM, the representative sequences were retrieved from the CAZy database and remaining metagenomic proteins queried against them with BLASTP. Proteins with relative hit alignment lengths >0.5 and *E* values $<10^{-15}$ were counted and classified.

PGA Degradation Experiments. Guts of two *A. mellifera* worker bees, collected from a colony at Yale West Campus, West Haven CT, were dissected, homogenized in PBS, and plated on tryptic soy agar plates containing 5% (vol/vol) sheep blood. Bacteria grew within 3–5 d at 37 °C in 5% (vol/vol) CO_2. Single colonies were randomly picked and passaged to new plates. For PGA degradation experiments, 2 μL of bacterial suspensions in PBS were spotted onto fresh medium plates. After 48 h, 10 mL agar containing 1.5% (wt/vol) PGA (Sigma) in 0.1 M Tris-HCl pH 8.6 was poured on top and incubated for another 24 h under growth conditions. Plates were floated with 1% (wt/vol) cetyltrimethylammonium bromide and incubated at room temperature until clearance zones became visible.

ACKNOWLEDGMENTS. We thank Gloria DeGrandis Hoffman and the Department of Agriculture's Carl Hayden Bee Research Center for allowing us to sample bees, Waldan Kwong for providing methods for culturing members of the bee microbiota, and Kim Hammond for her help with bee colonies and figures. This work was supported by National Science Foundation Award 1046153 (to N.A.M.), Yale University, and the Swiss National Science Foundation and the European Molecular Biology Organization (P.E.).

1. Aizen MA, Garibaldi LA, Cunningham SA, Klein AM (2008) Long-term global trends in crop yield and production reveal no current pollination shortage but increasing pollinator dependency. *Curr Biol* 18:1572–1575.
2. Klein AM, et al. (2007) Importance of pollinators in changing landscapes for world crops. *Proc Biol Sci* 274:303–313.
3. Evans JD, Schwarz RS (2011) Bees brought to their knees: Microbes affecting honey bee health. *Trends Microbiol* 19:614–620.
4. Genersch E (2010) Honey bee pathology: Current threats to honey bees and beekeeping. *Appl Microbiol Biotechnol* 87:87–97.
5. Seeley TD (1985) *Honeybee Ecology: A Study of Adaptation in Social Life* (Princeton Univ Press, Princeton).
6. Toth AL, Robinson GE (2005) Worker nutrition and division of labour in honeybees. *Anim Behav* 69:427–435.
7. Ament SA, Corona M, Pollock HS, Robinson GE (2008) Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proc Natl Acad Sci USA* 105:4226–4231.
8. Haydak MH (1970) Honey bee nutrition. *Annu Rev Entomol* 15:143–156.
9. Cox-Foster DL, et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318:283–287.
10. Koch H, Schmid-Hempel P (2011) Bacterial communities in central European bumblebees: Low diversity and high specificity. *Microb Ecol* 62:121–133.
11. Martinson VG, et al. (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* 20:619–628.
12. Martinson VG, Moy J, Moran NA (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl Environ Microbiol* 78: 2830–2840.
13. Moran NA, Hansen AK, Powell JE, Sabree ZL (2012) Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS ONE* 7: e36393.
14. Koch H, Schmid-Hempel P (2011) Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci USA* 108:19288–19292.
15. Barker RJ (1977) Some carbohydrates found in pollen and pollen substitutes are toxic to honey bees. *J Nutr* 107:1859–1862.
16. Mullin CA, et al. (2010) High levels of miticides and agrochemicals in North American apiaries: Implications for honey bee health. *PLoS ONE* 5:e9754.
17. Babendreier D, Joller D, Romeis J, Bigler F, Widmer F (2007) Bacterial community structures in honeybee intestines and their response to two insecticidal proteins. *FEMS Microbiol Ecol* 59:600–610.
18. Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* 68:518–537.
19. Hansen EE, et al. (2011) Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc Natl Acad Sci USA* 108 (Suppl 1):4599–4606.
20. Martens EC, et al. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* 9:e1001221.
21. Warnecke F, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450:560–565.
22. Pope PB, et al. (2010) Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc Natl Acad Sci USA* 107:14793–14798.
23. Lamendella R, Domingo JWS, Ghosh S, Martinson J, Oerther DB (2011) Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiol* 11:103.
24. Swanson KS, et al. (2011) Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J* 5:639–649.
25. Zhu L, Wu Q, Dai J, Zhang S, Wei F (2011) Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc Natl Acad Sci USA* 108:17714–17719.
26. Gill SR, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359.
27. 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.
28. Görke B, Stülke J (2008) Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nat Rev Microbiol* 6:613–624.
29. Law CJ, Maloney PC, Wang DN (2008) Ins and outs of major facilitator superfamily antiporters. *Annu Rev Microbiol* 62:289–305.
30. Kwakman PHS, et al. (2010) How honey kills bacteria. *FASEB J* 24:2576–2582.
31. González-Teuber M, Heil M (2009) Nectar chemistry is tailored for both attraction of mutualists and protection from exploiters. *Plant Signal Behav* 4:809–813.
32. van Loon LC, Rep M, Pieterse CM (2006) Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol* 44:135–162.
33. Martinez JL, et al. (2009) Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol Rev* 33:430–449.
34. Snodgrass RE (1910) *The Anatomy of the Honey Bee* (U. S. Department of Agriculture, Bureau of Entomology. *Technical Series* 18:1–162.
35. Satchell KJ (2011) Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu Rev Microbiol* 65:71–90.
36. Guarner F, Malagelada JR (2003) Gut flora in health and disease. *Lancet* 361:512–519.
37. Zúñiga M, et al. (2005) Horizontal gene transfer in the molecular evolution of mannose PTS transporters. *Mol Biol Evol* 22:1673–1685.
38. Chalcoff VR, Aizen MA, Galetto L (2006) Nectar concentration and composition of 26 species from the temperate forest of South America. *Ann Bot (Lond)* 97:413–421.
39. Barker RJ, Lehner Y (1974) Acceptance and sustenance value of naturally occurring sugars fed to newly emerged adult workers of honey bees (*Apis-mellifera-L*). *J Exp Zool* 187:277–285.
40. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res* 37(Database issue):D233–D238.
41. Aouali N, Laporte P, Clément C (2001) Pectin secretion and distribution in the anther during pollen development in *Lilium*. *Planta* 213:71–79.
42. Taylor LP, Hepler PK (1997) Pollen germination and tube growth. *Annu Rev Plant Physiol Plant Mol Biol* 48:461–491.
43. Roulston TH, Cane JH (2000) Pollen nutritional content and digestibility for animals. *Plant Syst Evol* 222:187–209.
44. Klungness LM, Peng YS (1984) A histochemical-study of pollen digestion in the alimentary canal of honeybees (*Apis-mellifera L*). *J Insect Physiol* 30:511–521.
45. Lagaert S, Beliën T, Volckaert G (2009) Plant cell walls: Protecting the barrier from degradation by microbial enzymes. *Semin Cell Dev Biol* 20:1064–1073.
46. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.
47. Markowitz VM, Kyrpides NC (2007) Comparative genome analysis in the integrated microbial genomes (IMG) system. *Methods Mol Biol* 395:35–56.
48. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M (2011) Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* 12(Suppl 2):S4.
49. Heath LS, Pati A (2007) Genomic signatures in de Bruijn chains. *Lect N Bioinformat* 4645:216–227.
50. Harholt J, Suttangkakul A, Vibe Scheller H (2010) Biosynthesis of pectin. *Plant Physiol* 153:384–395.