PNAS
www.pnas.org

Supplementary Information for

Division of labor in honey bee gut microbiota for plant polysaccharide digestion

Hao Zheng[a,b], Julie Perreau[b], J. Elijah Powell[b], Benfeng Han[a], Zijing Zhang[a], Waldan K. Kwong[b], Susannah G. Tringe[c], Nancy A. Moran[b]

[a]Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China
[b]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712
[c]Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598

Corresponding authors: Hao Zheng, Nancy A. Moran
Email: hao.zheng@cau.edu.cn; nancy. moran@austin.utexas.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

www.pnas.org/cgi/doi/10.1073/pnas.1916224116

Supplementary Materials and Methods

Isolation and genome sequencing of *Bifidobacterium asteroides* and *Gilliamella apicola*. Pure cultures of *B. asteroides* strains W8111 and W8102, and *G. apicola* strains W8127 and W8131 were isolated from the guts of *Apis mellifera* collected in Jilin, China in July 2018 (Dataset S1). The dissected guts were directly crushed in 19% (vol/vol) glycerol and frozen after sampling. The glycerol stocks were plated on heart infusion agar supplemented with 5% (vol/vol) defibrinated sheep's blood (Solarbio, Beijing, China), incubated at 35°C under a CO_2-enriched atmosphere (5%). Genomic DNA was extracted using the CTAB buffer method as previously described (1). Total genomic DNA was sequenced on the Illumina HiSeq platform from paired-end libraries and then assembled with the SPAdes genome assembler version 3.0 (2). The completeness of genomes was assessed by CheckM (version 1.0.12) (3). Whole-genome average nucleotide Identity was calculated using FastANI (version 2.0) (4). The genomes were annotated with the Prokka software version 1.14.0 (5). The genome assemblies were deposited at DDBJ/EMBL/GenBank, and the accession numbers are all shown in Dataset S1.

Metagenome sequencing. Bees were collected from a single colony in New Haven, CT, USA in April, 2012. Collected bees were put into 50-ml conical tube and placed on ice. The DNA was prepared as described by Engel *et al.* (6) with some modifications. The whole guts of 20 bees were dissected and placed in 1.5-ml tubes with 500 µl $MgSO_4$ (10mM, pH 6.0) on ice. The guts were then homogenized using a plastic pestle for 1 min. Gut homogenates of ~20 bees were pooled in a 50-ml conical tube with 15 ml (final volume) of $MgSO_4$ and were passed through a 100-µm filter. The replaced filter was then rinsed with 10 ml $MgSO_4$ by vortexing vigorously. The rinse was passed through a 100-µm filter again and was pooled to the main filtrate, with a total combined volume of 25 ml. The sample was passed through 20- and 8-µm filters in succession, and was then centrifuged at 500×*g*, 4°C for 2 min. The supernatant was transferred to a new 50-ml tube and centrifuged again at 1 700×*g* for 30 min. The pellet was resuspended in 1.6 ml $MgSO_4$, and 200 µl of the suspension was put into a 1.5-ml tube with 600 µl Percoll solution (80%). Then the samples were centrifuged at 17,000×*g* for 30 min. The interfaces between the upper and lower layers were recovered into a new 1.5 ml tube and the total volume was brought to 1.5 ml with $MgSO_4$. The tubes were centrifuged at 1,600×*g* for 5 min. The pellets were resuspended in 240 µl TE buffer (pH 8.0) with 24 µl RDD buffer and 2.4 µl DNase I (Qiagen) and incubated at 37°C for 30 min to digest free DNA. The DNase was inactivated by adding 24 µl of EDTA (500 mM) and incubating at 25°C for 10 min. The mixtures were centrifuged at 1,600×*g* for 5 min.

DNA was then extracted by resuspending the pellets in 500 µl Buffer AG (Qiagen) and 500 µl phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol; pH 7.9; Ambion) in a 2-ml tube containing 500 µl of sterile zirconia beads Ø 0.1 mm (BioSpec, Bartlesville, OK, USA). Cells were lysed by running MO BIO Vortex Genie® for 3 min. Samples were centrifuged at 12,000×*g* for 5 min. The supernatant was transferred to a clean 1.5-ml tube and mixed with 20 µl Proteinase K (Qiagen), and incubated at 56°C for 30 min. After cooling at room temperature for 3 min, 5 µl RNase A (Qiagen) was added, and incubated at 37°C for 30min. Samples were transferred to phase lock gel with 400 µl Phenol:Chloroform:Isoamyl Alcohol, and centrifuged at 12,000×*g* for 5 min. The supernatants were transferred to new 1.5-ml tubes. 50 µl sodium acetate (pH 5.2–5.5) and 500 µl isopropanol was added, then mixed well. After centrifuging at 17,000×*g* for 30 min, the pellets were washed twice with 70% ethanol. The DNA pellets were then dissolved in TE buffer (pH 8.0). DNA samples were sent to the DOE Joint Genome Institute for metagenome sequencing. Two libraries were constructed: a shotgun library with an insert size of 300 bp and a mate-pair library with 4 kb spacing. Both libraries were sequenced on Illumina HiSeq 2000 platform (2×150 paired end), for a total of ca. 111 Gbp of nucleotide sequence data (65 and 46 Gbp from the shotgun and mate-pair libraries, respectively). The metagenomic data were assembled with the JGI/LANL pipeline using the MeGAMerge tool (7). The assembled sequences were submitted to the Integrated Microbial Genomes and Microbiomes (IMG/M) system for comparative analysis, and the annotation was performed using the DOE-JGI Metagenome Annotation Pipeline (8). The JGI IMG accession number is 3300000333.

Phylogenomic analyses and GH and PL family annotation. Phylogenomic trees for *Gilliamella* and *Bifidobacterium* strains were generated from the concatenation of amino acid sequences of 106 and 101 bacterial core genes, respectively. TIGRFAM Hidden Markov Model (HMM) profiles from the bacterial core gene set (GenProp0799) of each protein were used to query the genome database (9), and sequences for the top hits were pulled using an in-house script. Sequences were aligned with MAFFT (10), manually inspected, and trimmed using a BLOSUM62 matrix in BMGE (11). Maximum likelihood trees were constructed using the GTRCAT model in RAxML (12). To describe the distribution of GH and PL families across bee gut bacterial genomes, CAZymes of all bee gut bacterial genomes were annotated using the dbCAN2 server with the HMMER search tool, which use the threshold of *E*-value <1e–15 and coverage >0.35 (Dataset S2) (13). The taxon affiliations of the metagenomic CAZyme amino acid sequences were analyzed through the IMG annotation pipeline, which queries them against the NCBI non-redundant protein database with BLASTP (8).

Bacterial growth and *in vitro* / *in vivo* analyses of gene expression. *B. asteroides* strain W8111 and W8102 were routinely grown on heart infusion agar plus 5% sheep blood in an anaerobic chamber (2% H_2, 5% CO_2, and 93% N_2; Coy Laboratory Products, Grass Lake, MI, USA). The ability of strain W8111 and W8102 to grow on pure carbohydrates was measured by cultivation in modified MRS liquid medium (14) composed of (per liter) 2 g peptone, 1 g yeast extract, 1 g polysorbate 80, 0.4 g NH_4Cl, 0.2 g $MgSO_4.7H_2O$, 0.07 g $MnCl_2.4H_2O$, 2 g KH_2PO_4, 0.4 g L-cysteine hydrochloride, 0.1 g pyridoxine hydrochloride, 0.5 g pantothenic acid, 0.1 g inositol, 0.01 g aminobenzoic acid, 0.02 g adenine. The medium was then loaded with sterilized carbohydrate stocks of arabinan (CAS: 11078-27-6; Megazyme, Wicklow, Ireland), galactan (CAS: 9037-55-2; Megazyme), beta-glucan (CAS: 9041-22-9; Megazyme), and xyloglucan (CAS: 37294-28-3; Megazyme) at different final concentrations according to the manuals. D-glucose (10 mM) was used as positive control. Cultures were inoculated with single colonies growing on heart infusion agar plates, then incubated in the anaerobic chamber. Growth was determined spectrophotometrically by measuring the optical density at 600 nm after 48 h of incubation. Bacterial cells were then used for RNA extraction.

Microbiota-free and gnotobiotic bees were obtained as described by Zheng *et al.* (14) with modifications. Late-stage pupae were removed manually from brood frames, and placed in sterile plastic bins. The pupae emerged in an incubator at 35°C, humidity 50%. Newly emerged bees were kept in axenic cup cages with sterilized sucrose syrup (0.5 M); however, pollen was withheld, in order to ensure that the bees did not obtain any plant glycans. *B. asteroides* strains W8111 and W8102, and *G. apicola* strains W8127 and W8131 were prepared in 20% glycerol stock, frozen at –80°C. Bacterial cells from stocks were resuspended in 1x PBS at a final OD_{600nm} of 1, and then supplemented with equal volume of sucrose solution (50%, wt/vol). For each setup, 20–25 microbiota-free bees were placed into one cup cage, and the bees were feeding on the bacterial suspensions for 24h. Colonization levels were determined by colony-forming units from dissected guts, as described by Kwong *et al.* (15). Then mono-inoculated bees were provided sucrose (0.5 M) with or without sterile pollen, or one of the pure hemicelluloses (arabinan, galactan, β-glucan, xyloglucan) or polygalacturonic acid (CAS: 9049-37-0; Megazyme) dissolved in sucrose syrup. Bee guts were dissected 24 h after feeding with the substrates, and whole guts were harvested for RNA extraction.

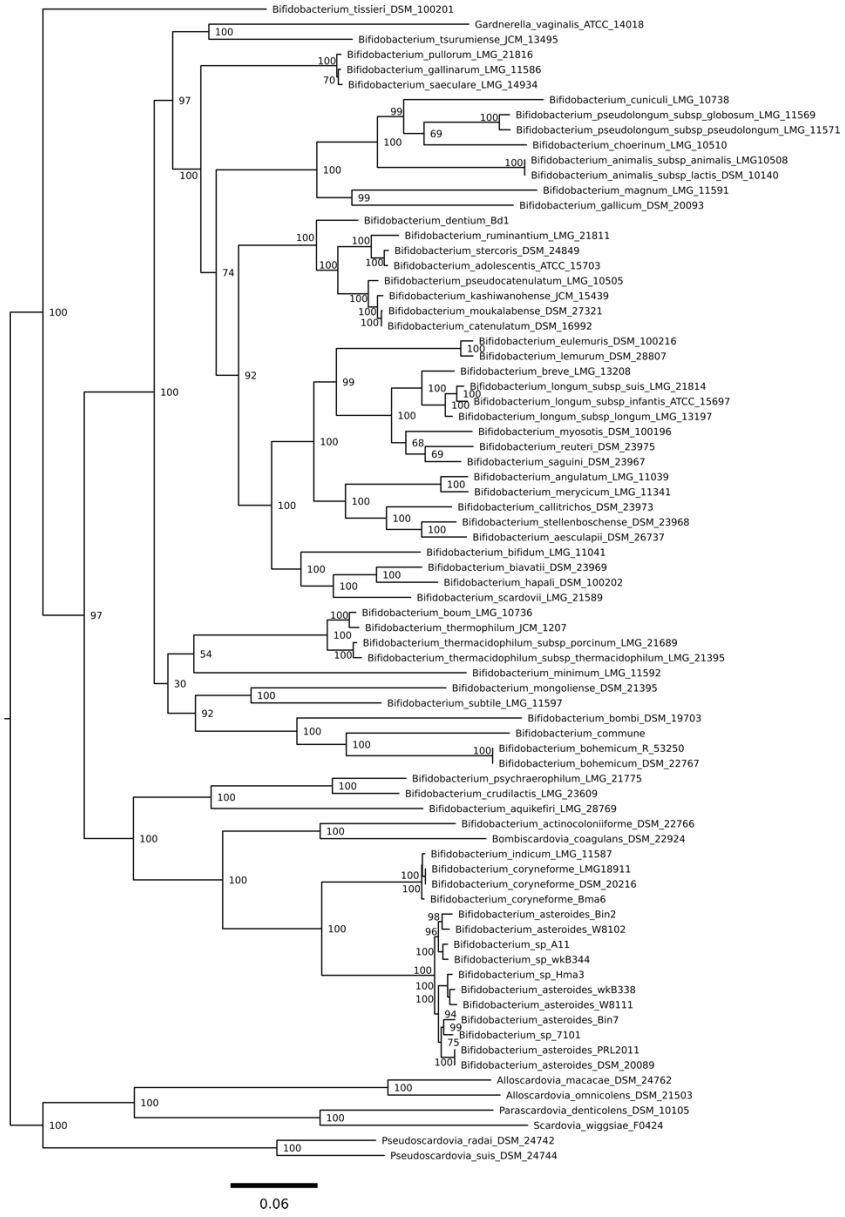
RNA of pure bacterial cultures and of whole bee guts was extracted using the Quick-RNA MiniPrep kit (Zymo Research, Irvine, CA, USA). cDNA was synthesized from 1 ng of each RNA sample using the HiScript® III 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China) with random hexamer primers from the kit. Quantitative PCR was performed using the ChamQ™ Universal SYBR® qPCR Master Mix (Vazyme) on the QuantStudio™ 1 real-time PCR system (Applied Biosystems) in a standard 96-well block (20 µl reactions; 40 cycles of denaturation at 95°C for 15 seconds, annealing/extension and plate read at 60°C for 60 seconds). The primers are listed in Table S2, and *recA* served as a reference gene. Relative expression was analyzed using the $2^{-\Delta\Delta CT}$ method (16). Three technical replicates for each sample were performed on the same plate. RNA samples that were not reverse transcribed served as negative controls.

Metabolomics. The concentrations of galacturonic-acid in whole gut of bees mono-colonized with *G. apicola* strains W8127 and W8131 were identified using the D-Glucuronic/D-Galacturonic Acid Assay Kit (Megazyme). Metabolomic profiles of gut homogenates of bees mono-colonized with *B. asteroides* strains W8111 and W8102 were identified as described in Zheng et al. (14) with some modifications. As described above, microbiota-free bees were mono-colonized with *B. asteroides* strain W8111 and W8102, respectively. The mono-colonized bees were then fed with arabinan, galactan, or xyloglucan for 24h. Individual bee guts were homogenized in 50μl distilled water. The homogenates were then centrifuged (2,500×*g*, 5 min) and filtered through 0.22-μm membrane. Six biological replicates of each group were shipped on dry ice to Novogene Cooperation (Beijing, China) for metabolomics analysis.

The gut homogenate supernatant was resuspended with prechilled 80% methanol and 0.1% formic acid. The samples were incubated on ice for 5 min and then were centrifuged at 12,000×*g*, 4°C for 5 min. The supernatant was diluted to final concentration containing 60% methanol by LC-MS grade water. The samples were subsequently transferred to a fresh Eppendorf tube with 0.22 μm filter and then were centrifuged at 15,000×*g*, 4°C for 10 min. Finally, the filtrate was injected into the LC-MS/MS system analysis. LC-MS/MS analyses were performed using a Vanquish UHPLC system (Thermo Fisher) coupled with an Orbitrap Q Exactive™ HF mass spectrometer (Thermo Fisher). Samples were injected onto an Hyperil Gold column (100×2.1 mm, 1.9 μm) using a 16-min linear gradient at a flow rate of 0.2 mL/min. The eluents for the positive polarity mode were eluent A (0.1% FA in Water) and eluent B (Methanol). The eluents for the negative polarity mode were eluent A (5 mM ammonium acetate, pH 9.0) and eluent B (Methanol). The solvent gradient was set as follows: 2% B, 1.5 min; 2-100% B, 12.0 min; 100% B, 14.0 min;100-2% B, 14.1 min; 2% B, 16 min. Q Exactive™ HF mass spectrometer was operated in positive/negative polarity mode with spray voltage of 3.2 kV, capillary temperature of 320°C, sheath gas flow rate of 35 arb and aux gas flow rate of 10 arb.

The raw data files generated by UHPLC-MS/MS were processed using the Compound Discoverer 3.0 (Thermo Fisher) to perform peak alignment, peak picking, and quantitation for each metabolite. The main parameters were set as follows: retention time tolerance, 0.2 minutes; actual mass tolerance, 5 ppm; signal intensity tolerance, 30%; signal/noise ratio, 3; minimum intensity, 100000. Peak intensities were normalized to the total spectral intensity. The normalized data was used to predict the molecular formula based on additive ions, molecular ion peaks, and fragment ions. Peaks were then matched with the mzCloud (https://www.mzcloud.org/) and ChemSpider (http://www.chemspider.com/) database to obtained the accurate qualitative and relative quantitative results. Metabolomics data analysis was performed using MetaboAnalyst 4.0 (17) as described in Zheng et al. (14). Briefly, raw peak intensity data were filtered using interquartile range approach and normalized with the options: normalization by the sum, logarithm transformation, and auto scaling. The fold change (W8111/W8102) of mono-colonized bees was calculated based on the normalized data and the significance were calculated using the t-test. The dataset was then subjected to a partial least square discriminant analysis (PLS-DA). 95% confidence interval ellipses (based on the standard deviation for each PC) were used to visually support the significance of each dataset cluster.

A



B

	Bin2	W8102	A11	wkB344	Hma3	wkB338	W8111	Bin7	7101	PRL2011	DSM20089
Bin2	/										
W8102	95.5	/									
A11	94.1	94.0	/								
wkB344	93.8	93.9	96.4	/							
Hma3	91.4	90.8	91.1	91.1	/						
wkB338	91.3	90.9	91.1	91.1		/					
W8111	91.0	91.0	90.6	90.5	97.2	96.8	/				
Bin7	92.7	92.3	92.0	92.0	92.0	91.9	91.6	/			
7101	91.9	91.5	91.9	91.8	91.5	91.5	91.0	93.4	/		
PRL2011	91.7	91.4	91.5	91.4	92.1	92.0	91.7	92.3	92.1	/	
DSM20089	91.5	91.5	91.3	91.1	91.7	91.6	91.7	92.3	91.8	100.0	/

Fig. S1. (A) Full phylogenetic tree of the *Bifidobacterium* genus based on the whole-genome sequences using the maximum-likelihood algorithm. (B) Average nucleotide identity values of *B. asteroides* strains.

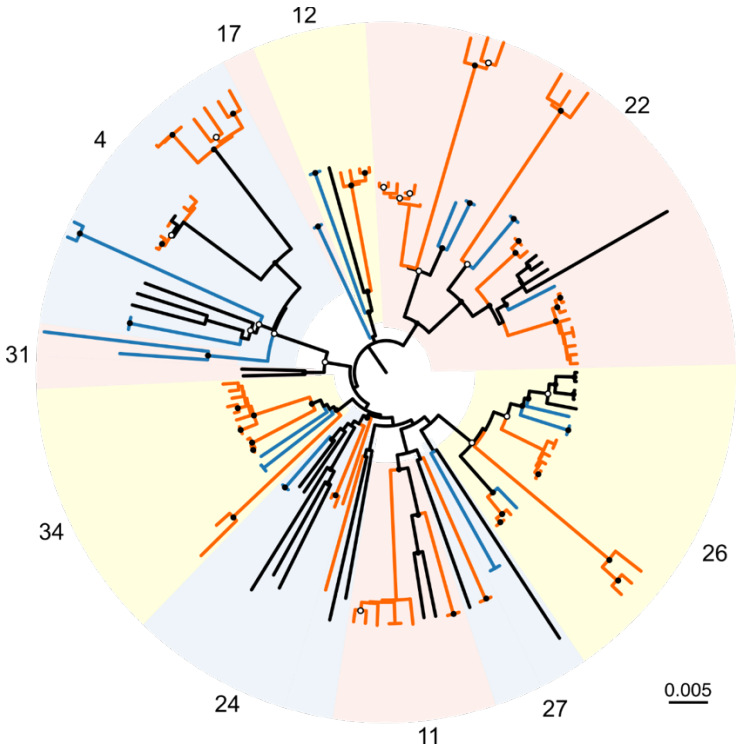


Fig. S2. Phylogenetic tree of genes belonging to GH43 family encoded in genomes of *Bifidobacterium* species. Sequences of strains from honey bees (orange) and from bumble bees (blue) are indicated as colored branches. Black branches are selected reference sequences. The number of subfamilies is shown next to the clusters. Circles indicate node bootstrap support (○>85%; ●=100%, 1,000 replicates)

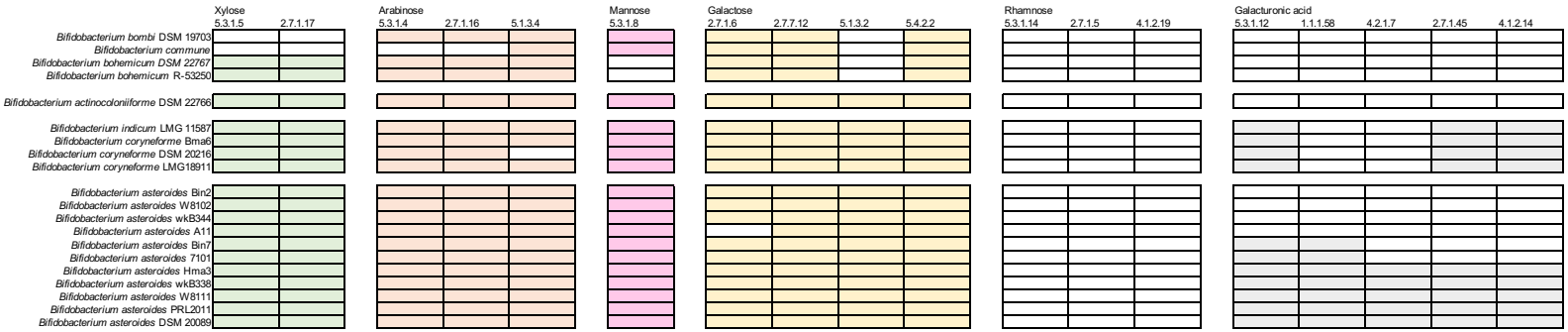


Fig. S3. Presence of the genes related to the metabolism of five monosaccharides and galacturonic acid in *Bifidobacterium* species isolated from honey bee guts. The Enzyme Commission (EC) numbers of the genes are indicated on top of each column. Colored boxes indicate gene presence, and white boxes indicate gene absence.

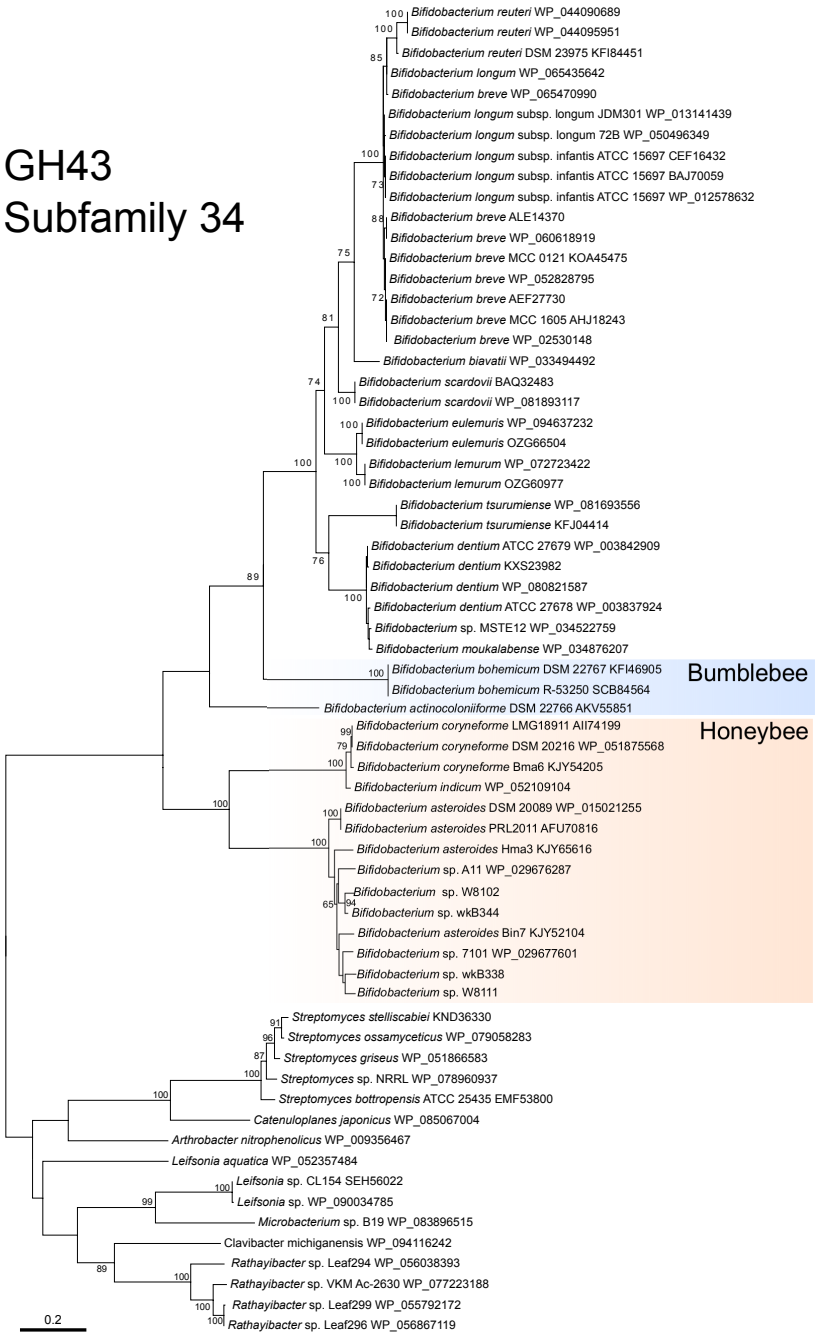


Fig. S4. Phylogenetic analysis of GH43 subfamily 34. The tree was built using the maximum-likelihood algorithm based on amino acid sequences (69 positions). Bootstrap values are shown next to the branches.

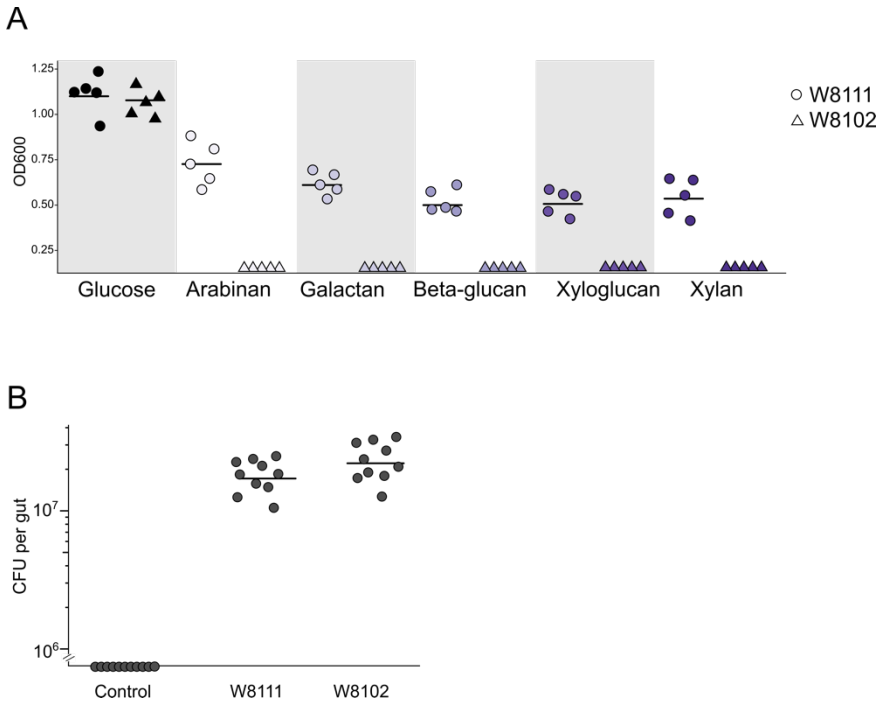


Fig. S5. (A) Growth of *Bifidobacterium asteroides* W8111 and W8102 on different polysaccharides as sole carbon source. (B) Mono-colonizations of microbiota-free bees with *B. asteroides* W8111 and W8102. After 3 days of colonization, colony forming units (CFUs) in whole gut homogenates were determined by plate count and calculated per gut. Control bees were fed with sucrose syrup amended with the same volume of 1×PBS instead of bacterial cell suspensions.

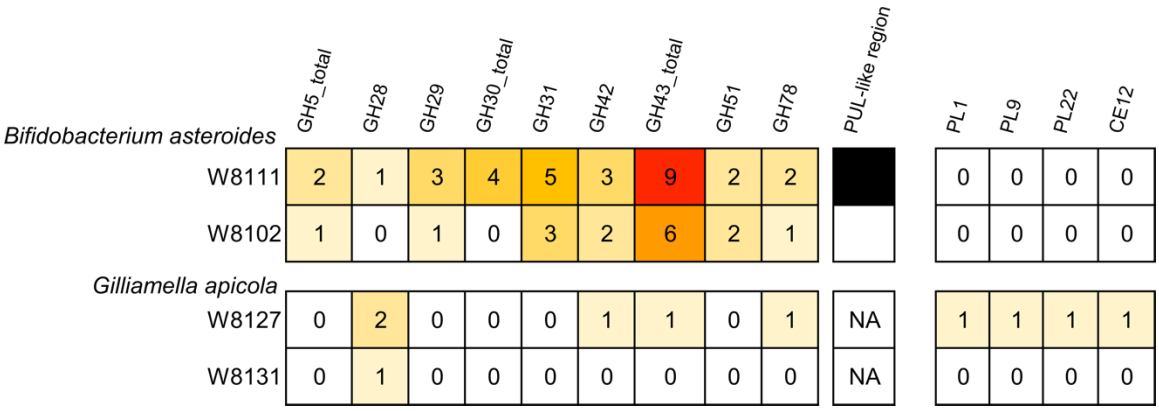


Fig. S6. The abundances of glycoside hydrolase (GH) and polysaccharide lyase (PL) families in the genomes of Bifidobacterium asteroides strain W8102 and W8111, and of *Gilliamella apicola* strain W8127 and W8131, which are used in colonization experiments. Strain W8111 possesses the PUL-like region (solid box), which is absent in strain W8102 (open box).

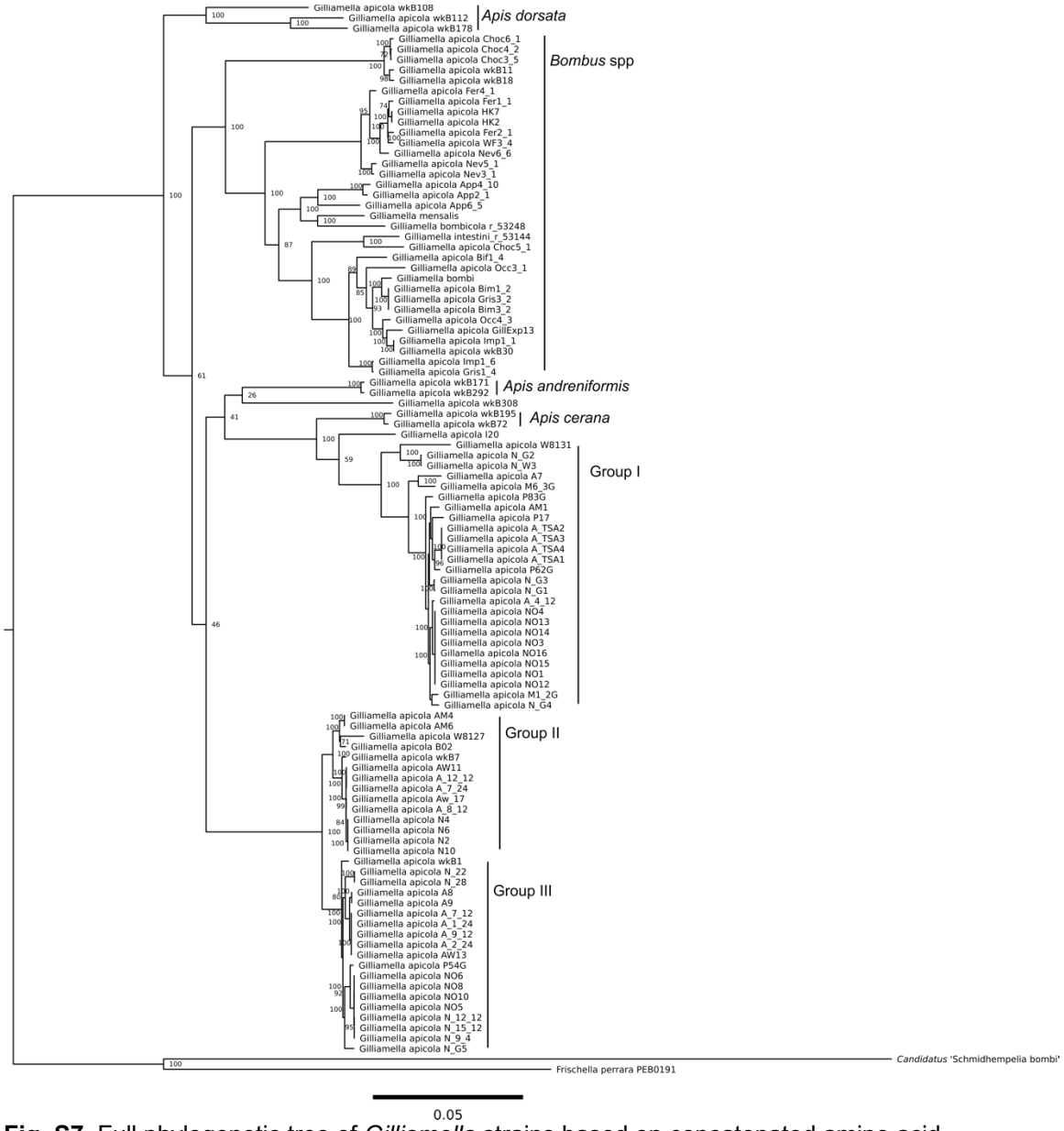


Fig. S7. Full phylogenetic tree of *Gilliamella* strains based on concatenated amino acid sequences. The bootstrap values are shown next to the nodes.

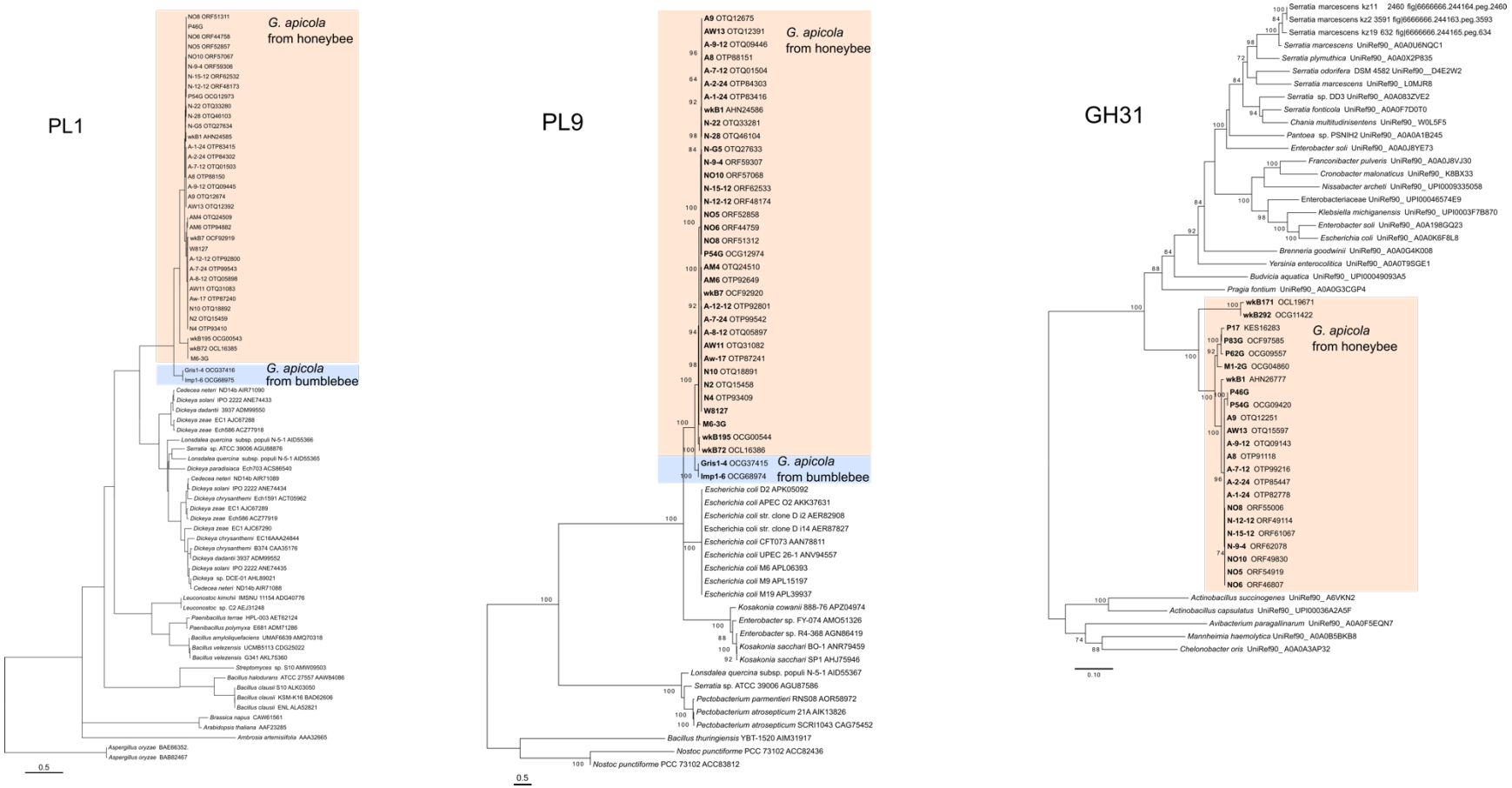


Fig. S8. Maximum-likelihood trees of the amino acid sequences of PL1, PL9, and GH31 from *Gilliamella* strains and from bacteria representing the top blast-hits in public databases.

Table S1. Comparison of the two metagenomic datasets for gut microbiota of Apis mellifera.

	This study	Engel et al., 2012
GOLD ID in IMG Database	Gs0067856	Gp0053272
Collection location	New Haven, CT, USA	Tucson, AZ, USA
Collection date	April, 2012	October, 2010
Sequencing platform	Illumina HiSeq	Illumina GAIIx
Number of raw data bases	111 Gbp	8 Gbp
Number of contigs	407,877	54,505
Number of protein-coding genes	614,276	125,637

Table S2. Primers used for qPCR analysis in this study.

Target strain	Target gene	Forward/Reverse	Sequence (5'–3')
Bifidobacterium asteroides W8111	GH31	F	TTCTGGCCAGATCCCGTCAGCA
		R	TCTTCGCCGTGGTCTGCTCGTA
	GH43-12	F	CCGACTTCCGCTGTACCTTC
		R	CAGGTTCCACCGACAGATCC
	GH42	F	TCGTTGCCAGCCTGATTCAT
		R	CGGTGAACGATTTCGGCAAG
	GH43-4-1	F	CAGGACGATGACCGAACGAT
		R	GCGCCTGACGTCTACTACAA
	GH43-26	F	TGCTCTCCAAGCCCGAATAC
		R	GAGGCCGAGTAGGTCAGGTA
	GH43-4-2	F	CGATGCGCTCTTGTTCGATG
		R	CCTAGCGACCAGAGGTAGGT
	GH43-27	F	CTCTACAGCGACGGGTCCAC
		R	TTCTGGAGAGACCAGGCTGA
	recA	F	ATGTCGAATTCGGCCACCTT
		R	CATCCGCCGTATCCAGACTC
Gilliamella apicola W8127	PL1	F	CCATGTTAATTGGGCACAGTG
		R	TCGTGGAGCTCGCTGAGTAA
	PL9	F	CGGACCTGAAGTGGATGGTA
		R	GCTGGTTGCGGTTCACTTCT
	PL22	F	GGTTGGTCGATCGTGATGGTA
		R	AAGCAAGTCCTGATCCATCTG
	GH28	F	CAGTCCACCCAATGGAGTCA
		R	ACCCGCTTTAATTGCGATGC
	recA	F	CGGCTAAACTAGGTGTGCAAG
		R	TGCCGCTACTGAGTCAACAA

Dataset S1 (separate file). The list of genomes of bacterial isolates from the gut of honey bee and bumble bee.

Dataset S2 (separate file). CAZyme annotation of the bacterial genomes isolated from the guts of honey and bumble bees.

Dataset S3 (separate file). The list of genes for amino acid synthesis in the genome of the bee gut isolates.

References

1. H. Zheng, C. Dietrich, R. Radek, A. Brune, *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. nov.(phylum *Elusimicrobia*)–an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ. Microbiol.* **18**, 191-204 (2016).
2. A. Bankevich*, et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
3. D. H. Parks*, et al.*, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055 (2015).
4. C. Jain*, et al.*, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
5. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* **30**, 2068-2069 (2014).
6. P. Engel, V. G. Martinson, N. A. Moran, Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. USA.* **109**, 11002-11007 (2012).
7. M. Scholz, C. C. Lo, P. S. Chain, Improved assemblies using a source-agnostic pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of contigs. *Sci. Rep.* **4**, 6480 (2014).
8. I. A. Chen*, et al.*, IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**, D666-D677 (2019).
9. D. H. Haft*, et al.*, TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res.* **41**, D387-D395 (2013).
10. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
11. A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
12. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* **30**, 1312-1313 (2014).
13. H. Zhang*, et al.*, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
14. H. Zheng*, et al.*, Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4775-4780 (2017).
15. W. K. Kwong, P. Engel, H. Koch, N. A. Moran, Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11509–11514 (2014).
16. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2–ΔΔCT method. *Methods.* **25**, 402–408 (2001).
17. J. Chong*, et al.*, MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486-W494 (2018).