AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Deep Divergence and Genomic Diversification of Gut Symbionts of Neotropical Stingless Bees

Garance Sarton-Lohéac,[a] Carlos Gustavo Nunes da Silva,[e] Florent Mazel,[b] Gilles Baud,[a] Vincent de Bakker,[a] Sudip Das,[a] Yassine El Chazli,[a] Kirsten Ellegaard,[a] Marc Garcia-Garcera,[d] Natasha Glover,[c] Joanito Liberti,[a,b] Lorena Nacif Marçal,[e] Aiswarya Prasad,[a] Vincent Somerville,[a] SAGE class 2019-2020 and 2020-2021, Germán Bonilla-Rosso,[a] Philipp Engel[a]

[a]Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland
[b]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[c]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[d]Société des produits Nestlé, Lausanne, Switzerland
[e]Department of Morphology, Instituto de Ciências Biológicas, Federal University of Amazonas, Manaus, Brazil

Garance Sarton-Lohéac and Carlos Gustavo Nunes da Silva contributed equally to this work. The author order was decided on mutual agreement. Garance Sarton-Lohéac led the analysis and prepared the manuscript; Carlos Gustavo Nunes da Silva took the lead in sample and data acquisition.

ABSTRACT Social bees harbor conserved gut microbiotas that may have been acquired in a common ancestor of social bees and subsequently codiversified with their hosts. However, most of this knowledge is based on studies on the gut microbiotas of honey bees and bumblebees. Much less is known about the gut microbiotas of the third and most diverse group of social bees, the stingless bees. Specifically, the absence of genomic data from their microbiotas presents an important knowledge gap in understanding the evolution and functional diversity of the social bee microbiota. Here, we combined community profiling with culturing and genome sequencing of gut bacteria from six neotropical stingless bee species from Brazil. Phylogenomic analyses show that most stingless bee gut isolates form deep-branching sister clades of core members of the honey bee and bumblebee gut microbiota with conserved functional capabilities, confirming the common ancestry and ecology of their microbiota. However, our bacterial phylogenies were not congruent with those of the host, indicating that the evolution of the social bee gut microbiota was not driven by strict codiversification but included host switches and independent symbiont gain and losses. Finally, as reported for the honey bee and bumblebee microbiotas, we found substantial genomic divergence among strains of stingless bee gut bacteria, suggesting adaptation to different host species and glycan niches. Our study offers first insights into the genomic diversity of the stingless bee microbiota and highlights the need for broader samplings to understand the evolution of the social bee gut microbiota.

IMPORTANCE Stingless bees are the most diverse group of the corbiculate bees and represent important pollinator species throughout the tropics and subtropics. They harbor specialized microbial communities in their gut that are related to those found in honey bees and bumblebees and that are likely important for bee health. Few bacteria have been cultured from the gut of stingless bees, which has prevented characterization of their genomic diversity and functional potential. Here, we established cultures of major members of the gut microbiotas of six stingless bee species and sequenced their genomes. We found that most stingless bee isolates belong to novel bacterial species distantly related to those found in honey bees and bumblebees and encoding similar functional capabilities. Our study offers a new perspective on the evolution of the social bee gut microbiota and presents a basis for characterizing the symbiotic relationships between gut bacteria and stingless bees.

KEYWORDS bacteria, diversification, genome, gut microbiome, insects, phylogeny, stingless bee, symbiosis

Editor Joerg Graf, University of Connecticut

Address correspondence to Philipp Engel, philipp.engel@unil.ch.

The authors declare no conflict of interest.

Received 20 December 2022
Accepted 23 February 2023
Published 20 March 2023

The eusocial corbiculate bees (referred to here as social bees) comprise more than 700 species distributed in three distinct tribes: honey bees (Apini), bumblebees (Bombini), and stingless bees (Meliponini). Stingless bees and bumblebees form a monophyletic clade, which is sister to the honey bees, with whom they shared a common ancestor 80 to 100 million years ago (1–3). Like mammals, social bees harbor dense and specialized bacterial communities in their gut that affect bee health and behavior (4–12). The composition of the gut microbiota of social bees is relatively simple, typically consisting of <10 bacterial phylotypes, i.e., sequence clusters sharing >97% identity in the 16S rRNA gene (13–20). Five of these phylotypes (*Snodgrassella*, *Gilliamella*, *Bombilactobacillus* Firm-4, *Lactobacillus* Firm-5, and *Bifidobacterium*) have been referred to as the core gut microbiota of the social bees (20), because they are prevalent and abundant across honey bees, bumblebees, and stingless bees. Most members of the bee gut microbiota are culturable, and gnotobiotic bees can be generated for several species (5, 21–23). Together, these distinctive characteristics make the social bee microbiota a versatile model system for studying the evolution and ecology of host-associated microbial communities. Moreover, social bees are important pollinators that suffer from severe population declines (24, 25) which makes studies of their microbiota relevant in their own right.

Most of what is currently known about the gut microbiota of social bees stems from studies on honey bees and bumblebees. Genomic and experimental approaches have revealed that their gut bacteria are usually saccharolytic fermenters that utilize plant glycans derived from the pollen and nectar/honey diet of the host (22, 23, 26–32). Further, it has been shown that the core members of the honey bee and bumblebee gut microbiotas have substantially diversified (27–30, 33–36). They consist of divergent sublineages (or species) and exhibit extensive strain-level diversity and gene content variation. Most sublineages are host specific (23, 27, 33, 36), and their phylogenetic relationships are to some extent congruent with the phylogeny of the host (16, 33). Therefore, it has been suggested that the core members of the microbiota were acquired in a common ancestor of the social bees (20) and possibly codiversified with the host (16, 33). In addition, studies in the Western honey bee (*Apis mellifera*) have shown that the diversification of the bee gut microbiota was also driven by adaptation to different spatial and metabolic niches within the gut (27, 30, 33–35). For example, strains of closely related sublineages of *Lactobacillus* Firm-5 and *Bifidobacterium* can coexist in individual bees. They carry distinct gene sets for the breakdown and utilization of pollen-derived carbohydrates, which allows them to partition the available dietary glycan niches in the gut (27, 30, 34).

In contrast to the microbiotas of honey bees and bumblebees, much less is known about the gut microbiota of the third group of social bees, the stingless bees (Meliponini). Previous studies have focused on determining the taxonomic composition of the gut microbiota of these bees using 16S rRNA gene sequencing (15, 17, 20, 37–39). However, only a few bacteria have been cultured from stingless bees (40–42), and except for two strains of *Bombilactobacillus* Firm-4 recently isolated from bees from Australia (43), no genomic data are currently available for core members of the gut microbiota of stingless bees.

With >500 described species, stingless bees present the largest and most diverse group of the social bees (44, 45). They are naturally distributed throughout the tropical and subtropical regions of Africa, Asia, Australia, and the Americas and exhibit great variation in morphology, diet, foraging range, social structure, and nesting habits (44, 45). As host phylogeny and ecology are both key determinants of gut microbiota composition (46–51), we hypothesize that genomic studies on bacterial isolates will help us to understand the functional diversity of gut bacteria of stingless bees and provide novel insights into the evolution of these bacteria across social bees, specifically in respect of the possible codiversification with the host.

To address these questions, we looked at the gut microbiotas of six neotropical species of stingless bees from Brazil: *Frieseomelitta varia* (Fv), *Scaptotrigona polysticta* (Sp), *Melipona fuliginosa* (Mf), *Melipona interrupta* (Mi), *Melipona seminigra* (Ms), and *Melipona*

lateralis (Ml). We determined the composition of the gut microbiota of these bees using 16S rRNA gene sequencing, established a comprehensive culture collection of bacterial isolates, and conducted genome sequencing and comparative genomics to determine the phylogenetic placement, genomic diversity, and functional capabilities of these bacteria relative to those previously isolated from honey bees and bumblebees.

RESULTS

Six neotropical stingless bee species from Brazil harbor distinct gut microbiotas dominated by nine bacterial families. We sampled three colonies of six stingless bee species (Fv, Sp, Mf, Mi, Ms, and Ml) from a meliponary located in the Amazonian rainforest near Manaus (see Table S1 in the supplemental material). For each colony, we pooled the guts of 15 to 60 worker bees (depending on the size of the bee species; see Materials and Methods) before DNA isolation. The V4 region of the 16S rRNA gene was amplified and sequenced with the Illumina MiSeq 2×250-bp platform, resulting in a median depth of 64,713 (52,479 to 95,774) reads per sample. In total, we identified 277 amplicon sequence variants (ASVs; 29 to 63 ASVs per sample), which belonged to 36 different bacterial families (Table S2). Despite this diversity, only nine families dominated the samples, together representing 97% of all quality-filtered reads (93 to 99% of the reads per sample): *Acetobacteraceae*, *Bifidobacteriaceae*, *Enterobacteriaceae*, *Lactobacillaceae*, *Neisseriaceae*, *Orbaceae*, *Prevotellaceae*, *Streptococcaceae*, and *Veillonellaceae* (Fig. 1A).

While *Lactobacillaceae* and *Bifidobacteriaceae* were abundant across all samples, there were clear differences in the distribution of some of the other bacterial families (Fig. 1A). *Neisseriaceae* were abundant in the samples of Sp and Fv but were detected in only three of 12 samples from the genus *Melipona* (Fig. 1A; Table S2). In contrast, *Acetobacteraceae* and *Streptococcaceae* were present in most *Melipona* samples but rare across samples from Fv and Sp. *Orbaceae* and *Enterobacteriaceae* were mostly detected in the three Fv samples. Intriguingly, a single *Enterobacteriaceae* ASV constituted the most abundant community member in this bee species (21 to 36% of the reads per sample). According to these compositional differences, nonmetric multidimensional scaling (NMDS) separated the samples into three distinct clusters: two clusters comprised all samples from Fv and Sp, and the third cluster comprised all samples from the four *Melipona* species (Mf, Mi, Ms, and Ml) (Fig. S1A).

Related stingless bee species have overlapping community profiles. We compared our results to a previously published amplicon sequencing data set from stingless bees (20) to assess the similarity of the communities to those of other stingless bee species. After discarding samples with <5,000 reads to control for variation in sequencing depths, our data set comprised 135 samples from 19 different host species and three different countries. We detected 688 ASVs in total with a median of 18 ASVs per sample (3 to 63 ASVs) (Table S3), spanning 53 bacterial families. Overall, the taxonomic patterns were similar across the analyzed bee species. Apart from one sample from *Tetragonula fuscobalteata*, for which 99% of the reads belonged to a single *Weeksellaceae* ASV, the nine families dominating in the six bee species from our study were also abundant in the microbiotas of the samples from the previous study and represented 34% to 100% of the total number of reads (Fig. S1B; Table S3).

NMDS based on ASV relative abundances separated the samples by location (i.e., samples from Brazil were different from those from Australia and Malaysia) (permutational multivariate analysis of variance [PERMANOVA]; location pseudo-$F = 20.17$, $P = 0.001$) and by bee genus (PERMANOVA; genus pseudo-$F = 10.49$, $P = 0.001$), although taxonomy and geography are not independent (Fig. 1B). In contrast, there was only weak clustering at the species level (species pseudo-$F = 1.96$, $P = 0.03$). Notably, only 30% (206 of 688) of all ASVs were shared across host species (i.e., 70% of all ASVs are found in only one species) and most of them (83.4%) only between 2 and 5 species (Fig. S1C). However, the shared ASVs belonged to the nine predominant bacterial families and represented a large fraction of the total number of reads per sample (53.3 to 99.7%; except that for the samples from Fv and *Partamona helleri*, the fraction was 10.6 to 15% of the reads) (Fig. S1D). In particular, bees sampled in the same country or belonging to the same bee genus shared

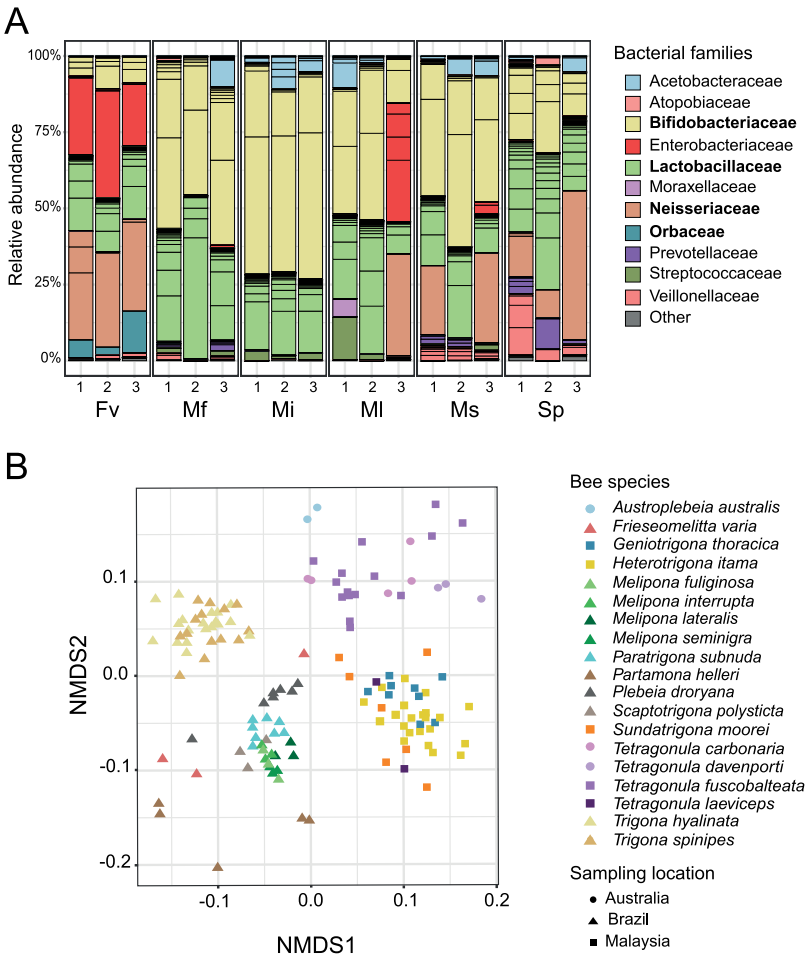

FIG 1 Community analysis of the gut microbiotas of stingless bees. (A) 16S rRNA gene-based community profiles of the gut microbiotas of three colonies of six stingless bee species collected in Brazil. Fv, *Frieseomelitta varia*; Ms, *Melipona seminigra*; Ml, *Melipona lateralis*; Mf, *Melipona fuliginosa*; Mi, *Melipona interrupta*; Sp, *Scaptotrigona polysticta*. Relative abundance of ASVs is shown. ASVs are ordered and colored at the family level (see the key); families of core members are in bold. ASVs with <1% relative abundances are summed up as "Other" and shown in gray. (B) NMDS based on ASV relative abundance (Bray-Curtis dissimilarity) across 136 16S rRNA gene amplicon sequence samples, including data from a previous study (20) and our study.

the same ASVs, explaining the clustering of these samples in the NMDS analysis. Together, these results show that despite the large variability observed, the gut microbiotas of most stingless bee species are dominated by a few bacterial families and that bee species of the same genus, or with overlapping geographic distribution, have similar community profiles at the 16S rRNA gene level.

Establishment of a strain collection of gut bacteria isolated from stingless bees. To enable genomic and experimental analyses of stingless bee gut bacteria, we established a culture collection of bacteria isolated from Fv, Sp, Mf, Mi, Ms, and Ml. We plated homogenized gut samples from the six bee species on eight different semisolid media and under three different atmospheres (microaerobic and anaerobic). This resulted in the cultivation of 98 distinct bacterial isolates (i.e., having different 16S rRNA genotypes or isolated from a different bee species or colony) from 11 bacterial families (Fig. 2A; Table S4). Most bacteria grew under both microaerobic and anaerobic conditions on generic growth media and formed colonies after 2 to 4 days of growth. The 16S rRNA genotypes of the isolated strains matched 32 ASVs, accounting for 16 to 87% of the overall community of the six stingless bee species and including many shared ASVs (Fig. 2A and B). BLASTN searches of the 16S rRNA gene sequences revealed that many of the isolates (55/98) were related to bacterial strains obtained from the gut of honey bees and bumblebees

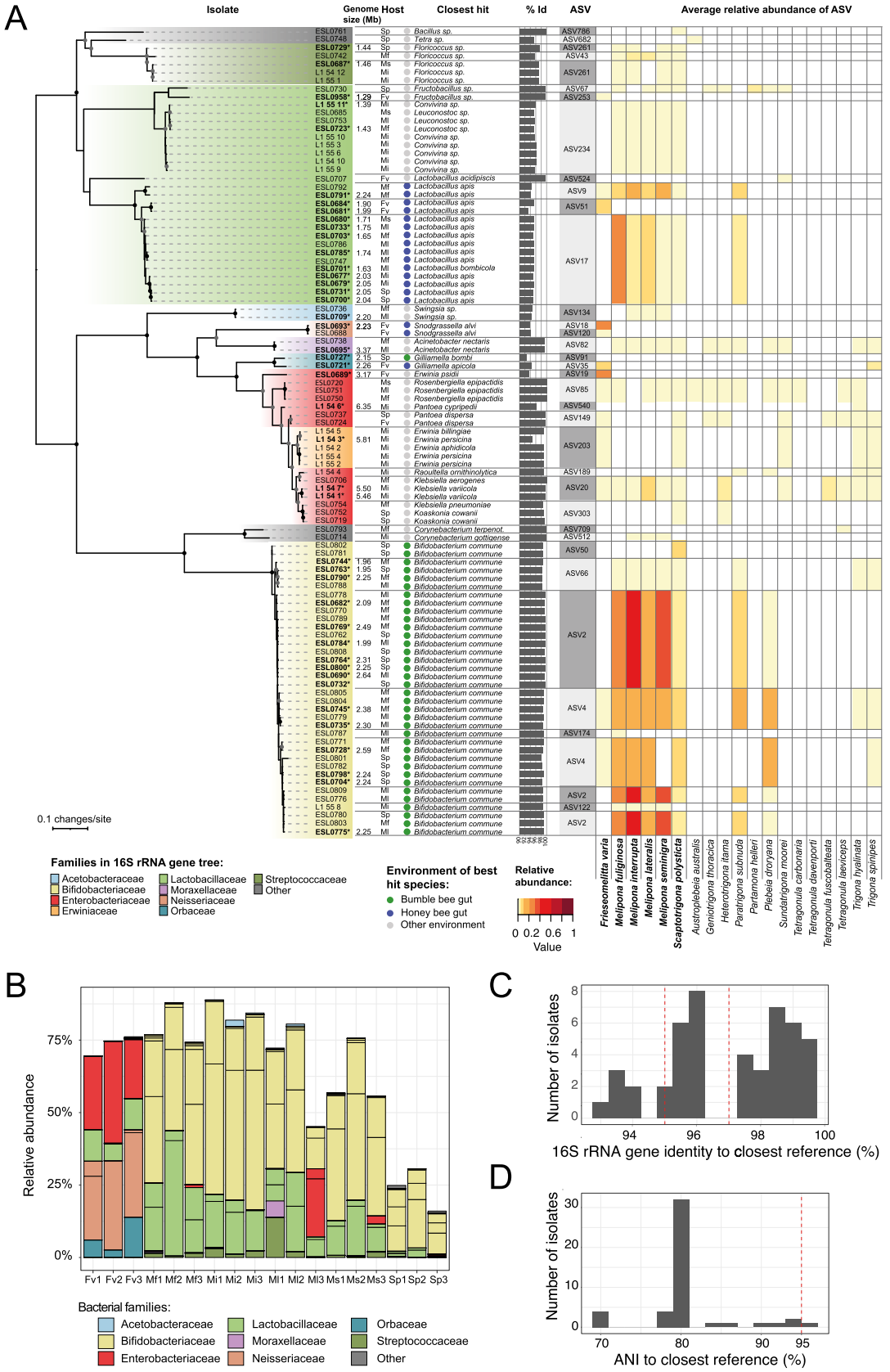

FIG 2 Bacterial strains isolated from the gut of six stingless bee species from Brazil. (A) Maximum-likelihood tree inferred from nearly complete sequences of the 16S rRNA gene of each isolate. Strain names of isolates for which we sequenced the

(Continued on next page)

(such as *Lactobacillus apis*, *Bifidobacterium commune*, *Gilliamella* sp., or *Snodgrassella alvi*) suggesting that they represent stingless bee isolates of core members of the social bee microbiota. Other isolates had best BLASTN hits to bacteria from other environments, such as a *Floricoccus* sp. isolated from flowers, various *Enterobacteriaceae* (e.g., *Pantoea* sp., *Klebsiella* sp., and *Rosenbergiella epipactidis*) isolated from humans and water, or *Fructobacillus* isolated from flowers and fruits. The percent identity of many of the BLAST hits was relatively low (<98%), suggesting that the isolated strains potentially correspond to new bacterial species (Fig. 2A and C; Table S4).

Stingless bee isolates form deep-branching phylogenetic lineages related to bacteria isolated from honey bees and bumblebees. To assess the phylogenetic placement of the isolated stingless bee gut bacteria relative to gut bacteria from honey bees and bumblebees, we selected 46 strains from 10 different bacterial families for genome sequencing (Fig. 2A; Table S4). Using a combination of Illumina and Oxford Nanopore sequencing, we obtained 23 complete and 23 draft genomes (2 to 66 contigs). The genome size of the cultured isolates ranged from 1.2 to 6.3 Mb. *Fructobacillus* ESL0730 (1.2 Mb) and the two *Streptococcaceae* strains ESL0687 and ESL0729 (1.4 Mb) harbored the smallest and *Leuconostocaceae* ESL0723 the largest (6.3 Mb) genomes of the sequenced strains (Table S4). Genome comparisons with other bacteria, including strains isolated from honey bees and bumblebees, showed that most stingless bee gut bacteria had 80% average nucleotide identity (ANI) with previously sequenced strains indicating that we isolated strains of novel bacterial species or genera (Fig. 2D).

Accordingly, genome-wide phylogenies based on single-copy orthologs showed that most isolates formed deep-branching, stingless bee-specific lineages, exclusive of any previously sequenced strain. However, consistent with the results of the 16S rRNA gene analysis, several of these lineages were related to major phylotypes of the honey bee and bumblebee gut microbiota (Fig. 3A to D; Fig. S2 to 6) such as *Snodgrassella*, *Gilliamella*, *Lactobacillus* Firm-5, *Bifidobacterium*, and *Bombella*. In the case of *Snodgrassella*, *Gilliamella*, and *Lactobacillus* Firm-5, the stingless bee-specific lineages formed a monophyletic clade with lineages of honey bee and bumblebee isolates (Fig. 3A to C and F). Notably, in all three cases, the bacteria from stingless bees presented the earliest-branching lineages, i.e., the honey bee and bumblebee gut bacteria diverged after the split from the stingless bee gut bacteria. While these results suggest that these bacteria are derived from a common ancestor that was already adapted to social bees, the bacterial phylogenies were incongruent with current phylogenies of the host, which show that the honey bees (Apini) diverged before the split of stingless bees (Meliponini) and bumblebees (Bombini) (Fig. 3E). A different pattern was observed for *Bifidobacterium*. In this case, strains isolated from stingless bees, honey bees, and bumblebees were not monophyletic. In fact, the stingless bee isolates belonged to a different clade than the honey bee isolates, while the strains isolated from bumblebees belonged to either of them (Fig. 3D and G). Similarly, the two *Acetobacteraceae* strains (ESL0695 and ESL0709) were not monophyletic with the honey bee isolates of the genus *Bombella*, although they belonged to the same Hymenoptera-associated clade within this family (Fig. S3). This suggests that in both cases, *Bifidobacterium* and *Acetobacteraceae*, bacteria of distinct lineages have independently adapted to the gut environment of social bees.

FIG 2 Legend (Continued)
genome are in bold. The genome size of each sequenced strain is noted next to it. The isolate host is indicated as follows: Fv, *Frieseomelitta varia*; Ms, *Melipona seminigra*; Ml, *Melipona lateralis*; Mf, *Melipona fuliginosa*; Mi, *Melipona interrupta*; Sp, *Scaptotrigona polysticta*. "Closest hit" indicates best BLASTN hit of the 16S rRNA gene sequence against the 16S rRNA (NCBI: *Bacteria* and *Archaea* type strains) database. Colored circles indicate if the strain of the best hit was isolated from the gut of a bumblebee or a honey bee or elsewhere. Bar plots indicate percent identity of the best BLASTN hit. The matching ASV and its average relative abundance across the 19 analyzed stingless bee species are indicated. The names of the stingless bee species from our study are in bold. Note that many isolates matched the same ASV. (B) Relative abundances of matching ASVs in each of the 18 samples of the six stingless bee species sampled in our study, i.e., representativeness of the isolates in the amplicon data. (C) Distribution of the isolated strains based on best 16S rRNA gene identity to the closest reference species. These are the values shown in panel A. The dashed lines indicate 95% and 97% identity corresponding to the genus and species thresholds, respectively. (D) Distribution of the isolated strains based on pairwise ANI with the closest reference genome publicly available. The dashed line indicates the species threshold at 95% ANI.

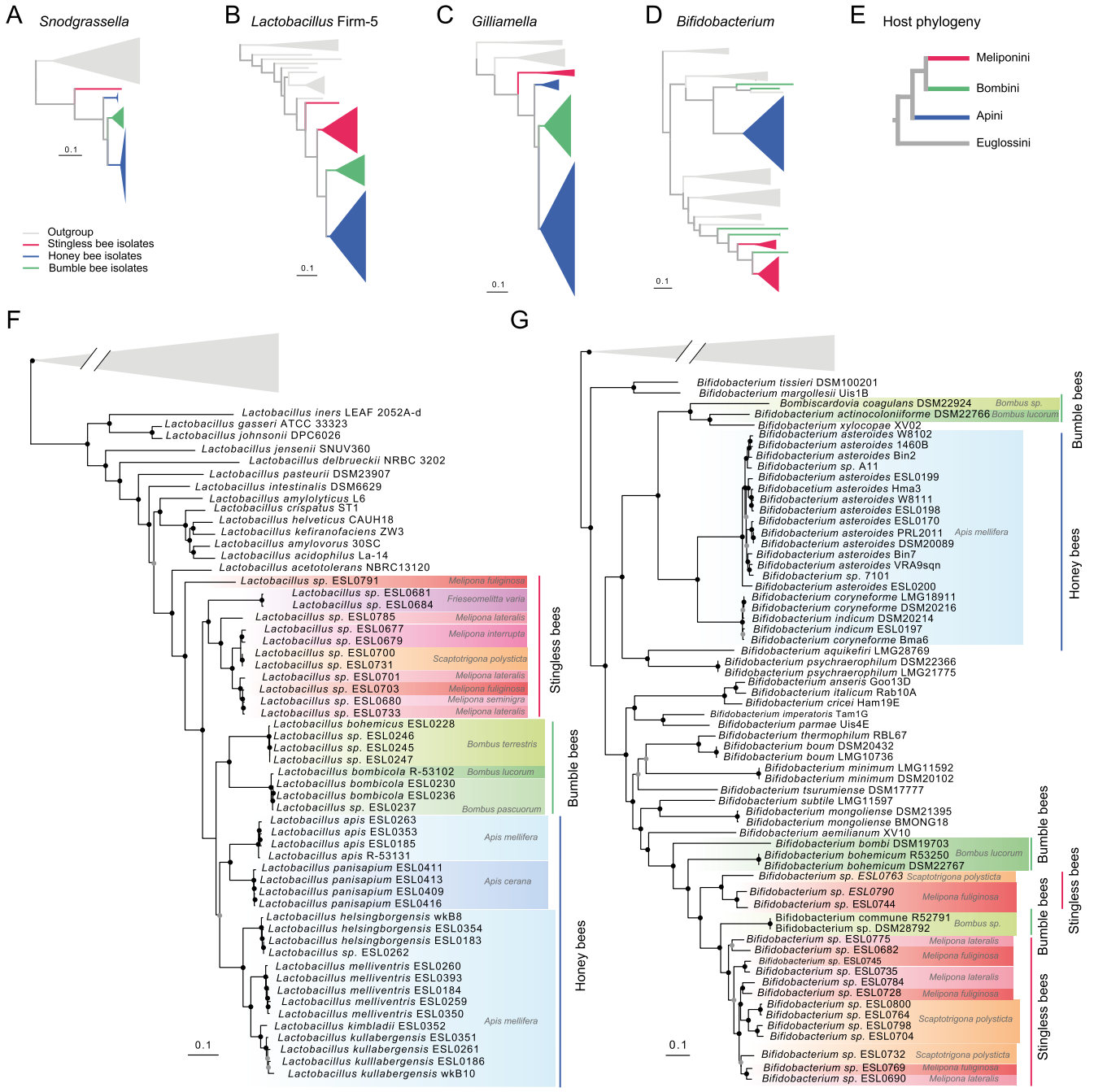

FIG 3 Isolates of stingless bee gut bacteria present novel species and belong to deep-branching phylogenetic lineages. (A to D) Simplified genome-wide maximum-likelihood phylogenies of *Snodgrassella*, *Lactobacillus* Firm-5, *Gilliamella*, and *Bifidobacterium* based on single-copy gene orthologs. All branches shown are supported by >95/100 bootstrap replicates. Most of the branches have been collapsed. (E) Dendrogram depicting the topology of the social bee phylogeny (adapted from references 52 and 53). (F) Detailed genome-wide phylogeny of the genus *Lactobacillus* with bacteria belonging to the social bee-specific phylotype *Lactobacillus* Firm-5 highlighted in different colors according to the host species/group. The maximum-likelihood tree was computed on the concatenated amino acid sequences of 355 single-copy core genes using the substitution model LG+F+I+G4. (G) Detailed genome-wide phylogeny of the genus *Bifidobacterium* with bacteria belonging to social bee-specific clades highlighted in different colors according to the host species/group. The maximum-likelihood tree was computed on the concatenated amino acid sequences of 151 single-copy core genes using the substitution model LG+F+I+G4.

Of the 46 strains selected for sequencing, 13 were not directly related to bacteria previously isolated from social bees. Most of these isolates matched minor ASVs in our community profiling analysis, with the exception of three strains (ESL0689, ESL0687, and ESL0729) (Fig. 2A). ESL0689 corresponded to the *Enterobacteriaceae* ASV19, which dominated the communities of all three Fv samples (20 to 35% of the

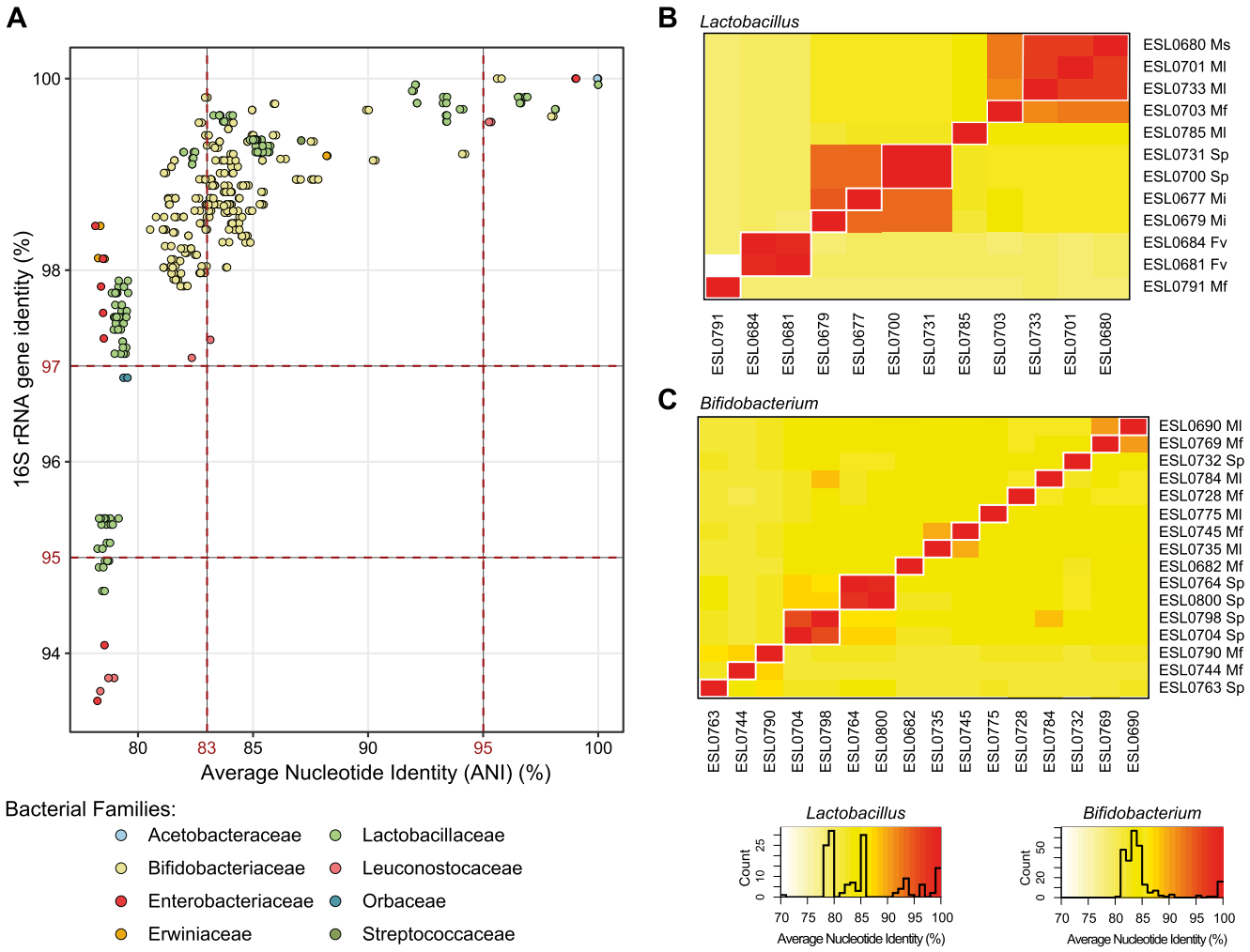

FIG 4 Genomic divergence among isolates with high 16S rRNA gene similarity. (A) The ANI versus 16S rRNA gene identity was plotted for each pair of isolate genomes belonging to the same bacterial family. The two vertical dashed bars indicate thresholds of 95% and 83%, demarcating the intraspecies and interspecies ANI values, respectively (86). The horizontal dashed lines delineate the commonly used species (>97%) and genus (>95%) thresholds for 16S rRNA gene similarity. (B and C) ANI heat maps of *Lactobacillus* Firm-5 isolates (B) and *Bifidobacterium* isolates (C). The 95% ANI clusters are outlined in white. The isolate host is indicated as follows: Fv, *Frieseomelitta varia*; Ms, *Melipona seminigra*; Ml, *Melipona lateralis*; Mf, *Melipona fuliginosa*; Mi, *Melipona interrupta*; Sp, *Scaptotrigona polysticta*.

reads per sample). This isolate was situated on a long branch diverging between the genera *Klebsiella* and *Raoultella* (Fig. S4). ESL0687 and ESL0729 corresponded to ASVs of the family *Streptococcaceae* which were detected in several stingless bee species in our studies as well as in previous studies (20, 37, 38). They formed a deep-branching sister clade of the flower-associated genus *Floricoccus* (Fig. S5A). All three strains seem to be novel species based on their divergence from previously sequenced bacteria and may represent specialized gut symbionts of stingless bees.

Stingless bee gut bacteria have diversified into distinct species and reveal a high extent of genomic diversity. Stingless bee isolates belonging to the same lineage were often separated by long branches in our phylogenies, indicating substantial genomic divergence (Fig. 3F and G; Fig. S2 and S6). This was confirmed by comparing pairwise 16S rRNA gene identity to genome-wide ANI between isolates of the same bacterial family. Despite high similarity in 16S rRNA gene identity, ANI was often <95%, suggesting that most lineages of stingless bee gut bacteria contain several divergent species (Fig. 4A). For example, the 12 sequenced strains of *Lactobacillus* Firm-5 fell into 8 distinct species-level clusters (i.e., ANI < 95%) (Fig. 4B). A similar pattern was observed for the 17 *Bifidobacterium* strains, which fell into 14 distinct species-level clusters (Fig. 4C), as well

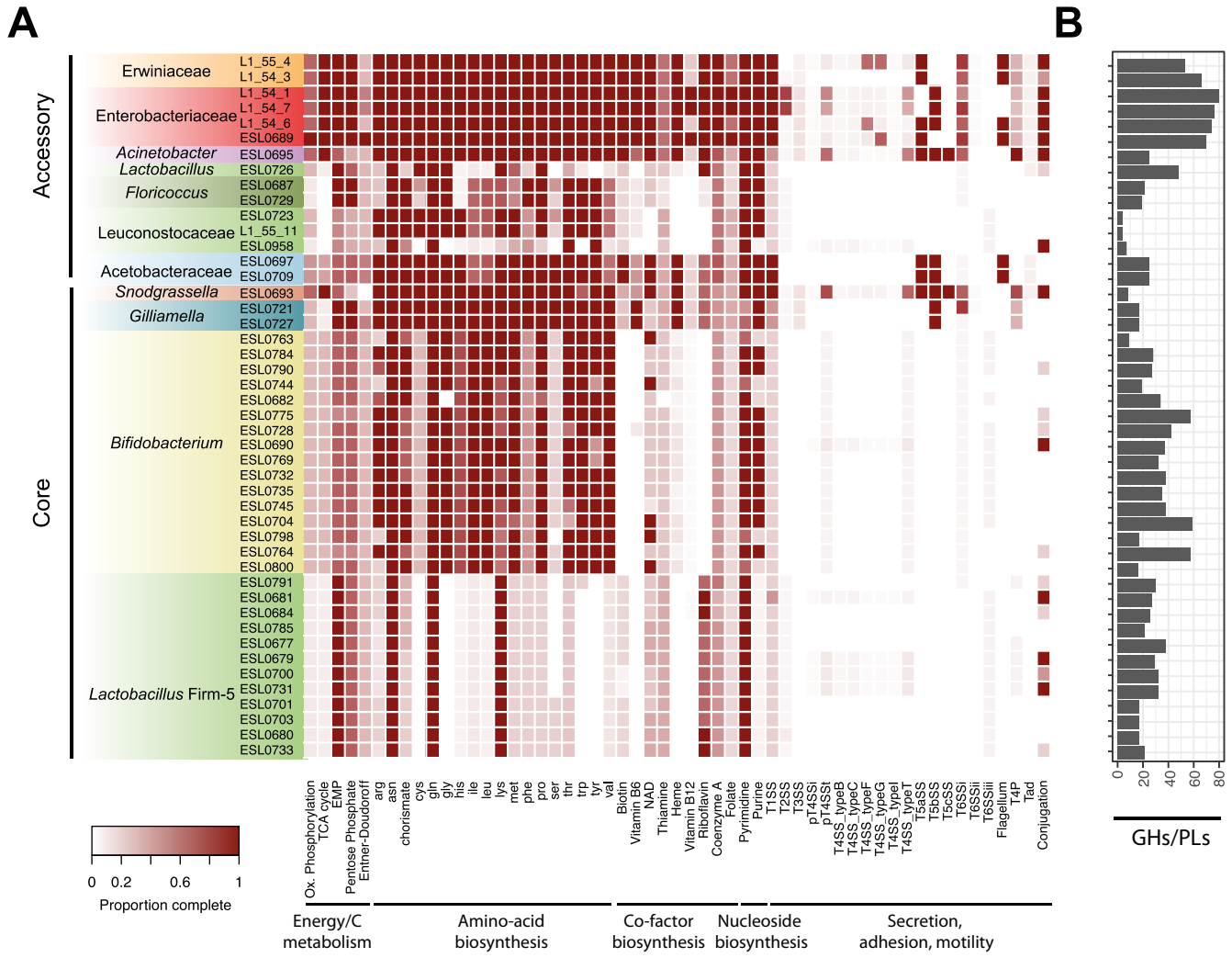

FIG 5 Metabolic capabilities of the sequenced stingless bee gut isolates. (A) The heat map indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism, biosynthesis of amino acids, cofactors, and nucleosides, secretion, adhesion, and motility across the sequenced bacterial isolates. The isolates are grouped into core and accessory members based on whether they are related to bacteria isolated from other social bees. (B) Total numbers of glycoside hydrolase (GH) and polysaccharide lyase (PL) genes in each isolate.

as the two *Gilliamella* and the two *Streptococcaceae* strains, which both also fell below the species-level ANI cutoff. Notably, some *Bifidobacterium* and *Lactobacillus* Firm-5 strains that were isolated from the same sample belonged to different ANI clusters, indicating that divergent bacterial species can co-occur in the same host species and colony. Inversely, strains belonging to the same ANI cluster were sometimes isolated from different bee species, suggesting that these bacterial species clusters are not necessarily host specific (Fig. 4B and C).

Core microbiota members in stingless bees have functional capabilities similar to those of microbiota members in honey bees and bumblebees. To assess the functional potential of stingless bee gut bacteria, we determined the genomic completeness of different metabolic pathways and functions in the genomes of the sequenced strains and compared it to that of related bacteria which had been isolated from honey bees, bumblebees, or elsewhere and which were included in our phylogenomic analysis. We specifically looked at energy and carbon metabolism, amino acid, cofactor, and nucleoside biosynthesis, secretion, motility, and adhesion (Fig. 5).

(i) Energy and carbon metabolism. Many of the sequenced strains (39/46), including all *Lactobacillaceae*, the *Bifidobacteriaceae*, and the two *Streptococcaceae* and *Gilliamella*

strains, were missing key genes of the tricarboxylic acid (TCA) cycle and for oxidative phosphorylation but encoded functions for the breakdown (mostly glycoside hydrolases [GH]) and oxidation (Embden-Meyerhof-Parnas [EMP], pentose phosphate [PPP], and/or Entner-Doudoroff [ED] pathways) of sugars. A detailed analyses of the enzymes used for carbohydrate breakdown by the core members *Lactobacillus* Firm-5, *Bifidobacterium*, and *Gilliamella* showed that the stingless bee gut bacteria carried glycoside hydrolase family genes similar to those in honey bee and bumblebee isolates (Fig. 6), including enzyme families for cleaving plant-derived glycans. For example, the glycoside families GH5, GH30, GH31, GH42, GH43, and GH51 are involved in the degradation of hemicellulose (27, 30, 34). GH78 can be responsible for the cleavage of rhamnose residues from rutin, a major pollen-derived flavonoid that was demonstrated to be deglycosylated by honey bee isolates of *Lactobacillus* Firm-5 which carry GH78 genes (34). Another example is GH13, which includes neopullulanases and α-amylases for the breakdown of plant-derived starch. Together, these results suggest that stingless bee gut isolates of *Lactobacillus* Firm-5, *Bifidobacterium*, and *Gilliamella* are saccharolytic fermenters that break down pollen- or nectar-derived glycans, as previously reported for the corresponding bacteria in the gut of honey bees and bumblebees. Notably, there was substantial variation in the number and type of glycoside hydrolase family genes between divergent strains, which is in line with the extensive genomic diversity detected between stingless bee gut isolates of these three phylotypes. A complete TCA cycle was found only in the genomes of *Neisseriaceae* strain ESL0693, *Acinetobacter* strain ESL0695, the *Enterobacteriaceae*, and the *Erwiniaceae*. The same strains also harbored the most complete gene sets for oxidative phosphorylation. Notably, *Neisseriaceae* strain ESL0693 also lacked key genes in the EMP, PPP, and ED pathways and contained very few GH family genes (Fig. 5). This suggests that this bacterium cannot utilize sugars and obtains energy via aerobic respiration, as previously found for *Snodgrassella* isolates of honey bees and bumblebees (23).

(ii) Amino acid, nucleoside, and cofactor biosynthesis. Differences between stingless bee gut isolates of different taxonomic groups were also found in terms of their biosynthetic potential. Strains belonging to the *Lactobacillus* Firm-5 clade were auxotrophic for the production of most amino acids (i.e., all except for Lys, Gln, and Asn) as well as purine and several cofactors (e.g., heme and vitamins B_6 and B_{12}) (Fig. 5). Isolates of the *Bifidobacteriaceae*, *Streptococcaceae*, and *Leuconostocaceae* were also auxotrophic for many cofactors but for many fewer amino acids than *Lactobacillus* Firm-5. Interestingly, there was variation in auxotrophies among the bifidobacterial strains, especially for the production of purine, NAD^+, Thr, Lys, Arg, Gly, and chorismate. Other strains (such as those of *Gilliamella*, *Snodgrassella*, *Acetobacteraceae*, *Enterobacteriaceae*, and *Erwiniaceae*) had fewer auxotrophies. Similar biosynthetic capability profiles were found in related strains included in our phylogenies, which suggests that these functional profiles are not specific to stingless bee gut bacteria but rather conserved across the entire phylotype (Fig. S2 to S6).

(iii) Secretion, adhesion, and motility. Secretion systems, pili, and flagella were mostly restricted to the Gram-negative bacteria of the isolated strains. Type I, type V, and type VI secretion systems were prevalent across these bacteria, whereas type II and type IV secretion systems were present in only a few strains (Fig. 5). Flagella were detected in the two *Acetobacteraceae*, all *Erwiniaceae*, and the *Enterobacteriaceae* strain ESL0689. Tad pili were not detected in any of the bacteria analyzed, while type IV pilus components were mostly found in the *Neisseriaceae* strain ESL0693 and *Acinetobacter* ESL0695 and to some extent also in *Orbaceae*, *Erwiniaceae*, and the *Enterobacteriaceae* strain ESL0689. Similar gene sets for secretion, adhesion, and motility were also found in related gut bacteria from honey bees or bumblebees, as shown by the functional profiles of all strains included in our phylogenies (Fig. S2 to S6).

Altogether, this first assessment of the gene content of the stingless bee gut bacteria shows that they have functional potential similar to that of isolates from honey bees and bumblebees, suggesting that they occupy similar ecological niches in the gut across social bees.

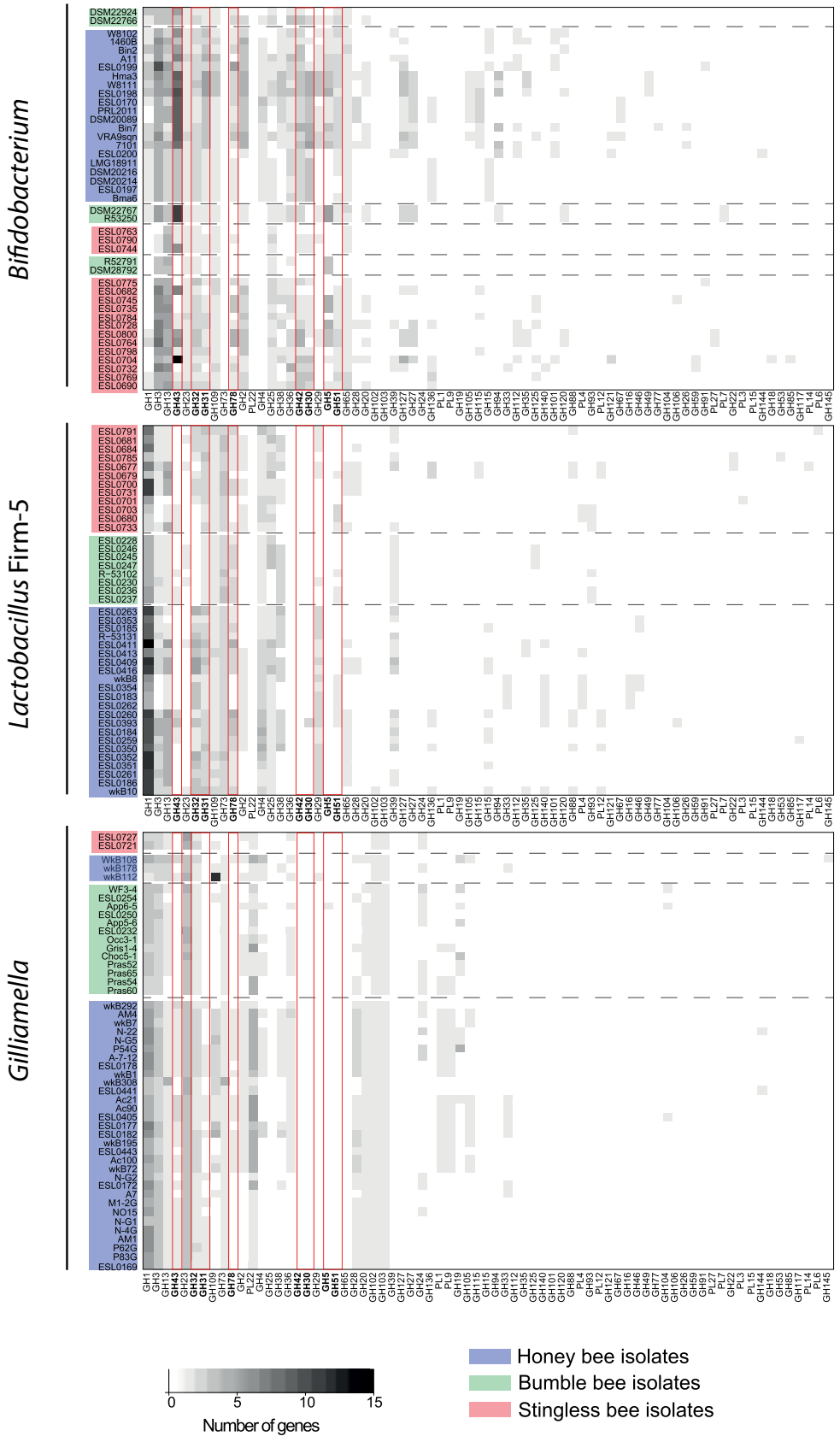

FIG 6 Carbohydrate-active enzyme profiles for *Bifidobacterium*, *Lactobacillus* Firm-5, and *Gilliamella* isolates. Distributions of genes in the GH and PL families for *Bifidobacterium*, *Lactobacillus* Firm-5, and *Gilliamella* isolates.

(Continued on next page)

DISCUSSION

Previous findings suggested that the core members of the bee gut microbiota were acquired in a common ancestor of the social bees (20) and possibly codiversified with their hosts over millions of years (16, 33). However, the lack of genomic data from gut bacteria of stingless bees has limited our view of the evolution of these specialized microbial communities. With the establishment of the genomes of diverse bacterial isolates from the stingless bee gut, our study fills an important knowledge gap and provides new lines of evidence that rule out strict codiversification between the core microbiota members and social bees.

Our genome-wide phylogenies of *Snodgrassella*, *Lactobacillus* Firm-5, and *Gilliamella* revealed that the identified lineages of stingless bee gut bacteria branched off before the divergence of the lineages of honey bee and bumblebee gut bacteria. This pattern is not congruent with the topology of the host phylogeny, in which honey bees diverged before the split of stingless bees and bumblebees (52, 53). The basal split of the stingless bee gut bacterial lineages depends on the rooting of our trees being correct. However, all phylogenies were supported by high bootstrap values at the critical nodes, suggesting robust phylogenetic signals in our data sets. Our phylogenies also showed that for both *Gilliamella* and *Snodgrassella*, honey bee isolates were not monophyletic, i.e., some lineages branched before and others after the divergence of the bumblebee clades, which is inconsistent with codiversification. This was already noted in a previous study (33), and similar results were also obtained for *Lactobacillus* Firm-5 based on the phylogenetic analysis of single protein-coding genes cloned from different social bee species (20). Finally, our phylogenomic analysis showed that honey bee and stingless bee isolates of *Bifidobacterium* belonged to two separate clades of the *Bifidobacteriaceae*, suggesting that these bacteria have independently adapted to the gut environment of social bees.

Codiversification can occur only when symbionts and hosts exhibit a high degree of partner fidelity and are transmitted vertically from one generation to the next over many generations (54). While most core microbiota members of social bees indeed seem to have a host-restricted distribution (20), examples of lineages with a broader host range exist as well (55). Moreover, some strains have been experimentally shown to be able to colonize nonnative hosts, demonstrating that host jumps are possible (20, 27). Our study showed that closely related stingless bee species (i.e., from the same or related genera) can have overlapping community profiles, with predominant ASVs being shared across hosts. Similar observations have been made in other 16S rRNA gene profiling studies of stingless bees (20, 37). While such analyses often provide insufficient resolution to discriminate between closely related strains or species, our genomic analyses confirmed that stingless bee isolates of *Bifidobacterium* and *Lactobacillus* do not necessarily cluster by host species. It is possible that the strength of host specificity varies across social bees or bacterial lineages, depending on the symbiotic function of the gut bacteria or the hosts' divergence, ecology, or geographic distribution. This would influence the extent to which gut bacteria can codiversify within certain host lineages. To test this hypothesis, future studies would need to compare the strength of host specificity across multiple bees using representative sets of host species in each of the three main bee clades.

Another piece of evidence indicating that the microbiotas across social bees may be more variable than previously assumed comes from the observation that some of the designated core members were not always detected in the sampled bees. For example, while *Lactobacillus* and *Bifidobacterium* were prevalent across all six stingless bee species sampled in our study, they were absent from the gut microbial communities of some of the previously sampled bee species (15, 17, 37, 56). Likewise, stingless

FIG 6 Legend (Continued)
The isolates were sorted according to their position in the phylogenies; the hosts are indicated by the background colors. Glycoside hydrolase families mentioned in the text are outlined in red. The color scale indicates the number of genes in each family.

bees of the genus *Melipona* were shown to systematically lack the two core members *Snodgrassella* and *Gilliamella* (37). Both taxa were also rare across the four *Melipona* species analyzed in our study. However, three of 12 colonies analyzed had high abundances of *Snodgrassella*. This suggests that this bacterium is not completely absent from this bee genus but may occasionally be acquired from other bee species, that it varies in prevalence depending on season, bee age, or development, or that it is restricted to only the *Melipona* species analyzed in our study.

Finally, representatives of core members of the social bee gut microbiota were recently also found in bees of the distant genus *Xylocopa* (carpenter bees) (57, 58), suggesting that these bacteria may have been associated with bees before the emergence of sociality or that they have a broader and less specific distribution across bees than previously suspected.

In summary, our results together with previous findings indicate a rather dynamic evolutionary background of the core members of the social bee gut microbiota. Rather than their having strictly codiversified with their hosts, extended periods of host-restricted evolution (and likely codiversification in some lineages) seem to have been interrupted by host switches and by independent symbiont gains and losses. Our observation that the stingless bee isolates repeatedly form a sister group to bumblebee and honey bee isolates (for *Gilliamella*, *Snodgrassella*, and *Lactobacillus* Firm-5) is intriguing given the host phylogeny. It may suggest that these core members have an origin in stingless bees and then spread to the other two groups, especially for *Lactobacillus*, where two stingless bee isolates clades split before the split between bumblebee and honey bee isolates. However, given the large diversity of social and solitary bees, it is clear that the currently available data sets are insufficient to explain the distribution and phylogenetic relationships of these gut symbionts across hosts. Broader samplings of stingless bees, honey bees, and bumblebees, combined with genome-resolved approaches, are needed to fully understand the diversity, distribution, and evolutionary trajectories of social bee gut bacteria and to accurately reconstruct the ancestral bee microbiome composition. Formal analysis should be applied to test for codiversification.

The importance of sampling biases when assessing patterns of co-diversification between hosts and their gut bacteria is highlighted by the analyses of Bacteroidaceae gut symbionts of hominids. While originally reported to have codiversified with their hosts (59), reexamination with increased sampling disrupted the codiversification pattern observed earlier (60). In contrast, a recent study identified strong signals of parallel evolutionary history between seven (of 56 tested) gut bacterial taxa and human populations (61), and phylogenetic congruency has also been found for certain stinkbug insects and their primary gut symbionts (62). This demonstrates that codiversification has occurred between certain gut bacteria and their hosts.

Besides offering new insights into the evolution of the social bee gut microbiota, our genomic analysis also revealed the functional potential of major gut symbionts of the analyzed stingless bee species. All isolates of *Lactobacillus* Firm-5, *Bifidobacterium*, and *Gilliamella* carried genes for the saccharolytic fermentation of diet-derived carbohydrates. In contrast, *Snodgrassella* ESL0689 lacked such functions in its genome but instead harbored genes for aerobic respiration. These results are consistent with findings from honey bees and bumblebees (22, 23, 26–28, 30) and hence suggest that the core microbiota members occupy similar ecological niches across the three groups of social bees.

Another parallel to findings from honey bees and bumblebees was the extensive genomic divergence present among strains of the core members *Lactobacillus* Firm-5, *Gilliamella*, and *Bifidobacterium*, even when isolated from the same host species. Moreover, we found genomic variation in carbohydrate breakdown and amino acid and nucleoside biosynthesis functions among these strains. This suggests that the diversification of these bacteria has been driven not only by isolation in different host species but also by adaptation to different ecological niches in the gut, similar to what has been shown for bumblebees and honey bees (28, 33, 35, 63). These parallels may not be surprising, as the dietary

preferences of the analyzed stingless bee species are similar to those of honey bees and bumblebees. It will be interesting to look at the functional potential of the core microbiota members in bees that have different dietary habits, such as the vulture bees (38) (i.e., stingless bees that feed on raw meat instead of pollen), yet share a subset of the core members with other social bees.

Some of the isolate genomes we sequenced in our study did not come from any of the core members of the social bee gut microbiota. They may represent transient community members, opportunistic pathogens, or host-specific gut symbionts with functions complementary to those of the core microbiota. Therefore, the established genomes present an important resource for future research. A particular strain that drew our attention was the *Enterobacteriaceae* strain ESL0689, as it belonged to an ASV that was present at high relative abundance in all three colonies of Fv. ESL0689 harbored a complete TCA cycle and a respiratory chain and could also synthesize most amino acids and cofactors, indicating a metabolic niche similar to that of *Snodgrassella* (23).

In conclusion, our study provides new insights into the evolution of the social bee gut microbiota and represents a first step toward characterizing the functional potential of major gut bacteria present in stingless bees. However, given the large diversity of stingless bees, with hundreds of different species distributed throughout the tropical and subtropical regions of the world, it is clear that our study presents only the starting point in characterizing their genomic diversity and functional potential. More detailed studies and larger genomic surveys combined with experimental analyses will be needed to understand their evolution and assess their impact on the host.

MATERIALS AND METHODS

Bee sampling. Bees were collected from three different nests of each of the following Meliponini species in February and March 2019: *Frieseomelitta varia*, *Scaptotrigona polysticta*, *Melipona fuliginosa*, *Melipona interrupta*, *Melipona lateralis*, and *Melipona seminigra*. All nests were located in a rural meliponary in the vicinities of Iranduba municipality (Iranduba, AM, Brazil; 3°10′52.7′′S 60°07′08.5′′W), and the sampling was carried out under SISGEN collection permission no. A256E82. Bees were sampled at the entrance of each nest and immediately immobilized by cooling at 4°C. Then the entire gastrointestinal tract was dissected, and the hindgut separated from the anterior gut parts. Hindguts of bees from the same nest were pooled in two separate tubes. One pool per nest was mixed with 1× phosphate-buffered saline (PBS) and glycerol, homogenized using bead beating, and subsequently cryopreserved at −80°C. This pool was used for bacterial culturing as described below. The other pool was cryopreserved at −80°C without homogenization and was subsequently used for DNA extraction and 16S rRNA gene analysis. For the sample used for sequencing, 20 bees per nest were pooled, while for the sample for culturing, only 3 bees per nest were pooled. For *Frieseomelitta varia* and *Scaptotrigona polysticta*, we pooled 40 and 10 bees, respectively, due to the small size of these bee species.

Bacterial culturing. For establishing a culture collection of primary isolates from the gut of the six sampled bee species, serial dilutions of the cryopreserved homogenates were plated on eight different media: CBA (Columbia blood agar supplemented with 5% defibrinated sheep blood [Thermo Fisher]), MRSA (De Man, Rogosa and Sharpe agar) supplemented with fructose (2%) and L-cysteine (0.1%), MRSA supplemented with mannitol (2%), chocolate agar, TSA (tryptone soy agar), TYG (tryptone glucose yeast extract agar), GC, LBA (Luria-Bertani agar) without NaCl, BHIA (brain heart infusion agar), and SDA (Sabouraud dextrose agar). Plates were incubated in two different conditions: in a microaerobic 5% CO_2-enriched atmosphere and in an anaerobic chamber (72% N_2, 8% H_2, 20% CO_2), both at 34°C. After 2 to 7 days of incubation, colonies of different size and appearance were picked and regrown on the same media with the same culturing conditions. Cryo-stocks of bacterial strains of interest were prepared by harvesting bacterial biomass in liquid media corresponding to the solid growth media and supplemented with 20% glycerol. For DNA isolation, bacteria were grown from the stocks, and a single colony was picked and regrown on fresh media before harvesting bacterial biomass.

Genotyping of bacterial isolates. All colonies that were selected for culturing were genotyped by PCR and Sanger sequencing of a 16S rRNA gene fragment. To this end, a small amount of bacterial material was transferred to a lysis buffer (1 M Tris-HCl [pH 7.5], 0.5 M EDTA, 10% SDS) containing 2.5 μL lysozyme (20 mg/mL) and 2.5 μL proteinase K (20 mg/mL) and incubated for 10 min at 37°C, 20 min at 55°C, and 10 min at 95°C. PCR was performed with universal bacterial primers that amplify the V1-V5 region of the 16S rRNA gene (27F [AGRGTTYGATYMTGGCTCAG] and 907R [CCGTCAATTCMTTTRAGTTT]) using the following reagents and thermocycler program: initial denaturing at 94°C for 5 min, followed by 32 cycles of denaturing at 94°C for 30 s, annealing at 56°C for 30 s, and extension 72°C for 1 min, and a final extension at 72°C for 7 min. PCR results were checked on a 1% agarose gel. PCR products selected for Sanger sequencing were purified using ExoSAP-IT (1 μL 5× ExoSAP, 4 μL double-distilled water [ddH_2O]) with the following thermocycler program: 30 min at 37°C followed by 15 min at 80°C. Purified samples were then sent to Eurofins for sequencing. Sanger sequences were analyzed with Geneious suite (Geneious) and compared to GenBank at NCBI using BLAST tools (64).

DNA isolation and genome sequencing. DNA isolation for Illumina sequencing was carried out using a customized SPRI bead-based extraction method or the FastPure bacterial DNA isolation minikit (Vazyme). For the SPRI bead method, bacteria were harvested and resuspended in tubes containing 200 mg of 0.1-mm acid-washed glass beads and 200 μL of TER buffer (10 mM Tris-HCl, 1 mM EDTA, 100 μg/mL RNase A [pH 8.0]). Samples were homogenized using a FastPrep-25 5G instrument (2 rounds of 30 s with the power set to 6) and subsequently centrifuged at maximum speed for 10 min at room temperature. Forty microliters of SPRI beads was added to 100 μL of supernatant, immediately mixed thoroughly by repeated pipetting (>20 times), and incubated for 5 min at room temperature. After the tubes were placed on a magnet stand, the liquid was removed and discarded and the beads were washed twice with 200 μL 80% ethanol. After air drying of the tubes on the magnetic stand, 22 μL of 5 mM Tris-HCl (pH 8) was added. For isolating bacteria with the FastPure bacterial DNA isolation minikit, the manufacturer's protocol for Gram-positive bacteria was followed.

DNA isolation for Oxford Nanopore Technologies (ONT) sequencing was carried out using a custom DNA extraction protocol for Gram-positive bacteria. Tubes were prepared with glass beads and 160 μL of buffer P1 (Qiagen). Then bacteria were harvested and resuspended in these tubes by intensive vortexing. Lysozyme (20 μL, 100 mg/mL) was added, and after gentle mixing, tubes were incubated at 56°C with shaking at 600 rpm for 30 min. Then, 4 μL RNase A (100 mg/mL) was added to the tubes, followed by 150 μL of buffer AL (lysis buffer; Qiagen). After mixing by vortexing, tubes were incubated in a thermal mixer (37°C, 900 rpm) for 20 min. Tubes were centrifuged for 10 min at 14,000 rpm to pellet the beads, and the supernatant was transferred to new tubes with 35 μL sodium acetate and 270 μL isopropanol and mixed by inverting. Following incubation for 1 h at 4°C, DNA was pelleted by centrifugation at 14,000 rpm for 10 min at 25°C. The supernatant was discarded, and the pellets were washed with 1 mL 80% EtOH. After a second centrifugation (14,000 rpm for 10 min at 25°C), the ethanol was removed, and the pellet was left to dry at room temperature. DNA pellets were solubilized with 50 μL TER (10 mM Tris-HCl, 1 M EDTA [pH 8.0], 2 mg/mL RNase A), and tubes incubated at 37°C for 15 min. The solution was then transferred to PCR tubes. Forty microliters of NGClean beads was added, and the solution was mixed by repeated pipetting. After a 5-min incubation, PCR tubes were placed on magnetic stands. When the solutions were clear, the liquid was removed, and the beads were washed twice with 200 μL of 80% EtOH. Upon complete drying, beads were resuspended with 22 μL of 5 mM Tris-HCl (pH 8.0). Finally, tubes were placed again on the magnetic stand, and when the solution were clear, the supernatant was transferred to new 1.5-mL Eppendorf tubes.

Illumina sequencing libraries were prepared using the Nextera DNA Flex library preparation kit following instructions of the Illumina reference guide. This was followed by indexing, dilution, and denaturation according to the Illumina documents Index Adapters Pooling Guide and the MiniSeq System Denature and Dilute Libraries Guide. Libraries were checked and quantified using a double-stranded DNA (dsDNA) fluorescent dye, before being loaded on a MiniSeq high-output flow cell (150PE). ONT libraries were prepared using the ligation-based approach (LSK109). The sequencing was conducted on a ONT MinION instrument for a duration of 72 h with high-accuracy base-calling using Guppy (v5.0.11).

Genome assembly. Forty-six isolates from the stingless bees were sequenced with Illumina MiniSeq (150PE). Raw reads were checked using FastQC v0.11.9 (65) and trimmed by Trimmomatic v0.39 (66) using the parameters: PE -phred33 AllIllumina-PEadapters.fa:3:25:7 LEADING:9 TRAILING:9 SLIDINGWINDOW:4:15 MINLEN:60. For *de novo* assembly, we used SPAdes (–careful option, v3.15.2) (67). For assembly quality control, reads were mapped back against the assembly with BWA v0.7.17 (68) and SAMtools v1.12 (69) and plotted with R v4.1.1. Genomes completeness was evaluated with checkM v1.0.13 (70).

Additionally, 26 isolates were also sequenced with Oxford Nanopore to produce long reads. Nanopore long reads were filtered with Filtlong v0.2.0 (https://github.com/rrwick/Filtlong) for a minimum length of 7,000 and minimum mean *q* score of 10 (min_length 7000, min_mean_q 10, length_weight 10). ONT-based assemblies were computed with Flye v2.7.1 (71) over 5 iterations. Graphmap v0.5.2 (72) and Racon v1.0.1 (73) were used to perform two rounds of polishing. Finally, the Racon-corrected assembly and Illumina reads were fed to Pilon v1.24 (74) for single-base and indel corrections.

Genome annotation and analysis. Genomes were annotated with Prokka v1.13 (75). Phylogenies were computed for the bacterial families for which we had an isolate. For each family, we identified a set of closely related strains and outgroup taxa and retrieved their genomes from NCBI and IMG/Mer (76). All genomes were reannotated with Prokka to ensure annotations consistency. Gene orthology was inferred with OrthoFinder v2.3.8 (77). Single-copy ortholog genes were selected, and their amino acid sequences were aligned (mafft v7.453) (78). An in-house script was used to trim the alignments by removing positions with more than 50% gaps, and sequences belonging to the same genome were concatenated to produce a core gene alignment. This alignment was used to infer the maximum-likelihood phylogeny using IQTree (v1.7.beta17, -st AA -bb 1000 -seed 12345 -m TEST) (79). For each phylogeny, the best evolutionary model was chosen according to the Bayesian information criterion (BIC): LG+I+G4, *Enterobacteriaceae*; LG+F+I+G4, *Acetobacteraceae*, *Bifidobacterium*, *Lactobacillus*, *Leuconostocaceae*, *Moraxellaceae*, *Neisseriaceae*, and *Streptococcaceae*; and JTTDMut+F+I+G4, *Orbaceae*. Branch support of the trees was inferred using 1,000 ultrafast bootstrap (UFBoot) repetitions. Clades can be trusted when UFBoot values are >95%.

16S rRNA gene based community profiling. Region V4 of the 16S rRNA gene was amplified using the primers 515F-Nex and 806R-Nex (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCCAGCMGCCGCGGTAA and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT). The primers include Nextera XT index adapter sequences and the primers for the 16S rRNA V4 region (80) described by Kešnerová et al. (22). The two-step PCR was performed as follows: the first PCR used 12.5 μL of 2× Phanta Max master mix (Vazyme, Nanjing, China), 5 μL of Milli-Q water, 2.5 μL of each primer (5 μM), and 2.5 μL of template DNA for a total volume of 25 μL. The PCR program started with a denaturation step at 98°C for 30 s, followed

by 25 cycles of amplification (10 s at 98°C, 20 s at 55°C, and 20 s at 72°C) and a 5-min final extension step at 72°C. The PCR products were verified by 2% agarose gel electrophoresis, purified with clean next-generation sequencing (NGS) purification beads in a 1:0.8 ratio of PCR product to beads and eluted in 27 μL Tris (10 mM, pH 8.5). A second PCR step was performed to append the unique dual indexes to each sample in a total volume of 25 μL using 12.5 μL of 2× Phanta Max master mix (Vazyme, Nanjing, China), 5 μL of Milli-Q water, 2.5 μL of Nextera XT index primers 1 and 2 (Nextera XT Index kit, Illumina), and 2.5 μL of templated DNA. The PCR program started with a denaturation step at 95°C for 3 min, followed by 8 cycles of amplification (30 s at 95°C, 30 s at 55°C, and 30 s at 72°C), and a 5-min final extension step at 72°C. The libraries were then cleaned using clean NGS purification beads (1:1.1 ratio of PCR product to beads) and were eluted in 27.5 μL Tris (10 mM, pH 8.5). Prior to sequencing, the PCR product concentrations were quantified by PicoGreen and pooled in equimolar concentrations; the negative controls and blank extractions were pooled in equal volume. Sequencing was performed on an Illumina MiSeq sequencer (2 × 250 bp) by the Genomic Technology Facility of the University of Lausanne. We followed the DADA2 (81) pipeline to analyze the sequencing data. For the first part of the analysis, we executed the pipeline only on the 18 samples from our study. To control for possible contaminants, we used blank extractions and water as negative controls during the PCR. For the second part of the analysis, where we combined our data set with the one from reference 20, we executed the pipeline independently a second time. Sequence quality control was performed with the DADA2 integrated function plotQualityProfile. Data from the three sequencing runs were processed independently for the filtering, dereplication, the error rate calculation, and the sample inference. After merging the denoised forward and reverse pairs, we merged the three sequence tables (mergeSequenceTables), and we applied the collapseNoMismatch function to unite similar ASVs with shifts or length variation. We selected sequences in the range from 250 to 256 bp and removed chimeric sequences. The Silva nonredundant small-subunit (SSU) database v138.1 (81) was used for the taxonomic assignment of the ASVs. We filtered out 62 ASVs for which the taxonomic assignment matched "Eukaryota," "Chloroplast," or "Mitochondria." Finally, we removed samples with fewer than 5,000 reads. NMDS plots were computed in R using Bray-Curtis distances (Phyloseq: ordinate). The adonis package was used to carry out PERMANOVA.

Functional profiler. We developed a genomic profiler using several software programs to annotate the genomes and compute the completeness of key pathways and functions. The genomes were reannotated with GhostKOALA (82) to obtain KEGG annotations of genes. We created a set of rules to define the steps of selected energy and metabolism pathways and cofactor/nucleoside biosynthesis pathways and computed their completeness. In brief, an in-house script parsed the KEGG annotations of the genomes and for each pathway evaluated the completeness of each step. The pathway completeness was then summarized by counting the number of steps present relative to total number of steps needed. To obtain the completeness of amino acid biosynthesis pathways, we used GapMind (83). Secretion systems and related appendages were detected in the genomes by MacSyFinder's TXSScan module (84). Finally, we ran dbCan (85) for the annotation of carbohydrate-active enzymes.

Data availability. The bacterial genomes and sequencing data are available in the NCBI's BioProject database under the accession number PRJNA906295. Scripts used for the amplicon sequencing analysis, the phylogenies, and the metabolic profiler can be found on GitHub https://github.com/gsartonl/Publication_Sarton-Loheac_2022.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, EPS file, 14.4 MB.
FIG S2, EPS file, 1 MB.
FIG S3, EPS file, 0.9 MB.
FIG S4, EPS file, 0.8 MB.
FIG S5, EPS file, 0.9 MB.
FIG S6, EPS file, 1.1 MB.
TABLE S1, XLSX file, 0.01 MB.
TABLE S2, XLSX file, 0.1 MB.
TABLE S3, XLSX file, 0.9 MB.
TABLE S4, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

We thank Julien Marquis and the team from the Lausanne Genomics Technology Facility and Alban Ramette and his team from the University of Bern for carrying out the Illumina sequencing and Oxford Nanopore sequencing, respectively. We also thank the Ecole de Biologie of the University of Lausanne for their financial support to sequence the presented genomes in the context of the course "Sequence-a-genome" (SAGE) as part of the Master of Science in Molecular Life Science.

The SAGE students of the years 2019-2020 and 2020-2021 include the following: Yami Ommar Arizmendi Cárdenas, Samuel Aubert, Alec Auston, Fabrice Battiston, Etienne Bellani, Valentina Benigno, Valentin Borgeat, Alessandro Brandulas Cammarata, Marion

Brechet, Marine Bugnon, Jessica Burnier, Théo Cavinato, Giacomo Ceracchini, Jérémy Cherbuin, Lucas Culebras, Audrey Daina, Hugues de Villiers de la Noue, Joe Dickinson, Alexandre Dudt, Elise Eray, Sara Ezzat, Christopher Forbes-Jaeger, Auriane Form, Léo Franchi, Pablo Guridi Fernández, Manon Henna, Aya Iizuka, Nicolas Jacquemin, Richie Kalusivikako, Maroussia Liechti, Simon Maréchal, Achille Mariotti, Aoife Mc Nally, Mam Malick Sy Ndiaye, Jade Nicolet, Astrid Oliva, Claire Paltenghi, Priyanka Parmar, Nicolas Pellaton, Katharina Pfaller, Brenda Ríos-Ochoa, Eric Risse, Daniel Rodriguez, Artemiy Saukin, Matthieu Simeoni, Miloš Stojanov, Marina Sudário, François Sutter, James Tan, Emanuele Tettamanti, Jamille Viray, Grazia Vizzarro, and Chaymae Ziyani.

This work was funded by an ERC Starting Grant (MicroBeeOme), NCCR Microbiomes, the National Centre of Competence in Research, the Swiss National Science Foundation (grant no. 180575), and a Swiss National Science Foundation Spirit grant (grant no. IZSTZ0_189496) to P.E. F.M. was funded by the Swiss National Science Foundation (grant no. 315230_184908: SOMETALP).

REFERENCES

1. Ascher JS, Pickering J. 2020. Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). http://www.discoverlife.org/mp/20q?guide=Apoidea_species.
2. Michner CD. 2007. The bees of the world. The Johns Hopkins University Press, Baltimore, MD.
3. Danforth BN, Cardinal S, Praz C, Almeida EA, Michez D. 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. Annu Rev Entomol 58:57–78. https://doi.org/10.1146/annurev-ento-120811-153633.
4. Bonilla-Rosso G, Engel P. 2018. Functional roles and metabolic niches in the honey bee gut microbiota. Curr Opin Microbiol 43:69–76. https://doi.org/10.1016/j.mib.2017.12.009.
5. Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. Nat Rev Microbiol 14:374–384. https://doi.org/10.1038/nrmicro.2016.43.
6. Raymann K, Moran NA. 2018. The role of the gut microbiome in health and disease of adult honey bee workers. Curr Opin Insect Sci 26:97–104. https://doi.org/10.1016/j.cois.2018.02.012.
7. Zhang Z, Mu X, Cao Q, Shi Y, Hu X, Zheng H. 2022. Honeybee gut Lactobacillus modulates host learning and memory behaviors via regulating tryptophan metabolism. Nat Commun 13:2037. https://doi.org/10.1038/s41467-022-29760-0.
8. Liberti J, Kay T, Quinn A, Kesner L, Frank ET, Cabirol A, Richardson TO, Engel P, Keller L. 2022. The gut microbiota affects the social network of honeybees. Nat Ecol Evol 6:1471–1479. https://doi.org/10.1038/s41559-022-01840-w.
9. Li L, Solvi C, Zhang F, Qi Z, Chittka L, Zhao W. 2021. Gut microbiome drives individual memory variation in bumblebees. Nat Commun 12:6588. https://doi.org/10.1038/s41467-021-26833-4.
10. Raymann K, Shaffer Z, Moran NA. 2017. Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. PLoS Biol 15:e2001861. https://doi.org/10.1371/journal.pbio.2001861.
11. Motta EVS, Mak M, De Jong TK, Powell JE, O'Donnell A, Suhr KJ, Riddington IM, Moran NA. 2020. Oral or topical exposure to glyphosate in herbicide formulation impacts the gut microbiota and survival rates of honey bees. Appl Environ Microbiol 86:e01150-20. https://doi.org/10.1128/AEM.01150-20.
12. Motta EVS, Raymann K, Moran NA. 2018. Glyphosate perturbs the gut microbiota of honey bees. Proc Natl Acad Sci U S A 115:10305–10310. https://doi.org/10.1073/pnas.1803880115.
13. Cariveau DP, Elijah Powell J, Koch H, Winfree R, Moran NA. 2014. Variation in gut microbial communities and its association with pathogen infection in wild bumblebees (Bombus). ISME J 8:2369–2379. https://doi.org/10.1038/ismej.2014.68.
14. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan P-L, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M, Pettis JS, Lipkin WI. 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318:283–287. https://doi.org/10.1126/science.1146498.
15. Hall MA, Brettell LE, Liu H, Nacko S, Spooner-Hart R, Riegler M, Cook JM. 2020. Temporal changes in the microbiome of stingless bee foragers following colony relocation. FEMS Microbiol Ecol 97:fiaa236. https://doi.org/10.1093/femsec/fiaa236.
16. Koch H, Abrol DP, Li J, Schmid-Hempel P. 2013. Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. Mol Ecol 22:2028–2044. https://doi.org/10.1111/mec.12209.
17. Leonhardt SD, Kaltenpoth M. 2014. Microbial communities of three sympatric Australian stingless bee species. PLoS One 9:e105718. https://doi.org/10.1371/journal.pone.0105718.
18. Li J, Powell JE, Guo J, Evans JD, Wu J, Williams P, Lin Q, Moran NA, Zhang Z. 2015. Two gut community enterotypes recur in diverse bumblebee species. Curr Biol 25:R652–R653. https://doi.org/10.1016/j.cub.2015.06.031.
19. 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. https://doi.org/10.1371/journal.pone.0036393.
20. Kwong WK, Medina LA, Koch H, Sing K-W, Soh EJY, Ascher JS, Jaffé R, Moran NA. 2017. Dynamic microbiome evolution in social bees. Sci Adv 3:e1600513. https://doi.org/10.1126/sciadv.1600513.
21. Engel P, James RR, Koga R, Kwong WK, McFrederick QS, Moran NA. 2013. Standard methods for research on Apis mellifera gut symbionts. J Apicultural Res 52:1–24. https://doi.org/10.3896/IBRA.1.52.4.07.
22. Kešnerová L, Mars RAT, Ellegaard KM, Troilo M, Sauer U, Engel P. 2017. Disentangling metabolic functions of bacteria in the honey bee gut. PLoS Biol 15:e2003467. https://doi.org/10.1371/journal.pbio.2003467.
23. Kwong WK, Engel P, Koch H, Moran NA. 2014. Genomics and host specialization of honey bee and bumblebee gut symbionts. Proc Natl Acad Sci U S A 111:11509–11514. https://doi.org/10.1073/pnas.1405838111.
24. 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. https://doi.org/10.1016/j.cub.2008.08.066.
25. Gallai N, Salles JM, Settele J, Vaissiere BE. 2009. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. Ecol Econ 68:810–821. https://doi.org/10.1016/j.ecolecon.2008.06.014.
26. Bottacini F, Milani C, Turroni F, Sánchez B, Foroni E, Duranti S, Serafini F, Viappiani A, Strati F, Ferrarini A, Delledonne M, Henrissat B, Coutinho P, Fitzgerald GF, Margolles A, van Sinderen D, Ventura M. 2012. Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. PLoS One 7:e44229. https://doi.org/10.1371/journal.pone.0044229.
27. Ellegaard KM, Brochet S, Bonilla-Rosso G, Emery O, Glover N, Hadadi N, Jaron KS, van der Meer JR, Robinson-Rechavi M, Sentchilo V, Tagini F, Engel P, SAGE class 2016-17. 2019. Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. Mol Ecol 28:2224–2237. https://doi.org/10.1111/mec.15075.
28. Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SGE, Vásquez A. 2015. Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. BMC Genomics 16:284. https://doi.org/10.1186/s12864-015-1476-6.
29. Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, Moran NA. 2016. Metabolism of toxic sugars by strains of the bee gut symbiont Gilliamella apicola. mBio 7:e01326-16. https://doi.org/10.1128/mBio.01326-16.

30. Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, Tringe SG, Moran NA. 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. Proc Natl Acad Sci U S A 116:25909–25916. https://doi.org/10.1073/pnas.1916224116.
31. Zheng H, Powell JE, Steele MI, Dietrich C, Moran NA. 2017. Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. Proc Natl Acad Sci U S A 114:4775–4780. https://doi.org/10.1073/pnas.1701819114.
32. Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A 109:11002–11007. https://doi.org/10.1073/pnas.1202970109.
33. Li Y, Leonard SP, Powell JE, Moran NA. 2022. Species divergence in gut-restricted bacteria of social bees. Proc Natl Acad Sci U S A 119:e2115013119. https://doi.org/10.1073/pnas.2115013119.
34. Brochet S, Quinn A, Mars RA, Neuschwander N, Sauer U, Engel P. 2021. Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. Elife 10:e68583. https://doi.org/10.7554/eLife.68583.
35. Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota. Nat Commun 10:446. https://doi.org/10.1038/s41467-019-08303-0.
36. Ellegaard KM, Suenami S, Miyazaki R, Engel P. 2020. Vast differences in strain-level diversity in the gut microbiota of two closely related honey bee species. Curr Biol 30:2520–2531.E2527. https://doi.org/10.1016/j.cub.2020.04.070.
37. Cerqueira AES, Hammer TJ, Moran NA, Santana WC, Kasuya MCM, da Silva CC. 2021. Extinction of anciently associated gut bacterial symbionts in a clade of stingless bees. ISME J 15:2813–2816. https://doi.org/10.1038/s41396-021-01000-1.
38. Figueroa LL, Maccaro JJ, Krichilsky E, Yanega D, McFrederick QS. 2021. Why did the bee eat the chicken? Symbiont gain, loss, and retention in the vulture bee microbiome. mBio 12:e02317-21. https://doi.org/10.1128/mBio.02317-21.
39. Haag KL, Caesar L, da Silveira Regueira-Neto M, de Sousa DR, Montenegro Marcelino V, de Queiroz Balbino V, Torres Carvalho A. 2022. Temporal changes in gut microbiota composition and pollen diet associated with colony weakness of a stingless bee. Microb Ecol https://doi.org/10.1007/s00248-022-02027-3. Epub ahead of print.
40. Ngalimat MS, Raja Abd Rahman RNZ, Yusof MT, Amir Hamzah AS, Zawawi N, Sabri S. 2020. A review on the association of bacteria with stingless bees. Sains Malays 49:1853–1863. https://doi.org/10.17576/jsm-2020-4908-08.
41. Tamarit D, Ellegaard KM, Wikander J, Olofsson T, Vásquez A, Andersson SGE. 2015. Functionally structured genomes in Lactobacillus kunkeei colonizing the honey crop and food products of honeybees and stingless bees. Genome Biol Evol 7:1455–1473. https://doi.org/10.1093/gbe/evv079.
42. Jaafar NAI, Mohamad SAS, Razak WRWA. 2019. Antimicrobial activity and antibiotic resistance of lactic acid bacteria isolated from Malaysian stingless bee's gut. Malays J Microbiol 15:333–341. https://doi.org/10.21161/mjm.1915410.
43. Oliphant SA, et al. 2022. Apilactobacillus apisilvae sp. nov., Nicolia spurrieriana gen. nov. sp. nov., Bombilactobacillus folatiphilus sp. nov. and Bombilactobacillus thymidiniphilus sp. nov., four new lactic acid bacterial isolates from stingless bees Tetragonula carbonaria and Austroplebeia australis. Int J Syst Evol Microbiol 72:5588. https://doi.org/10.1099/ijsem.0.005588.
44. Roubik DW. 2023. Stingless bee (Apidae: Apinae: Meliponini) ecology. Annu Rev Entomol 68:231–256. https://doi.org/10.1146/annurev-ento-120120-103938.
45. Grüter C. 2020. Stingless bees. Springer Nature, Cham, Switzerland.
46. Gaulke CA, Arnold HK, Humphreys IR, Kembel SW, O'Dwyer JP, Sharpton TJ. 2018. Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota. mBio 9:e01348-18. https://doi.org/10.1128/mBio.01348-18.
47. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat Commun 8:14319. https://doi.org/10.1038/ncomms14319.
48. Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, Ley RE, Farnleitner AH. 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nat Commun 10:2200. https://doi.org/10.1038/s41467-019-10191-3.
49. Reese AT, Dunn RR. 2018. Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. mBio 9:e01294-18. https://doi.org/10.1128/mBio.01294-18.
50. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332:970–974. https://doi.org/10.1126/science.1198719.
51. Levin D, Raab N, Pinto Y, Rothschild D, Zanir G, Godneva A, Mellul N, Futorian D, Gal D, Leviatan S, Zeevi D, Bachelet I, Segal E. 2021. Diversity and functional landscapes in the microbiota of animals in the wild. Science 372:eabb5352. https://doi.org/10.1126/science.abb5352.
52. Bossert S, Murray EA, Almeida EAB, Brady SG, Blaimer BB, Danforth BN. 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Mol Phylogenet Evol 130:121–131. https://doi.org/10.1016/j.ympev.2018.10.012.
53. Romiguier J, Cameron SA, Woodard SH, Fischman BJ, Keller L, Praz CJ. 2016. Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. Mol Biol Evol 33:670–678. https://doi.org/10.1093/molbev/msv258.
54. Groussin M, Poyet M, Sistiaga A, Kearney SM, Moniz K, Noel M, Hooker J, Gibbons SM, Segurel L, Froment A, Mohamed RS, Fezeu A, Juimo VA, Lafosse S, Tabe FE, Girard C, Iqaluk D, Nguyen LTT, Shapiro BJ, Lehtimäki J, Ruokolainen L, Kettunen PP, Vatanen T, Sigwazi S, Mabulla A, Domínguez-Rodrigo M, Nartey YA, Agyei-Nkansah A, Duah A, Awuku YA, Valles KA, Asibey SO, Afihene MY, Roberts LR, Plymoth A, Onyekwere CA, Summons RE, Xavier RJ, Alm EJ. 2021. Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell 184:2053–2067.E2018. https://doi.org/10.1016/j.cell.2021.02.052.
55. Powell E, Ratnayeke N, Moran NA. 2016. Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. Mol Ecol 25:4461–4471. https://doi.org/10.1111/mec.13787.
56. Tola YH, Waweru JW, Ndungu NN, Nkoba K, Slippers B, Paredes JC. 2021. Loss and gain of gut bacterial phylotype symbionts in Afrotropical stingless bee species (Apidae: Meliponinae). Microorganisms 9:2420. https://doi.org/10.3390/microorganisms9122420.
57. Holley JC, Jackson MN, Pham AT, Hatcher SC, Moran NA. 2022. Carpenter bees (Xylocopa) harbor a distinctive gut microbiome related to that of honey bees and bumble bees. Appl Environ Microbiol 88:e00203-22. https://doi.org/10.1128/aem.00203-22.
58. Handy MY, Sbardellati DL, Yu M, Saleh NW, Ostwald MM, Vannette RL. 2022. Incipiently social carpenter bees (Xylocopa) host distinctive gut bacterial communities and display geographical structure as revealed by full-length PacBio 16S rRNA sequencing. Mol Ecol https://doi.org/10.1111/mec.16736. Epub ahead of print.
59. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. 2016. Cospeciation of gut microbiota with hominids. Science 353:380–382. https://doi.org/10.1126/science.aaf3951.
60. Nishida AH, Ochman H. 2021. Captivity and the co-diversification of great ape microbiomes. Nat Commun 12:5632. https://doi.org/10.1038/s41467-021-25732-y.
61. Suzuki TA, Fitzstevens JL, Schmidt VT, Enav H, Huus KE, Mbong Ngwese M, Grießhammer A, Pfleiderer A, Adegbite BR, Zinsou JF, Esen M, Velavan TP, Adegnika AA, Song LH, Spector TD, Muehlbauer AL, Marchi N, Kang H, Maier L, Blekhman R, Ségurel L, Ko G, Youngblut ND, Kremsner P, Ley RE. 2022. Codiversification of gut microbiota with humans. Science 377:1328–1332. https://doi.org/10.1126/science.abm7759.
62. Kikuchi Y, Hosokawa T, Nikoh N, Meng XY, Kamagata Y, Fukatsu T. 2009. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. BMC Biol 7:2. https://doi.org/10.1186/1741-7007-7-2.
63. Engel P, Stepanauskas R, Moran NA. 2014. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. PLoS Genet 10:e1004596. https://doi.org/10.1371/journal.pgen.1004596.
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
65. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics.
66. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
67. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

68. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
69. Etherington GJ, Ramirez-Gonzalez RH, MacLean D. 2015. bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. Bioinformatics 31:2565–2567. https://doi.org/10.1093/bioinformatics/btv178.
70. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
71. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc Natl Acad Sci U S A 113:E8396–E8405. https://doi.org/10.1073/pnas.1604560113.
72. Sovic I, Sikic M, Wilm A, Fenlon SN, Chen S, Nagarajan N. 2016. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun 7:11307. https://doi.org/10.1038/ncomms11307.
73. Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
74. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
75. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
76. Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res 42:D568–D573. https://doi.org/10.1093/nar/gkt919.
77. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157. https://doi.org/10.1186/s13059-015-0721-2.
78. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436.
79. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
80. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108(Suppl 1):4516–4522. https://doi.org/10.1073/pnas.1000080107.
81. McLaren MRC, Callahan BJ. 2021. Silva 138.1 prokaryotic SSU taxonomic training data formatted for DADA2. Zenodo. https://doi.org/10.5281/zenodo.4587955.
82. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
83. Price MN, Deutschbauer AM, Arkin AP. 2020. GapMind: automated annotation of amino acid biosynthesis. mSystems 5:e00291-20. https://doi.org/10.1128/mSystems.00291-20.
84. Abby SS, Neron B, Menager H, Touchon M, Rocha EP. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One 9:e110726. https://doi.org/10.1371/journal.pone.0110726.
85. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451. https://doi.org/10.1093/nar/gks479.
86. Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351. https://doi.org/10.1099/ijs.0.059774-0.