# 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, 1015 Lausanne, Switzerland

[b]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

[c]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[d]Société des produits Nestlé, Lausanne, Switzerland

[e] Federal University of Amazonas, Instituto de Ciências Biológicas, Dept of Morphology, Manaus, Brazil

#equally contributed

Correspondence:

Prof. Philipp Engel

Department of Fundamental Microbiology

University of Lausanne, CH-1015 Lausanne, Switzerland

Tel.: +41 (0)21 692 56 12

e-mail: philipp.engel@unil.ch

## Abstract

Social bees harbor conserved gut microbiota that may have been acquired in a common ancestor of social bees and subsequently co-diversified with their hosts. However, most of this knowledge is based on studies on the gut microbiota of honey bees and bumble bees. Much less is known about the gut microbiota of the third and most diverse group of social bees, the stingless bees. Specifically, the absence of genomic data from their microbiota 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 bumble bee 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 co-diversification, but included host switches and independent symbiont gain and losses. Finally, as reported for the honey bee and bumble bee microbiota, we find 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.

## Introduction

The eusocial corbiculate bees (hereafter social bees) comprise more than 700 species distributed in three distinct tribes: honey bees (Apini), bumble bees (Bombini), and stingless bees (Meliponini). Stingless bees and bumble bees form a monophyletic clade, which is sister to the honey bees, with whom they shared a common ancestor 80-100 million years ago (1-3). Like mammals, social bees harbor dense and specialized bacterial communities in their gut that affect bee health and behavior (4-9). 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 (10-17). 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 (17), because they are prevalent and abundant across honey bee, bumble bee, and stingless bees. Most members of the bee gut microbiota are culturable and gnotobiotic bees can be generated for several species (5, 18-20). 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 (21, 22) which makes studies of their microbiota relevant in its own right.

Most of what is currently known about the gut microbiota of social bees stems from studies on honey bees and bumble bees. 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 (19, 20, 23-29). Further, it has been shown that the core members of the honey bee and bumble bee gut microbiota have substantially diversified (24-27, 30-33). They consist of divergent sub-lineages (or species) and contain

67 large extents of strain-level diversity and gene content variation. Most sub-lineages are host-
68 specific (20, 24, 30, 33), and their phylogenetic relationships are to some extent congruent
69 with the phylogeny of the host (13, 30). Therefore, it has been suggested that the core
70 members of the microbiota were acquired in a common ancestor of the social bees (17) and
71 possibly co-diversified with the host (13, 30). In addition, studies in the Western honey bee
72 (*Apis mellifera*) have shown that the diversification of the bee gut microbiota was also driven
73 by adaptation to different spatial and metabolic niches within the gut (24, 27, 30-32). For
74 example, strains of closely related sub-lineages of *Lactobacillus* Firm-5 and *Bifidobacterium*
75 can coexist in individual bees. They carry distinct gene sets for the breakdown and utilization
76 of pollen-derived carbohydrates which allows them to partition the available dietary glycan
77 niches in the gut (24, 27, 31).

78 In contrast to the microbiota of honey bees and bumble bees, much less is known about the
79 gut microbiota of the third group of social bees, the stingless bees (Meliponini). Previous
80 studies have focused on determining the taxonomic composition of the gut microbiota of
81 these bees using 16S rRNA gene sequencing (12, 14, 17, 34, 35). However, only a few bacteria
82 have been cultured from stingless bees (36-38) and, except for two strains of
83 *Bombilactobacillus* Firm-4 recently isolated from bees from Australia (39), no genomic data is
84 currently available for core members of the gut microbiota of stingless bees.

85 With >500 described species, stingless bees present the largest and most diverse group of the
86 social bees (40, 41). They are naturally distributed throughout the tropical and subtropical
87 regions of Africa, Asia, Australia, and the Americas and exhibit great variation in morphology,
88 diet, foraging range, social structure, and nesting habits (40, 41). As host phylogeny and
89 ecology are both key determinants of gut microbiota composition (42-47), we hypothesize

90 that genomic studies on bacterial isolates will help us to understand the functional diversity
91 of gut bacteria of stingless bees and provide novel insights into the evolution of these bacteria
92 across social bees, specifically in respect of the possible co-diversification with the host.

93 To address these questions, we looked at the gut microbiota of six neotropical species of
94 stingless bees from Brazil: *Frieseomelitta varia* (Fv), *Scaptotrigona polysticta* (Sp), *Melipona*
95 *fuliginosa* (Mf), *Melipona interrupta* (Mi), *Melipona seminigra* (Ms), and *Melipona lateralis*
96 (Ml). We determined the composition of the gut microbiota of these bees using 16S rRNA
97 gene sequencing, established a comprehensive culture collection of bacterial isolates, and
98 conducted genome sequencing and comparative genomics to determine the phylogenetic
99 placement, genomic diversity, and functional capabilities of these bacteria relative to those
100 previously isolated from honey bees and bumble bees.

101 ## Results

102 **Six neotropical stingless bee species from Brazil harbor distinct gut microbiota dominated**
103 **by nine bacterial families.**

104 We sampled three colonies of six stingless bee species (Fv, Sp, Mf, Mi, Ms, Ml) from a
105 meliponary located in the Amazonian rainforest near Manaus (**Supplementary Table 1**). For
106 each colony, we pooled the guts of 15-60 worker bees (depending on the size of the bee
107 species, see Methods) before DNA isolation. The V4 region of the 16S rRNA gene was
108 amplified and sequenced with Illumina MiSeq 2x250bp resulting in a median depth of 64,713
109 (52,479 to 95,774) reads per sample. In total, we identified 276 amplicon sequence variants
110 (ASVs, 29-63 ASVs per sample), which belonged to 37 different bacterial families
111 (**Supplementary Table 2**). Despite this diversity, only nine families dominated the samples,
112 representing together 97% of all quality-filtered reads (93-99% of the reads per sample):
113 Acetobacteraceae, Bifidobacteriaceae, Enterobacteriaceae, Lactobacillaceae, Neisseriaceae,
114 Orbaceae, Prevotellaceae, Streptococcaceae and Veillonellaceae (**Figure 1A**).

115

116 While Lactobacillaceae and Bifidobacteriaceae were abundant across all samples, there were
117 clear differences in the distribution of some of the other bacterial families (**Figure 1A).**
118 Neisseriaceae were abundant in the samples of Sp and Fv, but were only detected in three
119 out of 12 samples from the genus *Melipona* (**Figure 1A, Supplementary Table 2**). In contrast,
120 Acetobacteraceae and Streptococcaceae were present in most *Melipona* samples but rare
121 across samples of Fv and Sp. Orbaceae and Enterobacteriaceae were mostly detected in the
122 three Fv samples. Intriguingly, a single Enterobacteriaceae ASV constituted the most
123 abundant community member in this bee species (21 – 36% of the reads per sample).
124 According to these compositional differences, non-metric multidimensional scaling (NMDS)

125 separated the samples into three distinct clusters: two clusters comprised all samples of Fv
126 and Sp, respectively, and the third cluster comprised all samples of the four *Melipona* species
127 (Mf, Mi, Ms, Ml) (**Supplementary Figure 1**).
128
129 **Related stingless bee species have overlapping community profiles.**
130 We compared our results to a previously published amplicon sequencing dataset from
131 stingless bees (17) to assess the similarity of the communities to those of other stingless bee
132 species. After discarding samples with <5,000 reads to control for variation in sequencing
133 depths, our dataset comprised 136 samples from 19 different host species and three different
134 countries. We detected 688 ASVs in total with a median of 18 ASVs per sample (3 to 63 ASVs)
135 (**Supplementary Table 3**), spanning 53 bacterial families. Overall, the taxonomic patterns
136 were similar across the analyzed bee species. Apart from one sample of *Tetragonula*
137 *fuscobalteata,* for which 99% of the reads belonged to a single Weeksellaceae ASV, the nine
138 families dominating in the six bee species from our study were also abundant in the
139 microbiota of the samples from the previous study and represented between 34% to 100% of
140 the total number of reads (**Supplementary Figure 2**, **Supplementary Table 3**).
141
142 NMDS based on ASV relative abundances separated the samples by location (i.e. samples
143 from Brazil were different from those from Australia and Malaysia**,** Permutational
144 multivariate analysis of variance (PERMANOVA)**: '**Location' pseudo-F: 20.17, p-value=0.001**)**,
145 and by bee genus (PERMANOVA: 'Genus' pseudo-F=10.49, p-value=0.001), although
146 taxonomy and geography are not independent (**Figure 1B**). In contrast, there was only weak
147 clustering at the species level ('Species' pseudo-F=1.96, p-value=0.03). Notably, only 30% of
148 all ASVs (206 of 688 ASVs) were shared across host species (i.e. 70% of all ASVs are only found

149 in one species), and most of them (83.4%) only between 2-5 species (**Supplementary Figure**
150 **3A**). However, the shared ASVs belonged to the nine predominant bacterial families and
151 represented a large fraction of the total number of reads per sample (53.3-99.7 %; except for
152 the samples of Fv and *P. helleri*: 10.6-15 % of the reads) (**Supplementary Figure 3B**). In
153 particular, bees sampled in the same country or belonging to the same bee genus shared the
154 same ASVs, explaining the clustering of these samples in the NMDS analysis. Together, these
155 results show that despite the large variability observed, the gut microbiota of most stingless
156 bee species is dominated by a few bacterial families and that bee species of the same genus,
157 or with overlapping geographic distribution, have similar community profiles at the 16S rRNA
158 gene level.
159
160 **Establishment of a strain collection of gut bacteria isolated from stingless bees.**
161 To enable genomic and experimental analyses of stingless bee gut bacteria, we established a
162 culture collection of bacteria isolated from Fv, Sp, Mf, Mi, Ms, and Ml. We plated
163 homogenized gut samples of the six bee species on eight different semi-solid media and under
164 three different atmospheres (microaerophilic and anaerobic). This resulted in the cultivation
165 of 98 distinct bacterial isolates (i.e. different 16S rRNA genotype or isolated from a different
166 bee species or colony) from 11 bacterial families (**Figure 2A, Supplementary Table 4**). Most
167 bacteria grew under both microaerophilic or anaerobic condition on generic growth media
168 and formed colonies after 2-4 days of growth. The 16S rRNA genotypes of the isolated strains
169 matched to 32 ASVs, accounting for 16 – 87 % of the overall community of the six stingless
170 bee species and including many shared ASVs (**Figure 2A** and **B**). BLASTN searches of the 16S
171 rRNA gene sequences revealed that many of the isolates (55/98) were related to bacterial
172 strains obtained from the gut of honey bees and bumble bees (such as *Lactobacillus apis*,

173 *Bifidobacterium commune*, *Gilliamella sp.,* or *Snodgrassella alvi*) suggesting that they
174 represent stingless bee isolates of core members of the social bee microbiota. Other isolates
175 had best BLASTN hits to bacteria from other environments, such as *Floricoccus* sp. isolated
176 from flowers, various Enterobacteriaceae (e.g. *Pantoea* sp., *Klebsiella* sp., *Rosenbergiella*
177 *epipactidis*) isolated from humans and water, or *Fructobacillus* isolated from flowers and
178 fruits. The percent identity of many of the blast hits was relatively low (<98%) suggesting that
179 the isolated strains potentially correspond to new bacterial species **(Figure 2A** and **C,**
180 **Supplementary Table 4)**.
181
182 **Stingless bee isolates form deep-branching phylogenetic lineages related to bacteria**
183 **isolated from honey bees and bumble bees.**
184 To assess the phylogenetic placement of the isolated stingless bee gut bacteria relative to gut
185 bacteria from honey bees and bumble bees, we selected 46 strains from 10 different bacterial
186 families for genome sequencing **(Figure 2A, Supplementary Table 4)**. Using a combination of
187 Illumina and Oxford Nanopore sequencing, we obtained 23 complete and 23 draft genomes
188 (2-66 contigs). Genome size of the cultured isolates ranged from 1.2-6.3 Mb. *Fructobacillus*
189 ESL0730 (1.2Mb) and the two Streptococcaceae strains ESL0687 and ESL0729 (1.4 Mb)
190 harbored the smallest, and Leuconostocaceae ESL0723 the largest (6.3 Mb) genomes of the
191 sequenced strains **(Supplementary Table 4)**. Genome comparisons with other bacteria,
192 including strains isolated from honey bees and bumble bees, showed that most stingless bee
193 gut bacteria had 80% average nucleotide identity with previously sequenced strains indicating
194 that we have isolated strains of novel bacterial species or genera (**Figure 2D**).
195 Accordingly, genome-wide phylogenies based on single-copy orthologs showed that most
196 isolates formed deep-branching, stingless bee-specific lineages, exclusive of any previously

197 sequenced strain. However, consistent with the results of the 16s rRNA gene analysis, several
198 of these lineages were related to major phylotypes of the honey bees and bumble bee gut
199 microbiota (**Figure 3A-D, Supplementary Figure 4-12)** such as *Snodgrassella*, *Gilliamella*,
200 *Lactobacillus* Firm-5, *Bifidobacterium*, and *Bombella*. In the case of *Snodgrassella*, *Gilliamella*,
201 and *Lactobacillus* Firm-5, the stingless bee-specific lineages formed a monophyletic clade with
202 lineages of honey bee and bumble bee isolates (**Figure 3A-C** and **3F)**. Notably, in all three
203 cases, the bacteria from stingless bees presented the earliest branching lineages, i.e. the
204 honey bee and bumble bee gut bacteria diverged after the split from the stingless bee gut
205 bacteria. While these results suggest that these bacteria have derived from a common
206 ancestor that was already adapted to social bees, the bacterial phylogenies were incongruent
207 with current phylogenies of the host, which show that the honey bees (Apini) diverged before
208 the split of stingless bees (Meliponini) and bumble bees (Bombini) (**Figure 3E**). A different
209 pattern was observed for *Bifidobacterium*. In this case, strains isolated from stingless bees,
210 honey bees, and bumble bees were not monophyletic. In fact, the stingless bee isolates
211 belonged to a different clade than the honey bee isolates, while the strains isolated from
212 bumble bees belonged to either of them (**Figure 3D** and **3G**). Similarly, the two
213 Acetobacteraceae strains (ESL0695 and ESL0709) were not monophyletic with the honey bee
214 isolates of the genus *Bombella*, although they belonged to the same Hymenoptera-associated
215 clade within this family **(Supplementary Figure 6).** This suggests that in both cases,
216 *Bifidobacterium* and Acetobacteriaceae, bacteria of distinct lineages have independently
217 adapted to the gut environment of social bees.
218
219 Of the 46 strains selected for sequencing, 13 were not directly related to bacteria previously
220 isolated from social bees. Most of these isolates matched to minor ASVs in our community

221 profiling analysis, with the exception of three strains (ESL0689, ESL0687, and ESL0729, **Figure**
222 **2A**). ESL0689 corresponded to the Enterobacteriaceae ASV19 which dominated the
223 communities of all three Fv samples (20-35% of the reads per sample). This isolate was
224 situated on a long branch diverging between the genera *Klebsiella* and *Raoultella*
225 (**Supplementary Figure 7**). ESL0687 and ESL0729 corresponded to ASVs of the family
226 Streptococcaceae which were detected in several stingless bee species in our studies as well
227 as in previous studies (17, 34, 35). They formed a deep-branching sister clade of the flower-
228 associated genus *Floricoccus* (**Supplementary Figure 8**). All three strains seem to be novel
229 species based on their divergence to previously sequenced bacteria and may represent
230 specialized gut symbionts of stingless bees.
231
232 **Stingless bee gut bacteria have diversified into distinct species and reveal a high extent of**
233 **genomic diversity.**
234 Stingless bee isolates belonging to the same lineage were often separated by long branches
235 in our phylogenies indicating substantial genomic divergence (**Figure 3F** and **G**,
236 **Supplementary Figures 4-12**). This was confirmed by comparing pairwise 16S rRNA gene
237 identity to genome-wide average nucleotide identity (ANI) between isolates of the same
238 bacterial family. Despite high similarity in 16S rRNA gene identity, ANI was often <95%
239 suggesting that most lineages of stingless bee gut bacteria contain several divergent species
240 (**Figure 4A**). For example, the 12 sequenced strains of *Lactobacillus* Firm-5 fell into 8 distinct
241 species-level clusters (i.e. ANI <95%, **Figure 4B**). A similar pattern was observed for the 17
242 *Bifidobacterium* strains, which fell into 14 distinct species-level clusters (**Figure 4C**), as well as
243 the two *Gilliamella* and the two Streptococcaceae strains, which both also fell below the
244 species-level ANI cut-off. Notably, some *Bifidobacterium* and *Lactobacillus* Firm-5 strains that

245 were isolated from the same sample belonged to different ANI clusters, indicating that
246 divergent bacterial species can co-occur in the same host species and colony. Inversely,
247 strains belonging to the same ANI cluster were sometimes isolated from different bee species,
248 suggesting that these bacterial species clusters are not necessarily host-specific (**Figure 4B**
249 and **C**).
250
251 **Core microbiota members in stingless bees have similar functional capabilities as in honey**
252 **bees and bumble bees.**
253 To assess the functional potential of stingless bee gut bacteria, we determined the genomic
254 completeness of different metabolic pathways and functions in the genomes of the
255 sequenced strains and compared it to related bacteria isolated from honey bees, bumble
256 bees, or from elsewhere, and which were included in our phylogenomic analysis. We
257 specifically looked at energy and carbon metabolism, amino acid, co-factor, and nucleoside
258 biosynthesis, as well as secretion, motility, and adhesion (**Figure 5)**.
259
260 ***Energy and carbon metabolism:*** Many of the sequenced strains (39/46) - including all
261 Lactobacillaceae, the Bifidobacteria, and the two Streptococcaceae and *Gilliamella* strains –
262 were missing key genes of the TCA cycle and for oxidative phosphorylation, but encoded
263 functions for the breakdown (mostly glycoside hydrolases (GH)), and oxidation of sugars
264 (Embden-Meyerhof-Parnas (EMP), pentose phosphate (PPP), and/or Entner-Doudorrof (ED)
265 pathways). A detailed analyses of the enzymes used for carbohydrate breakdown by the core
266 members *Lactobacillus* Firm-5, *Bifidobacterium*, and *Gilliamella* showed that the stingless bee
267 gut bacteria encoded similar glycoside hydrolase family genes as honey bee and bumble bee
268 isolates (**Figure 6**), including enzyme families for cleaving plant-derived glycans. For example,

269 glycoside families GH5, GH30, GH31, GH42, GH43, and GH51 are involved in the degradation
270 of hemicellulose (24, 27, 48). GH78 can be responsible for the cleavage of rhamnose residues
271 from rutin, a major pollen-derived flavonoid that was demonstrated to be deglycosylated by
272 honey bee isolates of *Lactobacillus* Firm-5 which encode GH78 genes (48). Another example
273 is GH13 which includes neopullulanases and α-amylases for the breakdown of plant-derived
274 starch. Together, these results suggest that stingless bee gut isolates of *Lactobacillus* Firm-5,
275 *Bifidobacterium*, and *Gilliamella* are saccharolytic fermenters that breakdown pollen or
276 nectar-derived glycans, similar as previously reported for the corresponding bacteria in the
277 gut of honey bees or bumble bees. Notably, there was substantial variation in the number
278 and type of glycoside hydrolase family genes between divergent strains, which goes in line
279 with the extensive genomic diversity detected between stingless bee gut isolates of these
280 three phylotypes.A complete TCA cycle was only found in the genomes of Neisseriaceae sp.
281 ESL0693, *Acinetobacter* ESL0695, the Enterobacteriaceae, and the Erwiniaceae. The same
282 strains also harbored the most complete gene sets for oxidative phosphorylation. Notably,
283 Neisseriaceae sp. ESL0693 also lacked key genes in the EMP, PPP, and the ED pathways and
284 contained very few GH family genes (**Figure 5**). This suggests that this bacterium cannot utilize
285 sugars and obtains energy via aerobic respiration, as previously found for *Snodgrassella*
286 isolates of honey bees and bumble bees (20).
287
288 ***Amino acid, nucleoside, and co-factor biosynthesis***: Differences between stingless bee gut
289 isolates of different taxonomic groups were also found in terms of their biosynthetic
290 potential. Strains belonging to the *Lactobacillus* Firm-5 clade were auxotrophic for the
291 production of most amino acids (i.e. all except for Lys, Gln, and Asn) as well as purine and
292 several co-factors (e.g. heme, vitamin B6 and B12) **(Figure 5)**. Isolates of the Bifidobacteria,

293 Streptococcaceae, and Leuconostocaceae were also auxotrophic for many co-factors, but for
294 much fewer amino acids than *Lactobacillus* Firm-5. Interestingly, there was variation in
295 auxotrophies among the Bifidobacterial strains, especially for the production of purine, NAD+,
296 Thr, Lys, Arg, Gly, and chorismate. Other strains (such as those of *Gilliamella*, *Snodgrassella*,
297 Acetobacteraceae, Enterobacteriaceae, and Erwiniaceae) had fewer auxotrophies. Similar
298 biosynthetic capability profiles were found in related strains included in our phylogenies,
299 which suggests that these functional profiles are not specific to stingless bee gut bacteria but
300 rather conserved across the entire phylotype (**Supplementary Figures 4-12**).
301
302 ***Secretion, adhesion, motility:*** Secretion systems, pili, and flagella were mostly restricted to
303 the gram-negative bacteria of the isolated strains. Type 1, Type 5, and Type 6 secretion
304 systems were prevalent across these bacteria, whereas Type 2 and Type 4 secretion systems
305 were only present in a few strains **(Figure 5)**. Flagella were detected in the two
306 Acetobacteraceae, all Erwiniaceae, and Enterobacteriaceae strain ESL0689. Tad pili were not
307 detected in any of the analyzed bacteria, while Type IV pili components were mostly found in
308 Neisseriaceae sp. ESL0693 and *Acinetobacter* ESL0695, and to some extent also in Orbaceae,
309 Erwiniaceae, and Enterobacteriaceae ESL0689. Similar gene sets for secretion, adhesion, and
310 motility were also found in related gut bacteria from honey bees or bumble bees, as shown
311 by the functional profiles of all strains included in our phylogenies (**Supplementary Figures 4-**
312 **12**).
313
314 Altogether, this first assessment of the gene content of the stingless bee gut bacteria show
315 that they have similar functional potential as isolates from honey bees and bumble bees
316 suggesting that they occupy similar ecological niches in the gut across social bees.

317 ## Discussion

318 Previous findings suggested that the core members of the bee gut microbiota have been
319 acquired in a common ancestor of the social bees (17) and possibly co-diversified with their
320 hosts over millions of years (13, 30). However, the lack of genomic data from gut bacteria of
321 stingless bees has limited our view on the evolution of these specialized microbial
322 communities. With the establishment of isolate genomes of diverse stingless bee gut bacteria
323 our study fills an important knowledge gap and provides new lines of evidence that rule out
324 strict co-diversification between the core microbiota members and social bees.

325

326 Our genome-wide phylogenies of *Snodgrassella*, *Lactobacillus* Firm-5, and *Gilliamella*
327 revealed that the identified lineages of stingless bee gut bacteria branched off before the
328 divergence of the lineages of honey bee and bumble bee gut bacteria. This pattern is not
329 congruent with the topology of the host phylogeny, in which honey bees diverged before the
330 split of stingless bees and bumble bees (49, 50). The basal split of the stingless bee gut
331 bacterial lineages depends on the rooting of our trees being correct. However, all phylogenies
332 were supported by high bootstrap values at the critical nodes suggesting robust phylogenetic
333 signals in our datasets. Our phylogenies also showed that for both *Gilliamella* and
334 *Snodgrassella*, honey bee isolates were not monophyletic, i.e. some lineages branched before
335 and others after the divergence of the bumble bee clades, which is inconsistent with co-
336 diversification. This was already noted in a previous study (30), and similar results were also
337 obtained for Lactobacillus Firm5 based on the phylogenetic analysis of single protein-coding
338 genes cloned from different social bee species (17). Finally, our phylogenomic analysis
339 showed that honey bee and stingless bee isolates of *Bifidobacterium* belonged to two

340 separate clades of the Bifidobacteriaceae suggesting that these bacteria have independently
341 adapted to the gut environment of social bees.
342
343 Co-diversification can only occur when symbionts and hosts exhibit a high degree of partner
344 fidelity and are transmitted vertically from one generation to the next over many generations
345 (51). While most core microbiota members of social bees indeed seem to have a host-
346 restricted distribution (17), examples of lineages with a broader host range exist as well (52).
347 Moreover, some strains have been experimentally shown to be able to colonize non-native
348 hosts demonstrating that host jumps are possible (17, 24). Our study showed that closely
349 related stingless bee species (i.e. from the same or related genera) can have overlapping
350 community profiles with predominant ASVs being shared across hosts. Similar observations
351 have been made in other 16S rRNA gene profiling studies of stingless bees (17, 34). While
352 such analyses often provide insufficient resolution to discriminate between closely related
353 strains or species, our genomic analyses confirmed that stingless bees isolates of
354 *Bifidobacterium* and *Lactobacillus* do not necessarily cluster by host species. It is possible that
355 the strength of host specificity varies across social bees, or bacterial lineages, depending on
356 the symbiotic function of the gut bacteria, or the hosts' divergence, ecology, or geographic
357 distribution. This would influence the extent to which gut bacteria can co-diversify within
358 certain host lineages. To test this hypothesis, future studies would have to compare the
359 strength of host specificity across multiple bees using representative sets of host species in
360 each of the tree main bee clades.
361
362 Another piece of evidence indicating that the core microbiota across social bees may be more
363 variable than previously assumed comes from our finding that only *Lactobacillus* and

364 *Bifidobacterium* were prevalent across all stingless bee samples. Similar observations have
365 been made in other studies (12, 14, 34, 53). For example, stingless bees of the genus *Melipona*
366 were shown to systematically lack the two core members *Snodgrassella* and *Gilliamella* (34).
367 While both taxa were also rare across the four *Melipona* species analyzed in our study, three
368 out of 12 analyzed colonies had high abundances of *Snodgrassella*. This suggests that this
369 bacterium is not completely absent from this bee genus, but may be occasionally acquired
370 from other bee species, restricted to only the *Melipona* species analyzed in our study, or
371 varies in occurrence depending on season, bee age, or development.
372
373 Finally, representatives of core members of the social bee gut microbiota were recently also
374 found in bees of the distant genus *Xylocopa* (Carpenter bees) (54, 55), suggesting that these
375 bacteria may have been associated with bees before the emergence of sociality, or that they
376 have a broader and less-specific distribution across bees than previously suspected.
377
378 In summary, our results together with previous findings indicate a rather dynamic
379 evolutionary background of the core members of the social bee gut microbiota. Rather than
380 having strictly co-diversified with their hosts extended periods of host-restricted evolution
381 (and likely co-diversification in some lineages) seem to have been interrupted by host
382 switches, and independent symbiont gains and losses. Our observation that the stingless bee
383 isolates repeatedly form the most a sister group to bumble bee and honey bee isolates (for
384 *Gilliamella*, *Snodgrassella*, and *Lactobacillus* Firm-5) is intriguing given the host phylogeny. It
385 may suggest that these core members have an origin in stingless bees and then spread to the
386 other two groups, especially for *Lactobacillus* where two stingless bee isolates clades split
387 before the split between bumble bee and honey bee isolates. However, given the large

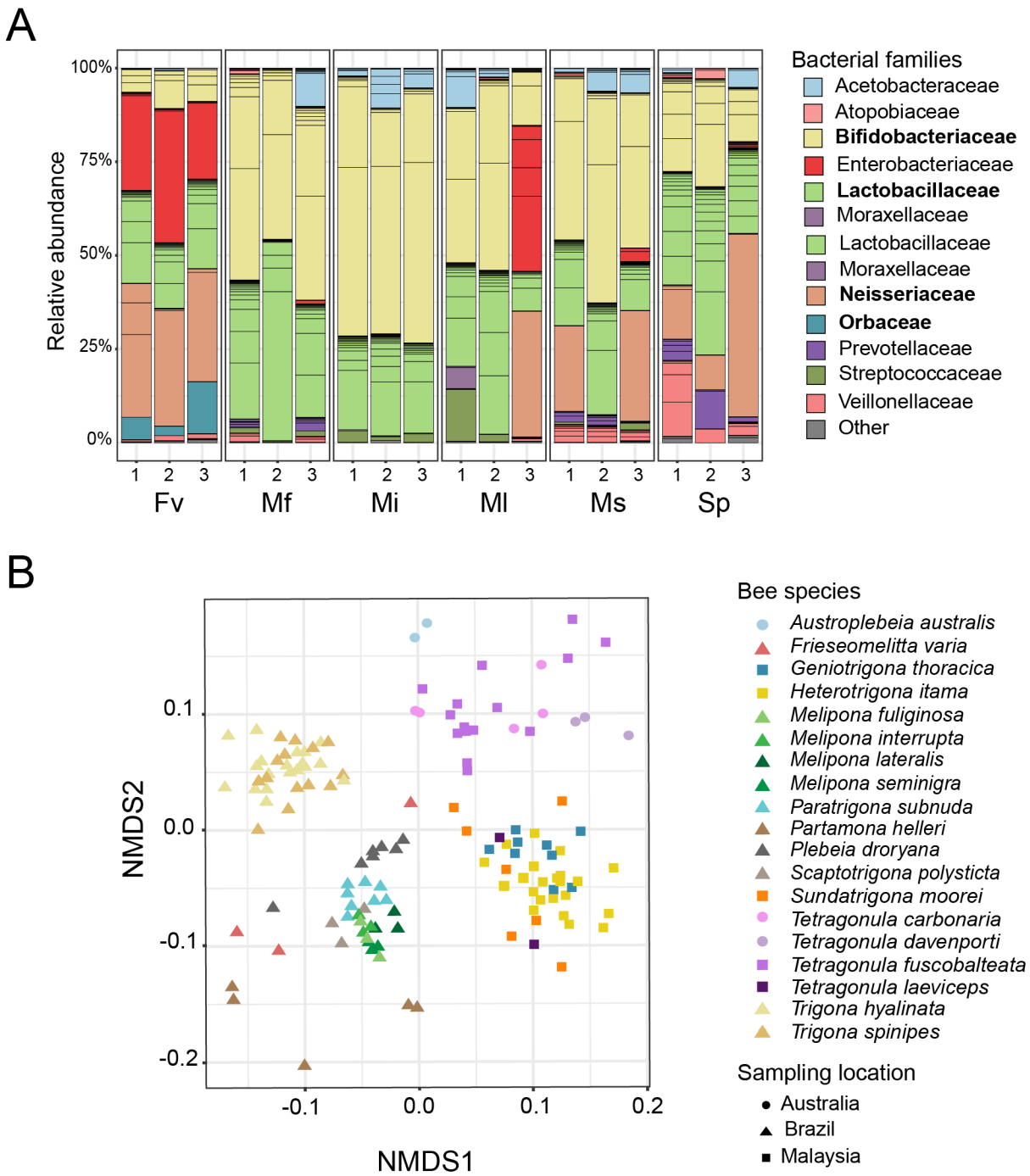
388 diversity of social and solitary bees, it is clear that the currently available datasets are
389 insufficient to explain the distribution and phylogenetic relationships of these gut symbionts
390 across hosts. Broader samplings of stingless bees and bumble bees, and genome-resolved
391 approaches are needed to fully understand the evolutionary trajectories of social bee gut
392 bacteria and to accurately reconstruct the ancestral bee microbiome composition; and formal
393 analysis should be applied to test for co-diversification.
394
395 The importance of sampling biases when assessing patterns of co-diversification between
396 hosts and their gut bacteria is highlighted by the analyses of Bacteroidaceae gut symbionts of
397 hominids. While originally reported to have co-diversified with their hosts (56), re-
398 examination with increased sampling disrupted the co-diversification pattern observed
399 earlier (57). In contrast, a recent study identified strong signals of parallel evolutionary history
400 between seven (out of 56 tested) gut bacterial taxa and human populations (58), and
401 phylogenetic congruency has also been found for certain stinkbug insects and their primary
402 gut symbiont (59). This demonstrates that co-diversification has occurred between certain
403 gut bacteria and their hosts.
404
405 Besides offering new insights into the evolution of the social bee gut microbiota, our genomic
406 analysis also revealed the functional potential of major gut symbionts of the analyzed
407 stingless bee species. All isolates of *Lactobacillus* Firm-5, *Bifidobacterium*, and *Gilliamella*
408 carried genes for the saccharolytic fermentation of diet-derived carbohydrates. In contrast,
409 *Snodgrassella* ESL0689 lacked such functions in its genome, but instead harbored genes for
410 aerobic respiration. These results are consistent with findings from honey bees and bumble

411 bees (19, 20, 23-25, 27) and hence suggest that the core microbiota members occupy similar
412 ecological niches across the three groups of social bees.
413
414 Another parallel to findings from honey bees and bumble bees was the extensive genomic
415 divergence present among strains of the core members *Lactobacillus* Firm-5, *Gilliamella*, and
416 *Bifidobacterium*, even when isolated from the same host species. Moreover, we found
417 genomic variation in carbohydrate breakdown and amino acid and nucleoside biosynthesis
418 functions among these strains. This suggests that the diversification of these bacteria has not
419 only been driven by isolation into different host species but also by the adaptation to different
420 ecological niches in the gut, similar as shown for bumble bees and honey bees (25, 30, 32,
421 60). These parallels may not be surprising as the dietary preferences of the analyzed stingless
422 bee species are similar to those of honey bees and bumble bees. In the future, it will be
423 interesting to look at the functional potential of the core microbiota members in bees with
424 different dietary habits, such as the vulture bees (35), i.e. stingless bees that feed on raw meat
425 instead of pollen yet share a subset of the core members with other social bees.
426
427 Some of the isolate genomes we sequenced in our study did not come from any of the core
428 members of the social bee gut microbiota. They may present transient community members,
429 opportunistic pathogens, or host-specific gut symbionts with complementary functions to the
430 core microbiota. Therefore, the established genomes present an important resource for
431 future research. A particular strain that drew our attention was Enterobacteriaceae ESL0689,
432 as it belonged to an ASV that was present at high relative abundance in all three colonies of
433 *Fv.* ESL0689 harbored a complete TCA cycle and a respiratory chain and could also synthesize
434 most amino-acids and co-factors indicating a similar metabolic niche as *Snodgrassella* (20).

435

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

444 **Figures**




445

446 **Figure 1. Community analysis of the gut microbiota of stingless bees. (A)** 16S rRNA gene-

447 based community profiles of the gut microbiota of three colonies of six different stingless

448 bee species collected in Brazil. Fv – *Frieseomelitta varia,* Ms – *Melipona seminigra*, Ml –

449 *Melipona lateralis*, Mf – *Melipona fuliginosa*, Mi – *Melipona interrupta*, Sp – *Scaptotrigona*

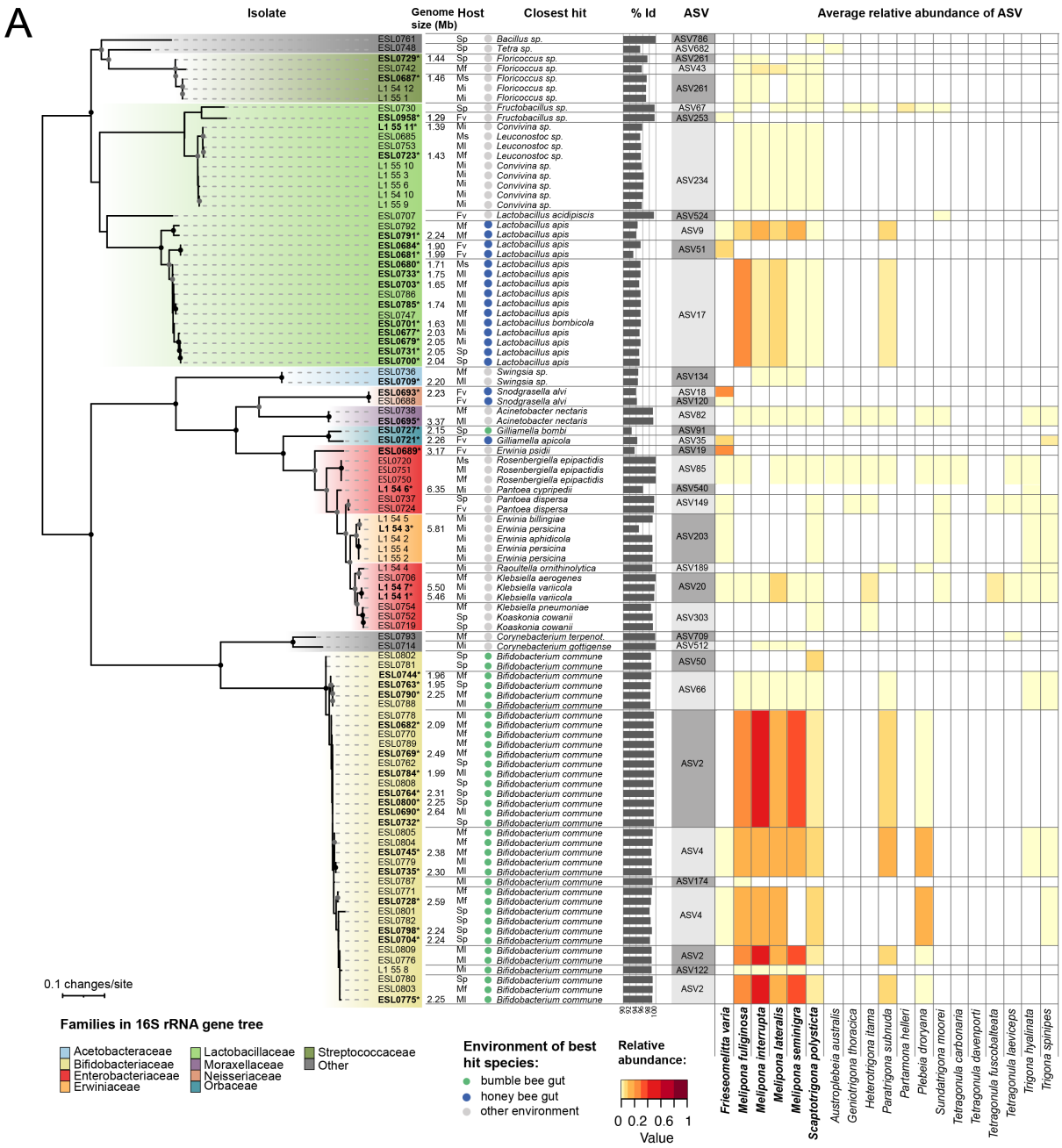
450 *polysticta*. Relative abundance of ASVs is shown. ASVs are ordered and colored at the family

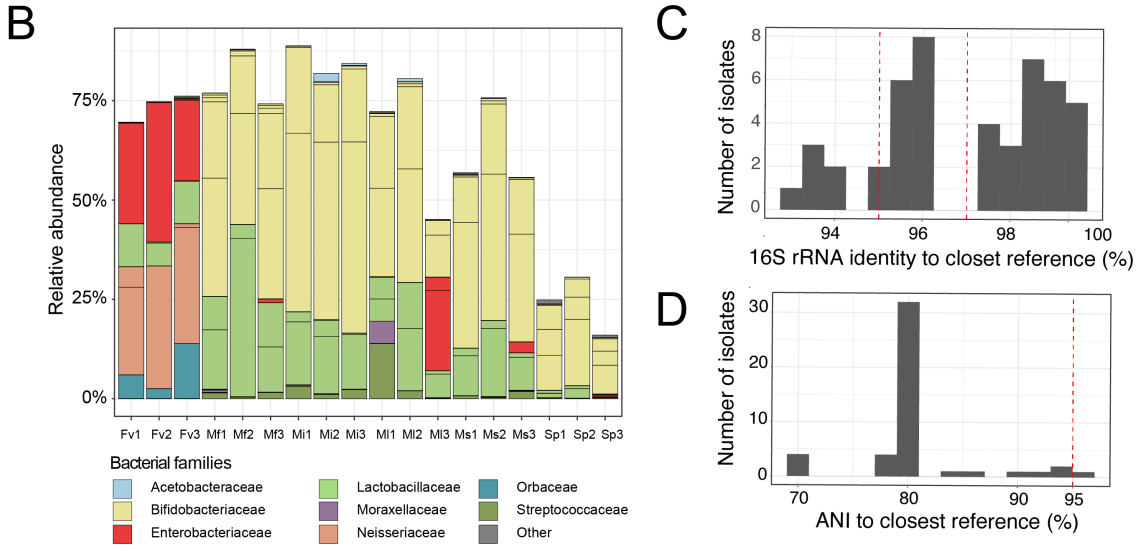
451 level according to the legend, families of core members are displayed in bold. ASVs with

452 <1% relative abundances are summed up as 'others' and shown in grey. **(B)** NMDS based on

453 ASV relative abundance (Bray-Curtis dissimilarity) across 136 16S rRNA gene amplicon

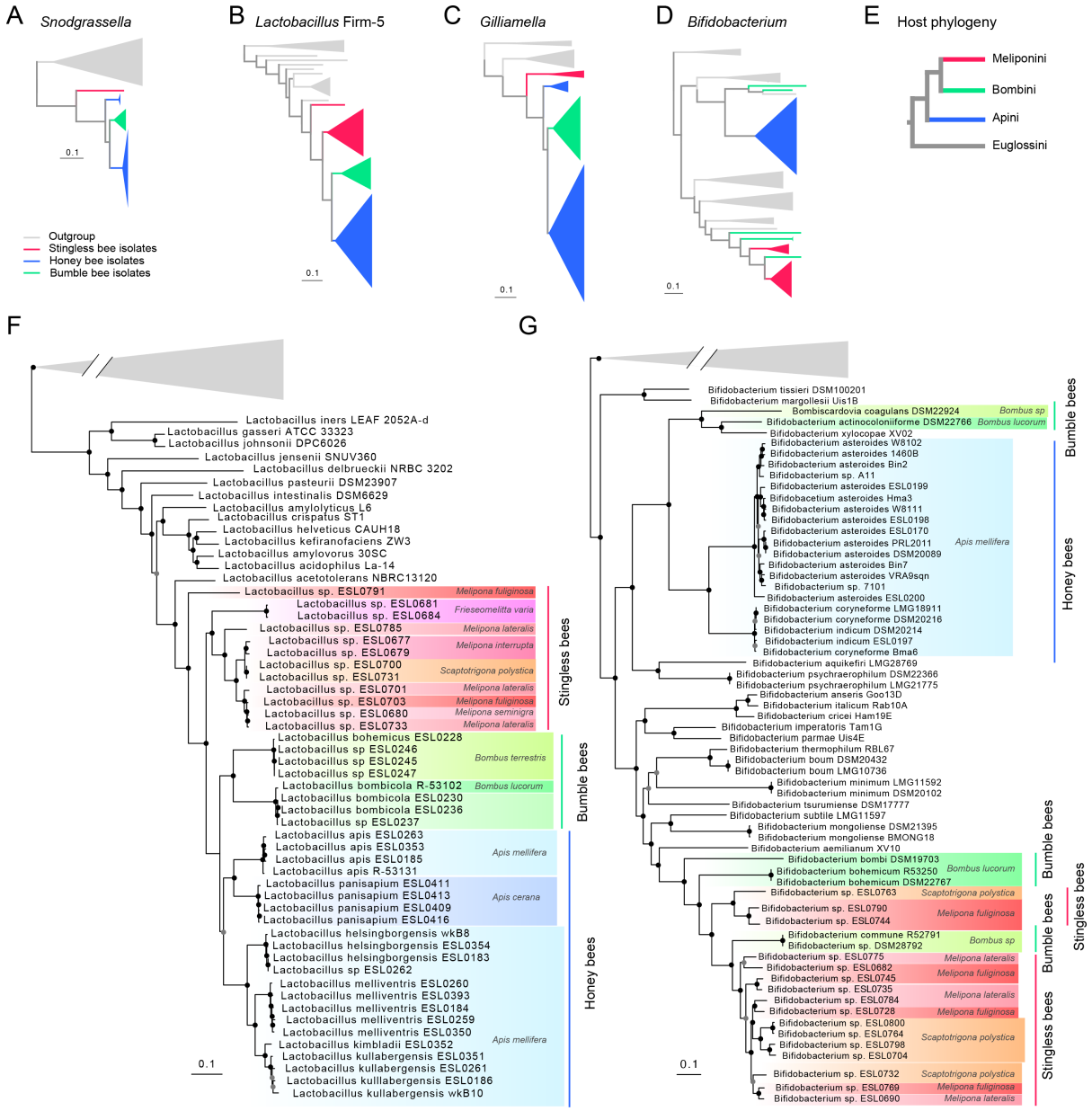
454 sequence samples including data from a previous study (17) and our study.


A
Isolate
Genome size (Mb)
Host
Closest hit
% Id
ASV
Average relative abundance of ASV
0.1 changes/site
Frieseomelitta varia
Melipona fuliginosa
Melipona interrupta
Melipona lateralis
Melipona seminigra
Scaptotrigona polysticta
Austroplebeia australis
Geniotrigona thoracica
Heterotrigona itama
Paratrigona subnuda
Partamona helleri
Plebeia droryana
Sundatrigona moorei
Tetragonula carbonaria
Tetragonula davenporti
Tetragonula fuscobalteata
Tetragonula laeviceps
Trigona hyalinata
Trigona spinipes
Families in 16S rRNA gene tree
Acetobacteraceae
Bifidobacteriaceae
Enterobacteriaceae
Erwiniaceae
Lactobacillaceae
Moraxellaceae
Neisseriaceae
Orbaceae
Streptococcaceae
Other
Environment of best hit species:
bumble bee gut
honey bee gut
other environment
Relative abundance:
0 0.2 0.6 1
Value


B
Relative abundance
75%
50%
25%
0%
Fv1 Fv2 Fv3 Mf1 Mf2 Mf3 Mi1 Mi2 Mi3 Ml1 Ml2 Ml3 Ms1 Ms2 Ms3 Sp1 Sp2 Sp3
Bacterial families
Acetobacteraceae
Bifidobacteriaceae
Enterobacteriaceae
Lactobacillaceae
Moraxellaceae
Neisseriaceae
Orbaceae
Streptococcaceae
Other
C
Number of isolates
16S rRNA identity to closet reference (%)
D
Number of isolates
ANI to closest reference (%)

455

456 **Figure 2. Bacterial strains isolated from the gut of six stingless bee species from Brazil. (A)**
457 Maximum likelihood tree was inferred from nearly complete sequences of the 16S rRNA gene
458 of each isolate. Strain names of isolates for which we sequenced the genome are indicated in
459 bold. The genome size of each sequenced strain is noted next to it. The isolate host is
460 indicated as follows: Fv – *Frieseomelitta varia,* Ms – *Melipona seminigra*, Ml – *Melipona*
461 *lateralis*, Mf – *Melipona fuliginosa*, Mi – *Melipona interrupta*, Sp – *Scaptotrigona polysticta*.
462 Closest hit indicates best BlastN hit of the 16S rRNA gene sequence against the 16S ribosomal
463 RNA (NCBI: Bacteria and Archaea type strains) database. Colored circles indicate if the strain
464 of the best hit was isolated from the gut of a bumble bee, a honey bee, or from elsewhere.
465 Bar plots indicate % identity of the best BlastN hit. The matching ASV and its average relative
466 abundance across the nineteen analyzed stingless bee species are indicated. The names of
467 the stingless bee species from our study are given in bold. Note that many isolates matched
468 the same ASV. **(B)** Relative abundances of matching ASVs in each of the 18 samples of the six
469 stingless bees species sampled in our study, i.e. representativeness of the isolates in the
470 amplicon data. **(C)** Distribution of the isolated strains based on best 16S rRNA gene identity
471 to closest reference species. These are the same values as shown in Figure 2A. The dashed
472 lines indicate 95% and 97% identity corresponding to the genus and species threshold,
473 respectively. **(D)** Distribution of the isolated strains based on pairwise ANI with the closest
474 reference genome publicly available. The dashed line indicates the species threshold at 95%
475 ANI.
476




477

478 **Figure 3. Isolates of stingless bee gut bacteria present novel species and belong to deep-**

479 **branching phylogenetic lineages. (A-D)** Simplified genome-wide maximum likelihood

480 phylogenies of *Snodgrassella*, *Lactobacillus* Firm-5, *Gilliamella*, and *Bifidobacterium* based on

481 single copy gene orthologs. All branches shown are supported by >95/100 bootstrap

482 replicates. Most of the branches have been collapsed. **(E)** Dendogram depicting the topology

483 of the social bee phylogeny (adapted from (49, 50))**. (F)** Detailed genome-wide phylogeny of

484 the genus *Lactobacillus* with bacteria belonging to the social bee-specific phylotype

485 *Lactobacillus* Firm-5 highlighted in different colors according to the host species/group. The

486 maximum likelihood tree was computed on the concatenated amino acid sequences of 355
487 single copy core genes using the substitution model LG+F+I+G4. **(G)** Detailed genome-wide
488 phylogeny of the genus *Bifidobacterium* with bacteria belonging to social bee-specific clades
489 highlighted in different colors according to the host species/group. The maximum likelihood
490 tree was computed on the concatenated amino-acid sequences of 151 single copy core genes
491 using the substitution model LG+F+I+G4.
492
493

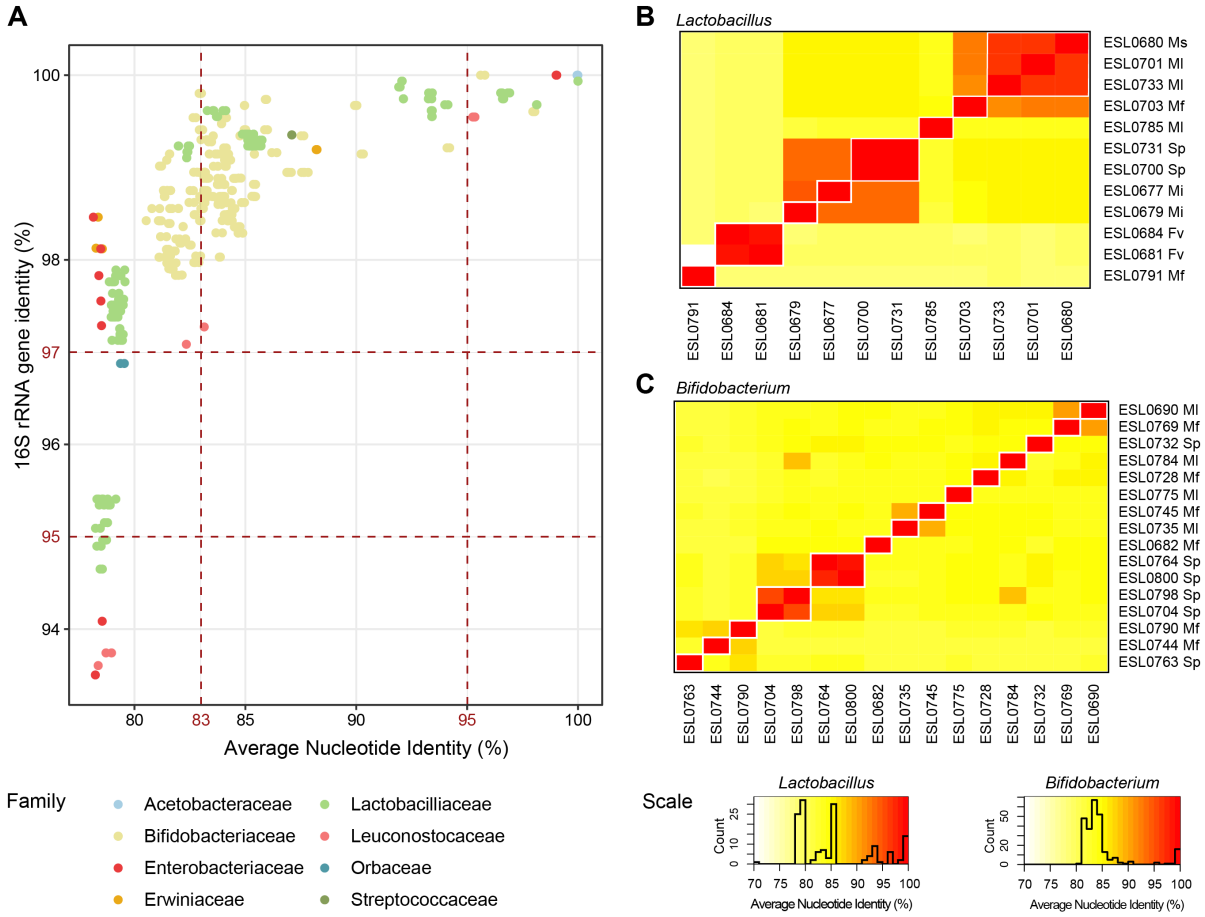



**Figure 4. Genomic divergence among isolates with high 16S rRNA gene similarity. (A)** The average nucleotide identity (ANI) versus 16S rRNA gene identity was plotted for each pair of isolate genomes belonging to the same bacterial family. The two vertical dashed bars indicate the ANI thresholds of 95% and 83% demarcating the intra-species and inter-species ANI values, respectively (61). The horizontal dashed lines delineate the commonly used species (>97%) and genus (>95%) thresholds for the 16S rRNA gene similarity. Average nucleotide identity heatmaps of **(B)** *Lactobacillus* Firm-5 isolates and **(C)** *Bifidobacterium* isolates. The 95% ANI clusters are delimited in white. The isolate host is indicated, Fv – *Frieseomelitta varia,* Ms – *Melipona seminigra*, Ml – *Melipona lateralis*, Mf – *Melipona fuliginosa*, Mi – *Melipona interrupta*, Sp – *Scaptotrigona polysticta*.

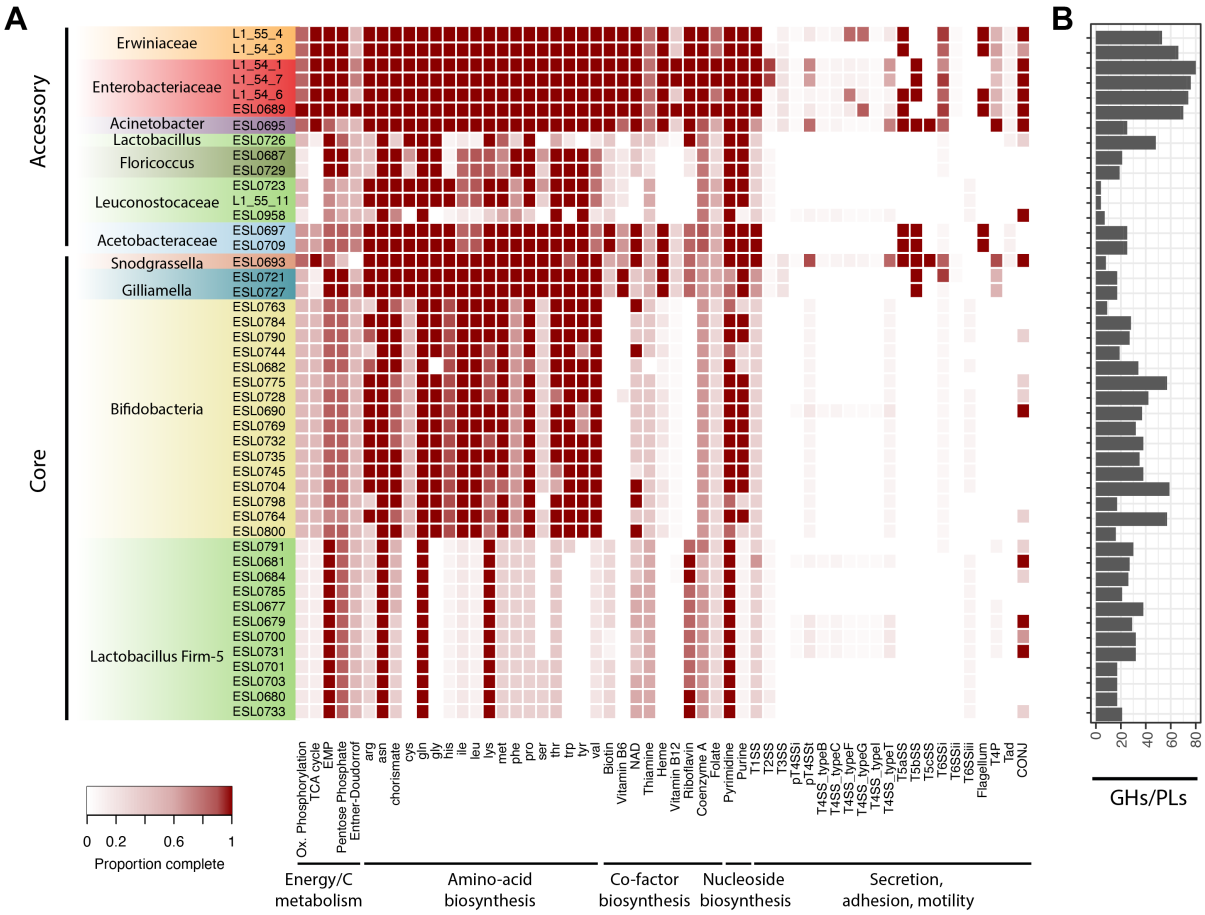
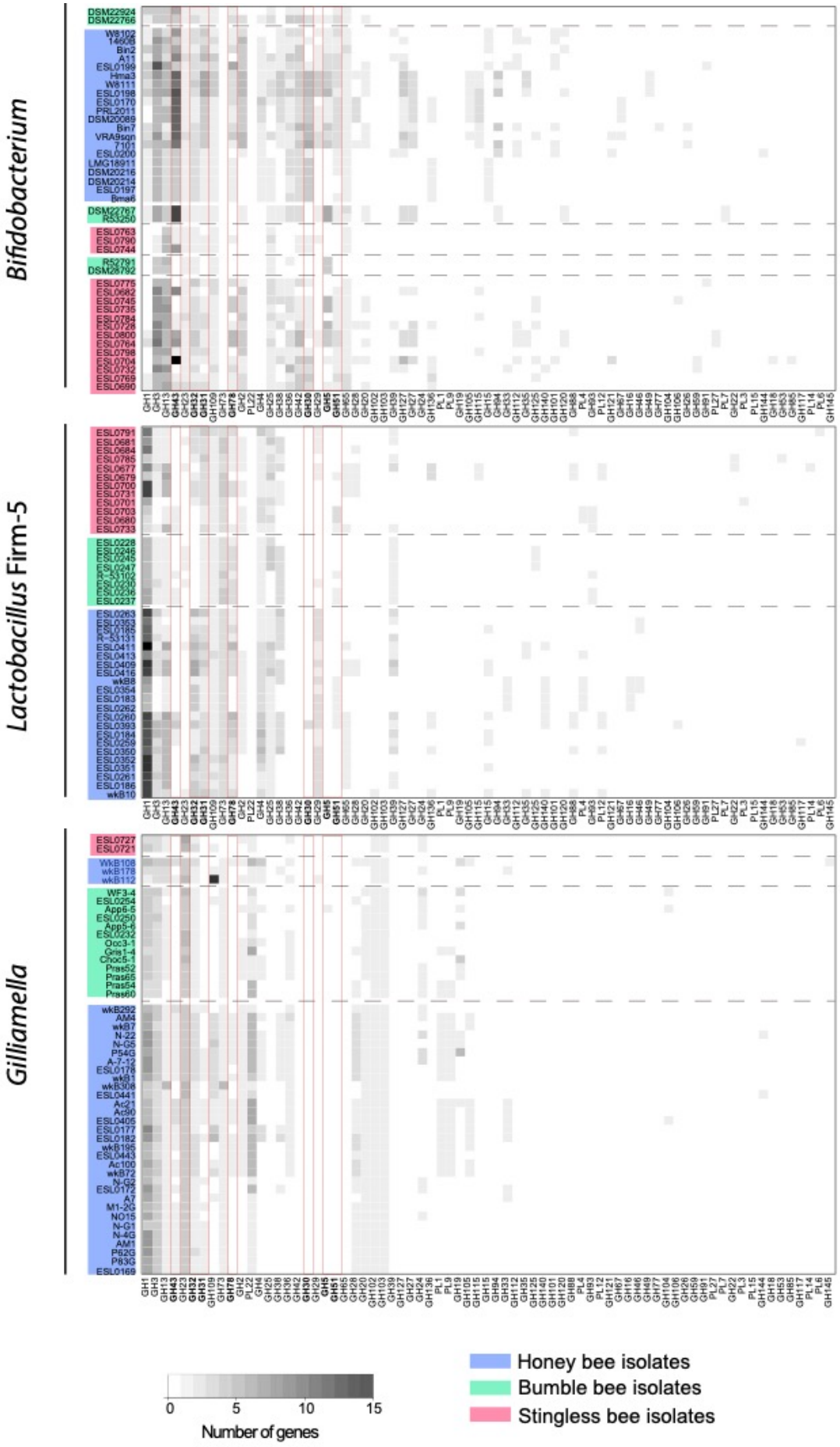



**Figure 5. Metabolic capabilities of the sequenced stingless bee gut isolates. (A)** The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates. The isolates are ordered into core and accessory members based on whether they are related to bacteria isolated from other social bees. **(B)** Total number of glycoside hydrolase (GH) and polysaccharide lyase (PL) genes in each isolate.


Bifidobacterium
Lactobacillus Firm-5
Gilliamella
Honey bee isolates
Bumble bee isolates
Stingless bee isolates
0
5
10
15
Number of genes

516

517 **Figure 6. Carbohydrate-active enzyme profiles for *Bifidobacterium, Lactobacillus* Firm-5 and**
518 ***Gilliamella* isolates.** Distribution of genes in each GH and PL families respectively for *Bifidobacterium,*
519 *Lactobacillus* Firm-5 and *Gilliamella* isolates. The isolates were sorted according to their position in
520 the phylogenies, the hosts are indicated by the background colors. Glycoside hydrolases families
521 mentioned in the text are outlined with a red square. The color scale indicates the number of genes
522 in each family.
523
524

525 ## Methods

526 **Bee sampling**

527 Bees were collected from three different nests of each of the following six Meliponini species
528 in February and March 2019: *Frieseomelitta varia*, *Scaptotrigona polysticta*, *Melipona*
529 *fuliginosa*, *Melipona interrupta*, *Melipona lateralis* and *Melipona seminigra*. All nests were
530 located in a rural meliponary in the vicinities of Iranduba municipality (Iranduba-AM, Brazil,
531 3°10'52.7"S 60°07'08.5"W). Bees were sampled at the entrance of each nest and immediately
532 immobilized by cooling at 4 °C. Then the entire gastrointestinal tract was dissected, and the
533 hindgut separated from the anterior gut parts. Hindguts of bees from the same nest were
534 pooled in two separate tubes. One pool per nest was mixed with 1x PBS and glycerol,
535 homogenized using bead-beating, and subsequently cryopreserved at -80°C. This pool was
536 used for bacterial culturing as described below. The other pool was cryopreserved at -80°C
537 without homogenization and was subsequently used for DNA extraction and 16S rRNA gene
538 analysis. For the sample used for sequencing, 20 bees per nest were pooled, while for the
539 sample for culturing, only 3 bees per nest were pooled. For *Frieseomelitta varia* and
540 *Scaptotrigona polysticta*, we pooled 40 and 10 bees, respectively, due to the small size of
541 these bee species.

542

543 **Bacterial culturing**

544 For establishing a culture collection of primary isolates from the gut of the six sampled bee
545 species, serial dilutions of the cryopreserved homogenates were plated on eight different
546 media: CBA (Columbia Blood Agar supplemented with 5% defibrinated sheep blood
547 (Thermofisher)), MRSA (De Man, Rogosa and Sharpe agar) supplemented with Fructose (2%)
548 and L-Cysteine (0.1%), MRSA supplemented with Mannitol (2%), Chocolate Agar, TSA

549 (tryptone soya agar), TYG (tryptone glucose yeast extract Agar), GC, LBA (Luria Bertani agar)
550 without NaCl, BHIA (brain heart infusion agar) and SDA (Sabouraud dextrose agar). Plates
551 were incubated in two different conditions: in a microaerophilic 5% $CO_2$-enriched atmosphere
552 and in an anaerobic chamber (72% $N_2$, 8% $H_2$, 20% $CO_2$), both at 34°C. After 2-7 days of
553 incubation, colonies of different size and appearance were picked and re-grown on the same
554 media and culturing condition. Cryo-stocks of bacterial strains of interest were prepared by
555 harvesting bacterial biomass in liquid media corresponding to the solid growth media and
556 supplemented with 20% glycerol. For DNA isolation, bacteria were grown from the stocks,
557 and a single colony was picked and re-grown on fresh media before harvesting bacterial
558 biomass.
559
560 **Genotyping**
561 All colonies that were selected for culturing were genotyped by PCR and Sanger sequencing
562 of a 16S rRNA gene fragment. To this end, a small amount of bacterial material was
563 transferred to a lysis buffer (Tris-HCl 1M pH=7.5, EDTA 0.5M, SDS 10%) containing 2.5 µl
564 lysozyme (20mg/mL) and 2.5 µl proteinase K (20mg/mL), and incubated for 10 min at 37°C,
565 20 min at 55°C, and 10 min at 95°C. PCR was performed with universal bacterial primers that
566 amplify the V1-V5 region of the 16S rRNA gene (27F - AGRGTTYGATYMTGGCTCAG and 907R-
567 CCGTCAATTCMTTTRAGTTT) using the following reagents and thermocycler program: initial
568 denaturing at 94°C for 5 min, followed by 32 cycles of denaturing at 94°C for 30 sec, annealing
569 at 56°C for 30 sec, and extension 72°C for 1 min, and a final extension at 72°C for 7 min. PCR
570 results were checked on a 1% agarose gel. PCR reactions selected for Sanger sequencing were
571 purified using ExoSAP-IT™ (1µl ExoSAP 5x, 4µl $ddH_2O$) with the thermocycler program: 30 min
572 at 37°C followed by 15 min at 80°C. Purified samples were then sent to Eurofins® for

573 sequencing. Sanger sequences were analyzed with Geneious suite (Geneious®) and compared
574 to GenBank at NCBI using BLAST tools (62).
575

576 **DNA isolation and genome sequencing**

577 DNA isolation for Illumina sequencing was carried out using a customized SPRI bead-based
578 extraction method or the FastPure Bacteria DNA Isolation Mini Kit (Vazyme). For the SPRI-
579 bead method, bacteria were harvested and resuspended in tubes containing 200 mg of 0.1
580 mm-sized acid washed glass beads and 200 µl of TER buffer (10 mM Tris-HCl, 1 mM EDTA, 100
581 µg /ml RNAse A, pH8.0). Samples were homogenized using a FastPrep-25 5G instrument (2
582 rounds of 30 sec with the power set to 6) and subsequently centrifuged at max speed for 10
583 min at room temperature. 40 µl of SPRI beads were added to 100 µl of supernatant,
584 immediately mixed thoroughly by repeated pipetting (>20x) and incubated for 5 min at room
585 temperature. After placing the tubes on a magnet stand the liquid was removed and
586 discarded and the beads washed two times with 200 µl 80% ethanol. After air drying the tubes
587 on the magnetic stand 22 µl of 5 mM Tris-HCl pH8 was added. For isolating bacteria with the
588 FastPure Bacteria DNA Isolation Mini Kit the manufacturer's protocol for gram-positive
589 bacteria was followed.
590 DNA isolation for Oxford Nanopore Technologies (ONT) sequencing was carried out using a
591 custom DNA extraction protocol for Gram-positive bacteria. Tubes were prepared with glass
592 beads and 160 µl of buffer P1 (Qiagen). Then bacteria were harvested and resuspended in
593 these tubes by intensive vortexing. Lysozyme (20 µl, 100 mg/ml) was added and after gentle
594 mixing, tubes were incubated at 56°C with shaking at 600 rpm for 30 min. Then 4 µl RNase A
595 (100 mg/ml) were added to the tubes followed by 150 µl of buffer AL (lysis buffer, Qiagen).

596 After mixing by vortexing, tubes were incubated in a thermomixer (37°C, 900 rpm) for 20 min.
597 Tubes were centrifuged 10 min at 14’000 rpm to pellet the beads, the supernatant was
598 transferred to new tubes with 35 µl Na-acetate and 270 µl isopropanol and mixed by
599 inverting. Following an incubation of 1 h at 4°C, DNA was pelleted by centrifugation at 14’000
600 rpm for 10 min at 25°C. The supernatant was discarded, and the pellets were washed with 1
601 ml EtOH 80%. After a second centrifugation (14’000 rpm, 10 min at 25°C) the ethanol was
602 removed, and the pellet left to dry at room temperature. DNA pellets were solubilized with
603 50 µl TER (10 mM Tris-HCl, 1 M EDTA, pH 8.0, 2 mg/ml RNase A) and tubes incubated at 37°C
604 for 15 min. The solution was then transferred to PCR tubes. 40 µl NGClean beads were added,
605 and the solution mixed by repeated pipetting. After a 5 min incubation, PCR tubes were
606 placed on magnetic stands. When the solutions were clear, the liquid was removed, and the
607 beads washed two times with 200 µl of 80% EtOH. Upon complete drying, beads were
608 resuspended with 22 µl of 5mM Tris-HCl (pH 8.0). Finally, tubes were placed again on the
609 magnetic stand and when the solution were clear, the supernatant was transferred to new
610 1.5 ml Eppendorf tubes.
611 Illumina sequencing libraries were prepared using the Nextera DNA Flex Library Preparation
612 kit following instructions of the Illumina Reference Guide. This was followed by indexing,
613 dilution, and denaturation according to the ‘Index Adapters Pooling Guide’ and the ‘MiniSeq
614 System Denature and dilute Libraries guide’. Libraries were checked and quantified using a
615 dsDNA fluorescent dye, before loading them on a MiniSeq High-output flowcell (150PE). ONT
616 libraries were prepared using the ligation-based approach (LSK109). The sequencing was
617 conducted on a ONT MinION for a duration of 72 hours with the high-accuracy base-calling
618 using Guppy (*v5.0.11*).
619

620 **Genome assembly**

621 Forty-six isolates from the stingless bees were sequenced with Illumina MiniSeq (150PE). Raw
622 reads were checked using FastQC *v0.11.9* (63) and trimmed by Trimmomatic v0.39 (64) using
623 parameters: PE -phred33 AllIllumina-PEadapters.fa:3:25:7 LEADING:9 TRAILING:9
624 SLIDINGWINDOW:4:15 MINLEN:60. For *de novo* assembly we used SPAdes (–careful option,
625 *v3.15.2*) (65). For assembly quality control, reads were mapped back against the assembly
626 with BWA *v0.7.17 (66)*, samtools *v1.12* (67) and plotted with R *v4.1.1*. Genomes completeness
627 was evaluated with checkM *v1.0.13* (68).

628 Additionally, 26 isolates were also sequenced with Oxford Nanopore to produce long reads.
629 Nanopore long reads were filtered with Filtlong *v0.2.0* (https://github.com/rrwick/Filtlong)
630 for a minimum length of 7'000 and minimum mean q-score of 10 (min_length 7000,
631 min_mean_q 10, length_weight 10). ONT-based assemblies were computed with Flye *v2.7.1*
632 (69) over 5 iterations. Graphmap *v0.5.2* (70) and Racon *v1.0.1* (71) were used to perform two
633 rounds of polishing. Finally, the racon-corrected assembly and Illumina reads were fed to
634 Pilon *v1.24* (72) for single base and indels corrections.

635

636 **Genome annotation and analysis**

637 Genomes were annotated with Prokka *v 1.13* (73). Phylogenies were computed for the
638 bacterial families for which we had an isolate. For each family, we identified a set of closely
639 strains and outgroup taxa and retrieved their genomes from NCIB and IMG/Mer (74). All
640 genomes were re-annotated with Prokka to ensure annotations consistency. Gene orthology
641 was inferred with OrthoFinder *v2.3.8 (75)*. Single copy ortholog genes were selected and their
642 amino acid sequences were aligned (mafft *v7.453) (76).* An in-house script was used to trim
643 the alignments by removing positions with more than 50% gaps and sequences belonging to

644 the same genome were concatenated to produce a core gene alignment. This alignment was
645 used to infer the maximum likelihood phylogeny using IQTree (*v1.7.beta17* , -st AA -bb 1000
646 -seed 12345 -m TEST) (77). For each phylogeny, the best evolutionary model was chosen
647 according to BIC: *LG+I+G4* – Enterobacteriaceae ; *LG+F+I+G4* – Acetobacteraceae,
648 *Bifidobacterium*, *Lactobacillus*, Leuconostocaceae, Moraxellaceae, Neisseriaceae and
649 Streptococcaceae ; and *JTTDMut+F+I+G4* – Orbaceae. Branch support of the trees were
650 inferred using 1000 ultrafast bootstrap (UFBoot) repetitions. Clades can be trusted when
651 UFBoot values are >95%.
652
653 **16S rRNA community profiling**
654 Region V4 of the 16S rRNA gene was amplified using the primers 515F-Nex and 806R-Nex
655 (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCCAGCMGCCGCGGTAA,
656 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT). The primers
657 include Nextera XT index adapter sequences and the primers for the 16S rRNA V4 region (78)
658 described by Kešnerová et al. (2020). The two step PCR was performed as following: the first
659 PCR was performed using 12.5 µL of 2× Phanta Max Master Mix (Vazyme, Nanjing, China), 5 µl
660 of MilliQ water, 2.5 µl of each primer (5 µM), and 2.5 µl of template DNA for a total volume
661 of 25 µl. The PCR program started with a denaturation step at 98°C for 30sec, followed by 25
662 cycles of amplification (10 sec at 98°C, 20 sec at 55°C, and 20 sec at 72°C) and a 5 min final
663 extension step at 72°C. The PCR products were verified by 2% agarose gel electrophoresis,
664 then purified with clean NGS purification-beads in a 1:0.8 ratio of PCR product to beads and
665 eluted in 27 µl Tris (10mM, pH 8.5). A second PCR step was performed to append the unique
666 dual indexes to each sample in a total volume of 25 µl using 12.5 µL of 2× Phanta Max Master
667 Mix (Vazyme, Nanjing, China), 5 µl of MilliQ water, 2.5 µl of Nextera XT index primers 1 and 2

668 (Nextera XT Index kit, Illumina) and 2.5 µl of templated DNA. The PCR program started with a
669 denaturation step at 95°C for 3 min, followed by 8 cycles of amplification (30 sec at 95°C, 30
670 sec at 55°C, and 30 sec at 72°C), and a 5 min final extension step at 72°C. The libraries were
671 then cleaned usingClean NGS purification beads (1:1.1 ratio of PCR product to beads) and
672 were eluted in 27.5 µl Tris (10mM, pH 8.5). Prior to sequencing, the PCR product
673 concentrations were quantified by PicoGreen and pooled in equimolar concentrations; the
674 negative controls and blank extractions were pooled in equal volume. Sequencing was
675 performed on an Illumina MiSeq sequencer (2 x 250bp) by the Genomic Technology Facility
676 of the University of Lausanne. We followed DADA2 (79) pipeline to analyze the sequencing
677 data. Sequence quality control was performed with DADA2 integrated function
678 *plotQualityProfile*. Data from the three sequencing runs were processed independently for
679 the filtering, dereplication, the error rates calculation and the sample inference. After
680 merging the denoised forward and reverse pairs, we merged the three sequence tables
681 (*mergeSequenceTables*) and we applied the *collapseNoMismatch* function to unite similar
682 ASVs with shifts or length variation. We selected sequences in range 250-256bp and removed
683 chimeric sequences. Silva non-redundant SSU database v138.1 (79) was used for the
684 taxonomic assignment of the ASVs. We filtered out 62 ASVs for which the taxonomic
685 assignment was matching 'Eukaryota', 'Chloroplast' or 'Mitochondria'. Finally, we removed
686 samples with less than 5,000 reads. Non-metric MultiDimensional Scaling plots were
687 computed in R using Bray-Curtis distances (Phyloseq: *ordinate*). The adonis package was used
688 to compute permanovas.
689
690 **Functional Profiler**

691 We developed a genomic profiler using several software to annotate the genomes and
692 compute the completeness of key pathways and functions. The genomes were re-annotated
693 with GhostKOALA (80) to get KEGG annotations of genes. We created a set of rules to define
694 selected energy and metabolism pathways, and co-factors/nucleoside biosynthesis pathways
695 and computed their completeness (steps present relative to total number of steps needed)
696 with an in-house script. To obtain the completeness of amino acid biosynthesis pathways, we
697 used GapMind (81). Secretion systems and related appendages were detected in the
698 genomes by MacSyFinder's TXSScan module (82). Finally, we ran dbCan (83) for the
699 annotation of carbohydrate-active enzymes.
700

## Acknowledgements

We would like to thank Julien Marquis and the team from the Lausanne Genomics Technology Facility and Alban Ramette and his team from the University of Berne for carrying out the Illumina sequencing and Oxford Nanopore sequencing, respectively. We also would like to 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. Finally, we would like to thank all SAGE students of the years 2019-2020 and 2020-2021 for their contribution to this course: 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 Lizuka, 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 the ERC Starting Grant (MicroBeeOme), the NCCR Microbiomes, a National Centre of Competence in Research, funded by 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 nr 315230_184908: SOMETALP).

724 **Data and materials availability:**

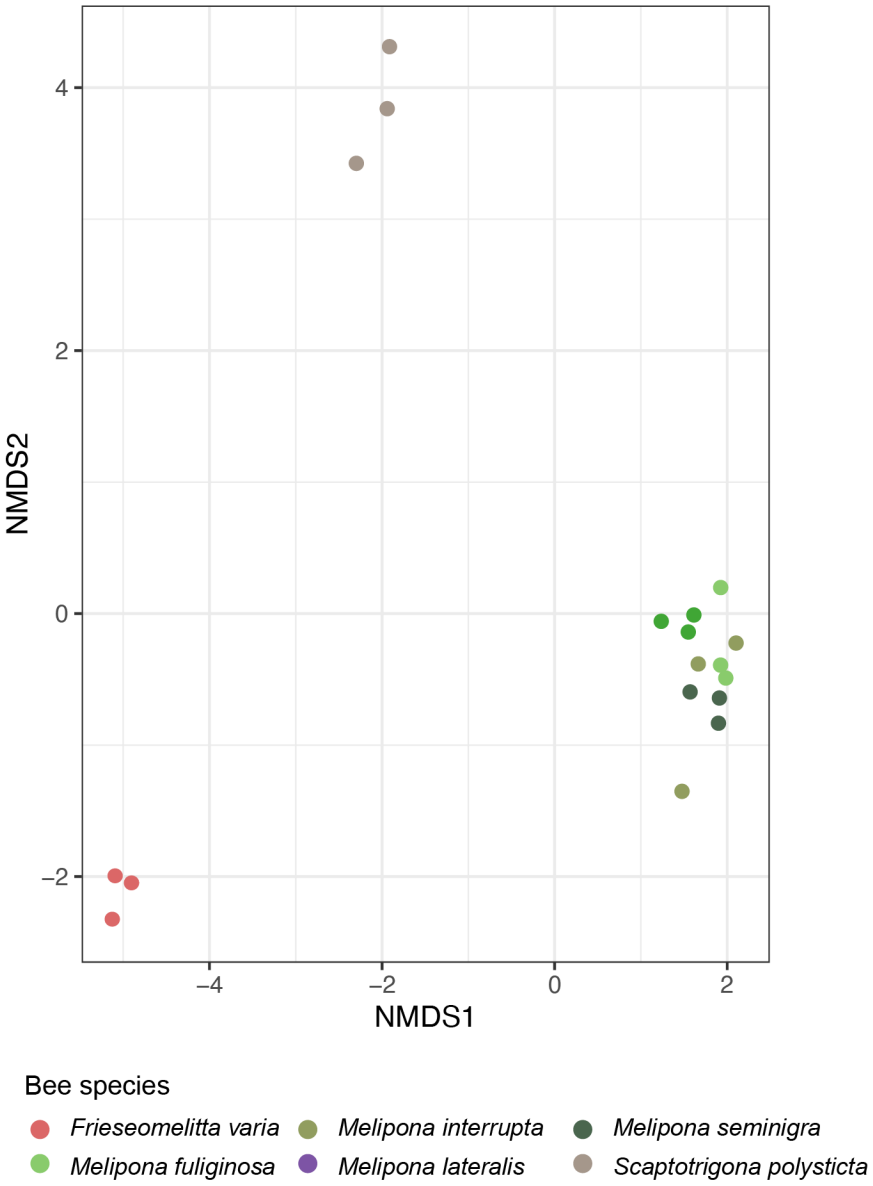
725 The bacterial genomes and sequencing data are available in the NCBI's BioProjects
726 PRJNA906295 and PRJNA897721. Scripts used for the amplicon sequencing analysis, the
727 phylogenies and the metabolic profiler can be found on GitHub
728 https://github.com/gsartonl/Publication_Sarton-Loheac_2022.

729

730

731 **Supplementary Figures and Tables**

732



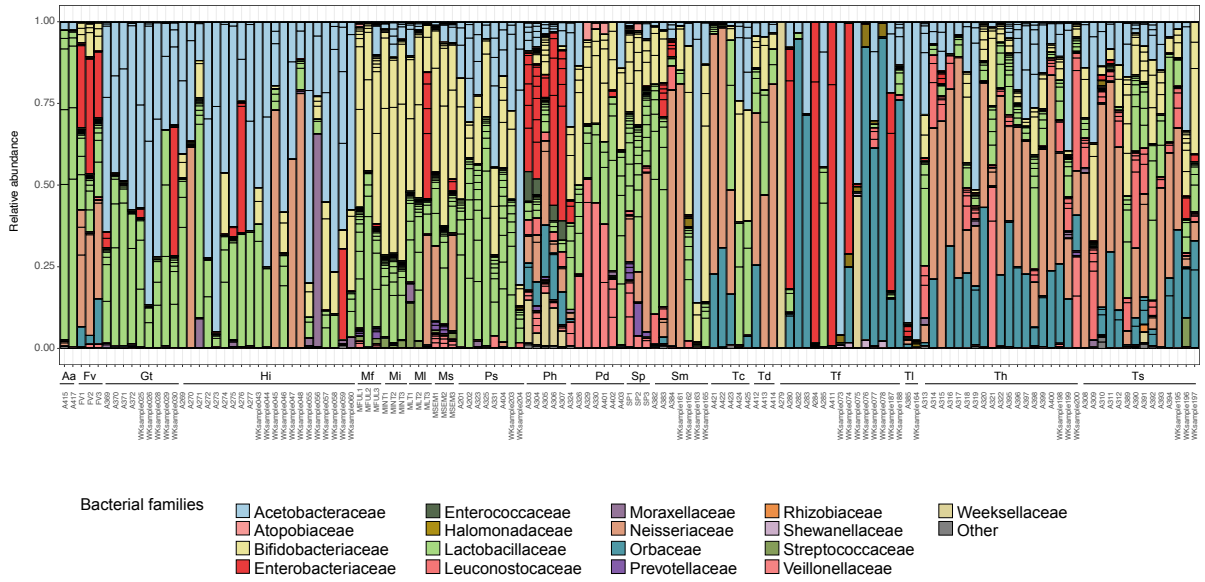

733

734 **Supplementary Figure 1.** NMDS based on ASV relative abundance in the 18 samples

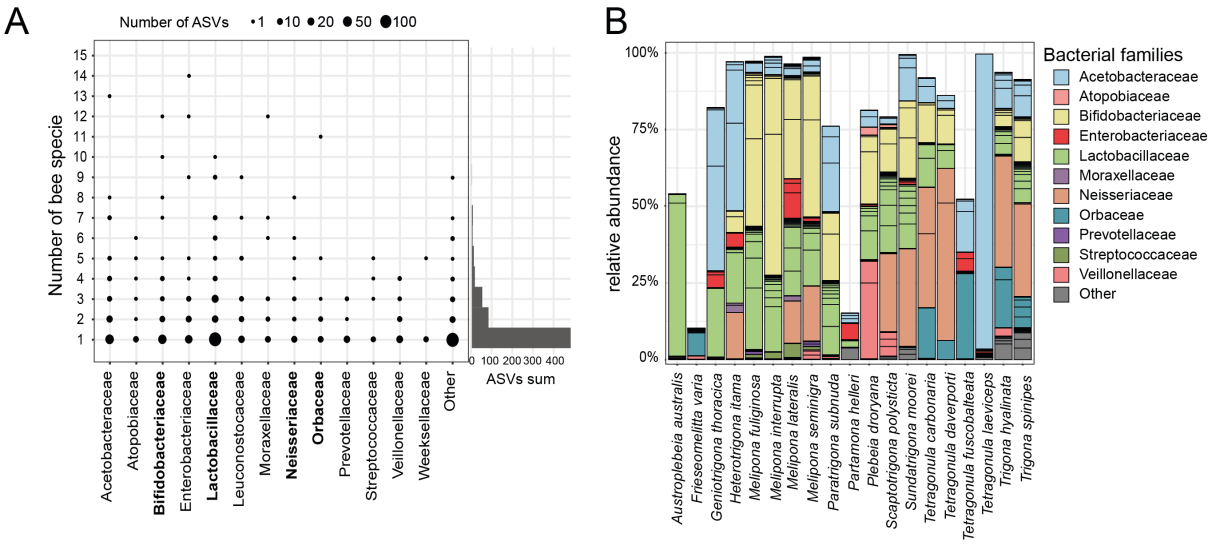
735 collected from six different stingless bee species in Brazil. The samples clustered by bee

736 genus (PERMANOVA**:** 'Genus' pseudo F=11.05, p-value< 0.001).

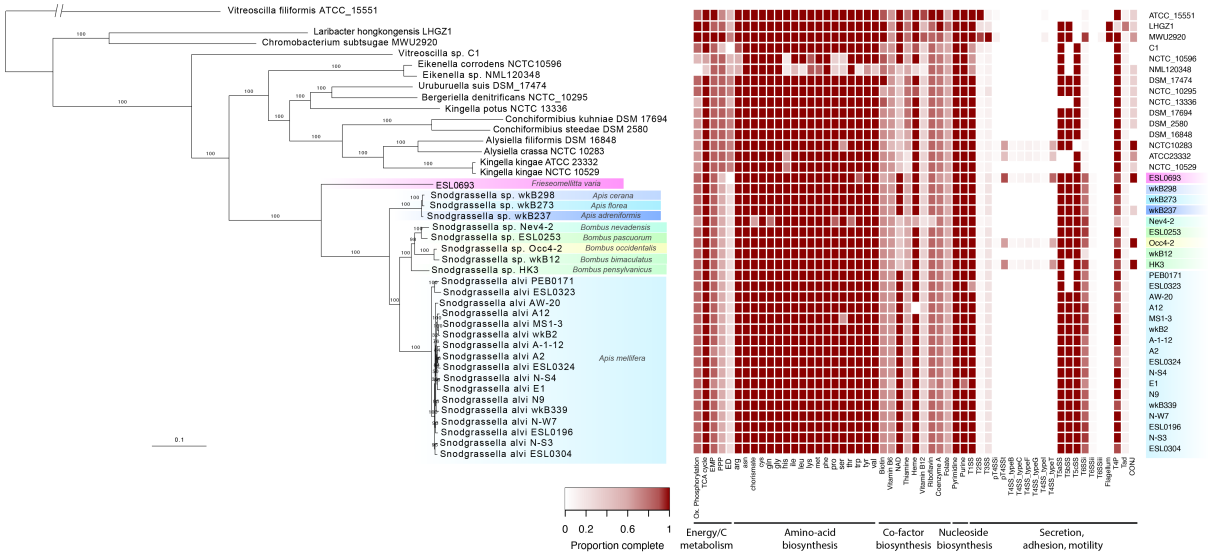
737




**Supplementary Figure 2.** Community profiles of the gut microbiota of 136 samples from 19 different stingless bee species collected in Brazil, Malaysia, and Australia. Relative abundance of ASVs is shown. ASVs are ordered and colored at the genus level according to the legend. ASVs with <1% relative abundances are summed up as 'others' and shown in grey. This plot includes the datasets from stingless bees from (17) and our study. The samples are ordered by bee species : Aa – *Austroplebeia australis*, Fv – *Frieseomelitta varia*, Gt – *Geniotrigona thoracica* , Hi – *Heterotrigona itama*, Mf – *Melipona fuliginosa*, Mi – *Melipona interrupta*, Ml – *Melipona lateralis*, Ms – *Melipona seminigra*, Ps – *Paratrigona subnuda*, Ph – *Partamona helleri*, Pd – *Plebeia droryana*, Sp – *Scaptotrigona polysticta*, Tc – *Tetragonula carbonaria*, Td – *Tetragonula davenporti*, Tf – *Tetragonula fuscobalteata*, Tl – *Tetragonula laeviceps*, Th – *Trigona hyalinata*, Ts – *Trigona spinipes*.




**Supplementary Figure 3. (A)** Number of shared ASVs. For each family, we reported how many ASVs were observed (indicated by the dot size) in how many bee species. Families with an average abundance <1% are grouped as 'Other'. The sum of species-specific or shared (by two or more bee species) ASVs is reported on the right side. Core members families are highlighted in bold. **(B)** Relative abundances of the shared ASVs across the 19 stingless bee species (samples from the same species were pooled).




760

761

762 **Supplementary Figure 4.** Core genome phylogeny of the Neisseriaceae including phylotype

763 *Snodgrassella*. The maximum likelihood tree was computed with the amino-acid sequences

764 of 428 concatenated single copy core genes using LG+F+I+G4 substitution model. Bee

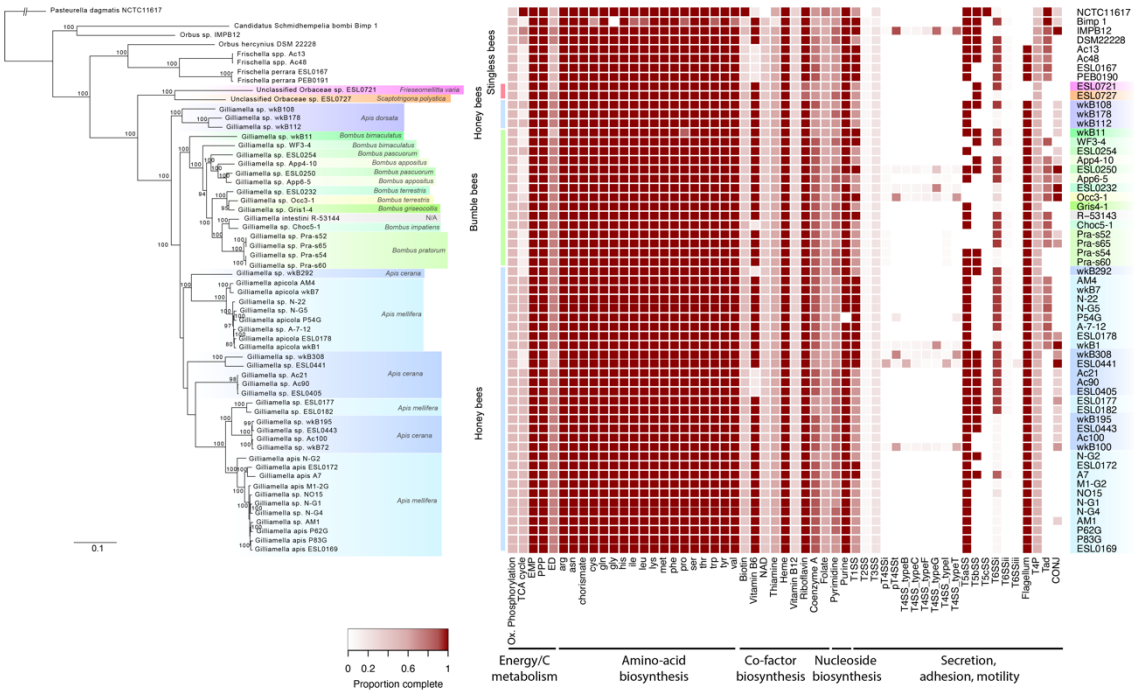
765 isolates form a monophyletic clade. The heatmap indicates the genomic completeness of

766 major metabolic pathways and functions related to energy and carbon metabolism (EMP:

767 Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as

768 well as secretion, adhesion, and motility across the sequenced bacterial isolates.

769



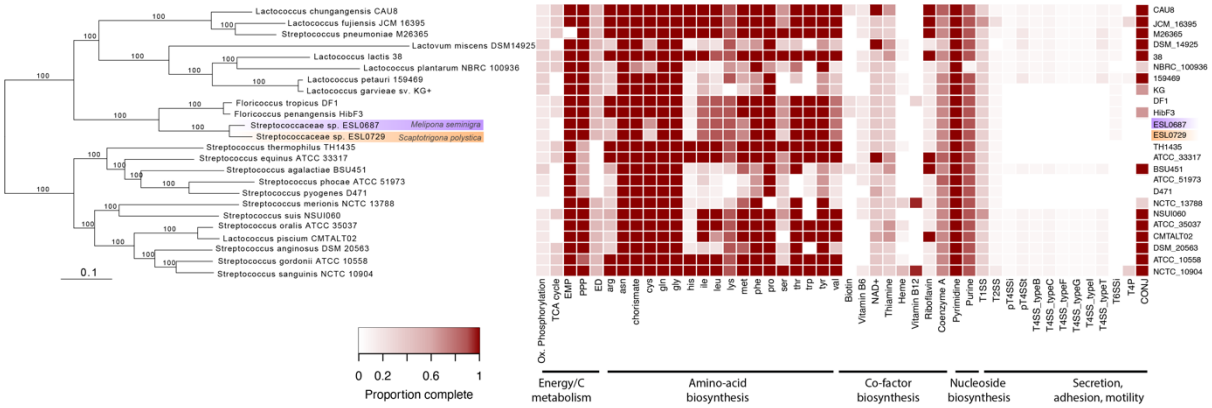

**Supplementary Figure 5.** Core genome phylogeny of the Orbaceae including *Gilliamella* phylotype. The maximum likelihood tree was computed on the concatenated amino-acid sequences of 782 single copy core genes using JTTDCMut+F+I+G4 substitution model. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.

Roseomonas gilardii U14-5
Acidiphilium cryptum JF-5
Rhodopila sp. LVNP
Acidisphaera rubrifaciens HS-AP2
Acidocella facilis ATCC35904
Acidibrevibacterium fodinaquatile G45-3
Granulibacter bethesdensis CGDNIH2
Commensalibacter intestini A911 (Drosophila melanogaster isolate)
Commensalibacter sp. MX-MONARCH01 (Danaus plexippus isolate)
Commensalibacter sp. ESL0366
Commensalibacter sp. ESL0379
Commensalibacter sp. ESL0392
Commensalibacter sp. ESL0382
Commensalibacter sp. ESL0390
Commensalibacter sp. ESL0367
Commensalibacter sp. ESL0284
Commensalibacter sp. AMU001
Apis mellifera
Honey bees
Gluconacetobacter diazotrophicus PA1 5
Gluconacetobacter liquefaciens DSM 5603
Komagataeibacter nataicola RZS01
Komagataeibacter medellinensis NBRC 3288
Komagataeibacter saccharivorans JH1
Komagataeibacter cocois WE7
Komagataeibacter hansenii NQ5
Komagataeibacter xylinus ATCC 53582
Acetobacter aceti ATCC 23746
Acetobacter nitrogenifigens DSM 23921
Acetobacter ghanensis LMG 23848
Acetobacter pasteurianus IFO 3283 01
Acetobacter pomorum SH
Acetobacter senegalensis 108B
Acetobacter malorum LMG 1746
Tanticharoenia sakaeratensis NBRC 10319
Acidomonas methanolica DSM 5432
Neoasaia chiangmaiensis NBRC 101099
Asaia platycodi SF2 1
Asaia sp. As1742
Asaia prunellae JCM 25354
Swaminathania salitolerans NBRC 104436
Kozakia baliensis NBRC 16680
Swingsia samuiensis AH83
Gluconobacter albidus TMW2 1191
Gluconobacter oxydans DSM 3504
Gluconobacter morbifer G707
Gluconobacter frateurii NBRC 103465
Neokomagataea thailandica NBRC 106555
Acetobacteraceae sp. AAB2-BBS1
Acetobacteraceae sp. AAB2-BBS3
Oecophyllibacter saccharovorans Ta1
Oecophyllibacter saccharovorans Jb2
Oecophyllibacter saccharovorans Ha5
Acetobacteraceae ESL0697
Melipona lateralis
Stingless bees
Saccharibacter sp. EH60
Saccharibacter sp. EH70
Saccharibacter sp. EH611
Saccharibacter sp. 17LHSD
Saccharibacter floricola DSM 15669
Bombella apis ESL0368
Bombella apis ESL0387
Parasaccharibacter apium G7_7_3c
Bee queen gut isolate ESL0380
Saccharibacter sp. M18
Bombella apis SME1
Parasaccharibacter apium A29
Saccharibacter sp. AM169
Parasaccharibacter apium C6
Saccharibacter sp. 3.A.1
Parasaccharibacter apium B8
Bombella intestini R-52487
Bombus lapidarius
Bombella apis ESL0378
Bombella apis ESL0385
Parasaccharibacter apium AS1
Honey bees
Bumble bees
0.1
Roseomonas_gilardii_U14-5
Acidiphilium_cryptum_JF-5
Rhodopila_sp_LVNP
Acidisphaera_rubrifaciens_HS_AP2
Acidocella_facilis_ATCC_35904
Acidibrevibacterium_fodinaquatile_G45-3
Granulibacter_bethesdensis_CGDNIH2
Commensalibacter_intestini_A911
Commensalibacter_sp_MX_MONARCH01
Commensalibacter_sp_ESL0366
Commensalibacter_sp_ESL0379
Commensalibacter_sp_ESL0392
Commensalibacter_sp_ESL0382
Commensalibacter_sp_ESL0390
Commensalibacter_sp_ESL0367
Commensalibacter_sp_ESL0284
Commensalibacter_sp_AMU001
Gluconacetobacter_diazotrophicus_PA1_5
Gluconacetobacter_liquefaciens_DSM_5603
Komagataeibacter_nataicola_RZS01
Komagataeibacter_medellinensis_NBRC_3288
Komagataeibacter_saccharivorans_JH1
Komagataeibacter_cocois_WE7
Komagataeibacter_hansenii_NQ5
Komagataeibacter_xylinus_ATCC53582
Acetobacter_aceti_ATCC_23746
Acetobacter_nitrogenifigens_DSM_23921
Acetobacter_ghanensis_LMG_23848
Acetobacter_pasteurianus_IFO_3283_01
Acetobacter_pomorum_SH
Acetobacter_senegalensis_108B
Acetobacter_malorum_LMG_1746
Tanticharoenia_sakaeratensis_NBRC_10319
Acidomonas_methanolica_DSM_5432
Neoasaia_chiangmaiensis_NBRC_101099
Asaia_platycodi_SF2_1
Asaia_sp_As1742
Asaia_prunellae_JCM_25354
Swaminathania_salitolerans_NBRC104436
Kozakia_baliensis_NBRC_16680
Swingsia_samuiensis_AH83
Gluconobacter_albidus_TMW2_1191
Gluconobacter_oxydans_DSM_3504
Gluconobacter_morbifer_G707
Gluconobacter_frateurii_NBRC_103465
Neokomagataea_thailandica_NBRC_106555
Acetobacteraceae_bacterium_860
Acetobacteraceae_bacterium_868
Neokomagataea_sp._Ta1
Neokomagataea_sp._Jb2
Oecophyllibacter_saccharovorans_Ha5
Acetobacteraceae_ESL0697
Acetobacteraceae_sp_ESL0709
Saccharibacter_sp_EH60
Saccharibacter_sp_EH70
Saccharibacter_sp_EH611
Saccharibacter_sp_17LHSD
Saccharibacter_floricola_DSM_15669
Bombella_apis_ESL0368
Bombella_apis_ESL_0387
Parasaccharibacter_apium_G773c
Bombella_apis_ESL_0380
Bombella_apis_SME1
Parasaccharibacter_apium_A29
Saccharibacter_sp_AM169
Parasaccharibacter_apium_C6
Saccharibacter_sp_3A1
Parasaccharibacter_apium_B8
Bombella_intestini_R-52487
Bombella_apis_ESL_0378
Bombella_apis_ESL_0385
Parasaccharibacter_apium_AS1
Ox. Phosphorylation
TCA cycle
EMP
PPP
ED
arg
asn
chorismate
cys
gly
his
ile
leu
lys
met
phe
pro
ser
thr
trp
tyr
val
Biotin
Vitamin B6
NAD
Thiamine
Heme
Vitamin B12
Riboflavin
Coenzyme A
Folate
Pyrimidine
Purine
T1SS
T2SS
T3SS
pT4SSi
pT4SSt
T4SS_typeB
T4SS_typeC
T4SS_typeE
T4SS_typeG
T4SS_typeI
T4SS_typeT
T5aSS
T5bSS
T6SSi
T6SSii
T6SSiii
Flagellum
T4P
Tad
CONJ
Energy/C metabolism
Amino-acid biosynthesis
Co-factor biosynthesis
Nucleoside biosynthesis
Secretion, adhesion, motility
0
0.2
0.6
1
Proportion complete

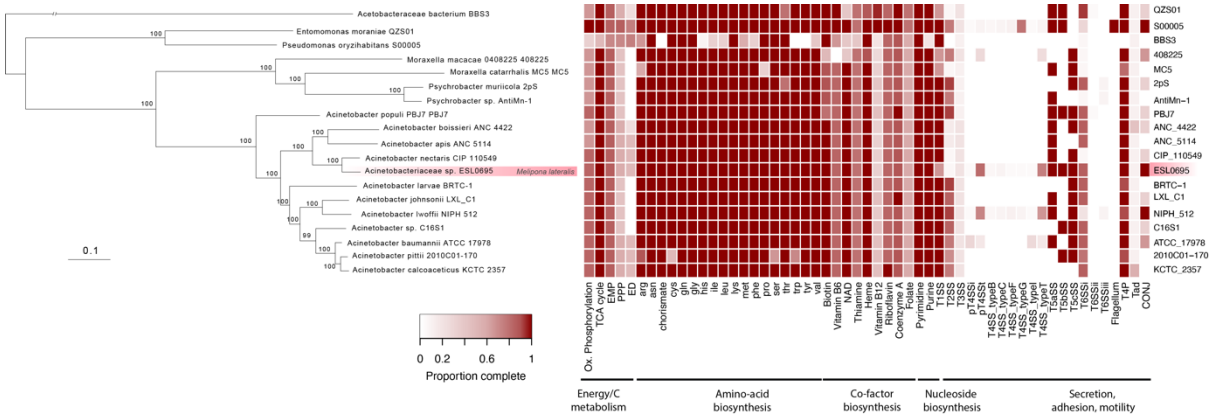
779

780 **Supplementary Figure 6.** Core genome phylogeny of the Acetobacteriaceae. The maximum
781 likelihood tree was computed on the concatenated amino-acid sequences of 303 single copy
782 core genes using LG+F+I+G4 substitution model. Bee isolates are distributed in two distant
783 monophyletic clades. The heatmap indicates the genomic completeness of major metabolic
784 pathways and functions related to energy and carbon metabolism (EMP: Embden–
785 Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as
786 secretion, adhesion, and motility across the sequenced bacterial isolates.
787

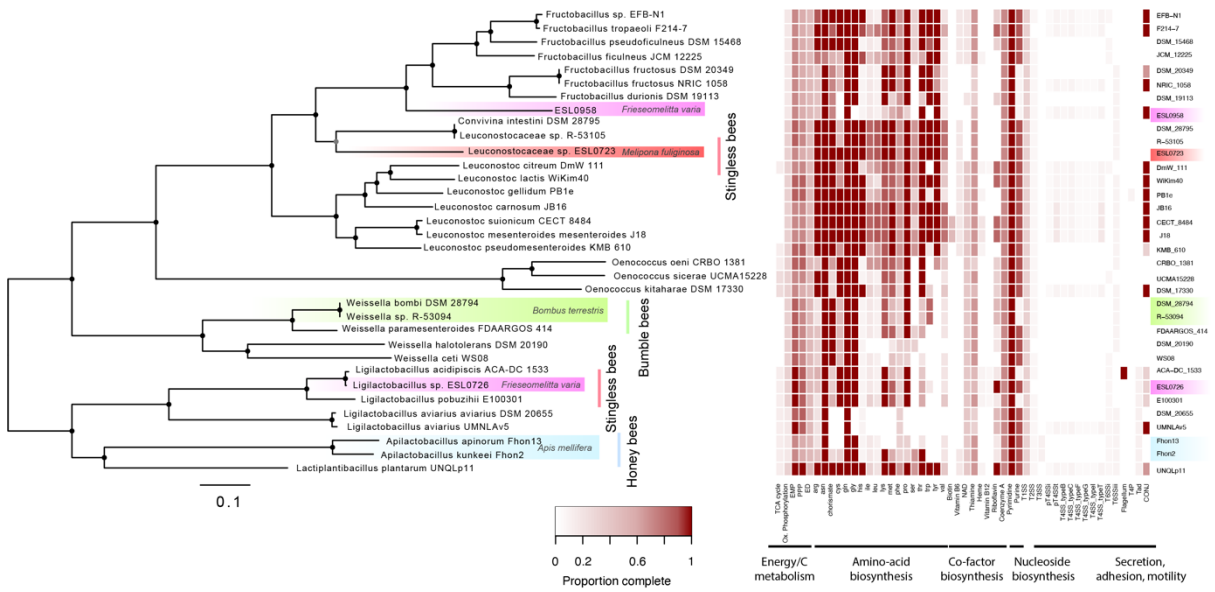
**Supplementary Figure 7.** Core genome phylogeny of the Enterobacteriaceae. The maximum likelihood tree was computed on the concatenated amino-acid sequences of 303 single copy core genes using LG+F+I+G4 substitution model. Metabolic heatmap of the Enterobacteriaceae isolates, the genomes are ordered according to the phylogeny. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.



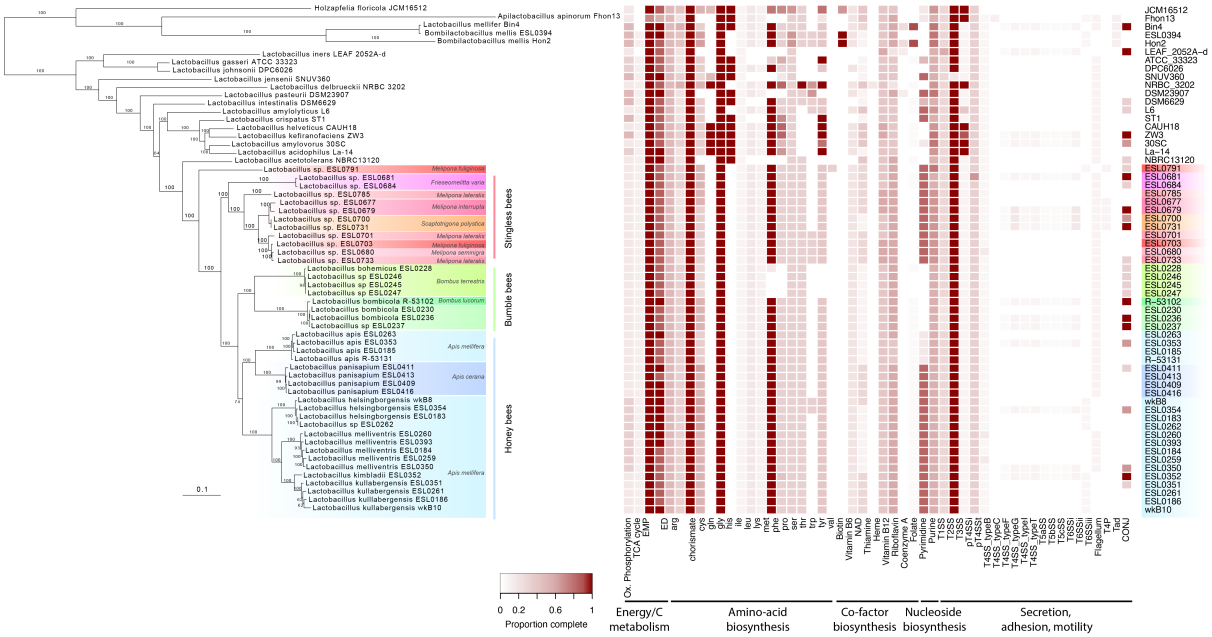

**Supplementary Figure 8.** Core genome phylogeny of the Streptococcaceae including *Floricoccus* isolates. The maximum likelihood tree was computed on the concatenated amino-acid sequences of 303 single copy core genes using LG+F+I+G4 substitution model. Metabolic heatmap of the Streptococcaceae isolates, the genomes are ordered according to the phylogeny. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.



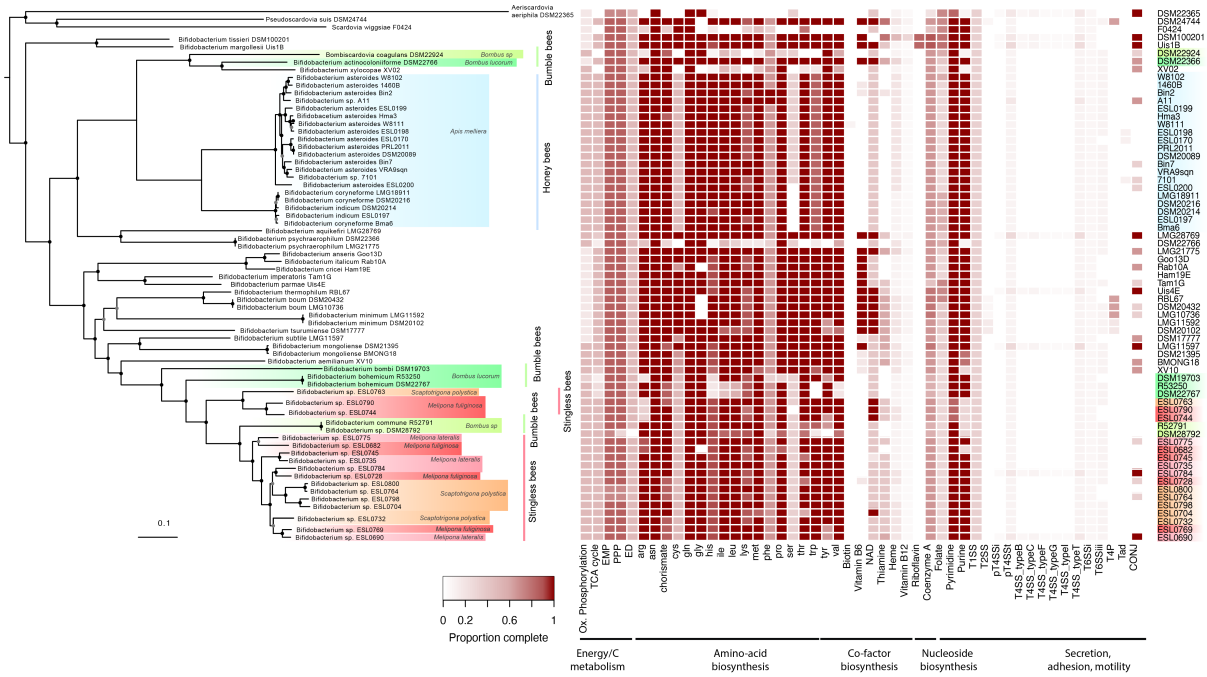

**Supplementary Figure 9.** Core genome phylogeny of the Moraxellaceae. The maximum likelihood tree was computed on the concatenated amino-acid sequences of 303 single copy core genes using LG+F+I+G4 substitution model. Metabolic heatmap of the Moraxellaceae isolates, the genomes are ordered according to the phylogeny. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.




**Supplementary Figure 10.** Core genome phylogeny of the Leuconostocaceae. The maximum likelihood tree was computed on the concatenated amino-acid sequences of 303 single copy core genes using LG+F+I+G4 substitution model. Bee isolates are distributed in two distant monophyletic clades. Metabolic heatmap of the *Leuconostocaceae* isolates, the genomes are ordered according to the phylogeny. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.




**Supplementary Figure 11.** Genome-wide phylogeny of the genus *Lactobacillus* including 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. Metabolic heatmap of the *Lactobacillus* isolates, the genomes are ordered according to the phylogeny. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.




**Supplementary Figure 12. (G)** Genome-wide phylogeny of the genus *Bifidobacterium* with bacteria belonging to social bee-specific clades. 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. Metabolic heatmap of the *Bifidobacterium* isolates, the genomes are ordered according to the phylogeny. The heatmap indicates the genomic completeness of major metabolic pathways and functions related to energy and carbon metabolism (EMP: Embden–Meyerhof–Parnas), the biosynthesis of amino-acid, co-factors, and nucleoside, as well as secretion, adhesion, and motility across the sequenced bacterial isolates.

850 **References**


851 1. P. J. Ascher J, Discover Life bee species guide and world checklist (Hymenoptera:
852 Apoidea: Anthophila). http://www.discoverlife.org/mp/20q?guide=Apoidea_species (2018).
853 2. C. D. Michner, *The bees of the world* (The Johns Hopkins University Press, Baltimore,
854 2007).
855 3. B. N. Danforth, S. Cardinal, C. Praz, E. A. Almeida, D. Michez, The impact of molecular
856 data on our understanding of bee phylogeny and evolution. *Annu Rev Entomol* **58**,
857 57-78 (2013).
858 4. G. Bonilla-Rosso, P. Engel, Functional roles and metabolic niches in the honey bee
859 gut microbiota. *Curr Opin Microbiol* **43**, 69-76 (2018).
860 5. W. K. Kwong, N. A. Moran, Gut microbial communities of social bees. *Nature reviews.*
861 *Microbiology* **14**, 374-384 (2016).
862 6. H. Zheng, M. I. Steele, S. P. Leonard, E. V. S. Motta, N. A. Moran, Honey bees as
863 models for gut microbiota research. *Lab Anim (NY)* **47**, 317-325 (2018).
864 7. I. Cho, M. J. Blaser, The human microbiome: at the interface of health and disease.
865 *Nat Rev Genet* **13**, 260-270 (2012).
866 8. F. Sommer, F. Backhed, The gut microbiota--masters of host development and
867 physiology. *Nature reviews. Microbiology* **11**, 227-238 (2013).
868 9. K. Raymann, N. A. Moran, The role of the gut microbiome in health and disease of
869 adult honey bee workers. *Curr Opin Insect Sci* **26**, 97-104 (2018).
870 10. D. P. Cariveau, J. Elijah Powell, H. Koch, R. Winfree, N. A. Moran, Variation in gut
871 microbial communities and its association with pathogen infection in wild bumble
872 bees (Bombus). *ISME J* **8**, 2369-2379 (2014).
873 11. D. L. Cox-Foster *et al.*, A metagenomic survey of microbes in honey bee colony
874 collapse disorder. *Science* **318**, 283-287 (2007).
875 12. M. A. Hall *et al.*, Temporal changes in the microbiome of stingless bee foragers
876 following colony relocation. *FEMS Microbiol Ecol* **97** (2020).
877 13. H. Koch, D. P. Abrol, J. Li, P. Schmid-Hempel, Diversity and evolutionary patterns of
878 bacterial gut associates of corbiculate bees. *Mol Ecol* **22**, 2028-2044 (2013).
879 14. S. D. Leonhardt, M. Kaltenpoth, Microbial communities of three sympatric Australian
880 stingless bee species. *PLoS One* **9**, e105718 (2014).
881 15. J. Li *et al.*, Two gut community enterotypes recur in diverse bumblebee species. *Curr*
882 *Biol* **25**, R652-653 (2015).
883 16. N. A. Moran, A. K. Hansen, J. E. Powell, Z. L. Sabree, Distinctive gut microbiota of
884 honey bees assessed using deep sampling from individual worker bees. *PLoS One* **7**,
885 e36393 (2012).
886 17. W. K. Kwong *et al.*, Dynamic microbiome evolution in social bees. *Sci Adv* **3**,
887 e1600513 (2017).
888 18. P. Engel *et al.*, Standard methods for research on Apis mellifera gut symbionts.
889 *Journal of Apicultural Research* **52** (2013).
890 19. L. Kesnerova *et al.*, Disentangling metabolic functions of bacteria in the honey bee
891 gut. *PLoS Biol* **15**, e2003467 (2017).
892 20. W. K. Kwong, P. Engel, H. Koch, N. A. Moran, Genomics and host specialization of
893 honey bee and bumble bee gut symbionts. *Proc Natl Acad Sci U S A* **111**, 11509-
894 11514 (2014).

21. M. A. Aizen, L. A. Garibaldi, S. A. Cunningham, A. M. Klein, Long-term global trends in crop yield and production reveal no current pollination shortage but increasing pollinator dependency. *Curr Biol* **18**, 1572-1575 (2008).
22. N. Gallai, J. M. Salles, J. Settele, B. E. Vaissiere, Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol Econ* **68**, 810-821 (2009).
23. F. Bottacini *et al.*, Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS One* **7**, e44229 (2012).
24. K. M. Ellegaard *et al.*, Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. *Mol Ecol* **28**, 2224-2237 (2019).
25. K. M. Ellegaard *et al.*, Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* **16**, 284 (2015).
26. H. Zheng *et al.*, Metabolism of Toxic Sugars by Strains of the Bee Gut Symbiont Gilliamella apicola. *MBio* **7** (2016).
27. H. Zheng *et al.*, Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc Natl Acad Sci U S A* **116**, 25909-25916 (2019).
28. H. Zheng, J. E. Powell, M. I. Steele, C. Dietrich, N. A. Moran, Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proc Natl Acad Sci U S A* **114**, 4775-4780 (2017).
29. P. Engel, V. G. Martinson, N. A. Moran, Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci U S A* **109**, 11002-11007 (2012).
30. Y. Li, S. P. Leonard, J. E. Powell, N. A. Moran, Species divergence in gut-restricted bacteria of social bees. *Proc Natl Acad Sci U S A* **119**, e2115013119 (2022).
31. S. Brochet *et al.*, Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *Elife* **10** (2021).
32. K. M. Ellegaard, P. Engel, Genomic diversity landscape of the honey bee gut microbiota. *Nat Commun* **10**, 446 (2019).
33. K. M. Ellegaard, S. Suenami, R. Miyazaki, P. Engel, Vast Differences in Strain-Level Diversity in the Gut Microbiota of Two Closely Related Honey Bee Species. *Curr Biol* **30**, 2520-2531 e2527 (2020).
34. A. E. S. Cerqueira *et al.*, Extinction of anciently associated gut bacterial symbionts in a clade of stingless bees. *ISME J* **15**, 2813-2816 (2021).
35. L. L. Figueroa, J. J. Maccaro, E. Krichilsky, D. Yanega, Q. S. McFrederick, Why Did the Bee Eat the Chicken? Symbiont Gain, Loss, and Retention in the Vulture Bee Microbiome. *Mbio* **12** (2021).
36. M. S. NGALIMAT *et al.*, A Review on the Association of Bacteria with Stingless Bees. *Sains Malaysiana* **49**, 1853-1863 (2020).
37. D. Tamarit *et al.*, Functionally Structured Genomes in Lactobacillus kunkeei Colonizing the Honey Crop and Food Products of Honeybees and Stingless Bees. *Genome Biol Evol* **7**, 1455-1473 (2015).
38. N. A. I. Jaafar, S. A. S. Mohamad, W. R. W. A. Razak, Antimicrobial activity and antibiotic resistance of lactic acid bacteria isolated from Malaysian stingless bee's gut. *Malays J Microbiol* **15**, 333-341 (2019).
39. S. A. Oliphant *et al.*, 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

941 Tetragonula carbonaria and Austroplebeia australis. *Int J Syst Evol Microbiol* **72**
942 (2022).
943 40. D. W. Roubik, Stingless Bee (Apidae: Apinae: Meliponini) Ecology. *Annu Rev Entomol*
944 10.1146/annurev-ento-120120-103938 (2022).
945 41. C. Grüter, *Stingless Bees* (Springer Nature, 2020), 10.1007/978-3-030-60090-7.
946 42. C. A. Gaulke *et al.*, Ecophylogenetics Clarifies the Evolutionary Association between
947 Mammals and Their Gut Microbiota. *mBio* **9** (2018).
948 43. M. Groussin *et al.*, Unraveling the processes shaping mammalian gut microbiomes
949 over evolutionary time. *Nat Commun* **8**, 14319 (2017).
950 44. N. D. Youngblut *et al.*, Host diet and evolutionary history explain different aspects of
951 gut microbiome diversity among vertebrate clades. *Nat Commun* **10**, 2200 (2019).
952 45. A. T. Reese, R. R. Dunn, Drivers of Microbiome Biodiversity: A Review of General
953 Rules, Feces, and Ignorance. *MBio* **9** (2018).
954 46. B. D. Muegge *et al.*, Diet drives convergence in gut microbiome functions across
955 mammalian phylogeny and within humans. *Science* **332**, 970-974 (2011).
956 47. D. Levin *et al.*, Diversity and functional landscapes in the microbiota of animals in the
957 wild. *Science* 10.1126/science.abb5352 (2021).
958 48. S. Brochet *et al.*, Niche partitioning facilitates coexistence of closely related honey
959 bee gut bacteria. *Elife* **10** (2021).
960 49. S. Bossert *et al.*, Combining transcriptomes and ultraconserved elements to
961 illuminate the phylogeny of Apidae. *Mol Phylogenet Evol* **130**, 121-131 (2019).
962 50. J. Romiguier *et al.*, Phylogenomics Controlling for Base Compositional Bias Reveals a
963 Single Origin of Eusociality in Corbiculate Bees. *Mol Biol Evol* **33**, 670-678 (2016).
964 51. M. Groussin *et al.*, Elevated rates of horizontal gene transfer in the industrialized
965 human microbiome. *Cell* **184**, 2053-2067 e2018 (2021).
966 52. E. Powell, N. Ratnayeke, N. A. Moran, Strain diversity and host specificity in a
967 specialized gut symbiont of honeybees and bumblebees. *Mol Ecol* **25**, 4461-4471
968 (2016).
969 53. Y. H. Tola *et al.*, Loss and Gain of Gut Bacterial Phylotype Symbionts in Afrotropical
970 Stingless Bee Species (Apidae: Meliponinae). *Microorganisms* **9** (2021).
971 54. J. C. Holley, M. N. Jackson, A. T. Pham, S. C. Hatcher, N. A. Moran, Carpenter Bees
972 (Xylocopa) Harbor a Distinctive Gut Microbiome Related to That of Honey Bees and
973 Bumble Bees. *Appl Environ Microbiol* **88**, e0020322 (2022).
974 55. M. Y. Handy *et al.*, Incipiently social carpenter bees (Xylocopa) host distinctive gut
975 bacterial communities and display geographical structure as revealed by full-length
976 PacBio 16S rRNA sequencing. *Mol Ecol* 10.1111/mec.16736 (2022).
977 56. A. H. Moeller *et al.*, Cospeciation of gut microbiota with hominids. *Science* **353**, 380-
978 382 (2016).
979 57. A. H. Nishida, H. Ochman, Captivity and the co-diversification of great ape
980 microbiomes. *Nat Commun* **12**, 5632 (2021).
981 58. T. A. Suzuki *et al.*, Codiversification of gut microbiota with humans. *Science* **377**,
982 1328-1332 (2022).
983 59. Y. Kikuchi *et al.*, Host-symbiont co-speciation and reductive genome evolution in gut
984 symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol* **7**, 2 (2009).
985 60. P. Engel, R. Stepanauskas, N. A. Moran, Hidden diversity in honey bee gut symbionts
986 detected by single-cell genomics. *PLoS Genet* **10**, e1004596 (2014).

61. M. Kim, H. S. Oh, S. C. Park, J. Chun, 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 (2014).
62. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
63. S. Andrews (2010) FastQC: a quality control tool for high throughput sequence data. (Babraham Bioinformatics).
64. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
65. A. Bankevich *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477 (2012).
66. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
67. G. J. Etherington, R. H. Ramirez-Gonzalez, D. MacLean, bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics* **31**, 2565-2567 (2015).
68. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).
69. Y. Lin *et al.*, Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci U S A* **113**, E8396-E8405 (2016).
70. I. Sovic *et al.*, Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun* **7**, 11307 (2016).
71. R. Vaser, I. Sovic, N. Nagarajan, M. Sikic, Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**, 737-746 (2017).
72. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
73. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).
74. V. M. Markowitz *et al.*, IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic acids research* **42**, D568-573 (2014).
75. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).
76. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059-3066 (2002).
77. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **35**, 518-522 (2018).
78. J. G. Caporaso *et al.*, 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 (2011).
79. M. R. C. McLaren, B. J. (2021) Silva 138.1 prokaryotic SSU taxonomic training data formatted for DADA2. (Zenodo).
80. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-731 (2016).

1033 81. M. N. Price, A. M. Deutschbauer, A. P. Arkin, GapMind: Automated Annotation of
1034 Amino Acid Biosynthesis. *mSystems* **5** (2020).
1035 82. S. S. Abby, B. Neron, H. Menager, M. Touchon, E. P. Rocha, MacSyFinder: a program
1036 to mine genomes for molecular systems with an application to CRISPR-Cas systems.
1037 *PLoS One* **9**, e110726 (2014).
1038 83. Y. Yin *et al.*, dbCAN: a web resource for automated carbohydrate-active enzyme
1039 annotation. *Nucleic acids research* **40**, W445-451 (2012).
1040