

### **Supplementary text summary:**

This .pdf file is composed by the Supplementary Material:

- Pg. 2 – 9: Supplementary Information – Supplementary Methods, References, Figure Legends, and Table Headings
- Pg. 10 – 11: colored versions of Fig.1 and Fig.2 for HTML version.
- Pg. 12 – 17: Supplementary Figures 1 – 6
- Pg. 18 – 20: Supplementary Tables 1 – 3

1 **Supplementary Information**

2 **Extinction of anciently associated gut bacterial symbionts in a clade of**  
3 **stingless bees**

4

5 Alan Emanuel Silva Cerqueira<sup>1</sup>, Tobin J. Hammer<sup>2</sup>, Nancy A. Moran<sup>2</sup>, Weyder  
6 Cristiano Santana<sup>3</sup>, Maria Catarina Megumi Kasuya<sup>1</sup>, Cynthia Canêdo da Silva<sup>1</sup>

7

8 Corresponding author: Alan Emanuel Silva Cerqueira: emanuelalansc@gmail.com /  
9 +55 31 36125056

10 <sup>1</sup> Department of Microbiology, Federal University of Viçosa, Viçosa, MG 36570-900,  
11 Brazil

12 <sup>2</sup> Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712,  
13 USA

14 <sup>3</sup> Department of Entomology, Federal University of Viçosa, Viçosa, MG 36570-900,  
15 Brazil

16 **Supplementary Methods**

17 We sampled bee colonies from 23 *Melipona* populations from four  
18 Brazilian biomes: Atlantic Forest, Caatinga (dry Tropical Forest in semiarid  
19 region), Cerrado (Brazilian Savanna) and Amazon (SISBIO/ICMBio authorization  
20 number 69500-1). For each population (i.e., bees from the same species living in  
21 the same site) 3 bee colonies, consisting of bees living in beekeeping boxes,  
22 were sampled. Additionally, 3 colonies of honeybees (*Apis mellifera*) and 8 of the  
23 stingless bee *Tetragonisca angustula* were sampled and used as comparison  
24 groups (Table S1).

25 Bee specimens with similar morphology were grouped into morphotypes  
26 and identified based on taxonomic keys [1] and by comparison with specimens  
27 previously identified by specialists. A total of 14 morphotypes, comprising 8  
28 identified species from the four *Melipona* sub-genera [2] and 6 morphotypes  
29 whose identification could not be confirmed (referred to as "*Melipona* cf. =  
30 *conferatum*") were sampled (see Table S1 for details).

31 Some bees were identified by collaborators: bees sampled in Amazonian  
32 biome (coded MSN and MIN) were identified by Professor Fernando Amaral da  
33 Silveira from the Federal University of Minas Gerais – Brazil; bees from the Santa  
34 Tereza city in the Atlantic forest biome (coded MQ7, MRF and MAN) were  
35 identified by Professor Eduardo Antonio Ferreira from the Federal Institute of  
36 Espírito Santo – Brazil; *Tetragonisca angustula* bees (coded TET) were identified  
37 by Professor Helder Canto Resende from the Federal University of Viçosa –  
38 Brazil. Voucher specimens were deposited in the scientific collection of the  
39 Regional Museum of Entomology of the Federal University of Viçosa (UFVB) -  
40 Brazil.

41 From each of the 80 bee colonies we sampled, 10-20 bees were taken  
42 from the box entrance and placed in sterile tubes containing 95% ethanol. Then,  
43 5 bees from each box were dissected using sterile forceps and iris scissors with  
44 a stereoscopic microscope. Each set of 5 guts was pooled together as a single  
45 sample, an approach used to ensure a sufficient yield of DNA for PCR  
46 amplification. Then, we extracted total DNA using the kit NucleoSpin® Soil  
47 (Macherey-Nagel) preceded by a proteinase K treatment for 4 hours at 56°C. The  
48 DNA was submitted for 250 bp paired-end amplicon sequencing at Novogene  
49 Corporation Inc (Sacramento, CA, USA) using an Illumina NovaSeq 6000

50 Sequencing System. The primer pairs 341F (CCTAYGGGRBGCASCAG) – 806R  
51 (GGACTACNNGGGTATCTAAT) and ITS5-1737F  
52 (GGAAGTAAAAGTCGTAACAAGG) – ITS2-2043R  
53 (GCTGCGTTCTTCATCGATGC) were used, targeting the 16S rRNA V3-V4 and  
54 ITS1 rDNA regions, respectively. Sequencing of the 16S rDNA gene region was  
55 successful for all the samples, but two samples (coded MRF and MSP3) were not  
56 sequenced for the fungal ITS region due to poor amplification during library  
57 preparation (Table S1). The data (SRA accession number PRJNA678404) were  
58 processed with the DADA2 [3] package in R following the DADA2 microbiome  
59 pipeline for 16S (available at: <https://benjjneb.github.io/dada2/tutorial.html>) and  
60 for ITS (available at: [https://benjjneb.github.io/dada2/ITS\\_workflow.html](https://benjjneb.github.io/dada2/ITS_workflow.html)). Briefly,  
61 DADA2 was used to filter, trim, de-replicate and denoise sequences. For fungi,  
62 trimming was not performed, and adapters were removed using Cutadapt [4].  
63 After removing chimeras, we then used these merged paired-end reads to  
64 construct an exact sequence variant (ESV) table. Finally, the taxonomy was  
65 assigned to ESVs by naive Bayesian classifier method implemented in DADA2  
66 using trained SILVA database for bacteria and trained UNITE database for fungi.  
67 For data analysis we used the R packages “mctoolsr” (available at:  
68 <https://github.com/leffj/mctoolsr>), “phyloseq” [5], “ggplot2” [6] and “vegan” [7].  
69 First, mitochondria, chloroplasts, and low abundance ESVs (those with <100 total  
70 reads across all samples) were removed from the dataset (Table S3). We then  
71 made rarefaction curves, which demonstrate that our sequencing depth  
72 adequately captured the majority of bacterial diversity (Fig. S6). Then, we  
73 visualized the differences in bacterial (Fig. 1A) and fungal (Fig. 2) community  
74 composition between bee species at the genus level and, for bacteria, we

75 compared our results to previously published data [8]. The significance of the  
76 differences in the relative abundance of microbial taxa between our *Melipona*,  
77 *Apis mellifera* and *Tetragonisca angustula* samples was determined using a  
78 Kruskal-Wallis test. For the core *Melipona* bacterial taxa (i.e., taxa containing  
79 ESVs present in all samples of all *Melipona* species) we compared the log-  
80 transformed mean relative abundances of the most abundant ESVs in each bee  
81 population (Fig. 1B). Furthermore, these ESVs were submitted to BLASTN  
82 similarity searches against GenBank at NCBI Reference Sequence Database at  
83 which we could identify and download sequences from type isolates aligned to  
84 them. Downloaded sequences (Table S2) were aligned using ClustalW and the  
85 Maximum Likelihood phylogenetic tree made with bootstrap of 1000 replications  
86 using MEGA software [9]. By this approach we could determine the possible  
87 origin of dominant ESVs in *Melipona*, specifically whether they are related to  
88 widespread environmental taxa or to specialized associates of social bees (Fig.  
89 S1-S4). A similar approach was used to determine the possible origin of the most  
90 abundant ESVs of *Starmerella*, the only core *Melipona* fungal taxon (Fig. S5).  
91 Finally, we made a simplified representation of the phylogenetic relationship  
92 between eusocial corbiculate bees (based on [2]) showing the gains [8, 10] and  
93 our hypothesized losses of corbiculate core symbionts (Fig. 1C).

## 94 **Supplementary References**

- 95 1. Camargo JMF, Pedro SRM. Meliponini Lepeletier, 1836. In Moure, J. S.,  
96 Urban, D. & Melo, G. A. R. (Orgs). Catalogue of Bees (Hymenoptera,  
97 Apoidea) in the Neotropical Region - online version. 2013.  
98 <http://www.moure.cria.org.br/catalogue>. Accessed 20 Jan 2021.
- 99 2. Ramírez SR, Nieh JC, Quental TB, Roubik DW, Imperatriz-Fonseca VL,  
100 Pierce NE. A molecular phylogeny of the stingless bee genus *Melipona*  
101 (Hymenoptera: Apidae). *Mol Phylogenet Evol.* 2010; **56**: 519–525.
- 102 3. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes  
103 SP. DADA2 : High-resolution sample inference from Illumina amplicon  
104 data. *Nat Methods.* 2016; **13**: 581–583.
- 105 4. Martin M. Cutadapt Removes Adapter Sequences From High-Throughput  
106 Sequencing Reads. *EMBnet J.* 2011; **17**: 5–7.
- 107 5. McMurdie PJ, Holmes S. phyloseq : An R Package for Reproducible  
108 Interactive Analysis and Graphics of Microbiome Census Data. *PLoS*  
109 *One.* 2013; **8**: e61217.
- 110 6. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. Springer-  
111 Verlag New York.
- 112 7. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et  
113 al. vegan: Community Ecology Package. R package version 2.5-7.  
114 <https://CRAN.R-project.org/package=vegan>.
- 115 8. Kwong WK, Medina LA, Koch H, Sing KW, Soh EJY, Ascher JS, et al.  
116 Dynamic microbiome evolution in social bees. *Sci Adv.* 2017; **3**: 1–17.

- 117 9. Kumar S, Stecher G, Tamura K. MEGA7 : Molecular Evolutionary  
118 Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;  
119 **33**: 1870–1874.
- 120 10. Kwong WK, Moran NA. Gut microbial communities of social bees. *Nat*  
121 *Rev Microbiol.* 2016; **14**: 374–384.

122

### 123 **Supplementary Figure Legends**

124 Fig S1. Phylogenetic tree of the most abundant *Lactobacillus* ESVs found in the  
125 samples of the *Melipona* bee populations. Sequences used in the tree were  
126 obtained in the GenBank database (accession numbers and other information in  
127 Table S2). Bootstrap values are shown in blue letters. ESVs in orange are  
128 phylogenetically closer to isolates obtained from environmental samples. ESVs  
129 in purple are closer to isolates obtained from honeybees or bumblebees  
130 (*Lactobacillus* Firm-4 or Firm-5). Only taxa most closely related to *Melipona* ESVs  
131 are labeled. The root was placed on the branch leading to *Pediococcus*  
132 *acidilactici*.

133 Fig S2. Phylogenetic tree of the most abundant Acetobacteraceae ESVs found in  
134 the samples of the *Melipona* bee populations. Sequences used in the tree were  
135 obtained in the Genbank database (accession numbers and other information in  
136 Table S2). Bootstrap values are shown in blue letters. ESVs in orange are  
137 phylogenetically closer to isolates obtained from environmental samples. ESVs  
138 in purple are closer to isolates obtained from bees. ESVs in gray are closer to  
139 isolates from alternative hosts (ESV 5) or have uncertain phylogenetic proximity  
140 to environmental – pollen/flowers or bee associated isolates (ESV 9 and ESV 12).

141 Only taxa most closely related to *Melipona* ESVs are labeled. The root was  
142 placed on the branch leading to *Acidomonas methanolica*.

143 Fig S3. Phylogenetic tree of the most abundant Bifidobacteriaceae ESVs found  
144 in the samples of the *Melipona* bee populations. Sequences used in the tree were  
145 obtained in the Genbank database (accession numbers and other information in  
146 Table S2). Bootstrap values are shown in blue letters. ESVs in purple are  
147 phylogenetically closer to bacteria isolated from bees. Only taxa most closely  
148 related to *Melipona* ESVs are labeled. The root was placed on the branch leading  
149 to *Gardnerella vaginalis*.

150 Fig S4. Phylogenetic tree of the core *Floricoccus* ESV found in the samples of  
151 the *Melipona* populations. Sequences used in the tree were obtained in the  
152 Genbank database (accession numbers and other information in Table S2).  
153 Bootstrap values are shown in blue letters. The ESV in orange is closely related  
154 to environmental isolates. Only the taxon most closely related to *Melipona* ESV  
155 is labeled. The root was placed on the branch leading to *Lactobacillus kunkeei*  
156 and *Lactobacillus apis*.

157 Fig S5. Phylogenetic tree of the most abundant *Starmerella* ESVs found in the  
158 samples of the *Melipona* populations. Sequences used in the tree were obtained  
159 in the Genbank database (accession numbers and other information in Table S2).  
160 Bootstrap values are shown in blue letters. ESVs in orange are phylogenetically  
161 closer to isolates obtained from environmental (beehive) samples. ESVs in purple  
162 are closer to isolates obtained from bees. Only taxa most closely related to  
163 *Melipona* ESVs are labeled. The root was placed on the branch leading to  
164 *Wickerhamiella azymoides*.

165



166 Fig S6. Rarefaction curves of the 80 samples after removing  
167 Mitochondria/Chloroplasts and low abundance ESVs (<100 reads across all  
168 samples) from the bacterial dataset. The curves indicate that the samples  
169 reached a plateau, thus, overall bacterial diversity is adequately captured by our  
170 sequencing.

171

## 172 **Supplementary Table Headings**

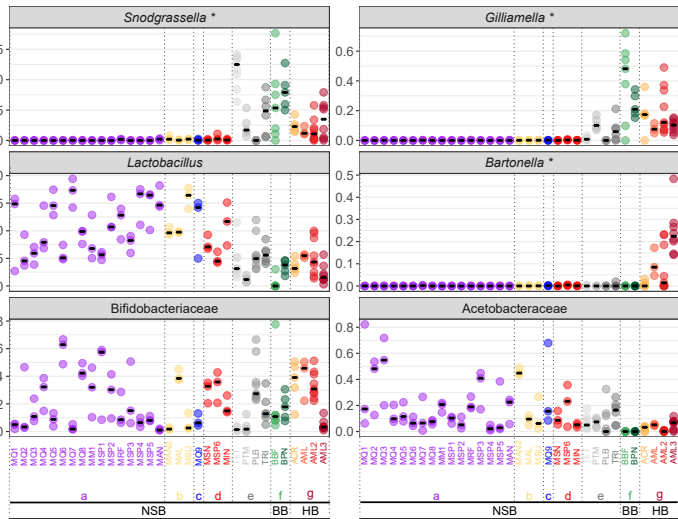
173 Table S1. Bee populations sampled in this study.

174 Table S2. GenBank sequences used for analyses.

175 Table S3. Number of ESVs and reads in the original bacterial dataset, after initial  
176 filtration of mitochondria/chloroplasts, and after subsequent low abundance ESVs  
177 filtering.

A

Proportion of sequences

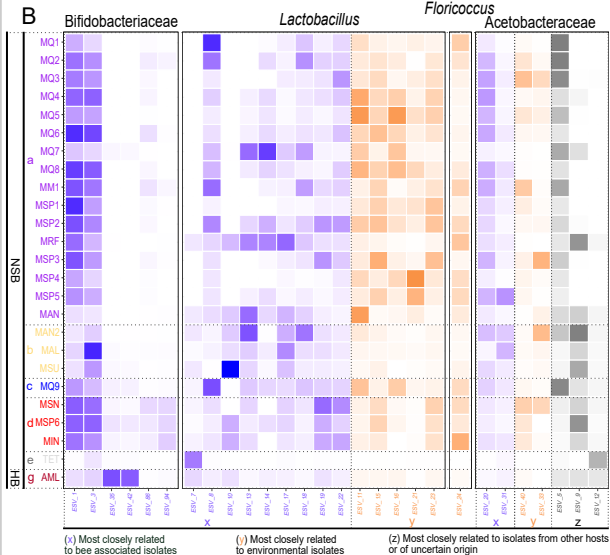


## Codes for Fig. 1A and Fig.1B:

- (a) *Melipona* spp. from the Atlantic Forest biome  
 (b) *Melipona* spp. from the Caatinga biome  
 (c) *Melipona* spp. from the Cerrado biome  
 (d) *Melipona* spp. from the Amazon biome  
 (e - TET) *Tetragonisca angustula* from Atlantic Forest biome  
 (e - PTM) *Partamona helleri* (Kwong et al., 2017)  
 (e - PLB) *Plebeia droyana* (Kwong et al., 2017)  
 (e - TRI) *Trigona spinipes* (Kwong et al., 2017)

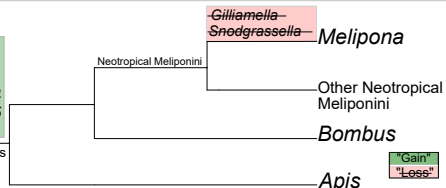
- (f - BBF) *Bombus bifarius* (Kwong et al., 2017)  
 (f - BPN) *Bombus pensylvanicus* (Kwong et al., 2017)  
 (g - ACR) *Apis cerana* (Kwong et al., 2017)  
 (g - AML) *Apis mellifera* from Atlantic Forest biome  
 (g - AML2) *Apis mellifera* - Queensland (Kwong et al., 2017)  
 (g - AML3) *Apis mellifera* - Texas (Kwong et al., 2017)  
 NSB - Neotropical Stingless bees  
 BB - Bumblebees  
 HB - Honeybees

B

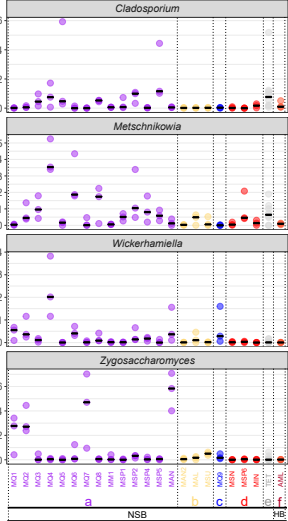
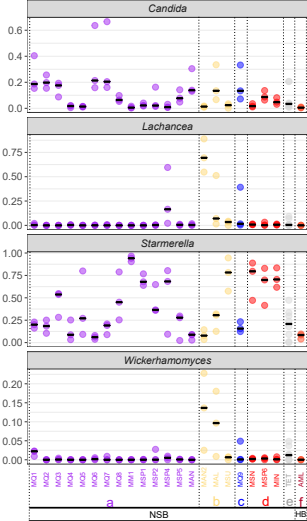


C

**Corbiculate core**  
*Gilliamella*  
*Snodgrassella*  
*Lactobacillus Firm-4*  
*Lactobacillus Firm-5*  
*Bifidobacteriaceae*  
 Eusocial Corbiculate bees



Proportion of sequences



(a) *Melipona* spp. from the Atlantic Forest biome

(b) *Melipona* spp. from the Caatinga biome

(c) *Melipona* spp. from the Cerrado biome

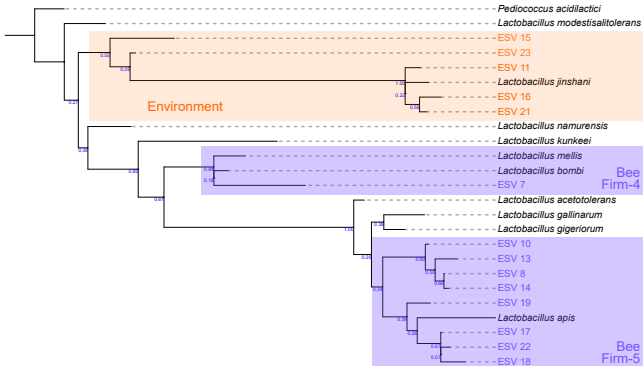
(d) *Melipona* spp. from the Amazon biome

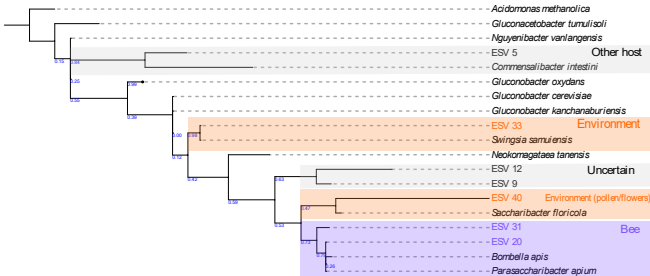
(e) *Tetragonisca angustula* from the Atlantic Forest biome

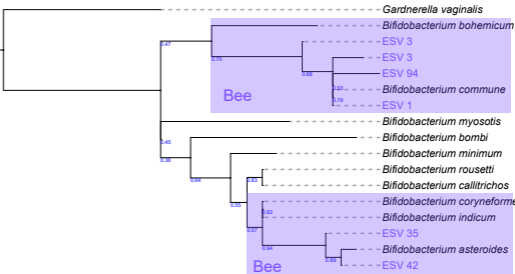
(f) *Apis mellifera* from Atlantic Forest biome

NSB - Neotropical Stingless bees

HB - Honeybees



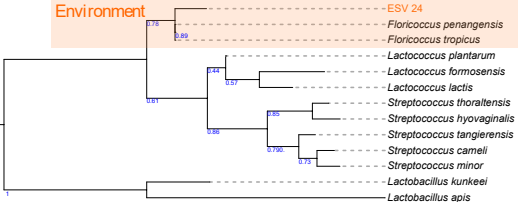




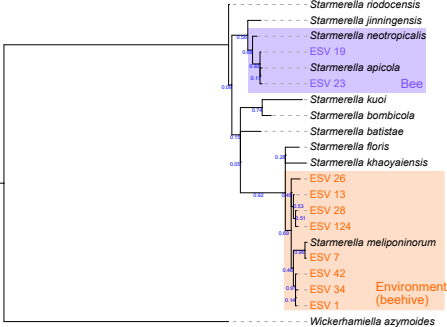
Tree scale: 0.01



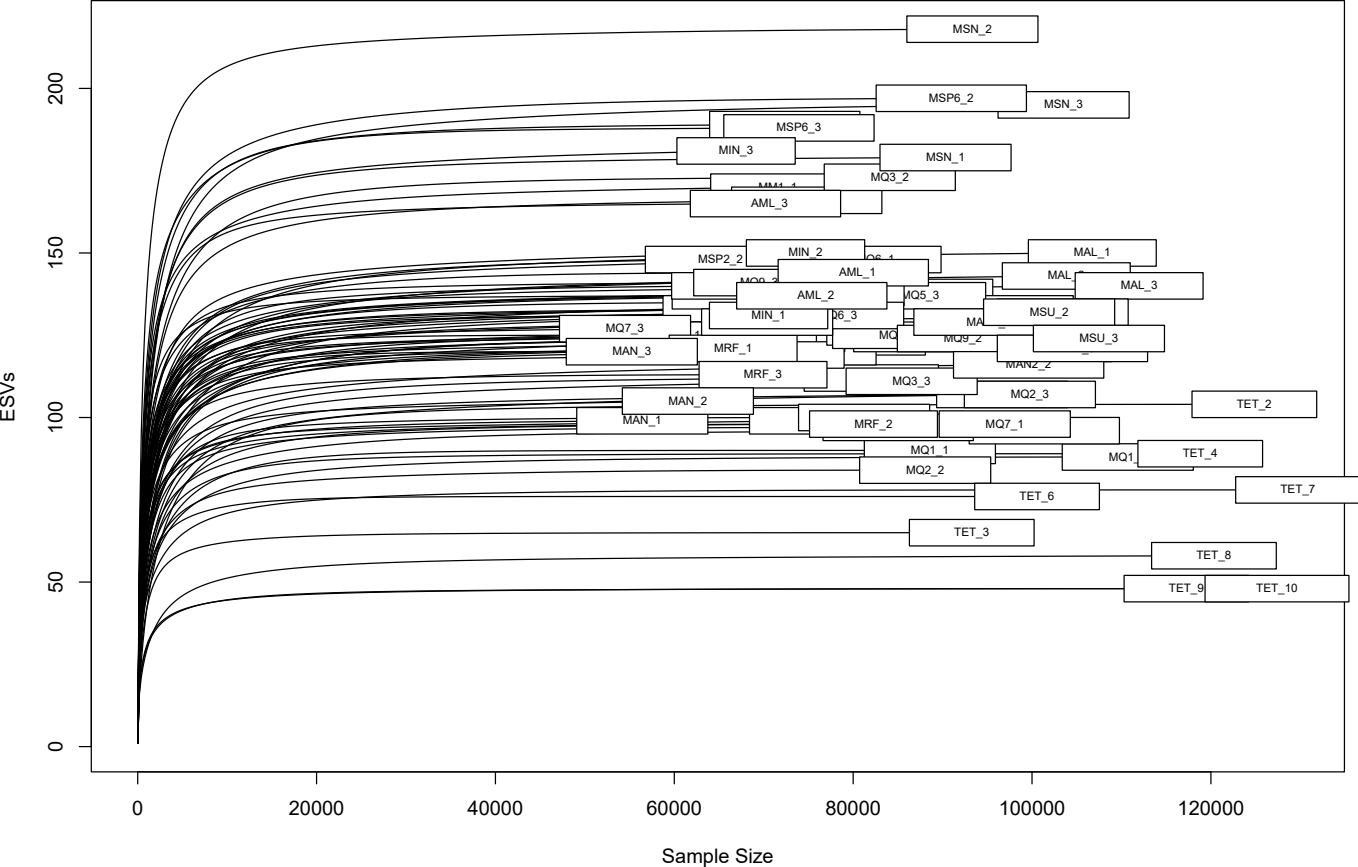
# Environment



Tree scale: 0.1







**Table S1. Bee populations sampled in this study.**

Code	Code NCBI*1	Sampling date	Sampling Biome	Sampling city	Sampling state	Population	Morphotype	Bee species	Sub-genus	Sequenced samples	
										16S	ITS
MQ1	MQA1	May-19	Atlantic Forest	Magé	Rio de Janeiro	1	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ2	MQA2	May-19	Atlantic Forest*2	São Sebastião do Paraíso	Minas Gerais	2	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ3	MQA3	May-19	Atlantic Forest*2	Antônio dos Santos/Caeté	Minas Gerais	3	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ4	MQA4	Apr-19	Atlantic Forest	Viçosa	Minas Gerais	4	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ5	MQA5	May-19	Atlantic Forest	Cotia	São Paulo	5	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ6	MQA6	May-19	Atlantic Forest	São Paulo	São Paulo	6	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ7	MQA7	Apr-19	Atlantic Forest	Santa Tereza	Espírito Santo	7	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MQ8	MQQ	May-19	Atlantic Forest	Cotia	São Paulo	8	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MM1	MM1	Aug-18	Atlantic Forest	Viçosa	Minas Gerais	9	2	<i>Melipona mondury</i>	<i>Michmelia</i>	3	3
MSP1	MM2	May-19	Atlantic Forest	Cotia	São Paulo	10	3	<i>Melipona cf. mondury</i>	<i>Michmelia</i>	3	3
MSP2	MRF-SP	May-19	Atlantic Forest	São Paulo	São Paulo	11	4	<i>Melipona cf. rufiventris</i>	<i>Michmelia</i>	3	3
MRF	MRF	Apr-19	Atlantic Forest	Santa Tereza	Espírito Santo	12	5	<i>Melipona rufiventris</i>	<i>Michmelia</i>	3	-
MSP3	MSC	May-19	Atlantic Forest	Cotia	São Paulo	13	6	<i>Melipona cf. scutellaris</i>	<i>Michmelia</i>	3	-
MSP4	MBS	May-19	Atlantic Forest	São Paulo	São Paulo	14	7	<i>Melipona cf. bicolor</i>	<i>Eomelipona</i>	3	3
MSP5	MRG	May-19	Atlantic Forest	Cotia	São Paulo	15	8	<i>Melipona cf. marginata</i>	<i>Eomelipona</i>	3	3
MAN	MAN	Apr-19	Atlantic Forest	Santa Tereza	Espírito Santo	16	9	<i>Melipona mandacaia</i>	<i>Melipona</i>	3	3
MAN2	MAN2	May-19	Caatinga	Petrolina	Pernambuco	17	9	<i>Melipona mandacaia</i>	<i>Melipona</i>	3	3
MAL	MAL	May-19	Caatinga	Petrolina	Pernambuco	18	10	<i>Melipona asilvai</i>	<i>Eomelipona</i>	3	3
MSU	MSU	Aug-19	Caatinga	Mossoró	Rio Grande do Norte	19	11	<i>Melipona subnitida</i>	<i>Melipona</i>	3	3
MQ9	MQA8	Dec-18	Cerrado	Passos	Minas Gerais	20	1	<i>Melipona quadrifasciata</i>	<i>Melipona</i>	3	3
MSN	MSN	Jun-19	Amazon	Irاندوبا	Amazonas	21	12	<i>Melipona seminigra merrillae</i>	<i>Michmelia</i>	3	3
MSP6	MFL	Jun-19	Amazon	Irاندوبا	Amazonas	22	13	<i>Melipona cf. rufiventris</i>	<i>Michmelia</i>	3	3
MIN	MIN	Jun-19	Amazon	Irاندوبا	Amazonas	23	14	<i>Melipona interrupta</i>	<i>Melikerria</i>	3	3
TET	TET	Oct-19	Atlantic Forest	Camargo/Ouro Preto	Minas Gerais	-	-	<i>Tetragonisca angustula</i>	-	8	8
AML	AML	Jul-19	Atlantic Forest	Viçosa	Minas Gerais	-	-	<i>Apis mellifera</i>	-	3	3

\*1 Population code deposited in NCBI-SRA (available at <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA678404>)

\*2 Atlantic forest in transition with Cerrado biome.

**Table S2. GenBank sequences used for analyses.**

Species name on the tree	Accession	Isolation substrate
<b>Lactobacillus tree</b>		
<i>Lactobacillus jinshani</i>	NR_165780.1	Solid-State Vinegar Culture of Zhenjiang Aromatic Vinegar
<i>Lactobacillus apis</i>	NR_125702.1	Stomach of Honeybees
<i>Lactobacillus bombi</i>	NR_134065.1	Digestive tract of a bumblebee queen ( <i>Bombus terrestris</i> )
<i>Lactobacillus mellis</i>	NR_126249.1	Honey stomach of the honeybee <i>Apis mellifera</i>
<i>Lactobacillus kunkeei</i>	NR_026404.1	Partially fermented grape juice.
<i>Lactobacillus modestisalitolerans</i>	NR_136786.1	Traditional Fermented Foods
<i>Lactobacillus namurensis</i>	NR_042514.1	Sourdough
<i>Lactobacillus gallinarum</i>	NR_042111.1	Chicken intestine
<i>Lactobacillus acetotolerans</i>	NR_117073.1	fermented broth of rice vinegar
<i>Lactobacillus gigeriorum</i>	NR_117057.1	Chicken Crop
<i>Pediococcus acidilactici</i>	NR_042057.1	Barley
<b>Acetobacteraceae tree</b>		
<i>Bombella apis</i>	NR_157653.1	The midgut of a honey bee.
<i>Commensalibacter intestini</i>	NR_116307.1	<i>Drosophila melanogaster</i> gut
<i>Gluconacetobacter tumulisoli</i>	NR_114383.1	Soil
<i>Gluconobacter cerevisiae</i>	NR_117735.1	The brewery environment
<i>Gluconobacter kanchanaburiensis</i>	NR_112822.1	Fermented fruits (spoiled fruits), <i>Artocarpus heterophyllus</i>
<i>Gluconobacter oxydans</i>	NR_118196.1	Involved in industrial vinegar
<i>Neokomagataea tanensis</i>	NR_112959.1	Flower of candle bush, <i>Senna alata</i>
<i>Nguyenibacter vanlangensis</i>	NR_125459.1	Rice rhizosphere
<i>Parasaccharibacter apium</i>	NR_133042.1	Isolated from Honeybee larvae, associated to nurse crops, hypopharyngeal glands (HG), and royal jelly (RJ).
<i>Saccharibacter floricola</i>	NR_024819.1	Pollen in flowers
<i>Swingsia samuiensis</i>	NR_132672.1	A flower of golden trumpet ( <i>Allamanda cathartica</i> )
<i>Tanticharoenia sakaeratensis</i>	NR_041601.1	Soil
<i>Acidomonas methanolica</i>	NR_112209.1	A septic methanol processing mixture
<b>Bifidobacteriaceae tree</b>		
<i>Bifidobacterium commune</i>	NR_136422.1	The bumble bee gut
<i>Bifidobacterium asteroides</i>	NR_044154.1	The Honey Stomach of the Honeybee <i>Apis Mellifera</i>
<i>Bifidobacterium bohemicum</i>	NR_108439.1	The digestive tract of three bumblebee species
<i>Bifidobacterium coryneforme</i>	NR_044690.2	The Hindgut of honeybee
<i>Bifidobacterium indicum</i>	NR_043439.1	The hindgut of honeybee
<i>Bifidobacterium minimum</i>	NR_044692.2	Sewage
<i>Bifidobacterium rousetti</i>	NR_164634.1	Faeces of the Egyptian fruit bat <i>Rousettus aegyptiacus</i>
<i>Bifidobacterium myosotis</i>	NR_147760.1	The faeces of common marmosets ( <i>Callithrix jacchus</i> )
<i>Bifidobacterium bombi</i>	NR_104872.1	The Digestive tract content of the bumblebee <i>Bombus lucorum</i>
<i>Bifidobacterium callitrichos</i>	NR_113172.1	The faeces of common marmoset ( <i>Callithrix jacchus</i> )
<i>Gardnerella vaginalis</i>	NR_118377.1	Human vaginal secretions
<b>Floricoccus tree</b>		
<i>Floricoccus penangensis</i>	NR_159225.1	Flowers of Durian tree and <i>Hibiscus</i>
<i>Floricoccus tropicus</i>	NR_159226.1	Flowers of Durian tree and <i>Hibiscus</i>
<i>Lactococcus formosensis</i>	NR_114366.1	Fermented broccoli stems
<i>Lactococcus plantarum</i>	NR_118903.1	Frozen peas
<i>Lactococcus lactis</i>	NR_113960.1	Human
<i>Streptococcus tangierensis</i>	NR_134818.1	Raw camel milk
<i>Streptococcus thoralensis</i>	NR_026368.1	The intestine of a pig
<i>Streptococcus hyovaginalis</i>	NR_044912.1	Isolated from the vagina of a pig
<i>Streptococcus cameli</i>	NR_134817.1	Raw camel milk
<i>Streptococcus minor</i>	NR_025729.1	The tonsils of a dog
<b>Starmarella tree</b>		
<i>Starmarella apicola</i>	NR_130681.1	Gut of bee
<i>Starmarella floris</i>	NR_155820.1	Adult of the stingless bee <i>Trigona</i> sp. in a <i>M. quinquefolia</i> flower
<i>Starmarella khaoyaiensis</i>	NR_155821.1	Flower
<i>Starmarella meliponinorum</i>	NR_137869.1	Pollen provisions of <i>Tetragonisca angustula</i>
<i>Starmarella neotropicalis</i>	NR_160316.1	Adult bee, <i>Melipona quinquefasciata</i>
<i>Starmarella riocensis</i>	NR_137870.1	Pollen-nectar provisions of the solitary bee <i>Megachile</i> sp
<i>Starmarella bombicola</i>	NR_121483.1	Honey of bumblebee ( <i>Bombus</i> )
<i>Starmarella kuoi</i>	NR_164377.1	Concentrated grape juice
<i>Starmarella batistae</i>	NR_155813.1	Larval provisions, larvae and pupae of the solitary bees <i>Diadasina distincta</i> and <i>Ptilotrix plumata</i>
<i>Starmarella jinningensis</i>	NR_155850.1	Flowers of <i>Erianthus rufipilus</i>
<i>Wickerhamiella azymoides</i>	NR_164509.1	Ripe fruits of <i>Eugenia uniflora</i> and the larva of the fly <i>Anastrepha mucronota</i> found in <i>Peritassa campestris</i> fruits

**Table S3. Number of ESVs and reads in the original bacterial dataset, after initial filtration of mitochondria/chloroplasts, and after subsequent low abundance ESVs filtering.**

	Original dataset	# / % after filtering mitochondria and chloroplasts	# / % after filtering low abundance ESVs
ESVs	2852	2806 / 98.39%	794 / 28.30%
READS	6,911,317	6,894,536 / 99.76%	6,865,138 / 99.57%