

RESEARCH

Open Access



A phylogenomic and comparative genomic analysis of *Commensalibacter*, a versatile insect symbiont

Juliana Botero¹, Atena Sadat Sombolestani¹, Margo Cnockaert¹, Charlotte Peeters¹, Wim Borremans², Luc De Vuyst², Nicolas J. Vereecken³, Denis Michez⁴, Guy Smaghe⁵, German Bonilla-Rosso⁶, Philipp Engel⁶ and Peter Vandamme^{1*}

Abstract

Background To understand mechanisms of adaptation and plasticity of pollinators and other insects a better understanding of diversity and function of their key symbionts is required. *Commensalibacter* is a genus of acetic acid bacterial symbionts in the gut of honey bees and other insect species, yet little information is available on the diversity and function of *Commensalibacter* bacteria. In the present study, whole-genome sequences of 12 *Commensalibacter* isolates from bumble bees, butterflies, Asian hornets and rowan berries were determined, and publicly available genome assemblies of 14 *Commensalibacter* strains were used in a phylogenomic and comparative genomic analysis.

Results The phylogenomic analysis revealed that the 26 *Commensalibacter* isolates represented four species, i.e. *Commensalibacter intestini* and three novel species for which we propose the names *Commensalibacter melissae* sp. nov., *Commensalibacter communis* sp. nov. and *Commensalibacter papalotli* sp. nov. Comparative genomic analysis revealed that the four *Commensalibacter* species had similar genetic pathways for central metabolism characterized by a complete tricarboxylic acid cycle and pentose phosphate pathway, but their genomes differed in size, G + C content, amino acid metabolism and carbohydrate-utilizing enzymes. The reduced genome size, the large number of species-specific gene clusters, and the small number of gene clusters shared between *C. melissae* and other *Commensalibacter* species suggested a unique evolutionary process in *C. melissae*, the Western honey bee symbiont.

Conclusion The genus *Commensalibacter* is a widely distributed insect symbiont that consists of multiple species, each contributing in a species specific manner to the physiology of the holobiont host.

Keywords *Commensalibacter*, Insect symbiont, Asian hornet, Bumble bee, Western honey bee, Butterfly, Phylogenomics, Functional genomics

*Correspondence:

Peter Vandamme

Peter.Vandamme@UGent.be

¹Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, K. L. Ledeganckstraat 35, 9000 Ghent, Belgium

²Research Group of Industrial Microbiology and Food Biotechnology, Department of Bioengineering Sciences, Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

³Agroecology Lab, Université libre de Bruxelles, Boulevard du Triomphe CP 264/02, 1050 Brussels, Belgium

⁴Laboratory of Zoology, Research Institute for Biosciences, University of Mons, Place du parc 20, 7000 Mons, Belgium

⁵Laboratory of Agrozoology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

⁶Department of Fundamental Microbiology, University of Lausanne, CH-1015 Lausanne, Switzerland



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Background

Bee health is endangered by various factors including pesticide exposure, habitat loss, and elevated loads of parasites [1, 2]. Symbiotic gut microbiota of insects play essential roles in the health of their hosts, through mechanisms that include the suppression of pathogens, and therefore contribute to gut homeostasis and host fitness [3–5]. Symbiotic associations between bacteria of the *Acetobacteraceae* family and their insect hosts have received great attention, particularly in pollinators because of their key contribution to ecosystem functioning and their role in agricultural production [3]. The genera *Commensalibacter* and *Bombella* are acetic acid bacteria that belong to the *Acetobacteraceae* family and are regarded as non-core gut symbionts of honey bees because they are generalists that are also able to colonize other bee-associated environments such as beebread and honeycombs as well as the honey bee crop and gut [2, 4].

Commensalibacter bacteria have been detected in and isolated from the intestines of several insects that feed on high carbohydrate diets including honey bees (*Apis mellifera*, *Apis florea* and *Apis dorsata*) [2, 6–10], bumble bees (*Bombus hypnorum* and *Bombus pascuorum*) [11], small carpenter bees (*Ceratina calcarata*) [12], firebugs (*Probergrothius angolensis*) [13], and butterflies (*Heliconius* and several related genera) [14–17]. In honey bees, *Commensalibacter* has a particular caste association since it is more commonly found in larvae, drones and queen guts [18, 19], especially in early stages of gut microbiome colonization [20, 21] and Kesnerova et al. [2] reported an increase in relative abundance of *Commensalibacter* in long-lived winter bees. Today, only a single species, isolated from the gut of *Drosophila melanogaster*, has been formally named, i.e. *Commensalibacter intestini* [22, 23]. Additionally, strain MX01, an isolate from the gut of a monarch butterfly (*Danaus plexippus*) was shown to represent a novel *Commensalibacter* species that was tentatively named “*Commensalibacter papalotli*” [24], but this name has no standing in bacterial nomenclature [25]. Its whole-genome sequence, along with that of several honey bee isolates [26], is publicly available.

Little is known about the taxonomic and functional diversity of *Commensalibacter* isolates from different insect hosts, or its cohesiveness as a genus. Several reports suggest that *Commensalibacter* strains are associated with the health of their respective insect hosts. For example, *C. intestini* was reported to be involved in modulating *Drosophila* immunity and suppressing the proliferation of *Gluconobacter morbifer* by competition [3]. Similarly, Hubert et al. showed that the relative abundance of *Commensalibacter* was increased in adult honey bees infested with varroosis, a disease caused by mites [8]. Moreover, an increased abundance of *Commensalibacter* was correlated with longer host life span in

Speyeria mormonia butterflies [17]. Finally, comparative genomic analyses of *Commensalibacter* and *Bombella* isolates from honey bees suggested that *Commensalibacter* has an advantage in the worker bee hindgut compared to *Bombella*, due to its broader metabolic range [26]. Despite their potential importance, the mechanistic understanding of these functional associations, especially considering the wide range of hosts which *Commensalibacter* can interact with, remains elusive.

In the present study we used comparative genomic analysis to revisit the taxonomy and functional potential of the genus *Commensalibacter*. This analysis includes the genomes of 14 publicly available honey bee [26, 27], fruit fly [28, 29] and butterfly isolates [24], complemented with draft genomes of 12 novel *Commensalibacter* isolates from bumble bees, rowan berries, hornets and butterflies.

Methods

Commensalibacter isolates and cultivation conditions

Novel *Commensalibacter* isolates were obtained in the course of several large-scale isolation campaigns in Belgium ([11, 30] and unpublished data), which involved the use of multiple isolation media and incubation conditions, and the application of MALDI-TOF mass spectrometry for isolate dereplication and preliminary identification [31, 32] (Table 1). This dereplication step allowed to limit the number of isolates for subsequent identification analyses, as isolates with distinct mass spectra are considered to represent genetically distinct strains [11, 32]. Twelve of these isolates were selected for whole genome sequence analysis in the present study (Table 1). Table 1 gives an overview of the *Commensalibacter* isolates obtained in these isolation campaigns, their sources, growth media and atmospheric conditions used for primary isolation, and the strain designation of the isolates selected on the basis of MALDI-TOF MS pattern diversity for whole-genome sequence analysis in the present study.

Commensalibacter reference strains LMG 31900^T (=ESL0284^T) and LMG 27436^T (=A911^T) were obtained from the BCCM/LMG Bacteria Collection (Ghent, Belgium). The former was isolated from a Western honey bee gut sample in Switzerland [26]; the latter is the *C. intestini* type strain, and was isolated from a gut sample of *Drosophila melanogaster* in South Korea [23]. All strains were routinely cultivated on LMG agar medium 404 [50 g/l d-glucose, 10 g/l yeast extract, and 15 g/l agar] and incubated under aerobic conditions at 28 °C for 48 h.

DNA extraction, sequence analysis and genome assembly

Genomic DNA of isolates LMG 28296, LMG 31819^T, LMG 32512^T, R-53529, R-79671, R-79672, R-79673 and R-79674 was extracted using the Maxwell Tissue DNA

Table 1 Overview of new *Commensalibacter* isolates, their isolation source and other isolation details. All isolates originated from Belgium. Insect samples were whole gut samples

Source	N ¹	Isolation medium ²	Isolation year	Strain
<i>Bombus pascuorum</i>	3	M13-A, M404-A	2013	LMG 28296
<i>Bombus hypnorum</i>	3	AC-MA	2013	LMG 31819 ^T and R-53529
<i>Vespa velutina</i> ³	12	M13-A, M404-A	2019	LMG 32512 ^T
<i>Sorbus aucuparia</i> berries	12	MRS-A	2020	R-79671 to R-79674
<i>Aglais io</i>	1	TSA-A	2022	R-83493
<i>Vespa velutina</i> ³	3	M13-A	2022	R-83526
<i>Pieris rapae</i> ⁴	55	M13-A, M404-A	2022	R-83534
<i>Vespa velutina</i> ³	1	M13-A	2022	R-83540

¹Number of isolates with indistinguishable MALDI-TOF mass spectra

²AC, all culture agar (20 g/l tryptose, 3 g/l beef extract, 3 g/l yeast extract, 3 g/l malt extract, 5 g/l dextrose, 0.2 g/l ascorbic acid, 15 g/l agar); M13, LMG medium 13 (25% g/l D-mannitol, 5 g/l yeast extract, 3 g/l bacteriological peptone, 15 g/l agar)

M404, LMG medium 404 (50 g/l D-glucose; 10 g/l yeast extract, 15 g/l agar); MRS, Man, Rogosa and Sharpe agar (Oxoid); TSA, Tryptone Soya Agar (Oxoid); A, aerobic incubation; MA, microaerobic incubation. All isolation media comprised 10 ppm cycloheximide (Sigma-Aldrich) to inhibit fungal growth

³LMG 32512^T, R-83526 and R-83540 were isolated from different *V. velutina* hornets each

⁴Gut samples from two *P. rapae* butterflies were analyzed

kit (Promega, USA) and the Maxwell RSC instrument according to the manufacturer's instructions; genomic DNA of isolates R-83493, R-83526, R-83534, R-83540 was extracted using the Maxwell Cultured Cells DNA kit (Promega, USA). DNA extracts were treated with RNase (2 mg/mL, 5 µL per 100 µL of extract) and incubated at 37 °C for one hour. DNA quality was checked using 1% agarose gel electrophoresis and DNA quantification was performed using the QuantiFluor ONE dsDNA system and the Quantus fluorometer (Promega, USA). Whole-genome sequencing was carried out on the Illumina HiSeq 4000 (R-53529) or NovaSeq 6000 (LMG 28296, LMG 31819^T, LMG 32512^T, R-79671, R-79672, R-79673 and R-79674) platform at the Oxford Genomics Centre (Oxford, UK), or on the NextSeq 2000 platform (R-83534, R-83493, R-83526 and R-83540) at MiGS center (Pittsburgh, USA).

A quality check of the reads was performed using FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and the results were compiled into a single report by using MultiQC 1.9 [33]. Reads were filtered by removing low-quality sequences using fastp v0.20.1 [34] in simple usage. *De novo* assemblies

were obtained with Shovill v1.0.4 (<https://github.com/tseemann/shovill>) [35] with disabled error correction and default settings. Contigs shorter than 500 bp were removed from the final assembly. Reads were mapped to the assemblies using BWA v0.7.17 [36] and the resulting summary statistics, including mapped reads and coverage, were calculated with SAMtools v1.11 [37] and Qualimap v2.2.1 [38]. PlasmidHunter was used for the identification of plasmids [39].

In addition, all 29 publicly available *Commensalibacter* genome assemblies and the genome assemblies of type strains representing additional acetic acid bacterial genera (Supplementary Tables S1 and S2) were downloaded from the NCBI database (June 3, 2022) by using the E-utilities Command [40]. CheckM v1.1.2 was used to estimate genome completeness and contamination [41]. The G+C content and genome size were calculated using QUAST v5.0.2 [42]. The 16S rRNA gene sequences were extracted from the draft genomes using the BAsic Rapid Ribosomal RNA Predictor software (Barrnap) (<https://github.com/tseemann/barrnap>) and were submitted to the EzBioCloud identification tool [43].

Phylogenomic analyses

The whole-genome sequences of *Commensalibacter* isolates and of representative phylogenetic neighbors were used to construct a phylogenomic tree based on 107 single-copy genes using bcgTree [44] and a partitioned maximum-likelihood analysis in RAxML v8.2.12 [45]. Visualization and annotation of the tree were performed using iTOL [46]. To verify taxonomy, genomes were submitted to the Type Strain Genome Server (TYGS) [47], and digital DNA-DNA hybridization (dDDH, formula d4) values were calculated using the Genome-to-Genome Distance Calculator (GGDC 2.1) with recommended settings [48]. In addition, Average Nucleotide Identity (ANI) values were calculated by using the OrthoANIu algorithm using the standalone tool [49].

Annotation and comparative genomic analyses

The Anvi'o pangenomics suite was used to perform annotation (COG and KEGG) and a comparative genomics analysis of *Commensalibacter* genomes. An MCL inflation parameter of 8 was used to assign the protein-coding DNA sequences (CDSs) to clusters of orthologous genes. CDS generated by anvi'o were also annotated using EggNOG-mapper v2.1.7 with the EggNOG database v5.0.2 [50, 51]. Based on COG [52] and KEGG orthology [53, 54], each CDS was assigned to its respective COG, COG category, KEGG, KEGG module, KEGG reaction and KEGG pathway. The COG and KEGG annotations obtained by both tools (i.e. anvi'o and EggNOG-mapper) were combined to obtain a higher proportion of annotated CDSs. Genus core genes and species core genes

were inferred from the gene clusters. A gene cluster was considered to belong to the genus or species core if it was present in all the members of the group. COG annotation was used to assess defense mechanisms. KEGG annotation was used for the calculation of the KEGG module completeness fraction (mcf) by using the KO_mapper script from MicrobeAnnotator [55] and the KEGGREST was used to determine the reactions from the KEGG numbers for the identification of carbohydrate-utilizing enzymes. MacSyFinder v2 was used to identify bacterial secretion systems [56]. Finally, Virsorter2 [57] was used to identify prophages sequences and the resulting sequences were compared to each other using blast+ [58].

Data analysis

Anvi'o results, the COG20 database, MicrobeAnnotator results and KEGG hierarchy were imported in R 4.1.3 and analyzed using tidyverse, imputeTS, matrixStats, fuzzyjoin, ggnewscale, ComplexHeatmap and ggven packages.

Phenotypic tests

Cell morphology and phenotypic characteristics of *C. intestini* LMG 27436^T and *Commensalibacter* isolates LMG 31819^T, LMG 32512^T and LMG 31900^T were examined as described before [59]. For testing growth in the presence of 1 and 2% NaCl and in the presence of 10% ethanol, standard medium (SM) [0.5% (m/v) yeast extract and 5% (m/v) d-glucose] was used. To test the growth on nitrate, all isolates were grown on trypticase soy agar (TSA, Oxoid) supplemented with 0.1% KNO₃ and incubated for 7 days at 28 °C under anaerobic conditions. As a control experiment, isolates were also incubated on TSA without KNO₃ and were incubated under the same conditions.

Results and discussion

Genome characteristics

The genome characteristics of all *Commensalibacter* isolates included in the present study were listed in Supplementary Table S2. The 12 new genomes resulted in assemblies of 16 to 85 contigs and genome sizes of 2.34 to 2.58 Mbp. CheckM analysis (with marker lineage *Rhodospirillales*) revealed more than 99% completeness and less than 0.75% contamination in each of these 12 draft assemblies. For ESL0284^T, only one of the two publicly available assemblies was retained for further analysis. Only 14 of the publicly available *Commensalibacter* genomes comprised 16S rRNA gene sequences and therefore the remaining 14 *Commensalibacter* genome assemblies were excluded from further analysis.

A single copy of the rRNA operon was detected in each of the retained assemblies (Supplementary Table S2), except in *Commensalibacter* sp. ESL0284^T (GCF_009734185.1) and *Commensalibacter* sp. AMU001

(GCF_003691365.1), which are complete genome sequences, and both genome assemblies comprised four identical copies of the rRNA operon. Upon remapping reads of each of the 12 new genome assemblies, we noted that the depth of coverage of the rRNA operon was approximately four times that of the remainder of the genome, suggesting that these *Commensalibacter* genome assemblies comprised four identical copies of the rRNA operon. Plasmid sequences were detected in each of the 12 new *Commensalibacter* genomes. These sequences were present in 3 to 7 contigs in the assemblies (data not shown).

Phylogenomic analyses

The phylogenomic analysis confirmed that all 26 *Commensalibacter* isolates represented a single line of descent within the acetic acid bacteria, with *Entomobacter blattae* as nearest neighbor taxon (Fig. 1). The *Commensalibacter* lineage comprised four well-separated clusters with high bootstrap support. A first cluster (cluster A) comprised *Commensalibacter* sp. ESL0284^T and ten additional honey bee isolate genomes. A second cluster (cluster B) comprised the two *C. intestini* isolate genomes. A third cluster (cluster C) comprised the "*C. papalotli*" MX01, the Asian hornet isolate *Commensalibacter* sp. LMG 32512^T and butterfly isolate *Commensalibacter* sp. R-83534 genomes. Finally, a fourth cluster (cluster D) comprised the genomes of three bumble bee isolates (LMG 28296, LMG 31819^T and R-53529), the rowan berry isolates (R-79671 through R-79674), two Asian hornet isolates (R-83526 and R-83540) and a single butterfly isolate (R-83493).

OrthoANIu and dDDH values were calculated between each pair of *Commensalibacter* genomes (Supplementary Fig. S1) and revealed that each of the four clusters represented a group of isolates sharing dDDH and orthoANIu values above the species delineation thresholds of 70% dDDH [48] and 95–96% orthoANIu [60]. In addition, dDDH and orthoANIu between isolates of different clusters were well below both species delineation thresholds (Supplementary Fig. S1). Together, these data demonstrated that the four clusters corresponded with four *Commensalibacter* species. Below, we formally propose the names *Commensalibacter melissae* sp. nov. for the cluster A isolates, *Commensalibacter papalotli* sp. nov. for all cluster C isolates and *Commensalibacter communis* sp. nov. for all cluster D isolates (Supplementary information).

Supplementary Fig. S2 presents the estimated G+C content and genome size of each of the genomes analyzed and revealed that the four *Commensalibacter* species were characterized by distinct G+C content and genome size ranges. *Commensalibacter melissae* genomes were characterized by the highest G+C

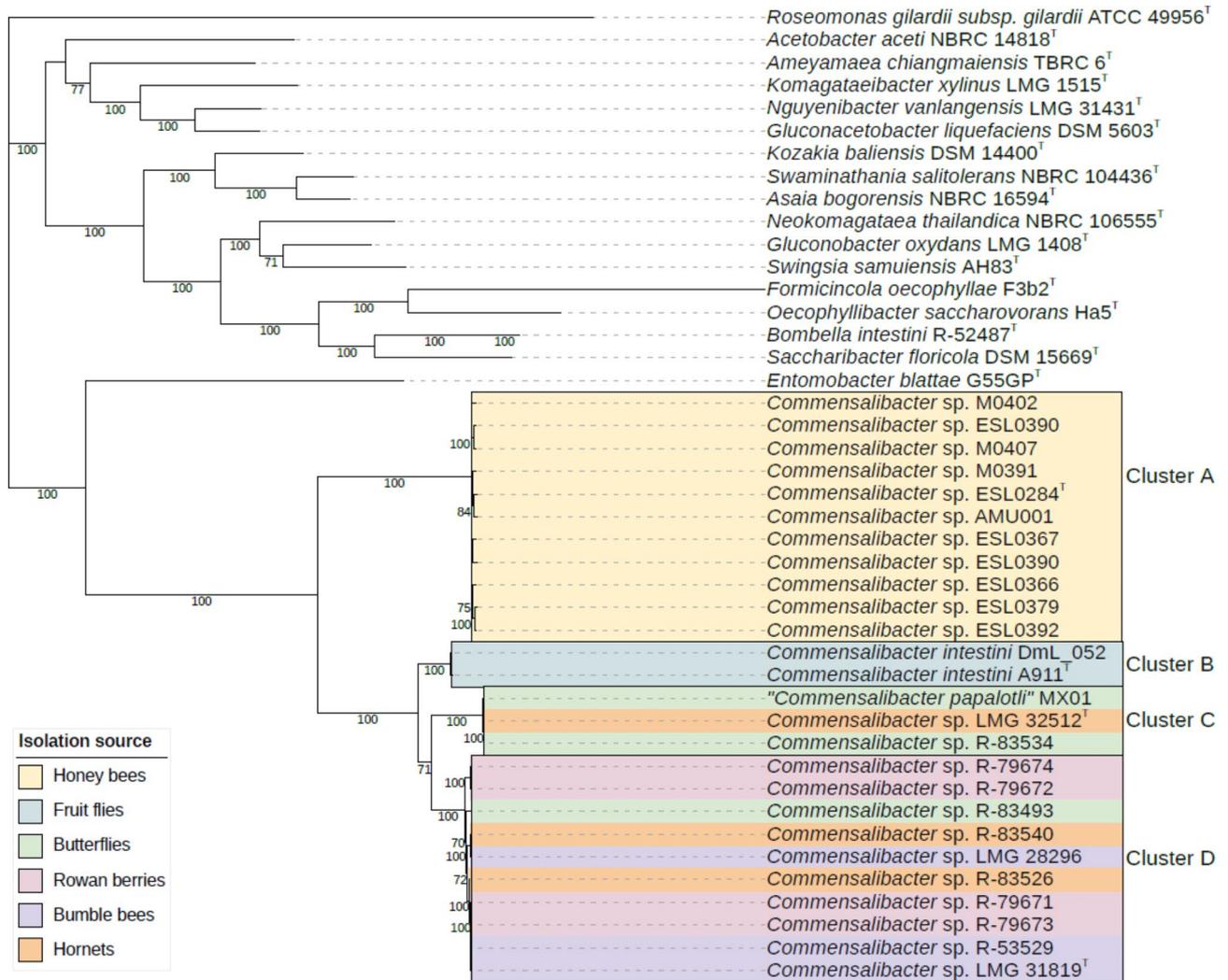


Fig. 1 Maximum-likelihood bcgTree tree based on 107 core genes showing the phylogenomic relationships between the *Commensalibacter* and representative phylogenetic neighbor taxa. Bootstrap percentages above 70% (1000 replicates) are shown next to the branch points. The color bar indicates isolation sources. Superscript 'T' denotes taxonomic type strains

content (37.67 ± 0.08 mol %) and the smallest genome sizes (1.99 ± 0.03 Mbp), suggesting a genomic reduction that may reflect features of their ecology and their specialized association with honey bees [61]. *Commensalibacter communis* genomes had a G+C content (37.40 ± 0.02 mol %) that was slightly lower than that of the *C. melissae* genomes, but had the largest genome sizes (2.51 ± 0.04 Mbp). In contrast, *C. papalotli* and *C. intestini* genomes were similar in size (2.35 ± 0.02 Mbp and 2.44 ± 0.01 Mbp, respectively) and had the lowest G+C content (36.68 ± 0.02 mol % and 36.83 ± 0.02 mol %, respectively).

Comparative genomic analysis

The *Commensalibacter* pangenome consisted of 4,523 gene clusters (54,280 CDS) which included a genus core set of 1,054 gene clusters (30,040 CDS), and 1,219 gene

clusters (9,899 CDS) that were part of the species cores (Fig. 2). A total of 4,153 (68%) and 2,954 (48%) gene clusters were assigned to COG categories and KEGG ortholog groups, respectively. The distribution of COG categories among the *Commensalibacter* genus core genome and the *Commensalibacter* species core genomes (Fig. 3) showed that gene clusters with unknown function (S) were the largest category and represented 12.6% of the *Commensalibacter* genus core genome, and between 22.7% and 25.2% of each of the *Commensalibacter* species core genomes.

The distribution of gene clusters among COG categories (Fig. 3) in each of the *Commensalibacter* species core genomes was fairly homogeneous, with the *C. melissae* core genome as the most aberrant one in which cell wall/membrane/envelope (M) (13.2%), coenzyme transport and metabolism (H) (8.7%) and replication

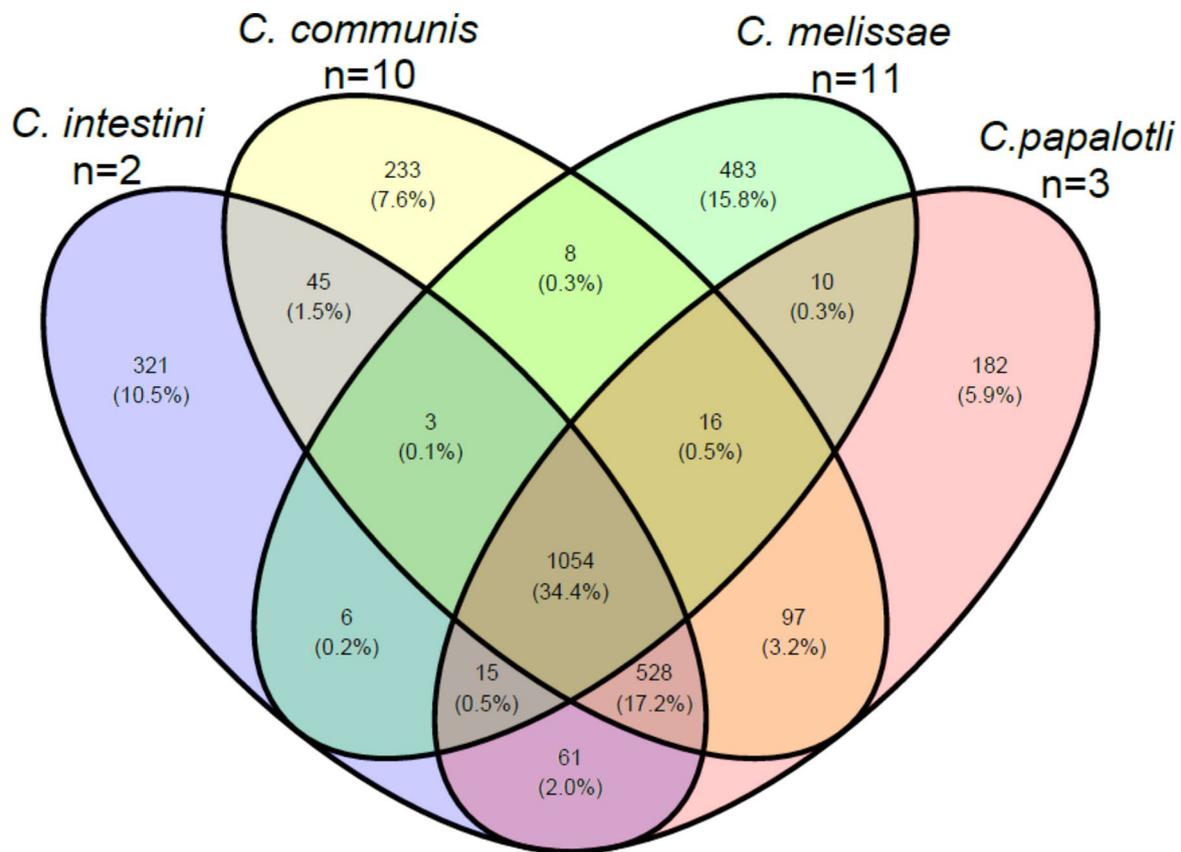


Fig. 2 Venn diagram representing the species and genus core gene clusters within *Commensalibacter*. N is the number of genomes analyzed

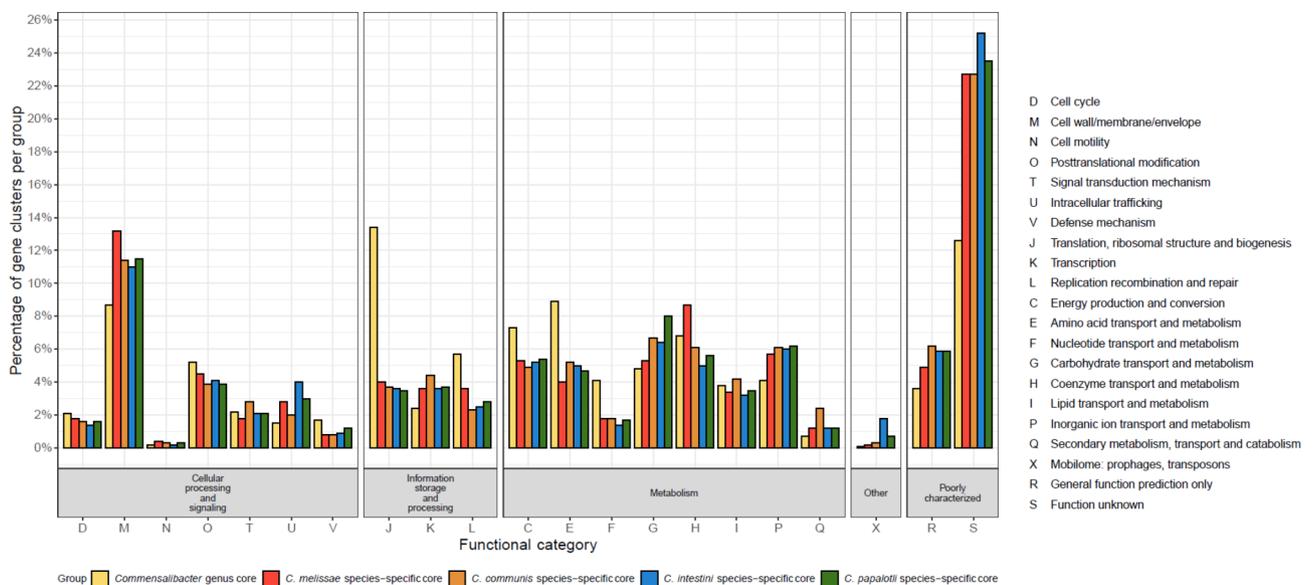


Fig. 3 Distribution of gene clusters over COG categories of the *Commensalibacter* genus core and *Commensalibacter* species core genomes. Percentages of gene clusters per group are presented

recombination and repair (L) (3.6%) gene clusters were overrepresented, and in which carbohydrate transport metabolism (G) (5.3%) was underrepresented. However, the number of gene clusters classified in each of these

categories was lower in *C. melissae* compared to other groups (Supplementary Fig. S3). The *C. communis* core genome was relatively rich in transcription (K) (4.4%), secondary metabolism (Q) (2.4%) and signal transduction

mechanism (T) (2.8%), possibly suggesting a superior environmental adaptability, as category K genes contain many transcriptional regulators (Fig. 3 and Supplementary Fig. S3). The *C. papalotli* core genome appeared particularly enriched in carbohydrate transport metabolism (G) (6.7%) (Fig. 3). Finally, the *C. intestini* core genome was particularly enriched in intracellular trafficking and secretion (U) (5.9%) and mobilome (X) (1.8%). The latter

category included prophage and transposase genes (Fig. 3 and Supplementary Fig. S3).

Amino acid metabolism

An analysis of the completeness of KEGG metabolic pathways revealed a species-specific occurrence and degree of completeness of various metabolic pathways (Fig. 4). Several insect symbionts have been reported to



Fig. 4 Heatmap illustrating the level of completeness of KEGG metabolic modules annotated by MicrobeAnnotator based on the presence and absence of genes. The color scale ranges from zero to 100 indicating the percentage of module completeness. Black and dark colors (the color scale is shown at the bottom) represent a complete or highly complete modules, while white and light colors refer to areas where a module is absent and highly incomplete. The dendrogram (at the top) shows that the species are clustered by the Pearson correlation

hydrolyze proteins and synthesize amino acids that are essential for proper growth and development of their insect host [62, 63]. *Commensalibacter* genomes encoded the biosynthesis of 11 amino acids, which included 8 out of 10 essential amino acids for bees (Fig. 4) [64, 65]. The pathway for methionine biosynthesis, which is an essential amino acid involved in the initiation of protein translation, was incomplete in the *Commensalibacter* genomes (M00017) because the *metB* gene was missing, as also reported in other insect symbionts with reduced genome sizes [66, 67]. It is unclear how these bacteria produce methionine. In addition, putrescine, a polyamine derived from the decarboxylation of ornithine (M00134), was exclusively encoded in *C. intestini* and *C. communis* genomes. The latter species was also capable of producing betaine (M00555), a powerful osmoprotectant that allows bacteria to survive and compete in environments with variable external osmolarity [68]. It has been suggested that nitrogenous waste products such as uric acid that are produced in the honey bee rectum may support persistence of *Commensalibacter* [20]. Uric acid degradation genes were detected in the *C. communis* genomes, but not in *C. melissae*, *C. papalotli* or *C. intestini* genomes. The former genomes encoded genes for the degradation of uric acid to ureidoglycine, which can be further converted into glycine (EC:1.14.13.113, EC:3.5.2.17, EC:4.1.1.97, EC: 3.5.2.5, EC: 3.5.3.9 and EC: 2.6.1.112).

Carbohydrate metabolism

As in other acetic acid bacterial genomes [69], none of the *Commensalibacter* genomes encoded 6-phosphofructokinase which suggested that the Embden–Meyerhof–Parnas pathway (M00001) is non-functional (Fig. 4). In contrast, the oxidative and non-oxidative phases of the pentose phosphate pathway (M00006 and M00007) were complete, suggesting that this pathway is functional. The Entner–Doudoroff pathway is likely not functional (M00008) due to the absence of the enzyme that catalyzes 6-P-gluconate to D-glyceraldehyde-3-phosphate (EC: 4.1.2.14). All tricarboxylic acid cycle genes (M00009 and M00011) were present. Pathways for galactose catabolism (both M00554 and M00632) were complete only in *C. intestini*, *C. papalotli* and *C. communis* genomes.

An analysis of the presence of genes encoding 119 carbohydrate-utilizing enzymes and KEGG reactions belonging to carbohydrate metabolism again revealed a species-specific occurrence of these carbohydrate-utilizing enzymes (Supplementary Fig. S4). Genes related to 14 carbohydrate-utilizing enzymes were identified in *Commensalibacter* genomes. Glucose and fructose are the main carbohydrates in nectar, and hence in pollinator diets [70]. Fructose-utilizing enzymes were abundantly encoded in each of the genomes analyzed, but genes

encoding glucose-utilizing enzymes were largely lacking in *C. melissae*, as were many of the other carbohydrate-utilizing enzymes. Indeed, some *Commensalibacter* genomes encoded enzymes that utilize less common carbohydrates such as mannose, lactose, arabinose, and melibiose, which are indigestible or toxic to many pollinators [71–73]. The detection of these genes suggests that *Commensalibacter* symbionts might mitigate effects induced by such carbohydrates [4, 74]. Furthermore, the presence of β -glucosidase (GH3) and β -mannanase (GH26) which hydrolyze glycosidic bonds in complex gluco- or manno-configured polysaccharides [75] such as hemicellulose polymers [76], suggested that *Commensalibacter* participates with other symbionts in the digestion of polysaccharides [77].

Energy metabolism

Acetic acid bacteria have an unusual metabolism that is characterized by the oxidization of carbohydrates, sugar alcohols and ethanol to produce the corresponding sugar acids or acetic acid, a process executed by primary dehydrogenases located on the periplasmic side of the cytoplasmic membrane [78]. Such oxidation reactions are referred to as ‘oxidative fermentation’ reactions because they result in incomplete oxidation of compounds, which can eventually be further assimilated -or overoxidized- in a later growth phase [79, 80]. However, *Commensalibacter* bacteria can utilize a tricarboxylic acid cycle coupled to oxidative phosphorylation, which is energetically more efficient (Fig. 4). The 16 dehydrogenases/reductases reported by Bonilla-Rosso et al. [26] in honey bee *Commensalibacter* isolate genomes were also detected in the present study except for some slight differences. All *Commensalibacter* isolate genomes shared five dehydrogenases/reductases: three were able to oxidize metabolites (i.e. D-lactate dehydrogenase [EC 1.1.2.5], putative membrane-bound dehydrogenase [EC:1.5.5.1] and bifunctional proline dehydrogenase [EC 1.5.5.2]), cytochrome bc1 that transfers electrons from quinol to a higher potential acceptor protein (complex III), and cytochrome o ubiquinol oxidase that is the terminal electron acceptor oxidase (complex IV). The entire cytochrome bd ubiquinol gene complex (alternative complex IV) was detected in *C. intestini*, *C. papalotli* and *C. melissae* genomes. Moreover, *C. communis* and *C. papalotli* genomes all encoded the same dihydro-orotate dehydrogenase (DHOD), where a different DHOD was detected in *C. melissae* genomes. In total, five dehydrogenases/reductases were specific to *Commensalibacter* (four of which could oxidize succinate, NADH, glycerol 3-phosphate and malate, and one of which could reduce nitrate).

Strikingly, all *Commensalibacter* genomes encoded nitrate reductase (EC: 1.7.5.1), suggesting the capacity

to gain energy through anaerobic respiration. Only *C. intestini* and *C. papalotli* genomes encoded the complete pathway for dissimilatory nitrate reduction to ammonia (M00530, Fig. 4), where *C. melissae* and *C. communis* genomes encoded the conversion of nitrate to nitrite, and nitric oxide to nitrous oxide (nitric oxide reductase norB gene, EC: 1.7.2.5). When inoculated on TSA supplemented with 0.1% KNO₃ and incubated anaerobically, growth was observed in *C. papalotli*, weak growth was noted in *C. intestini* and *C. communis*, and no growth was observed in *C. melissae* (data not shown). In the microaerobic environment of an insect gut where oxygen remains the main electron acceptor, it is likely that nitric oxide reductase has a detoxifying role [81].

Interactions with host cells and other gut microorganisms

Commensalibacter genomes contained ~14 defense mechanism gene clusters (Supplementary Fig. S5). Type 1 and type 5 secretion systems, multidrug efflux pump genes (COG1132 and COG2076) and a bacteriocin exporter gene (COG2274) were present in all genomes studied, where CRISPR-cas genes were uniformly absent, as previously reported [26]. Other defense mechanism genes were occasionally detected, but not in a species-specific manner [82, 83]. These include genes related to type I (COG0610) and type II (COG1002) restriction-modification systems [84] and toxin-antitoxin systems (i.e. YeeF-YezG and RelBE) [85]. In addition, *C. melissae* genomes carried genes for the detoxification of formaldehyde by catalyzing S-formylglutathione into formate (COG0627). Formaldehyde is highly toxic to animals and bacteria, but can be detoxified by some organisms [86]. The presence of formaldehyde detoxification genes may suggest that formaldehyde can be produced by the host or by other host microbiota as a defense mechanism, as reported during the *Varroa destructor* infection process in honey bees [87], for which the *Commensalibacter* relative abundance increased by increasing varroosis levels [8].

COG category X comprised some prophage-associated genes, particularly in the *C. intestini* and *C. papalotli* genomes (Fig. 3), and multiple prophage sequences were detected using Virsorter2 [57]: between one and five in *C. melissae*, between one and nine in *C. communis*, between four and 15 in *C. papalotli*, and six each in *C. intestini* genomes. Prophage sequences that occurred in multiple genomes were consistently 100% identical within species, but differed between species (data not shown), except for the *C. papalotli* LMG 32512^T genome which comprised prophage sequences that were 100% identical and with more than 90% of query coverage, towards prophage sequences observed in the *C. communis* R-79673, R-79671, R-53529 and LMG 31819^T genomes.

Finally, all *Commensalibacter* genomes encoded genes for the production of biotin (vitamin H, M00123), riboflavin (vitamin B2, M00125), niacin (vitamin B3, EC:6.3.4.21), pantothenic acid (vitamin B5, EC:6.3.2.1), pyridoxal 5-phosphate, pyridoxal and pyridoxine (vitamin B6, EC:1.1.1.65 and EC:4.3.3.6), and cobalamin (vitamin B12, M00122) (Fig. 4). In addition, the thiamine (vitamin B1) biosynthesis pathway was completely encoded in the *C. melissae*, *C. communis*, and *C. papalotli*, but not the *C. intestini*, genomes. We failed to detect the *folA* gene and therefore could not confirm that *Commensalibacter* genomes encode folic acid biosynthesis (vitamin B9, M00126, M00841 and M00840) [26].

Commensalibacter is a widely distributed insect symbiont

Commensalibacter bacteria are increasingly reported as members of the gut microbiota of pollinator and other insect species [2, 6–17]. Here, we analyzed genome sequences of a representative selection of new *Commensalibacter* isolates, along with publicly available reference strains and genome sequences. Together, the genus *Commensalibacter* was composed of four taxonomically and functionally different species (Figs. 1, 2, 3 and 4, Supplementary Figs. S1–S5) [22–24, 26, 27].

We hypothesized that the detection of two *Commensalibacter* species in Asian hornet samples reflected its predatory behavior on other insect species. We revisited publicly available 16S rRNA amplicon sequencing data reported in an Italian study of *V. velutina* hornets of different castes, life stages and colonies as well as colony samples [88]. Where the authors did not report or discuss *Commensalibacter* sequences in their study, a reanalysis of their amplicon sequencing variants (ASVs) revealed the presence of nine *Commensalibacter* ASVs (Supplementary Table S3) in their data set. These ASVs were detected in low abundances (<1%) in workers, gynes, larvae and nest paper, thus supporting a hypothesis of non-colonizing bacteria that are in transit. However, in meconium samples an abundance of 13% was detected. Four ASVs were 100% identical to genome-derived 16S rRNA gene sequences of *C. melissae* ESL0284^T, *C. communis* LMG 31819^T, *C. papalotli* LMG 32512^T, and *C. intestini* A911^T. The remaining ASVs highlighted additional *Commensalibacter* sequence diversity, suggesting predation on insect hosts that carried other, hitherto unreported, *Commensalibacter* species. To the best of our knowledge, the isolation of *C. communis* from rowan berry samples in a small-scale study of acetic acid bacteria in fruit samples (unpublished data), represented the first report of environmentally isolated *Commensalibacter* bacteria. Although it is unclear where microbiota that are shared between flowers and pollinator species originate from, it is well-known that flowers are hubs of microbial transmission [89] and a study in Belgium

showed that the *Sorbus* group is an important food source for bumble bees in anthropogenic environments [90].

The taxonomic and functional diversity within other bee symbiont genera is poorly understood

The microbiota of honey bees and other eusocial corbiculate bees have been studied intensely, not only because these pollinators fulfill critical roles in ecosystem services and agriculture, but also because the bee microbiome serves as a model for evolution and ecology of host-microbe interactions [4, 91, 92]. While the honey bee microbiome in particular is simple and highly conserved there is misconception in literature regarding its taxonomic complexity. Five core phylotypes -or 'species'- have consistently been reported, along with considerable strain-level variation within each of the core species [93–96]. In most of these phylotypes sequence discrete populations have now been observed, some of which corresponded with named species [95, 96]. While it has become gradually clear that the five core phylotypes in the three major corbiculate bee clades, i.e. *Apis*, *Bombus* and stingless bee species, correspond more with named bacterial genera rather than with single named species, authors continue to treat phylotypes and species as synonymous terms [92, 97]. From a taxonomic perspective, the five core phylotypes correspond with the genera *Snodgrassella* (the so-called Beta phylotype [98]), *Gilliamella* (Gamma-1), *Lactobacillus* (Firm-5), *Bombilactobacillus* (Firm-4) and *Bifidobacterium* (Bifido-1 and Bifido-2), while the *Bartonella* (Alpha-1), *Commensalibacter* (Alpha 2.1), *Bombella* (Alpha 2.2), *Frischella* (Gamma-2), *Apilactobacillus* (Lacto-3), *Bombiscardovia* (Bifido), and *Apibacter* (Bacteroides) phylotypes are considered non-core bacteria [2, 4].

The functional potential of these symbionts is gradually being explored. *Bifidobacterium* and *Gilliamella* are considered primary degraders of hemicellulose [77]. The genomes of these two bacteria, along with *Snodgrassella*, *Bartonella*, *Lactobacillus*, and *Bombilactobacillus*, encode genes that catalyze the reactions of a wide variety of polysaccharides and monosaccharides, including pectin-degrading enzymes and glycoside hydrolases, and therefore have the potential to aid in the breakdown of pollen, the release of nutrient-rich components thereof, and the removal of toxic sugars [99, 100]. In contrast, *Apibacter* [101, 102] and *Bombella* [103, 104] genomes mainly encode enzymes for the utilization of simple mono-saccharides or sucrose. Bee symbionts also have diverging capacities for the biosynthesis of amino acids and other vitamins. *Gilliamella*, *Snodgrassella*, *Apibacter*, *Bifidobacterium*, *Bartonella* and *Bombella* genomes encode the genes required for the synthesis of at least 18 amino acids; in contrast, *Bombilactobacillus*

and *Lactobacillus* genomes present genes for the biosynthesis of some five amino acids [77, 101, 105]. Furthermore *Snodgrassella*, *Frischella*, and *Gilliamella* genomes encode genes for the biosynthesis of vitamins B2 and B9, and thiamine [106]; *Bombella* genomes comprise complete gene sets for the biosynthesis of vitamins B2, B3, B5, B6 and B9 [26]; *Bifidobacterium* genomes encode genes for the production of vitamins B6 and B9 [107], and *Apibacter* genomes contained genes involved in the biosynthesis of vitamin B2 only [102, 108].

While many other functional capacities and differences have been reported [97, 99, 101, 109, 110], very few comparative genomic or physiological studies systematically addressed functional differences between all species of a single symbiont genus, as in the present study [111–114]. The metabolic repertoire of a bacterial species is encoded in a core genome that is conserved within species and that typically comprises 75 to 90% of the gene content of any strain therein, and in an accessory genome that is strain specific [115], and functional analyses of bee symbionts are therefore best modeled on state-of-the-art taxonomic information. The number of named species in each of the bee symbiont genera ranges now from two (*Frischella* and *Apibacter*) to 14 (*Bifidobacterium*) (Supplementary Table S4), and the metabolic capacities and differences of most of these species are yet to be explored through comparative genomic or physiological studies. While some of these species are clearly host-specific [97, 110, 116], many others co-occur in a single host [96]. The observation that genetically distinct but closely related strains partition their environment at fine phylogenetic and phylogenomic scales is not well understood [117]. The existence and function of such microdiversity, i.e. the co-occurrence of closely related but ecologically and physiologically distinct taxonomic groups, has been documented for about two decades and is an intrinsic property of many microorganisms [118]. A deeper mining of the genomic and physiological potential of different species of each of these bee symbiont genera will be required to improve our understanding of their functional roles and differences.

Conclusion

The present study demonstrated that the genus *Commensalibacter* comprises at least four insect-associated species. Comparative genomic analyses revealed that the four *Commensalibacter* species had a similar genomic potential for central metabolism that was characterized by complete tricarboxylic acid cycle and pentose phosphate pathways, but their genomes differed in size, G+C content, and amino acid metabolism and carbohydrate-utilizing enzyme repertoires. *Commensalibacter melissae* genomes were most reduced in size. This was reflected in the loss of several metabolic pathways and

even in pathways that encode metabolism of D-glucose, a key component of nectar (Fig. 4 and Supplementary Fig. S4). In concert with this, *C. melissae* genomes comprised the largest number of species-specific gene clusters and shared very few (10 or less) (Fig. 2) gene clusters with each of the three remaining *Commensalibacter* species. In contrast, *C. communis*, *C. papalotli*, and *C. intestini* shared 528 gene clusters which were absent in *C. melissae* (Fig. 2). There were clear metabolic differences between the former three species as well. *Commensalibacter communis* and *C. intestini* encoded the biosynthesis of putrescine, a commonly produced microbial metabolite that regulates multiple biological processes in the large intestine of humans and mice [119]. *Commensalibacter communis* was also capable of producing betaine, a powerful osmoprotectant that allows bacteria to survive and compete in environments with variable external osmolarity [68]. In particular the distribution of some carbohydrate-utilizing enzymes (i.e. D-galactose, xylitol, L-sorbose, D-xylose, D-mannose, L-rhamnose, lactose, L-arabinose, D-tagatose, D-galacturonic acid, L-ribulose, and L-xylulose) (Supplementary Fig. S4) differed markedly between these three species, and likely revealed a potential for detoxification or revealed cross-feeding mechanisms [74, 120]. Together, the reduced genome size, the large number of species-specific gene clusters, and the small number of gene clusters shared between *C. melissae* and other *Commensalibacter* species suggested a unique evolutionary process in *C. melissae*, the Western honey bee symbiont.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42523-023-00248-6>.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the sequencing data.

Authors' Contributions

JB and PV conceived the study and wrote the first draft of the manuscript. ASS, GB-R, PE and PV provided the biological material. ASS, MC, WB performed the practical work. JB and CP performed all bioinformatics analyses. LDV and PV provided infrastructure. NJV, DM, GS and PV provided funding. All authors read, revised and approved the submitted manuscript.

Funding

This work was supported through funding of the FNRS/FWO joint program "EOS – Excellence Of Science" for the project "Climate change and its impact on pollination services" (CLIPS, no.3094785).

Data Availability

The annotated genome assemblies were submitted to the European Nucleotide Archive (ENA) and are publicly available under project PRJEB54578.

Declarations

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Received: 3 February 2023 / Accepted: 21 April 2023

Published online: 29 April 2023

References

1. Barron AB. Death of the bee hive: understanding the failure of an insect society. *Curr Opin Insect Sci.* 2015;10:45–50. <https://doi.org/10.1016/j.cois.2015.04.004>.
2. Kešnerová L, Emery O, Troilo M, et al. Gut microbiota structure differs between honeybees in winter and summer. *ISME J.* 2020;14:801–14. <https://doi.org/10.1038/s41396-019-0568-8>.
3. Crotti E, Rizzi A, Chouaia B, et al. Acetic acid bacteria, newly emerging symbionts of insects. *Appl Environ Microbiol.* 2010;76:6963–70. <https://doi.org/10.1128/AEM.01336-10>.
4. Kwong WK, Moran NA. Gut microbial communities of social bees. *Nat Rev Microbiol.* 2016;14:374–84. <https://doi.org/10.1038/nrmicro.2016.43>.
5. Maebe K, Vereecken NJ, Piot N, et al. The Holobiont as a key to the adaptation and conservation of wild bees in the Anthropocene. *Front Ecol Evol.* 2021;9:781470. <https://doi.org/10.3389/fevo.2021.781470>.
6. Dong Z-X, Li H-Y, Chen Y-F, et al. Colonization of the gut microbiota of honey bee (*Apis mellifera*) workers at different developmental stages. *Microbiol Res.* 2020;231:126370. <https://doi.org/10.1016/j.micres.2019.126370>.
7. Gruneck L, Khongphinitbunjong K, Popluechai S. Gut microbiota associated with two species of domesticated honey bees from Thailand. *Symbiosis.* 2021;83:335–45. <https://doi.org/10.1007/s13199-021-00754-8>.
8. Hubert J, Bicianova M, Ledvinka O, et al. Changes in the Bacteriome of Honey Bees Associated with the Parasite *Varroa destructor*, and Pathogens *Nosema* and *Lotmaria passim*. *Microb Ecol.* 2017;73:685–98. <https://doi.org/10.1007/s00248-016-0869-7>.
9. Liu P, Zhu Y, Ye L, et al. Overwintering honeybees maintained dynamic and stable intestinal bacteria. *Sci Rep.* 2021;11:22233. <https://doi.org/10.1038/s41598-021-01204-7>.
10. Ribière C, Hegarty C, Stephenson H, et al. Gut and whole-body microbiota of the Honey Bee separate thriving and non-thriving hives. *Microb Ecol.* 2019;78:195–205. <https://doi.org/10.1007/s00248-018-1287-9>.
11. Praet J, Parmentier A, Schmid-Hempel R, et al. Large-scale cultivation of the bumblebee gut microbiota reveals an underestimated bacterial species diversity capable of pathogen inhibition. *Environ Microbiol.* 2018;20:214–27. <https://doi.org/10.1111/1462-2920.13973>.
12. Graystock P, Rehan SM, McFrederick QS. Hunting for healthy microbiomes: determining the core microbiomes of Ceratina, Megalopta, and Apis bees and how they associate with microbes in bee collected pollen. *Conserv Genet.* 2017;18:701–11. <https://doi.org/10.1007/s10592-017-0937-7>.
13. Martínez AJ, Onchuru TO, Ingham CS, et al. Angiosperm to Gymnosperm host-plant switch entails shifts in microbiota of the *Welwitschia* bug, *Probergrothius angolensis* (distant, 1902). *Mol Ecol.* 2019;28:5172–87. <https://doi.org/10.1111/mec.15281>.
14. Hammer TJ, McMillan WO, Fierer N. Metamorphosis of a butterfly-associated bacterial community. *PLoS ONE.* 2014;9:e86995. <https://doi.org/10.1371/journal.pone.0086995>.
15. Hammer TJ, Dickerson JC, McMillan WO, Fierer N. *Heliconius* butterflies host characteristic and phylogenetically structured adult-stage microbiomes. *Appl Environ Microbiol.* 2020;86:e02007–20. <https://doi.org/10.1128/AEM.02007-20>.

16. Ravenscraft A, Berry M, Hammer T, et al. Structure and function of the bacterial and fungal gut microbiota of neotropical butterflies. *Ecol Monogr*. 2019;89. <https://doi.org/10.1002/ecm.1346>.
17. Ravenscraft A, Kish N, Peay K, Boggs C. No evidence that gut microbiota impose a net cost on their butterfly host. *Mol Ecol*. 2019;28:2100–17. <https://doi.org/10.1111/mec.15057>.
18. Tarpy DR, Mattila HR, Newton ILG. Development of the Honey Bee Gut Microbiome throughout the queen-rearing process. *Appl Environ Microbiol*. 2015;81:3182–91. <https://doi.org/10.1128/AEM.00307-15>.
19. Anderson KE, Ricigliano VA, Mott BM, et al. The queen's gut refines with age: longevity phenotypes in a social insect model. *Microbiome*. 2018;6:108. <https://doi.org/10.1186/s40168-018-0489-1>.
20. Copeland DC, Anderson KE, Mott BM. Early Queen Development in Honey Bees: Social Context and Queen Breeder Source Affect Gut Microbiota and Associated Metabolism. *Microbiol Spectr*. 2022;10:e00383–22. <https://doi.org/10.1128/spectrum.00383-22>.
21. Powell JE, Eiri D, Moran NA, Rangel J. Modulation of the honey bee queen microbiota: Effects of early social contact. *PLoS ONE*. 2018;13:e0200527. <https://doi.org/10.1371/journal.pone.0200527>.
22. Oren A, Garrity G. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol*. 2019;69:3313–4. <https://doi.org/10.1099/ijsem.0.003740>.
23. Roh SW, Nam YD, Chang HW, et al. Phylogenetic characterization of two novel commensal bacteria involved with innate immune homeostasis in *Drosophila melanogaster*. *Appl Environ Microbiol*. 2008;74:6171–7. <https://doi.org/10.1128/AEM.00301-08>.
24. Servin-Garciduenas LE, Sanchez-Quinto A, Martinez-Romero E. (2014) Draft Genome Sequence of Commensalibacter papalotli MX01, a Symbiont Identified from the Guts of Overwintering Monarch Butterflies. *Genome Announc* 2. <https://doi.org/10.1128/genomeA.00128-14>.
25. (2019) International Code of Nomenclature of Prokaryotes: Prokaryotic Code (2008 Revision.). *Int J Syst Evol Microbiol* 69:S1–S111. <https://doi.org/10.1099/ijsem.0.000778>.
26. Bonilla-Rosso G, Paredes Juan C, Das S et al. (2019) Acetobacteraceae in the honey bee gut comprise two distant clades with diverging metabolism and ecological niches. *bioRxiv*. <https://doi.org/10.1101/861260>.
27. Sizios S, Moran J, Chege M, et al. Complete reference Genome Assembly for Commensalibacter sp. Strain AMU001, an Acetic Acid Bacterium isolated from the gut of Honey Bees. *Microbiol Resour Announc*. 2019;8. <https://doi.org/10.1128/MRA.01459-18>.
28. Kim EK, Kim SH, Nam HJ, et al. Draft genome sequence of Commensalibacter intestini A911T, a symbiotic bacterium isolated from *Drosophila melanogaster* intestine. *J Bacteriol*. 2012;194:1246. <https://doi.org/10.1128/JB.06669-11>.
29. Winans NJ, Walter A, Chouaia B, et al. A genomic investigation of ecological differentiation between free-living and *Drosophila*-associated bacteria. *Mol Ecol*. 2017;26:4536–50. <https://doi.org/10.1111/mec.14232>.
30. Li L, Praet J, Borremans W, et al. Bombella intestini gen. nov., sp. nov., an acetic acid bacterium isolated from bumble bee crop. *Int J Syst Evol Microbiol*. 2015;65:267–73. <https://doi.org/10.1099/ijms.0.068049-0>.
31. Wieme AD, Spitaels F, Aerts M, et al. Effects of Growth Medium on Matrix-Assisted laser desorption–ionization time of Flight Mass Spectra: a case study of acetic acid Bacteria. *Appl Environ Microbiol*. 2014;80:1528–38. <https://doi.org/10.1128/AEM.03708-13>.
32. Dumolin C, Aerts M, Verheyde B, et al. Introducing SPeDE: high-throughput dereplication and Accurate determination of Microbial Diversity from Matrix-Assisted laser desorption–ionization time of Flight Mass Spectrometry Data. *mSystems*. 2019;4:e00437–19. <https://doi.org/10.1128/mSystems.00437-19>.
33. Ewels P, Magnusson M, Lundin S, Källner M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32:3047–8. <https://doi.org/10.1093/bioinformatics/btw354>.
34. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
35. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a New Genome Assembly Algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77. <https://doi.org/10.1089/cmb.2012.0021>.
36. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25:1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
37. Li H, Handsaker B, Wysoker A, et al. The sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl*. 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
38. Okonechnikov K, Conesa A, García-Alcalde F. (2015) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* btv566. <https://doi.org/10.1093/bioinformatics/btv566>.
39. Tian R, Imanian B. (2023) PlasmidHunter: Accurate and fast prediction of plasmid sequences using gene content profile and machine learning. *Bioinformatics*
40. Kans J. Entrez Direct: E-utilities on the Unix Command line. National Center for Biotechnology Information (US); 2022.
41. Parks DH, Imelfort M, Skennerton CT, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55. <https://doi.org/10.1101/gr.186072.114>.
42. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–5. <https://doi.org/10.1093/bioinformatics/btt086>.
43. Yoon S-H, Ha S-M, Kwon S, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol*. 2017;67:1613–7. <https://doi.org/10.1099/ijsem.0.001755>.
44. Ankenbrand MJ, Keller A. bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome*. 2016;59:783–91. <https://doi.org/10.1139/gen-2015-0175>.
45. Kozlov AM, Darriba D, Flouri T, et al. RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019;35:4453–5. <https://doi.org/10.1093/bioinformatics/btz305>.
46. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47:W256–9. <https://doi.org/10.1093/nar/gkz239>.
47. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun*. 2019;10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
48. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. 2013;14:60. <https://doi.org/10.1186/1471-2105-14-60>.
49. Yoon S-H, Ha S, Lim J, et al. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek*. 2017;110:1281–6. <https://doi.org/10.1007/s10482-017-0844-4>.
50. Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al. eggNOG-mapper v2: functional annotation, Orthology assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*. 2021;38:5825–9. <https://doi.org/10.1093/molbev/msab293>.
51. Huerta-Cepas J, Szklarczyk D, Heller D, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47:D309–14. <https://doi.org/10.1093/nar/gky1085>.
52. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 2015;43:D261–9. <https://doi.org/10.1093/nar/gku1223>.
53. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci*. 2022;31:47–53. <https://doi.org/10.1002/pro.4172>.
54. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci*. 2020;29:28–35. <https://doi.org/10.1002/pro.3711>.
55. Ruiz-Perez CA, Conrad RE, Konstantinidis KT. MicrobeAnnotator: a user-friendly, comprehensive functional annotation pipeline for microbial genomes. *BMC Bioinformatics*. 2021;22:11. <https://doi.org/10.1186/s12859-020-03940-5>.
56. Néron B, Denise R, Coluzzi C et al. (2022) MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. *Bioinformatics*
57. Guo J, Bolduc B, Zayed AA, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*. 2021;9:37. <https://doi.org/10.1186/s40168-020-00990-y>.
58. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.
59. Sombolostani AS, Cleenwerck I, Cnockaert M, et al. Characterization of novel *Gluconobacter* species from fruits and fermented food products: *Gluconobacter cadivus* sp. nov., *Gluconobacter vitians* sp. nov. and *Gluconobacter potus* sp. nov. *Int J Syst Evol Microbiol*. 2019;71. <https://doi.org/10.1099/ijsem.0.004751>.

60. Chun J, Oren A, Ventosa A, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol*. 2018;68:461–6. <https://doi.org/10.1099/ijsem.0.002516>.
61. Gupta A, Nair S. Dynamics of Insect–Microbiome Interaction influence host and Microbial Symbiont. *Front Microbiol*. 2020;11:1357. <https://doi.org/10.3389/fmicb.2020.01357>.
62. Pilon FM, Visóto LE, Guedes RNC, Oliveira MGA. Proteolytic activity of gut bacteria isolated from the velvet bean caterpillar *Anticarsia gemmatilis*. *J Comp Physiol B*. 2013;183:735–47. <https://doi.org/10.1007/s00360-013-0744-5>.
63. Parish AJ, Rice DW, Tanquary VM, et al. Honey bee symbiont buffers larvae against nutritional stress and supplements lysine. *ISME J*. 2022;16:2160–8. <https://doi.org/10.1038/s41396-022-01268-x>.
64. Paoli PP, Donley D, Stabler D, et al. Nutritional balance of essential amino acids and carbohydrates of the adult worker honeybee depends on age. *Amino Acids*. 2014;46:1449–58. <https://doi.org/10.1007/s00726-014-1706-2>.
65. Stabler D, Paoli PP, Nicolson SW, Wright GA. Nutrient balancing of the adult worker bumblebee (*Bombus terrestris*) depends on the dietary source of essential amino acids. *J Exp Biol*. 2015;218:793–802. <https://doi.org/10.1242/jeb.114249>.
66. Ferla MP, Patrick WM. Bacterial methionine biosynthesis. *Microbiology*. 2014;160:1571–84. <https://doi.org/10.1099/mic.0.077826-0>.
67. McCutcheon JP, McDonald BR, Moran NA. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci*. 2009;106:15394–9. <https://doi.org/10.1073/pnas.0906424106>.
68. Wargo MJ. Homeostasis and catabolism of Choline and Glycine Betaine: Lessons from *Pseudomonas aeruginosa*. *Appl Environ Microbiol*. 2013;79:2112–20. <https://doi.org/10.1128/AEM.03565-12>.
69. Arai H, Sakurai K, Ishii M. Metabolic features of *Acetobacter aceti*. In: Matsushita K, Toyama H, Tonouchi N, Okamoto-Kainuma A, editors. *Acetic acid Bacteria: Ecology and Physiology*. Tokyo: Springer Japan; 2016. pp. 255–71.
70. Vaudo AD, Tooker JF, Grozinger CM, Patch HM. Bee nutrition and floral resource restoration. *Curr Opin Insect Sci*. 2015;10:133–41. <https://doi.org/10.1016/j.cois.2015.05.008>.
71. Romeis J, Wackers FL. Feeding responses by female *Pieris brassicae* butterflies to carbohydrates and amino acids. *Physiol Entomol*. 2000;25:247–53. <https://doi.org/10.1046/j.1365-3032.2000.00188.x>.
72. Barker RJ, Lehner Y, IN COOPERATION WITH THE UNIVERSITY OF ARIZONA. Acceptance and sustenance value of naturally occurring sugars fed to newly emerged adult workers of honey bees (*Apis mellifera* L.). *J Exp Zool*. 1974;187:277–85. <https://doi.org/10.1002/jez.1401870211>.
73. Kusano T, Sato H. The sensitivity of tarsal chemoreceptors for sugars in the Cabbage Butterfly, *Pieris rapae crucivora* BOISDUVAL. *Appl Entomol Zool*. 1980;15:385–91. <https://doi.org/10.1303/aez.15.385>.
74. Zheng H, Nishida A, Kwong WK, et al. Metabolism of toxic sugars by strains of the Bee Gut Symbiont *Gilliamella apicola*. *mBio*. 2016;7:e01326–16. <https://doi.org/10.1128/mBio.01326-16>.
75. Taylor EJ, Goyal A, Guerreiro CIPD, et al. How Family 26 Glycoside Hydro-lases Orchestrate Catalysis on different polysaccharides. *J Biol Chem*. 2005;280:32761–7. <https://doi.org/10.1074/jbc.M506580200>.
76. Broeker J, Mechelke M, Baudrexel M, et al. The hemicellulose-degrading enzyme system of the thermophilic bacterium *Clostridium stercorarium*: comparative characterisation and addition of new hemicellulolytic glycoside hydrolases. *Biotechnol Biofuels*. 2018;11:229. <https://doi.org/10.1186/s13068-018-1228-3>.
77. Zheng H, Perreau J, Powell JE, et al. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc Natl Acad Sci*. 2019;116:25909–16. <https://doi.org/10.1073/pnas.1916224116>.
78. Matsushita K, Toyama H, Tonouchi N, Okamoto-Kainuma A. *Acetic acid Bacteria: Ecology and Physiology*. Tokyo: Springer Japan; 2016.
79. Sainz F, Jesús Torija M, Matsutani M, et al. Determination of dehydrogenase activities involved in D-Glucose oxidation in *Glucobacter* and *Acetobacter* strains. *Front Microbiol*. 2016;7. <https://doi.org/10.3389/fmicb.2016.01358>.
80. Lynch KM, Zannini E, Wilkinson S, et al. Physiology of acetic acid Bacteria and their role in vinegar and fermented beverages. *Compr Rev Food Sci Food Saf*. 2019;18:587–625. <https://doi.org/10.1111/1541-4337.12440>.
81. Büsch A, Friedrich B, Cramm R. Characterization of the *norB* Gene, encoding nitric oxide reductase, in the Nondenitrifying Cyanobacterium *Synechocystis* sp. Strain PCC6803. *Appl Environ Microbiol*. 2002;68:668–72. <https://doi.org/10.1128/AEM.68.2.668-672.2002>.
82. Kobayashi I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res*. 2001;29:3742–56. <https://doi.org/10.1093/nar/29.18.3742>.
83. Zhang S-P, Wang Q, Quan S-W, et al. Type II toxin–antitoxin system in bacteria: activation, function, and mode of action. *Biophys Rep*. 2020;6:68–79. <https://doi.org/10.1007/s41048-020-00109-8>.
84. Sitaraman R. The role of DNA restriction-modification Systems in the Biology of *Bacillus anthracis*. *Front Microbiol*. 2016;7. <https://doi.org/10.3389/fmicb.2016.00011>.
85. Singh G, Yadav M, Ghosh C, Rathore JS. Bacterial toxin-antitoxin modules: classification, functions, and association with persistence. *Curr Res Microb Sci*. 2021;2:100047. <https://doi.org/10.1016/j.crmicr.2021.100047>.
86. Chen NH, Djoko KY, Veyrier FJ, McEwan AG. Formaldehyde stress responses in bacterial pathogens. *Front Microbiol*. 2016;7. <https://doi.org/10.3389/fmicb.2016.00257>.
87. Erban T, Sopko B, Kadlikova K, et al. *Varroa destructor* parasitism has a greater effect on proteome changes than the deformed wing virus and activates TGF- β signaling pathways. *Sci Rep*. 2019;9:9400. <https://doi.org/10.1038/s41598-019-45764-1>.
88. Cini A, Meriggi N, Bacci G, et al. Gut microbial composition in different castes and developmental stages of the invasive hornet *Vespa velutina nigrithorax*. *Sci Total Environ*. 2020;745:140873. <https://doi.org/10.1016/j.scitotenv.2020.140873>.
89. Vannette RL. The Floral Microbiome: Plant, Pollinator, and microbial perspectives. *Annu Rev Ecol Evol Syst*. 2020;51:363–86. <https://doi.org/10.1146/annurev-ecolsys-011720-013401>.
90. Parmentier L, Meeus I, Cheroutre L, et al. Commercial bumblebee hives to assess an anthropogenic environment for pollinator support: a case study in the region of Ghent (Belgium). *Environ Monit Assess*. 2014;186:2357–67. <https://doi.org/10.1007/s10661-013-3543-2>.
91. Engel P, Kwong WK, McFrederick Q, et al. The Bee Microbiome: impact on Bee Health and Model for Evolution and Ecology of host-microbe interactions. *mBio*. 2016;7:e02164–15. <https://doi.org/10.1128/mBio.02164-15>.
92. Bobay L-M, Wissel EF, Raymann K. Strain structure and Dynamics revealed by targeted deep sequencing of the Honey Bee Gut Microbiome. *mSphere*. 2020;5:e00694–20. <https://doi.org/10.1128/mSphere.00694-20>.
93. Engel P, Stepanauskas R, Moran NA. Hidden diversity in Honey Bee Gut Symbionts detected by single-cell Genomics. *PLoS Genet*. 2014;10:e1004596. <https://doi.org/10.1371/journal.pgen.1004596>.
94. Powell E, Ratnayake N, Moran NA. Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. *Mol Ecol*. 2016;25:4461–71. <https://doi.org/10.1111/mec.13787>.
95. Ellegaard KM, Engel P. Genomic diversity landscape of the honey bee gut microbiota. *Nat Commun*. 2019;10:446. <https://doi.org/10.1038/s41467-019-08303-0>.
96. Ellegaard KM, Suenami S, Miyazaki R, Engel P. Vast differences in strain-level diversity in the gut microbiota of two closely related Honey Bee Species. *Curr Biol*. 2020;30:2520–2531e7. <https://doi.org/10.1016/j.cub.2020.04.070>.
97. Kwong WK, Engel P, Koch H, Moran NA. Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc Natl Acad Sci*. 2014;111:11509–14. <https://doi.org/10.1073/pnas.1405838111>.
98. Martinson VG, Danforth BN, Minckley RL, et al. A simple and distinctive microbiota associated with honey bees and bumble bees: THE MICROBIOTA OF HONEY BEES AND BUMBLE BEES. *Mol Ecol*. 2011;20:619–28. <https://doi.org/10.1111/j.1365-294X.2010.04959.x>.
99. Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci*. 2012;109:11002–7. <https://doi.org/10.1073/pnas.1202970109>.
100. Lee FJ, Miller KI, McKinlay JB, Newton ILG. Differential carbohydrate utilization and organic acid production by honey bee symbionts. *FEMS Microbiol Ecol*. 2018;94. <https://doi.org/10.1093/femsec/fiy113>.
101. Zhang W, Zhang X, Su Q, et al. Genomic features underlying the evolutionary transitions of *apibacter* to honey bee gut symbionts. *Insect Sci*. 2022;29:259–75. <https://doi.org/10.1111/1744-7917.12912>.
102. Praet J, Aerts M, Brandt ED, et al. *Apibacter mensalis* sp. nov.: a rare member of the bumblebee gut microbiota. *Int J Syst Evol Microbiol*. 2016;66:1645–51. <https://doi.org/10.1099/ijsem.0.000921>.
103. Härer L, Hilgarth M, Ehrmann MA. Comparative Genomics of Acetic acid Bacteria within the Genus *Bombella* in Light of Beehive Habitat Adaptation. *Microorganisms*. 2022;10:1058. <https://doi.org/10.3390/microorganisms10051058>.

104. Smith EA, Newton ILG. Genomic signatures of Honey Bee Association in an Acetic Acid Symbiont. *Genome Biol Evol.* 2020;12:1882–94. <https://doi.org/10.1093/gbe/evaa183>.
105. Li L, Illegghems K, Van Kerrebroeck S, et al. Whole-genome sequence analysis of *Bombella intestini* LMG 28161T, a Novel Acetic Acid Bacterium isolated from the crop of a red-tailed Bumble Bee, *Bombus lapidarius*. *PLoS ONE.* 2016;11:e0165611. <https://doi.org/10.1371/journal.pone.0165611>.
106. Alberoni D, Di Gioia D, Baffoni L. Alterations in the microbiota of caged honeybees in the Presence of *Nosema ceranae* infection and related changes in functionality. *Microb Ecol.* 2022. <https://doi.org/10.1007/s00248-022-02050-4>.
107. Alberoni D, Gaggia F, Baffoni L, et al. *Bifidobacterium xylocopae* sp. nov. and *Bifidobacterium aemilianum* sp. nov., from the carpenter bee (*Xylocopa violacea*) digestive tract. *Syst Appl Microbiol.* 2019;42:205–16. <https://doi.org/10.1016/j.syapm.2018.11.005>.
108. Kwong WK, Steele MI, Moran NA. Genome sequences of *Apibacter* spp., Gut Symbionts of Asian Honey Bees. *Genome Biol Evol.* 2018;10:1174–9. <https://doi.org/10.1093/gbe/evy076>.
109. Li Y, Leonard SP, Powell JE, Moran NA. Species divergence in gut-restricted bacteria of social bees. *Proc Natl Acad Sci.* 2022;119:e2115013119. <https://doi.org/10.1073/pnas.2115013119>.
110. Ellegaard K, Brochet S, Bonilla-Rosso G et al. (2018) Genomic changes underlying host specialization in the bee gut symbiont *Lactobacillus Firm5*. *Microbiology.*
111. Liu Y, Chen J, Lang H, Zheng H. *Bartonella choladocola* sp. nov. and *Bartonella apihabitans* sp. nov., two novel species isolated from honey bee gut. *Syst Appl Microbiol.* 2022;45:126372. <https://doi.org/10.1016/j.syapm.2022.126372>.
112. Lugli GA, Fontana F, Tarracchini C, et al. Exploring the biodiversity of *Bifidobacterium asteroides* among honey bee microbiomes. *Environ Microbiol.* 2022;24:5666–79. <https://doi.org/10.1111/1462-2920.16223>.
113. Zhang Z, Huang M, Qiu L, et al. Diversity and functional analysis of chinese bumblebee gut microbiota reveal the metabolic niche and antibiotic resistance variation of *Gilliamella*. *Insect Sci.* 2021;28:302–14. <https://doi.org/10.1111/1744-7917.12770>.
114. Cornet L, Cleenwerck I, Praet J, et al. Phylogenomic analyses of *Snodgrassella* isolates from Honeybees and Bumblebees reveal taxonomic and functional diversity. *mSystems.* 2022;7:e01500–21. <https://doi.org/10.1128/mSystems.01500-21>.
115. The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC Genomics.* 2017;18:151. <https://doi.org/10.1186/s12864-017-3543-7>.
116. Bohlin J, Eldholm V, Pettersson JHO, et al. Dynamic microbiome evolution in social bees. *Sci Adv.* 2017;3:e1600513. <https://doi.org/10.1126/sciadv.1600513>.
117. Jaspers E, Overmann J. Ecological significance of Microdiversity: identical 16S rRNA gene sequences can be found in Bacteria with highly divergent genomes and ecophysiologicals. *Appl Environ Microbiol.* 2004;70:4831–9. <https://doi.org/10.1128/AEM.70.8.4831-4839.2004>.
118. Larkin AA, Martiny AC. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep.* 2017;9:55–70. <https://doi.org/10.1111/1758-2229.12523>.
119. Nakamura A, Kurihara S, Takahashi D, et al. Symbiotic polyamine metabolism regulates epithelial proliferation and macrophage differentiation in the colon. *Nat Commun.* 2021;12:2105. <https://doi.org/10.1038/s41467-021-22212-1>.
120. Quinn A, Chazli YE, Escrig S et al. (2023) Foraging on host synthesized metabolites enables the bacterial symbiont *Snodgrassella alvi* to colonize the honey bee gut. *Microbiology.*

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.