Genetic divergence and functional convergence of gut bacteria between the Eastern honey bee *Apis cerana* and the Western honey bee *Apis mellifera*

Yuqi Wu [a,1], Yufei Zheng [a,b,1], Shuai Wang [a,1], Yanping Chen [c], Junyi Tao [d], Yanan Chen [a], Gongwen Chen [a], Hongxia Zhao [e], Kai Wang [f], Kun Dong [g], Fuliang Hu [a,*], Ye Feng [b,h,*], Huoqing Zheng [a,*]

[a] *College of Animal Sciences, Zhejiang University, Hangzhou, China*
[b] *Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China*
[c] *USDA-ARS Bee Research Laboratory, Beltsville, MD, USA*
[d] *Department of Surgery, Miller School of Medicine, University of Miami, Miami, FL, USA*
[e] *Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Guangdong Institute of Applied Biological Resources, Guangzhou, China*
[f] *Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China*
[g] *Eastern Bee Research Institute, College of Animal Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, China*
[h] *Institute for Translational Medicine, Zhejiang University School of Medicine, Hangzhou, China*

HIGHLIGHTS

- The inter-species diversity of *A. cerana* and *A. mellifera* core gut bacteria was revealed.
- Core bacterial species of *A. cerana* and *A. mellifera* are distinctive in function.
- Functional profile of overall gut community of *A. cerana* and *A. mellifera* are similar.
- Metabolome showed that *A. cerana* and *A. mellifera* gut bacteria have similar metabolic capability.
- *A. cerana* and *A. mellifera* core gut bacteria have no strict host specificity.

GRAPHICAL ABSTRACT
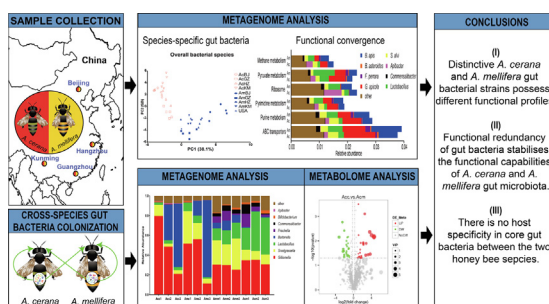
ARTICLE INFO

ABSTRACT

Introduction: The functional relevance of intra-species diversity in natural microbial communities remains largely unexplored. The guts of two closely related honey bee species, *Apis cerana* and *A. mellifera*, are colonised by a similar set of core bacterial species composed of host-specific strains, thereby providing a good model for an intra-species diversity study.

Abbreviations: ABC, ATP-binding cassette; Ac, *Apis cerana*; Am, *Apis mellifera*; Acc, *A. cerana* workers with *A. cerana* gut microbiota; Acm, *A. cerana* workers with *A. mellifera* gut microbiota; Amc, *A. mellifera* workers with *A. cerana* gut microbiota; Amm, *A. mellifera* workers with *A. mellifera* gut microbiota; AcBJ, *A. cerana* workers sampled in Beijing; AcGZ, *A. cerana* workers sampled in Guangzhou; AcHZ, *A. cerana* workers sampled in Hangzhou; AcKM, *A. cerana* workers sampled in Kunming; AmBJ, *A. mellifera* workers sampled in Beijing; AmGZ, *A. mellifera* workers sampled in Guangzhou; AmHZ, *A. mellifera* workers sampled in Hangzhou; AmKM, *A. mellifera* workers sampled in Kunming; CAZyme, Carbohydrate-active enzyme; CTAB, Cetyltrimethyl Ammonium Bromide; FDR, False Discovery Rate; GO, Gene ontology; HMDB, Human metabolites database; KEGG, Kyoto Encyclopedia of Genes and Genomes; KO, KEGG ortholog; LCA, Lowest common ancestor-based algorithm; LC-MS, Liquid chromatography mass spectrometry; NCBI, National Center of Biotechnology Information, USA; ORFs, Open reading frames; PBS, Phosphate Buffer Saline; PCA, Principal component analysis; PCoA, Principal coordinates analysis; PTS, Phosphotransferase system; SNP, Single nucleotide polymorphism; SNVs, Single nucleotide variants; SRA, Sequence read archive; VIP, Variable important in projection.

Peer review under responsibility of Cairo University.

* Corresponding authors.

E-mail addresses: flhu@zju.edu.cn (F. Hu), pandafengye@zju.edu.cn (Y. Feng), hqzheng@zju.edu.cn (H. Zheng).

[1] These authors contributed equally.

Objectives: We aim to assess the functional relevance of intra-species diversity of *A. cerana* and *A. mellifera* gut microbiota.

Methods: Honey bee workers were collected from four regions of China. Their gut microbiomes were investigated by shotgun metagenomic sequencing, and the bacterial compositions were compared at the species level. A cross-species colonisation assay was conducted, with the gut metabolomes being characterised by LC-MS/MS.

Results: Comparative analysis showed that the strain composition of the core bacterial species was host-specific. These core bacterial species presented distinctive functional profiles between the hosts. However, the overall functional profiles of the *A. cerana* and *A. mellifera* gut microbiomes were similar; this was further supported by the consistency of the honey bees' gut metabolome, as the gut microbiota of different honey bee species showed rather similar metabolic profiles in the cross-species colonisation assay. Moreover, this experiment also demonstrated that the gut microbiota of *A. cerana* and *A. mellifera* could cross colonise between the two honey bee species.

Conclusion: Our findings revealed functional differences in most core gut bacteria between the guts of *A. cerana* and *A. mellifera*, which may be associated with their inter-species diversity. However, the functional profiles of the overall gut microbiomes between the two honey bee species converge, probably as a result of the overlapping ecological niches of the two species. Our findings provide critical insights into the evolution and functional roles of the mutualistic microbiota of honey bees and reveal that functional redundancy could stabilise the gene content diversity at the strain-level within the gut community.

Introduction

Animals carry massive and diverse communities of symbiotic microbes in their gastrointestinal tracts [1]. Some of these communities play vital roles in host health [2-5]. There has been growing evidence that intra-species level or strain level diversity is omnipresent in the gut microbiota because the same species of bacteria consists of a coherent group of strains [6,7], which may display gene content variation and nucleotide polymorphism [8,9]. However, the functional contribution of intra-species genomic variation to the gut microbiome remains obscure in natural microbial communities. The overall functional profile (see Glossary) of the gut microbiota can change significantly, with only minor changes being displayed in the taxonomic profiles [10,11], providing indirect evidence for the functional relevance of intra-species diversity [12]. Distinct gut microbiomes can also lead to similar microbial genetic profiles. Shan et al. [13] found that high-energy diets with different fat-to-sugar ratios can induce different gut microbiota that share similar genetic and metabolite compositions. Contradictory findings have shown that strain-level diversity can lead to a similar or different functional gene profile among the gut microbiomes of different hosts; however, what drives these different outcomes is still unknown. Addressing this question can significantly improve our understanding of the functional relevance of intra-species diversity.

Honey bees (*Apis* spp.) have unique advantages as model organisms for gut microbiota studies [14], particularly in terms of intra-species diversity. Firstly, honey bees harbour a simple and conservative gut community, mainly consisting of 8–10 phylotypes (see Glossary) [15,16]. Secondly, multiple studies have revealed that high strain diversity exists in the major lineages of *Apis mellifera* guts [7,17-19], and host-specific intra-species diversity exists among bee species [15,20,21]. Finally, the honey bee gut microbiota reveals parallels with the human gut microbiota; they have both presumably coevolved with their host over millions of years and are transmitted primarily through social interactions [14,15,22,23].

The Eastern honey bee *Apis cerana* and the Western honey bee *A. mellifera* are the only two honey bee species that have been commercially reared for food crop pollination; they have immense economic and ecological value [24,25]. Historically, *A. cerana* and *A. mellifera* were geographically isolated in Asia and Europe-Africa, respectively. As a result of allopatric speciation, these two honey bee species demonstrate different behavioural and physiological characteristics, as reviewed by Oldroyd et al. [26]. *A. mellifera* was introduced into Asia with the development of the modern beekeeping industry. However, in recent years, the health and populations of *A. cerana* and *A. mellifera* have been declining in many parts of the world [27-29]. While the historically geographical isolation might result in the differentiation of gut microbiota between the two honey bee species, the modern sympatry has provided opportunities for the cross-transmission of gut microbiota between the two species. The comparative study of the gut microbiota of these two bee species would provide important insights into the differentiation and host specificity of gut microbiota and also help understand the function of these gut bacteria in the health of honey bee individuals and their colonies [30-33].

Recently, Ellegaard et al. [21] compared the composition of gut bacteria between *A. cerana* and *A. mellifera*. While compositional difference at the strain level was identified for core gut bacteria, it remains unclear whether the bacterial difference would bring about metabolic differences between the two honey bee species and then affect their physiology. Because the two honey bee species have adapted to similar niches and played similar ecological roles, we hypothesised that the gut bacteria of them are similar in overall functional profiles. To test this hypothesis, we employed shotgun metagenomic sequencing to comprehensively analyse the resident gut microbiota of *A. cerana* and *A. mellifera* colonies in China (Fig. 1A). Furthermore, we conducted a cross-species colonisation assay and characterised the metabolomes of workers harbouring different gut microbiota to uncover whether these strain-specific bacteria have host specificity (Fig. 1B). Our study confirmed that the core bacterial species of honey bees consist of distinctive strains, which differ significantly in their functional profiles. Interestingly, pathway reconstruction based on the overall gut microbiomes and metabolic analyses suggested that the overall gut communities of *A. cerana* and *A. mellifera* have similar capabilities. Our results also suggested that there is no host specificity in core gut bacteria between *A. cerana* and *A. mellifera*.

Materials and methods

Ethics statement

The owners of the sampled apiaries gave us permissions to collect honey bee samples. No additional permits were required because the study did not involve endangered or protected species.
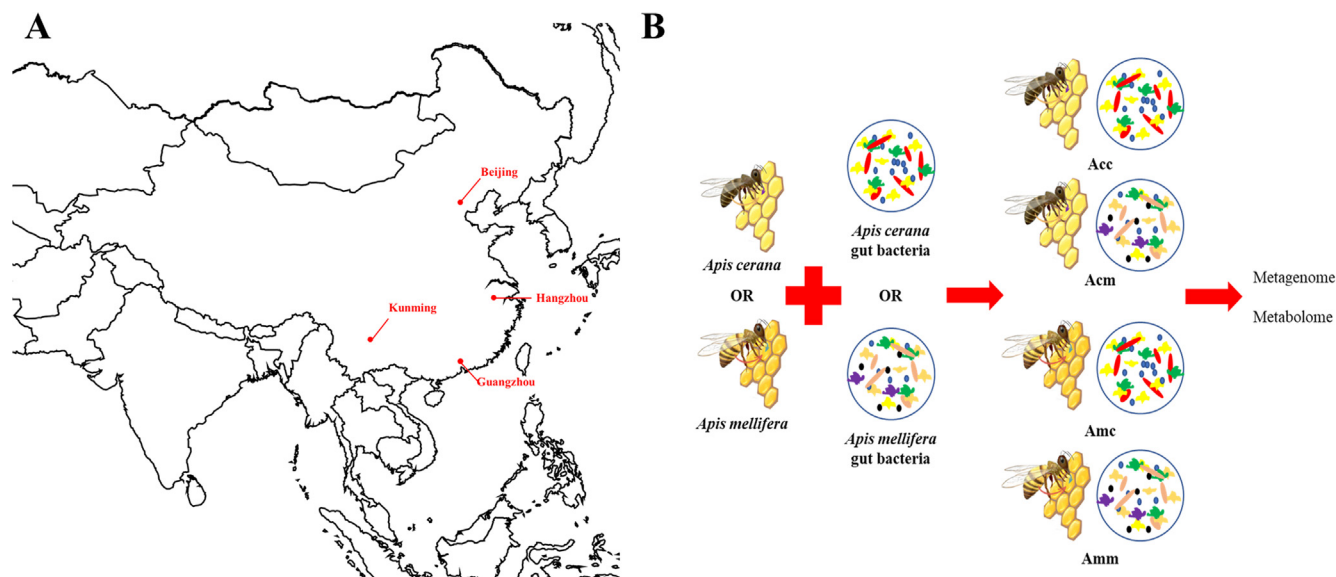
Fig. 1. The experimental design of our study. (A) The sampling sites of field honey bees were Beijing, Guangzhou, Hangzhou, and Kunming. In each location, five *Apis cerana* colonies and five *A. mellifera* colonies were sampled from the same apiary. (B) The experimental procedures of cross-species colonisation assay. Gnotobiotic workers were colonised with *A. cerana* or *A. mellifera* gut bacteria, and then used for metagenomic and metabolic analyses.

Field honey bee sample collection

Honey bee workers from 20 *A. cerana* colonies and 20 *A. mellifera* colonies were collected from four apiaries located in Kunming, Hangzhou, Beijing, and Guangzhou in June 2019. In each apiary, five *A. cerana* and five *A. mellifera* colonies were sampled. To collect foraging workers, colony entrances were blocked, and approximately 100 workers carrying pollen were sampled. After chilling bees at 4 °C to immobilise them, the ileum and rectum (hindgut) of individual bees were dissected using sterile forceps and iris scissors. A total of 50 dissected guts from each colony were pooled together, submerged in 50% glycerol, and stored at −80 °C.

DNA extraction, library preparation, and sequencing

Before DNA extraction, the gut bacteria were enriched according to Ellegaard et al. [7], in which the gut tissue was homogenised with a bead-beater using glass beads. The homogenates were centrifuged at 500 g for 5 min to remove debris, and the supernatant was collected into new Eppendorf tubes. The samples were then centrifuged at 10000g for 10 min to pellet the bacterial cells. The supernatant was removed, and the bacterial pellets were resuspended in PBS and centrifuged again at 500g for 5 min. Finally, the samples were centrifuged at 10,000g for 15 min to pellet the bacterial cells.

Bacterial DNA was extracted using a CTAB-based DNA extraction protocol. For the DNA sample preparation, 200 ng DNA per sample was used as input material. First the DNA sample was fragmented by sonication to a size of 350 bp. Then, sequencing libraries were constructed using the NEBNext Ultra DNA Library Prep Kit (NEB, USA), and index codes were added to attribute sequences to each sample. Specifically, the Chip DNA was purified using the AMPure XP system (Beckman Coulter, USA). After adenylation of the 3′ ends of DNA fragments, the NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridisation. Following this, 3 μL USER Enzyme (NEB, USA) was added to the adaptor-ligated DNA, and the mixture was incubated at 37 °C for 15 min followed by 95 °C for 5 min. PCR was carried out with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) primers provided in NEBNext Multiplex (NEB, USA).

Finally, PCR products were purified (AMPure XP system) and library quality was assessed using the Agilent Bioanalyzer 2100 system. Sequencing was performed on an Illumina NovaSeq system by Novagene Co., Ltd.

Sequence quality control and assembly

To obtain high-quality reads, the raw fastq data from the Illumina NovaSeq platform were trimmed to remove low-quality (30% of bases with Q < 15 per read), ambiguous bases (N > 5% per read) and the overlap with adapter sequences of ≥ 5 bp. For each sample, scaffolds were then assembled using Megahit v1.2 with parameters "- k-list 45,55,67,73".

Metagenome assembly and gene prediction

For scaftigs (see Glossary) >500 bp in each sample, the open reading frames (ORFs) were predicted using MetaGeneMark with the MetaGeneMark_v1.mod model. The CD-HIT program [34] was used to obtain an initial non-redundant gene catalogue (nrGC) from the predicted ORFs (>100 bp) with parameters "-c 0.95, -aS 0.9". The abundance of non-redundant genes (unigenes) was measured by realigning the reads to the nrGC using SoapAligner [35]. Genes with ≤ 2 mapped sequences in all samples were excluded from further analysis. To calculate the relative abundance of a gene, the number of reads that were aligned to the gene was scaled by the gene length and the total number of reads that were mapped to the entire non-redundant gene catalogue for each sample.

Taxonomic and functional assignment of genes

To obtain the taxonomic information of unigenes, DIAMOND v0.9 was used to search the assembly against the NCBI microNR database (including bacteria, fungi, archaea, and viruses, version 13 July 2019) with a blastp, e value $\leq 1 \times 10^{-5}$ threshold. The taxonomic information of each gene was determined using the lowest common ancestor-based algorithm (LCA) implemented in MEGAN (version 4, Tübingen, Baden-Württemberg, Germany). Alpha diversity (see Glossary) was calculated using vegan [36] and picante

packages [37] in the R statistical software [38]. The taxonomic compositions at the genus and species levels were organised as 0/1 (presence/absence) matrix and were input into the PanGP program [39] for core/pan analyses with the default parameter.

Kyoto Encyclopedia of Genes and Genomes (KEGG) [40] annotation and Carbohydrate-Active enzymes Database (CAZy) [41] of each gene were also performed using DIAMOND against the KEGG database and CAZy database (version: 25 November 2014), respectively. The relative abundance of the annotated functional features was calculated by summing the relative abundances of the genes annotated to the features.

Strain-level nucleotide diversity analysis

The strains of *Bartonella apis* BBC0122 (NZ_CP015625.1) [42], *Bifidobacterium asteroides* PRL2011 (NC_018720.1) [43], *Frischella perrara* PEB0191 (NZ_CP009056.1) [44], *Gilliamella apicola* wkB1 (NZ_CP007445.1) [45], and *Snodgrassella alvi* wkB2 (NZ_CP007446.1) [45] were selected as representatives of the five core species according to Ellegaard et al. [7]. Their complete genomes were employed as references for mapping the raw reads of the five bacterial species in our study. The reference genes, which were defined in this study as the genes conserved among the five representative genomes, were identified by pairwise BLAST searches (blastp, e value $\leq 1 \times 10^{-20}$) (see additional file 8 at https://doi.org/10.17605/OSF.IO/SEMYG). Bowtie2 was used for mapping the raw reads against the nucleotide sequences of these reference genes [46]. Based on the alignment results, single nucleotide polymorphism (SNP) were identified using Samtools, mpileup, and Vcftools [47]. The SNPs located within the reference genes were used as the input for principal component analysis (PCA) by using PLINK v2.0 with default parameters [48]. Furthermore, the dominant allele sequences of these reference genes for each sample were concatenated for their use as inputs for the subsequent evolutionary analyses. MEGA v10.0.5, was used to construct the neighbour-joining trees, with the Jukes-Cantor (JC) substitution model and 100 bootstrap replicates [49]. The pairwise nucleotide distances were also calculated by using MEGA with the JC model.

Strain-level gene content analysis

For each of the five core bacterial species, the ORFs assigned to each species were extracted according to the taxonomic assignment. Within each species, the homologous relationship among the predicted ORFs was determined by comparing the protein sequences of the ORFs with the blastclust program in the NCBI BLAST package; the parameters 0.5 length coverage (-L) and 80% protein identity (-S) were applied. After binning the ORFs into gene clusters, the abundance of each gene cluster was obtained by summing the abundances of all ORFs belonging to the gene cluster. A gene cluster with an abundance > 0 was considered to be present in the sample; otherwise, it was considered absent. The pairwise similarity of gene content between samples was defined as the number of shared gene clusters divided by the number of identified gene clusters for the compared samples.

Strain-level functional analysis

For each of the five core bacterial species, the ORFs assigned to each species were extracted according to the taxonomic assignment. Subsequently, the ORFs were binned into functional units according to their functional annotations. The abundance of each functional unit was obtained by summing the abundances of ORFs belonging to the functional unit. Finally, principal coordinate analysis (PCoA) and principal component analysis (PCA) were performed on the abundances of functional units among the samples.

Cross-species colonisation assays

A. cerana and *A. mellifera* gut bacteria were prepared as follows: 10 honey bee forager workers were sampled near the entrance of a colony of each honey bee species in Hangzhou. Their hindguts were dissected immediately and homogenised in 1 mL PBS. The gut homogenate was centrifuged at 10,000 g for 10 min, and the supernatant was removed to eliminate possible virus contamination. To prepare pollen containing gut bacteria, the isolated bacteria were resuspended in 1 mL PBS, and 100 μL suspension was mixed with sterilised pollen.

Germ-free *A. cerana* and *A. mellifera* workers were obtained using the protocol described by Zheng et al. [31]. The germ-free workers were divided into four groups for the cross-species colonisation assays: 1) Acc: *A. cerana* workers inoculated with *A. cerana* gut microbiota; 2) Amc: *A. mellifera* workers inoculated with *A. cerana* gut microbiota; 3) Amm: *A. mellifera* workers inoculated with *A. mellifera* gut microbiota; and 4) Acm: *A. cerana* workers inoculated with *A. mellifera* gut microbiota. Each group consisted of 60 workers placed in a single cage. Different groups of workers were supplied with pollen containing respective gut bacteria described for 5 days and then switched to sterilised pollen. Approximately 1×10^5 bacterial cells were fed to each honey bee worker. In addition, another cage of germ-free *A. cerana* and germ-free *A. mellifera* workers with no inoculation of gut bacteria were used as a negative control to confirm that honey bees were not contaminated by bacteria from other sources, and contamination was not observed. Honey bees were housed according to standard methods [50]. On day ten, from each cage, forty workers were sampled and pooled for metagenome sequencing as described above, ten workers were sampled and pooled for untargeted metabolomics analysis, and five workers were sampled and used for 16s rDNA quantification individually. The experiment was conducted in triplicate using three *A. cerana* and three *A. mellifera* colonies.

Quantification of the absolute bacterial loads of cross-species colonisation honey bee samples

The absolute bacterial loads of Acc, Acm, Amm, and Amc workers were determined by qPCR using universal bacterial 16S rRNA primers (F: 5′-AGAGTTTGATCCTGGCTCAG-3′, R: 5′-CTGCTGCCTCC CGTAGGAGT-3′) [22]. Workers were frozen at −20 °C until immobilised, and their guts were immediately dissected in RNase-free water and used for DNA extraction. DNA was extracted using a TIA-Namp Stool DNA Kit (Tiangen Biotech Co., Ltd, Beijing, China) following the manufacturer's recommendations. 16 s rRNA copy numbers were quantified using the StepOne Plus real-time PCR system with thermal cycling conditions as follows: initial denaturation at 95 °C for 30 s, 40 cycles of 95 °C for 5 s and 60 °C for 30 s, followed by a melting curve from 60 °C to 95 °C at 0.5 °C/5 s increments.

Metabolite extraction

The replicates of each group were shipped on dry ice to Novogene Corporation (Beijing, China) for metabolomic analysis. Pooled honey bee hindguts were ground with liquid nitrogen, and the homogenate was resuspended with pre-chilled 80% methanol and 0.1% formic acid. The suspensions were cooled on ice for 5 min and then centrifuged at 15,000g at 4 °C for 5 min. The supernatant was diluted to a final concentration containing 53% methanol using LC-MS grade water. The solution was transferred to a fresh Eppendorf tube and centrifuged at 15,000g at 4 °C for 10 min to obtain the supernatant for analysis.

LC-MS/MS analysis was conducted using a Vanquish UHPLC system (Thermo Fisher, Waltham, Massachusetts, United States) coupled with an Orbitrap Q Exactive series mass spectrometer (Thermo Fisher, Waltham, Massachusetts, United States). Samples were injected onto a Hyperil Gold column (100 × 2.1 mm, 1.9 μm) using a 16-min linear gradient at a flow rate of 0.2 mL/min. The eluents for the positive polarity mode were eluent A (0.1% FA in water) and eluent B (methanol). The eluents for the negative polarity mode were eluent A (5 mM ammonium acetate, pH 9.0) and eluent B (methanol). The gradient elution was set as follows: 1.5 min, 2% B; 12.0 min, 2–100% B; 14.0 min, 100% B; 14.1 min, 100–2% B; 17 min, 2% B. The Q Exactive series mass spectrometer was operated in positive/negative polarity mode with a spray voltage of 3.2 kV, capillary temperature of 320 °C, sheath gas flow rate of 35 arb, and aux gas flow rate of 10 arb.

Untargeted metabolomics data analysis

Compound Discoverer 3.1 (Thermo Fisher, Waltham, Massachusetts, United States) was used to perform peak alignment, peak picking, and quantitation for each metabolite, with the following settings: retention time tolerance, 0.2 min; actual mass tolerance, 5 ppm; signal intensity tolerance, 30%; signal/noise ratio, 3; and minimum intensity, 100,000. The peak intensities were then normalised to the total spectral intensity. The normalised data were used to predict the molecular formula based on the additive ions, molecular ion peaks, and fragment ions. Peaks were then matched with the mzCloud, mzVault, and MassList databases to obtain accurate qualitative and relative quantitative results. The area normalisation method was applied to the data that were not normally distributed. The metabolites were annotated using the HMDB database (http://www.hmdb.ca/), Lipidmaps database (http://www.lipidmaps.org/), and KEGG database (http://www.genome.jp/kegg/). PCA and hierarchical clustering were performed using MetaboAnalyst [51]. Statistical significance (P-value) was calculated using univariate analysis (*t*-test).

Statistical analysis

Statistical analysis was performed using the R software v3.6.1 [38]. The significant taxonomic and functional differences between *A. cerana* and *A. mellifera* samples were determined by Student's t-tests. For comparisons of the pairwise distance of SNP, gene content similarity, and 16 s rDNA copy number, Kruskal-Wallis tests were used. All P values were adjusted using the false discovery rate (FDR) method, and adjusted P values < 0.05, were considered as statistically significant [52]. Hierarchical clustering was conducted using the hclust function in the stats package. PCA was performed using the prcomp function and PCoA based on the Bray-Curtis distance was performed using the ape library [53]. The Adonis test was used to distinguish group similarity, and the statistic R^2 represents effect size, which reveals the percentage of variation in distances explained by the grouping being tested. The Mantel test was used to test for differences across geographical populations. Metabolites with VIP > 1, P value < 0.05, and fold change ≥ 2 or fold change ≤ 0.5, were considered to be differential metabolites. The differential abundance of taxonomy and functional category was tested using Metastat [54].

Availability of data and material

The raw sequence data from metagenome shotgun sequencing were submitted to NCBI SRA under BioProject accession numbers PRJNA599289 (field honey bees) and PRJNA599288 (cross-species colonisation assay). All data generated or analysed during this study are included in this published article and its supplementary information files at OSF, https://doi.org/10.17605/OSF.IO/SEMYG.

Results

Characterisation of the Apis cerana and A. mellifera gut community

High-throughput sequencing metagenomics and the *de novo* assembly of paired-end reads resulted in 1,765,642 scaffolds with an average length of 1.36 kb. Detailed quality metrics, including the number of reads, reads mapped to contigs, reads that were successfully annotated, and reads that were mapped to honey bee genomes are listed in supplementary file Table S1 in the Open Science Framework, https://doi.org/10.17605/OSF.IO/SEMYG. Rarefaction analysis of our samples revealed a curve approaching saturation (Fig. S1), demonstrating that the vast majority of *A. cerana* and *A. mellifera* gut microbial genes were present in our gene catalogues.

Consistent with previous studies [7,15,21], the overall community profile showed that *A. cerana* and *A. mellifera* guts were both dominated by six genera in our samples, namely, *Lactobacillus*, *Bifidobacterium*, *Snodgrassella*, *Bartonella*, *Gilliamella,* and *Frischella*. The guts of the *A. cerana* workers also harboured a significant amount of the genus *Apibacter*, whereas the *A. mellifera* gut was colonised by the genus *Commensalibacter* (Fig. 2A).

Taking advantage of the high resolution of the metagenome, we analysed and compared gut bacteria compositions at the species level (Fig. S2). Of the six core bacterial genera, *Bifidobacterium asteroides*, *Snodgrassella alvi*, *Bartonella apis*, *Gilliamella apicola,* and *Frischella perrara* were the dominant species identified in both *A. cerana* and *A. mellifera* (Fig. S3), and we referred to these five bacterial species as the core bacterial species of *A. cerana* and *A. mellifera* throughout our study. Contrasting this, we found that the genus *Lactobacillus* was composed of multiple species, which varied significantly in the gut biota of the two honey bee species (Fig. 2B). PCA and PCoA of the bacterial communities at the species level showed that the gut communities of sympatric *A. cerana* and *A. mellifera* were distinguishable at the species level (Fig. 2C, S4, Adonis test: P = 0.005). Furthermore, samples within each honey bee species were clustered together based on their geographic locations (Fig. 2C, Mantel test: P = 0.02). Interestingly, when we reanalysed the metagenome data of *A. mellifera* from the USA, which was published by Engel et al. [4], this sample was grouped within *A. mellifera* samples from Guangzhou (Fig. 2C), suggesting that the geographical separation of the samples might not be solely determined by the geographical distance between the sampling locations.

Strain-level diversity of Apis cerana and A. mellifera core bacteria species

The intra-specific diversity of the *A. cerana* and *A. mellifera* core bacterial species was demonstrated through the single nucleotide polymorphisms present in the *B. apis*, *B. asteroides*, *F. perrara*, *G. apicola*, and *S. alvi*, of *A. cerana* and *A. mellifera* species living in the same habitat in China.

PCA and PCoA analyses based on the SNP patterns revealed that different hosts carry diverse strains. Specifically, *B. asteroides*, *S. alvi*, *G. apicola*, and *F. perrara* strains in *A. cerana* and *A. mellifera* colonies were distinctly host-specific, and *B. apis* strains were also largely host-specific (Fig. 3A, S4, Adonis test: all P = 0.001). However, the geographical region failed to correlate with any notable difference between these five core bacterial species (Fig. 3A, Mantel test: all P > 0.05). We also constructed a phylogenetic tree using the dominant alleles as representative sequences for each sample
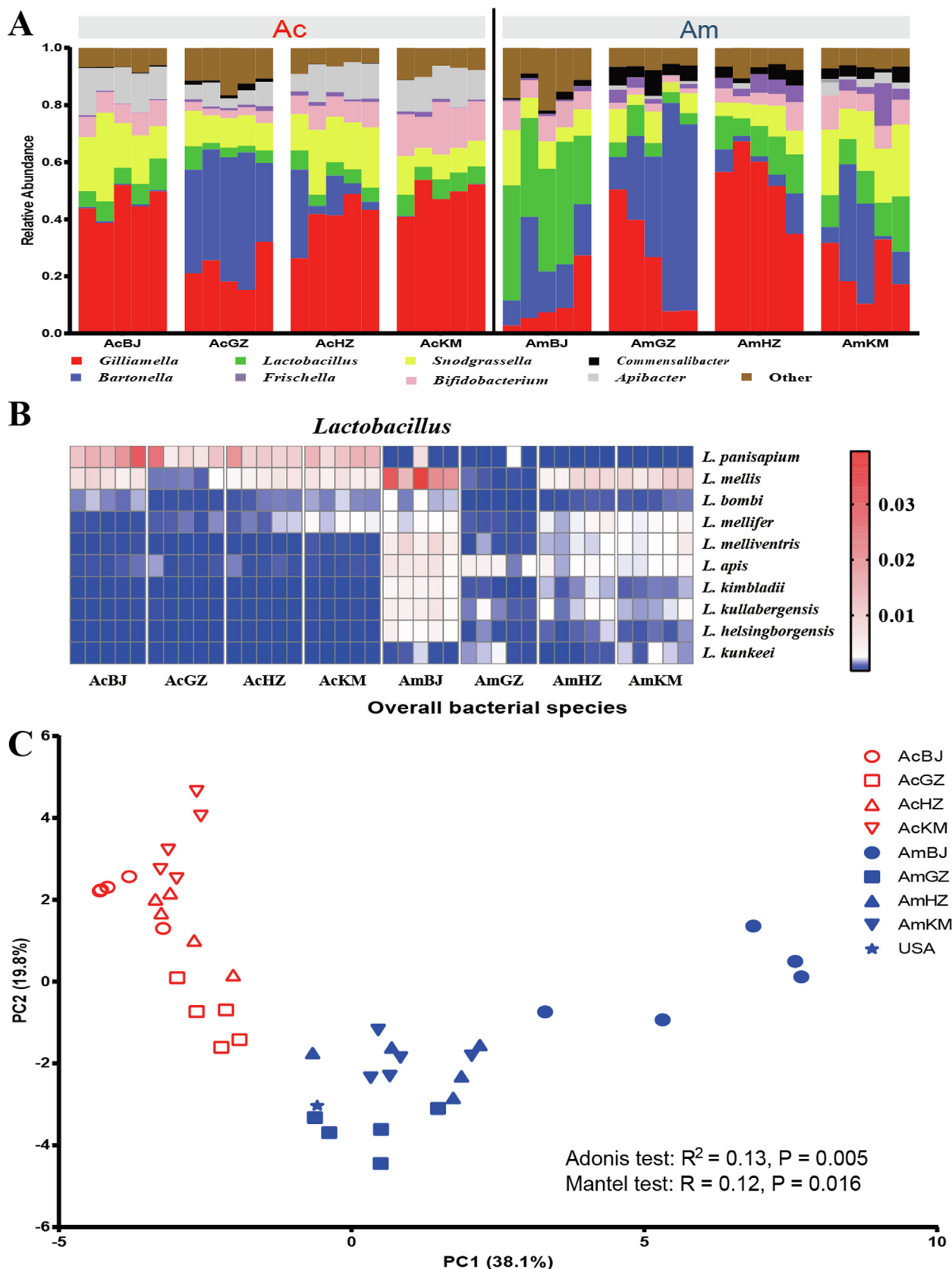
Fig. 2. Gut community of *Apis cerana* and *A. mellifera* workers. (A) Composition at genus level. (B) Heatmap of the composition of *Lactobacillus* at species level. The colours represent the relative abundance of each bacterial species. (C) Principal Component Analysis of overall bacteria at species level. Symbols are coloured according to the host species and shaped according to the region of sampling. AcBJ: *A. cerana* workers sampled in Beijing; AcGZ: *A. cerana* workers sampled in Guangzhou; AcHZ: *A. cerana* workers sampled in Hangzhou; AcKM: *A. cerana* workers sampled in Kunming; AmBJ: *A. mellifera* workers sampled in Beijing; AmGZ: *A. mellifera* workers sampled in Guangzhou; AmHZ: *A. mellifera* workers sampled in Hangzhou; AmKM: *A. mellifera* workers sampled in Kunming; USA: metagenome sample of *A. mellifera* from United States (50).

and found that the dominant strains of each *A. cerana* and *A. mellifera* samples formed separate branches (Fig. 3B). The only exception was *B. apis*; the predominant strains of four *A. mellifera* samples (from three different sampling locations) and *A. cerana* samples were twined together (Fig. 3B). We then compared their SNP similarity and stratified them into intra- and inter-host comparisons (Fig. 3C). We found that the intra-host comparisons possessed more similar strains within *A. cerana* as compared to *A. mellifera*, which demonstrated that the diversity of the gut microbiota was higher in *A. mellifera* at the strain level, but with the exception of *G. apicola*. These findings are consistent with those of previous studies [15,21].

Functional variation in the distinctive strains of Apis cerana and A. mellifera

The aforementioned strain-level distinction between *A. cerana* and *A. mellifera* was also reflected in its accessory genome variation. First, we compared the gene content similarity, which is the percentage of shared genes between two genomes, of these five bacterial species within and between these two honey bee species.

We noticed that the inter-species gene content similarities were significantly smaller than intra-species comparisons, revealing a significant differentiation between the accessory genomes of the *A. cerana* and *A. mellifera* strains.

We hypothesised that, at the strain-level, gene content differences would bring about functional differences. In accordance with a previous study [21], the KEGG profiles of these host-specific strains, based on the functional assignments and relative gene abundances, displayed a difference in the functions of these host-specific strains. The PCA and PCoA assessments of the KEGG orthologs (KOs) revealed a clear separation between the core bacterial species of *A. cerana* and *A. mellifera* (Fig. 4B, S5, Adonis test: all P = 0.001). When we further grouped the genes into KEGG pathways, most of them revealed a significant difference between the strains hosted by *A. cerana* and *A. mellifera* (Fig. S6). We also analysed the genes related to carbohydrate digestion using the CAZy database. In line with KOs, the carbohydrate-active enzyme (CAZyme) profiles of *B. asteroides*, *S. alvi*, *G. apicola*, and *F. perrara* also diverged between the host groups (Fig. 4C, S5, Adonis test: all P = 0.001). Detailed lists of KOs and CAZymes are provided in Supplementary Dataset 2.
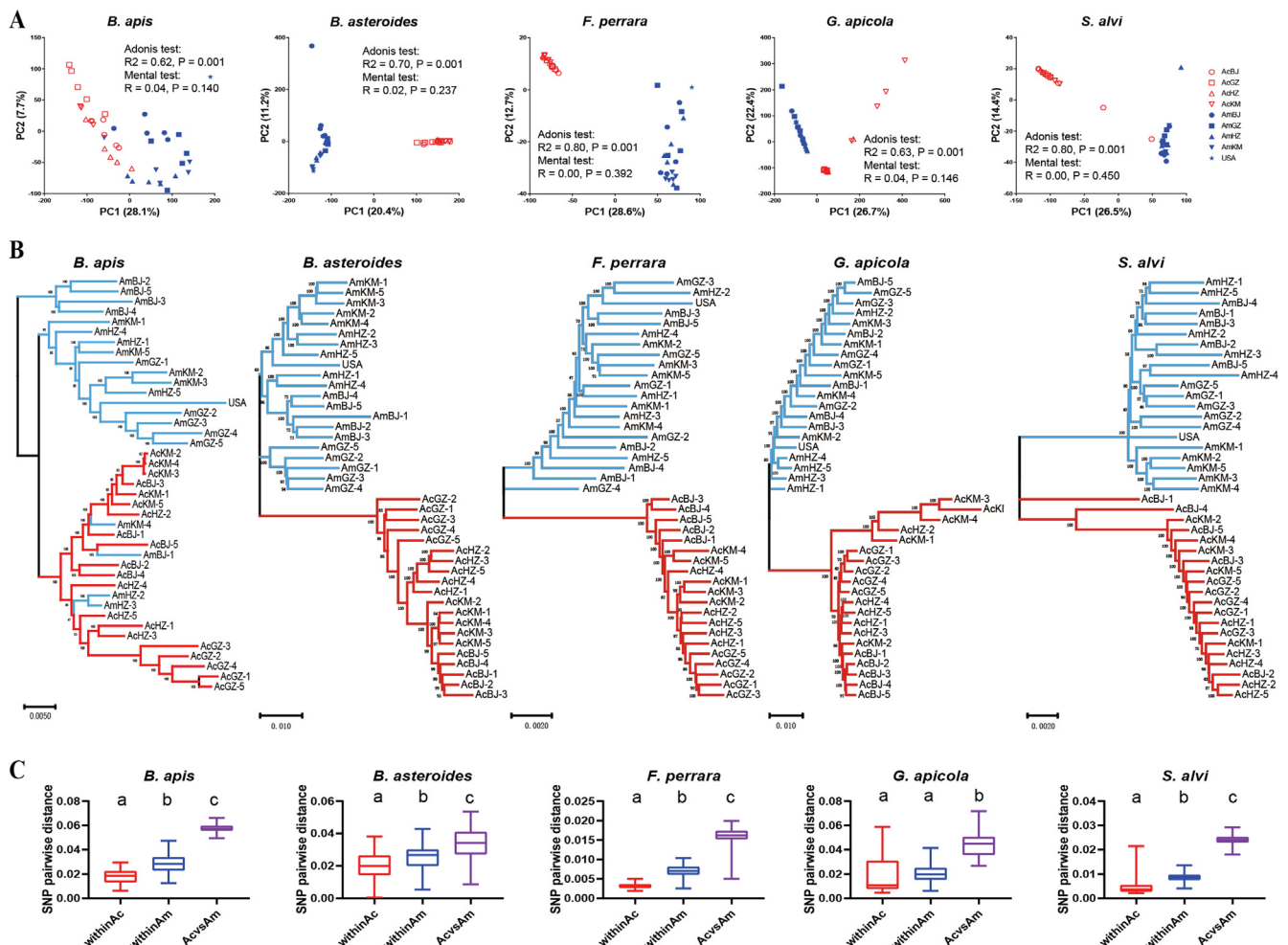
Fig. 3. Strain-level diversity of *Apis cerana* and *A. mellifera* core bacterial species. (A) Principal Component Analysis of *Bartonella apis*, *Bifidobacterium asteroides*, *Frischella perrara*, *Gilliamella apicola*, and *Snodgrassella alvi* strains. (B) Phylogenetic tree of *B. apis*, *B. asteroides*, *F. perrara*, *G. apicola*, and *S. alvi* strains. The geographical origin of the strains is indicated to the right of the tree. (C) Average SNP pairwise distance per genome for *B. apis*, *B. asteroides*, *F. perrara*, *G. apicola*, and *S. alvi* strains. For each boxplot, the center line displays the median and whiskers span minimum to maximum. Different letters represent significant differences (Kruskal-Wallis test, P < 0.001). Within Ac: intra-host comparison within *A. cerana*; within Am: intra-host comparison within *A. mellifera*; Ac VS Am: inter-host comparison between *A. cerana* and *A. mellifera*. AcBJ: *A. cerana* workers sampled in Beijing; AcGZ: *A. cerana* workers sampled in Guangzhou; AcHZ: *A. cerana* workers sampled in Hangzhou; AcKM: *A. cerana* workers sampled in Kunming; AmBJ: *A. mellifera* workers sampled in Beijing; AmGZ: *A. mellifera* workers sampled in Guangzhou; AmHZ: *A. mellifera* workers sampled in Hangzhou; AmKM: *A. mellifera* workers sampled in Kunming; USA: metagenome sample of *A. mellifera* from the United States (50). Throughout this figure, the colour red has been used for *A. cerana* and blue for *A. mellifera*.
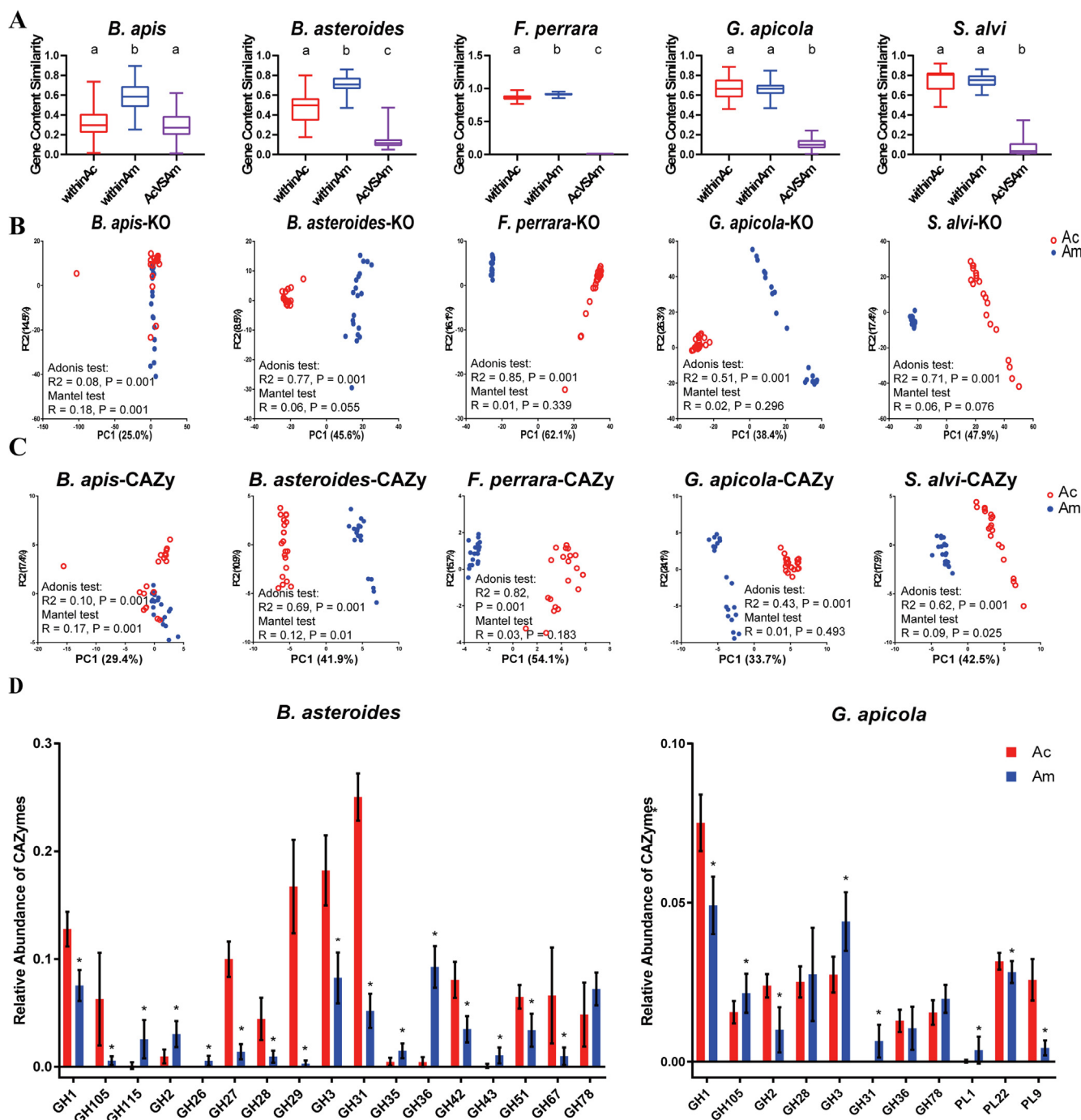
Fig. 4. Functional gene profile of *Apis cerana* and *A. mellifera* core bacterial species. (A) Gene content similarities of *Bartonella apis*, *Bifidobacterium asteroides*, *Frischella perrara*, *Gilliamela Apicola*, and *Snodgrassella alvi*. For each boxplot, the center line displays the median and whiskers span minimum to maximum. Different letters represent significant differences (Kruskal-Wallis test, P < 0.001). Within Ac: intra-host comparison within *A. cerana*; within Am: intra-host comparison within *A. mellifera*; Ac VS Am: inter-host comparison between *A. cerana* and *A. mellifera*. (B and C) Principal Component Analysis of the functional profiles of *B. apis*, *B. asteroides*, *F. perrara*, *G. apicola*, and *S. alvi* based on the KEGG orthologies (KOs) and carbohydrate-active enzymes (CAZymes) subfamilies. Symbols are coloured according to the host species. (D) Relative abundance of CAZymes subfamilies targeting the plant cell wall in *B. asteroides* and *G. apicola*. Data are shown as mean ± SD "*" represents significant differences between *A. cerana* and *A. mellifera* (Kruskal-Wallis test, P$_{maximum}$ = 0.003). Ac: *A. cerana*; Am: *A. mellifera*. Throughout this figure, the colour red has been used for *A. cerana* and blue for *A. mellifera*.

Plant polysaccharides are abundant in the honey bee diet, especially cellulose, hemicellulose, and pectin [55]. Of the CAZymes targeting the plant cell wall [4,56], 24 CAZyme subfamilies have been identified in honey bee core bacterial species. In accordance with a previous study [57], most of the related enzymes were identified in *B. asteroides* and *G. apicola* (Table S2), and their relative abundances significantly differed between *A. cerana* and *A. mellifera* (Fig. 4D).

Gene functions enriched in the metagenome of honey bees

The identification of clear compositional and functional differences between *A. cerana* and *A. mellifera* core bacteria at the strain level led us to speculate whether the overall gut metagenome of these two honey bee species also differed in function. Interestingly, analyses of the functional profile of the overall gut metagenome of

A. cerana and *A. mellifera* revealed that the percentage of significantly different KOs was notably higher among most individual core bacteria than in the overall microbial ecosystem, except for *B. apis* (Fig. 5A). While the KO profiles of *A. cerana* and *A. mellifera* were still significantly different (Fig. 5B, S7, Adonis test: all P = 0.001), we noticed that the R^2 value of the overall gut microbiomes in *A. cerana* and *A. mellifera* was much smaller than that of most of the core bacterial species ($R^2_{overall\ gut\ community}$ = 0.18, $R^2_{B.\ apis}$ = 0.08, $R^2_{B.\ asteroides}$ = 0.77, $R^2_{F.\ perrara}$ = 0.85, $R^2_{G.\ apicola}$ = 0.51, and $R^2_{S.\ alvi}$ = 0.71).

The majority of KOs of both the gut microbiomes of *A. cerana* and *A. mellifera* was enriched in transportation and metabolic pathways including "ABC transporters", "phosphotransferase system (PTS)", "purine metabolism", "pyrimidine metabolism", "amino sugar and nucleotide sugar metabolism" and "fructose and mannose metabolism", among other pathways (Fig. 5C). For these pathways, while their overall abundances of pathways were similar between *A. cerana* and *A. mellifera*, the contributions from individual gut bacteria were significantly different (Fig. 5C). In addition, we found that *Apibacter*, *Commensalibacter*, and other non-core bacteria also played an important role in the functional profile of the honey bee microbial ecosystem, as they contributed significantly to the KEGG pathways (Fig. 5C). The functional profile using the CAZyme annotation showed a similar situation, as the overall CAZyme profile was indistinguishable between the hosts (Fig. S8).

We then attempted to reconstruct the shift from a functional distinction at the species level to functional convergence at the microbiota level. First, we summed all the significantly different KOs in each core bacteria and analysed their changes in the overall microbiome, and nearly half of these KOs (1354/2989) no longer had a significant influence at the microbiota level (Fig. 5D). We then binned the species based on the profiles of these KOs and assessed their contributions to diversity. As shown in Fig. 5E, 5F, these KOs were contributed by varied combinations of multiple community members in *A. cerana* and *A. mellifera*.

Metabolite profiles of workers harbouring diverse gut microbiota

It is challenging to predict the metabolic output from only shotgun metagenomic sequencing data. Therefore, we characterised the robust metabolic profiles of workers harbouring different gut microbiota. To eliminate the possible influences and disturbances of the metabolites from host species, we attempted to colonise germ-free *A. cerana* and *A. mellifera* workers with complete *A. cerana* and *A. mellifera* gut microbiota (Fig. 1B), with the aim of obtaining honey bees that were harbouring gut bacteria from another honey bee species.

Initially, we analysed the absolute and relative bacterial compositions and found no significant differences between Acc and Amc or between Amm and Acm (Fig. 6A, 6B, S9). In addition, alpha diversity analysis revealed that host species had no significant influence on microbiome diversity based on Chao 1 and Shannon indices (Fig. S10). These results suggested that *A. cerana* and *A. mel-*

*lifera* gut bacteria can successfully colonise both *A. cerana* and *A. mellifera* workers.

We then analysed homogenates of gut samples of Acc, Acm, Amm, and Amc through untargeted metabolomics (complete metabolome profiles are provided in Supplementary Dataset 6). To uncover the metabolomic differences brought by the *A. cerana* and *A. mellifera* gut microbiota, we mainly focused on two comparisons: Acc vs. Acm and Amm vs. Amc. In accordance with our metagenomic analyses, although some differences existed, the metabolomes of the different samples were mainly clustered according to their host species rather than their gut microbiota. PCA and hierarchical clustering revealed that the samples were mainly separated according to their host species rather than the gut microbiota they harboured (Fig. 6C, S11). In addition, only 48 and 67 metabolites of the 1,212 identified metabolites were significantly different in Acc vs. Acm and Amm vs. Amc, respectively (Fig. 6D, Tables S3, S4).

Discussion

Bacterial evolution is shaped not only by mutations but also by horizontal gene transfer [58]. *A. cerana* and *A. mellifera* were geographically isolated for millions of years [59], which limited possible interaction between their gut bacteria and enabled the development of specialised gut microbiota that coevolved with their hosts. Based on targeted amplicon sequencing, host-specific strains of *Lactobacillus* Firm-5 and *S. alvi* were identified in the guts of *A. mellifera* and *A. cerana*. Ellegaard et al. [21] found that the core phylotypes colonising *A. mellifera* and *A. cerana* are distinct at the sequence-discrete population level. Consistent with their findings, our results also showed that the core bacterial species of *A. mellifera* and *A. cerana* in China were composed of distinct strains.

Previous genomic sequencing of several isolated strains of honey bee gut microbiota has demonstrated an extensive sequence divergence and gene content variation [17,18]. In addition, metagenomic comparison using *A. cerana* and *A. mellifera* in Japan also revealed that *A. mellifera* displayed a more diverse gene content in the microbiota than *A. cerana* [21]. Thus, it was expected that these distinctive strains of *A. cerana* and *A. mellifera* in the present study possessed different functional gene profiles. Interestingly, we found that the functional profiles of the overall gut microbiomes of *A. cerana* and *A. mellifera* were rather similar, which was further supported by our metabolic analysis. The animal gut is populated by complex assemblages of microorganisms, and functional redundancy (see Glossary) is a common aspect of many microbial systems [60-62]. This functional redundancy of the gut microbiome has been mentioned by Vatanen et al. [63], in which most gene ontology terms displayed high within-sample functional diversity in the human gut microbiome. Comparisons between the metagenomes of *A. mellifera* workers revealed age-related changes in the distribution of the core gene families, despite the variability in strain composition among honey bee individuals of the same age, which indicated functional redun-

Fig. 5. Functional gene profiles of *Apis cerana* and *A. mellifera* gut microbiota. (A) Numbers of significantly different (Sig) and not significantly different (None-Sig) KEGG orthologies (KOs) of core bacteria and overall microbiota between *A. cerana* and *A. mellifera*, symbols and line were the percentages of significantly differing KOs. (B) Principal Component Analysis of the functional profiles of overall *A. cerana* and *A. mellifera* gut microbiota based on the KOs. Symbols are coloured according to the host species. (C) Comparison of KEGG functional profiles (15 pathways with the highest abundance) of gut microbiota of *A. cerana* and *A. mellifera*. (D) The changes of significance of KOs that significantly differed at core bacteria level, Sig represents KOs that significantly differed at both core bacteria and microbiota level, None-Sig represents KOs that significantly differed at core bacteria level but not at the microbiota level. (E) Heatmap of the distribution of species contributing to the KOs that significantly differed at core bacteria level but not at the microbiota level. The contribution of a given bacteria to KO is calculated by dividing its relative abundance by the total relative abundance of this KO. 'Other' here includes *Apibacter* and *Commensalibacter*. (F) The distribution of species contributing to K00104 (glycolate oxidase) and K00101 (L-lactate dehydrogenase (cytochrome)). AcBJ: *A. cerana* workers sampled in Beijing; AcGZ: *A. cerana* workers sampled in Guangzhou; AcHZ: *A. cerana* workers sampled in Hangzhou; AcKM: *A. cerana* workers sampled in Kunming; AmBJ: *A. mellifera* workers sampled in Beijing; AmGZ: *A. mellifera* workers sampled in Guangzhou; AmHZ: *A. mellifera* workers sampled in Hangzhou; AmKM: *A. mellifera* workers sampled in Kunming. Throughout this figure, colour red has been used for *A. cerana* and blue for *A. mellifera*.
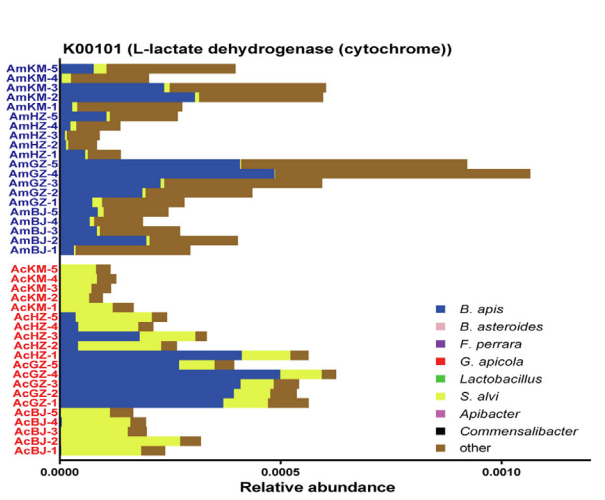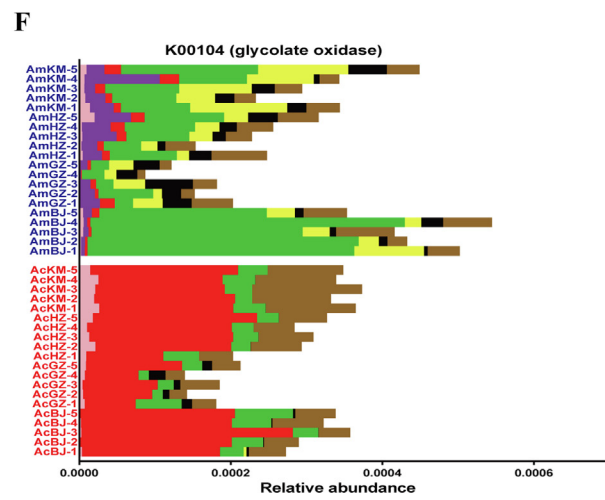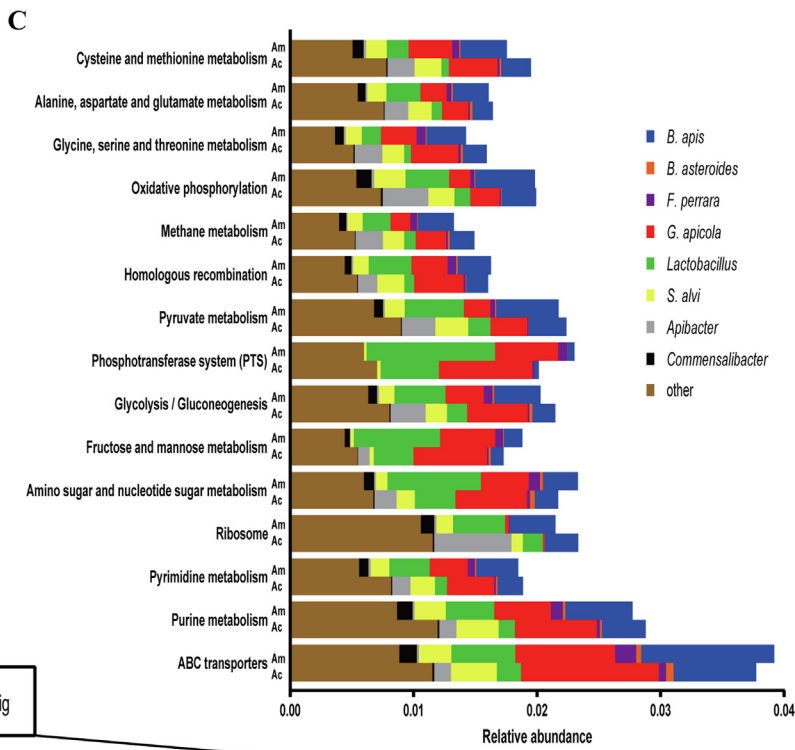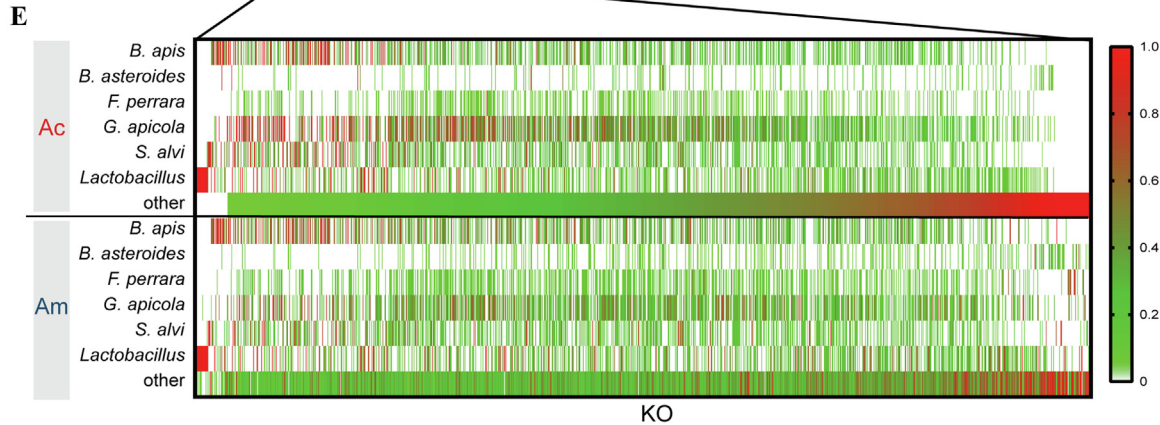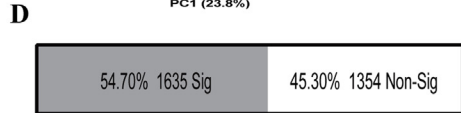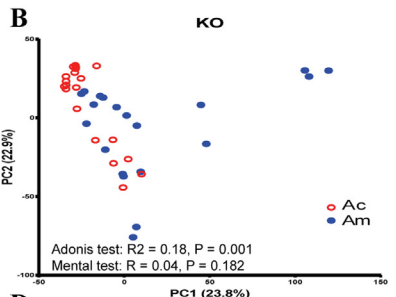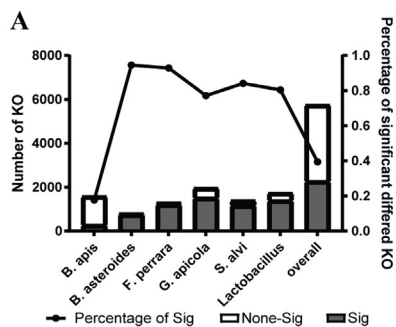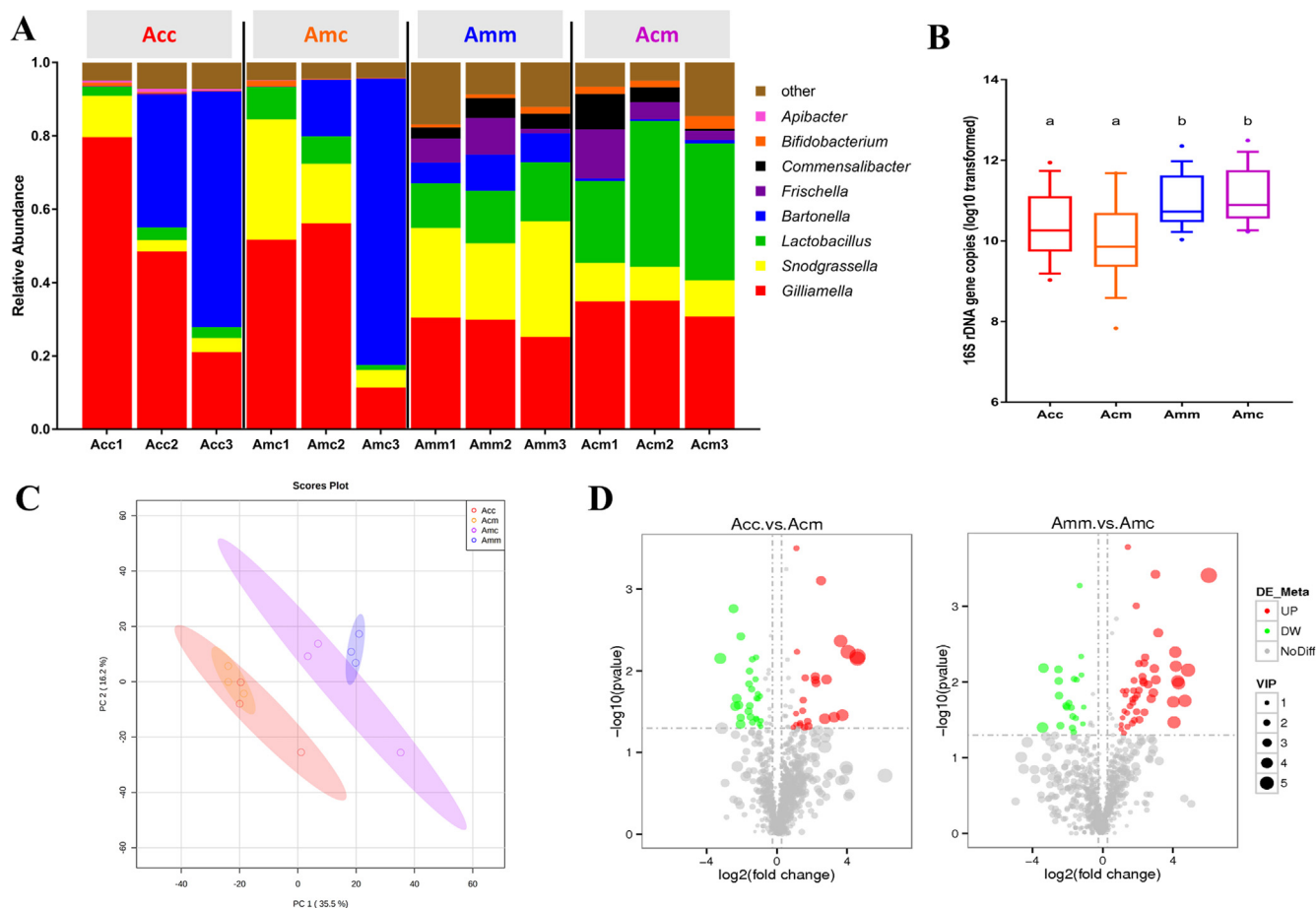
Fig. 6. Gut community and metabolite profiles of workers harbouring different gut microbiota. (A) Gut community composition of Acc, Acm, Amm, and Amc at genus level. (B) Total bacterial loads in the gut of Acc, Acm, Amm, and Amc (n = 15 individual workers). 16 s rRNA copy number were assessed by quantitative PCR (qPCR) with universal bacterial 16S rRNA primers. Lines represent mean and SD. Different letters represent significant differences (Kruskal-Wallis test, $P_{maximum}$ = 0.02). (C) Principal Component Analysis of the metabolomes of Acc, Acm, Amm, and Amc with 95% confidence regions. (D) Metabolite changes between workers harbouring different bacteria. Fold change (FC) of A vs. B is calculated as A/B, up represents FC > 2 and down represents FC < 2. Acc: *A. cerana* workers with *A. cerana* gut microbiota; Amc: *A. mellifera* workers with *A. cerana* gut microbiota; Amm: *A. mellifera* workers with *A. mellifera* gut microbiota; Acm: *A. cerana* workers with *A. mellifera* gut microbiota.

dancy across strains [7]. In this study, we found that functional redundancy existed across the gut bacteria species in both *A. cerana* and *A. mellifera*. Most functions of the honey bee gut microbiota were fulfilled by not only one but multiple bacterial species, and the contributions of these bacteria to KOs differed between *A. cerana* and *A. mellifera*. For example, we found that K00104 (glycolate oxidase) was mainly encoded by *G. apicola* in *A. cerana*, whereas *Lactobacillus* spp. were the main contributor in *A. mellifera*. Moreover, these non-core bacteria also contribute significantly to the functional redundancy of the honey bee gut community. Another example is of K00101 (L-lactate dehydrogenase (cytochrome)), which is mainly encoded by *S. alvi* and *B. apis* in *A. cerana*, but by *B. apis* and non-core bacteria in *A. mellifera*. Eventually, functional redundancy, which can confer resilience and stabilise ecosystem functionality, contributes to the generation of similar functional profiles between these two honey bee species. In contrast to environmental ecosystems, the evolution of the gut microbiome is not only driven from the bottom up by interactions between microbes, but the host is also under strong natural selection to shape the microbiota from the top down and foster a community that is beneficial to it [1]. We speculate that the similarity between the *A. cerana* and *A. mellifera* microbiomes is largely due to the similarities between these two species. Although the Eastern and Western honey bees have been geographically isolated and

have developed several different behavioural traits [64], they still share some ecological and physiological aspects in common, particularly those involving certain important factors that shape gut community function, for example, a similar diet [26]. Therefore, these two honey bee species have provided a stable environment and similar selective pressure on their gut microbiota, which has eventually shaped a similar metabolic potential.

The colonisation ability of exogenous microbes in a host environment has been extensively studied, revealing that the host genotype exerts different selective pressures on exogenous colonisers [15,65-67]. Interestingly, our cross-species gut bacteria transplantation assay results suggested that there is no host specificity in core gut bacteria between *A. cerana* and *A. mellifera*, since *A. cerana* core bacteria could successfully colonise *A. mellifera* and vice versa, and following colonisation, the gut microbial community structure was not dramatically changed by the host. In fact, we discovered that certain strains of *B. apis* from *A. mellifera* were more similar to *A. cerana* strains in our field honey bee samples, indicating a certain degree of putative host switches in the natural environment.

The above findings raise an interesting question: if there is no host specificity, how do *A. cerana* and *A. mellifera* maintain a distinctive gut community while their colonies are located in the same habitat? This is of particular interest when considering the

fact that *A. cerana* and *A. mellifera* honey bees for this study were sampled from the same apiaries, in which they shared overlapping ranges of activity. One possible explanation is that the sociality of corbiculate bees provides a semi-closed and stable system for gut bacterial transmission between generations [15]. Honey bee workers acquire their gut microbiota within hours after emergence through the faecal-oral route and fully establish a stable gut community within 5–6 days before they start to forage outside the hive [22]. Moreover, it has been suggested that priority effects in the early life assembly of the gut microbiota contribute to individualised gut community profiles of *A. mellifera* workers [7]. Co-inoculation trials using *S. alvi* strains revealed that strains of *A. cerana* were able to simultaneously colonise *A. mellifera* workers alongside a native strain, albeit with less efficiency [15], suggesting that microbe competition may be essential to the assemblage of honey bee gut microbiota. In addition, differences in gene function profiles between the strains revealed in our study indicated differences in adaptability. Thus, we hypothesise that the highly conserved social transmission route between generations within the honey bee colony, along with the priority effects and the possible competition between native and foreign strains, have ensured the existence of a unique and stable gut community of *A. cerana* and *A. mellifera*. However, our knowledge concerning the transmission of gut bacteria in honey bees and the competition between different strains is still quite limited. Thus, future studies are required to determine the mechanisms involved in obtaining, assembling, and stabilising diverse honey bee microbial communities.

Conclusion

In conclusion, through our systematic metagenomic comparison between *A. cerana* and *A. mellifera* gut microbiomes, we found that the distinctive *A. cerana* and *A. mellifera* gut bacterial strains possess different functions, except in *B. apis*. However, our analysis of the overall metagenomes and the metabolomes emphasises the similarity in functional capabilities of the *A. cerana* and *A. mellifera* gut microbiota, due to the functional redundancy of honey bee gut bacteria. These findings not only reveal the functional importance of strain-level diversity in host-associated bacterial communities, but also provide fundamental insights into the functional contribution of intra-species diversity to the overall gut community.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Glossary

Alpha diversity

The diversity within a particular area or ecosystem. There are several metrics for measurement of alpha diversity, such as Observed number of species, Chao1, ACE, Shannon and Simpson.

Functional profile

The profile of the function of a certain bacteria or microbiome. In this study, the functional profile was inferred by gene annotation of the Carbohydrate-Active EnZymes (CAZy) database or Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Functional redundancy

The ecological phenomena that multiple species representing a variety of taxonomic groups can share similar, if not identical, roles in ecosystem functionality.

Phylotype

A biological type that classifies an organism by its phylogenetic relationship to other organisms. In prokaryotic microbiology, phylotypes, often referred to as Operational Taxonomic Units (OTUs), are a proxy for species.

Scaftigs

Continuous sequences within scaffolds.

References

[1] Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. Nature 2017;548(7665):43–51.

[2] Engel P, Moran NA. The gut microbiota of insects-diversity in structure and function. FEMS Microbiol Rev 2013;37(5):699–735.

[3] Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet 2012;13(4):260–70.

[4] Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci USA 2012;109(27):11002–7.

[5] Wu Y, Zheng Y, Chen Y, Wang S, Chen Y, Hu F, et al. Honey bee (*Apis mellifera*) gut microbiota promotes host endogenous detoxification capability via regulation of P450 gene expression in the digestive tract. Microbiol Biotechnol 2020;13(4):1201–12.

[6] Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. Nature 2013;493 (7430):45–50.

[7] Ellegaard KM, Engel P. Genomic diversity landscape of the honey bee gut microbiota. Nat Commun 2019;10(1):446.

[8] Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. Cell 2015;160(4):583–94.

[9] Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, et al. Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola*. mBio 2016;7 (6). doi: https://doi.org/10.1128/mBio.01326-16.

[10] Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 2012;13(9):R79. doi: https://doi.org/10.1186/gb-2012-13-9-r79.

[11] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 2012;490(7418):55–60.

[12] Ellegaard KM, Engel P. Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. Front Microbiol 2016;7:1475.

[13] Shan K, Qu H, Zhou K, Wang L, Zhu C, Chen H, et al. Distinct gut microbiota induced by different fat-to-sugar-ratio high-energy diets share similar pro-obesity genetic and metabolite profiles in prediabetic mice. mSystems. 2019;4 (5):e00219-19.

[14] Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA. Honey bees as models for gut microbiota research. Lab Anim 2018;47(11):317–25.

[15] Kwong WK, Medina LA, Koch H, Sing K-W, Soh EJY, Ascher JS, et al. Dynamic microbiome evolution in social bees. Sci. Adv. 2017;3(3):e1600513.

[16] Martinson VG, Moy J, Moran NA. Establishment of characteristic gut bacteria during development of the honeybee worker. Appl Environ Microbiol 2012;78 (8):2830–40.

[17] Engel P, Stepanauskas R, Moran NA. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. PLoS Genet. 2014;10(9): e1004596.

[18] Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vásquez A. Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. BMC Genom. 2015;16(1):284.

[19] Powell E, Ratnayeke N, Moran NA. Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. Mol Ecol 2016;25 (18):4461–71.

[20] Koch H, Abrol DP, Li J, Schmid-Hempel P. Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. Mol Ecol 2013;22(7):2028–44.

[21] 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(13):2520–2531.e7.

[22] Powell JE, Martinson VG, Urban-Mead K, Moran NA. Routes of acquisition of the gut microbiota of the honey bee Apis mellifera. Appl Environ Microbiol 2014;80(23):7378–87.

[23] Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, et al. Cospeciation of gut microbiota with hominids. Science 2016;353 (6297):380–2.

[24] Gallai N, Salles J-M, Settele J, Vaissière BE. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. Ecol Econ 2009;68(3):810–21.

[25] Chantawannakul P, Williams G, Neumann P. Asian beekeeping in the 21st century. Singapore: Springer; 2018.

[26] Oldroyd BP, Wongsiri S. Asian honey bees: biology, conservation, and human interactions. Boston: Harvard University Press; 2009.

[27] Brodschneider R, Gray A, Adjlane N, Ballis A, Brusbardis V, Charrière J-D, et al. Multi-country loss rates of honey bee colonies during winter 2016/2017 from the COLOSS survey. J Apic Res 2018;57(3):452–7.

[28] Brodschneider R, Gray A, van der Zee R, Adjlane N, Brusbardis V, Charrière J-D, et al. Preliminary analysis of loss rates of honey bee colonies during winter 2015/16 from the COLOSS survey. J Apic Res 2016;55(5):375–8.

[29] Theisen-Jones H, Bienefeld K. The Asian honey bee (Apis cerana) is significantly in decline. Bee World 2016;93(4):90–7.

[30] Evans JD, Schwarz RS. Bees brought to their knees: microbes affecting honey bee health. Trends Microbiol 2011;19(12):614–20.

[31] Zheng H, Powell JE, Steele MI, Dietrich C, Moran NA. Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. Proc Natl Acad Sci USA 2017;114(18):4775–80.

[32] Klein A-M, Vaissiere BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, et al. Importance of pollinators in changing landscapes for world crops. P Roy Soc B-Biol Sci 2006;274(1608):303–13.

[33] Raymann K, Shaffer Z, Moran NA. Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. PLoS Biol. 2017;15(3): e2001861.

[34] Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 2010;26(5):680–2.

[35] Gu S, Fang L, Xu X. Using SOAPaligner for short reads alignment. Curr Protoc. Bioinform. 2013;44(1):11.11.1-7.

[36] Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci 2006;14(6):927–30.

[37] Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 2010;26(11):1463–4.

[38] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2017.

[39] Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, et al. PanGP: a tool for quickly analyzing bacterial pan-genome profile. Bioinformatics 2014;30(9):1297-9.

[40] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.

[41] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42(D1):D490–5.

[42] Segers FHID, Kešnerová L, Kosoy M, Engel P. Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. ISME J 2017;11(5):1232–44.

[43] Bottacini F, Milani C, Turroni F, Sánchez B, Foroni E, Duranti S, et al. Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. PLoS One 2012;7(9):e44229.

[44] Engel P, Vizcaino MI, Crawford JM. Gut symbionts from distinct hosts exhibit genotoxic activity via divergent colibactin biosynthesis pathways. Appl Environ Microbiol 2015;81(4):1502–12.

[45] Kwong WK, Engel P, Koch H, Moran NA. Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc Natl Acad Sci US 2014;111 (31):11509–14.

[46] Langdon WB. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. BioData Min 2015;8(1):1.

[47] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25(16):2078–9.

[48] Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.

[49] Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. Bioinformatics 1994;10(2):189–91.

[50] Williams GR, Alaux C, Costa C, Csáki T, Doublet V, Eisenhardt D, et al. Standard methods for maintaining adult Apis mellifera in cages under in vitro laboratory conditions. J Apic Res 2013;52(1):1–36.

[51] Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. Curr. Protoc. Bioinformatics 2019;68 (1):e86.

[52] Jafari M, Ansari-Pour N. Why, when and how to adjust your p values? Cell J 2019;20(4):604–7.

[53] Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 2004;20(2):289–90.

[54] White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. PLoS Comput Biol. 2009;5 (4):e1000352.

[55] Roulston TH, Cane JH. Pollen nutritional content and digestibility for animals. Plant Syst Evol 2000;222:187–209.

[56] Rytioja J, Hildén K, Yuzon J, Hatakka A, de Vries RP, Mäkelä MR. Plant-polysaccharide-degrading enzymes from basidiomycetes. Microbiol Mol Biol Rev 2014;78(4):614–49.

[57] Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, et al. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. Proc Natl Acad Sci USA 2019;116(51):25909–16.

[58] Koskella B, Vos M. Adaptation in natural microbial populations. Annu Rev Ecol Evol Syst 2015;46:503–22.

[59] Arias MC, Sheppard WS. Phylogenetic relationships of honey bees (Hymenoptera: Apinae: Apini) inferred from nuclear and mitochondrial DNA sequence data. Mol Phylogenet Evol 2005;37(1):25–35.

[60] Heintz-Buschart A, Wilmes P. Human gut microbiome: function matters. Trends Microbiol 2018;26(7):563–74.

[61] Moya A, Ferrer M. Functional redundancy-induced stability of gut microbiota subjected to disturbance. Trends Microbiol 2016;24(5):402–13.

[62] Louca S, Polz MF, Mazel F, Albright MB, Huber JA, O'Connor MI, et al. Function and functional redundancy in microbial systems. Nat Ecol Evol 2018;2 (6):936–43.

[63] Vatanen T, Plichta DR, Somani J, Münch PC, Arthur TD, Hall AB, et al. Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. Nat Microbiol 2019;4(3):470.

[64] Park D, Jung JW, Choi B-S, Jayakodi M, Lee J, Lim J, et al. Uncovering the novel characteristics of Asian honey bee, Apis cerana, by whole genome sequencing. BMC Genom 2015;16(1). doi: https://doi.org/10.1186/1471-2164-16-1.

[65] Zhou W, Chow K-H, Fleming E, Oh J. Selective colonization ability of human fecal microbes in different mouse gut environments. ISME J. 2019;13(3):805.

[66] Ellegaard KM, Brochet S, Bonilla-Rosso G, Emery O, Glover N, Hadadi N, et al. Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. Mol Ecol 2019;28(9):2224–37.

[67] Bobay L-M, Raymann K. Population genetics of host-associated microbiomes. Curr Mol Biol Rep 2019;5(3):128–39.