Supplementary information

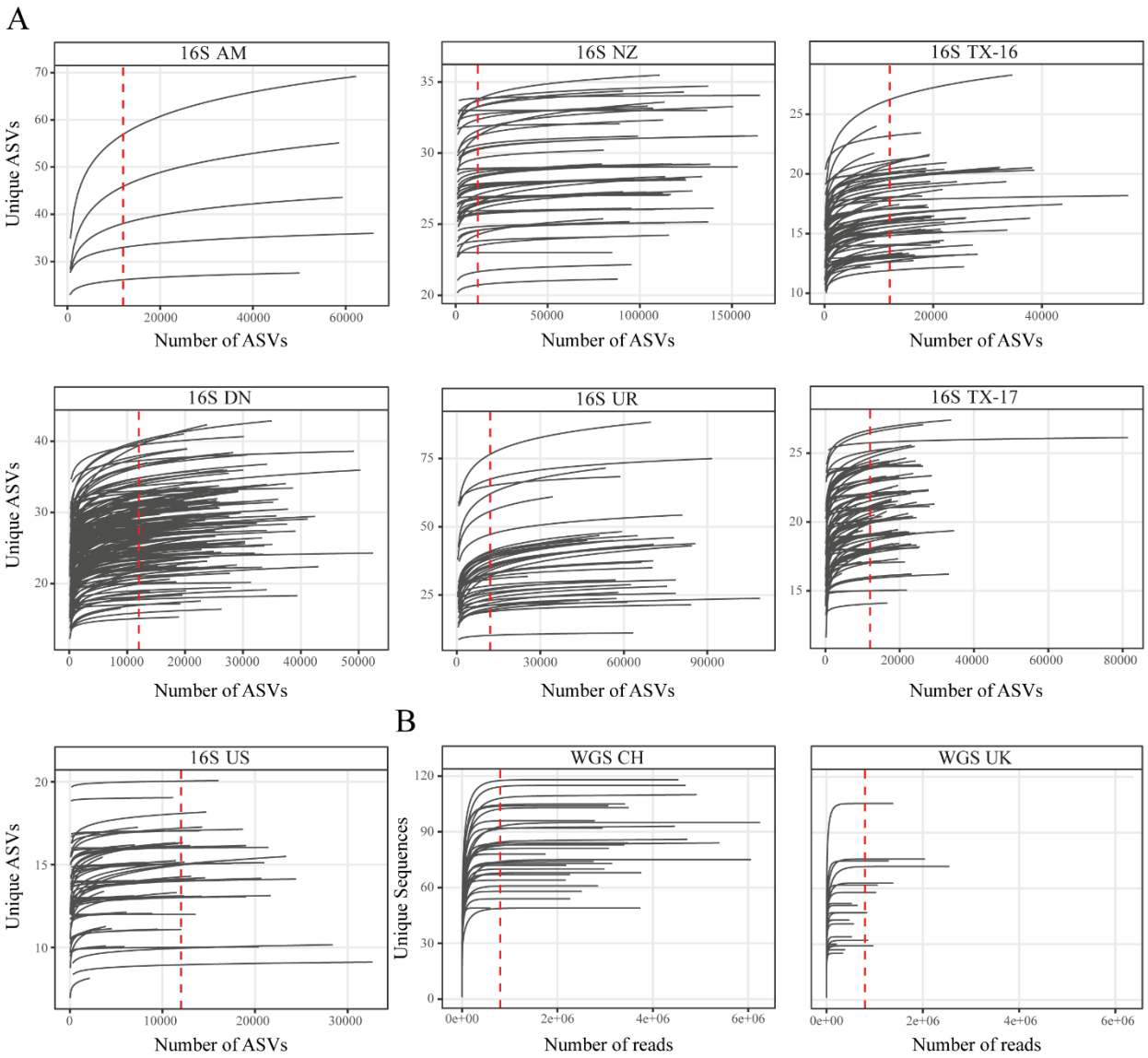

Figure S1. Rarefaction curves illustrating the number of identified ASVs as a function of library size for each study. Dashed lines represent the rarefaction **(A)** threshold of 12 000 reads for 16S rRNA and **(B)** 800 000 for WGS, used to normalize sequencing depth across samples.

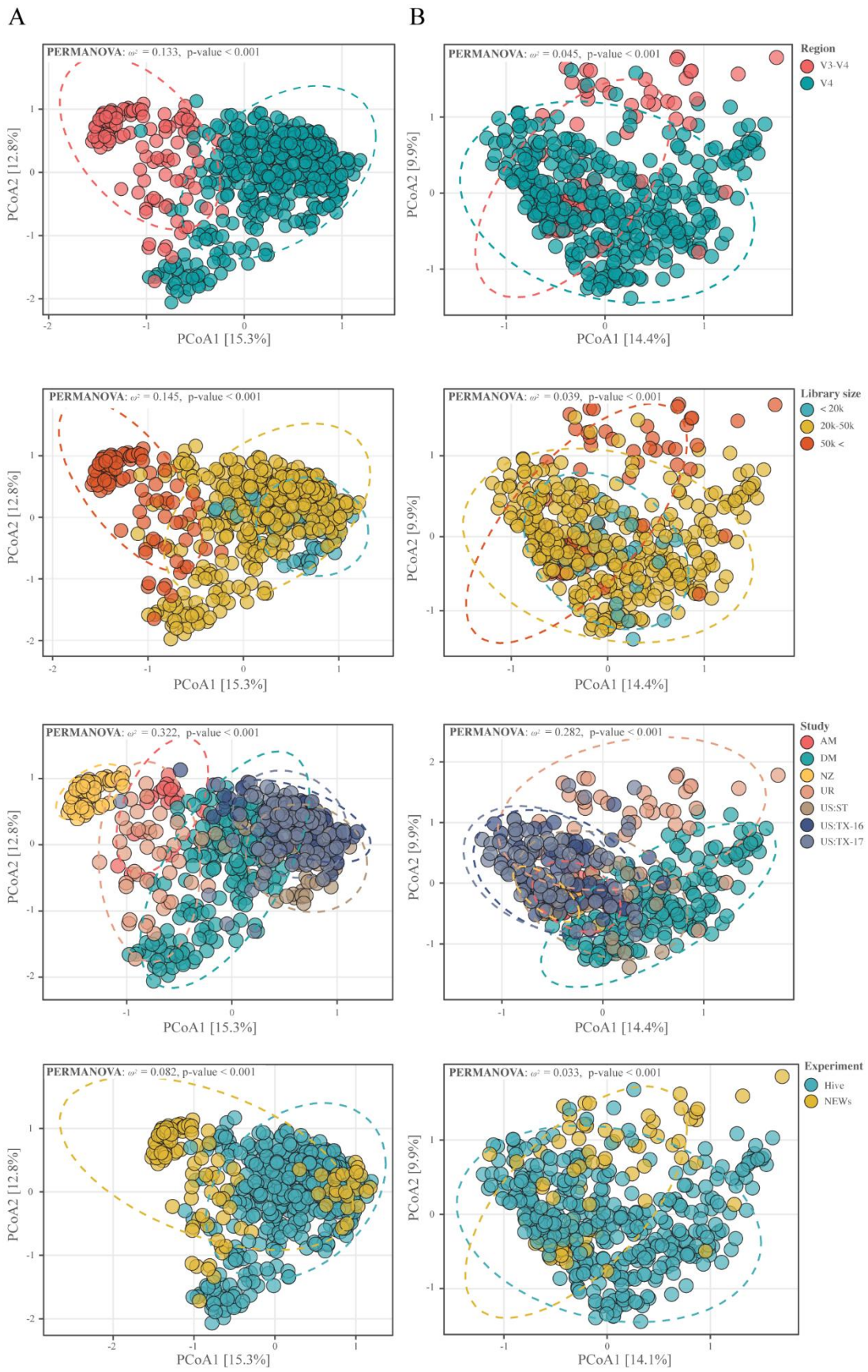
A
B
PERMANOVA: ω² = 0.133, p-value < 0.001
PERMANOVA: ω² = 0.045, p-value < 0.001
Region
V3-V4
V4
PERMANOVA: ω² = 0.145, p-value < 0.001
PERMANOVA: ω² = 0.039, p-value < 0.001
Library size
< 20k
20k-50k
50k <
PERMANOVA: ω² = 0.322, p-value < 0.001
PERMANOVA: ω² = 0.282, p-value < 0.001
Study
AM
DM
NZ
UR
US:ST
US:TX-16
US:TX-17
PERMANOVA: ω² = 0.082, p-value < 0.001
PERMANOVA: ω² = 0.033, p-value < 0.001
Experiment
Hive
NEWs
PCoA1 [15.3%]
PCoA2 [12.8%]
PCoA1 [14.4%]
PCoA2 [9.9%]
PCoA1 [14.1%]

Figure S2. Principal coordinate analysis (PCoA) of Bray-Curtis dissimilarity, colored by hypervariable regions of the 16S rRNA gene **(A)** before and **(B**) after trimming the libraries to a common V4 region and applying library size reduction using rarefaction, colored by the amplicon region, library size, study and experiment type.

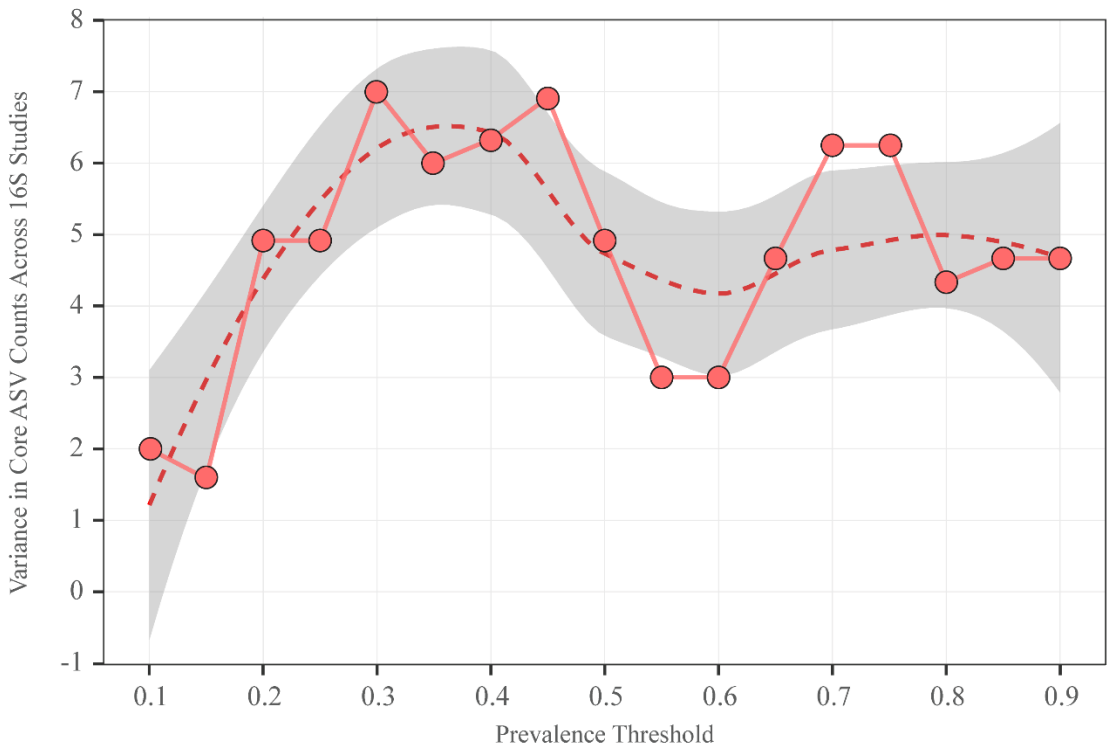

Figure S3. Sensitivity prevalence threshold selection for the core microbiome definition. For each prevalence threshold, we have computed the number of overlapping taxa between dataset pairs. The x axis indicates prevalence thresholds as fractions, and the y axis indicates the variance of overlapping taxa sizes.

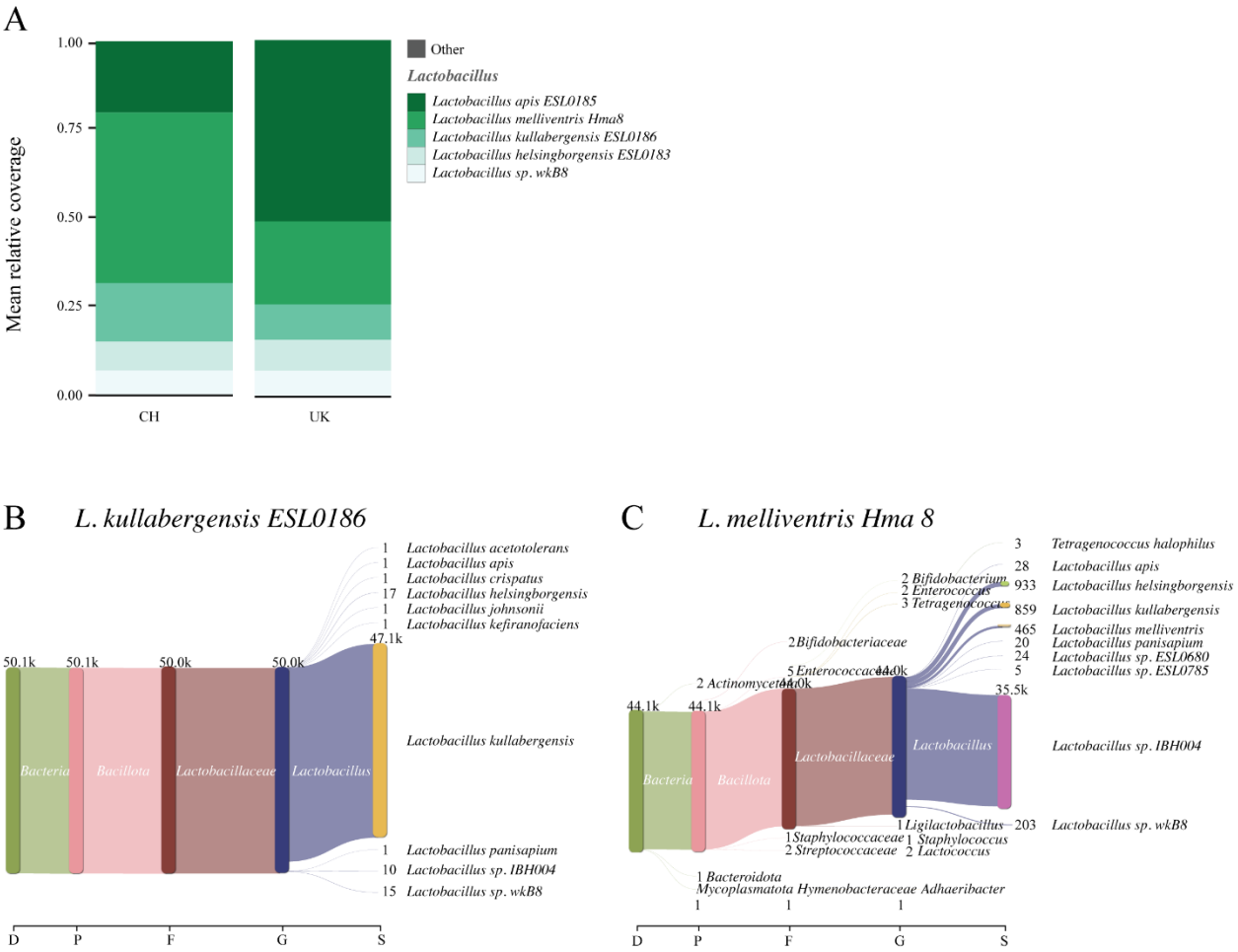

Figure S4. (A) Sankey diagram with the total number of corrected reads as estimated with Kraken 2 for all sites combined at the level of kingdom, phylum, family, genus and species. Mean relative coverage of MAGs within WGS studies for **(B)** *Lactobacillus kullabergensis ESL0186* and **(C)** *Lactobacillus melliventris Hma 8*.

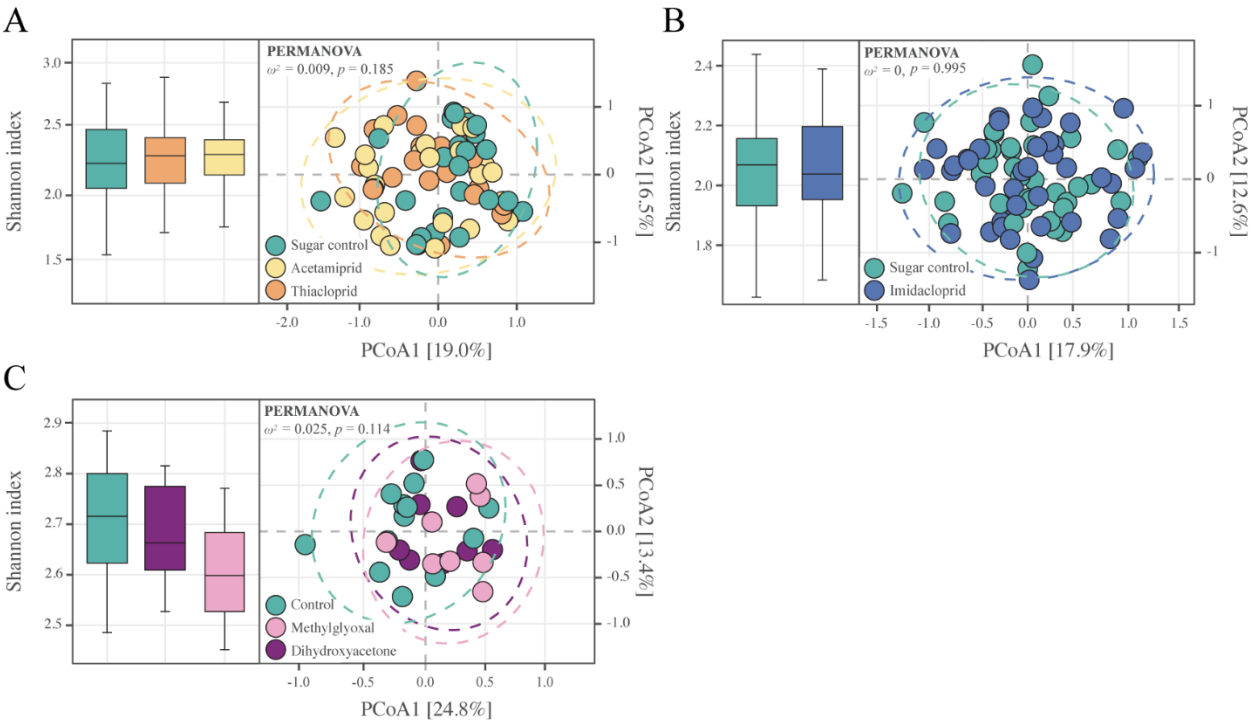

Figure S5. Alpha and beta diversity analyses showed that the treatments did not alter the structure and membership of the gut microbiota based on Shannon index and Bray-curtis dissimilarities **(A)** acetamiprid and thiacloprid, **(B)** imidacloprid and **(C)** methylglyoxal and dihydroxyacetone (p > 0,05 in all cases, **A-C**).

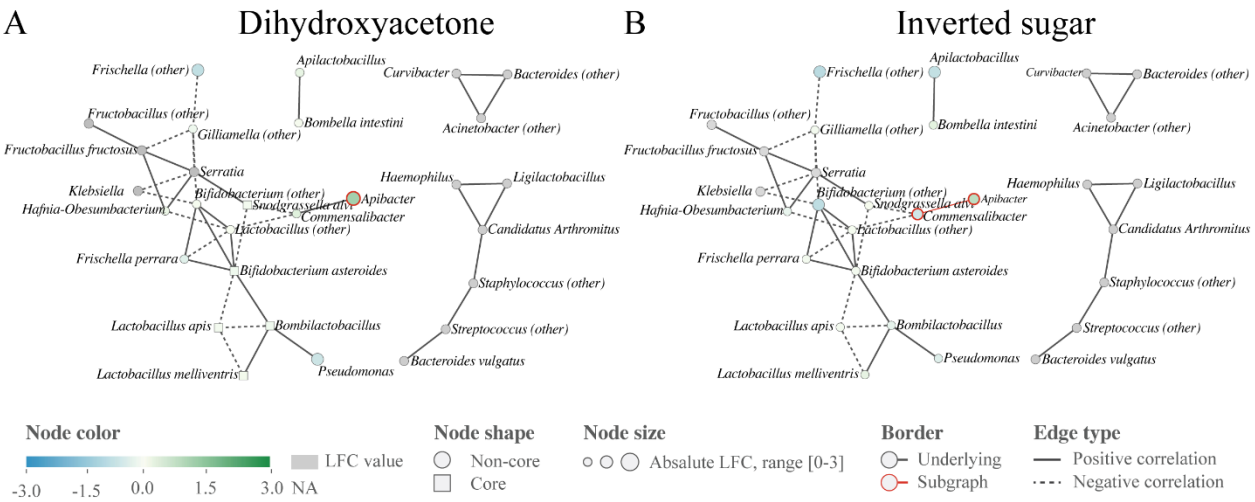

Figure S6. Co-abundance networks of the honeybee gut microbiome under **(A)** Dihydroxyacetone and **(B)** Inverted sugar. For each treatment, the upper panel shows the co-abundance network and the lower panel displays the corresponding functional enrichment results. In the networks, nodes represent taxa and edges denote significant Spearman correlations ($\rho > 0.6$, adjusted $p < 0.2$); solid lines indicate positive correlations, and dashed lines negative correlations. Core taxa are represented by rectangular nodes, and non-core taxa by circles. Node size reflects the absolute log fold change (LFC) between treatment and control groups, with a green-to-blue color scale indicating up- and down-regulation, respectively. Nodes for which LFC values were not calculated by ANCOM-BC2 due to sample size limitations are colored gray. Nodes and edges highlighted with red borders indicate the treatment-specific subnetwork identified using the maximum weight connected subgraph algorithm (see Methods).

Supplementary note 1: *Lactobacillus melliventris* misclassification

The absence of *L. melliventris* in WGS datasets, despite its documented presence in honeybee microbiomes and consistent detection in 16S rRNA datasets, raised questions about potential misclassification in the bioinformatics pipeline. To address this, we conducted a series of targeted analyses to identify the underlying cause. To reject the absence of the L. melliventris genome in the WGS datasets we constructed a custom reference genome database including *L. melliventris* and other representative *Lactobacillus* species from NCBI. Alignment with bowtie2 (v. 2.4.4) confirmed that a lot of reads (48.1% in China and 23.2% in UK) from WGS datasets mapped to the *L. melliventris Hma 8* genome, suggesting that its absence in initial classifications was not due to its complete absence in the data (Fig. S3).

Afterwords, to investigate potential classification errors in Kraken2, with the art_illumina tool (v.2016.06.05) [1] we artificially fragmented the reference genomes of *L. melliventris* representative strain *Hma8* and *L. kullabergensis ESL0186* (another highly abundant *Lactobacillus* strain identified in our sample with Kraken 2), creating paired-end FASTQ files with 5x coverage. Kraken analysis of these files,showed that while *L. kullabergensis ESL0186* reads were classified correctly, *L. melliventris Hma8* reads were predominantly misclassified as *Lactobacillus sp. IBH004*, an unclassified *Lactobacillus* species (Fig. S3). However, we did not find *Lactobacillus sp. IBH004* in the WGS samples either.

Further analysis using the kraken2-inspect tool verified that the taxonomic identifiers for both species were accurate and consistent with NCBI Taxonomy. Reference genome quality checks revealed high completeness and low contamination for both *L. melliventris Hma 8* (97.74% completeness, 0.99% contamination) and *Lactobacillus sp. IBH004* (97.68% completeness, 0.81% contamination), ruling out poor assembly or contamination as potential causes.

We conclude that database annotation inconsistencies can significantly impact taxonomic accuracy, as demonstrated by the misclassification of *L. melliventris* in the WGS dataset. The absence of *Lactobacillus sp. IBH004* in the samples further suggests that these misclassifications were due to limitations in the reference database, rather than the biological presence of this species.

References

1. Huang W, Li L, Myers JR *et al.* ART: a next-generation sequencing read simulator. *Bioinformatics* 2012;28:593–4.

Supplementary note 2: PRISMA flow diagram