

RESEARCH

Open Access



Strain-level profiling with picodroplet microfluidic cultivation reveals host-specific adaptation of honeybee gut symbionts

Yujie Meng¹, Shuang Li², Chong Zhang^{2*} and Hao Zheng^{1*}

Abstract

Background: Symbiotic gut microbes have a rich genomic and metabolic pool and are closely related to hosts' health. Traditional sequencing profiling masks the genomic and phenotypic diversity among strains from the same species. Innovative droplet-based microfluidic cultivation may help to elucidate the inter-strain interactions. A limited number of bacterial phylotypes colonize the honeybee gut, while individual strains possess unique genomic potential and critical capabilities, which provides a particularly good model for strain-level analyses.

Results: Here, we construct a droplet-based microfluidic platform and generated $\sim 6 \times 10^8$ droplets encapsulated with individual bacterial cells from the honeybee gut and cultivate in different media. Shotgun metagenomic analysis reveals significant changes in community structure after droplet-based cultivation, with certain species showing higher strain-level diversity than in gut samples. We obtain metagenome-assembled genomes, and comparative analysis reveal a potential novel cluster from *Bifidobacterium* in the honeybee. Interestingly, *Lactobacillus panisapium* strains obtained via droplet cultivation from *Apis mellifera* contain a unique set of genes encoding L-arabinofuranosidase, which is likely important for the survival of bacteria in competitive environments.

Conclusions: By encapsulating single bacteria cells inside microfluidic droplets, we exclude potential interspecific competition for the enrichment of rare strains by shotgun sequencing at high resolution. The comparative genomic analysis reveals underlying mechanisms for host-specific adaptations, providing intriguing insights into microbe-microbe interactions. The current approach may facilitate the hunting for elusive bacteria and paves the way for large-scale studies of more complex animal microbial communities.

Keywords: *Apis mellifera*, *Apis cerana*, Microfluidic droplet, High-throughput cultivation, Gut microbiota, Strain diversity, Host specificity

Background

The animal gut is inhabited by billions of bacterial cells from a wide range of taxa with a rich genomic and metabolic pool. The gut microbiota can profoundly affect hosts' physiology, metabolism, immunity, and behaviors

[1]. Over the years, investigations based on marker gene amplification and shotgun metagenomic technologies have expanded our understanding of the diversity of the complex gut microbial communities and revealed intriguing associations with hosts' health [2, 3]. However, a great deal of genomic and phenotypic diversity exist among strains of the same species, which causes a massive disconnect between the sequencing results and the actual existence of bacteria strains. Therefore, culture is vital for studies of the intestinal microbiome. By culturing specific bacterial individuals, we can recover the complete

*Correspondence: chongzhang@tsinghua.edu.cn; hao.zheng@cau.edu.cn

¹ College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China

² Department of Chemical Engineering, Institute of Biochemical Engineering, Tsinghua University, Beijing 100084, China



reference genome and accurately identify the taxonomic and functional potential of specific rare strains [4, 5]. However, traditional culture methods are often limited by substrates and growth conditions. Slow-growing bacteria present in low abundance are significantly affected by inter-species competition [6]. Thus, only a few have been effectively characterized [7]. In addition, most cultivation efforts relying on traditional strategies require manual selection of large numbers of colonies; the cost and throughput severely limit the exploration progress of rare microbial taxa [7].

To circumvent or minimize the potential limitations of traditional culture strategies, innovative technologies have broadened the toolkits for microbial isolation and cultivation. Droplet-based microfluidics is a novel technology for manipulating and processing small amounts of droplets carried by corresponding immiscible phases [8]. The effects of overgrown fast-growing populations in community culture can be eliminated by compartmentalizing microbes in droplets of media that are tens to hundreds of micrometers in diameter and separated by immiscible oils and engineered surfactants [9]. Since microfabricated physical pores or channels do not confine droplets, millions of independent culture systems can be created quickly. So far, platforms for high-throughput automated isolation, culture, and sorting of gut microbial members in microfluidic droplets have been developed and used for isolating microbes from seawater and soil communities [10, 11], high-coverage genome sequencing of single cells [12], and targeted screening of microbiome resource searching for probiotics and physiologically active compounds [13].

Animals carry complex communities of symbiotic microbiota that typically encompass strains with highly variable gene content and existence a rich strain-level diversity [14]. Among the human health conditions linked to microbial communities, phenotypes are often associated with only a subset of strains within causal microbial groups [15]. However, important strains are deficient in abundance and often require a high sequencing depth to be detected [16]. Their isolations are often affected by the competition from fast-growing individuals [7]. Therefore, it is promising to apply the innovative microfluidic droplet method to study the strain-level composition and functional diversity of gut microbiota. Combined with 16s rRNA amplicon sequencing, preculture of microfluidic droplets exhibited broad applicability for investigating the dietary carbohydrate metabolism [17] and antibiotic-resistance of intestinal bacteria [18].

Honeybees (*Apis mellifera*) are important pollinators of plants in both natural and agricultural landscapes [19]. Studies have shown that honeybee gut microbiota affects host nutrition, weight gain, and endocrine

signaling [20]. Moreover, gut microbial members also have immunomodulation effects on bees [21] and protect honeybees from the opportunistic pathogens [22], which is critical to their survival and health in a complex environment [23]. Furthermore, there are known interaction between honeybee intestinal microorganisms and some honeybee diseases, for example, *Nosema ceranae* infection can promote proliferation of yeasts in honeybee gut and cause disease [24]. Therefore, intestinal bacteria are important for the bee hosts, and more investigations on the roles of different gut microbial members are needed. Compared with other animals, *A. mellifera* harbors a simple, recurring, and stable set of gut bacteria, including shared core phylotypes of *Gilliamella*, *Snodgrassella*, *Bifidobacterium*, *Lactobacillus* Firm4 and Firm5, and several host-specific phylotypes [25]. These bacteria are host-adapted, and each species cluster occupies particular niches and spatial locations in the host [21]. Although closely related bacterial species co-colonize stably in the bee gut, competitions exist between species. For example, food polysaccharide is an important factor in determining the coexistence state of *Lactobacillus* species in the gut of honeybee [26]. While the honeybee gut is composed of a limited number of bacterial phylotypes, metagenomic and single-cell genomic analyses revealed significant strain-level diversity [27, 28]. Moreover, individual strains with unique genomic potentials possess different capabilities, which are functionally relevant to hosts' nutrition metabolism and health [29, 30].

In this study, we first constructed a microfluidic droplets platform and generated droplets encapsulated with individual bacterial cells from the honeybee gut. Subsequently, we performed incubation and determined the growability of microorganisms within microfluidic droplets. To demonstrate the utility of our platform for measuring strain-level diversities, we employed metagenomic sequencing and obtained potentially novel strains from different bacteria genera. The comparative genomic analysis of *Lactobacillus panisapium* found that strains from *A. mellifera* contain a set of genes encoding arabinofuranosidase, which is likely important for the survival of bacteria in competitive environments.

Methods

Honeybee gut sample collection

The honeybees (*A. mellifera*) used in this study were obtained from an apiary in Kunming, Yunnan Province, China, and all individuals were hive bees. Newborn adult bees were collected from one single frame ~ 10 days after emergence. As described in previous studies [31], the entire guts were aseptically dissected by gently pulling the strings without touching the abdomen surface using

sterilized forceps. Subsequently, dissected guts were directly crushed in 25% (*v/v*) glycerol using an electric tissue grinder (OSE-Y30; Tiangeng Biotech Co., Ltd., Beijing, China). We obtained intestinal samples from a total of 30 individual bees, and the dissected guts were pooled and thoroughly mixed. Then the gut homogenate was aliquoted and stored at -80°C for subsequent droplet generation processes.

Microfluidic droplets generation

Droplets were made on a droplet entrapping microfluidic cell-sorter (DREM cell; Yuanqing Tianmu Biotechnology, Ltd., Wuxi, Jiangsu, China). The microfluidic chip was designed by AutoCAD and manufactured via soft lithography (Additional file 1: Figure S1). The negative photoresist SU-8 2015 (Westborough, MA, USA) was rotated and coated on a 4-inch silicon wafer to obtain the $15\ \mu\text{m}$ height channel. The Polydimethylsiloxane (Midland, MI, USA) prepolymer was mixed with the curing agent and poured onto the silicon mold. After removing bubbles, it was heated overnight at 65°C . The microchannel pattern was peeled off from the silicon mold and punched at a set position to form an inlet and outlet for the sample and reagent. The chips and glass slides were exposed to 140 W of oxygen plasma (PDC002; Harrick Plasma, Ithaca, NY, USA) for 60 s and heated for 24 h at 120°C . The chip consists of two inlets for the continuous oil phase and aqueous phase, respectively, and one outlet for collecting highly monodisperse water-in-oil emulsions. On the chip, the two liquids confined within microfluidic channels are brought together using pressure pumps, followed by the subsequent formation of droplets at the flow-focusing junction due to shear stress.

To guide the subsequent microfluidic droplet generation process, we first roughly estimated the bacterial concentration of the mixed intestinal sample. It was diluted and plating on Brain Heart Infusion (BHI; Oxoid, Basingstoke, UK) supplemented with 5% (*v/v*) defibrinated sheep blood (Solarbio, Beijing, China) at 35°C under a CO_2 -enriched atmosphere (5%) for 2 days for plate counting, and the concentration of the sample was about 10^8 CFU/mL. The droplet generation oil for EvaGreen[®] (Bio-Rad Laboratories, Inc., Hercules, CA, USA) was used for the continuous oil phase. According to the Poisson distribution $P(X=n) = e^{-\lambda}(\lambda^n/n!)$ (Additional file 1: Figure S2), the droplet occupancy (n) is related to the average number of cells per droplet (λ) given by the equation $\lambda = \rho V$, where V is droplet volume, and ρ is cell density. In this study, assays were performed using λ values of 0.3 to minimize the number of droplets loaded with two more bacterial cells; therefore, for a given droplet volume, there is only one corresponding suspension concentration (Additional

file 1: Table S1). In this study, two pressure pumps were used to control the oil and cell suspension flow rates, and we controlled the droplet volume by adjusting the flow rate ratio to a diameter of $30\ \mu\text{m}$ ($\sim 14\ \text{pL}$). Therefore, the frozen intestinal samples were centrifuged ($5000\times g$, 5 min), washed, and resuspended in BHIB (CM1135, Oxoid Ltd., Basingstoke, UK) and MRS broth (CM1175, Oxoid Ltd., Basingstoke, UK) respectively and diluted to $\sim 2 \times 10^7$ CFU/mL for droplets generation. In addition, the suspension was incubated with $1\ \mu\text{g/mL}$ DAPI stain solution (E607303; Sangon Biotech Co., Ltd., Shanghai, China) at 35°C for droplet generation. Then, the droplets were introduced into a hemocytometer plate and then visualized by an inverted fluorescence microscope (Eclipse Ts2; Nikon Instech Co., Ltd., Tokyo, Japan) to test whether the distribution of bacteria was consistent with the theoretical calculation under these parameters. For the incubation of the droplets, $\sim 3\ \text{mL}$ of the emulsion was loaded into Teflon tubes (about 7.7 m, $\phi = 0.71\ \text{mm}$; Suzhou Volsun Electronics Technology, Suzhou, China) to maintain stability during the incubation process. Three samples were generated separately for each type of medium, each containing around 2×10^8 droplets. All samples were incubated for 120 h at 35°C under a CO_2 -enriched atmosphere (5%). After incubation, the droplets were introduced into a cell counting chamber slide (Bodboege, Shenzhen, China) for observation. The inverted fluorescence microscope (Nikon) was also used to examine the stability of microfluidic droplets and microbial growth.

Droplet DNA extraction and shotgun metagenomic sequencing

After cultivation, we pipetted out all the droplet emulsion using sterile syringes, mixed it with an equal volume of 1H,1H,2H,2H-Perfluoro-1-octanol (PFO; CAS647-42-7, Shanghai Aladdin Bio-Chem Technology Co., Ltd., Shanghai, China), vortexed and centrifuged the solution, and removed the oil and PFO carefully. The Ezup Column Bacteria Genomic DNA Purification Kit (Sangon Biotech Co., Ltd., Shanghai, China) was used for DNA extraction. The gut DNA was extracted using the cetyltrimethyl ammonium bromide (CTAB) buffer method [33]. All DNA samples were submitted to the Novogene Company (Beijing, China) for shotgun metagenome sequencing. NEBNext UltraTM II DNA Library Prep Kit for Illumina (New England Biolabs, MA, USA) was used for the generation of sequencing libraries, and Qubit 3.0 Fluorometer (Life Technologies, Grand Island, NY, USA) and Agilent 4200 (Agilent, Santa Clara, CA, USA) system were used for library quality assessment. The libraries

were then sequenced on the Illumina HiSeq platform with 150-bp paired-end reads.

Species- and strain-level community profiling

After the sequencing results were obtained, fastp [32] was used for adaptor trimming and quality control of the raw sequencing data. Then reads are aligned to the genome of *A. mellifera* (GCA_003254395), and any *A. mellifera* reads were removed from the metagenomic data. The species- and strain-level community profiling and gene content estimation were performed using the Metagenomic Intra-Species Diversity Analysis System (MIDAS) pipeline [34]. As described in our previous study [35], the custom database included genomes of pure isolates from the guts of *A. mellifera*, *Apis cerana*, *Apis dorsata*, and *Bombus* species. The relative abundance of species clusters was estimated by mapping quality-filtered reads to the database of phylogenetic marker genes using HS-BLASTN with the “run_midass.py species” module. Then, the results across all samples were combined using “merge_midass.py species”.

We subsequently used the “run_midass.py snps” module of the MIDAS pipeline to profile single nucleotide polymorphisms (SNPs) by identifying single nucleotide variants (SNVs) diversity for each species cluster in metagenomic samples. The genomes with the highest completeness and lowest contamination were selected as representatives for each species cluster, and our metagenomic reads were aligned to the reference genomes using Bowtie2 [36]. Pileups of each sample were generated using SAMtools [37], and the nucleotide variation statistics were then counted at each genomic site. The results were merged using the “merge_midass.py snps” module to generate core genomic SNP matrices to compare nucleotide variants in genomic loci and metagenomic samples present in different cultured samples. We focused on the bi-allelic SNVs prevalent in more than 5% of the samples. Strain-level diversity within species clusters in each sample was estimated by quantifying the fraction of SNVs in protein-coding genes (number of SNVs/length of genes). We also generated a Jaccard distance matrix based on shared polymorphic sites and performed principal coordinate analysis and visualization based on the pairwise fractions of shared SNVs for different samples using the vegan package [38].

Metagenome binning and functional annotation of MAGs

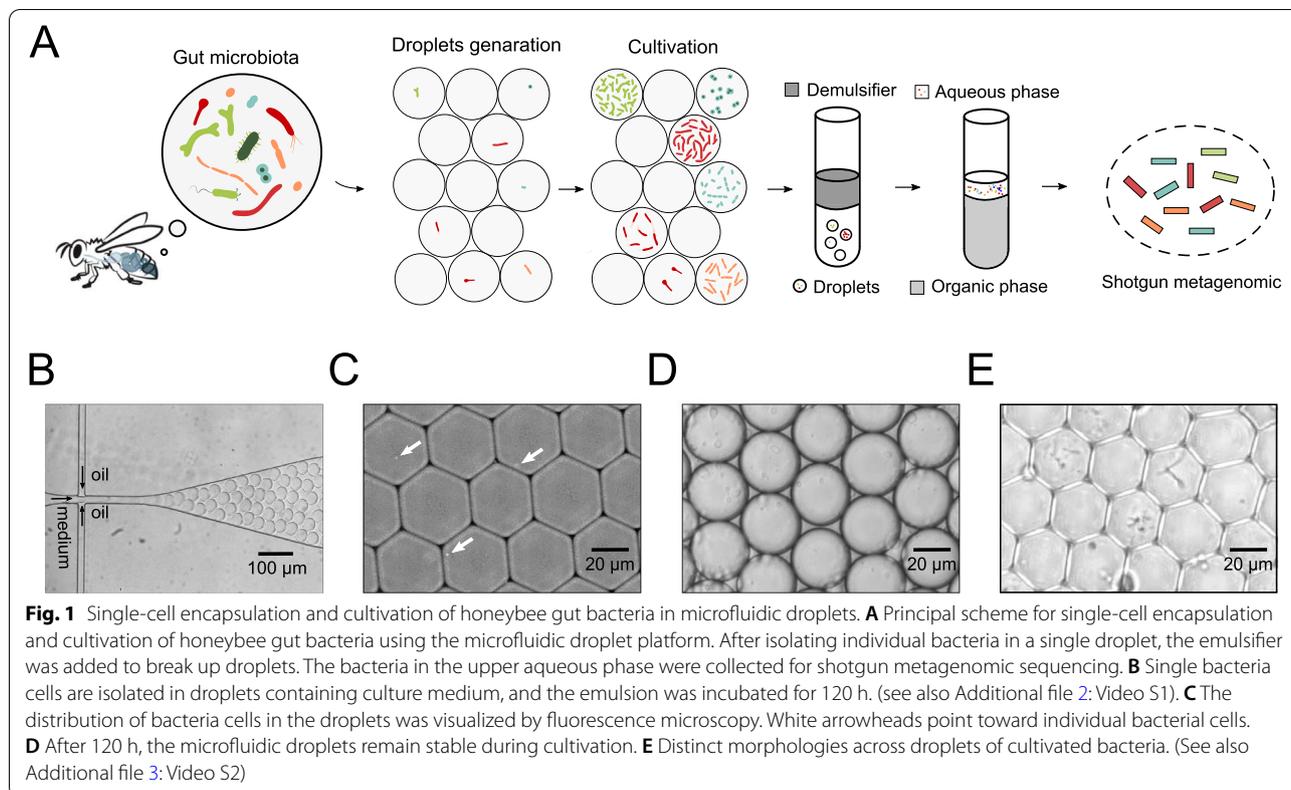
Metagenomic binning was performed using the metaWRAP pipeline [39]. Following de novo assembly with the metaSPAdes [40], the quality-controlled reads (about 30 million reads per sample) were mapped to the assembled contigs using Bowtie2 to generate a coverage score for individual contigs. The metagenome-assembled genomes

(MAGs) were recovered from each sample independently using three different tools: CONCOCT, MaxBin, and metaBAT. Subsequently, the three final bin sets produced were consolidated into a single and more robust bin set with the minimum completion (-c 50) and maximum contamination (-x 30) parameters using the “Bin_refinement” module in metaWRAP. The completeness and contamination of each MAG were estimated using CheckM software [41], and each bin was taxonomically assigned against the Genome Taxonomy Database with GTDB-tk [42]. To get a complete picture of the exact distribution of microbial members in the samples, we did not perform de-duplication. In a further study, phylogenetic analyses of genomes were conducted with the PhyloPhlAn 3.0 under the “--diversity low” parameter [43], and the iTOL web-based software [44] was used for the visualization of phylogenetic trees. Whole-genome average nucleotide identity (ANI) was calculated using the web service JSpeciesWS [45]. After protein-coding gene prediction using Prodigal software [46], the grouped amino acid sequences were submitted separately to the Orthovenn2 tool web server [47] for orthologous analysis to compare and annotate the orthologous cluster between genomes under the parameters of “E-value 10^{-15} ” and “Inflation value 1.0”. For *Lactobacillus* Firm5, after identifying unique gene clusters, cblaster software [48] was used to search and visualize for collocated protein-coding regions locally within our Firm5 database. We also performed carbohydrate-active enzymes (CAZymes) annotation of genomes against the dbCAN2 database using HMM search approach as reported by Zhang et al. [49]. All heatmaps were visualized by the pheatmap package [50] in R software.

Results and discussion

Microfluidic single-cell encapsulation and cultivation of honeybee gut microbiota

To isolate and culture individual bacterial cells from *A. mellifera* microbiota communities, an array of high-throughput droplet microfluidic technologies was developed (Fig. 1A). We generated microfluidic droplets of bacterial cells by integrating the commercial droplet entrapping microfluidic cell-sorter (DREM cell) with two pressure pumps and a high frame rate camera (Additional file 1: Figure S1). We selected MRS and BHI as the liquid medium because they are the most commonly used culture media for the gut bacteria of honey bees [51]. We mainly used the classic continuous droplet generation technique named “flow-focusing” [52]. The droplets were generated at the flow-focusing junction from the liquid culture medium into oil (Fig. 1B, Additional file 2: Video S1). The number of cells contained in the formed droplet is determined by the probability that a given volume of



initial cell suspension contains a given number of cells. It follows a Poisson distribution (Additional file 1: Figure S2, Table S1). Thus, the frozen intestinal suspensions were resuspended and diluted to 2×10^7 CFU/mL for droplets generation so that, in principle, $\sim 22\%$ of encapsulated microbial cells in the droplets (~ 14 pL) initially contain only one cell, and less than 5% of droplets contain two or more live bacterial cells stochastically. To ensure the single-cell deposition per droplet, we encapsulated DAPI-stained bacteria cells under the same parameters. The distribution of bacteria cells in the droplets was visualized using fluorescence microscopy (Fig. 1C). We found that most of the droplets encapsulated with bacteria contained only one cell, and the proportion of droplets with more than two cells was negligible. Notably, due to the random distribution of bacteria under current droplet generation methods, $\sim 70\%$ of the droplets contained no bacterial cells. Such a distribution of bacterial cells is likewise an obstacle to the production of single-cell droplets [53]. To obtain pure populations of positive droplets, the generation of droplets is always coupled with detection and sorting by flow-cytometry technologies [54]. Here, we did not sort bacteria to track single bacterial cells but performed shotgun metagenomes for the whole community to secure all genomic diversity. We generated approximately 2×10^8 droplets for each sample, which

means that more than 4×10^7 droplets only contain individual bacteria, sufficient for the subsequent sequencing analysis.

The generated droplets were then collected in a Teflon tube (Additional file 1: Figure S3) and incubated for 120 h at 35°C . Since most honey bee intestinal bacterial members can only grow in an elevated CO_2 environment [51], the droplets were incubated under a CO_2 -enriched atmosphere (5%). The droplets remained stable for several days in culture, and the diameter of the droplet did not change significantly (Fig. 1D). After 120 h of cultivation, microscopy showed the presence of live bacteria moving in the droplet (Fig. 1E, Additional file 3: Video S2), suggesting that a single bacterial cell could reproduce in the droplets. The consistent morphology of the bacteria in each droplet indicated that isolated viable strains could clonally replicate within a droplet. Moreover, the living cells from individual droplets showed different morphologies, indicating the segregation of diverse clonal populations from the microbial communities. Thus, our platform encapsulated individual honeybee gut members into microfluidic droplets, and the droplet environment could support the growth and metabolism of microorganisms. Ecological competition is prevalent in natural gut communities, and bacteria compete for space and resources [55]. In the traditional

culture process, certain microorganisms quickly dominate the culture system and prevent the growth of others using the same substrate but with a lower affinity [7]. The droplet-based cultivation isolated bacteria within an enclosed single droplet. The slow-growing bacteria avoid the competitive overgrowth, providing an opportunity to cultivate slow-growing microorganisms. However, many microorganisms rely on products of syntrophic partners and the accumulation of signals for quorum sensing. Thus, the microfluidic chips that grow bacteria are fully sealed chambers, likely prohibiting the cultivation of dependants [7].

After incubation, we pipetted out the droplet emulsion, mixed it with an equal amount of emulsifier (PFO), vortexed and centrifuged the solution, removed the organic phase (oil and PFO) located in the lower phase, and collected the bacteria in the upper aqueous phase for DNA extraction (Fig. 1A). In addition, the DNA extracted from preculture intestinal homogenates was also submitted for sequencing.

Community structure after microfluidic droplet cultivation

Approximately 30 million pair-end reads (150 bp) per sample were produced using shotgun metagenomic sequencing technology. After base quality control, reads were aligned to the genome of the host, and ~7% of reads derived from the bee host were removed. For samples after microfluidic-droplets incubation, almost all reads were from bacteria without host contamination. Our data analysis consists of two major steps (Fig. 2A). We first estimated bacterial species abundance and strain-level genomic variation, including SNPs from shotgun metagenomic reads, using the MIDAS pipeline with a custom database for bee microbiome [35]. Then, we restored the bacterial genomes using genome-resolved metagenomic approaches based on the metagenomic assembly and clustering of contigs through the metagenomic binning procedure.

The results of the MIDAS pipeline show that the uncultured gut community samples were dominated by five core bee gut members, and most of the common gut microbial species in *A. mellifera* could be detected (Fig. 2B). However, after our droplet-based cultivation, genus-level community compositions changed significantly, and the results differed depending on the medium

(Fig. 2B). *Bifidobacterium* and *Gilliamella* were significantly enriched after 120-h incubation in BHI-droplets, while *Lactobacillus* and *Apilactobacillus* were enriched within the MRS-cultured groups. We then focused on species-level changes in these genera. After cultivation, the alpha diversities of communities from each sample were significantly reduced (Additional file 1: Figure S4). Principal coordinates analysis (PCoA) revealed distinct clustering of microbiota composition for the different medium groups (Additional file 1: Figure S4). Further, the relative abundance of specific species was significantly increased after cultivation, such as *Bifidobacterium apousia* and *Bifidobacterium choladohabitans* of the BHI-cultured group. Likewise, *Apilactobacillus kunkelii* and *Lactobacillus helsingborgensis* were accumulated in MRS-droplets, and some rare species (< 0.01% relative abundance) in the community have been enriched to higher abundance in droplets, such as *Gilliamella apicola* (> 19% in BHI droplets), suggesting that they may be more adapted to the incubation conditions. Correspondingly, some species may not grow or grow slowly in the droplets, exhibiting a decrease in relative abundance during cultivation. Therefore, droplet-based culture reduced the species-level diversities compared to the original honeybee gut communities, consistent with previous studies [18].

Since it has been suggested that the high strain-level diversities from *A. mellifera* species, especially for *Lactobacillus* [56] and *Bifidobacterium* [57], we compared the strain-level genomic variation by calculating the fraction of single nucleotide variants (SNVs) sites among all profiled sites for each species (Fig. 2C, D). We first focused on *Lactobacillus* species: *Lactobacillus apis*, *L. helsingborgensis*, and *Lactobacillus melliventris* harbor more than 2% SNVs in most samples, while the *Lactobacillus kullabergensis* showed a lower level of variations than other *Lactobacillus* species. In addition, we observed a higher proportion of SNVs for *L. apis* in the BHI-cultured group. In MRS-cultured groups, *L. helsingborgensis* and *L. melliventris* were detected with more single-nucleotide variants, indicating the adaptability of the medium varies significantly among different strains. As for *Bifidobacterium*, there were few differences between different species from droplets-cultured groups; nearly all the bacterial species had

(See figure on next page.)

Fig. 2 Strain-level compositions of honeybee gut shift after droplet-based cultivation. **A** Schematic of the data analysis workflow. After base quality control, the bacterial species abundance and strain-level genomic variation were estimated using the MIDAS pipeline, and metagenomic binning was performed to restore the metagenome-assembled genomes. **B** Species-level profiles for the gut sample (GUT) and the picodroplet samples using the Brain Heart Infusion (BHI) or the MRS broth (MRS) after cultivation. **C, D** Fraction of single-nucleotide variants (SNVs) within core genes in each sample for *Lactobacillus* (**C**) and *Bifidobacterium* (**D**). **E–H** Principal coordinate analysis plots based on the pairwise fractions of shared SNVs (Jaccard distance) for the species *Lactobacillus melliventris* (**E**), *Lactobacillus helsingborgensis* (**F**), *Bifidobacterium choladohabitans* (**G**), and *Bifidobacterium polysaccharolyticum* (**H**). Dots represent individual samples, color-coded by the medium used for cultivation

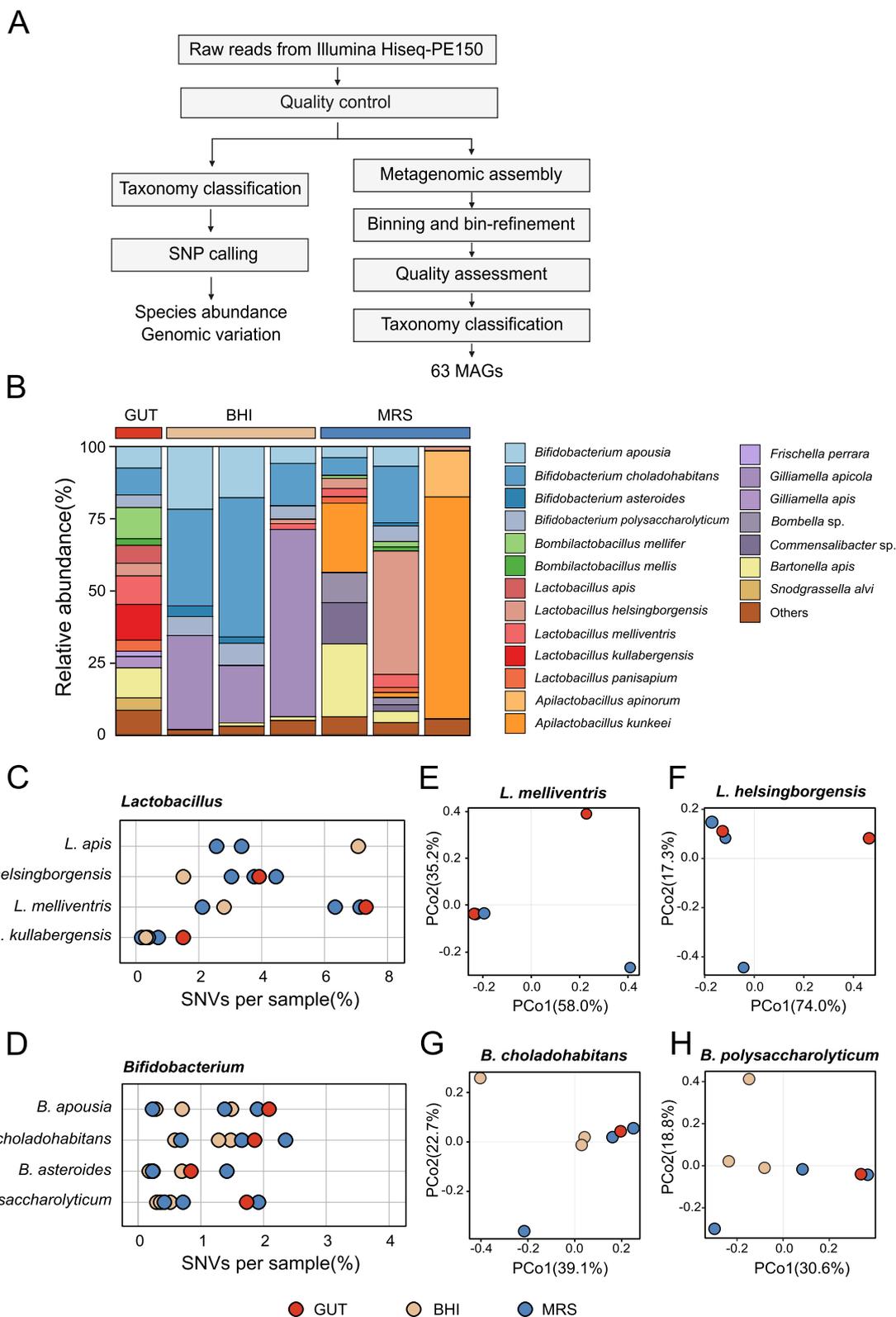
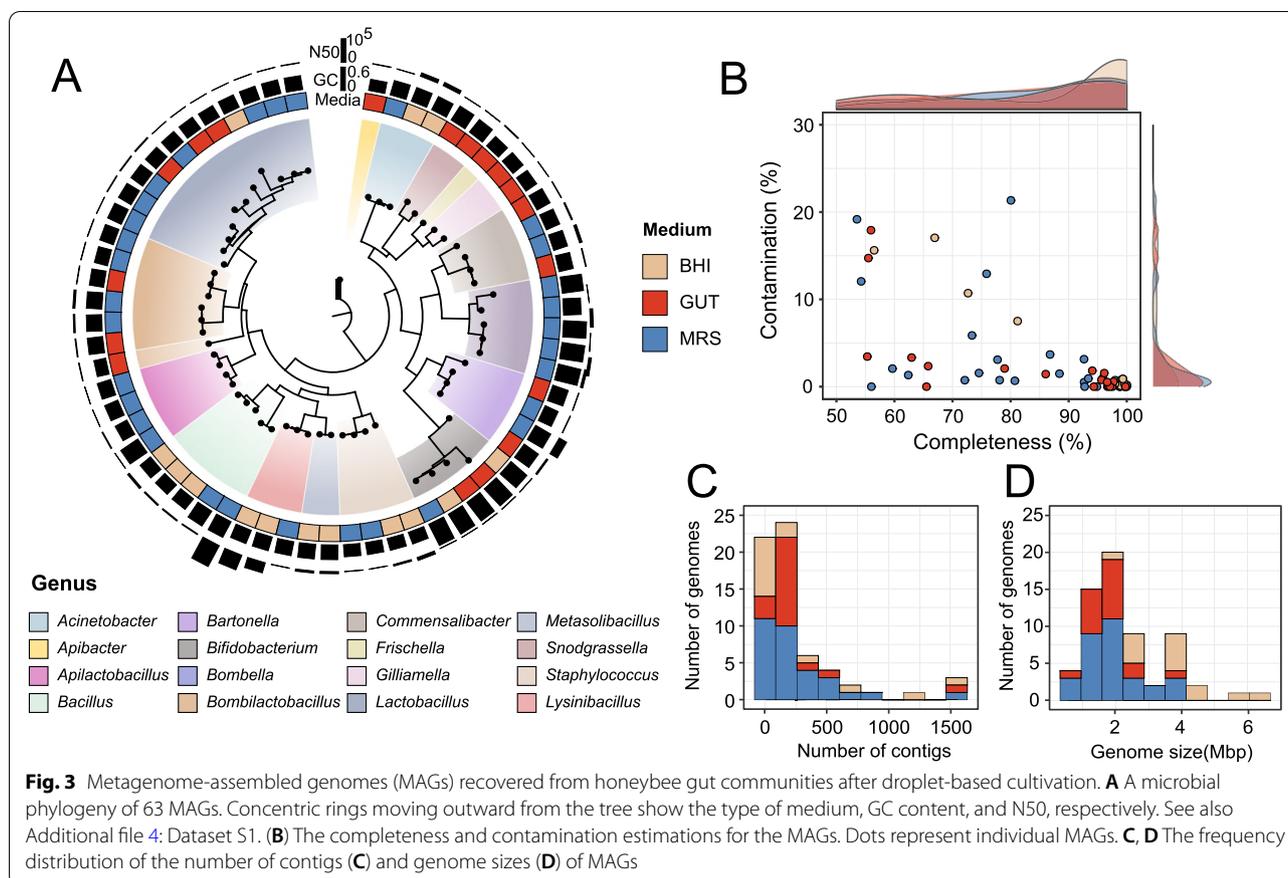


Fig. 2 (See legend on previous page.)



less than 2.5% SNVs. Interestingly, we found that for *L. kullabergensis* and *L. melliventris* in MRS-cultured groups, *Bifidobacterium coryneforme*, *B. choladohabitans*, and *Bifidobacterium polysaccharolyticum* in BHI groups, the fractions of SNVs were even higher than the gut samples. This implied that some rare strains were enriched by droplet-based cultivations, and they were not detected by the previous depth of sequencing. To visualize the distribution of SNVs across samples, we calculated Jaccard distances between all pairs of samples based on shared SNVs. PCoA revealed that the composition of SNVs was significantly different among the same species under different culture conditions, illustrating that distinctly different strains were specifically enriched under different culture conditions (Fig. 2E–H).

In summary, we obtained multiple morphologies of microorganisms through our high throughput microfluidic droplets cultivation. Despite the number of species being close to the uncultured group, the strain-level variations differed among samples, demonstrating that our culture strategy did enrich different individual bacterial cells of honeybee gut communities in a single droplet. For some species, we even could observe higher

strain-level diversities after droplet cultivation relative to uncultured samples, demonstrating the potential of our platform for the isolation and enrichment of rare microbial strains. These strains are often challenging to detect and quantify because of their low abundance in natural communities, and high sequencing depth is essential for their investigation.

Sixty-three draft MAGs were recovered after microfluidic droplet cultivation

Due to some rare strains being detected in our droplets-cultured samples, de novo assembly and binning were used on shotgun metagenomic reads. Metagenomes were assembled independently to reduce the influence of strain variation and improve the recovery of closely related genomes [58]. Sixty-three MAGs were identified after refinement and filtering of the resulting population genomes (Fig. 3A, Additional file 4: Dataset S1), and we evaluated their quality (Fig. 3B). About 55 MAGs had completion scores above 60%, and almost all had less than 20% genomic contamination. Further, recovered MAGs had a median genome size of 1.9 Mbp, 133 contigs, and a median N50 of 20.6 kbp (Fig. 3C, D, Additional file 4: Dataset S1). Taxonomic annotation using GTDB

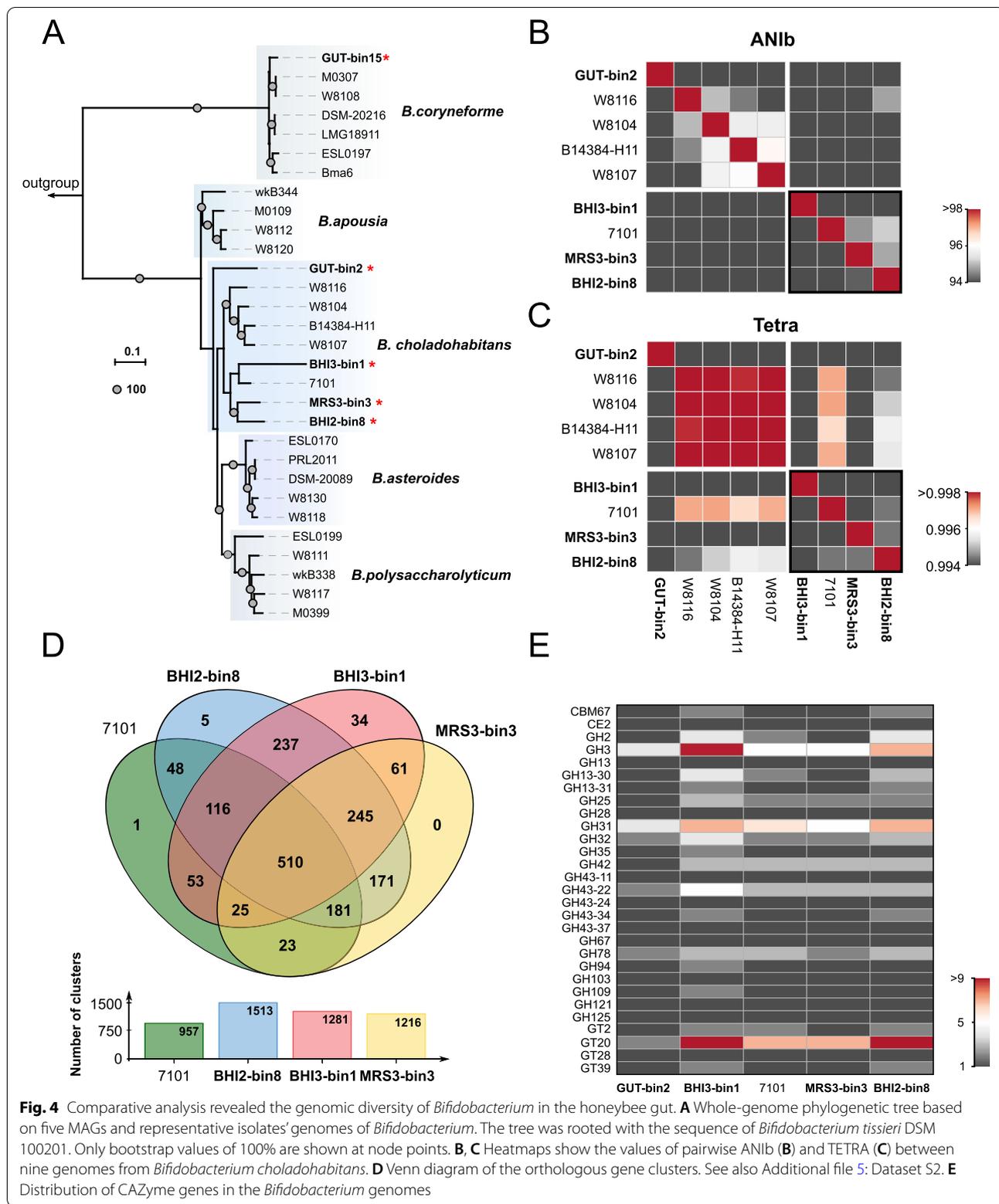


Fig. 4 Comparative analysis revealed the genomic diversity of *Bifidobacterium* in the honeybee gut. **A** Whole-genome phylogenetic tree based on five MAGs and representative isolates' genomes of *Bifidobacterium*. The tree was rooted with the sequence of *Bifidobacterium tissieri* DSM 100201. Only bootstrap values of 100% are shown at node points. **B**, **C** Heatmaps show the values of pairwise ANiB (**B**) and TETRA (**C**) between nine genomes from *Bifidobacterium choladohabitans*. **D** Venn diagram of the orthologous gene clusters. See also Additional file 5: Dataset S2. **E** Distribution of CAZyme genes in the *Bifidobacterium* genomes

indicated the majority of our MAGs belonged to phyla *Proteobacteria* and *Firmicutes* (Fig. 3A, Additional file 4: Dataset S1). We also obtained five bacterial genomes from *Actinobacteria*, all of which were *Bifidobacterium*. To further characterize assembled genomes, we classified MAGs and reconstructed genomic phylogenetic trees with genomes from our database (Fig. 4A, Fig. 5A, Additional file 1: Figure S5–S9).

Investigation of samples after droplet-based cultivation revealed the diversity of *Bifidobacterium* in honeybee gut

Bifidobacterium spp. is one of the core gut members of *A. mellifera*. So far, five species (*B. coryneforme*, *B. apousia*, *B. choladohabitans*, *Bifidobacterium asteroides*, and *B. polysaccharolyticum*) have been isolated and characterized from *A. mellifera* [57]. However, based on the phylogenetic tree (Fig. 4A), all four MAGs from droplets-cultured samples were clustered with *B. choladohabitans*. Moreover, we noted that three MAGs were clustered with the strain of *B. choladohabitans* 7101, implying the existence of potential novel cluster species. To quantify the magnitude of differences between genomes, we evaluated the Overall Genome Relatedness Index based on average nucleotide identity (ANI) analysis (ANIB based on BLASTN) and tetranucleotide usage patterns. The ANIB values between strains from different clusters ranged from 89 to 94% (Fig. 4B). The pairwise comparisons are almost > 95% within one cluster (W8116, W8104, B14384-H11, and W8107), indicating that this cluster represents a species-level taxon. The TETRA values also illustrated this point (Fig. 4C). However, for another cluster mainly consisted of our MAGs (BHI3-bin1, 7101, MRS3-bin3, and BHI2-bin8), the pairwise comparisons were between 92 and 95%, such findings implied the genomic diversity of these strains. Indeed, Ellegaard et al. [59] examined the density distribution of similarity within this specific SDP from *Bifidobacterium*. They found a significant percentage of ANI values between 90 and 95% and speculated that additional equally divergent strains are presented in the bee gut community. Based on this, four different bifidobacterial species within this SDP were identified and characterized [57]. But we found a new cluster within the discontinuous regions, and the genomes of this cluster had almost all paired ANI values <95% nucleotide sequence identity for species by Richter and Rosselló-Móra [60], which further demonstrated the diversity of *Bifidobacteria* in the honeybee gut, and suggested the presence of species not previously detected by conventional methods.

To gain further insights into the differences between our MAGs and *B. choladohabitans* 7101, we used OrthoVenn2 [47] to identify orthologous genes among our cluster (Fig. 4D). The four strains of *Bifidobacterium*

possessed 1710 gene families. In contrast, a core genome comprised 510 clusters of orthologous (Additional file 5: Dataset S2), only accounting for 29.8% of all gene families, indicating that there are apparent genetic differences between all these strains. Most of the annotation functions of core homologous clusters were involved in metabolic process, biological process, cellular metabolic process, hydrolase activity, and molecular function, which may be closely related to the survival of these strains.

Remarkably, the number of unshared clusters of the MAG "BHI3-bin1" was higher than related species. The subsequent GO enrichment analysis showed that six gene clusters were related to transmembrane transport (GO: 0055085) and fatty acid biosynthetic process (GO: 0006633) (Additional file 5: Dataset S2). These clusters encode multiple sugar-binding transport system permease protein (MsmG) and 3-oxoacyl-ACP synthase II (FabF), which were associated with the protein-dependent transport system responsible for the uptake of melibiose and raffinose. This implied the unique advantages of carbohydrate transport and fatty acid synthesis compared with the other strains.

In addition, we compared MAGs from the different culture mediums, and the two BHI-associated MAGs shared 237 unique homologous gene clusters (Additional file 5: Dataset S2). GO analysis indicated that 16 were related to transmembrane transport, including formylaminopyrimidine transport permease ThiX, polygalacturonan/rhamnogalacturonan transport system permease YtcP, and putative ABC transporter permease ORF2. In addition, seven gene clusters were related to the inositol catabolic process (GO:0019310), encoding inositol 2-dehydrogenase, 3D-(3,5/4)-trihydroxycyclohexane-1,2-dione hydrolase, and inosose dehydratase. Interestingly, inositol utilization was reported as part of cell mass generation of *Corynebacterium glutamicum* during growth on the BHI [61].

The members of *Bifidobacterium* have been identified as the key polysaccharide degrader in the bee gut community [30]. However, it has been reported that strains from different phylogenetic clusters vary in the CAZyme repertoires for hemicellulose metabolism [57]. Therefore, we comprehensively analyzed the composition of CAZyme genes in our assembly genomes. Generally, numerous carbohydrate-binding modules (CBMs), glycoside hydrolases (GHs), carbohydrate esterases (CEs), and glycosyltransferases (GTs) were identified in all genomes. The genomes from the same cluster possessed similar CAZyme profiles (Fig. 4E), which agrees with the previous genomic study [30]. We further focused on the suspected novel clusters. MAGs from BHI post-culture samples tended to have more GH3, GH31, and GT20,

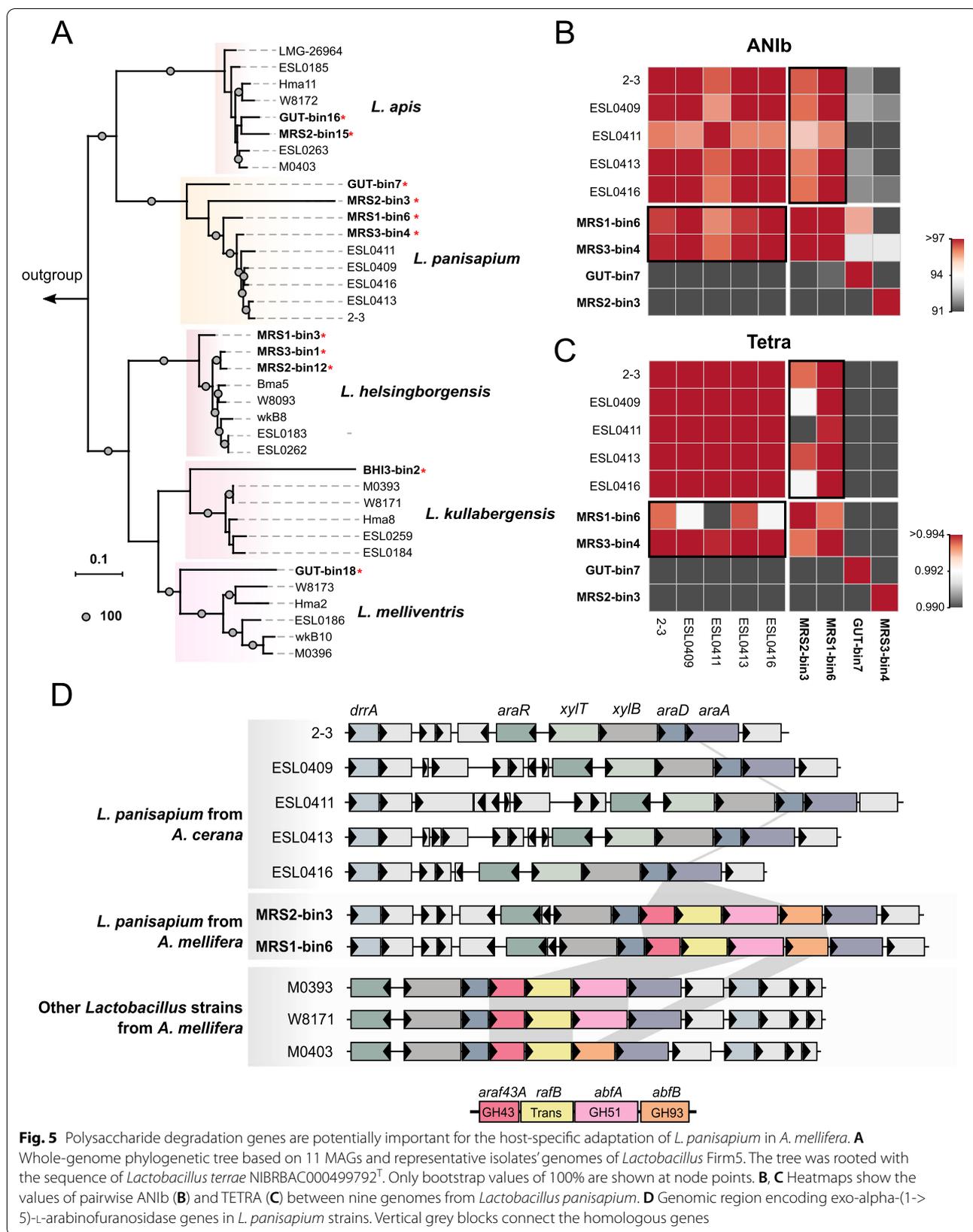


Fig. 5 Polysaccharide degradation genes are potentially important for the host-specific adaptation of *L. panisapium* in *A. mellifera*. **A** Whole-genome phylogenetic tree based on 11 MAGs and representative isolates' genomes of *Lactobacillus* Firm5. The tree was rooted with the sequence of *Lactobacillus terrae* NIBRBAC000499792^T. Only bootstrap values of 100% are shown at node points. **B, C** Heatmaps show the values of pairwise ANIb (**B**) and Tetra (**C**) between nine genomes from *Lactobacillus panisapium*. **D** Genomic region encoding exo-alpha-(1->5)-L-arabinofuranosidase genes in *L. panisapium* strains. Vertical grey blocks connect the homologous genes

revealing the different performances of some bifidobacterial strains in polysaccharide degradation.

Polysaccharide utilization is important for the survival of *L. panisapium* in *A. mellifera*

Lactobacillus Firm5 is the most widely distributed and abundant phylotype in the bee gut microbiota [62]. Four deep-branching species of Firm5 have been identified in the gut of *A. mellifera*, with pairwise average ANI values below 90% [59, 63]. Here, we obtained 11 MAGs from Firm5, seven of which were derived from the samples of MRS-droplet cultivation. Unlike *Bifidobacterium*, the assembled MAGs are distributed in different species clusters (Fig. 5A). Specially, we obtained MAGs of *L. helsingborgensis* from three different MRS replicating samples, while only one MAG of *L. kullabergensis* was from the BHI sample. Notably, three MAGs from the MRS group and one assembly from the gut sample fell into the *L. panisapium* cluster, with five isolates from *A. cerana* [62]. To examine the genomic similarities, we compared the intra-species genomic variation of *L. panisapium* by pairwise comparison of ANIb and Tetra values (Fig. 5B, C). Except for one genome from the gut sample (GUT-bin7) and one genome (MRS3-bin4) with relatively low completeness (53.53%), two MAGs obtained from the MRS-droplet cultivation showed high similarity to the other strains from *A. cerana* (> 95% for ANIb, > 0.991 for Tetra).

Remarkable strain-level diversities were observed within *Lactobacillus* Firm5, and the genomic variation was associated with the functions in carbohydrate metabolism [56]. Thus, we compared the homologous gene clusters of different strains. Interestingly, genes specific to *A. mellifera* strains were related to the L-arabinose metabolic process (GO:0046373; GO:0019569) (Additional file 6: Dataset S3), including genes of *araf43A* (extracellular exo-alpha-(1-> 5)-L-arabinofuranosidase), *rafB* (raffinose permease), *abfA* (intracellular exo-alpha-(1-> 5)-L-arabinofuranosidase), and *abfB* (extracellular exo-alpha-L-arabinofuranosidase). These genes were explicitly inserted between *araA* and *araD*, which are part of the L-arabinose operon [65], forming a co-locate gene locus with *araR* (LacI-type transcriptional regulator), *xylT* (xylose transporter), and *xylB* (xylulokinase). However, these genes were totally absent in strains from *A. cerana*, while other monosaccharide degradation genes showed synteny (Fig. 5D). Moreover, the gene set of L-arabinofuranosidase was present in other Firm5 species from western honeybees, which implied that they were important for the colonization of *A. mellifera*.

L. panisapium was first identified from the honey bread of *A. cerana* [66] and formed monophyletic clusters from

other species, suggesting that it might be a specific species to *A. cerana*. Although trace amounts of shotgun metagenomic reads from *L. panisapium* were detected in gut samples from *A. mellifera*, it was hypothesized to be the bacteria transfer due to contact with non-native symbionts [62]. Here, we obtained bacterial genomes that showed high similarity with the known *L. panisapium* strains from *A. mellifera*. However, significant variation was identified between genomes from different bee hosts, suggesting that these *L. panisapium* strains enriched by microfluidic droplets were native to *A. mellifera* rather than transfer or contamination. Our microfluidic droplet cultivation strategy enriched the low-abundance bacteria, probably because of the exclusion of their competition with other high abundant strains.

The honeybee diet has various polysaccharide components, and intestinal bacteria are the main agents in the degradation of these polysaccharides [30]. *Lactobacillus* Firm5 are significant fermenters of dietary carbohydrates for bees [56]. However, gut bacteria of bees have obvious distinct repertoires of carbohydrate-active enzymes and occupy different glycan niches. For Firm5, Brochet et al. [67] demonstrated that polysaccharide fractions are the main determinants for the structure of different species. We showed that *L. panisapium* strains from *A. mellifera* contain unique genes of polysaccharide metabolism, specifically for the hydrolysis of arabinoxylans to oligosaccharides. Moreover, these genes cluster with monosaccharide metabolism genes, forming CAZyme gene clusters. Similar structures have been found in *Bifidobacterium* from *A. mellifera* [30]. Although *A. mellifera* and *A. cerana* have a similar dietary regimes, there may be differences in the specific composition of their diets [64]. *L. panisapium* in *A. mellifera* possibly compete to utilize polysaccharides from the host's diet, which confers a selective advantage for colonization.

Gut ecosystems often contain strains with highly variable genetic contents, and the strain-level diversity is substantial in host-associated bacterial communities [68]. Genomic analysis has shown that strain-level variants within microbial species are essential in determining functional capacities in honeybees [29, 30], which is also documented for humans [69]. It is crucial to identify and assess the strains with distinct genetic repertoires, probably affecting the establishment of a healthy symbiosis and host biology [70]. However, the gut microbial community is often complex, requiring a high sequencing depth to achieve a satisfactory resolution [16]. Through the microfluidic droplet cultivation platform, we were able to identify rare strains. Separate encapsulation excludes the effect of microbial competition during the culture process, allowing for the enrichment of strains that are hardly detected. Inevitably, certain strains

may not be cultured due to the absence of co-dependent individual or population sensing signals. Overall, we established a microfluidic cultivation strategy combined with metagenomic analysis for honeybee gut symbionts, encouraging potential applications in other complex microbiota communities.

Conclusions

In this study, we established a droplet microfluidic platform for the high-throughput culture of honeybee gut bacteria combined with shotgun metagenomic and binning strategy. Individual droplet encapsulation excludes the effect of competition during cultivation. However, it should be noted that the cultivation output would be modulated due to the selected culture medium. Strain-level analysis revealed potential novel species from the honeybee. In particular, a comparative analysis of *L. panisapium* found that strains from *A. mellifera* contain a set of genes related to the utilization of diet polysaccharides, which is likely important in competitive environments. In addition, cultures of Eukaryotes from the gut, such as fungi, will likely be studied in the future by adjusting droplet size, culture conditions, and sequencing approaches, which will further expand our understanding of the complex members of the gut. Overall, our results demonstrate the adaptability of droplet-based cultivation in investigating microbial diversity in the honeybee gut. This approach is also applicable for other complex communities, which may validate functional analyses predicted by the genomes.

Abbreviations

PFO: 1H,1H,2H,2H-Perfluoro-1-octanol; DAPI: Diamidino-2-phenylindole; CTAB: Cetyl trimethylammonium bromide; MIDAS: Metagenomic intra-species diversity analysis system; SNP: Single nucleotide polymorphism; SNV: Single nucleotide variant; BHI: Brain heart infusion; ANI: Whole-genome average nucleotide identity; MAG: Metagenome-assembled genome; PCoA: Principal coordinates analysis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-022-01333-9>.

Additional file 1: Figure S1. Devices for microfluidic droplet generation. The microfluidic droplets of bacterial cells were generated by the droplet entrapping microfluidic cell-sorter integrating with two pressure pumps. The chip consists of two inlets for continuous and aqueous phases and one outlet for collecting the water-in-oil emulsions. The two liquids confined within microfluidic channels are brought together at the flow-focusing junction, and the generated droplets were collected in a Teflon tube. **Figure S2.** Poisson distribution curve. When the average number of cells per droplet (λ) is determined, the distribution of the number of bacteria in the droplet conforms to the Poisson distribution. **Figure S3.** The generated droplets are introduced into Teflon tubes for incubation. **Figure S4.** (A-B) The Shannon (A) and Simpson (B) diversity metrics of the data sets. (C-D) Microbiome compositions in different developmental stages are plotted on Jaccard (C) and Bray-Curtis (D) PCoA graphs. **Figure**

S5. Whole-genome phylogenetic tree based on MAGs and representative isolates' genomes from *Apibacter*. **Figure S6.** Whole-genome phylogenetic tree based on MAGs and representative isolates' genomes from *Bartonella*. **Figure S7.** Whole-genome phylogenetic tree based on MAGs and representative isolates' genomes from *Lactobacillus Firm4*. **Figure S8.** Whole-genome phylogenetic tree based on MAGs and representative isolates' genomes from *Lactobacillus kunkeei*. **Figure S9.** Whole-genome phylogenetic tree based on MAGs and representative isolates' genomes from *Acetobacteraceae*. **Table S1.** The bacterial concentration is required to generate a given volume of droplets.

Additional file 2: Video S1. Generation of microfluidic droplets.

Additional file 3: Video S2. Morphologies of gut bacteria in droplets after cultivation.

Additional file 4: Dataset S1. Information of metagenome-assembled genomes.

Additional file 5: Dataset S2. Reports for homologous gene cluster analysis of the novel cluster from *Bifidobacterium*.

Additional file 6: Dataset S3. Reports for homologous gene cluster analysis of *Lactobacillus panisapium*.

Acknowledgements

Not applicable.

Authors' contributions

YM, HZ, and CZ designed the research. YM and SL collected the samples. YM analyzed the data. YM and HZ wrote the manuscript. All authors revised the manuscript and approved the final version.

Funding

This work was funded by the National Natural Science Foundation of China Project 32170495.

Availability of data and materials

Sequencing data of the metagenomes and the Metagenome-assembled genomes have been deposited under BioProject PRJNA82536.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 28 April 2022 Accepted: 26 July 2022

Published online: 31 August 2022

References

- Lynch JB, Hsiao EY. Microbiomes as sources of emergent host phenotypes. *Science*. 2019;365:1405–9.
- Lee W-J, Hase K. Gut microbiota-generated metabolites in animal health and disease. *Nat Chem Biol*. 2014;10:416–24.
- Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, et al. The microbiome and human biology. *Annu Rev Genomics Hum Genet*. 2017;18:65–86.
- Zou Y, Xue W, Luo G, Deng Z, Qin P, Guo R, et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol*. 2019;37:179–85.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol*. 2016;1:1–6.

6. Janssen PH. Selective enrichment and purification of cultures of *Methanosaeta* spp. *J Microbiol Methods*. 2003;52:239–44.
7. Lewis WH, Tahon G, Geesink P, Sousa DZ, Ettema TJG. Innovations to culturing the uncultured microbial majority. *Nat Rev Microbiol*. 2021;19:225–40.
8. Zhu Y, Fang Q. Analytical detection techniques for droplet microfluidics—a review. *Anal Chim Acta*. 2013;787:24–35.
9. Kaminski TS, Scheler O, Garstecki P. Droplet microfluidics for microbiology: techniques, applications and challenges. *Lab Chip*. 2016;16:2168–87.
10. Xu B, Hu B, Wang J, Lan Y, Zhu Y, Dai X, et al. *Virgibacillus indicus* sp. nov. and *Virgibacillus profundi* sp. nov, two moderately halophilic bacteria isolated from marine sediment by using microfluidic streak plates. *Int J Syst Evol Microbiol*. 2018;68:2015–23.
11. Jiang C, Dong L, Zhao J, Hu X, Shen C, Qiao Y, et al. High-throughput single-cell cultivation on microfluidic streak plates. *Appl Environ Microbiol*. 2016;82:2210–8.
12. Li C, Gong Y, Wang X, Xu J, Ma B. Integrated addressable dynamic droplet array (aDDA) as sub-nanoliter reactors for high-coverage genome sequencing of single yeast cells. *Small*. 2021;17:2100325.
13. Terekhov SS, Smirnov IV, Malakhova MV, Samoilov AE, Manolov AI, Nazarov AS, et al. Ultrahigh-throughput functional profiling of microbiota communities. *Proc Natl Acad Sci*. 2018;115:9551–6.
14. Tremaroli V, Bäckhed F. Functional interactions between the gut microbiota and host metabolism. *Nature*. 2012;489:242–9.
15. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res*. 2017;27:626–38.
16. Jovel J, Patterson J, Wang W, Hotte N, O’Keefe S, Mitchel T, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front Microbiol*. 2016;7:459.
17. Villa MM, Bloom RJ, Silverman JD, Durand HK, Jiang S, Wu A, et al. Inter-individual variation in dietary carbohydrate metabolism by gut bacteria revealed with droplet microfluidic culture. *mSystems*. 2020;5:e00864–19.
18. Watterson WJ, Tanyeri M, Watson AR, Cham CM, Shan Y, Chang EB, et al. Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes. *eLife*. 2020;9:e56998.
19. Gallai N, Salles J-M, Settele J, Vaissière BE. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecol Econ*. 2009;68:810–21.
20. Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA. Honey bees as models for gut microbiota research. *Lab Anim*. 2018;47:317.
21. Kwong WK, Moran NA. Gut microbial communities of social bees. *Nat Rev Microbiol*. 2016;14:374–84.
22. Lang H, Duan H, Wang J, Zhang W, Guo J, Zhang X, et al. Specific strains of honeybee gut *Lactobacillus* stimulate host immune system to protect against pathogenic *Hafnia alvei*. *Microbiol Spectr*. 2022;10:e01896–21.
23. Ptaszyńska AA, Latoch P, Hurd PJ, Polaszek A, Michalska-Madej J, Grochowalski Ł, et al. Amplicon sequencing of variable 16S rRNA from bacteria and ITS2 regions from fungi and plants, reveals honeybee susceptibility to diseases results from their forage availability under anthropogenic landscapes. *Pathogens*. 2021;10:381.
24. Ptaszyńska AA, Paleolog J, Borsuk G. *Nosema ceranae* infection promotes proliferation of yeasts in honey bee intestines. *PLoS One*. 2016;11:e0164477.
25. 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:e1600513.
26. Mahler L, Niehs SP, Martin K, Weber T, Scherlach K, Hertweck C, et al. Highly parallelized droplet cultivation and prioritization of antibiotic producers from natural microbial communities. *eLife*. 2021;10:e64774.
27. Engel P, Martinson VG, Moran NA. Functional diversity within the single gut microbiota of the honey bee. *Proc Natl Acad Sci*. 2012;109:11002–7.
28. Philipp, Engel Ramunas, Stepanauskas Nancy A, Moran Paul M, Richardson. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. *PLoS Genetics*. 2014;10(9):e1004596.
29. 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:e01326-16.
30. 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*. 2019;116:25909–16.
31. Grunec L, Gentekaki E, Khongphinitbunjong K, Popluechai S. Distinct gut microbiota profiles of Asian honey bee (*Apis cerana*) foragers. *Arch Microbiol*. 2022;204:187.
32. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
33. Zheng H, Dietrich C, Radek R, Brune A. *Endomicrobium proavitum*, the first isolate of *Endomicrobia* class. Nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a group IV nitrogenase. *Environ Microbiol*. 2016;18:191–204.
34. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res*. 2016;26:1612–25.
35. Wu J, Lang H, Mu X, Zhang Z, Su Q, Hu X, et al. Honey bee genetics shape the strain-level structure of gut microbiota in social transmission. *Microbiome*. 2021;9:225.
36. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9:357–9.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
38. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’hara R, et al. Package ‘vegan’. *Community Ecol Package Version*. 2013;2:1–295.
39. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 2018;6:158.
40. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34.
41. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55.
42. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*. 2020;36:1925–7.
43. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun*. 2020;11:2500.
44. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;44:W242–5.
45. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*. 2016;32:929–31.
46. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11:119.
47. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2019;47:W52–8.
48. Gilchrist CLM, Booth TJ, van Wersch B, van Grieken L, Medema MH, Chooi Y-H. Cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinforma Adv*. 2021;1:vbab016.
49. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46:W95–101.
50. Kolde R, Kolde MR. Package ‘pheatmap’. *R Package*. 2015;1:790.
51. Romero S, Nastasa A, Chapman A, Kwong WK, Foster LJ. The honey bee gut microbiota: strategies for study and characterization. *Insect Mol Biol*. 2019;28:455–72.
52. Anna SL, Bontoux N, Stone HA. Formation of dispersions using “flow focusing” in microchannels. *Appl Phys Lett*. 2003;82:364–6.
53. Collins DJ, Neild A, DeMello A, Liu A-Q, Ai Y. The Poisson distribution and beyond: methods for microfluidic droplet production and single cell encapsulation. *Lab Chip*. 2015;15:3439–59.
54. Chabert M, Viovy J-L. Microfluidic high-throughput encapsulation and hydrodynamic self-sorting of single cells. *Proc Natl Acad Sci*. 2008;105:3191–6.
55. Coyte KZ, Rakoff-Nahoum S. Understanding competition and cooperation within the mammalian gut microbiome. *Curr Biol*. 2019;29:R538–44.
56. 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;28:2224–37.

57. Chen J, Wang J, Zheng H. Characterization of *Bifidobacterium apousia* sp. nov., *Bifidobacterium choladohabitans* sp. nov., and *Bifidobacterium polysaccharolyticum* sp. nov., three novel species of the genus *Bifidobacterium* from honey bee gut. *Syst Appl Microbiol*. 2021;44:126247.
58. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell*. 2019;176:649–62.
59. Ellegaard KM, Engel P. Genomic diversity landscape of the honey bee gut microbiota. *Nat Commun*. 2019;10:446.
60. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci*. 2009;106:19126–31.
61. Krings E, Krumbach K, Bathe B, Kelle R, Wendisch VF, Sahm H, et al. Characterization of myo-inositol utilization by *Corynebacterium glutamicum*: the stimulon, identification of transporters, and influence on l-lysine formation. *J Bacteriol*. 2006;188:8054–61.
62. 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;30:2520–31.
63. Olofsson TC, Alsterfjord M, Nilson B, Butler È, Vásquez A. *Lactobacillus apinorum* sp. nov., *Lactobacillus mellifer* sp. nov., *Lactobacillus mellis* sp. nov., *Lactobacillus melliventris* sp. nov., *Lactobacillus kimbladii* sp. nov., *Lactobacillus helsingborgensis* sp. nov. and *Lactobacillus kullabergensis* sp. nov., isolated from the honey stomach of the honeybee *Apis mellifera*. *Int J Syst Evol Microbiol*. 2014;64(Pt 9):3109.
64. Dolezal AG, Carrillo-Tripp J, Judd TM, Allen Miller W, Bonning BC, Toth AL. Interacting stressors matter: diet quality and virus infection in honeybee health. *R Soc Open Sci*. 2019;6:181803.
65. Xu W, Zhang W, Zhang T, Jiang B, Mu W. L-arabinose isomerases: characteristics, modification, and application. *Trends Food Sci Technol*. 2018;78:25–33.
66. Wang C, Huang Y, Li L, Guo J, Wu Z, Deng Y, et al. *Lactobacillus panisapium* sp. nov., from honeybee *Apis cerana* bee bread. *Int J Syst Evol Microbiol*. 2018;68:703–8.
67. Silvia, Brochet Andrew, Quinn Ruben AT, Mars Nicolas, Neuschwander Uwe, Sauer Philipp, Engel. Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife*. 2021;10:e68583.
68. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol*. 2014;12:263–73.
69. Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell*. 2015;160:583–94.
70. Yassour M, Jason E, Hogstrom LJ, Arthur TD, Tripathi S, Siljander H, et al. Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe*. 2018;24:146–54.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

