

RESEARCH

Open Access



Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers

Gilles L. C. Baud¹, Aiswarya Prasad¹, Kirsten M. Ellegaard¹ and Philipp Engel^{1*} 

*Correspondence:
philipp.engel@unil.ch

¹ Department of Fundamental Microbiology, University of Lausanne, CH-1015 Lausanne, Switzerland

Abstract

Background: Strain-level diversity is widespread among bacterial species and can expand the functional potential of natural microbial communities. However, to what extent communities undergo consistent shifts in strain composition in response to environmental/host changes is less well understood.

Results: Here, we used shotgun metagenomics to compare the gut microbiota of two behavioral states of the Western honeybee (*Apis mellifera*), namely nurse and forager bees. While their gut microbiota is composed of the same bacterial species, we detect consistent changes in strain-level composition between nurses and foragers. Single nucleotide variant profiles of predominant bacterial species cluster by behavioral state. Moreover, we identify strain-specific gene content related to nutrient utilization, vitamin biosynthesis, and cell–cell interactions specifically associated with the two behavioral states.

Conclusions: Our findings show that strain-level diversity in host-associated communities can undergo consistent changes in response to host behavioral changes modulating the functional potential of the community.

Keywords: Metagenomics, Gut microbiota, Strain diversity, Honey bee, Symbiosis, Social insects

Background

Microbial communities are highly dynamic. They undergo compositional changes in response to environmental factors, which can profoundly influence their functions. For example, antibiotic treatment can disturb the human gut microbiota, altering its effect on the host metabolism and immune system, and facilitating the invasion of pathogens [1]. Such microbiota shifts are difficult to predict as they often occur irregularly and in different contexts. However, when linked to recurrent or predetermined environmental changes, microbial communities can undergo consistent compositional transitions, providing excellent opportunities to study ecological interactions



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

governing community shifts. For example, bacterioplankton communities in the ocean that depend on sunlight intensity [2–4], or gut microbial communities of animals that have a seasonal diet, or that hibernate during the winter months [5–8], turnover as a function of season. Similarly, the assembly of the human gut microbiota after birth follows a predictable pattern characterized by the succession of different microbial taxa during the first few months of life [9, 10].

Compositional changes associated with the turnover of microbial communities have mostly been studied at the genus- or species-level using 16S rRNA gene analysis. However, natural microbial communities usually harbor a large extent of strain-level diversity, with strains of the same species carrying distinct functions. In the human gut, strain-level differences can influence the metabolism of dietary compounds [11, 12] or drugs [13] or modulate their interactions with the host immune system [14]. Moreover, several recent studies have suggested that strains of the same species can exhibit different dynamics and govern the eco-evolutionary dynamics in microbial communities [15–18]. However, if natural microbial communities undergo consistent compositional transitions at the strain level in response to environmental changes is less well understood.

Honeybees (*Apis mellifera*) can be useful models for studying strain-level dynamics in host-associated microbial communities. They are eusocial animals that live in large colonies composed of a single queen, a few hundred male bees, and thousands of sterile female worker bees. They harbor a relatively simple, but highly specialized microbial community in their gut, composed of < 10 bacterial phylotypes (i.e., bacteria sharing > 97% sequence identity in the 16S rRNA gene) [19, 20]. Most of these phylotypes have diverged into sequence discrete populations (sequence clusters of > 95% ANI; hereafter referred to as species), each harboring a high degree of strain-level diversity and variation in functional gene content [21–26].

Using shotgun metagenomics, we have previously shown that while phylotypes, and species within phylotypes, co-exist in individual bees, strains within species segregate, suggesting that each worker bee harbors a distinct community at the strain level [22]. These differences in strain-level composition could not be attributed to any variation in age, time of sampling, or colony affiliation. However, in our previous study, only a few colonies were sampled, and bees were exclusively sampled from within the hive between September and February (i.e., autumn and winter) raising the possibility that most microbiota samples originated from nurse bees or long-lived winter bees. During the summer months, worker bees undergo an age-related behavioral maturation [27–29]. Young nurse bees feed on pollen and fulfill tasks inside the hive related to brood care, nest building, and food processing. However, when 2–3 weeks old, nurses transition to forager bees, which feed on honey, and collect pollen and nectar outside the hive. Besides the dramatic change in dietary intake, this behavioral maturation is accompanied by a series of physiological changes, such as endocrine secretion, fat body metabolism, and brain activity [30, 31]. Community profiling based on 16S rRNA gene sequencing has shown that the nurse-to-forager transition is associated with a consistent change in gut microbiota composition [32, 33]. While the microbiota of both bee types is dominated by the same bacterial phylotypes, the

composition of the community shifts with a marked decrease in the absolute abundance of most phylotypes in forager versus nurse bees.

Here, we tested if the gut microbiota of honeybees during the transition from nurse to forager bees also undergoes consistent compositional changes at the strain level. To this end, we analyzed shotgun metagenomics data from nurse and forager bees sampled across 15 colonies in Western Switzerland. Our results show that the two behavioral states harbor distinctive bacterial communities across colonies, characterized by consistent shifts in strain composition and functional gene content of major bacterial species. These findings show that strains are ecologically distinct units that undergo a predictable ecological transition when bees turn from nurses to foragers. We speculate that this microbiota shift is driven by nutrient availability disparities or physiological changes in the gut of foragers relative to nurses.

Results

Bacterial loads and species-level community composition differ between nurse and forager bees

To investigate compositional changes with genomic resolution in the honeybee gut microbiota in response to behavioral state, we collected metagenomic samples from larvae-feeding nurses and pollen foragers (Additional file 2: Table S1). Samples consisted of pooled honeybee guts to ensure representative sampling of the strain-level diversity for each colony. In total, we sampled 15 colonies from five different sampling locations (three each), to account for possible variation between sites and colonies, not related to behavioral state.

Due to the different dietary habits of the two behavioral states, the hindguts of nurses were usually pollen-filled, as indicated by the enlarged gut size, and the opacity and yellow color of the gut content. In contrast, the hindguts of foragers were smaller and contained mostly nectar based on their translucent appearance. Accordingly, nurse guts were on average $2.9 \times$ heavier than their forager counterparts ($p = 3.3e - 09$, Wilcoxon signed-rank test, Additional file 1: Fig. S1).

To estimate the absolute abundance of bacteria relative to the amount of total DNA in each pooled gut sample, we performed qPCR on genomic DNA targeting the bacterial 16S rRNA gene. This revealed that the bacterial loads in the forager samples were on average $2.6 \times$ lower than the ones in the nurse samples ($p = 1.83e - 04$, Wilcoxon signed-rank test, Fig. 1A), confirming previous results by Kešnerová et al. [33].

Shotgun metagenomics sequencing resulted in 58.4–76.7 million reads per sample. To compare the bacterial community composition of nurses and foragers, the metagenomic reads were mapped to a reference database of isolate genomes of strains from the gut microbiota of social bees, as in Ellegaard et al. [34] and Ellegaard and Engel [22]. The non-redundant database was composed of 198 isolate genomes and encompassed all major community members described to be present in the honeybee gut microbiota. Between 19.8 and 57.4% of all metagenomic reads mapped to the bacterial reference database. Of the remaining reads, 30.6–76.0% mapped to the host genome of *A. mellifera*, and only 4.2–20.8% of the reads could not be assigned (Additional file 1: Fig. S2A). Forager samples overall had a lower proportion of reads mapping to the bacterial database than the nurse samples (forager average 35.6%, nurse average 50.8%,

reads (0.15–3.1%) in nine samples mapped to bacteria from other bee species (including Lactobacilli isolated from *Apis cerana*, and Gilliamella isolated from bumble bees and *Apis cerana*; see Additional file 2: Table S3), which may be due to spurious mapping.

Non-metric multidimensional scaling (NMDS) using Bray–Curtis dissimilarity of the bacterial relative abundances separated nurse from forager samples at both the phylotype- (PERMANOVA, Host— $R^2=0.40279$, p -value=0.001) and the species-level (PERMANOVA, Host— $R^2=0.45201$, p -value=0.001, Fig. 1C and D). Samples are also clustered by location (Additional file 1: Fig. S4). However, it is important to note that our study was not designed to test for this, as only three colonies were sampled per location with two samples per colony.

In terms of absolute abundances (as determined by multiplying the proportions of each community member by the total bacterial biomass in each sample, see the “Material and methods” section), most community members were less abundant in foragers than in nurse bees (Additional file 1: Fig. S5), a direct consequence of the lower total bacterial biomass present in the gut of foragers (Fig. 1A). However, there were clear differences in the strength of these changes between community members, even among species of the same phylotype (Fig. 1E). For example, the total abundances of Lactobacilli Firm5_1 and Firm5_4 changed to a lesser extent (average log₂ fold change = -1.45 and -0.71) than those of Firm5_2 and Firm5_3 (average log₂ fold change = -2.82 and -2.99) in foragers relative to nurses. This explains the consistent shift towards a less even species composition of this phylotype in foragers compared to nurses as determined with the Shannon diversity index (Wilcoxon signed-rank test, p -value = $8.8e-06$, Additional file 1: Fig. S3). Likewise, Bifido_1.2 was more affected (average log₂ fold change = -2.32) than any of the other species of the *Bifidobacterium* phylotype (average log₂ fold change across species of the *Bifidobacterium* phylotype = -0.37) by the behavioral state transition of the host (Additional file 1: Fig. S3). Notably, four community members (Bapis, Snod_1, Com_1 and Lkun) did not change in absolute bacterial load between nurses and forager samples, suggesting that they are not affected in their abundance by the behavioral transition.

In summary, although the gut microbiota of forager and nurse bees were composed of the same community members, we found consistent differences in the absolute abundance of many of them, not only at the phylotype level as previously shown [32, 33], but also at the species level. Moreover, some species of a given phylotype were more strongly affected than others suggesting species-specific responses to the nurse-to-forager transition.

Consistent differences in strain-level composition between nurse and forager bees

Previous studies have shown that most honeybee gut microbiota species harbor high levels of strain diversity and that the strain composition differs between individual bees [22, 34]. To assess strain-level differences between nurses or foragers, we mapped the metagenomic reads against a reduced reference database containing one genome per species and quantified the number of polymorphic sites in the reads mapped onto the single-copy core genes of these reference genomes.

We detected between 1.85 and 11.19% polymorphic sites per species per sample, with Firm4_2 showing the lowest, and Firm5_4 the highest average proportion of

polymorphic sites per sample (Fig. 2A). These differences in proportions of polymorphic sites among community members are comparable to those previously found by Ellegaard et al. [34] and Ellegaard and Engel [22]. To assess the total amount of strain-level diversity detected per species across all samples, we calculated cumulative proportions of polymorphic sites as a function of the number of bee samples. Firm4_2 showed the lowest (4.91%) and Bifido1_1 the highest (12.8%) level of total diversity across all samples (Fig. 2B). In all cases, we observed that the slope of the curves rapidly decreased with the addition of new samples, indicating that many of the polymorphic sites are shared across samples.

When comparing the two behavioral states, we found that five of the 17 analyzed species showed higher strain-level diversity in nurses than in foragers (Wilcoxon signed-rank test, Firm4_1: q -value = $4.20e-3$, Firm5_1: q -value = $6.45e-3$, Firm5_2: q -value = $1.08e-2$, Firm5_3: q -value = $3.56e-4$, Gilli_1: q -value = $6.45e-3$) whereas only two species showed lower strain-level diversity in nurses than in foragers (Wilcoxon signed-rank test, Gilli_2: q -value = $2.91e-2$, Snod_1: q -value = $2.91e-2$, Fig. 2A). The

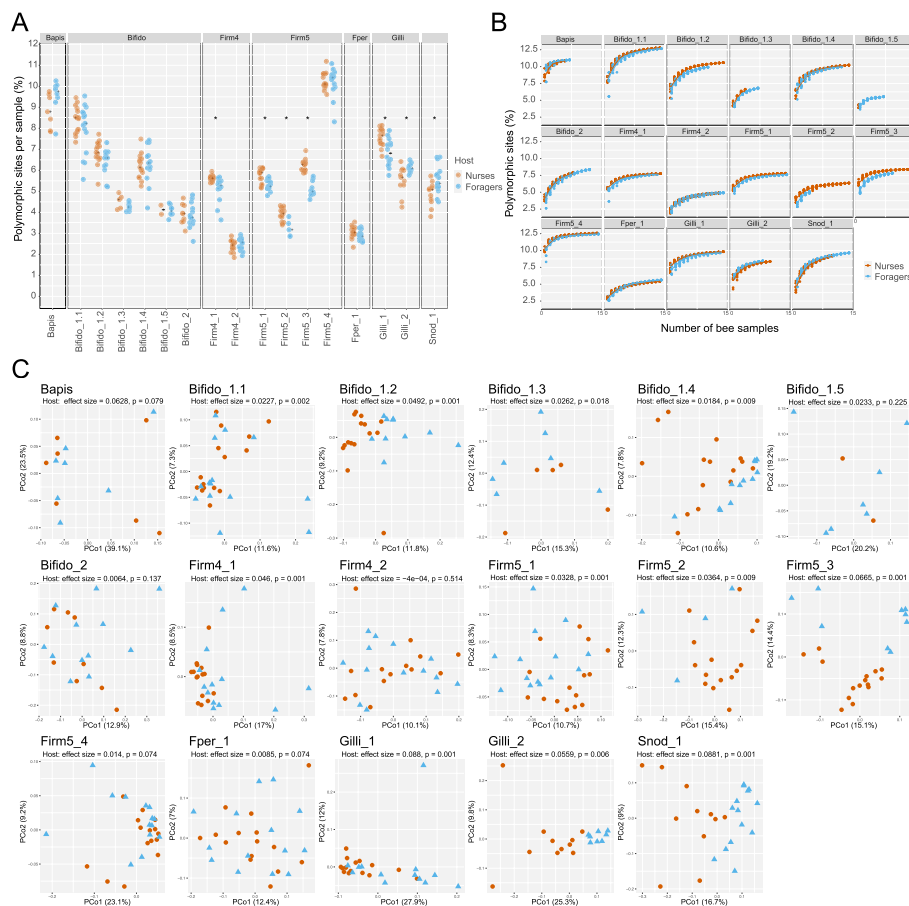


Fig. 2 Differences in strain-level diversity between nurse and forager bees. **A** Proportion of polymorphic sites in the single-copy core genes for each bacterial species, expressed as a percentage of the total single-copy core genome length. Asterisks symbolize a q -value < 0.05 for the Wilcoxon signed-rank test between the proportion of polymorphic sites between nurse and forager samples. **B** Curves of the cumulative proportions of polymorphic sites for each species. **C** Principal coordinate analysis plots based on the Jaccard distance shared polymorphic sites between each pair of samples for each of the species

forager samples had on average 2.6 times lower bacterial load than nurse samples, but on average only 1.5 times smaller number of reads per sample. This means that a larger fraction of the total bacterial population was analyzed for forager relative to nurse samples, a difference that could potentially affect diversity estimates. However, no significant correlation was found between the proportion of polymorphic sites detected and the read coverage of our samples (Additional file 1: Fig. S6), suggesting that we had sufficiently deep coverage across the study.

Next, we assessed the difference in strain profiles between forager and nurse samples by comparing the distribution of single nucleotide variants (i.e., alleles) at the detected polymorphic sites for each species across samples. To do so, we computed the Jaccard distances between all sample pairs based on the number of shared polymorphic sites. Principal coordinate analysis of these similarity indices resulted in clustering of the samples by host behavioral state for 11 of the 17 analyzed species (PERMANOVA, p -value < 0.05, Fig. 2C), including all species of Bifido_1 and Firm5, except for Bifido_1.5 and Firm5_4, Firm4_1, Gilli_1, Gilli_2, and Snod_1. Notably, we also observed clustering by location in the principal coordinate analysis (PERMANOVA, p -value < 0.05, Additional file 1: Fig. S7). However, as already stated before, our study was not designed to test for the effect of location as two samples per location came from the same colony.

In summary, our analysis shows that nurse bees and foragers have distinct strain profiles for many bacterial species both in terms of the amount of diversity and type of variants detected.

Intraspecific differences in functional gene content between nurse and forager bees

Previous studies have shown that most species of the honeybee gut microbiota harbor large accessory gene pools and that individual bees vary in terms of the functional gene content encoded in a given species [22]. Because nurses and foragers have different physiology and diet, we thus hypothesized that the observed changes in strain-level composition between the two behavioral states may be driven by functional differences among strains.

To identify strain-specific genes that may be enriched in one behavioral state or the other, we compared the read coverage of all genes assigned to a given species. Strain-specific genes should have a lower read coverage than the single-copy core genes of a given species, and those that predominate in one behavioral state relative to the other should show a consistent coverage discrepancy between forager and nurse samples. To identify such genes, we assembled all reads not mapping to the host genome and binned the resulting contigs into bacterial species bins using our reference database (see the “[Material and methods](#)” section for details and Additional file 1: Fig. S8). Then, we clustered the ORFs encoded in the binned contigs into species-specific ortholog groups (OGs) and identified those OGs that showed consistent coverage discrepancy between forager and nurse samples. In total, 81.7% (65.3–90.9% per sample) of all assembled bases could be binned and 83.4% (73.2–86.1% per sample) of the non-host reads mapped to the binned contigs indicating that we had included most of the bacterial diversity present in the samples. Out of 45,863 species-specific OGs, 1551 revealed significant differences between the nurse and forager samples (q -value < 0.05, Wilcoxon signed-rank test, and absolute log₂ fold change > 1, Additional file 2: Table S4), which indicated that these

genes were more abundant in one host behavioral state versus the other. Notably, the number of differential abundant OGs substantially varied between species ranging from 15 to 231 (Fig. 3A, Additional file 2: Table S4). For example, in Bifido_1.1, we detected very few consistent gene content differences between foragers and nurses (6 and 9 OGs, respectively). This species had the highest proportion of polymorphic sites across samples, but showed no clustering of the samples by behavioral state based on the number of shared polymorphic sites (Fig. 2C). The largest differences in gene content between nurses and foragers were found for three *Lactobacillus* Firm5 species (Firm5_2: 184 OGs, Firm5_3: 131 OGs, Firm5_4: 177 OGs), two *Gilliamella* species (Gilli_1: 64 OGs, Gilli_2: 161 OGs), Snod_1 (197 OGs), and Bifido_1.2 (159 OGs). Consistent with these findings, all of these species, except for Firm5_4, also showed clustering by behavioral state based on the number of shared polymorphic sites (Fig. 2C). Interestingly, a larger fraction of the state-specific OGs were found in foragers than in nurses (in total 1019 versus 445 significantly different OGs, respectively).

To determine the functions encoded by the behavioral state-specific gene content, we annotated all OGs using eggnoG and used COG categories as a first level of functional characterization. Of all behavioral state-specific OGs, 33.1% had no COG annotation, while 17.3% of them fell into the COG category S (unknown function). For the rest of the significant OGs, we categorized 20.6% of them as metabolism-related; 11.7% as linked to cellular processes, signaling, and structure; and 18.3% as information storage and processing (Fig. 3B). Among the significant categorized OGs (excluding COG S), category G (carbohydrate metabolism and transport) was the most important category, representing 11.4% of all significant OGs. However, the proportion of different COG categories contributing to the significant OGs varied between species (Fig. 3B). In case of Firm4_1, COG category H (coenzyme metabolism) was specifically enriched in the behavioral state-specific OGs relative to the total number of OGs per species (Fisher’s exact test q -value = $2.5e - 06$, Additional file 2: Table S5).

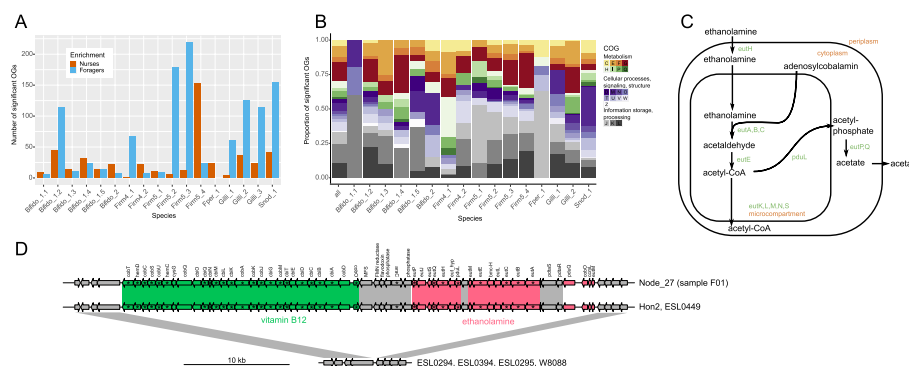


Fig. 3 Differences in functional gene content within species of the gut microbiota between nurse and forager bees. **A** Number of orthologous groups (OGs) significantly differing in coverage for each of the species. **B** Repartition of the COG categories of the significant OGs, nurse-enriched and forager-enriched included. OGs that were not categorized or categorized as COG S were not included in the plot. **C** Figure adapted from [37]. Possible ethanolamine utilization pathway enabled by the operon shown in **D**. **D** Genomic island containing the ethanolamine utilization and vitamin B12 biosynthesis genes, and its presence in the isolate genomes of the bacterial genome database. Genes enriched in the forager samples are in bold and flagged with an asterisk

Behavioral state-specific OGs of COG category G included many phosphotransferase system genes, in particular in *Lactobacillus* and *Gilliamella* species, but also glycoside hydrolases, sugar-binding, and metabolism proteins. In most species, these OGs were more abundant in foragers relative to nurses, except for Bifido_1.2 and Firm5_3, where the enrichment was in the nurse samples, or spread equally between nurse and forager samples (Additional file 2: Table S4).

The overrepresentation of the COG category H in Firm4_1 corresponded to 27 OGs involved in the biosynthesis of cobalamin (vitamin B12) and adenosylcobalamin (coenzyme B12), which had a higher coverage in forager than in nurse samples (see Additional file 2: Table S4). In the same species, we also found 18 genes for the utilization of ethanolamine among the OGs enriched in foragers versus nurses. Processing of ethanolamine by the enzyme ammonia-lyase is dependent on adenosylcobalamin (Fig. 3C). To further explore the possible link between the identified cobalamin biosynthesis and ethanolamine utilization OGs, we analyzed the available isolate genomes of Firm4. Homologs of the cobalamin biosynthesis and ethanol utilization pathways were only present in two (Hon2 and ESL0449) of the six Firm4_1 strains (Fig. 3D) confirming that these genes are strain-specific. Moreover, the genes of both pathways were encoded in the same genomic island, integrated into a conserved chromosomal region, providing further evidence that they are functionally linked and have likely evolved together.

Another interesting set of behavioral state-specific OGs was identified in Gilli_1, Gilli_2, and Snod_1, and encoded various components of Type VI secretion systems (T6SSs) involved in interbacterial killing. This included genes annotated as effector proteins with PAAR or RHS domains and immunity proteins. In Gilli_2, several genes encoding the structural components of an entire T6SS were among the behavioral state-specific OGs. They clustered in a genomic island which was present in only one (strain A7) out of 17 strains of Gilli_2 in our reference database (Additional file 1: Fig. S9). Strikingly, all these behavioral state-specific T6SS-related OGs were more abundant in forager than in nurse samples.

In Snod_1, we also found other surface structure-related functions than T6SS among the behavioral state-specific OGs, including genes encoding major subunits of Type IV pili, RTX toxins, and proteins involved in capsular polysaccharide biosynthesis (Additional file 2: Table S4).

Discussion

There is accumulating evidence that natural microbial communities harbor a high extent of strain-level diversity [38–41]. How such fine-scale diversity varies across space and time in nature is a key question in microbial ecology and evolution [15, 42, 43].

Here, we find that the strain composition of the honeybee gut microbiota undergoes a consistent ecological turnover when bees transition from nurses to foragers. The single nucleotide variant profiles of several community members clustered by behavioral state. Moreover, for most community members, we found consistent differences in intraspecific gene content between nurses and foragers, which included functions related to nutrient utilization, vitamin biosynthesis, and cell–cell interactions.

In the human gut microbiota, strain-level variation in individual community members has been linked to diet [44] or health state [43, 45–47]. Our results show that a

radical shift in host lifestyle can change the strain composition of a large fraction of the gut microbiota, demonstrating that strains represent ecologically distinct units and emphasizing the importance of studying diversity in host-microbe associations at the strain-level.

Previous shotgun metagenomics studies have linked variation in species- and strain-level diversity in the gut microbiota to host traits, such as age, colony origin, or genotype, but not to behavioral state [22, 34, 48–50]. Some of these studies analyzed individual bees while others focused on pooled gut samples. In the current work, we analyzed pooled samples of 20 bees rather than individual bees. The advantage of pooling is that a larger number of colonies can be surveyed, increasing the statistical power of the study. However, pooling precludes the detection of variants present at low prevalence across the sampled bees explaining the overall lower levels of polymorphic sites observed in our samples as compared to previous studies [22, 34]. However, rare strains have likely a minor impact on the whole community or the host.

Another possible limitation of our study is that the sampled nurse and forager bees were not age-matched. As foragers are usually older than nurses, age could have been a confounding factor in our study. However, our previous findings have shown that the age of bees has little impact on the strain level composition of the gut microbiota [22]. Therefore, we are confident that most of the observed differences are due to the nurse to forager transition and the associated changes in dietary intake, physicochemical gut conditions, or host physiology.

In contrast to nurses, foragers do not consume pollen. This is likely the reason why foragers were found to harbor lower bacterial biomass in the gut than nurses and exhibit a consistent decrease of *Bombilactobacillus*, *Lactobacillus*, and *Bifidobacterium* relative to the other phylotypes. This was already reported by two previous studies based on 16S rRNA gene analysis [32, 33], one of which used age-matched forager and nurse bees, corroborating the results of our shotgun metagenomics analyses. Pollen constitutes an important nutrient source for bee gut-associated *Bombilactobacillus*, *Lactobacillus*, and *Bifidobacterium* explaining their decrease in absolute abundance. All three phylotypes harbor genes to degrade pollen-derived carbohydrates [21, 26, 51, 52], and they readily grow in minimal media supplement with pollen extract or entire pollen grains [53, 54]. Moreover, experimentally colonized bees kept on a sugar water diet harbor substantially fewer bacteria of these phylotypes in their gut than bees that also receive pollen [33, 53].

As previous studies were based on 16S rRNA gene sequencing, their analyses were limited to the phylotype level. Hence, a key finding of our study is that not only phylotypes, but also closely related species within phylotypes, and even different strains of the same species respond differently to the nurse-to-forager transition. For example, one species of *Lactobacillus* Firm5 was similar in total abundance between foragers and nurses, while others significantly differed (> 8 times). Likewise, some species showed consistent changes in strain composition between nurses and foragers, while others did not. This suggests that some species and strains are more successful than others in thriving in the gut environment of one state versus the other. The reasons underlying these differences can be manifold. Some species and strains may be adapted to utilize specific pollen compounds and hence have an advantage in the nurse gut. Others may not depend on pollen nutrients for growth that much or may be better at withstanding physicochemical

changes in the gut as bees transition from one state to the other, explaining why they persist in the forager gut. Also, the fact that we find changes in strain-level diversity in opposite directions between the two host states suggests that different ecological processes are at play in different species. For example, in one species, the reduced nutrient availability in the forager gut may hinder strains to overgrow and dominate the community, resulting in decreased total abundance but increased strain-level diversity in foragers. For another species, the nutrient-rich pollen diet may expand ecological niches or open new ones that can be colonized by functionally divergent strains explaining the observed increase in strain-level diversity in nurses relative to foragers.

If the observed shifts in strain composition were due to changes in the ecological niches in the gut, we would expect to find functional differences between strains of the same species dominating in nurses versus foragers. Previous studies have successfully identified such condition-specific genes within host-associated bacterial species by comparing metagenomic datasets [17, 38, 44]. For example, strains of the human gut symbiont *Prevotella copri* harbored distinct gene repertoires for drug metabolism and complex carbohydrate degradation in Western and non-Western individuals [44]. Likewise, strains of an intracellular symbiont of *Bathymodiolus* deep-sea mussels showed substantial variation in gene content for nutrient and energy acquisition indicating adaptation to the local conditions present in the environment they were sampled from [38].

We detected strain-specific gene sets (i.e., OGs) associated with one of the two behavioral states in almost all species of the gut microbiota. Carbohydrate transport and metabolism functions were prevalent among the behavioral state-specific gene content, which was somehow expected as most species carry large accessory gene sets of this functional category in their genomes [22, 51, 52]. Moreover, previous studies have shown that carbohydrate transport and metabolism are important for niche partitioning of diet-derived carbohydrates in *Lactobacillus* and *Bifidobacterium* and that many of these genes are differentially expressed when bees are fed pollen as compared to sugar water as diet [26, 53].

Somewhat more surprising was the finding that vitamin B12 biosynthesis and ethanolamine utilization genes of *Bombilactobacillus* were specifically enriched in foragers versus nurses. Some of these genes have already been found to be variably present among individual bees in our previous study [22], but they could not be linked to any host differences. Vitamin B12 is an important co-factor that cannot be synthesized by honeybees (or other animals) [55], but needs to be taken up from the diet or synthesized by bacterial symbionts [56]. It is possible that *Bombilactobacillus* strains synthesizing this vitamin compensate for the lack of vitamin B12 in the pollen-deprived diet of foragers. However, a more plausible explanation for the increased abundance of these genes in foragers may be that vitamin B12 is an essential co-factor for a major enzyme of the ethanolamine utilization pathway [55]. Ethanolamine is derived from the membrane phospholipid phosphatidylethanolamine and constitutes an important bacterial nitrogen and carbon source in the animal gut due to the high turnover of the host epithelium and the microbiota [57]. It is possible that ethanolamine accumulates in the forager gut due to the presence of dead bacterial material or increased epithelial shedding, providing a behavioral state-specific nutrient niche for the gut microbiota. The fact that the genes for utilizing ethanolamine are strain-specific indicates that they are costly to maintain,

and that host state-specific nutritional niches may contribute to the maintenance of strain-level diversity in the gut microbiota.

We also found genes mediating bacteria-host interactions and interbacterial killing among the behavioral state-specific gene content. Of particular interest was the finding of several T6SS genes associated with foragers in two of the three *Gilliamella* species and in *Snodgrassella*. The primary niche of these bacteria is the ileum, where they form multispecies biofilms on the host epithelium [58]. Diverse T6SS gene repertoires have been identified in almost all sequenced strains of *Gilliamella* and *Snodgrassella* indicating that they play key roles in interbacterial competition and niche establishment for these bacteria [59, 60]. Our findings suggest that the forager gut environment selects strains with expanded capabilities in interbacterial warfare, possibly as a consequence of the increased competition for nutrients and space. In *Snodgrassella*, we also found genes encoding other surface structures (Type IV pili, LPS modifying enzymes, RTX toxins) to be enriched in either nurses or foragers. Some of these genes are known colonization factors of *Snodgrassella* [61], likely because they enable host attachment and biofilm formation. However, why these functions are specifically enriched in one of the two behavioral states remains to be determined.

Conclusions

In summary, using shotgun metagenomics analysis, we show that the gut microbiota of honeybees undergoes consistent shifts in strain-level diversity when bees transition between two major behavioral states (i.e., from nurses to foragers). These strain-level differences are associated with changes in intraspecific gene content of several bacterial species suggesting adaptation to host state-specific niches. The possibility of carrying out gnotobiotic bee experiments provides excellent opportunities for future research to probe the functional relevance of the link between strain-specific gene content and host state. Since the association of social bees with their microbiota is evolutionarily ancient, it will also be interesting to expand these analyses to closely related bee species which show similar division of labor and whose microbiota has diversified to similar extents as in *A. mellifera*.

Material and methods

Sampling

Adult female worker bees of *A. mellifera* were collected over the Summer 2019 from 15 different colonies in 5 different locations across Western Switzerland (Additional file 2: Table S1). For each sampled hive, we collected 20 larvae-feeding nurses and 20 pollen foragers, in order to encompass most of the strain-level diversity within each colony. Samplings were performed between 7 and 9 AM, when worker bees were beginning to return from their foraging trips. Larvae-feeding nurses were identified as bees putting their heads in larvae-containing cells. Pollen foragers were identified as bees returning to their hive with pollen pellets on their corbiculae. The bees were anesthetized using CO₂ and put on ice. They were then dissected in sterile 1 × PBS to extract the entire hindgut (pylorus, ileum, rectum) with tools sterilized in 70% ethanol. Gut content was visually inspected to be consistent with the life stage, i.e., nurse guts had to be pollen-filled and forager guts had to be nectar-filled. Bees for which the gut content could not be verified

were excluded. Finally, the guts were pooled by hive of origin and host phenotype (nurse and forager, respectively) (Additional file 2: Table S1).

DNA extraction

DNA extraction from the pooled gut samples was performed using the Qiagen All-prep Powerfecal DNA/RNA extraction kit (Qiagen) following the kit handbook except for the bead-beating steps. After dissection, the pooled guts were transferred into the kit's bead-beating tubes. Lysis buffer and DDT were added following protocol instructions, and samples underwent two bead-beating steps of 4 s at 6 m/s on a FastPrep-24 5G bead-beater (MP Biomedicals). Gut samples with a total weight over 220 mg were diluted to reach a concentration of 220 mg in 650 μ l lysis buffer + 25 μ l DDT; then, the excess volume was removed (i.e., for a gut weight of 440 mg, 337.5 μ l was removed from the sample, and an identical volume of lysis buffer + 3.8% DDT 1 M was added to it). Subsequently, all samples underwent a third bead-beating step for 45 s at 6 m/s. All the remaining steps were performed following the manufacturer's protocol. DNA was diluted in 2 \times 30 μ l Buffer EB, its concentration and quality were assessed with Qubit and Nanodrop, and samples were stored at -20°C until further processing.

Quantitative real-time PCR

We assessed the bacterial loads of the DNA samples using quantitative real-time PCR (qPCR) following the methods of Ellegaard et al. [34]. We targeted both the V3-V4 region of the 16S rRNA bacterial gene and the actin gene of *A. mellifera* using the primers published by Kešnerová et al. [54]. We performed each qPCR reaction using a reaction of 1 μ l extracted DNA (10 ng/ μ l), 3.2 μ l nuclease-free water, 5 μ l SYBR Select Master Mix (Life Technologies), 0.2 μ l forward primers (5 μ M), and 0.2 μ l reverse primers (5 μ M) in 96-well plates (MicroAmp Fast 96-well Reaction Plate, Applied Biosystems for Life Technologies). The reactions were performed in triplicates on a QuantStudio 5 Real-Time PCR System (Thermo Fischer Scientific) using the parameters from Kešnerová et al. [54].

The data was extracted using the QuantStudio Design & Analysis Software v1.4 (Applied Biosystems) and exported in csv format.

We computed the absolute number of each gene copy using the formula:

$$\text{Gene copy number} = 10^{\frac{C_t - \text{Intercept}}{\text{Slope}}}$$

Intercept and slope parameters were previously determined by our group [54].

In order to compare the microbial gut loads of the different samples, we normalized the number of 16S rRNA gene copies by the total DNA yield per sample. Normalization by the number of host actin gene copies gave similar results. However, foragers had more actin gene copies than nurses. Therefore, we did not use this type of normalization in the final analysis (see Additional file 1: Fig. S10). The figures were generated using R (version 4.0) and the ggplot2 library.

Shotgun metagenomics sequencing

The libraries were prepared with the Illumina Nextera Flex library kit (Illumina) with unique dual indices (UDI) and sequenced on a HiSeq 4000 instrument (PE150) at the

Genomic Technologies Facility (GTF) of the University of Lausanne. All samples were multiplexed together and sequenced on three lanes. The quality of the reads was assessed with FastQC (v0.11.4, Babraham Institute) and subsequently trimmed and filtered to remove low-quality sequences and short reads using the tool Trimmomatic v0.35 [62] with settings PE ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 TRAILING:20 SLIDING-WINDOW:4:20 MINLEN:50.

Reference genome database

The reference genome database containing bacterial isolate genomes was constructed in the same way as in Ellegaard and Engel [22] and Ellegaard et al. [34]. A total of 198 high-quality genomes of strains isolated from the gut of honeybees and other social bee species such as bumble bees and other *Apis* species were included in the database. Strains having more than 98.5% pairwise ANI with other already included strains were not added to the database to limit redundancy. The pairwise genomic ANI was computed using fastANI [63]. In general, multiple genomes were included for each bacterial phylotype as listed in Additional file 2: Table S2. Each genome includes a fasta file of nucleotide sequences of all contigs (.fna), nucleotide sequences (.ffn), and amino acid sequences (.faa) of predicted genes and concatenated (.fna) files where the contigs were concatenated into one sequence for each genome. Bed files were made to summarize the updated positions of genes in the concatenated files which were used for further analysis. We predicted gene families of homologous genes using OrthoFinder (v2.3.5, [64]) and identified all single-copy core genes, i.e., genes present in one copy in all genomes of a given phylotype. The resulting files listing each single-copy core orthologous group and the constituent genes of each genome of that phylotype are included in the reference genome database along with reduced bed files for each genome indicating the location of these core genes to be used for further analysis. The total length of these single-copy core genes relative to the total genome length of a given species is given in Additional file 1: Fig. S11.

Community composition analysis

The community composition analysis was based on the work of Ellegaard and Engel [22]. We mapped the reads using bwa mem (v0.7.15-r1142-dirty) [65] against the reference genome database and used the samtools (v1.9) view -F 0 × 800 command [66] to remove chimeric alignments. Reads mapping to multiple locations were not included as bwa chooses one of the many alignments as primary and only includes those in the output. Alignments with 50 matches or fewer were removed using a custom perl script that parses alignment length from the CIGAR string. The read coverage information was extracted for each gene using the command samtools bedcov and summed up across all orthologs of each single-copy core gene family for genomes coming from the same species. Since genes located close to the origin of replication had higher coverage than genes located close to the terminus (as an indication of replicating bacteria [67]), the coverage at the terminus was inferred. To this end, the coverage of each gene family was plotted relative to its genomic position using one selected reference genome per species and a segmented regression line was fitted. The relative abundance of each species in each sample was determined by dividing the terminus coverage by the summed

terminus coverage of all species, respectively. For the species Bifido_1.1, Bifido_1.2, and Bifido_1.4, the segmented lines did not show a proper fit (terminus location inferred far from estimated breakpoint, or coverage at *Ori* lower than *Ter*). In these cases, the median coverage of all single-copy core gene families was used as an estimate for the relative abundance of that species.

Strain-level analysis

To quantify strain-level diversity within species, we followed the approach established in Ellegaard and Engel [22]. In short, we mapped the metagenomic reads using `bwa mem` to a reduced version of the reference genome database, including only one representative member per species (see Additional file 2: Table S2). Then, we used `samtools -F 0 × 800` to remove chimeric alignments. Reads mapped with less than 50 matches registered in their CIGAR were also removed, as well as those whose number of mismatches was higher than 5 to reduce errors in variant calling. Only species with more than 20 × terminus coverage were included. Genes with less than 10 × coverage as estimated in the community profiling pipeline were excluded. To perform variant calling, we used the tool `FreeBayes (v1.3.1-19-g54bf409)` [68] with options `--haplotype-length 0, --min-alternate-fraction 0.1, --min-coverage 10, -C 5, -l, -X, -u` and `--pooled-continuous`. We used the tools `vcfbreakmulti` and `vcfallelicprimitives` from the `vcflib` library [69] to split complex observations into multiple lines. The bed files containing gene positions of single-copy core genes of each genome were subset to obtain the corresponding locations to be considered for each species. These filtered bed files were used to focus downstream analysis on the phylotype-core region for each species.

We computed the fraction of polymorphic sites for each bacterial species using the number of identified sites divided by the cumulative length of all single-copy core genes. An allele was considered detected if there was more than one read supporting the allele and the allele was at a frequency of at least 0.01. A position was considered polymorphic if more than one allele was detected within the metagenomic reads of that sample mapped to the reference regardless of what allele was in the reference genome. We also computed the cumulative number of polymorphic sites across samples for each species using 10 random orders of samples each time.

To assess the similarity between samples in terms of strain composition, we used the Jaccard index based on the proportion of shared polymorphic sites, i.e., sites that were polymorphic in at least one of the samples in each pair and where each sample had an allele and an alternate allele shared and at a frequency above the detection threshold (0.01) were considered shared between the two samples. So, for example, if the reference genome has the allele T at a given position, but sample 1 has the alleles C and G and sample 2 has the alleles T and C, then the two samples would share one allele (C) and they would have two non-shared alleles. This site would not be counted as shared. On the other hand, if sample 1 had the alleles C and T the polymorphic site would be considered shared with sample 2. This was implemented in a custom python script as cases where the site was polymorphic in at least one sample in the pair were scored and shared positions were those with an intersection > 1 in the set of detected alleles in the sample pair. The distance between all sample pairs was computed as $Distance(S1, S2) = 1 - Shared\ fraction(S1, S2)$ where shared fraction for a given pair of

samples was calculated as the number of shared polymorphic sites divided by the number of sites polymorphic in at least one of the two samples. The matrix of pairwise distances between the samples was plotted using principal coordinate analysis to identify samples that share similar polymorphic sites. We computed the principal coordinates using the ape R library (v5.6–2).

Functional gene content composition

For the functional profiling of the gene content of the microbiota, we first extracted all the reads which mapped to the bacterial database, plus those that did not map to the host genome (*A. mellifera* genome (PRJNA471592, version Amel_Hav3.1), using the picard tools (v2.7.1-SNAPSHOT) SamToFastq with parameters VALIDATION_STRATEGY=SILENT [70]. We merged those reads for each sample and assembled them using SPAdes [71] with parameters --meta and -t 20. The obtained contigs were filtered to keep only the ones with a length of at least 500 bp and a k-mer coverage of at least 1.

We predicted the open-reading frame (ORF) positions using Prodigal (version 2.6.3, Hyatt et al.) [72] with parameters -m -p meta. The ORFs with the flag partial=00 or length smaller than 300 bp were filtered out. We annotated the remaining ORFs using eggNOG (v1.0.3–40-g41a8498) [73, 74] with the script emapper.py and the options --annotate_hits_table --no_file_comments --cpu 20.

To assign the contigs to the major bacterial species of the bee gut microbiota, we blasted the predicted ORFs against a nucleotide database of all ORFs from our bacterial database using the blastn tool (version 2.2.31+) [75] (evalue < 1e-05, pident > 50 and qcovs > 50). Contigs were assigned to a given species when at least 80% of the ORFs were mapped against genes of that species, as well as when at least 10% of the ORFs were mapped against genes of that species and no mapping was found for any other species. We filtered out all non-unanimous contigs (15,198 contigs out of a total of 257,015 (5.9%), representing 13.6% of the bp).

For each species, we used the amino acid sequences from the ORFs and the genes from the database strains to compute orthologous groups (OG) using OrthoFinder [64]. We created a database of all ORF's nucleotide sequences and mapped the reads against it using bwa mem (version 0.7.15-r1142-dirty) [65]. We removed chimeric alignments using samtools view option -F 0 × 800 (version 1.9) [66]. Alignments with less than 50 matches were also removed. For each sample, we extracted the coverage using the samtools coverage command (v1.15) and summed the coverage for the ORFs belonging to the same OG. We normalized the coverage information for every sample and every species using the OGs containing single-copy core genes: for each species of each sample, the summed coverage of all single-copy core OGs was set to the same number (10000). To determine which OGs were underrepresented or overrepresented in either the nurses or the foragers, we computed the paired Wilcoxon signed rank test for each OG based on the normalized coverage values and then used *q*-values to correct for multiple comparisons. We also computed the fold-change by dividing the mean forager normalized coverage value by the nurse counterpart. To be considered as significantly different between the treatments, an ORF had to have both an absolute log₂ fold change of at least 1, and a *q*-value < 0.05. We finally annotated the ORFs belonging to the significant OGs using the

tool eggno-mapper (1.0.3–40-g41a8498) [73, 74]. The whole analysis workflow is summarized in Additional file 1: Fig. S8.

Statistics and data visualization

We performed the statistical analyses in R (v4.2.1) using the libraries ape (v5.6–2), data.table (v1.14.8), and vegan (v2.6–4) and in python (v3.10.11) using the package scipy (v1.11.1). We used the package qvalue (v2.28.0) to adjust *p*-values to *q*-values where stated. Shannon indices were calculated using the vegan package. For PERMANOVA, we used the function adonis2 from the vegan package using the model “Distance ~ Host + Location” with 999 permutations and set to estimate the marginal effects of the terms. R^2 values were adjusted using the function adonisOmegaSq implemented in the MicEco package (<https://github.com/Russel88/MicEco>, commit 91f8e6f).

We generated all plots using R and the following R libraries: ggplot2 (v3.3.6), reshape (v0.8.9), gridExtra (v2.3), ggrepel (v0.9.1), RColorBrewer (v1.1–3), and scales (v1.2.1) and in python (v3.10.11) using the package matplotlib (v3.7.1) and seaborn (v0.12.2). Figures have been generated using Inkscape (v1.1) or Adobe Illustrator. A comprehensive list of all packages used and their version is available in the GitHub repository (https://github.com/Aiswarya-prasad/Publication_Baud_metaG_NvsF_2023) [76].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-03131-4>.

Additional file 1: Fig. S1. Per sample average gut mass, computed as each sample's gut pool mass, divided by the number of guts in the pool. **Fig. S2.** (A) Stacked bar plot of the proportions of reads mapping to the bacterial genome database and to the *A. mellifera* genome, as well as non-mapped reads, for each of the samples. (B) Taxonomic profiling of the unmapped reads at the species-level with mOTU2. Colors correspond to genera. Percentages below each distribution show fraction of unmapped reads in each sample (also see panel A). **Fig. S3.** Stacked bar plots of the proportions of the different species and corresponding Shannon diversity index across nurse and forager samples for the four multi-species phylotypes (A) Bifidobacterium, (B) Lactobacillus, (C) Bombilactobacillus, and (D) Gilliamella. **Fig S4.** NMDS plot of the samples' bacterial phylotype composition (A) and species composition (B). Samples are colored according to sampling location. PERMANOVA results: Phylotype-level, Location – $R^2=0.40279$, p -value=0.001; Species-level, Location – $R^2=0.40642$, p -value=0.001. **Fig. S5.** Absolute abundance of the different species detected in the samples approximated by the normalized estimated number of 16S rRNA gene copies. Asterisks indicate q -values < 0.05. **Fig. S6.** Pearson correlation coefficient and non-corrected p -value for the correlation of the percentage of polymorphic sites per species with the coverage of the species normalized by the bacterial load for each sample. **Fig. S7.** Principal Coordinate Analyses of the Jaccard distance between all samples for the different species based on proportions of shared polymorphic sites. The color scheme shows information on the sampling sites of the different samples. The connecting lines indicates the two samples from the same colony. **Fig. S8** Workflow used for the functional gene content analysis. **Fig. S9.** Genomic island of Gilli_2 encoding T6SS genes enriched in forager versus nurse samples. The genomic island is only present in strain A7 of Gilli_2 in our genome database and is integrated in a conserved genomic region. Genes in green indicate T6SS genes, genes in pink depict strain-specific genes belonging to the same genomic island, and genes in grey are other other genes present in the flanking regions of genomes of Gilli_2. Genes enriched in forager samples are labeled in bold and flagged with an asterisk. **Fig. S10** (A) Actin gene copies in 10ng of each sample. (B) 16S rRNA gene copy number per actin gene copy per sample. (C) rRNA gene copy number per actin gene copy per ng of total DNA. The normalization with actin gene copies amplifies the differences between the two groups, as there are fewer copies in foragers than in nurses. **Fig. S11.** Cumulative length of all single-copy core genes (Core genome length, in green) relative to the total genome length of the reference strain (Total genome length, in blue) of each species. Reference strain is the strain used in the reduced reference database. The filtered core genome length represents the portion of the core genes which were used for the SNV analysis as they had >10x read coverage across all samples.

Additional file 2: Table S1. Sample collection information. **Table S2.** Table of all strain genomes included in the reference database, their Species/SDP abbreviation used in this study, their corresponding full species name (if existing) and the host where they were isolated. Genomes used as representative in the reduced database are marked by 'y' in the last column. **Table S3.** Proportions of reads mapping to species isolated in non-*A. mellifera* hosts. The proportions are computed as the number of reads mapping to the species, divided by the number of reads mapping to the same phylotype for this sample. **Table S4.** Table of differentially abundant orthogroups showing which behavioral state they were enriched in along with the associated fold-change, p -values, their COG category and functional

annotation. **Table S5.** Matrix of the counts of the OGs annotated as H category significantly enriched in the forager samples compared to the al OGs and all COG categories.

Additional file 3. Review history.

Acknowledgements

We would like to thank Mam Malick Sy Ndiaye, Florent Mazel, Garance Sarton-Lohéac, and Shinichi Sunagawa for providing helpful comments on the manuscript. We would like to thank Julien Marquis and the team from the Lausanne Genomics Technology Facility for carrying out the Illumina sequencing. We would like to thank Benjamin Dainat, Jan Dudda, Olivier Emery, and Chloé Peillon for letting us sample bees from their apiaries.

Review history

The review history is available as Additional file 3.

Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

P.E. and G.B. were involved in conceptualization; G.B. and A.P. carried out the formal analysis, investigation, and data curation. P.E. was responsible for funding acquisition, supervision, and providing resources. The methodology was developed by K.E., G.B. and A.P. Visualization and validation were carried out by G.B., P.E., A.P., and K.E. The original manuscript was written by G.B. and P.E.; then, P.E., A.P., and K.E. contributed to reviewing and editing. All authors approved the final manuscript.

Funding

Open access funding provided by University of Lausanne This work was funded by the University of Lausanne, a Mathilde Agassiz Scholarship awarded to A.P., an ERC Starting Grant (MicroBeeOme) and a Swiss National Science Foundation SPIRIT grant (grant no. IZSTZ0_189496) both awarded to P.E., and the NCCR Microbiomes, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant no. 180575).

Availability of data and materials

The sequencing data is available under the NCBI BioProject ID PRJNA904667 [77]. The honeybee gut bacterial strain genome database is available for download on Zenodo: <https://zenodo.org/records/10182034> [78]. The scripts required to run the analyses and produce the figures are available on GitHub: https://github.com/Aiswarya-prasad/Publication_Baud_metaG_NvsF_2023 [76] and Zenodo: <https://zenodo.org/records/10183178> [79]. Both the repositories are released under the MIT license.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 12 January 2023 Accepted: 28 November 2023

Published online: 08 December 2023

References

- Keeney KM, Yurist-Doutsch S, Arrieta MC, Finlay BB. Effects of antibiotics on human microbiota and subsequent disease. *Annu Rev Microbiol.* 2014;68:217–35.
- Martin-Platero AM, Cleary B, Kauffman K, Preheim SP, McGillicuddy DJ, Alm EJ, Polz MF. High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nat Commun.* 2018;9:266.
- Teeling H, Fuchs BM, Bennis CM, Kruger K, Chafee M, Kappelmann L, Reintjes G, Waldmann J, Quast C, Glockner FO, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife.* 2016;5:e11888.
- Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennis CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science.* 2012;336:608–11.
- Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, Knight R, Manjurano A, Chagalucha J, Elias JE, et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science.* 2017;357:802–6.
- Sommer F, Stahlman M, Ilkayeva O, Arnemo JM, Kindberg J, Josefsson J, Newgard CB, Frobert O, Backhed F. The gut microbiota modulates energy metabolism in the hibernating brown bear *Ursus arctos*. *Cell Rep.* 2016;14:1655–61.
- Maurice CF, Knowles SC, Ladau J, Pollard KS, Fenton A, Pedersen AB, Turnbaugh PJ. Marked seasonal variation in the wild mouse gut microbiota. *ISME J.* 2015;9:2423–34.
- Davenport ER, Mizrahi-Man O, Michelini K, Barreiro LB, Ober C, Gilad Y. Seasonal variation in human gut microbiome composition. *PLoS ONE.* 2014;9:e90731.

9. Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, Rakoff-Nahoum S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*. 2021;591:633–8.
10. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe*. 2015;17:690–703.
11. Patnode ML, Guruge JL, Castillo JJ, Couture GA, Lombard V, Terrapon N, Henrissat B, Lebrilla CB, Gordon JL. Strain-level functional variation in the human gut microbiota based on bacterial binding to artificial food particles. *Cell Host Microbe*. 2021;29:664–673 e665.
12. Barboza M, Sela DA, Pirim C, Locascio RG, Freeman SL, German JB, Mills DA, Lebrilla CB. Glycoprofiling bifidobacterial consumption of galacto-oligosaccharides by mass spectrometry reveals strain-specific, preferential consumption of glycans. *Appl Environ Microbiol*. 2009;75:7319–25.
13. Haiser HJ, Gootenberg DB, Chatman K, Sirasani G, Balskus EP, Turnbaugh PJ. Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science*. 2013;341:295–8.
14. Bron PA, van Baarlen P, Kleerebezem M. Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa. *Nat Rev Microbiol*. 2011;10:66–78.
15. Goyal A, Bittleston LS, Leventhal GE, Lu L, Cordero OX. Interactions between strains govern the eco-evolutionary dynamics of microbial communities. *Elife*. 2022;11:e74987.
16. Leventhal GE, Boix C, Kuechler U, Enke TN, Sliwerska E, Holliger C, Cordero OX. Strain-level diversity drives alternative community types in millimetre-scale granular biofilms. *Nat Microbiol*. 2018;3:1295–303.
17. Lou YC, Olm MR, Diamond S, Crits-Christoph A, Firek BA, Baker R, Morowitz MJ, Banfield JF. Infant gut strain persistence is associated with maternal origin, phylogeny, and traits including surface adhesion and iron acquisition. *Cell Rep Med*. 2021;2:100393.
18. Wolff R, Shoemaker W, Garud N. Ecological stability emerges at the level of strains in the human gut microbiome. *MBio*. 2023;14:e0250222.
19. Bonilla-Rosso G, Engel P. Functional roles and metabolic niches in the honey bee gut microbiota. *Curr Opin Microbiol*. 2018;43:69–76.
20. Kwong WK, Moran NA. Gut microbial communities of social bees. *Nat Rev Microbiol*. 2016;14:374–84.
21. Ellegaard KM, Brochet S, Bonilla-Rosso G, Emery O, Glover N, Hadadi N, Jaron KS, van der Meer JR, Robinson-Rechavi M, Senthil V, et al. Genomic changes underlying host specialization in the bee gut symbiont *Lactobacillus Firm5*. *Mol Ecol*. 2019;28:2224–37.
22. Ellegaard KM, Engel P. Genomic diversity landscape of the honey bee gut microbiota. *Nat Commun*. 2019;10:446.
23. Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vasquez A. Extensive intra-phylo-type diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics*. 2015;16:284.
24. Segers FH, Kesnerova 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;11:1232–44.
25. Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, Moran NA. Metabolism of toxic sugars by strains of the bee gut symbiont *Gilliamella apicola*. *MBio*. 2016;7:e01326–e1416.
26. Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, Tringe SG, Moran NA. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc Natl Acad Sci U S A*. 2019;116:25909–16.
27. Brodschneider R, Crailsheim K. Nutrition and health in honey bees. *Apidologie*. 2010;41:278–94.
28. Seeley TD. *The Wisdom of the Hive: the social physiology of honey bee colonies*. Cambridge, Massachusetts London, England: Harvard University Press; 1996.
29. Smith CR, Toth AL, Suarez AV, Robinson GE. Genetic and genomic analyses of the division of labour in insect societies. *Nat Rev Genet*. 2008;9:735–48.
30. Robinson GE. Genomics and integrative analyses of division of labor in honeybee colonies. *Am Nat*. 2002;160(Suppl 6):S160–172.
31. Whitfield CW, Ben-Shahar Y, Brillet C, Leoncini I, Crauser D, Leconte Y, Rodriguez-Zas S, Robinson GE. Genomic dissection of behavioral maturation in the honey bee. *Proc Natl Acad Sci U S A*. 2006;103:16068–75.
32. Jones JC, Fruciano C, Marchant J, Hildebrand F, Forslund S, Bork P, Engel P, Hughes WOH. The gut microbiome is associated with behavioural task in honey bees. *Insectes Soc*. 2018;65:419–29.
33. Kesnerova L, Emery O, Troilo M, Liberti J, Erkosar B, Engel P. Gut microbiota structure differs between honeybees in winter and summer. *ISME J*. 2020;14:801–14.
34. 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;30:2520–2531 e2527.
35. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, Hingamp P, Alves R, Costea PI, Coelho LP, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun*. 2019;10:1014.
36. Kwong WK, Medina LA, Koch H, Sing KW, Soh EJ, Ascher JS, Jaffe R, Moran NA. Dynamic microbiome evolution in social bees. *Sci Adv*. 2017;3:e1600513.
37. Garsin DA. Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nat Rev Microbiol*. 2010;8:290–5.
38. Ansoorge R, Romano S, Sayavedra L, Porras MAG, Kupczok A, Tegetmeyer HE, Dubilier N, Petersen J. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat Microbiol*. 2019;4:2487–97.
39. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwiientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J*. 2016;10:1589–601.
40. Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, Armanini F, Manghi P, Bonham K, Zolfo M, et al. The *Prevotella copri* complex comprises four distinct clades underrepresented in westernized populations. *Cell Host Microbe*. 2019;26:666–679 e667.
41. Zhu A, Sunagawa S, Mende DR, Bork P. Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol*. 2015;16:82.
42. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. *Nat Rev Microbiol*. 2020;18:491–506.

43. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J, Wijmenga C, Zhernakova A, Segal E. Structural variation in the gut microbiome associates with host health. *Nature*. 2019;568:43–8.
44. De Filippis F, Pasolli E, Tett A, Tarallo S, Naccarati A, De Angelis M, Neviani E, Cocolin L, Gobbetti M, Segata N, Ercolini D. Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets. *Cell Host Microbe*. 2019;25:444–453 e443.
45. Carrow HC, Batachari LE, Chu H. Strain diversity in the microbiome: lessons from *Bacteroides fragilis*. *PLoS Pathog*. 2020;16:e1009056.
46. Chen L, Wang D, Garmaeva S, Kurilshikov A, Vich Vila A, Gacesa R, Sinha T, Lifelines Cohort S, Segal E, Weersma RK, et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell*. 2021;184:2302–2315 e2312.
47. Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell*. 2015;160:583–94.
48. Wu J, Lang H, Mu X, Zhang Z, Su Q, Hu X, Zheng H. Honey bee genetics shape the strain-level structure of gut microbiota in social transmission. *Microbiome*. 2021;9:225.
49. Li C, Tang M, Li X, Zhou X. Community dynamics in structure and function of honey bee gut bacteria in response to winter dietary shift. *MBio*. 2022;13:e0113122.
50. Wu Y, Zheng Y, Wang S, Chen Y, Tao J, Chen Y, Chen G, Zhao H, Wang K, Dong K, et al. Genetic divergence and functional convergence of gut bacteria between the Eastern honey bee *Apis cerana* and the Western honey bee *Apis mellifera*. *J Adv Res*. 2022;37:19–31.
51. Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci U S A*. 2012;109:11002–7.
52. Li Y, Leonard SP, Powell JE, Moran NA. Species divergence in gut-restricted bacteria of social bees. *Proc Natl Acad Sci U S A*. 2022;119:e2115013119.
53. Brochet S, Quinn A, Mars RAT, Neuschwander N, Sauer U, Engel P. Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *Elife*. 2021;10:e68583.
54. Kesnerova L, Mars RAT, Ellegaard KM, Troilo M, Sauer U, Engel P. Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biol*. 2017;15:e2003467.
55. Kaval KG, Garsin DA. Ethanolamine Utilization in Bacteria. *MBio*. 2018;9:e00066–e118.
56. Elsayeh WA, Cook C, Wright GA. B-vitamins influence the consumption of macronutrients in honey bees. *Front Sustain Food Syst*. 2022;6:804002.
57. Tsoy O, Ravcheev D, Mushagian A. Comparative genomics of ethanolamine utilization. *J Bacteriol*. 2009;191:7157–64.
58. Martinson VG, Moy J, Moran NA. Establishment of characteristic gut bacteria during development of the honey-bee worker. *Appl Environ Microbiol*. 2012;78:2830–40.
59. Steele MI, Kwong WK, Whiteley M, Moran NA. Diversification of type VI secretion system toxins reveals ancient antagonism among bee gut microbes. *MBio*. 2017;8:10–128.
60. Steele MI, Moran NA. Evolution of interbacterial antagonism in bee gut microbiota reflects host and symbiont diversification. *mSystems*. 2021;6:10–128.
61. Powell JE, Leonard SP, Kwong WK, Engel P, Moran NA. Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. *Proc Natl Acad Sci U S A*. 2016;113:13887–92.
62. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
63. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9:5114.
64. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 2015;16:157.
65. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S: The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
67. Bremer H, Churchward G. An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J Theor Biol*. 1977;69:645–54.
68. Garrison EM, G. Haplotype-based variant detection from short-read sequencing. *arXiv*. 2012.
69. Garrison EK, Z. N.; Dawson, E. T.; Pedersen, B. S.; Prins, P.: Vcfliib and tools for processing the VCF variant call format. *bioRxiv*. 2021.
70. Picard Tools. Broad Institute. [<http://broadinstitute.github.io/picard/>].
71. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34.
72. 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.
73. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol*. 2017;34:2115–22.
74. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47:D309–14.
75. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
76. Baud G, Prasad A, Ellegaard K, Engel P. Code repository: Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers. *GitHub*. 2023. https://github.com/Aiswarya-prasad/Publication_Baud_meta_NvsF_2023.

77. Baud G, Prasad A, Ellegaard K, Engel P. Comparison of gut microbiota of honeybee nurse and forager workers. Datasets. Sequence Read Archive. 2023. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA904667>.
78. Baud G, Prasad A, Ellegaard K, Engel P. Genome Database: Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers. Zenodo. 2023. <https://zenodo.org/records/10182034>.
79. Baud G, Prasad A, Ellegaard K, Engel P. Code repository: Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers. Zenodo. 2023. <https://zenodo.org/records/10183288>.

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

