Italian Journal of Animal Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tjas20

Effect of host genetics on gut microbiota composition in an italian honeybee breeding population

M. G. De Iorio, G. Pagnacco, F. Tiezzi & G. Minozzi

To cite this article: M. G. De Iorio, G. Pagnacco, F. Tiezzi & G. Minozzi (2024) Effect of host genetics on gut microbiota composition in an italian honeybee breeding population, Italian Journal of Animal Science, 23:1, 1732-1740, DOI: 10.1080/1828051X.2024.2424404

To link to this article: https://doi.org/10.1080/1828051X.2024.2424404

ITALIAN JOURNAL OF ANIMAL SCIENCE
2024, VOL. 23, NO. 1, 1732–1740
https://doi.org/10.1080/1828051X.2024.2424404

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Effect of host genetics on gut microbiota composition in an italian honeybee breeding population

M. G. De Iorio[a], G. Pagnacco[b], F. Tiezzi[c] and G. Minozzi[a]

[a]Dipartimento di Medicina Veterinaria e Scienze Animali, Università degli Studi di Milano, Lodi, Italy; [b]Consiglio Nazionale delle Ricerche (CNR), Istituto di Biologia e Biotecnologia Agraria (CNR-BBA), Milano, Italy; [c]Dipartimento di Scienze e Tecnologie Agrarie, Alimentari, Ambientali e Forestali, Università degli Studi di Firenze, Firenze, Italy

ABSTRACT

Honeybees host a specialised gut microbiota composed of five ever-present bacteria, which undergo seasonal variation in their proportions. The objective of this study was to investigate how host genetics can affect the microbial composition of the honeybee gut over a 5-month sampling period. Seventy-seven colonies from eight maternal genetic lines were sampled from a breeding population located in Lombardy, Italy. The colonies have been selected since 2015 for three different traits, using isolated mating stations for reproduction. Worker bees were sampled three times: June, July and October 2021. From each colony, gut from ten worker bees were pooled, and the V3-V4 region of the 16S rRNA gene was amplified and sequenced. Statistical analyses were conducted in the R environment to assess the impact of genetic lines, timepoints, and their interaction on α-diversity and β-diversity. Results for α-diversity showed a significant effect of genetic line on the Simpson index, as well as a consistent effect of the timepoint. Specifically, genetic line H showed a lower Simpson index, and this line also produced less honey in the same years, suggesting a genetic influence on both microbiota richness and honey yield. For β-diversity, results revealed that the month of sampling had the strongest effect, while no significant differences were identified for the genetic line or their interaction.

ARTICLE HISTORY
Received 15 July 2024
Revised 27 October 2024
Accepted 28 October 2024

KEYWORDS
Apis mellifera; gut microbiota; seasonal variation; genetics; breeding population

Introduction

Honeybees harbour a specialised gut microbiota primarily located in the distal part of the gastrointestinal tract (ileum and rectum), with fewer bacteria present in the proximal area (crop and midgut), which are mainly environmental species (Martinson et al. 2012). Differences between queens and workers and between adult bees and larvae are present, due to caste-specific factors and distinct diets (Kapheim et al. 2015; Tarpy et al. 2015).

However, all adult workers host a core microbiota, composed of 5 ubiquitous bacterial genera that remain consistently present despite environmental variations: *Bifidobacterium*, *Lactobacillus*, *Bombilactobacillus* (formerly *Lactobacillus Firm-4*), *Gilliamella*, and *Snodgrassella* (Kwong and Moran 2016). Additionally, bacteria such as *Bartonella apis*, *Apibacter adventoris*, *Frischella perrara*, and *Acetobacteraceae* are often found in the gut of many worker honeybees (Kwong and Moran 2016).

Various studies conducted worldwide in different climatic areas agree on a significant seasonal shift in honeybee gut microbiota composition, with a decreased level of α-diversity in winter compared to summer (Ludvigsen et al. 2015; Subotic et al. 2019; Kešnerová et al. 2020; Bleau et al. 2020; Almeida et al. 2023). Furthermore, as in mammals, the gut microbiota of bees is socially transmitted and serves vital symbiotic functions, including complementing host nutrition, facilitating dietary breakdown, and supporting colony health and resilience against pathogens (Kwong and Moran 2016; Engel and Moran 2013).

Given the pivotal role of microbiota in maintaining bee well-being, it is increasingly important to study the intricate relationship between honeybees and microbiota to mitigate the current global decline in bee populations (Le Conte and Navajas 2008; Le Conte et al. 2010; Henry et al. 2012). This decline is alarming due to the indispensable contribution of

CONTACT G. Minozzi giulietta.minozzi@unimi.it

Supplemental data for this article can be accessed online at https://doi.org/10.1080/1828051X.2024.2424404.

bees to the preservation of ecosystems and biodiversity, serving as primary pollinators: 80% of the main food crops depend on their pollination work (Klein et al. 2007). Consequently, their survival is increasingly dependent on beekeeping practices and the adoption of selective breeding schemes is becoming essential (Le Conte and Navajas 2008; Le Conte et al. 2010; Henry et al. 2012).

Host genetics has been found to influence gut microbiota composition in humans, explaining 1.9-8.1% of the variation (Kurilshikov et al. 2021; Lopera-Maya et al. 2022). Similarly, studies in bees have observed the influence of genetics, such as when comparing *Apis mellifera* with *Apis cerana* or when comparing different subspecies of *Apis mellifera* (Ellegaard et al. 2020; Yang et al. 2021; Wu et al. 2021). Studies have also shown that individuals from more genetically diverse colonies had more diverse gut microbiota (Mattila et al. 2012; Bridson et al. 2022). Moreover, worker bees belonging to the same colony exhibit a higher similarity in their microbiota, with colony membership explaining 41% of the observed variation in the bacterial community among samples (Bridson et al. 2022).

The purpose of this study was therefore to investigate whether and how genetic factors can influence the variation in microbiota composition of selected bee lines over a 5-month period.

Materials and methods

Data collection

Each year since 2015, 108 colonies are phenotyped to estimate their breeding value (EBV), for three different traits: docility, honey production, and hygienic behaviour. From the best queen, selected on a combined index of the three traits, a group of 12 sisters is obtained through grafting, a technique where larvae are transferred from their original cells into artificial queen cups to be reared as queens. These sisters are called drone-producing queens and serve as "fathers". The other best 6 queens serve as “mothers” to produce, by grafting, 18 virgin queens each. Finally, the 108 virgin queens are mated in isolated mating stations with the drone-producing queens to produce the new 108 colonies. In this study, colonies from the described breeding population were sampled three times between June and October 2021, with ten worker bees collected per colony and stored at −80 °C in 50 ml empty Falcon tubes.

Despite the initial breeding population consisting of 108 colonies, only 77 colonies survived to the first sampling in June 2021, as colonies with queen loss or replacement were excluded from the selection program. Over the course of the five months of sampling, there were further discrepancies in data collection or instances (e.g. the loss of the queen during the month of sampling) led to variations. Consequently, the total number of samples was 190, distributed as follows: 75 colonies in June (timepoint 1), 69 in July (timepoint 2) and 46 in October (timepoint 3). The queens of these 77 colonies all have the same “sire” but descend from 8 different dams or 8 different maternal genetic lines. Table 1 shows the distribution of the 77 colonies among the 8 maternal genetic lines at each timepoint.

DNA extraction, library preparation and sequencing

Gut samples from ten worker bees per colony were pooled together, and DNA extraction was performed using the ‘Blood & Tissue Genomic DNA Extraction kit’ from Fisher Molecular Biology. Subsequently, the V3-V4 regions of the 16S rRNA gene were amplified using the Illumina ‘16S Metagenomic Sequencing Library Preparation’ kit following the manufacturer’s instructions. The primers used for amplification were "16S Amplicon PCR Forward Primer" (sequence:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG) and "16S Amplicon PCR Reverse Primer" (sequence: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC). Sequencing was conducted on an Illumina Novaseq sequencer, generating paired-end 250–bp reads, with an initial target depth of 700,000 reads per sample.

Bioinformatic analysis

The sequencing data were processed using the Divisive Amplicon Denoising Algorithm 2 pipeline from the R package ‘DADA2’ (Callahan et al. 2016), with default settings, except the forward reads were trimmed at 245 bp and the reverse reads at 240 bp. Taxonomy assignments were performed using the SILVA reference database

Table 1. Distribution of the sampled colonies among the 8 maternal genetic lines at each timepoint.

Genetic line	Timepoint 1	Timepoint 2	Timepoint 3
A	12	11	11
B	11	10	9
C	10	10	5
D	8	9	7
E	7	6	3
F	11	11	4
G	6	4	2
H	10	8	5

(v132) with a confidence cut-off of 2. A core microbiota was constructed, including only the amplicon sequence variants (ASVs) present in at least 50% of the samples. This core was normalised through rarefaction to a total sum of 10,000 reads per sample for accurate comparison. Rarefaction curves were generated using the 'Vegan' package (Oksanen et al. 2015) to verify the representativeness of bacterial communities within the core rarefied microbiome.

The α-diversity indexes and stress value were estimated using the 'Phyloseq' and 'Vegan' packages (Oksanen et al. 2015; McMurdie and Holmes 2013) and were normalised using z-score transformation with the "scale" function in R. The study examined the potential correlation between gut microbiota and genetic lines across different sampling months. The six normalised α-diversity indexes (Shannon, Observed ASV, Chao1, Simpson, ACE and Fisher) were treated as dependent variables, while timepoint, genetic lines and a composite variable representing their combination were used as independent variables. Additionally, following the Shapiro-Wilk test results that indicated the variables were not normally distributed, separate Kruskal-Wallis tests were performed to evaluate the effect of timepoint on α-diversity within each genetic line and the effect of genetic lines on α-diversity across the three timepoints.

Following significant findings from Kruskal-Wallis tests, post-hoc pairwise comparisons were conducted using the Dunn method to identify which genetic lines and timepoint-genetic line combinations exhibited differentiation in α-diversity indexes. Kruskal-Wallis tests were conducted utilising the 'kruskal.test' function in the R software environment (R Core Team). The post-hoc tests were performed using the 'dunn.test' function from the 'dunn.test' package in R.

Additionally, permutational multivariate analysis of variance (PERMANOVA) was employed, utilising the 'adonis2' function within the Vegan package (Oksanen et al. 2015). In this analysis, the β-diversity of microbiota compositions, calculated using the Bray-Curtis distance metric, was considered the dependent variable, while timepoint, genetic lines and their interaction as independent variables.

Moreover, in the same year of sampling, the honey production of the colonies was measured in kilograms, combining two distinct honey yields: the first from Acacia obtained in May 2021 and the second from wildflowers produced in July 2021. To calculate the honey yield, the upper part of each hive was weighed before being placed on the hive, and then weighed at the end of the production period. The honey yield was determined by the difference in the two weights.

ANOVA was conducted to investigate the association between genetic lines and honey production for both the first harvest and the total harvest, as these variables were found to be normally distributed in the Shaprio-Wilk test. In contrast, since the second harvest variable was not normally distributed, the Kruskal Wallis test was used. In these analyses, genetic lines were considered the independent variable, while the honey production variables served as the dependent variables.

Results

From the total ASVs identified, only those showing non-zero counts in at least 50% of the samples were selected to define the core microbiome. This approach aims to focus on the most representative taxa with significant ecological functions, resulting in a reduction of the total number of ASVs identified across the entire dataset of 190 samples and at each timepoint. Further details are provided in Table 2.

Moreover, as the number of reads may be due to discrepancies in library sizes or sequencing efficiency (Kim 2023), rarefaction was performed. This process ensured that the total sum of reads in each sample was standardised to 10,000.

Seasonal variation and genetic host association

In all genetic lines and at all timepoints, the most prevalent bacterial genera were *Lactobacillus*, *Bifidobacterium*, and *Bombilactobacillus*, followed by *Gilliamella* and *Snodgrasella*. Additionally, although less abundant, *Frischella* and *Commensalibacter* were consistently identified. Bacteria belonging to the *Apilactobacillus* genera and *Rhizobiaceae* family (unknown genera) were only identified in timepoints 1 and 2 across all genetic lines, but not in timepoint 3. Similarly, the genera *Lelliotia* and *Escherichia-Shigella* were found only in timepoint 1 and timepoint 2 but not across all genetic lines. *Enterobacter*, while present in all timepoints, was not consistently found across all genetic lines. Further details on the relative abundance of these taxa at both family and genus levels are provided in Supplementary Tables 1, 2 and 3 and illustrated in Figures 1 and 2.

Table 2. Number of amplicon sequence variant (ASV) before (total ASV) and after (core ASV) the selection for the identification of the core ASV, in each dataset.

Dataset	Total ASVs	Core ASVs
190 samples	17,823	2,036
Timepoint 1	8,232	2,148
Timepoint 2	8,788	2,225
Timepoint 3	7,926	1,912

To explore differences in microbiota α-diversity between genetic lines and timepoints, Kruskal-Wallis was conducted with timepoint, genetic line, and a composite variable representing their combination as independent variables. The 6 different α-diversity indexes were examined as dependent variables (Supplementary Table 4). The results revealed a significant effect of the genetic line on the Simpson index and a significant effect of timepoints in all the 6 measures. The compositive variable of timepoint and genetic lines showed significant difference for the Shannon, Simpson and ACE indexes.

First, a post-hoc comparison was performed to discern which genetic lines exhibited significant differentiation in Simpson indexes.

The results of the Dunn test using the entire dataset showed significant difference between genetic line A and genetic line H for the Simpson index with a p-value of 0.0054 (Table 3). This difference in diversity for genetic line H is clearly visible in Figure 3a, which shows the Simpson indexes for all genetic lines, highlighting the lower diversity exhibited by genetic line H.

Post-hoc analysis was conducted to explore the significant differences in the Shannon, Simpson, and ACE indexes among the groups of the compositive variable. No significant differences were observed for the Shannon index. For the Simpson index, a significant difference was observed between genetic line A at timepoint 1 and genetic line H at timepoint 3 (p-value = 0.0315). This aligns with our previous findings indicating that the greatest differences occur between genetic lines A and H. For the ACE index, the only significant difference was found between timepoint 2 and timepoint 3 of the same genetic line, H (p-value = 0.01825). This suggests that the difference in ACE is primarily driven by the timepoint effect rather than the combined genetic line and timepoint effect.

Given the significant differences observed in the overall dataset for the genetic lines in the Simpson index, a Kruskal-Wallis test was conducted to examine the effect of the genetic line at each timepoint. This implied defining a dataset for each time point, which was used for the analysis. A significant impact of the genetic lines was identified at timepoint 2 ($p = 0.038$),

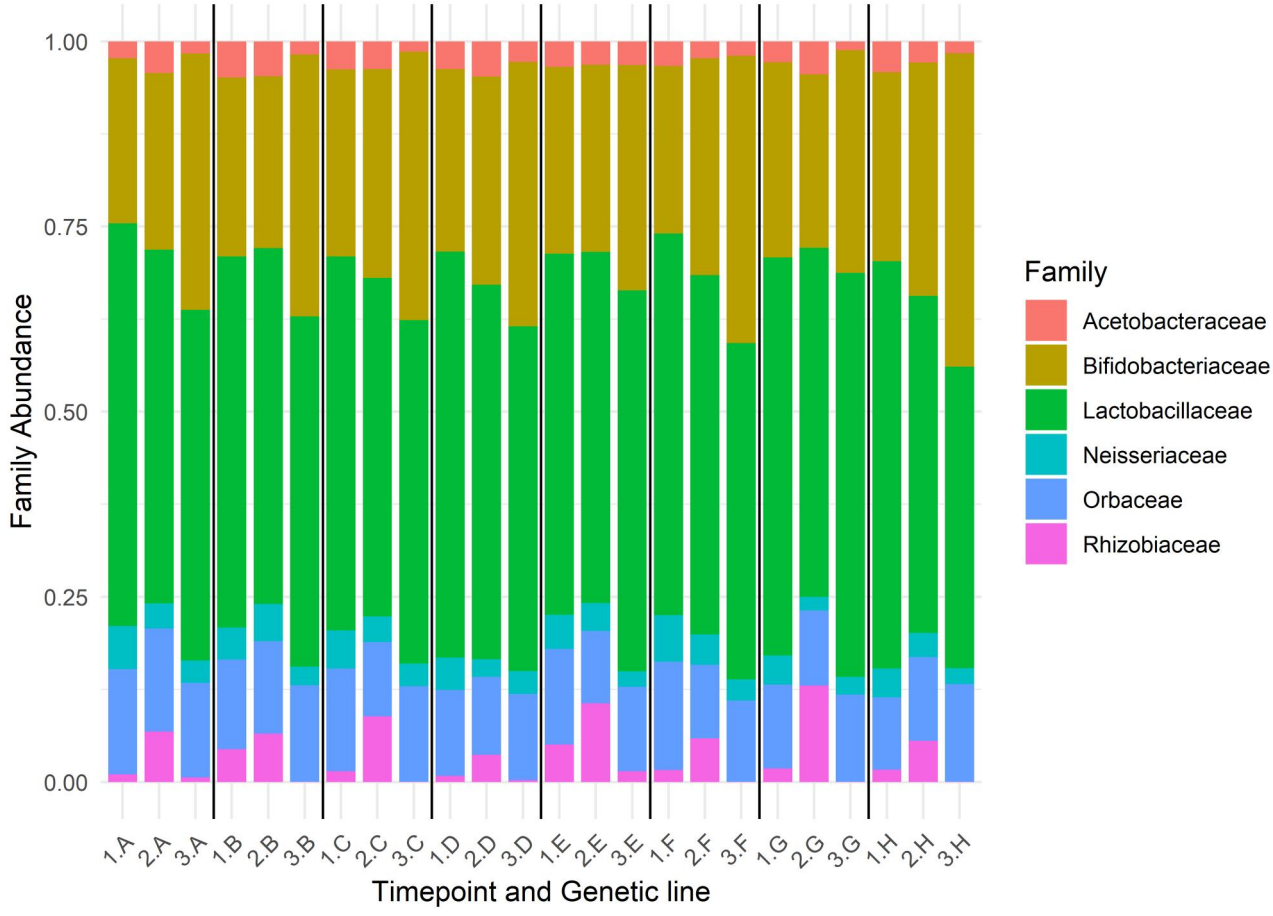

Figure 1. Bar chart of the taxa showing the relative abundance of bacterial taxa at the family level across samples grouped by timepoint (1, 2 and 3) and genetic lines (A, B, C, D, E, F, G, H).

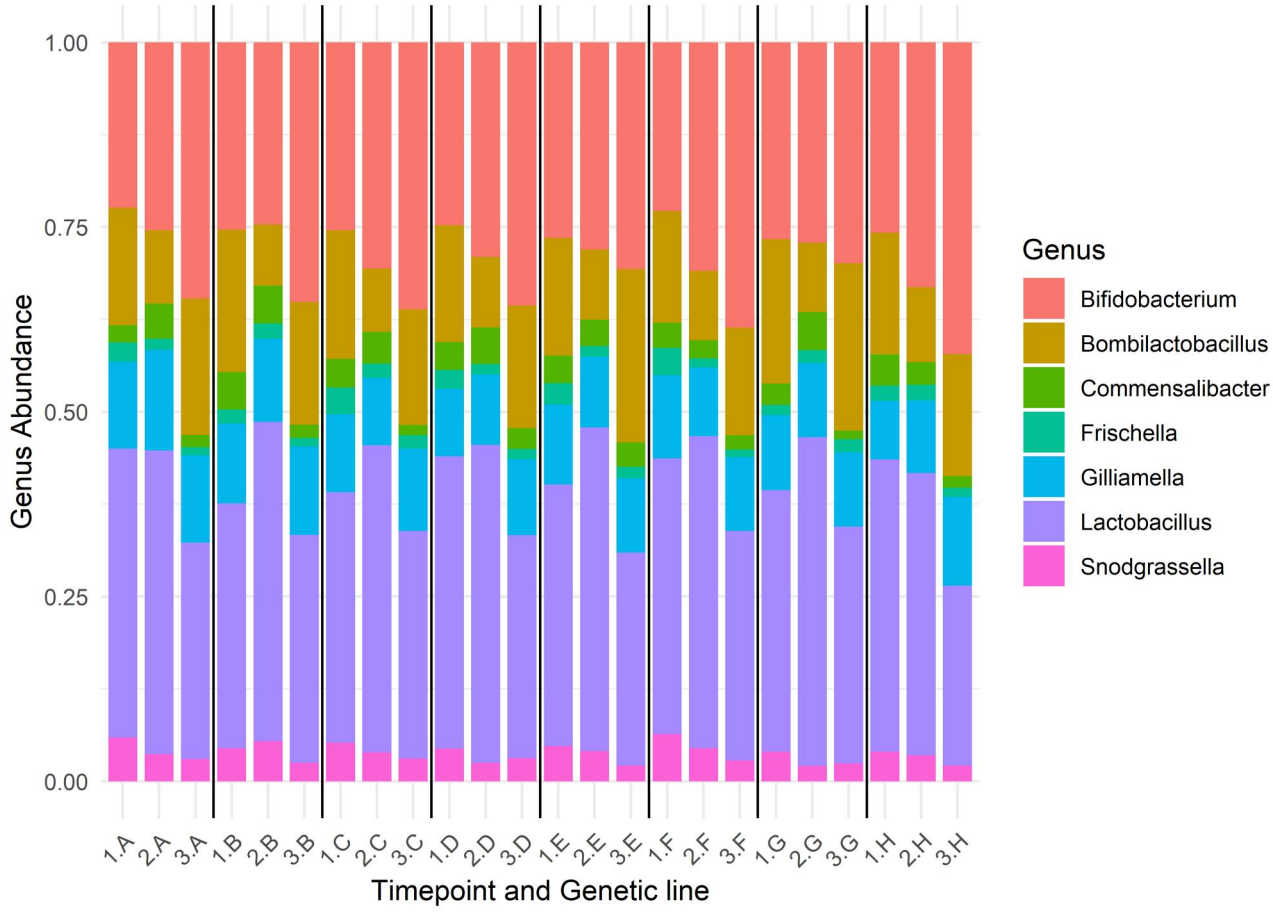

Figure 2. Bar chart of the taxa showing the relative abundance of bacterial taxa at the genus level across samples grouped by timepoint (1, 2 and 3) and genetic lines (A, B, C, D, E, F, G, H).

Table 3. Results of dunn test for pairwise comparison of Simpson index among genetic lines using the entire dataset of 190 samples. The table shows the test statistic values for each comparison. Comparisons with statistically significant differences ($p < 0.05$) are shown in bold and with "*".

Genetic line	A	B	C	D	E	F	G	H
A	–							
B	0.85	–						
C	1.38	0.55	–					
D	1.42	0.6	0.05	–				
E	0.31	−0.39	−0.84	−0.88	–			
F	1.9	1.05	0.47	0.41	1.27	–		
G	0.9	0.26	−0.17	−0.22	0.54	−0.56	–	
H	**3.55***	2.68	2.06	1.99	2.66	1.61	1.84	–

as shown in Table 4. However, the post-hoc comparisons did not reveal any significant pairwise differences.

Additionally, since the overall dataset showed a significant effect of timepoint across all alpha diversity indices, Kruskal-Wallis tests were conducted to evaluate the effect of timepoint within each of the eight genetic lines. The results showed that timepoint had a significant effect on some α-diversity indices in 5 out of 8 genetic lines (A, B, C, D, H), while no significant effect was observed on any α-diversity index in the remaining 3 lines (E, F, G). Detailed results are provided in Supplementary Table 5. Post-hoc tests did not identify significant differences between timepoint 1 and timepoint 2 in any genetic line. Nonetheless, 10 out of 13 total post-hoc comparisons showed a significant difference between timepoint 1 and timepoint 3, and between timepoint 2 and timepoint 3, as detailed in Supplementary Table 6.

PERMANOVA analysis explored differences across genetic lines, timepoints, and their interaction in β-diversity using Bray-Curtis dissimilarity. Results showed no significant effect of genetic line or the interaction between timepoint and genetic lines on β-diversity (stress value 0.241). This finding is supported by Figure 3. Additionally, temporal changes accounted for 12.79% of the total variation in microbiota composition, with a P-value of 0.001.

Furthermore, in the same years under consideration, the analysed colonies yielded two distinct honey crops: the first of Acacia in May (prior to sampling), and the second of wildflower in June (between timepoint 1 and 2 of sampling).

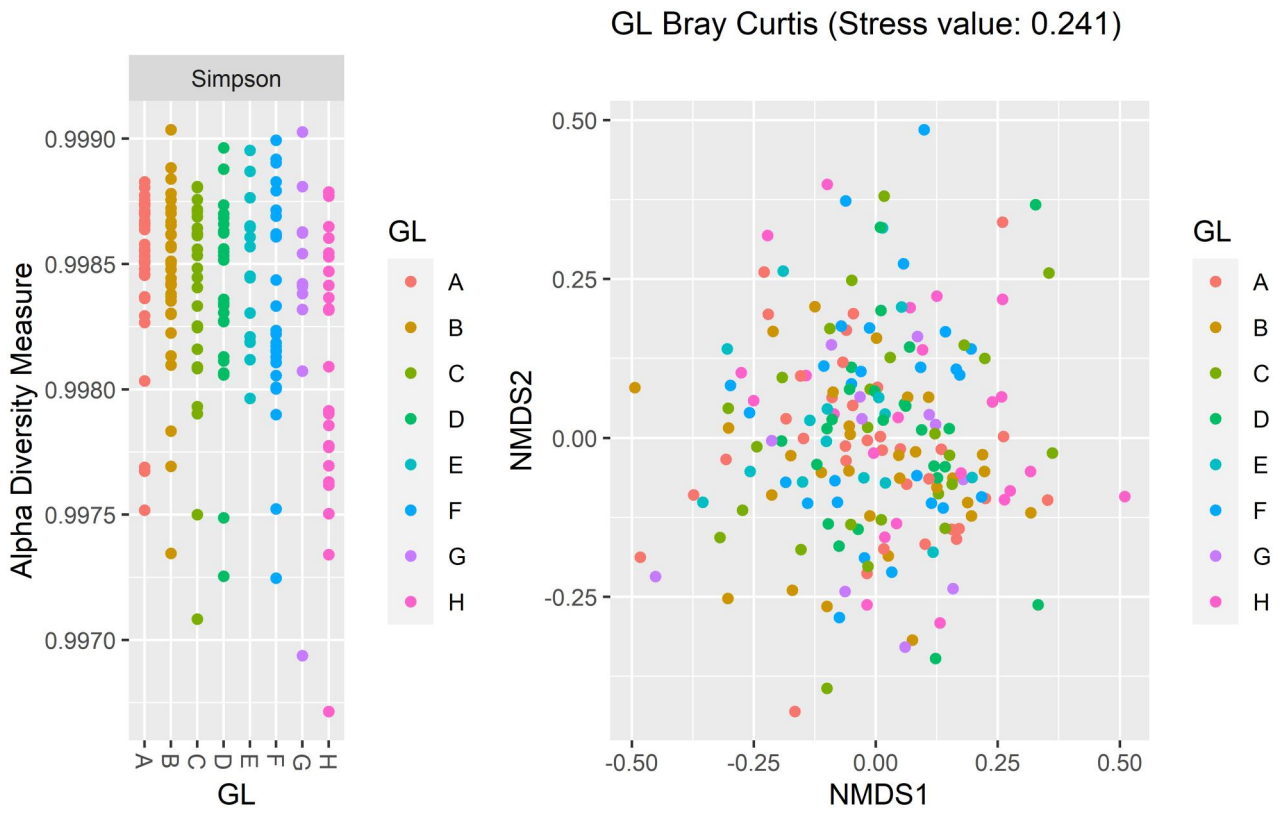

Figure 3. Graphical representation of the Simpson indexes (a) and of β diversity based on Bray Curtis distance with relative stress value (b), grouping samples by genetic lines. Samples from different genetic lines are shown with different colours.

Table 4. Results of Kruskal-Wallis test on Simpson index among genetic lines at the three timepoints. The table shows the timepoint, chi-squared (chi2), degrees of freedom (df) and p-values.

Timepoint	chi^2	df	p-value
1	5.27	7	0.627
2	14.85	7	**0.038**
3	10.09	7	0.183

Values in bold indicate statistical significance.

From the statistical tests conducted using genetic line as the independent variable and honey production as the dependent variable, ANOVA results showed a significant association between genetic line and total honey production ($p = 0.0034$). Similarly, the Kruskal-Wallis revealed a significant association with wildflower production ($p = 0.025$).

Comparing the results of α-diversity analysis and honey production, genetic line H showed both lower diversity and lower honey yields compared to the other genetic lines, with a mean of 3.66 kg for Acacia and 21.90 kg for wildflower honey yield. Supplementary Figure 1 show box plots of the honey production for each genetic line, while Supplementary Tables 7 and 8 give the detailed results of ANOVA and Kruskal-Wallis between genetic lines and honey yield.

Discussion

The main objective of the study was to investigate the influence of host genetics on the gut microbial community across a 5-month period. At all three timepoints and across genetic lines, the most prevalent bacteria belonged to the five genera previously identified as comprising the core honeybee microbiota (Kwong and Moran 2016). The most abundant taxa were *Bifidobacterium* and *Lactobacillus,* followed by other core microbiota genera such as *Bombilactobacillus, Gilliamella* and *Snodgrasella*. The heritability of this genera was estimated to range from 0.4 to 0.6 (Wu et al. 2021).

Other two taxa that are not part of the most common core microbiota were identified at each timepoints and genetic lines: *Frischella* and *Commensalibacter*. This finding is in line with previous studies that identified these two taxa as very frequently observed in *Apis mellifera* across different seasons (Bleau et al. 2020; Subotic et al. 2019; Damico et al. 2021; Ludvigsen et al. 2015; Nowak et al. 2021). Moreover, *Commensalibacter* seems to be positively correlated with frames of bees and brood (Almeida et al. 2023).

Other taxa were identified only in specific timepoints or genetic lines. Overall, timepoints 1 and 2 showed greater similarity in microbiota composition to

each other compared to timepoint 3, which, being the latest in time and outside the pollination period, demonstrated more distinct differences, as indicated by the post hoc analysis.

Bacteria belonging to the *Apilactobacillus* genus and the *Rhizobiaceae* family (unknown genera) were exclusively identified at timepoints 1 and 2 across all genetic lines, but not at timepoint 3. *Apilactobacillus* (formerly referred to as *Lactobacillus kunkeei*), is a flower-associated bacteria and an obligately fructophilic lactic acid bacterium (Neveling et al. 2012). Therefore, it is reasonable to find these taxa in June and July, during pollination, but not in October. Additionally, it seems to contribute to bee health by inhibiting the proliferation of certain pathogens, such as Serratia marcescens, in the gut (Chege et al. 2023). The *Rhizobiaceae* family was particularly abundant, especially at timepoint 2, where it accounted for 3.59% to 10.07% of the family taxa. Seasonal variation in the relative abundance of *Rhizobiaceae*, particularly from April to September, has also been reported by other studies (Bleau et al. 2020; Almeida et al. 2023).

Escherichia-Shigella was identified in all genetic lines at timepoint 2 in very low proportions (0.02-0.08), and only in five genetic lines (A, B, C, D, and H) at timepoint 1. The lower presence of *Escherichia-Shigella* may be attributed to a significant decrease of these taxa within the gut microbiota typically observed between 19 and 25 days post-emergence of worker bees (Dong et al. 2020). Finally, *Enterobacter*, while slightly present in all months of sampling, showed variation across timepoints and genetic lines. Specifically, it was absent in some genetic lines at timepoint 1, only appearing in six out of eight lines (A, B, D, F, G, and H), while it was found in all genetic lines by timepoints 2 and 3. This pattern suggests a temporal and possibly genetic influence on the colonisation or persistence of *Enterobacter*, which could be linked to environmental factors or developmental stages within the colonies (Anderson and Maes 2022).

Statistical analyses revealed a significant effect of genetic lines on Simpson index, indicating that certain genetic lines tend to exhibit lower or higher α-diversity values. Previous studies have found that greater genetic diversity within a colony correlates with higher α diversity (Mattila et al. 2012; Bridson et al. 2022). In our study, the genetic line H was identified as significantly different from the other due to its lower Simpson index. The major differentiation among genetic lines occurred in July (timepoint 2). Additionally, the most significant difference in Simpson diversity was observed between genetic line A at timepoint 1 (June) and genetic line H in timepoint 3 (October). This finding suggests that genetic line H tends to have a lower microbial diversity, especially at later timepoints in the study period.

The effect of timepoints on α-diversity was further investigated within each genetic line using Kruskal-Wallis tests, which showed a significant impact on some α-diversity indices in five out of eight genetic lines (A, B, C, D, H). The post-hoc analysis revealed that, while no significant differences were observed between timepoint 1 (June) and timepoint 2 (July), significant differences were found between timepoint 1 and timepoint 3, as well as between timepoint 2 and timepoint 3 for most of the comparisons. This suggests that the most pronounced shifts in microbial diversity occurred between July and October, highlighting the influence of seasonal or time-dependent factors on microbial communities across different genetic lines. This supports findings reported by Almeida et al. (2023), who demonstrated a greater shift in September compared to other timepoints during the foraging period.

This lower microbial diversity in genetic line H may also be linked to differences in colony productivity. Given that Acacia honey production in 2021 was significantly lower than wildflower honey, with three colonies not producing any Acacia honey at all (likely due to unfavourable environmental conditions during that year), the statistical analysis still confirmed a significant association between genetic lines and total honey production, as well as wildflower honey production. The lack of a significant association between genetic line and Acacia honey yield may reflect the unusually poor harvest of the colonies under study for that crop in 2021.

Notably, colonies of genetic line H, which had lower α diversity, also produced the least honey in both May and June. These findings together suggest that host genetics impacts both gut microbiota diversity and bee productivity.

A recent study comparing the gut microbiota of active and inactive foragers found that inactive foragers had a higher relative abundance of *Lactobacillus* and lower levels of *Bombilactobacillus mellis* (Vernier et al. 2024). In our study, genetic line H, during the foraging period (timepoint 1), exhibited the highest relative abundance of *Lactobacillus* (41.2%) among all genetic lines, while its *Bombilactobacillus* abundance (15.89%) was relatively low. Together, these findings support the hypothesis that both host genetics and microbiota composition are critical factors in determining foraging efficiency and, consequently, honey yield.

Finally, our findings underscore the significant influence of timepoint as the primary factor, highlighting its pivotal role in shaping the diversity of honeybee gut microbiota. The sampling timepoint not only indicates different climatic conditions but also relates to different diet composition, as June (timepoint 1) and July (timepoint 2) concur within the pollination period, whereas in October (timepoint 3), honeybees predominantly consume sugar syrup. All these findings suggest that both environmental and genetic factors play a crucial role in determining the gut microbiota composition in honeybees (Wu et al. 2021).

Conclusion

Our investigation focuses on the pivotal role of host genetics in shaping the composition of honeybee gut microbiota. Consistent patterns across genetic lines and timepoints showed that core genera such as *Lactobacillus* and *Bifidobacterium* dominate the microbial community.

A significant association between genetic lines and α-diversity indexes was observed, with genetic line H showing a significantly lower Simpson index, indicative of less diverse gut microbiota. Moreover, the significant association between genetic lines and honey production underscores the multifaceted influence of host genetics on both gut indexes had also the lowest honey yields.

Contrarily, β-diversity analysis revealed no significant differences among genetic lines, suggesting that genetic factors primarily influence microbial diversity within individuals, rather than distinct community structures. Instead, temporal dynamics were a significant driver of microbiota variation, accounting for a notable portion of the total variability.

These results highlight the potential of integrating microbiota analysis into bee breeding programs, where selecting for genetic lines with more favourable microbial communities could enhance honey production and overall colony health. Future research is needed to validate these results and to explore these interactions further. Despite the limited sample size, this study involves a unique population where both maternal and paternal pairings are controlled, and detailed pedigree and breeding value data are available. Phenotyping such populations requires significant time and effort, which underscores the rarity and value of the insights gained from this dataset. Future studies should continue to build on this foundation to optimise breeding strategies for improved productivity and resilience in bee populations.

Author contributions

GM and FT designed the study. MGDI conducted the data analyses. FT, GM and GP supervised and validated the analyses. MGDI wrote the first draft. All authors revised the manuscript. GM funded the work.

Disclosure statement

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

Funding

This research was funded by the BEENOMIX 2.0 and BEENOMIX 3.0 projects funded by the Lombardy Region (FEASR program). PSR (grant number 201801057971—G44I19001910002 and 202202375596—E87F23000000009).

Data availability statement

These data (sequences and phenotypes) are based on a breeding population used for selection by commercial breeders and have commercial value. Therefore, restrictions apply to the availability of these data, which are not publicly available. The authors can be contacted for a specific request.

References

Almeida EL, Ribiere C, Frei W, Kenny D, Coffey MF, O'Toole PW. 2023. Geographical and seasonal analysis of the honeybee microbiome. Microb Ecol. 85(2):765–778. doi: 10.1007/s00248-022-01986-x.

Anderson KE, Maes P. 2022. Social microbiota and social gland gene expression of worker honey bees by age and climate. Sci Rep. 12(1):10690. doi: 10.1038/s41598-022-14442-0.

Bleau N, Bouslama S, Giovenazzo P, Derome N. 2020. Dynamics of the honeybee (*Apis mellifera*) gut microbiota throughout the overwintering period in Canada. Microorganisms. 8(8):1146. doi: 10.3390/microorganisms8081146.

Bridson C, Vellaniparambil L, Antwis RE, Müller W, Gilman RT, Rowntree JK. 2022. Genetic diversity of honeybee colonies predicts gut bacterial diversity of individual colony members. Environ Microbiol. 24(12):5643–5653. doi: 10.1111/1462-2920.16150.

Callahan B, McMurdie P, Rosen M, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 13(7):581–583. doi: 10.1038/nmeth.3869.

Chege M, Kinyua J, Paredes JC. 2023. *Lactobacillus kunkeei* impacts the health of honey bees, *Apis mellifera scutellata*, and protects the bees against the opportunistic pathogen *Serratia marcescens*. Int J Trop Insect Sci. 43(6):1947–1955. doi: 10.1007/s42690-023-01103-6.

Damico ME, Rueppell O, Shaffer Z, Han B, Raymann K. 2021. High royal jelly production does not impact the gut microbiome of honey bees. Anim Microbiome. 3(1):60. doi: 10.1186/s42523-021-00124-1.

Dong ZX, Li HY, Chen YF, Wang F, Deng XY, Lin LB, Zhang QL, Li JL, Guo J. 2020. Colonization of the gut microbiota of honey bee (*Apis mellifera*) workers at different developmental stages. Microbiol Res. 231:126370. doi: 10.1016/j.micres.2019.126370.

Ellegaard KM, Suenami S, Miyazaki R, Engel P. 2020. Vast differences in strain-level diversity in the gut microbiota of two closely related honey bee species. Curr Biol. 30(13): 2520–2531.e7. doi: 10.1016/j.cub.2020.04.070.

Engel P, Moran NA. 2013. The gut microbiota of insects - diversity in structure and function. FEMS Microbiol Rev. 37(5):699–735. doi: 10.1111/1574-6976.12025.

Henry M, Béguin M, Requier F, Rollin O, Odoux J-F, Aupinel P, Aptel J, Tchamitchian S, Decourtye A. 2012. A common pesticide decreases foraging success and survival in honeybees. Science. 336(6079):348–350. doi: 10.1126/science.1215039.

Kapheim KM, Rao VD, Yeoman CJ, Wilson BA, White BA, Goldenfeld N, Robinson GE. 2015. Caste-specific differences in hindgut microbial communities of honey bees (*Apis mellifera*). PLoS One. 10(4):e0123911. doi: 10.1371/journal.pone.0123911.

Kešnerová L, Emery O, Troilo M, Liberti J, Erkosar B, Engel P. 2020. Gut microbiota structure differs between honeybees in winter and summer. Isme J. 14(3):801–814. doi: 10.1038/s41396-019-0568-8.

Kim Y. 2023. Bioinformatic and statistical analysis of microbiome data. Methods Mol Biol. 2629:183–229.

Klein A-M, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T. 2007. Importance of pollinators in changing landscapes for world crops. Proc Biol Sci. 274(1608):303–313. doi: 10.1098/rspb.2006.3721.

Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, et al. 2021. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 53(2):156–165. doi: 10.1038/s41588-020-00763-1.

Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. Nat Rev Microbiol. 14(6):374–384. doi: 10.1038/nrmicro.2016.43.

Le Conte Y, Ellis M, Ritter W. 2010. Varroa mites and honeybee health: can Varroa explain part of the colony losses? Apidologie. 41(3):353–363. doi: 10.1051/apido/2010017.

Le Conte Y, Navajas M. 2008. Climate change: impact on honey bee populations and diseases. Rev Sci Tech. 27(2): 485–497.

Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sánchez S, Chen L, Vila AV, Gacesa R, Sinha T, Collij V, et al. 2022. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet. 54(2):143–151. doi: 10.1038/s41588-021-00992-y.

Ludvigsen J, Rangberg A, Avershina E, Sekelja M, Kreibich C, Amdam G, Rudi K. 2015. Shifts in the midgut/pyloric microbiota composition within a honey bee apiary throughout a season. Microbes Environ. 30(3):235–244. doi: 10.1264/jsme2.ME15019.

Martinson VG, Moy J, Moran NA. 2012. Establishment of characteristic gut bacteria during development of the honeybee worker. Appl Environ Microbiol. 78(8):2830–2840. doi: 10.1128/AEM.07810-11.

Mattila HR, Rios D, Walker-Sperling VE, Roeselers G, Newton IL. 2012. Characterization of the active microbiotas associated with honey bees reveals healthier and broader communities when colonies are genetically diverse. PLoS ONE. 7(3):e32962. doi: 10.1371/journal.pone.0032962.

McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 8(4):e61217. doi: 10.1371/journal.pone.0061217.

Neveling DP, Endo A, Dicks LMT. 2012. Fructophilic *Lactobacillus kunkeei* and *Lactobacillus brevis* isolated from fresh flowers, bees and bee-hives. Curr Microbiol. 65(5): 507–515. doi: 10.1007/s00284-012-0186-4.

Nowak A, Szczuka D, Górczyńska A, Motyl I, Kręgiel D. 2021. Characterization of *Apis mellifera* gastrointestinal microbiota and lactic acid bacteria for honeybee protection—a review. Cells. 10(3):701. doi: 10.3390/cells10030701.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens MHH, Wagner H. 2015. Vegan: community Ecology Package. R package version 2.2-1.

R Core Team. 2022. R: A language and environment for statistical computing. Vienna. Austria; https://www.r-project.org/.

Subotic S, Boddicker AM, Nguyen VM, Rivers J, Briles CE, Mosier AC. 2019. Honey bee microbiome associated with different hive and sample types over a honey production season. PLoS One. 14(11):e0223834. doi: 10.1371/journal.pone.0223834.

Tarpy DR, Mattila HR, Newton IL. 2015. Development of the honey bee gut microbiome throughout the queen-rearing process. Appl Environ Microbiol. 81(9):3182–3191. doi: 10.1128/AEM.00307-15.

Vernier CL, Nguyen LA, Gernat T, Ahmed AC, Chen Z, Robinson GE. 2024. Gut microbiota contribute to variations in honey bee foraging intensity. Isme J. 18(1): wrae030. doi: 10.1093/ismejo/wrae030.

Wu J, Lang H, Mu X, Zhang Z, Su Q, Hu X, Zheng H. 2021. Honey bee genetics shape the strain-level structure of gut microbiota in social transmission. Microbiome. 9(1):225. doi: 10.1186/s40168-021-01174-y.

Yang J, Zhong Y, Xu L, Zeng B, Lai K, Yang M, Li D, Zhao Y, Zhang M, Li D. 2021. The dominating role of genetic background in shaping gut microbiota of honeybee queen over environmental factors. Front Microbiol. 12:722901. doi: 10.3389/fmicb.2021.722901.