Imperial College London

Microbial associations and inferred interactions in wild bumble bees

Congjia Chen

August 2022

Formatted in the journal style of **Journal of Evolutionary Biology**

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Imperial College London

Submitted for the MSc in Computational Methods in Ecology and Evolution

Declaration

I declare that the raw data is provided by my supervisor from Imperial college London, Dr. Peter Graystock.

I declare that I am responsible for all the data processing , analysis and visualization.

I declare that all third-party software and works are appropriately referenced.

Abstract

Bacterial gut symbiont communities are critical for the health of social insects. However, the differences in microbial communities between different host species and how they react to parasite invasion are poorly understood. Some *Bombus* populations are experiencing the declines, probably due to gut pathogens and loss of genetic diversity. Next generation sequencing data was utilized to study *Bombus* microbial association and interaction with host species and parasite infection. Microbial community composition differs among bumble bee species. Genetically related *Bombus* species tend to have similar microbial communities. *B. hortorum* and *B. ruderatus* are dominated by relatively higher abundance of environmental bacteria. In contrast, *B. terrestris* is best characterized by core bacteria that have only been sampled from the guts of *Apis* and *Bombus*. Bumble bees are known to have a high incidence of the pathogens *Crithidia*, and inbreeding increases the likelihood of this parasite spread. The *Crithidia* load was negatively associated with the abundance of core *Bifidobacterium* bacteria. *Apibacter*, while not well characterized, was found negatively associated with both *Nosema* and *Crithidia* infection, and interacted significantly with other core bacteria. Our results indicate that *Bombus* species have distinctive gut bacterial communities, and that genus-level gut bacterial variation is an important determinant associated with parasite infection. As inbreeding mainly induced by population decline or isolated population will lead to higher parasite prevalence and threaten to the population, the microbial communities may warrant more research into the probiotic benefits against the parasite.

Keywords

Bumble Bee; Microbiome diversity and interactions; Parasite-host interaction; Parasite-microbiome interaction

1 Introduction

Many bumble bee populations are currently experiencing declines (Whitehorn et al., 2022), with the increasing pathogenic infection and loss of genetic diversity considered to be potential reasons (Figueroa et al., 2021; Whitehorn et al., 2014; Wallberg et al., 2014; Potts et al., 2010). Bees are naturally infected by a diverse range of parasites and pathogens including *Nosema sp.* (Kurze et al., 2016; Emsen et al., 2020), *Crithidia sp.* (Elias, 2022; Mockler et al., 2018), *Apicystis sp.*, (Schoonvaere et al., 2020; Tian et al., 2018) and Deformed Wing Virus (Graystock et al., 2016). They are known to lead to reduced *Bombus* direct survival (Kurze et al., 2016), shorter foraging flights (Koch et al., 2017), reduced homing ability (Koch et al., 2017) and lower reproductive fitness such as the reduction of fatbody in overwintering queens (Grupe & Quandt, 2020; Graystock et al., 2016). Whilst there is some co-evolution between hosts and parasites (Moro et al., 2021), the long foraging distance and movement of infected bees has facilitated novel parasites transmission into new areas where hosts may not have natural defences (Grupe & Quandt, 2020; Graystock et al., 2016). In addition to novel pathogen transmission, the loss of genetic context (inbreeding) in which selection can act will lead to lower co-evolution capacity (Whitehorn et al., 2009, 2014; Isagi et al., 2020). While remain largely unstudied in the bumble bee's gut parasites, reduced population-level genetic diversity, were blamed for increasingly severe respiratory parasitic infections and population decline (Whitehorn et al., 2014). Therefore, the inbreeding mainly induced by habitats isolation and population reduction may aggravate parasites infection and pose an additional threat to bumble bees population Whitehorn et al. (2014, 2009).

Although the effect of parasites can be evident, we are still learning about the breadth and depth of the significance of symbiotic gut bacteria to the health of bees. Gut microbiome of bumble bees, which was found to share the digestive niche with pathogenic organisms (Cox-Foster et al., 2007), influences host health in a number of ways including energy supply (Elias, 2022), polysaccharide digestion (Zheng et al., 2019), detoxification(Wu et al., 2020), and providing protection against parasites infections (Mockler et al., 2018; Koch & Schmid-Hempel, 2011). Therefore, when considering the direct effects of stressors on bees, it is important to also know if these stressors have disturbed the important microbiome or if the microbiome is responding positively or negatively to a particular stressor, since these may lead to indirect effects on bee health. Bumble bees collected microbiome

from variety of sources including shared food source in the hive such as fermenting honey (Simone et al., 2009), and external environment such as surface of flowers (Corby-Harris et al., 2014). Five bacterial phylotypes (*Lactobacillus Firm-4*, *Lactobacillus Firm-5*, *Snodgrassella*, *Gilliamella*, and *Bifidobacterium*) are repeatedly found in individuals of both *Apis* and *Bombus* and are defined as 'core' bacteria phylotypes of corbiculate bees (Kwong & Moran, 2016). It is useful to differentiate 'core' taxa with 'non-core' taxa, as core bacterial communities may have co-evolved with their hosts (Kwong & Moran, 2013; Martinson et al., 2012; Kwong et al., 2017), and have been linked to the health of bees (Zheng et al., 2019; Kwong & Moran, 2016). By contrast, non-core taxa frequently occur in non-host-specific environments, and the presence of a higher proportion of non-core taxa , on the other hand, may indicate dysbiosis (Cariveau et al., 2014; Raymann et al., 2018; Zheng et al., 2018). In contrast to the uniform, core microbial community observed in healthy honey bees, the bumble bee individuals harbouring gut communities range from having uniform abundances to being dominated by a few phylotypes (e.g. core bacteria taxa), making them more unpredictable than honey bee gut communities (Kwong et al., 2017). Moreover, the *Bombus* gut microbiota seems to be more susceptible to displacement by environmental microbes for unknown reasons (Li et al., 2015; Kwong et al., 2017; Cariveau et al., 2014). Therefore, It is a great challenge to identify the healthy gut microbiome community for bumble bees, as many variables, including host species, host genetics, habitat, parasites infections and other environmental parameters, may contribute to the variation in the microbial community observed among hosts (Cariveau et al., 2014; Kwong & Moran, 2016; Jabal-Uriel et al., 2022).

Previous researches on *Bombus* gut bacteria have mostly focused on specific taxa (Koch & Schmid-Hempel, 2011, 2012; Mockler et al., 2018), and few comprehensive surveys about bumble bees reveal the interactions within *Bombus* gut community. Infection by *Nosema* and *Crithidia* could modify the state of the digestive tract of bees, leading to changes in gut microbial communities, with increases or decreases in some core bacteria (Jabal-Uriel et al., 2022; Huang & Evans, 2020; Koch & Schmid-Hempel, 2011, 2012; Mockler et al., 2018). However, the vast majority of microorganisms do not exist in isolation (Ma et al., 2020; Faust & Raes, 2012). Instead, they create intimate connections in module that yield enhanced advantages for the community. Due to a lack of mechanistic knowledge of microbial community assembly, inconsistencies in microbial community prediction over various fields have been discovered (Ma et al., 2020). Nevertheless, modules in microbial associa-

tion networks have been found to assist in identifying lifestyle similarities (Berry & Widder, 2014) and might be indicative of ecological processes, such as habitat preference (Lima-Mendez et al., 2015). To resolve microbial community unpredictability and explore how the interactions will benefit the bees, a thorough knowledge of all features of microbiome community, including microbial interaction and co-occurrence networks, is needed.

In this study, I investigate if and how the microbiome differ among three *Bombus* species. I then investigate the effect of two common parasites ,*Nosema* and *Crithidia*, and genetic heterogeneity on the microbiome structure. Finally, as genetic diversity loss may facilitate the parasitism, I evaluate the relationship between genetic diversity loss and parasitism. In the UK, both *B. hortorum* and *B. terrestris* are found widespread, whilst *B. ruderatus* are found in a declining pockets of the UK (Williams, 2005; Whitehorn et al., 2022; Williams et al., 2007). Therefore, despite being closely related to the common *B. hortorum*, *B. ruderatus* is now listed as a priority species for conservation in the UK Biodiversity Action Plan (UKBAP) (Bee, 2022). To understand the interactions between the parasites and its niche (i.e. host and gut microbiome) and to better conserve the bumble bees, I specifically, investigated three questions: (1) Whether and how gut microbial community differ among three *Bombus* species and among different habitats? (2) Can any interactions between microbiomes and parasites of bumblebees be inferred? (3) Does bumble bee's host genetic variation influence parasitism and the microbial community?

2 Material and Methods

2.1 Experimental design and sample selection

This study uses data from previously collected and sequenced samples. The design of the original collections is summarized as follows. Four sites across England, where all three species (*B. hortorum*, *B. ruderatus*, and *B. terrestris*) coexist were chosen: Caistor, Hillesden, Ouse Washes and Upton (Figure 1). Given that bumblebee workers are thought to forage at distances up to 16 km (Rao et al., 2019), spatial distance was also considered to reduce the possibility of specimens being collected from the same colony at various locations (Figure 1a). At the four locations, 253 bumblebees were collected for further study. Visual identification using morphological characteristics resulted in the identification of 93 *B. hortorum*, 75 *B. ruderatus*, and 85 *B. terrestris*. Taking advantage of multi-omics data including Restriction site-associated DNA sequencing (RAD-seq) host genomics data, microbiome 16S amplicon sequencing data, and parasites counts, I tested associations among host genetics (species), microbiome and parasitism (Figure 1b). To avoid unequal sample sizes, three *Bombus* species evenly collected for different experiment use (Figure 1c). 90 of the bees sampled for their microbiomes also contains host genetics and parasite data (Figure 1d).

2.2 Detection and quantification of *Crithidia*,*Nosema*

The entire gut and Malpighian tubules of each bee were dissected and homogenized in RNAlater Stabilization Solution (200 μl). Prior to DNA extraction, a 5μl aliquot of this homogenate was diluted 20-fold with phosphate-buffered saline for spore (or choanomastigote) quantification by microscopy. Although molecular detection methods can be more sensitive, spore and choanomastigote counts are significant indicators of infection. Unlike molecular results which can be influenced by non infectious nucleic acids inside the gut lumen, leading to false-positive, the spores of*Nosema* and the choanomastigotes of *Crithidia* are easily recognized because of their distinct size and appearance. Glassic slides, manufactured by KOVA International Inc., were used to count spores (or choanomastigote) in triplicate, based on which the loadings per bee were calculated.

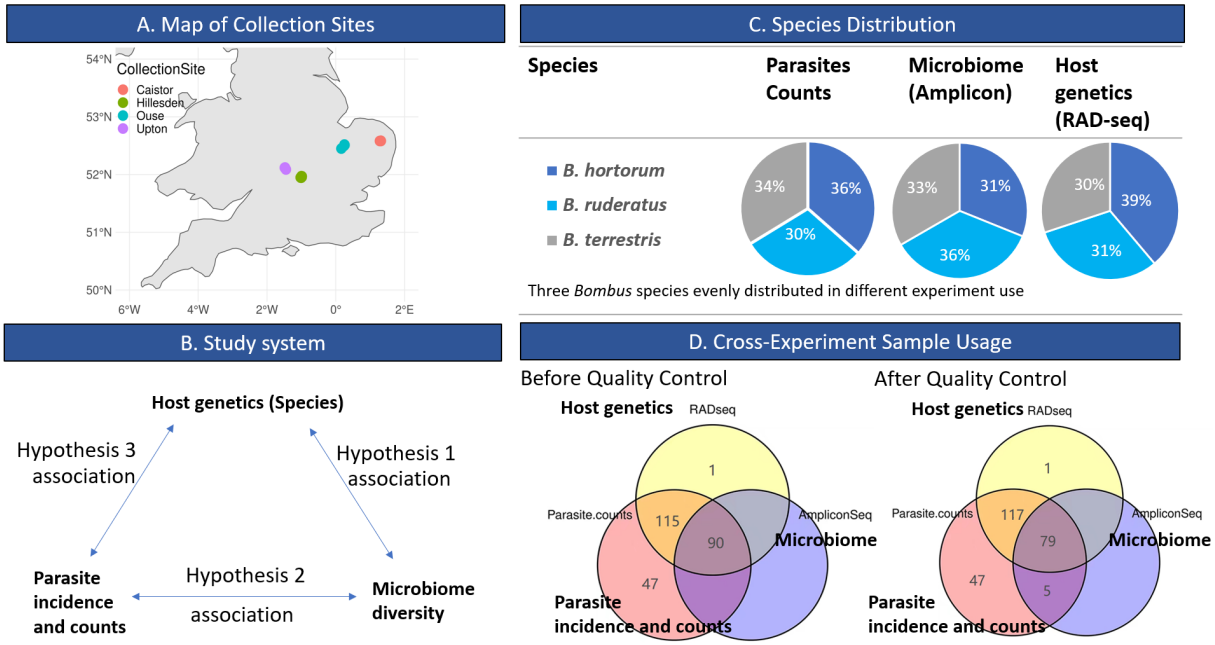


Figure 1: Experimental design and sample usage. (A) Map of Sample Collection. Three species evenly distributed at four collection sites in the UK were collected for different experiments. Pairwise spatial distance > 50 km.(B) Study system. Three main hypothesis were tested in this study. Association between metrics in (1) Host genetics (Species) and Microbiome (2) Microbiome and parasitism (3) Host genetics and parasitism were tested. (C) Species Distribution. Pie chart shows the percent of species used in different analysis. Three *Bombus* species evenly distributed in different experiment use. (D) Cross-Experiment Sample Usage. The venn graph shows that samples were selected from the total bees collected for use in different analyses. In the quality control and data preprocessing steps, 9 samples were discarded from the RAD-seq analysis, and 6 samples were discarded from the amplicon sequencing analysis.

2.3 library preparation and sequencing

DNA was extracted using DNeasy Blood & Tissue kits (Qiagen) before the 206 samples were prepared for RAD-seq and a subset of 90 samples also prepared for 16S sequencing together with a blank control sample (Figure 1c). This was done at the Microbiome Core Facility in North Carolina, USA. The raw data were sent back for further bioinformatics analysis.

2.4 Bioinformatics and statistics

I was provided with the raw paired-end 16s sequencing data, which I processed using QIIME2-2021.11 (Bolyen et al., 2019). I demultiplexed (via q2-demux plugin), visualized and trimmed reads with low quality end using QIIME2 and then used DADA2 (Callahan et al., 2016) (via q2-dada2) to obtain merged, denoised, chimaera-free, exact amplicon sequence variants (ASVs). Taxonomic assignment to the ASVs was performed using the naïve Bayesian-based q2-feature-classifier (Bokulich et al., 2018) trained with the 8F338R region of the 16S rRNA sequence annotated in the SILVA database (Quast et al., 2013).

I also conducted BLAST searches by q2-feature-classifier against the NCBI 16S microbial database (accessed 25 April, 2022) based on identity > 97%. After comparing the results, any discordance were resolved based on confidence and BLAST percentage identity. To reduce the error caused by contaminants, I subtracted the the abundance of the corresponding blank control from the feature abundance and then filtered out features that were assigned as contaminants (e.g. sequences from the host, chloroplasts sequence of flowering plants). To reduce sequencing errors, I chose to remove ASVs with fewer than 3 reads and ASVs that were present in only one sample. As the diversity analysis would be influenced by the uneven library size, to normalize my data, I use an alpha rarefaction plot to determine the subsampling depth. Based on the saturation point of rarefaction curves, I selected 3355 as the normalized library size and 84 samples were retained. I used the MAFFT aligner (Katoh et al., 2002) and FastTree2 (Price et al., 2010) to generate a phylogenetic tree of sequences. I used the resulting tree and ASV table to analyse alpha diversity (Faith phylogenetic diversity (Faith, 1992)) and beta diversity (weighted Unifrac distance (Lozupone et al., 2007) and Bray–Curtis dissimilarity) via q2-diversity.

Whether and how gut microbial community differ among three *Bombus* species and among different habitats?

To differentiate the 'core' taxa with 'non-core' taxa, I define *Snodgrassella*, *Gilliamella*, and *Bifidobacterium* and *Lactobacillus* as the 'core' taxa in my work (Kwong & Moran, 2016).

To test the microbial variation among different host species, collection sites and group with varying infection status, the statistical significance of alpha and beta diversity was tested using the Kruskal–Wallis test and PERMANOVA test, respectively, in QIIME2. To get a genus-level abundance table, I use R package phyloseq (McMurdie & Holmes, 2013) to aggregating the original ASV table. To determine which bacterial genus differs in relation to host species, I performed pairwise Kruskal–Wallis test on top 10 abundant bacterial genus. As genus with low relative abundance may also contribute to the variation among host species, to indicate how specific and widespread an bacteria genus is within a particular group, I performed indicator species analysis based on point-biserial correlation coefficients (Halme et al., 2009; Chytrý et al., 2002; Cáceres & Legendre, 2009) in the R package indicspecies (De Cáceres et al., 2010).

Can any interactions between microbiomes and parasites of bumblebees be inferred?

One of the difficulties in the study of parasite community ecology is determining whether count data (CD) or incidence data (presences/absences) (ID) should be used as input data for analysis (Alexander, 2012). In general, there is no consensus about the requirement of abundance data and the justification of utilising data on parasites presence and absence (Krasnov et al., 2021). For clarity, I used count data (parasite loads) only from infected samples and used incidence data (parasite prevalence) from all samples in this report.

To investigate the presence/absence of the *Crithidia* or *Nosema* given exposure to several variable of interests including core gut bacteria, different collection sites and host species, I used odds ratio from a generalised linear model with a binomial distribution. To test whether the abundance of core gut bacteria changed with parasite loads among different bee species, the lme4 (v1.1-17) package in R was applied to run a linear mixed model (LMM) (Bates et al., 2015). Before running LMM, to normalize the skewed count data and stabilize variation, I first added 1 pseudo-count to the parasite and bacteria abundances and then performed log10 transformation (Alexander, 2012). The lmerTest (version 3.1-0) (Kuznetsova et al., 2017) package was employed to calculate p value using Satterthwaite's degrees of freedom in the LMM. I took parasite loads (infected samples) as the response variable. The fixed effects were the abundance of certain bacteria including *Bifidobacterium*, *Lactobacillus*, *Snodgrassella*, *Gilliamella* and *Apibacter* per bee. The host species were determined as random effects.

To build the association network between microbiomes and parasite loads, I utilised CoNet v1.1.1 (Faust & Raes, 2016) in Cytoscape (Shannon et al., 2003). ASV abundance table were first aggregated into species-level abundance table. To explore the correlation between higher-level taxa, for example, correlation between genus, I enabled "explore links between higher-level taxa" in CoNet (Faust & Raes, 2016). Associations (edges) between microbiome taxa (nodes), *Crithidia* and *Nosema* loads were generated by using a variety of statistical methods, including Pearson and Spearman correlation, mutual information, Bray–Curtis dissimilarity and Kullback–Leibler dissimilarity. To reduce compositionality and for extra stringency (Faust & Raes, 2016), I renormalized the data and employed a mix of permutations and bootstrapping to identify network edges. Following the application of the Brown p value merging technique, the interdependence of different measures were taken

into consideration (Brown, 1975). To account for multiple testing, a Benjamini–Hochberg correction was applied to p value (Benjamini & Hochberg, 1995). Unstable edges ($p < 0.05$) were removed throughout network building. The general topological properties of the inferred networks (number of nodes and edges, average path length, clustering coefficient) were analysed using NetworkAnalyzer (Assenov et al., 2008).

To identify non-random co-occurrence patterns, the checkerboard score (C-score,(Stone & Roberts, 1990)) was used under a null model in the R packages vegan (Oksanen et al., 2022) and bipartite (Dormann et al., 2009). After finding nonrandom co-occurrence pattern, to identify highly overlapping and hierarchical co-occurrence modules within the microbial association network, I removed negative edges and applied the OH-PIN algorithm (Wang et al., 2012) with the default settings (threshold = 2; overlapping score = 0.5) in CytoCluster (Li et al., 2017). A hub is a node having significantly above-average number of connections with other nodes (Ma et al., 2020). I identified two hubs with the highest degree from each modules.

Does bumblebee host genetic variation have association with parasitism and the microbial community?

I was provided with 206 raw paired-end RAD-seq data. To ensure that only high quality samples were used, I filtered out two that were deemed to have low quality retained reads after trimming the adaptor by process_radtags (Stacks 2.61 pipeline, (CATCHEN et al., 2013)). To assess contaminants, sequencing reads were then aligned to the host genome (*Bombus Hortorum*, GenBank: GCA_905332935.1) using BWA-MEM with the default parameters (Li, 2013). SAMtools (version 1.9) was used to obtain the ratio of alignment (Danecek et al., 2021). Thereafter, 4 samples with abnormally low reads aligned to the reference were filtered out. VCF files were derived from 'populations' (Stacks 2.61 pipeline, (CATCHEN et al., 2013)) and then analyzed by VCFtools 0.1.17 (Danecek et al., 2011) to obtain the individual-based heterozygosity (F value). In contrast to female worker bees, drones (male bumblebees) carry only one type of allele at each chromosomal position. To reduce the bias caused by genetic and sexual mechanisms, I removed 3 samples that appeared to be nearly haploid (F value = 1). Finally, I used 197 samples in downstream analysis. Then, to obtain the inbreeding coefficients of different populations (in this study, populations were defined as the specific species at each collection site), I performed SNP calling by ref_map.pl (Stacks 2.61 pipeline, (CATCHEN et al., 2013)) incorporating population information. To

obtain the genetic distance matrix, I imported the VCF file into PLINK 1.9 (Purcell et al., 2007; Purcell, 2022) to generate the Hamming genetic distance matrix (Frederick et al., 1993) between each individual.

To explore the relationships between host genetic variation, parasite intensity and microbial diversity, I used both genetic distance was further compared with microbial beta diversity (Bray–Curtis distance) and parasites load distance (Euclidean metric). The distance correlation was assessed with the Mantel test implemented in the R package ade4 (Dray & Dufour, 2007). To assess the impact of genetic context loss (inbreeding), the inbreeding coefficient (Fis) of different populations were further used as a fixed effects to explore the correlation with parasite incidence by linear mixed model (Kuznetsova et al., 2017). Collection sites and species were used as random effect.

Unless otherwise specified, all statistical analyses were conducted in the R environment (R Core Team, 2022).

3 Results

3.1 Bee species identity drives differences in gut microbiomes and core bacteria abundance

Overall community-level microbial diversity did not differ between collection sites, but did differ between hot species. No interaction effects between species and collection cites were found (Kruskal–Wallis test, H value = 11.33, d.f = 11 , p = 0.42). There were no significant differences among different collection sites in terms of either alpha diversity (Kruskal–Wallis test, H value = 1.07, d.f = 3, p = 0.785; Figure 2b) and beta diversity (PERMANOVA test, pseudo-F value = 0.69, Permutations = 999, p = 0.798; Figure 2d). However, Faith's phylogenetic alpha diversity (Faith's PD) estimates for *Bombus* gut microbiome differed among different host species (Kruskal–Wallis test, H value= 7.64, d.f. = 2 , p < 0.05; Figure 2a). *B. ruderatus* (Faith's PD = 7.17 ± 0.61) had a richer gut microbiota than *B. hortorum* and *B. terrestris* (Faith's PD = 5.92 ± 0.62 s.e., Faith's PD = 4.62 ± 0.15 s.e., respectively), indicating that *B. ruderatus* recruited more genetically distinct gut community than *B. terrestris* and *B. hortorum* (Kruskal–Wallis test, H value = 6.68, d.f = 1, p < 0.01; Figure 2a). When comparing the alpha diversity between three species, *B. terrestris* and *B. ruderatus* accounted for the largest difference. *B. terrestris* showed significantly lower alpha diversity than *B. ruderatus* (Kruskal–Wallis test, H value = 8.29, d.f = 1, p < 0.01; Figure 2a). Whereas, *B. hortorum* showed no significant difference of alpha diversity (Faith's PD) from those of both *B. terrestris* and *B. ruderatus* (Kruskal–Wallis test, H value = 0.01, d.f = 1, p = 0.92; Kruskal–Wallis test, H value = 3.30, p = 0.07,respectively; Figure 2a).

The beta diversity (weighted UniFrac) of microbial community among different host species was also significantly different in terms of relative abundance (PERMANOVA test, pseudo-F value = 5.66, Permutations = 999, p < 0.05; Figure 2c). Specifically, the beta diversity (weighted UniFrac) of *B. terrestris* microbiome differed from those of *B. ruderatus* and *B. hortorum* (PERMANOVA test, pseudo-F value = 11.27, Permutations = 999, p < 0.001; PERMANOVA test, pseudo-F value = 6.05, Permutations = 999, p < 0.001 respectively; Figure 2c), while *B. ruderatus* and *B. hortorum* showed no significant difference (PERMANOVA test, pseudo-F value = 1.52, Permutations = 999, p = 0.155). Comparing across bee species with top 10 abundant genus-level microbiome, the differences of microbiome between *B. terrestris* and other two *Bombus* species were significant and de-

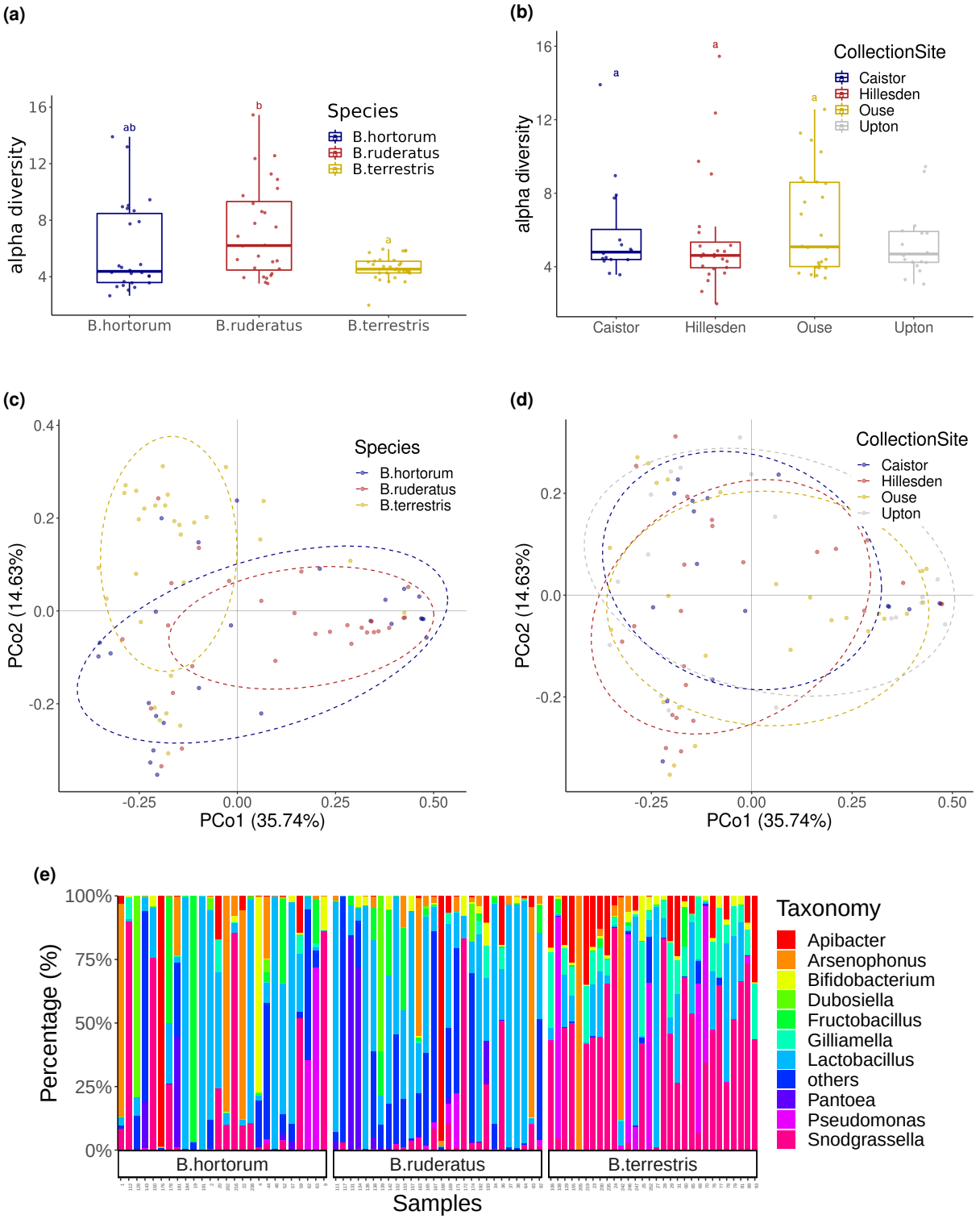


Figure 2: Microbiome alpha and beta diversity and microbial composition differ among *Bombus* species, rather than among habitats. Distribution of alpha diversity (Faith phylogenetic diversity) in each study, stratified by species (a) and collection site (b), where statistical differences are indicated by the letters above the boxes. Principal coordinates PCo1 (account for 35.74% of the variance) and PCo2 (account for 14.63% of the variance) with weighted phylogenetic UniFrac distances. Ellipses cover 68% of the data for each *Bombus* species in (c) and collection sites in (d). (e) Stack plot of top 10 genus (taxonomy labelled with different colors) showing the relative abundance in different host species. The numbers of replicate samples for each species in the Figure 2 are as follows: *B. hortorum* (n=26), *B. ruderatus* (n=28), *B. terrestris* (n=30).

tectable at the genus level (Figure 2e).*B. terrestris* showed higher relative abundance of *Snodgrasella*,*Gilliamella* and *Apibacter* in the community, than those of *B. ruderatus* and *B. hortorum* (Supplementary Table 1; Supplementary Figure 1). Whereas, *B. hortorum* and *B. ruderatus* showed higher proportion of *Pantoea* and *Dubosiella* in the community than those of *B. terrestris* (Figure 2e; Supplementary Table 1; Supplementary Figure 1). *B. ruderatus* showed higher proportion of *Fructobacillus* and *Lactobacillus* than those of *B. terrestris* (Figure 2e; Supplementary Table 1; Supplementary Figure 1). Additionally, among top 10 bacterial genus, no taxa differed between *B. ruderatus* and *B. hortorum* in terms of relative abundance (Figure 2e; Supplementary Table 1; Supplementary Figure 1).

Table 1 Results of indicator species analysis showing the associations between gut bacteria at the genus level and host species. Results are based on point-biserial correlation coefficients > 0.25 and significant at p < 0.05.

Indicator	Core	*B. terrestris* coefficients	*B. hortorum* coefficients	*B. ruderatus* coefficients	p value
Lactobacillus	Yes/No^*	—	—	0.265	<0.05
Gilliamella	Yes	0.595	—	—	<0.001
Snodgrassella	Yes	0.427	—	—	<0.001
Curvibacter	No	—	0.306	0.306	<0.001
Devosia	No	—	0.250	0.250	<0.05
Ralstonia	No	—	—	0.265	<0.05
Sphingomonas	No	—	—	0.284	<0.01
Leifsonia	No	—	—	0.346	<0.001
Brevundimonas	No	—	—	0.317	<0.01
Aquabacterium	No	—	—	0.307	<0.01
Staphylococcus	No	—	—	0.299	<0.01
Phyllobacterium	No	—	—	0.277	<0.05
Bradyrhizobium	No	—	—	0.256	<0.05

* *Lactobacillus* includes both core species found in bee and environmental species associated with flower products

Moreover, as genus with low relative abundance may also contribute to the variation among host species, to better understand the microbiome community difference, I characterized different *Bombus* species with indicative gut bacteria through the Indicator species analysis (ISA). ISA revealed that different gut bacteria genus prefer different *Bombus* species (Table 1). *B. ruderatus* was best characterized by *Lactobacillus* and other environmental bacteria including *Curvibacter* and *Sphingomonas* which are also common in floral products, including nectar and pollen. *B. hortorum* were best characterized by *Curvibacter* and *Devosia*. Although *B. hortorum* and *B. ruderatus* were both preferred by bacteria acquired from environmental sources, *B. ruderatus* showed a wider range of environmental bacteria according to ISA (Table 1). In contrast, *B. terrestris* was best characterized by the core bacteria that have only been sampled from the guts of *Apis* and *Bombus*, *Snodgrassella* and *Gilliamella*. Similar results were shown in the Figure 2e, where the core bacteria, such as *Snodgressella* showed higher relative abundance in samples from *B. terrestris*, while

the samples from *B. hortorum* and *B. ruderatus* showed higher relative abundance of other environmental bacteria.

3.2 The infection with *Crithidia* or *Nosema* is not associated with the community-level microbiome diversity but might be associated with several key bacteria

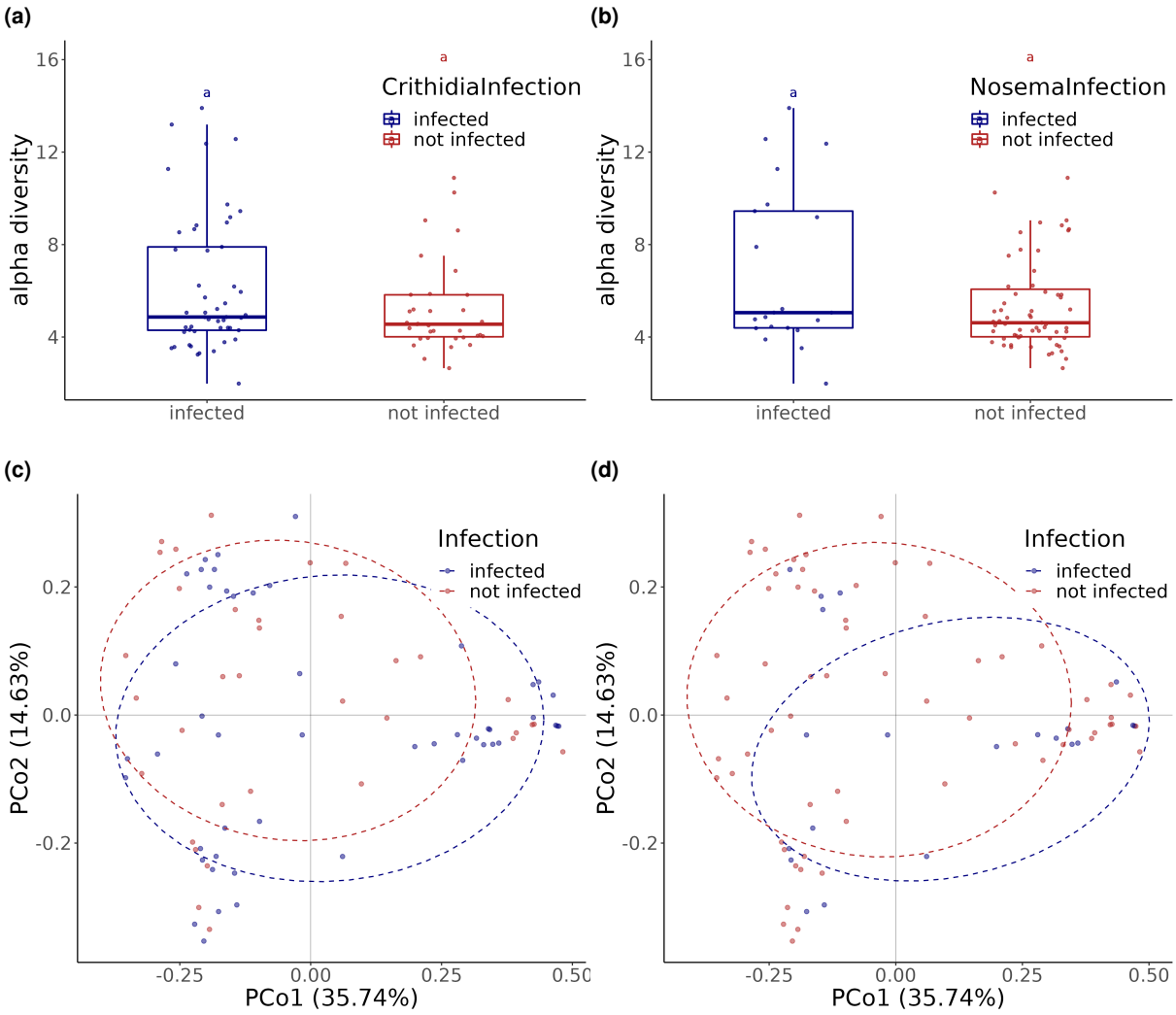


Figure 3: Comparison of alpha and beta diversity between infected and uninfected samples. Boxplot showing the alpha diversity (Faith's phylogenetic diversity) based on presence/absence of *Crithidia* (a) and *Nosema* (b). Principal coordinates (weighted phylogenetic UniFrac distance). PCo1 (accounting for 35.74% of the variance) and PCo2 (accounting for 14.63% of the variance) based on presence/absence of *Crithidia* (c) and *Nosema* (d)

Community-level diversity analysis did not reveal a significant association for the parasites presence/absence and microbial alpha and beta diversity. *Crithidia*-infected and uninfected samples did not differ significantly in terms of alpha diversity (Kruskal-Wallis test, H value = 1.44, d.f = 1, p = 0.23, Figure 3a). *Nosema* exhibited a similar pattern to *Crithidia*, but with a smaller p value (Kruskal-Wallis test, H value = 2.45, d.f = 1, p = 0.12, Figure 3b). In ad-

dition, beta diversity in terms of relative abundance (weighted UniFrac) indicated no significant differences between infected and uninfected samples for *Crithidia* and *Nosema* (PERMANOVA, pseudo-F value = 1.10, Permutations = 999, p = 0.33; PERMANOVA, pseudo-F value = 1.62, Permutations = 999, p = 0.109, respectively).

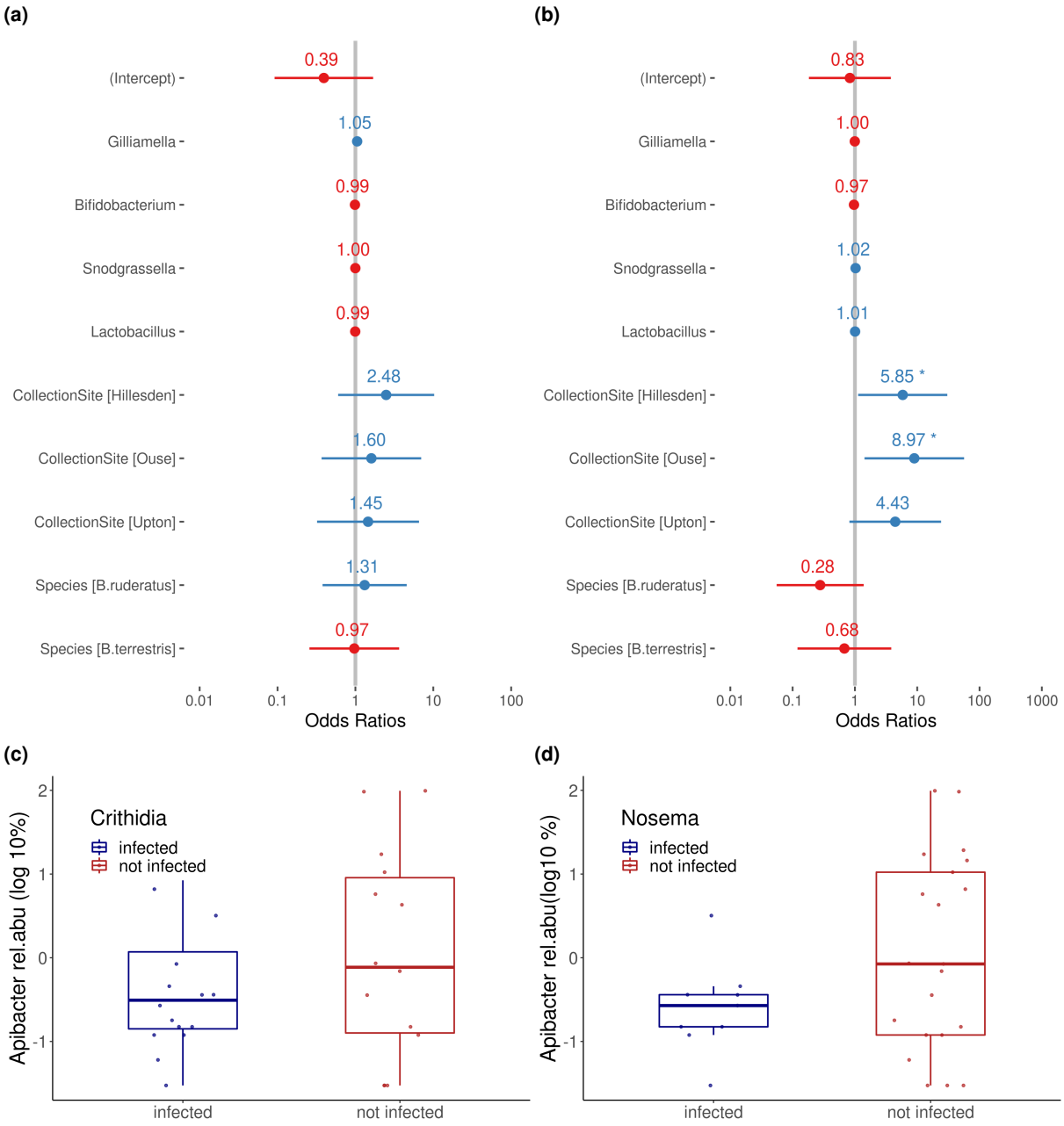


Figure 4: Relative abundance of core bacteria taxa have no association with both *Nosema* and *Crithidia* presence/absence. Forest plots of the coefficient from the binomial generalized linear model assessing *Crithidia* presence/absence (a) and *Nosema* presence/absence (b). A horizontal line representing the 95% confidence intervals of the Odds Ratio. Odds Ratios(OR) = 1: Exposure to variable of interest (e.g. certain bacteria) does not affect odds of the *Nosema* or *Crithidia* presence/absence. The intercept used here is as follows: Host species: *B.hortorum* and Collection site: Caistor. Boxplots showing the relative abundance distribution of *Apibacter* in infected and not infected samples of *Crithidia* (c) and *Nosema* (d)

As core bacteria including *Snodgrassella*,*Gilliamella*, *Bifidobacterim* and *Lactobacillus* were

found to have profound effects on the bees health (Zheng et al., 2019), I included those core bacteria in the generalised linear model with binomial distribution to see if the exposure of specific bacteria taxa have association with the odds of *Nosema* or *Crithidia* presence/absence. Collection sites and species were used as the categorical variables. For coupling with the microbiome data, I only used 84 samples of parasite incidence data (Figure 1d).

Generalised linear model analysis of *Nosema* and *Crithidia* presence/absence did not reveal a significant association given different *Bombus* species, or relative abundance of core bacteria, but suggested a significant association with collection sites.

The presence or absence of *Nosema* and *Crithidia* was not significantly correlated with the selection of *Bombus* species or the relative abundance of core bacteria, but there was a correlation with the collection sites, as determined by a generalised linear model analysis. Odds ratio(OR) of *Crithidia* infection or *Nosema* infection showed range from 0.97 to 1.05, given exposure to core bacteria, indicating that the abundance of core bacteria do not have significantly association with the parasites absence/presence. Additionally, the result showed that *Nosema* incidence was associated with collection sites, indicating that samples in different collection sites had different probability of *Nosema* infection ($\beta_{Hillesden}$ = 1.77, se = 0.84, z value = 2.101 , d.f = 83 , $p < 0.05$; β_{Ouse} = 2.19, se = 0.93, z value = 2.34 , d.f = 83 , $p < 0.05$; Figure 4b). *Crithidia* incidence was neither influenced by collection sites and species nor by bacteria relative abundance (Figure 4a). However, interestingly, the indicator species analysis (ISA) revealed a weak but significant pattern that *Apibacter* preferred *Crithidia* uninfected and *Nosema* uninfected specimens rather than infected specimens (point-biserial correlation coefficient = 0.218, $p < 0.05$;point-biserial correlation coefficient = 0.249, $p < 0.05$, respectively; Figure 4 cd).

To further explore the relationship between parasite loads and core bacterial abundances, I used a linear mixed model (LMM) to explore the microbial associations. *Apibacter* was chosen as a fixed effect in this analysis, because of its association found previously between *Apibacter* and parasites presence/absence. Moreover, as the bacteria differs significantly between species, I use species as the random effect.

In terms of the *Crithidia* infection, only *Bifidobacterium* showed a significant negative corre-

Table 2 Linear mixed model reveals several bacteria significantly associated with parasites loads.

	Crithidia loads per bee			*Nosema* loads per bee		
Predictors	Estimates(95 CI)	t value	p	Estimates(95 CI)	t value	p
(Intercept)	**5.13(4.54 – 5.72)**	**17.66**	**< 0.001**	**3.72(3.00 – 4.43)**	**11.22**	**< 0.001**
Snodgrassella	0.05(-0.13 – 0.24)	0.57	0.572	0.07(-0.33 – 0.46)	0.36	0.723
Bifidobacterium	**-0.20(-0.39 – -0.02)**	**-2.22**	**0.032**	0.11(-0.11 – 0.33)	1.11	0.286
Apibacter	0.06 (-0.12 – 0.24)	0.64	0.523	**-0.33(-0.65 – -0.02)**	**-2.29**	**0.039**
Gilliamella	-0.17(-0.40 – 0.05)	-1.54	0.132	0.02(-0.36 – 0.39)	0.09	0.931
Lactobacillus	0.02(-0.15 – 0.20)	0.28	0.780	**0.23(0.02 – 0.44)**	**2.34**	**0.036**
Random effect	Variance	Standard deviation		Variance	Standard deviation	
Species	0.02141	0.1463		0.04769	0.2184	
Residual	0.22261	0.4718		0.14325	0.3785	

lation with *Crithidia* loads ($\beta_{Bifidobacterium}$ = -0.20, se = 0.09 , t value = -2.218 , d.f = 42.57 , p < 0.05 ;Table 2). In addition, *Apibacter* abundance had a substantial negative correlation with *Nosema* infection loads, while *Lactobacillus* showed a positive correlation with *Nosema* infection loads ($\beta_{Apibacter}$ = -0.33, se = 0.14 , t value = -2.30, d.f = 14.89 , p < 0.05; $\beta_{Lactobacillus}$ = 0.23, se = 0.10, t value = 2.34 , d.f = 13.34 , p < 0.05; Table 2).

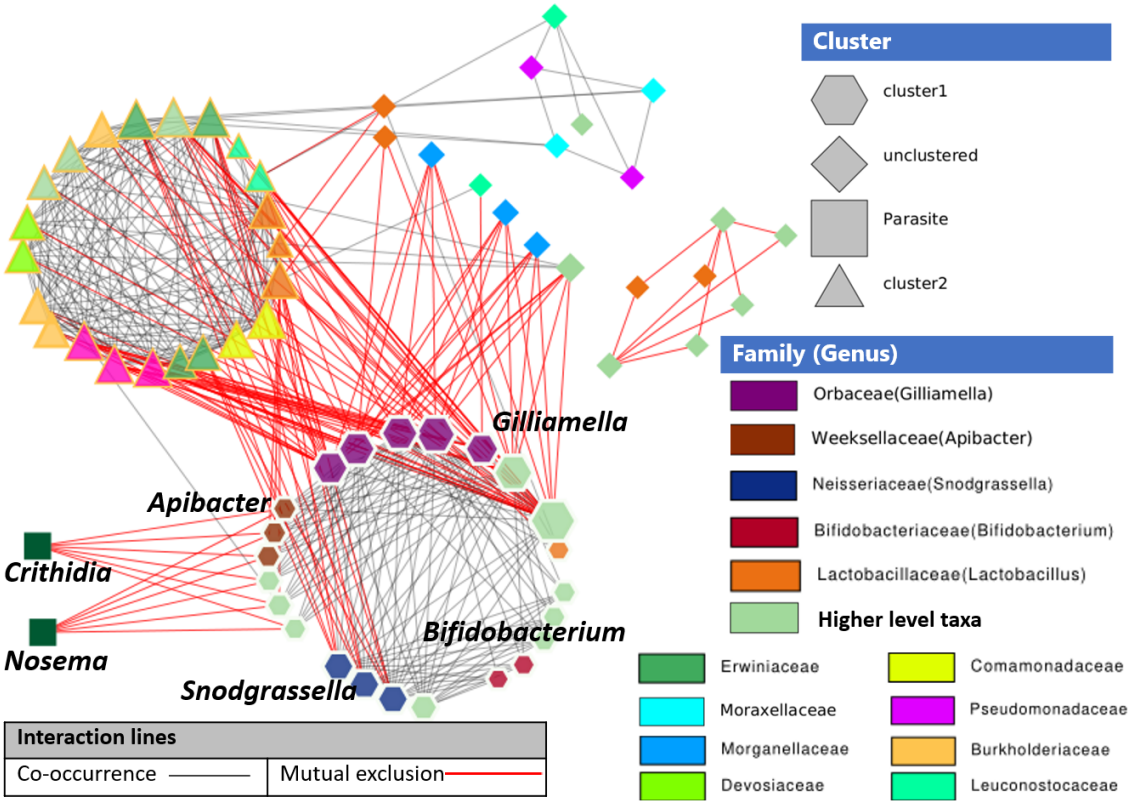


Figure 5: Co-occurrence network shows two distinct co-occurrence clusters in bumble bee's microbial community. Cluster 1 (hexagon) consists of the *Gilliamella*,*Sndograssella*, *Apibacter*, *Bifidobacterium*, and *Lactobacillus*. Cluster 2 (triangle) consists of *Vibrio*, *Pantoea* and other environmental/'non-core' bacteria. A edge stands for a statistical significant (Brown merged p value < 0.05) correlation. The size of each node is proportional to the degree. Cluster 1 and Cluster 2 shows mutual exclusion (red line) with each other. Higher-level taxa, such as phylum, are presented with the matching family-level nodes

Once I observed that the bumble bee's microbial community assemble patterns were non-random (C-score=67.03, Permutations = 999, p <0.01; Figure 5), I further explored co-occurrence modules using OH-PIN algorithm based on a combination of different correlations measures with the application of the Brown p value merging technique (Faust & Raes, 2016). The resulting network (Figure 5) consisted of 67 nodes (including higher-level taxa; only 14 family-level taxa retained, as shown in Figure 5) and 387 edges (average number of neighbours 12.57). Two distinct co-occurrence modules, including one module rich in core bacteria taxa and another one rich in non-core bacteria taxa, were observed. When analysing individual taxa throughout the module, *Gilliamella* and *Apibacter* demonstrated a significant impact on module topology and an association with parasite infection, respectively. When looking at the effects of specific taxa across the module, I find that *Gilliamella* had a major bearing on module topology and *Apibacter* was linked to parasite infection.

The core taxa including *Gilliamella*, *Sndograssella*, *Bifidobacterium*, *Lactobacillus* and *Apibac-*

ter comprised cluster 1 (referred as 'core' module; Figure 5). Whereas, *Pantoea*, *Curvibacter* and other environmental/non-core taxa comprised cluster 2 (referred as 'non-core' module; Figure 5). There were a number of mutual-exclusion (negative association) between taxa in 'core' module and those in 'non-core' module. The difference between non-random and random networks could be explained by the emergence of hubs. *Pantoea*(degree = 21), *Curvibacter*(degree = 19) were found as hubs in 'non-core' module, and *Gilliamella* (degree = 27), *Sndograssella*(degree = 16) were found as hubs in 'core' module. As a hub, the *Gilliamella* served as bridges between the nodes with smaller degree, including *Sndograssella*, *Bifidobacterium* and *Apibacter*. Surprisingly, in addition to the core taxa discovered in the 'core' module, *Apibacter* demonstrated a significant positive connection with *Gilliamella*, resulting in its involvement in the 'core' module ($p < 0.001$; Figure 5; Supplementary table 2). As another hub in the 'core' module, *Sndograssella* was positively linked with both *Gilliamella* and *Bifidobacterium* ($p < 0.001$, Figure 5, Supplementary table 2). Additionally, in my network, *Apibacter* was negatively associated with both *Crithidia* loads and *Nosema* loads. As *Apibacter* was positively associated with hub taxa *Gilliamella*, the *Nosema* and *Crithidia* loads might subsequently associated with the 'core' module and the taxa involved (Supplementary table 2).

3.3 Genetically close *Bombus* species tend to have similar microbial communities and the loss of genetic context will lead to higher parasite prevalence

Samples with sufficient genetic distance (i.e different species) exhibited significantly larger microbial distance (Mantel test, $r = 0.22$, Permutations = 999 , $p < 0.01$; Figure 6a). This result extended the previous finding that the bee species identity drives the differences in gut bacteria, and the genetically related *Bombus* tend to have more similar gut bacterial communities. However, genetic variation could not explain the microbial diversity observed when comparing samples within species (within *B. hortorum*: Mantel test, $r = 0.106$, Permutations = 999, $p = 0.112$; within *B. ruderatus*: Mantel test, $r = -0.235$, Permutations = 999, $p = 0.997$;within *B. terrestris*: Mantel test, $r = -0.039$, Permutations = 999, $p = 0.662$; Figure 6b). Additionally, the inbreeding coefficient of bumble bee's populations showed no significant correlation with alpha diversity (Faith's pd) (Pearson's correlation test, $r = 0.23$, t value = 0.71 , d.f = 9, $p = 0.50$).

146 out of 197 (74%) samples were infected by *Crithidia*, while 63 out of 197 (31%) samples

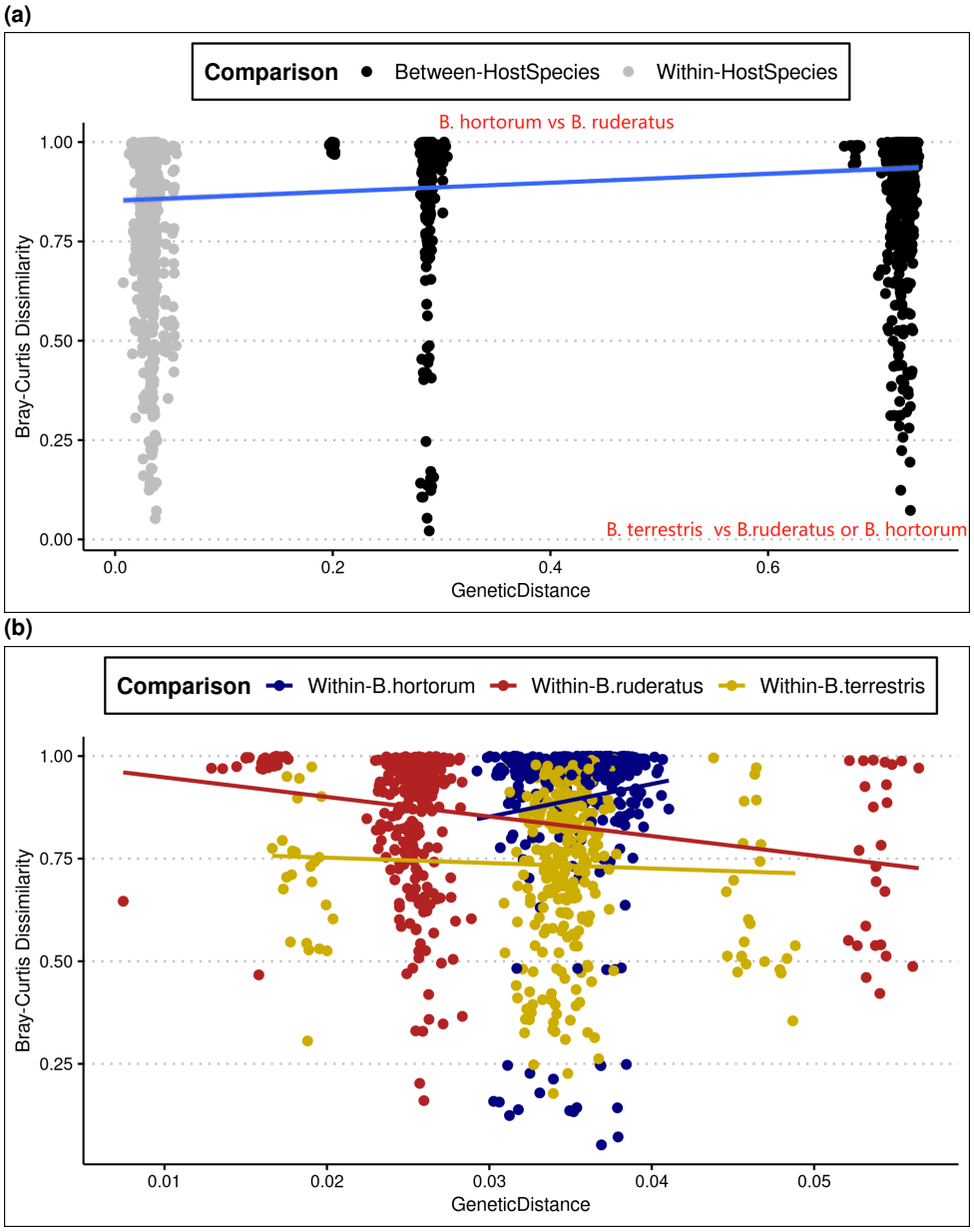


Figure 6: Microbiome dissimilarity as a function of host genetic distance, the line shows the best fitted simple linear model. (a) Bray–Curtis microbiome distance as a function of host divergence (Hamming genetic distance). Each points represents for a pairwise comparison between two samples including comparison within species (grey points) and comparison between host species (black points). Comparison between host species are separated clearly with the former labelled with comparison between *B. hortorum* and *B. ruderatus* and latter labelled with comparison between *B. terrestris* and *B. hortorum*, *B. ruderatus* (b) Bray-Curtis microbiome dissimilarity as a function of host divergence (Hamming Genetic distance) within each *Bombus* species.

were infected by *Nosema*. As no *B.ruderatus* was collected in Caistor, 11 different population showed various parasite prevalence ranging from 52% to 94% of *Crithidia* and 0% to 43% of *Nosema* (Supplementary Table 3). Additionally,the inbreeding coefficient showed a significant positive correlation with *Crithidia* prevalence, indicating that a higher degree of inbreeding would lead to higher parasite prevalence (β_{Fis} = 1.87, se = 0.28, t value = 6.56, d.f = 4.71, p < 0.001; Table 3). In contrast, *Nosema* prevalence showed a negative but non

Table 3 Linear mixed model reveals *Crithidia* prevalence is significantly correlated with inbreeding coefficient. Nosema prevalence is not correlated with inbreeding coefficient.

	Crithidia prevalence			*Nosema* prevalence		
Predictors	Estimates(95 CI)	t value	p	Estimates(95 CI)	t value	p
(Intercept)	**0.61 (0.41 – 0.81)**	**7.44**	**< 0.001**	**0.31 (0.17 – 0.46)**	**5.21**	**< 0.001**
Inbreeding coefficient (Fis)	**1.87 (1.17 – 2.56)**	**6.56**	**< 0.001**	-0.09(–1.54 – 1.37)	-0.15	0.888
Random effect	Variance	Standard deviation		Variance	Standard deviation	
Collection sites	0.02	0.15		0.000	0.000	
Species	0.0007	0.02		0.0012	0.03	
Residual	0.002	0.05		0.015	0.12	

siginificant correlation (β_{Fis} = -0.09, se = 0.59, t value = -0.15, d.f = 4.71, p = 0.888; Table 3). However, the Euclidean distance of parasites load was not associated with host genetic distance (*Crithidia*, Mantel test, r = -0.0143, Permutations = 999,p = 0.654; *Nosema*, Mantel test, r = 0.05, Permutations = 999, p = 0.261).

4 Discussion

To improve the conservation of the wild bumble bees, the aim of the study was to determine the associations and interactions between the gut parasites and its niche (i.e Host and gut microbiome). My results showed that (1) The largest source of variation in the gut microbiome was strongly associated with host species, rather than habitats and infection status, where genetically related *Bombus* species tended to have similar microbial community. (2) Several key microbes were found to be associated with parasite infection (3) As bumble bee populations lose heterozygosity, the prevalence of *Crithidia* will increase (Supplementary Figure 2).

I found evidence in this study to conclude that species are strongly associated with host species, rather than with habitats, where genetically related *Bombus* species tended to have similar microbial community. These results do not rule out the possibility of geographical influences, but they show that the species of the host is a considerably more essential element in determining microbiome composition. Although the present work was constrained by the short geographic distances considered, a study on a larger geographical scale with multiple social insects, produced a consistent result (Kwong et al., 2017). When comparing the alpha diversity of *B. ruderatus* and the other two species, *B. ruderatus* showed higher Faith phylogenetic diversity (Faith, 1992), indicating that *B. ruderatus* recruits more genetically diverse bacteria than other two *Bombus* species. When comparing across bee species with genus-level microbiome, the results showed that the *B. terrestris* have relatively higher abundance of *Snodgrassella* and *Gilliamella*. Whereas, *B. ruderatus* and *B. hortorum* shows relatively higher abundance of 'non-core' bacteria such as *Pantoea*. These findings are further supported by the indicator species analysis. *B. ruderatus* is dominated by a wide range of 'non-core' bacteria. *B. hortorum* is similarly dominated by 'non-core' bacteria, but to a lesser extent. In contrast, *B. terrestris* is best characterized by core bacteria that have only been sampled from the guts of social bees. Considering as yet unknown mechanisms, the higher phylogenetic alpha diversity and wider range of 'non-core' microbiome may indicate that *B. ruderatus* is easily invaded by environmental bacteria, which may lead to dysbiosis (Cariveau et al., 2014; Raymann et al., 2018; Zheng et al., 2018). The difference between species were further reinforced and extended by the association found between host genetic distance and microbial diversity. Genetically related *Bombus* species tend to have more similar microbial communities. These findings provide compelling evidence of a conserved coevolutionary pattern between the social bees and

their microbiome.

To better understand the interaction of microbiomes in wild bumble bees, co-occurrence modules were detected in order to uncover similarities of microbiomes in life style (Faust & Raes, 2016). Two distinct non-random co-occurrence modules were found based on the whole data set, one dominated by core bacteria and another one dominated by environmental bacteria. This result is consistent to the previous finding of two different robust clusters, also called 'Enterotypes'(Arumugam et al., 2011), existed in bumble bee's microbial community (Li et al., 2015). Nonetheless, *Apibacter* was also included in the core bacteria enriched module in this study, indicating that *Apibacter* may share a similar life style with other core bacteria such as *Gilliamella*. However, as a note of caution, the formation of two distinct co-occurrence modules were likely driving by the microbiome difference found between host species to some extent, but the results did show general microbial profiles for a broader range of *Bombus*.

The *Crithidia* and *Nosema* infection incidence caused no significant modulations of the bee's microbiome community, as no difference in alpha, beta diversity were found in my experiments. However, multiple lines of evidence in my results revealed two bacterial taxa were associated with the *Crithidia* and *Nosema* infection: *Bifidobacterium* and *Apibacter*. The biological role of *Apibacter* within the gut remains uncharacterized, despite the fact that it is increasingly found in social bees (Mockler et al., 2018; Kwong & Moran, 2016). In the network analysis, *Apibacter* was proven to show negative association with both *Nosema* and *Crithidia* parasite loads, with the former finding being supported by the result from linear mixed model. Additionally, as *Apibacter* was positively associated with hub taxa Gilliamella, the *Nosema* and *Crithidia* loads might subsequently associated with the 'core' module and the taxa involved. Moreover, from indicator species analysis, *Apibacter* can also characterize both *Crithidia* and *Nosema* uninfected samples with a relatively higher *Apibacter* abundance, indicating that the abundance of *Apibacter* is likely linked to the health of bee. My results offer evidence consistent with another experimental microbiome study (Mockler et al., 2018), pointing out that *Apibacter* plays an important role in the pathogen-modulated microbial communities. *Lactobacillus* was found to have a positive relationship with *Nosema* loads. However, I also noted that my study cannot determine whether these associations are causal, because *Lactobacillus* contains both bee-gut-specific species and environmental species associated with flower products such as nectar and other substrates. The roles of *Gilliamella* and *Snodgrassella* as hub bacteria in the network demonstrated their poten-

tial functional importance with the microbial community, despite the absence of a significant correlation between parasite infection and bacterial abundance, which is consistent with the previous findings (Koch & Schmid-Hempel, 2011; Schwarz et al., 2016). However, the results of my experiments do not permit me to assert that there is a definite mechanism linking the bacteria and the parasites. This is because the 16S sequencing data can only assist me in identifying bacteria at the species level or higher, rather than providing the precise function profile of each microbe. To validate such mechanisms, investigations such as metagenomic sequencing, to fully sample all genes that are present in a particular complex environment sample, might be useful to explore interactions at the gene and protein level.

Further analyses were performed to assess the role of host genetic diversity related to parasites infection. It has been reported in other organisms that a loss of genetic diversity will lead to lower resistance to parasite infection, lower fitness and lower microbial diversity (Ekroth et al., 2019; Wei et al., 2020; Whitehorn et al., 2014). As numerous species of bumble bees are inbred (Bogo et al., 2018) and adverse effects that might be negatively influencing bumble bees fitness have been discovered (Gosterit, 2016; Whitehorn et al., 2009), empirical research focusing on the inbreeding of bumble bees is urgently needed. About 74% samples were infected by *Crithidia* in this study. As bumble bee populations lose heterozygosity, *Crithidia* prevalence was found to increase significantly, indicating that loss of genetic context will aggravate parasite infection, and further push threaten to the whole population. Similarly, previous research showed that genetically diverse populations tend to have lower prevalence of tracheal mite *L. buchneri* in two species of bumble bees (Whitehorn et al., 2014). However, There is no significant correlation between *Nosema* prevalence and loss of genetic context. This might be explained by the limited number of *Nosema* infection (31% samples infected) observed (Whitehorn et al., 2009). Therefore, inbred and geographically isolated population is more vulnerable and susceptible to parasite infection (Whitehorn et al., 2009).

In this study, host species was the variable that most substantially influenced the diversity of the wild bumble bee gut communities and the genetically related *Bombus* species tended to have similar microbial communities. The priority species *B. ruderatus* showed a higher alpha diversity and a more 'non-core' microbiome, indicating the higher possibility of gut niche easily invaded by external microbes and pathogens, which might be a possible reason of the population decline. As pathogen infection was identified as the primary cause of

the population reduction (Figueroa et al., 2021), inbreeding, which can aggravate *Crithidia* infection and exacerbate the threat to the entire population, must be promptly considered when devising conservation strategies for bumble bees. Meanwhile, several key bacteria emerged with significant correlation with parasite infection. When social bees are infected with disease-causing organisms, the bacteria in their guts may experience an indicative dysbiosis state (Koch & Schmid-Hempel, 2011; Cariveau et al., 2014; Maes et al., 2016; Jabal-Uriel et al., 2022), which might be remedied by probiotics tailored to bees (Arredondo et al., 2018; Borges et al., 2021). Two of the key bacteria, *Apibacter* and *Bifidobacterium*, may well be scope for further investigation to exploring probiotic effects. This might be a viable future strategy for reducing the consequence of bumble bee pathogen's infection (Borges et al., 2021). However, more empirical and lab-based experiments are necessary to validate the association. As numerous *Bombus* species have seen precipitous population decreases, increasingly severe pathogen infection with loss of genetic diversity may be to blame (Graystock et al., 2016; Elias, 2022; Whitehorn et al., 2014). Understanding the factors influencing the health of bumble bees is crucial for the sustainability of natural environments. Evaluations of the impact of pathogen exposure on wild bee populations and the approaches for mitigating such stresses need to be implemented with greater urgency.

5 Data Availability Statement

The data and code underlying this study are available from https://github.com/nedchen2/CMEECourseWork/tree/master/MscProject. Raw fastq file are available from my supervisor Peter, Graystock (p.graystock@imperial.ac.uk) .

6 Acknowledgements

I would like to express my deepest appreciation to my supervisor, Peter Graystock for all the care, attention and patience during my project. Special thanks to anonymous colleagues for collecting data for this project. I am also grateful to master's student Jiayue, Wang, whose recommendations considerably enhanced the final version of this project.

7 References

2022. Bees, Wasps & Ants Recording Society. Bombus ruderatus. Accessed: 2022-08-06.

Alexander, N. 2012. Review: analysis of parasite and other skewed counts. *Tropical Medicine & International Health* **17**: 684–693. doi:10.1111/j.1365-3156.2012.02987.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3156.2012.02987.x.

Arredondo, D., Castelli, L., Porrini, M.P., Garrido, P.M., Eguaras, M.J., Zunino, P. & Antúnez, K. 2018. Lactobacillus kunkeei strains decreased the infection by honey bee pathogens Paenibacillus larvae and Nosema ceranae. *Beneficial Microbes* **9**: 279–290. doi:10.3920/BM2017.0075.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Doré, J., Weissenbach, J., Ehrlich, S.D. & Bork, P. 2011. Enterotypes of the human gut microbiome. *Nature* **473**: 174–180. doi:10.1038/nature09944. Number: 7346 Publisher: Nature Publishing Group.

Assenov, Y., Ramírez, F., Schelhorn, S.E., Lengauer, T. & Albrecht, M. 2008. Computing

topological parameters of biological networks. *Bioinformatics* **24**: 282–284. doi:10.1093/bioinformatics/btm554.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**: 1–48. doi:10.18637/jss.v067.i01.

Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**: 289–300. Publisher: [Royal Statistical Society, Wiley].

Berry, D. & Widder, S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology* **5**: 219. doi:10.3389/fmicb.2014.00219.

Bogo, G., de Manincor, N., Fisogni, A., Galloni, M., Zavatta, L. & Bortolotti, L. 2018. No evidence for an inbreeding avoidance system in the bumble bee Bombus terrestris. *Apidologie* **49**: 473–483. doi:10.1007/s13592-018-0575-1.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A. & Gregory Caporaso, J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**: 90. doi:10.1186/s40168-018-0470-z.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K., Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M., Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K.B., Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A.V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von

Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R. & Caporaso, J.G. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* **37**: 852–857. doi:10.1038/s41587-019-0209-9. Number: 8 Publisher: Nature Publishing Group.

Borges, D., Guzman-Novoa, E. & Goodwin, P.H. 2021. Effects of Prebiotics and Probiotics on Honey Bees (Apis mellifera) Infected with the Microsporidian Parasite Nosema ceranae. *Microorganisms* **9**: 481. doi:10.3390/microorganisms9030481.

Brown, M.B. 1975. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics* **31**: 987–992. doi:10.2307/2529826. Publisher: [Wiley, International Biometric Society].

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**: 581–583. doi:10.1038/nmeth.3869. Number: 7 Publisher: Nature Publishing Group.

Cariveau, D.P., Elijah Powell, J., Koch, H., Winfree, R. & Moran, N.A. 2014. Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (Bombus). *The ISME Journal* **8**: 2369–2379. doi:10.1038/ismej.2014.68. Number: 12 Publisher: Nature Publishing Group.

CATCHEN, J., HOHENLOHE, P.A., BASSHAM, S., AMORES, A. & CRESKO, W.A. 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology* **22**: 3124–3140. doi:10.1111/mec.12354.

Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, Z. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* **13**: 79–90. doi:https://doi.org/10.1111/j.1654-1103.2002.tb02025.x.

Corby-Harris, V., Maes, P. & Anderson, K.E. 2014. The Bacterial Communities Associated with Honey Bee (Apis mellifera) Foragers. *PLOS ONE* **9**: e95056. doi:10.1371/journal.pone.0095056. Publisher: Public Library of Science.

Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.L., Briese, T., Hornig, M., Geiser, D.M., Martinson, V., vanEngelsdorp, D., Kalkstein, A.L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S.K., Simons, J.F., Egholm, M.,

Pettis, J.S. & Lipkin, W.I. 2007. A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. *Science* **318**: 283–287. doi:10.1126/science.1146498. Publisher: American Association for the Advancement of Science.

Cáceres, M.D. & Legendre, P. 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566–3574. doi:10.1890/08-1823.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1890/08-1823.1.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. & 1000 Genomes Project Analysis Group 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158. doi:10.1093/bioinformatics/btr330.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. & Li, H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**. doi:10.1093/gigascience/giab008. Giab008.

De Cáceres, M., Legendre, P. & Moretti, M. 2010. Improving indicator species analysis by combining groups of sites. *Oikos* **119**: 1674–1684. doi:10.1111/j.1600-0706.2010.18334.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0706.2010.18334.x.

Dormann, C.F., Frueund, J., Bluethgen, N. & Gruber, B. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal* **2**: 7–24.

Dray, S. & Dufour, A.B. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software* **22**: 1–20. doi:10.18637/jss.v022.i04.

Ekroth, A.K.E., Rafaluk-Mohr, C. & King, K.C. 2019. Host genetic diversity limits parasite success beyond agricultural systems: a meta-analysis. *Proceedings of the Royal Society B: Biological Sciences* **286**: 20191811. doi:10.1098/rspb.2019.1811.

Elias, S.A. 2022. Plight of the Bumblebees. In: *Imperiled: The Encyclopedia of Conservation* (D.A. DellaSala & M.I. Goldstein, eds), pp. 549–565. Elsevier, Oxford. doi: https://doi.org/10.1016/B978-0-12-821139-7.00240-3.

Emsen, B., De la Mora, A., Lacey, B., Eccles, L., Kelly, P.G., Medina-Flores, C.A., Petukhova, T., Morfin, N. & Guzman-Novoa, E. 2020. Seasonality of Nosema ceranae Infections and Their Relationship with Honey Bee Populations, Food Stores, and Survivorship in a North American Region. *Veterinary Sciences* **7**: 131. doi:10.3390/vetsci7030131. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Faith, D.P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**: 1–10. doi:10.1016/0006-3207(92)91201-3.

Faust, K. & Raes, J. 2012. Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**: 538–550. doi:10.1038/nrmicro2832. Number: 8 Publisher: Nature Publishing Group.

Faust, K. & Raes, J. 2016. CoNet app: inference of biological association networks using Cytoscape. *F1000Research* **5**: 1519. doi:10.12688/f1000research.9050.2.

Figueroa, L.L., Grincavitch, C. & McArt, S.H. 2021. Crithidia bombi can infect two solitary bee species while host survivorship depends on diet. *Parasitology* **148**: 435–442. doi:10.1017/S0031182020002218. Publisher: Cambridge University Press.

Frederick, W.G., Sedlmeyer, R.L. & White, C.M. 1993. The Hamming metric in genetic algorithms and its application to two network problems. In: *Proceedings of the 1993 ACM/SIGAPP symposium on Applied computing: states of the art and practice*, SAC '93, pp. 126–130. Association for Computing Machinery, New York, NY, USA. doi:10.1145/162754.162835.

Gosterit, A. 2016. Adverse effects of inbreeding on colony foundation success in bumblebees, Bombus terrestris (Hymenoptera: Apidae). *Applied Entomology and Zoology* **51**: 521–526. doi:10.1007/s13355-016-0427-2.

Graystock, P., Meeus, I., Smagghe, G., Goulson, D. & Hughes, W.O.H. 2016. The effects of single and mixed infections of Apicystis bombi and deformed wing virus in Bombus terrestris. *Parasitology* **143**: 358–365. doi:10.1017/S0031182015001614. Publisher: Cambridge University Press.

Grupe, A.C. & Quandt, C.A. 2020. A growing pandemic: A review of Nosema parasites in globally distributed domesticated and native bees. *PLoS Pathogens* **16**: e1008580. doi:10.1371/journal.ppat.1008580.

Halme, P., Mönkkönen, M., Kotiaho, J.S., Ylisirniö, A.L. & Markkanen, A. 2009. Quantifying the indicator power of an indicator species. *Conservation Biology: The Journal of the Society for Conservation Biology* **23**: 1008–1016. doi:10.1111/j.1523-1739.2009.01206.x.

Huang, Q. & Evans, J.D. 2020. Targeting the honey bee gut parasite Nosema ceranae with siRNA positively affects gut bacteria. *BMC Microbiology* **20**: 258. doi:10.1186/s12866-020-01939-9.

Isagi, Y., Makino, T., Hamabata, T., Cao, P.L., Narita, S., Komaki, Y., Kurita, K., Naiki, A., Kameyama, Y., Kondo, T. & Shibabayashi, M. 2020. Significant loss of genetic diversity and accumulation of deleterious genetic variation in a critically endangered azalea species, Rhododendron boninense, growing on the Bonin Islands. *Plant Species Biology* **35**: 166–174. doi:10.1111/1442-1984.12270. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1442-1984.12270.

Jabal-Uriel, C., Alba, C., Higes, M., Rodríguez, J.M. & Martín-Hernández, R. 2022. Effect of Nosema ceranae infection and season on the gut bacteriome composition of the European honeybee (Apis mellifera). *Scientific Reports* **12**: 9326. doi:10.1038/s41598-022-13337-4. Number: 1 Publisher: Nature Publishing Group.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066. doi:10.1093/nar/gkf436.

Koch, H., Brown, M.J. & Stevenson, P.C. 2017. The role of disease in bee foraging ecology. *Current Opinion in Insect Science* **21**: 60–67. doi:10.1016/j.cois.2017.05.008.

Koch, H. & Schmid-Hempel, P. 2011. Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proceedings of the National Academy of Sciences* **108**: 19288–19292. doi:10.1073/pnas.1110474108. Publisher: Proceedings of the National Academy of Sciences.

Koch, H. & Schmid-Hempel, P. 2012. Gut microbiota instead of host genotype drive the specificity in the interaction of a natural host-parasite system. *Ecology Letters* **15**: 1095–1103. doi:10.1111/j.1461-0248.2012.01831.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2012.01831.x.

Krasnov, B.R., Spickett, A., Junker, K., Bugmyrin, S.V., Ieshko, E.P., Bespyatova, L.A., Stanko, M., Khokhlova, I.S. & Matthee, S. 2021. Parasite counts or parasite incidences? Testing differences with four analyses of infracommunity modelling for seven parasite–host associations. *Parasitology Research* **120**: 2569–2584. doi:10.1007/s00436-021-07217-5.

Kurze, C., Routtu, J. & Moritz, R.F.A. 2016. Parasite resistance and tolerance in honeybees at the individual and social level. *Zoology* **119**: 290–297. doi:10.1016/j.zool.2016.03.007.

Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* **82**: 1–26. doi:10.18637/jss.v082.i13.

Kwong, W.K., Medina, L.A., Koch, H., Sing, K.W., Soh, E.J.Y., Ascher, J.S., Jaffé, R. & Moran, N.A. 2017. Dynamic microbiome evolution in social bees. *Science Advances* **3**: e1600513. doi:10.1126/sciadv.1600513. Publisher: American Association for the Advancement of Science.

Kwong, W.K. & Moran, N.A.. 2013. Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of Snodgrassella alvi gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria, and Gilliamella apicola gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order 'Enterobacteriales' of the Gammaproteobacteria. *International Journal of Systematic and Evolutionary Microbiology* **63**: 2008–2018. doi:10.1099/ijs.0.044875-0. Publisher: Microbiology Society,.

Kwong, W.K. & Moran, N.A. 2016. Gut Microbial Communities of Social Bees. *Nature reviews. Microbiology* **14**: 374–384. doi:10.1038/nrmicro.2016.43.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. doi:10.48550/arXiv.1303.3997. ArXiv:1303.3997 [q-bio].

Li, J., Powell, J.E., Guo, J., Evans, J.D., Wu, J., Williams, P., Lin, Q., Moran, N.A. & Zhang, Z. 2015. Two gut community enterotypes recur in diverse bumblebee species. *Current Biology* **25**: R652–R653. doi:10.1016/j.cub.2015.06.031.

Li, M., Li, D., Tang, Y., Wu, F. & Wang, J. 2017. CytoCluster: A Cytoscape Plugin for Cluster Analysis and Visualization of Biological Networks. *International Journal of Molecular Sciences* **18**: 1880. doi:10.3390/ijms18091880.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A.M., Coppola, L., Cornejo-Castillo, F.M., d'Ovidio, F., De Meester, L., Ferrera, I., Garet-Delmas, M.J., Guidi, L., Lara, E., Pesant, S., Royo-Llonch, M., Salazar, G., Sánchez, P., Sebastian, M., Souffreau, C., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans coordinators, Gorsky, G., Not, F., Ogata, H., Speich, S., Stemmann, L., Weissenbach, J., Wincker, P., Acinas, S.G., Sunagawa, S.,

Bork, P., Sullivan, M.B., Karsenti, E., Bowler, C., de Vargas, C. & Raes, J. 2015. Determinants of community structure in the global plankton interactome. *Science* **348**: 1262073. doi:10.1126/science.1262073. Publisher: American Association for the Advancement of Science.

Lozupone, C.A., Hamady, M., Kelley, S.T. & Knight, R. 2007. Quantitative and Qualitative Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology* **73**: 1576–1585. doi:10.1128/AEM.01996-06.

Ma, B., Wang, Y., Ye, S., Liu, S., Stirling, E., Gilbert, J.A., Faust, K., Knight, R., Jansson, J.K., Cardona, C., Röttjers, L. & Xu, J. 2020. Earth microbial co-occurrence network reveals interconnection pattern across microbiomes. *Microbiome* **8**: 82. doi:10.1186/s40168-020-00857-2.

Maes, P.W., Rodrigues, P.A.P., Oliver, R., Mott, B.M. & Anderson, K.E. 2016. Diet-related gut bacterial dysbiosis correlates with impaired development, increased mortality and Nosema disease in the honeybee (Apis mellifera). *Molecular Ecology* **25**: 5439–5450. doi:10.1111/mec.13862.

Martinson, V.G., Moy, J. & Moran, N.A. 2012. Establishment of Characteristic Gut Bacteria during Development of the Honeybee Worker. *Applied and Environmental Microbiology* **78**: 2830–2840. doi:10.1128/AEM.07810-11. Publisher: American Society for Microbiology.

McMurdie, P.J. & Holmes, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**: e61217. doi:10.1371/journal.pone.0061217. Publisher: Public Library of Science.

Mockler, B.K., Kwong, W.K., Moran, N.A. & Koch, H. 2018. Microbiome Structure Influences Infection by the Parasite Crithidia bombi in Bumble Bees. *Applied and Environmental Microbiology* **84**: e02335–17. doi:10.1128/AEM.02335-17. Publisher: American Society for Microbiology.

Moro, A., Blacquière, T., Panziera, D., Dietemann, V. & Neumann, P. 2021. Host-Parasite Co-Evolution in Real-Time: Changes in Honey Bee Resistance Mechanisms and Mite Reproductive Strategies. *Insects* **12**: 120. doi:10.3390/insects12020120. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J. & Weedon, J. 2022. *vegan: Community Ecology Package*. R package version 2.6-2.

Potts, S.G., Biesmeijer, J.C., Kremen, C., Neumann, P., Schweiger, O. & Kunin, W.E. 2010. Global pollinator declines: trends, impacts and drivers. *Trends in Ecology & Evolution* **25**: 345–353. doi:10.1016/j.tree.2010.01.007.

Price, M.N., Dehal, P.S. & Arkin, A.P. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**: e9490. doi:10.1371/journal.pone.0009490.

Purcell, S. 2022. PLINK: Whole genome data analysis toolset. Available from: http://pngu.mgh.harvard.edu/purcell/plink/.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., de Bakker, P., Daly, M. & Sham, P. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* **81**: 559–575.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F.O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**: D590–D596. doi:10.1093/nar/gks1219.

R Core Team 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, S., Hoffman, G., Kirby, J. & Horne, D. 2019. Remarkable long-distance returns to a forage patch by artificially displaced wild bumble bees (Hymenoptera: Apidae). *Journal of Apicultural Research* **58**: 522–530. doi:10.1080/00218839.2019.1584962. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00218839.2019.1584962.

Raymann, K., Bobay, L.M. & Moran, N.A. 2018. Antibiotics reduce genetic diversity of core species in the honeybee gut microbiome. *Molecular ecology* **27**: 2057–2066. doi:10.1111/mec.14434.

Schoonvaere, K., Brunain, M., Baeke, F., De Bruyne, M., De Rycke, R. & de Graaf, D.C. 2020. Comparison between Apicystis cryptica sp. n. and Apicystis bombi (Arthrogregarida, Apicomplexa): Gregarine parasites that cause fat body hypertrophism in bees. *European Journal of Protistology* **73**: 125688. doi:10.1016/j.ejop.2020.125688.

Schwarz, R.S., Moran, N.A. & Evans, J.D. 2016. Early gut colonizers shape parasite susceptibility and microbiota composition in honey bee workers. *Proceedings of the National Academy of Sciences* **113**: 9345–9350. doi:10.1073/pnas.1606631113. Publisher: Proceedings of the National Academy of Sciences.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**: 2498–2504. doi:10.1101/gr.1239303.

Simone, M., Evans, J.D. & Spivak, M. 2009. Resin Collection and Social Immunity in Honey Bees. *Evolution* **63**: 3016–3022. doi:10.1111/j.1558-5646.2009.00772.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1558-5646.2009.00772.x.

Stone, L. & Roberts, A. 1990. The checkerboard score and species distributions. *Oecologia* **85**: 74–79. doi:10.1007/BF00317345.

Tian, T., Piot, N., Meeus, I. & Smagghe, G. 2018. Infection with the multi-host micro-parasite Apicystis bombi (Apicomplexa: Neogregarinorida) decreases survival of the solitary bee Osmia bicornis. *Journal of Invertebrate Pathology* **158**: 43–45. doi:10.1016/j.jip.2018.09.005.

Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., Simões, Z.L.P., Allsopp, M.H., Kandemir, I., De la Rúa, P., Pirk, C.W. & Webster, M.T. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee Apis mellifera. *Nature Genetics* **46**: 1081–1088. doi:10.1038/ng.3077. Number: 10 Publisher: Nature Publishing Group.

Wang, J., Ren, J., Li, M. & Wu, F.X. 2012. Identification of Hierarchical and Overlapping Functional Modules in PPI Networks. *IEEE Transactions on NanoBioscience* **11**: 386–393. doi:10.1109/TNB.2012.2210907. Conference Name: IEEE Transactions on NanoBioscience.

Wei, L., Zeng, B., Zhang, S., Li, F., Kong, F., Ran, H., Wei, H.J., Zhao, J., Li, M. & Li, Y. 2020. Inbreeding Alters the Gut Microbiota of the Banna Minipig. *Animals : an Open Access Journal from MDPI* **10**: 2125. doi:10.3390/ani10112125.

Whitehorn, P.R., Seo, B., Comont, R.F., Rounsevell, M. & Brown, C. 2022. The effects of climate and land use on British bumblebees: Findings from a decade of citizen-science observations. *Journal of Applied Ecology* **59**: 1837–1851. doi:10.1111/1365-2664.14191. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2664.14191.

Whitehorn, P.R., Tinsley, M.C., Brown, M.J., Darvill, B. & Goulson, D. 2009. Impacts of inbreeding on bumblebee colony fitness under field conditions. *BMC Evolutionary Biology* **9**: 152. doi:10.1186/1471-2148-9-152.

Whitehorn, P.R., Tinsley, M.C., Brown, M.J.F., Darvill, B. & Goulson, D. 2014. Genetic diversity and parasite prevalence in two species of bumblebee. *Journal of Insect Conservation* **18**: 667–673. doi:10.1007/s10841-014-9673-1.

Williams, P. 2005. Does specialization explain rarity and decline among British bumblebees? A response to Goulson et al. *Biological Conservation* **122**: 33–43. doi:10.1016/j.biocon.2004.06.019.

Williams, P.H., Araújo, M.B. & Rasmont, P. 2007. Can vulnerability among British bumblebee (Bombus) species be explained by niche position and breadth? *Biological Conservation* **138**: 493–505. doi:10.1016/j.biocon.2007.06.001.

Wu, Y., Zheng, Y., Chen, Y., Wang, S., Chen, Y., Hu, F. & Zheng, H. 2020. Honey bee (Apis mellifera) gut microbiota promotes host endogenous detoxification capability via regulation of P450 gene expression in the digestive tract. *Microbial Biotechnology* **13**: 1201–1212. doi:10.1111/1751-7915.13579.

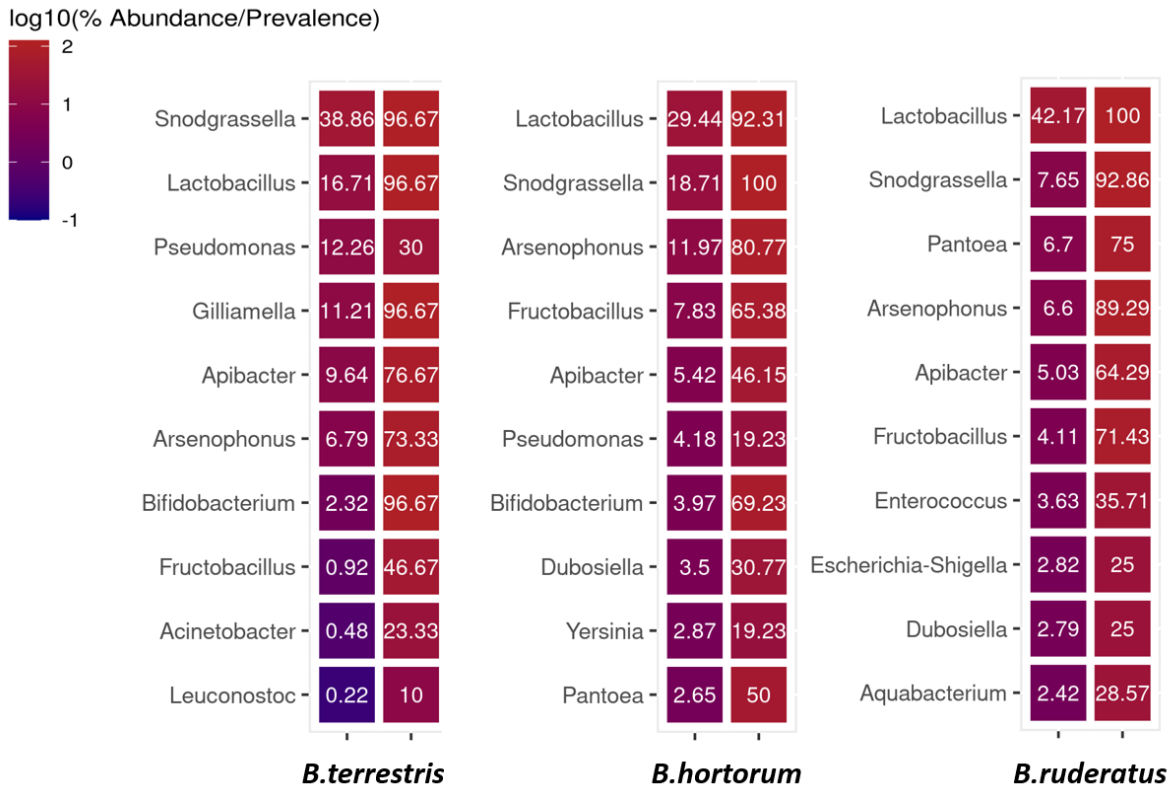
Zheng, H., Perreau, J., Powell, J.E., Han, B., Zhang, Z., Kwong, W.K., Tringe, S.G. & Moran, N.A. 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proceedings of the National Academy of Sciences* **116**: 25909–25916. doi:10.1073/pnas.1916224116. Publisher: Proceedings of the National Academy of Sciences.

Zheng, H., Steele, M.I., Leonard, S.P., Motta, E.V.S. & Moran, N.A. 2018. Honey bees as models for gut microbiota research. *Lab animal* **47**: 317–325. doi:10.1038/s41684-018-0173-x.

8 Supplementary Information

Supplementary Table 1 Relative abundance of top 10 (Here I rank the order by the whole data set including all three *Bombus* species) abundant bacterial genus in different species. Kruskal-Wallis test were performed to compare bacterial genus. FDR method was employed to correct the p value.

genus	Bh rel.abun	Bh_sd	Br rel.abun	Br_sd	Bt rel.abun	Bt_sd	p value	d.f	stat	Pairwise comparison(p value)		
										Bt_vs_Bh	Bt_vs_Br	Bh_vs_Br
Snodgrassella	18.71	30.75	7.65	18.12	38.86	26.6	0.001	2	16.72	0.014	0	0.69
Lactobacillus	29.44	37.76	42.17	35.23	16.71	24.1	0.012	2	10.29	0.278	0.003	3.56
Arsenophonus	11.97	27.15	6.6	18.15	6.79	23.73	0.229	2	3.16	0.202	0.132	0.05
Pseudomonas	4.18	15.39	1.27	4.25	12.26	27.12	0.38	2	1.94	0.278	0.949	1.88
Apibacter	5.42	19.53	5.03	18.39	9.64	11.21	0.017	2	9.21	0.013	0.036	1.03
Fructobacillus	7.83	21.59	4.11	8.31	0.92	3	0.048	2	6.53	0.211	0.018	0.92
Pantoea	2.65	9.11	6.7	20.38	0.04	0.12	0	2	19.76	0.032	0	4.86
Dubosiella	3.5	15.53	2.79	11.75	0	0.02	0.028	2	7.89	0.013	0.023	0.19
Bifidobacterium	3.97	15.15	1.62	2.64	2.32	3.32	0.004	2	12.9	0.002	0.022	0.13
Gilliamella	1.23	3.18	1.25	2.97	11.21	10.3	0	2	36.99	0	0	0.07



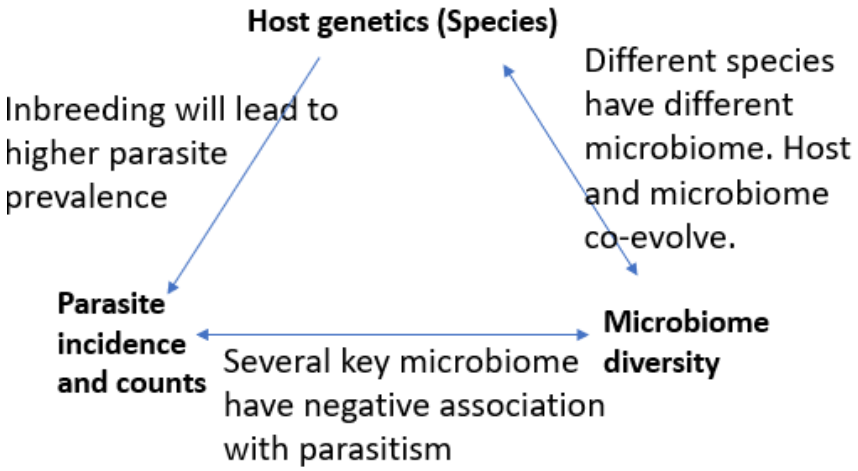
Supplementary Figure 1 Heatmap showing the top10 abundant bacterial genus found in different species (The rank here is ordered by the abundance within one specific host species). In each different species, relative abundance were shown in the left side and bacterial genus prevalence (%) were shown in the right side.

Supplementary Table 2 I use 5 different correlation measures including Pearson (correl_pearson) and Spearman correlation (correl_spearman), mutual information (sim_mutInfo), Bray–Curtis dissimilarity (dist_bray) and Kullback–Leibler dissimilarity (dist_kull). They were further merged by Brown p value merging technique.

Edge	interaction	methodname_score	pval-brown-merge
Nosema-total-per-bee->Apibacter	mutualExclusion	correl_pearson=-0.11|correl_spearman=-0.12 sim_mutInfo=0.068|dist_kull=30.79|dist_bray=0.94	< 0.001
Apibacter->Crithidia-total-per-bee	mutualExclusion	correl_spearman=-0.14|correl_pearson=-0.19 sim_mutInfo=0.17|dist_kull=24.30	< 0.001
Gilliamella->Apibacter	copresence	correl_pearson=0.17|correl_spearman=0.47 dist_kull=7.91|sim_mutInfo=0.32|dist_bray=0.47	< 0.01
Gilliamella->Snodgrassella	copresence	correl_pearson=0.36|correl_spearman=0.58 dist_kull=3.15|dist_bray=0.45|sim_mutInfo=0.43	< 0.001
Snodgrassella->Bifidobacteriaceae	copresence	correl_pearson=0.02|correl_spearman=0.51 dist_bray=0.68|dist_kull=5.66|sim_mutInfo=0.33	< 0.01
Gilliamella->Bifidobacteriaceae	copresence	correl_spearman=0.42 dist_kull=7.58|dist_bray=0.75|sim_mutInfo=0.31	<0.01

Supplementary Table 3 Table demonstrating the parasite incidence and inbreeding coefficient of 11 populations. No *ruderatus* found in Caistor.

population	CrithidiaIncidence	NosemaIncidence	Fis
B.hortorum_Caistor	0.944444444	0.277777778	0.18908
B.hortorum_Hillesden	0.523809524	0.380952381	0.07401
B.hortorum_Ouse	0.944444444	0.388888889	0.09753
B.hortorum_Upton	0.777777778	0.277777778	0.02425
B.ruderatus_Hillesden	0.65	0.3	0.08078
B.ruderatus_Ouse	0.769230769	0.384615385	0.02502
B.ruderatus_Upton	0.857142857	0.428571429	0.06511
B.terrestris_Caistor	0.538461538	0.384615385	0.01636
B.terrestris_Hillesden	0.714285714	0.214285714	0.19661
B.terrestris_Ouse	0.777777778	0.333333333	0.02356
B.terrestris_Upton	0.7	0	0.01204



Supplementary Figure 2 Schematic shows the result of hypothesis.