2021

# Microbiome community and parasitic infections in wild bees

Mark G. Young
*Colby College*

# Microbiome community and parasitic infections in wild bees

Mark Young

Honors Thesis 2021

Colby College Department of Computational Biology

Microbiome community and parasitic infections in wild bees

An Honors Thesis

Presented to

The Faculty of The Department of Biology

Colby College

in partial fulfillment of the requirements for the

Degree of Bachelor of Arts with Honors

by

Mark Young

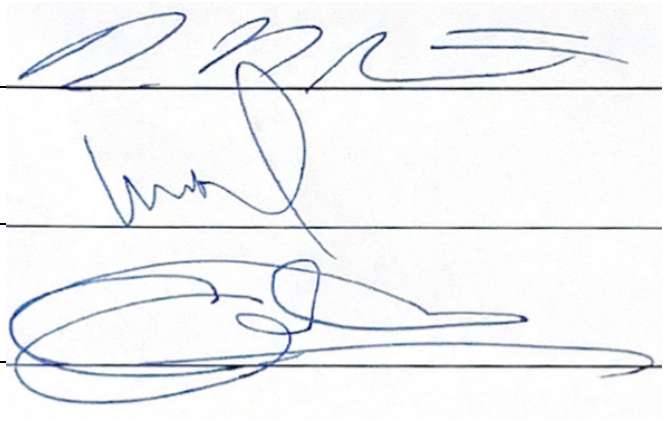
Waterville, ME



Advisor: David Angelini

Reader: Suegene Noh

Reader: Christopher Moore



**CONTENTS**

ABSTRACT

The microbiome is increasingly recognized for its complex relationship with host fitness. Akin to primates and other social animals, bumblebees harbor a specific microbiome derived from social contact. The bumblebee microbiome is characteristically species poor, with just a few "core" phylotypes accounting for the majority of total abundance. Genomic analyses reveal phylogenetic congruence and adaptation of the core endosymbionts to the bumblebee gut, indicating a shared evolutionary history. Prior investigations reveal that the relative abundance of core microbes is negatively associated with infection by a trypanosome, *Crithidia bombi*, and that the microbiome has a mechanistic role in immunity. As an extension of these studies, we have deeply sampled wild bumblebee populations throughout central and down east Maine in a three-year microbiome field survey. The wild bees of the field survey consistently host high abundances of the previously recognized core microbes, as well as low abundant, environmentally derived taxa. Community composition appears to be dependent on colony lifecycle but is largely robust to the horizontal exchange of microbes, supporting the narrative of a robust and coevolved bumblebee microbiome. Compositional variation additionally shows a relationship with *Crithidia bombi* infection severity. Bumblebees with high relative abundances of core microbes have lower infection loads. The relative abundances of two core phylotypes, *Lactobacillus* and *Apibacter*, are particularly strong indicators of host health. However, as relative abundance does not directly reflect absolute abundance, total community composition may be a better indicator of health than the relative abundance of single taxa.

# CHAPTER ONE

# INTRODUCTION

## 1.1 Bumblebees harbor a socially derived microbiome

The *Bombus* genus enjoys a cosmopolitan range, with species found from the tropics of southeast Asia to Maine's down east coast (Cameron et al. 2007; Goulson 2010). Within the greater bee tree of life, *Bombus* shares a clade with *Euglossini,* the orchid bee, *Meliponini,* the stingless bee, and *Apis,* the honeybee (Hines 2008). Together, they are known as the *corbiculate* bees for their hind-leg corbicula, or pollen baskets. *Bombus* has diverged from the other corbiculates as recently as 25-40 mya, with much speciation within subgenera occurring in the last 15 mya (Hines 2008). Species are morphologically similar and overlap in range, but divide into monophyletic short and long-faced clades (Cameron et al. 2007). In comparison to the other corbiculates, *Bombus* species are large, adapted to colder climates, and have primarily dispersed throughout cool alpine and Holarctic regions (Hines 2008; Goulson 2010).

The corbiculate bees receive attention for a significant adaptation, eusociality. Eusociality is the most advanced form of social organization, in which some individuals reduce their own reproductive potential for the benefit of their kin. Eusociality has arisen throughout the animal tree of life, with certain characteristics predisposing organisms for emergent social behavior. Necessary are the formation of nests or groups and the silencing of dispersal behavior. Eusociality further emerges through selection acting on interactions between group members, leading to development of elaborate social behavior and colony lifecycles (Nowak et al. 2010).

Bumblebees, and other eusocial Hymenoptera, are notable for their strictly defined caste systems (Nowak et al. 2010). Colonies have a single reproductive queen and tens to hundreds of

female workers (Goulson 2010). Within the worker caste, there is evidence for behavioral specialization and alloethism. Larger workers excel at foraging while their smaller sisters handle tasks within the nest (O'Donnell et al. 2000; Goulson et al. 2002). Foraging behavior is coordinated, with observed communication between workers (Dornhaus and Chittka 2001). Through coordination, colonies are able to maximize efficiency of resource collection. As indicated by homing experiments, workers are equipped to forage distances up to 10 kilometers from the nest, but are more likely to stay more close by (Goulson and Stout 2001). Surplus resources are diverted to the rearing of reproductive males and daughter queens. Males exist solely for reproductive purposes, lacking pollen baskets for resource collection and stingers for nest defense (Michener 1974). Female larvae chosen to be daughter queens are withheld exposure to a juvenile hormone released by the queen (Röseler 1991). Upon reaching maturity, reproductives leave the colony to mate and begin the lifecycle anew.

Sociality and the development of a robust and coevolved microbiome are interconnected. Endosymbionts can be divided into two general classes, obligate and facultative. Obligates are common in insects and often rely on specialized cell types for endosymbiont storage, known as bacteriocytes (Moran et al. 2008). Bacteriocyte-associated endosymbionts cannot infect naïve hosts and are thus dependent on vertical transmission. In contrast, for facultative, non-bacteriocyte-associated endosymbionts, horizontal transmission through social contact is another method of infection. For social organisms, horizontal transmission may be an even more common mode of infection. Increased population densities increase the numbers of contacts between individuals, providing more opportunities for horizontal exchange of microbes (Ferrari et al. 2011). Additionally, by allowing for reinfection of the same individual, horizontal transmission facilitates opportunities for recombination and gene transfer among strains of the

same phylotype (Moran et al. 2008). As evidence for the dominant role of horizontal transmission in social animals, social networks are predictive of gut microbiome composition in primates (Tung et al. 2015).

Like primates, the bumblebee microbiome appears to be socially derived. Through living in colonies with dense social networks, bumblebees facilitate frequent transmission of microbiota between numerous individuals. As evidence for the prominent role of social transmission, bees lack bacteriocytes and do not receive their microbiome through direct vertical inheritance. Rather, social contact between adults is necessary for colonization by the typical gut microbiome (Koch and Schmid-Hempel 2011). The bee gut undergoes a major reorganization during metamorphosis, so without social contact, it is entirely sterile (Martinson et al. 2012). Additionally, the core bumblebee microbiome shares an evolutionary history with social behavior. The three major corbiculate tribes host a set of five coevolved phylotypes (*Snodgrassella, Gilliamella, Lactobacillus* Firm-4, *Lactobacillus* Firm-5, *Bifidobacterium*) that are not observed in closely related, non-eusocial outgroups. By parsimony, it is likely that these core phylotypes were present with the most recent ancestor of the social bees and are adapted for endosymbiosis of social organisms (Kwong et al. 2017).

Besides direct contact within the nest, foraging provides a route for transmission of microbes between hosts. When pollinators visit flowers, they deposit and pick up microbiota (Keller et al. 2021). Bumblebees do not pollinate randomly, exhibiting ability to discriminate between the quality of flowers and consistently seek out species providing high rewards (Müller 1883; Cresswell and Robertson 1994; Chittka et al. 1997). Pollination preference is a mutually beneficial trait for both pollinators and flowers. It increases foraging efficiency and the likelihood that plants will receive conspecific pollen. Analogously, it increases the likelihood

that deposited microbiota will be picked up by a coevolved host. Floral transmission has been observed indirectly in two ways. First, the composition of floral microbiomes is dependent on the species composition of pollinators, indicating that pollinators deposit microbiota (Zemenick et al. 2021). Second, gut microbiome-adapted bacteria are able to persist on flowers, indicating that coevolved taxa are transmitted through pollination (McFrederick et al. 2017).

Despite the potential for the exchange of microbiota between sympatric bee species, the bumblebee microbiome is distinct from the other social bees. Besides the corbiculate core microbiota, bumblebees have unique symbioses with *Apibacter, Lactobacillus* Firm-3, *Bombiscardovia,* and *Schmidhempelia* (Kwong et al. 2017). Bumblebees, honeybees, and stingless bees are differentiable by microbiome community composition, and there is a strong relationship between pairwise community dissimilarity and host divergence (Kwong et al. 2017). At the strain level, the phylogenies of *Apibacter, Snodgrassella,* and *Gilliamella* show significant correlations with their hosts' phylogeny (Koch et al. 2013; Kwong et al. 2017). In contrast, host geographic range shows minimal or no correlation with the phylogenies of *Gilliamella* and *Snodgrassella,* indicating host specificity (Koch et al. 2013; Kwong et al. 2017). Although *Apis* and *Bombus* specific strains of *Snodgrassella* are able to colonize non-native hosts under laboratory conditions, they are unsuccessful at achieving high abundances and are rapidly outcompeted by native strains, even at numerical advantages of 10:1 (Kwong et al. 2014). Due to their shared history, it is likely that the other bumblebee core microbiota would exhibit similar host specificity, if investigated.

### 1.2 The microbiome is central to host health

Failure rates of bumblebee colonies are high. With rare exceptions, bumblebees have

annual life cycles, with queens founding colonies each spring. (Goulson 2010). The success of a queen is measured in her ability to found a colony of workers industrious enough to support a new generation of males and queens (Heinrich 1979). In historical field surveys, success rates of as low as 15 to 30 percent are observed (Cumber 1953; Müller and Schmid-Hempel 1993). The high energetic cost of the bumblebee life cycle may be responsible for these observed rates. While flying, bees consume between 40 and 70 ml $O_2$/gram/h (Ellington et al. 1990), corresponding to a metabolic rate 75% higher than a hummingbird and 120 times that of a jogging human male (Heinrich 1996). Before gaining experience, young workers frequently return from foraging lighter than when they left (Peat and Goulson 2005). Colonies survive on thin margins and store little food. As short as a week without efficient foraging is likely to be disastrous (Goulson 2010).

A healthy gut microbiome may increase the efficiency of individual workers, benefitting the entire colony. Primarily, the microbiome plays a role in digestion. Bumblebees survive exclusively on pollen and nectar that they gather from a broad range of native plants, agricultural crops, and non-native garden flowers (Michener 1974; Goulson et al. 2004). Nectar contains carbohydrates necessary for adult workers, while pollen is a rich protein source for the developing brood (Goulson 2010). Correspondingly, the adult bee gut metagenome is significantly enriched for orthologs annotated for carbohydrate metabolism and transport (Engel et al. 2012). Likely, these aid in the digestion of the sugars found in nectar and pollen. Additionally, the metagenome shows enrichment for classes of proteins that may support resistance to plant-derived antimicrobial compounds and aid in metabolism of various toxic or non-bioavailable nectar and pollen-derived sugars (Engel et al. 2012). Through their metabolism of these compounds, endosymbionts can detoxify and increase nutritional density of their host

diets.

The gut microbiome also plays a role in host immunity. The two most abundant endosymbionts of honeybees and bumblebees, *Snodgrassella* and *Gilliamella*, form a biofilm along the epithelium of the hindgut (Martinson et al. 2012). *Snodgrassella* directly associates with host tissue and is covered with a thick layer of *Gilliamella.* Metabolically, the two may partition resources in a syntrophic network. *Gilliamella* is a facultative anaerobe possessing genes for saccharolytic fermentation while *Snodgrassella* possesses a complete TCA cycle for oxidation of carboxylic acids (Kwong et al. 2014). Both show significant reduction in genome size and horizontal exchange of genes involved in bacteriophage defense, indicating an evolutionary history of host specific endosymbiosis (Kwong et al. 2014). *Snodgrassella* is additionally enriched for RTX proteins (Kwong et al. 2014). RTX proteins are involved with host cell interaction and may additionally form a protective surface layer between host cells and pathogens (Satchell 2011). Thus, the biofilm most likely plays a role in modulation of the host immune system, as well as in physical defense.

Other core microbiota may also play a role in biofilm development. The genomes of bee-associated *Apibacter, Lactobacillus, Bifidobacterium,* and *Schmidhempelia* all show enrichment for genes putatively involved in biofilm formation (Martinson et al. 2014; Ellegaard et al. 2015; Kwong et al. 2018). Although these four phylotypes have not yet been directly observed in a biofilm, genomic evidence is suggestive of their niches within the gut. Like *Snodgrassella, Apibacter* possesses a complete TCA cycle, indicating preference for aerobic portions of the gut (Engel et al. 2012; Kwong et al. 2018). Like *Gilliamella, Schmidhempelia* is equipped for saccharolytic fermentation in anaerobic environments (Engel et al. 2012; Martinson et al. 2014). *Bifidobacterium* also possesses genes for aerobic fermentation, but *Lactobacillus* phylotypes are

more variable (Ellegaard et al. 2015). Likely, core phylotypes dependent on aerobic respiration associate closely with host tissue, while anaerobic fermenters radiate further into the lumen.

### 1.3 Microbiome community is associated with trypanosome infection outcomes

Bumblebee populations are held in check by numerous parasites and parasitoids. Protozoa cause the most well understood bumblebee diseases, but nematodes, fungi, mites, prokaryotes, and viruses have been implicated in colony health (Goulson 2010; Pascall et al. 2021). In a natural experiment, four European bumblebee species were introduced to New Zealand in 1885, without most of their native parasites (Donovan and Wier 1984). In the years since, two have gone extinct in their native ranges, while the parasite-free, invasive populations flourish at greater densities than ever observed in Europe (Goulson 2010).

Behavior may predispose bumblebees to the development of host parasite systems. A common trypanosome, *Crithidia bombi*, is spread through ingestion of parasite cells found in feces (Schmid-Hempel and Schmid-Hempel 1993). Bees live in densely populated nests, facilitating rapid social transmission (Otterstatter and Thomson 2007). Bees are further exposed to inter-colony infection through foraging of the same flowers (Durrer and Schmid-Hempel 1994). By the end of a summer, *Crithidia bombi* can reach within-colony infection rates of 100%, and analyses of field populations reveal that up to 80% of colonies in an area can become infected. (Shykoff and Schmid-Hempel 1991; Imhoof and Schmid-Hempel 1999). *Crithidia* shows signs of interaction with host genotype, spreading faster through genetically homogenous hosts (Yourth and Schmid-Hempel 2006). Naïve host populations have greater mortality rates, indicating antagonistic coevolution and strain specific interactions (Imhoof and Schmid-Hempel 1998; Brown et al. 2003).

As a gut parasite, *Crithidia* has particularly direct interactions with the host microbiome. The sterilization of the bumblebee gut during metamorphosis provides a route for non-invasive microbiome transplants and comparison of infection outcomes between gut communities. In a 2011 study, *C. bombi* infection severity of workers with typical microbiota was compared to that of others raised in sterile environments or colonized with just *Gilliamella* (Koch and Schmid-Hempel 2011). The three groups were challenged by a mixture of *C. bombi* strains. Workers with typical microbiota had significantly less diversity and lower total abundance of *C. bombi* than the sterile workers. In particular, relative abundances of *Gilliamella* and *Snodgrassella* showed negative relationships with infection severity. However, workers colonized by just *Gilliamella* showed no differences from the sterile workers, indicating that the microbiome-pathogen interaction may be more than the sum of its parts.

Mockler *et al.* used a similar experimental methodology to test the relationship between microbial abundances and *C. bombi* infection severity (Mockler et al. 2018). Newly eclosed workers were transferred microbiota from combinations of wild and commercial bees. Filtrate was used as a control. All microbiome treatment groups conferred significant *C. bombi* resistance in comparison to the control. Total diversity and bacterial load, as well as the relative abundances of core microbiota showed negative correlations with infection severity. *Apibacter* had the most significant correlation, followed by *Lactobacillus* Firm-5, and *Gilliamella.* Abundance of *Snodgrassella* did not show a relationship with infections. Interestingly, the composition of posttreatment microbiomes closely resembled their respective inoculum, indicating that composition did not change with *C. bombi* infection. Thus, it is likely that the microbiome has a causative relationship with *C. bombi* infection severity, as it does not appear that *C. bombi* is responsible for the observed variation.

The predictions of experimental outcomes are observed in wild populations. Cariveau *et al.,* found that relative abundances of *Gilliamella* were negatively associated with infection by *C. bombi* (Cariveau et al. 2014). In addition, they found that the richness of "non-core" microbiota was positively associated with infection. However, total bacterial abundance may confound this relationship. It's unclear if bees with high non-core richness truly were colonized by high numbers of non-core microbiota, or if an absence of core microbiota facilitated deeper sequencing of the non-core population (Cariveau et al. 2014).

The mechanism of immunity is unclear but may be linked to biofilm produced by core microbiota. From the results of Mockler *et al.,* it is apparent that the bumblebee microbiome has a mechanistic role in host immunity (Mockler et al. 2018). The biofilm produced by *Snodgrassella* and *Gilliamella* lines the epithelium of the hind gut, making it a potential deterrent to *C. bombi* infection (Martinson et al. 2012). However, it's unclear how the biofilm prevents infection, if other core microbiota contribute to it, and why different studies have implicated different core microbiota as most associated with infection outcomes.

Regardless of mechanism, the association between microbiome composition and *Crithidia* infection outcomes is indicative of host-endosymbiont coevolution. As defined by John Thompson, diversifying coevolution is the process of reciprocal adaptations between interacting species contributing to eventual diversification (Thompson 1989). From phylogenetic analyses, this co-diversification is clearly present (Kwong et al. 2017). On its own however, phylogenetic congruence is not sufficient evidence for coevolutionary diversification. For host-microbe systems, host-filtering is an adequate null model for phylosymbiosis (Groussin et al. 2020).

Evidence for coevolution is provided by fitness dependencies and the results of reciprocal transplants (Althoff et al. 2014; Groussin et al. 2020). Fitness dependency is clearly

demonstrated by the association between microbiome community and host health. Hosts with atypical microbiota are significantly more susceptible to *Crithidia* infection, indicating adaptation of the host immune system to its typical endosymbionts (Koch and Schmid-Hempel 2011; Mockler et al. 2018). Reciprocally, corbiculate core-microbiota have adapted to their hosts, as indicted by the host preference of honeybee and bumblebee native strains. Non-native strains fail to achieve high abundances and are rapidly outcompeted by their native counterparts, even at numerical advantages (Kwong et al. 2014). If host filtering was solely responsible for phylogenetic congruence, these reciprocal fitness effects would likely not be present (Groussin et al. 2020). Reciprocal adaptations is also apparent between endosymbionts. *Snodgrassella* and *Gilliamella* show signs of genomic reduction and specialization to complementary niches within the bumblebee gut (Kwong et al. 2014). The evidence for fitness dependencies between bees and their endosymbionts supports the central role of the microbiome to host fitness.

### 1.4 Inference on wild populations

There are fewer bumblebees than there used to be. Observational ecological studies show 20$^{th}$ century declines in species richness and range throughout Europe and North America (Kosior et al. 2007; Colla and Packer 2008; Goulson et al. 2008; Williams and Osborne 2009; Cameron et al. 2011). Individual species are not affected equally. Rare species tend to be long tongued with narrower diets (Goulson et al. 2006). The combined stresses of parasites, pesticides, and habitat loss appear to share responsibility for observed declines (Goulson et al. 2015).

Modern agricultural intensification has increased natural habitat fragmentation, leading to a loss of biodiversity (Wilcove et al. 1998). Although many agricultural crops provide

pollination rewards for bumblebees, foraging economics require that they must be within a short distance of natural nesting habitats. For Californian tomato fields, this was measured to be just 300 meters, with 40% of the surrounding 2 kilometers suitable for nesting (Greenleaf and Kremen 2006). Monocultures enabled by modern fertilizers may also decrease the usefulness of agriculture for wild bees. Long-tongued species especially depend on leguminous rotations, which are made obsolete by fertilizer (Goulson and Darvill 2004).

Fragmented populations are disconnected and effectively smaller, making them more vulnerable to be pushed to local extinction through stochastic processes (Caughley 1994). Haplodiploid sex determination and monoandrous eusociality (Schmid-Hempel and Schmid-Hempel 2000) make bumblebees especially vulnerable to local extinction. Effective population size is just 1.5 times the number of colonies. Inbred queens have decreased success in colony formation and hibernation (Gerloff and Schmid-Hempel 2005). Diploid males arise in inbred populations. Unlike in some other haplodiploid hymenopterans, these are reared to adulthood in bumblebee colonies (Duchateau et al. 1994). Diploid males are sterile, expensive to rear, and provide no return on investment for the colony (Packer and Owen 2001). Distances as short as 10 km may cause significant fragmentation, as indicated by studies of microsatellite frequencies in a declining British species (Darvill et al. 2006).

Agriculture also brings exposure to systemic pesticides through nectar and pollen consumption (Dively and Kamel 2012). Neonicotinoids, the most widely used class of insecticide, account for roughly one third of the global market (Simon-Delso et al. 2015). Even sublethal exposures have a marked effect on pollinators. Exposed individuals have effected physiology, behavior, and learning ability (Desneux et al. 2007). In simple homing experiments, honeybees have significantly less success at returning to the nest when exposed to neonicotinoids

(Henry et al. 2012). These disturbances to worker efficiency and lifespan translate to a loss of resources for the colony as a whole. Exposed bumblebee colonies have slower growth rates and produce few queens (Whitehorn et al. 2012).

Anthropogenic threats most likely exacerbate existing bumblebee diseases. Mounting an immune response is extremely metabolically costly (Moret and Schmid-Hempel 2000). Thus, concurrent pressures on host health may prove to be overwhelming. Virulence of *Crithidia bombi* infections has already been shown to be context dependent, increasing in severity during stressful life stages (Brown et al. 2003). This context dependent virulence can be replicated experimentally. Under stress induced by starvation, infected worker mortality increases by 50% (Brown et al. 2000). Habitat fragmentation directly replicates food stress in wild populations. Neonicotinoid exposure reduces foraging efficiency and alters bumblebee immune regulation (Simmons and Angelini 2017). Taken together, they may increase virulence of context dependent parasite systems.

As a putative regulator of host immunity and digestion, the bumblebee microbiome is central to bumblebee health and conservation. Experimental results support a mechanistic role in *Crithidia bombi* immunity, and observational findings indicate that these results may translate to wild populations (Cariveau et al. 2014; Mockler et al. 2018). However, the true diversity of wild bumblebee microbiota is not well understood. To this end, we have conducted a three-year field survey, deeply sampling populations throughout central and down east Maine. Broadly, we investigate how microbiome communities change with the landscape, and examine associations between composition and *Crithidia bombi* infection severity.

CHAPTER TWO

MATERIALS AND METHODS

## 2.1 Sample collection

Foraging bumblebees were collected throughout the state of Maine during the summers of 2017 through 2019 (Table 1). Research assistants recorded sample species, caste, and geographic location, and imaged samples upon capture. Bees were stored at -80$^{o}$C before dissection.

Specimen mouthparts and wings were dissected and imaged for concurrent projects. The entire gut was removed and stored in 70% ethanol at -80$^{o}$C until DNA extraction. Maxwell 16 DNA purification kits were used for DNA extraction. Sample DNA was stored at -20$^{o}$C before sequencing. DNA extraction for samples from 2017 and 2018 were performed during the fall semester of the following school year. Samples from 2019 were processed in the fall of 2020.

**Table 1:** Species counts per year of the *Bombus* field survey.

| | **2017** | **2018** | **2019** | **Total** |
|---|---|---|---|---|
| ***bimaculatus*** | 28 | 65 | 92 | 185 |
| ***ternarius*** | 38 | 38 | 68 | 144 |
| ***vagans*** | 38 | 50 | 52 | 140 |
| ***impatiens*** | 19 | 28 | 71 | 118 |
| ***borealis*** | 26 | 22 | 11 | 59 |
| ***terricola*** | 1 | 13 | 17 | 31 |
| ***fervidus*** | 5 | 8 | 5 | 18 |
| ***rufocinctus*** | 8 | 6 | 4 | 18 |
| ***perplexus*** | 3 | 5 | 1 | 9 |
| ***sandersoni*** | 2 | 0 | 7 | 9 |
| ***griseocollis*** | 0 | 0 | 1 | 1 |
| ***Total*** | 168 | 235 | 329 | 732 |

## 2.1 Sequencing

Paired end amplicon sequencing of the 16S v6-v8 region was run on an Illumina MiSeq by the Dalhousie University Integrated Microbiome Resource. Samples from each year were run separately, following DNA extraction. Samples were barcoded for multiplexed sequencing runs and demultiplexed by the Dalhousie University IMR.

Samples with nucleic acid concentrations less than 10 $ng\ /\ \mu l$ were dried before sequencing. Twelve samples from 2017, 29 from 2018, and 51 from 2019 were classified as low quality by sequencing quality control and have been excluded (Table 2).

**Table 2:** Species counts per year adjusted by sequencing quality control.

| | **2017** | **2018** | **2019** | **Total** |
|---|---|---|---|---|
| ***bimaculatus*** | 28 | 54 | 79 | 161 |
| ***ternarius*** | 37 | 37 | 64 | 138 |
| ***vagans*** | 31 | 40 | 34 | 105 |
| ***impatiens*** | 16 | 22 | 63 | 101 |
| ***borealis*** | 26 | 22 | 9 | 57 |
| ***terricola*** | 1 | 13 | 15 | 29 |
| ***fervidus*** | 5 | 8 | 4 | 17 |
| ***rufocinctus*** | 6 | 5 | 3 | 14 |
| ***perplexus*** | 2 | 5 | 1 | 8 |
| ***sandersoni*** | 2 | 0 | 5 | 7 |
| ***Total*** | 154 | 206 | 278 | 638 |

## 2.2 Sequence processing

I used the QIIME2 API (v. 2019.4) for processing of the raw amplicon sequencing data (Bolyen et al. 2019). I first applied DADA2 (v. 1.1.0) to denoise, dereplicate, remove chimeras,

and trim reads (Callahan et al. 2016). I processed and trimmed the three sequencing runs separately to account for batch-specific variability in read quality.

In total, DADA2 recognized 3331 unique amplicon sequence variants (ASVs). The relative abundance estimates for individual ASVs followed an exponential distribution (Figure 1). The top 10 most observed ASVs account for 75% of total abundance, while the bottom 3000 account for just 0.3%. Many of the low abundant features may be lingering sequence errors or contaminants. Since samples were drawn from wild populations, we also expect that the sequencing dataset is enriched for low abundant, environmentally derived ASVs that do not provide relevant signal for comparing microbiome composition. To reduce noise, I applied the covariance based filter implemented in PERFect (Smirnova et al. 2019). PERFect fits a skew normal distribution to log-transformed differences in filtering loss. The test assumes that over half of features are uninformative and removes features until a specified significance threshold is reached. I selected a conservative 5% significance threshold to control for the large size of the bumblebee field survey and wild origins of samples. PERFect removed 2574 features, corresponding to 0.43% of total dataset abundance (Table 3).

**Figure 1:** *Abundance Accounted for by Sorted Features*. Features were sorted by the proportion of total dataset abundance they account for. Bars represent the summed abundance of the $n^{th}$ most abundant feature and all less abundant features. Cumulative abundance follows an exponential distribution, with the 3000 least abundant features accounting for 0.3% of total abundance and the 10 most abundant accounting for 75%.

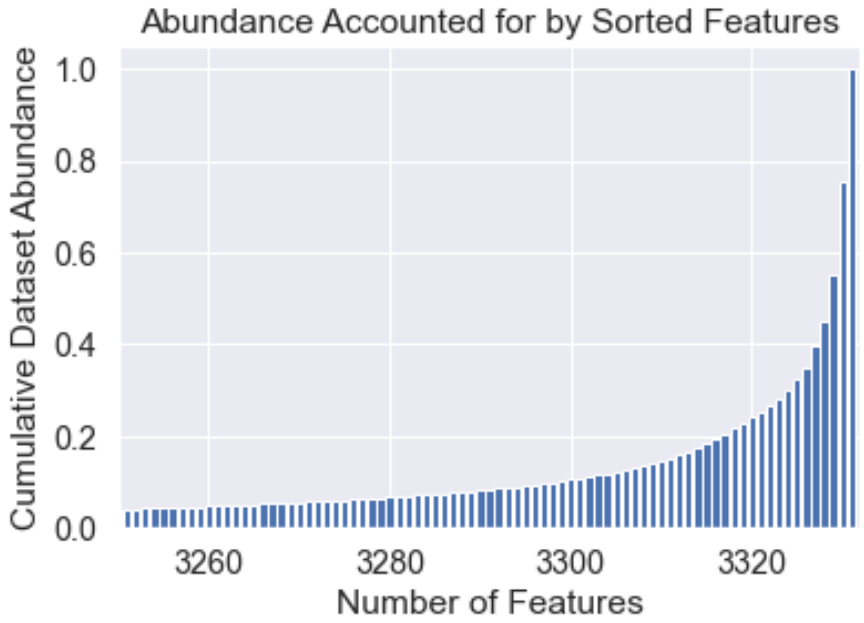



**Table 3: Feature Table Metrics**

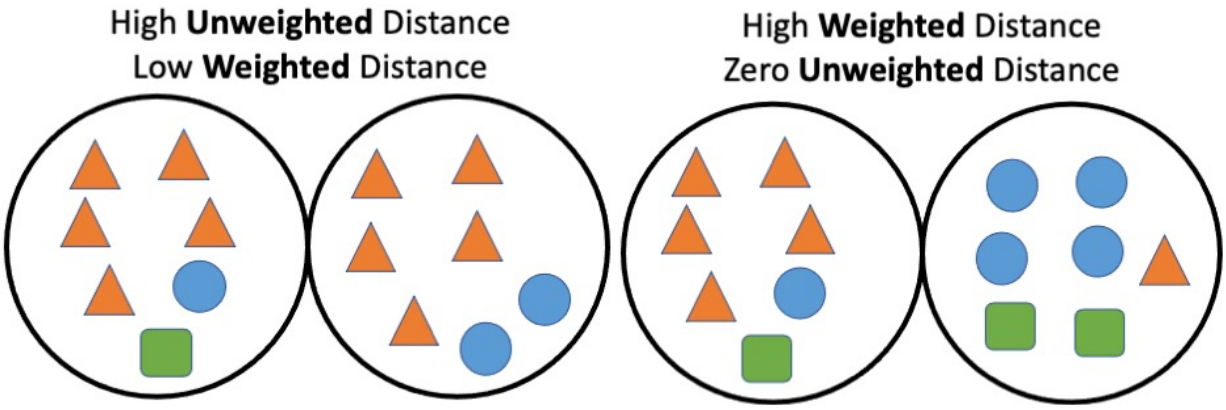
| | **Features** | **Total Reads** | **Mean Depth** | **Median Depth** | **Depth IQR** | **Depth Range** | **Sparsity** |
|---|---|---|---|---|---|---|---|
| **Full Dataset** | 3331 | $3.313 \times 10^7$ | 51927 | 44371 | [23686, 67751] | [988, 275933] | 0.992 |
| **PERFect Filtered** | 757 | $3.299 \times 10^7$ | 51704 | 44369 | [23513, 67230] | [988, 275894] | 0.972 |

### 2.3 $\beta$ Diversity calculation

$\beta$ diversity metrics quantify the dissimilarity between communities. I calculated UniFrac distance for pairwise comparisons of microbiomes (Lozupone and Knight 2005). Designed specifically for microbial communities, UniFrac distance is derived from a feature table and corresponding phylogeny. UniFrac is bounded between 0 and 1 and defined as the fraction of branch lengths of the phylogeny leading to one community but not the other. The weighted version of the metric, weighted UniFrac, scales branch lengths by the relative abundance of features they lead to (Lozupone et al. 2007). Unweighted UniFrac highlights the distribution of rare taxa, while weighted UniFrac detects compositional variation between samples colonized by related taxa (Figure 2). Neither UniFrac metric is sensitive to the method used for phylogeny building (Lozupone et al. 2007).

To construct a phylogeny, I first aligned representative sequences returned from DADA2 with MAFFT (v. 7) and masked hypervariable regions in QIIME2 (v. 2020.11.1) (Katoh and Standley 2013; Bolyen et al. 2019). I then calculated the maximum likelihood phylogeny from the masked alignment with FastTree2 (Price et al. 2010). Species richness and unweighted UniFrac distance are dependent on library size but compositionality and weighted UniFrac are not (Weiss et al. 2017). To control for variation in library size, I rarefied samples to a depth of 4000 reads before calculating unweighted UniFrac. I did not normalize the feature table for weighted UniFrac calculation to maximize the amount of available information.

**Figure 2:** ***Weighted vs Unweighted UniFrac.*** In this mock example, the triangles, circles, and squares represent three different bumblebee endosymbionts. On the left, one of the microbiome communities entirely lacks the green square phylotype. This difference would contribute to a high unweighted distance between communities. However, because both communities are primarily colonized by orange squares, weighted distance would be low. In the example on the right, both communities are colonized by the same phylotypes. They would have an unweighted distance of 0. Compositionally, the two communities are very different, so weighted distance would be high.



## 2.4 Mantel testing

I used Mantel (scikit-bio v. 0.5.6) and Partial Mantel (vegan v. 2.5.6) tests to compute the spearman rank correlation between the UniFrac dissimilarity matrices and covariates from the sample metadata. Specifically, I considered host phylogeny, geographic distance between samples, and difference in capture date. I estimated phylogenetic relationships as the sum of branch lengths from *Cameron et al.'s* bumblebee phylogeny realigned with ClustalW and RAxML (Thompson et al. 1994; Cameron et al. 2007; Stamatakis 2014). For calculation of geographic distance, I applied the haversine formula with the radius of the earth estimated as 6371 km. I tested for relationships between diversity and difference in capture date individually for each summer. I also tested for correlation with difference in capture year, with separate linear and binary encoding. I calculated linear year dissimilarities from the difference in capture year. I scored binary year dissimilarities as 0 for samples collected in the same year, and 1 for samples collected in different years. The linear year dissimilarities tests if samples become increasingly more different over the years, while the binary dissimilarities test if each year has a unique composition, without imposing a multi-year trend.

Mantel testing requires complete data. All samples have recorded capture dates and sequencing data, so the time difference and weighted UniFrac matrices are complete. However, 37 samples lack geographic coordinates and 8 were sequenced at depths lower than 4000 reads and were unable to be rarefied for unweighted UniFrac calculations. Tests using geographic and unweighted UniFrac matrices were run on subsets of the full data (Table 4).

**Table 4:** ***Incomplete geographic and unweighted UniFrac data by year***

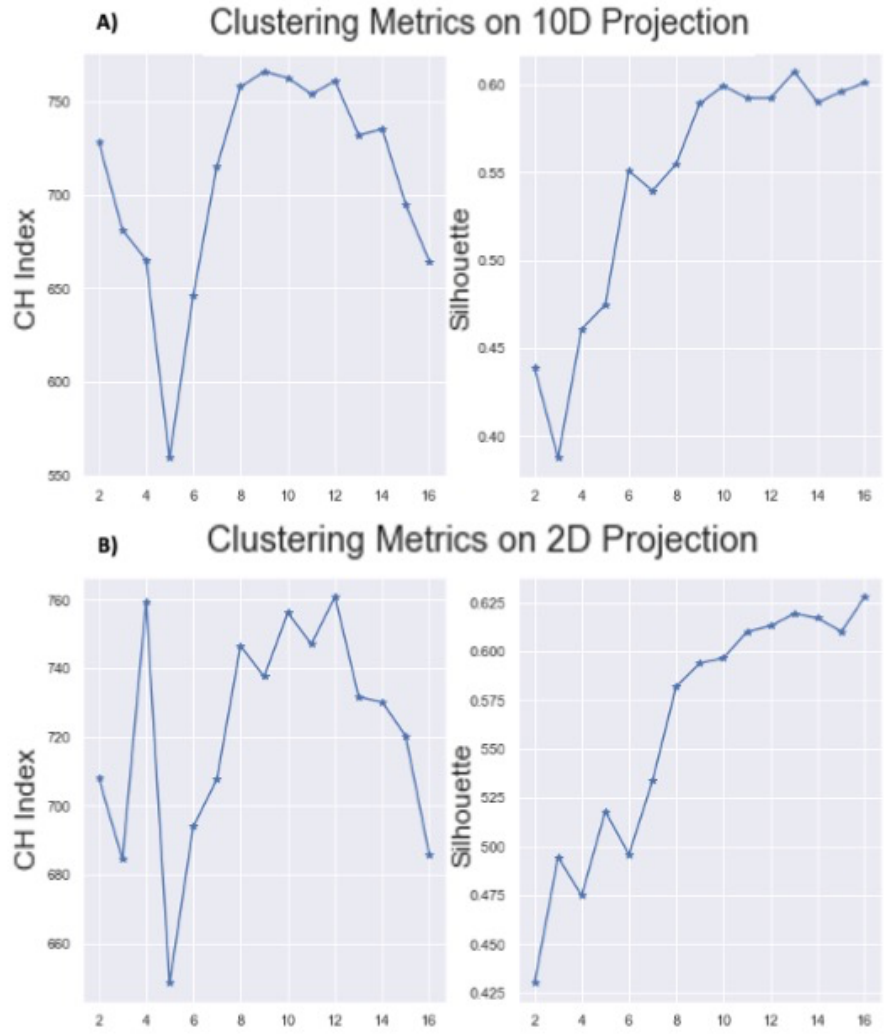
| | **2017** | **2018** | **2019** | **Total** |
|---|---|---|---|---|
| **Geographic Data** | 134 | 195 | 272 | 601 |
| **unweighted UniFrac** | 152 | 204 | 274 | 630 |
| **Total** | 154 | 206 | 278 | 638 |

### 2.5 Clustering on microbiome composition

I clustered samples to compare microbiome communities with similar composition. Prior to clustering, I applied the Uniform Manifold Approximation and Projection algorithm (v. 0.4.6) (McInnes et al. 2018). UMAP outperforms comparable dimensionality reduction algorithms on sensitivity to classification problems involving biological count data (Dorrity et al. 2020). The algorithm functions by constructing a low dimensional representation based upon local topologies in the full data. I chose to consider the 30 nearest neighbors of each point and specified a minimum distance of 0.05. I clustered on a 10-dimensional projection of the weighted UniFrac data and produced a 2-dimensional projection for plotting (Figure 9A).

I clustered samples with the K-medoids algorithm (scikit-learn extra v. 0.1.0b2) (Park and Jun 2009). For assessment of cluster number, I calculated the Calinski-Harabasz index and mean silhouette coefficients for $2 \leq k \leq 16$ (Caliński and Harabasz 1974; Rousseeuw 1987). Both metrics compare within to between cluster variation. The Calinski-Harabasz index is scaled by the number of clusters and shows a global peak at $k = 9$ in the 10-dimensional data (Figure 3A). Silhouette coefficients are bounded between -1 and 1 and are not scaled by the number of clusters. At $k = 9$ the mean silhouette is 0.59 in the 10-dimensional data, indicating strong support for clustering. The 2-dimensional projection used for plotting shows slightly reduced, but strong support for clusters (Figure 3B).

**Figure 3:** ***Clustering on microbiome composition. A)*** *Clustering metrics on 10D projection.* Clustering was performed with the K-medoids algorithm on a 10-dimensional UMAP of the weighted UniFrac distance matrix. A cluster number of $k = 9$ was empirically selected with the Calinski-Harabasz index and mean silhouette coefficients. The CH index peaks at 9 clusters. At $k = 9$, the mean silhouette is 0.59, indicating strong support for clustering. ***B)*** *Clustering metrics on 2D projection.* Although clustering was performed on a 10-dimensional UMAP, a separate 2-dimensional projection has been used for visualization. Clustering metrics are slightly less robust on the 2-dimesional data.



### 2.6 Taxa classification

Taxonomic classification is relevant for downstream analyses and interpretability. I trained a Naïve Bayes classifier (scikit-learn v. 0.23.1) on the Silva database of full length 16s sequences (release 132) (Quast et al. 2013). I used the 99% NR database with a consensus taxonomy to reduce ambiguity in assignment. I classified representative sequences in the PERFect filtered feature table and collapsed taxonomic labels to the genus level, finding 151 unique phylotypes. At the genus level, the Silva database is not able to differentiate between different *Lactobacillus* phylotypes, nor does it contain annotations for *Parasaccaribacter* (Quast et al. 2013).

### 2.7 Differential abundance testing

I tested for differential abundance between microbiome clusters with analysis of composition of microbiomes with bias correction (ANCOM-BC) (Lin and Peddada 2020). ANCOM-BC uses a linear regression framework to test for differential relative abundance between samples. It assumes that observed samples are unknown fractions of the entire microbiome and corrects for difference in sampling fraction with an offset term. In comparison to other differential abundance tests, ANCOM-BC minimizes FDR while preserving high power (Lin and Peddada 2020). I used the R implementation (v. 1.0.5) on the PERFect filtered feature table collapsed to genus level assignments.

### 2.8 Cluster observational frequency

Latitude and longitude were recorded for 601 of the 638 samples (Table 4). I created categorical collection sites to group samples from the same geographic areas. I selected a

diameter of 10 km, corresponding to the maximum efficient nectar foraging range (Cresswell et al. 2000). The dataset contains 54 unique sites. Allen Island and Colby College are the most deeply sampled, with over 100 specimens each. Most sites were visited only once or twice over the course of the field survey, with 40 having less than 10 samples. I denoted collection trips as unique site and date pairs. The field survey spans 119 trips. I used a $\chi^2$ test to compare observational frequency across microbiome compositional clusters.

### 2.9 *Crithidia bombi* quantification

I quantified *Crithidia bombi* infections with qPCR targeting the GAPDH gene. I flagged samples as positive if their $C_T$ was less than the $10^3\ template\ molecule/\mu l$ standard. I processed all samples during the winter of 2020. Preceding qPCR, purified DNA samples were stored at -20$^0$C. I used one microliter of template for reactions.

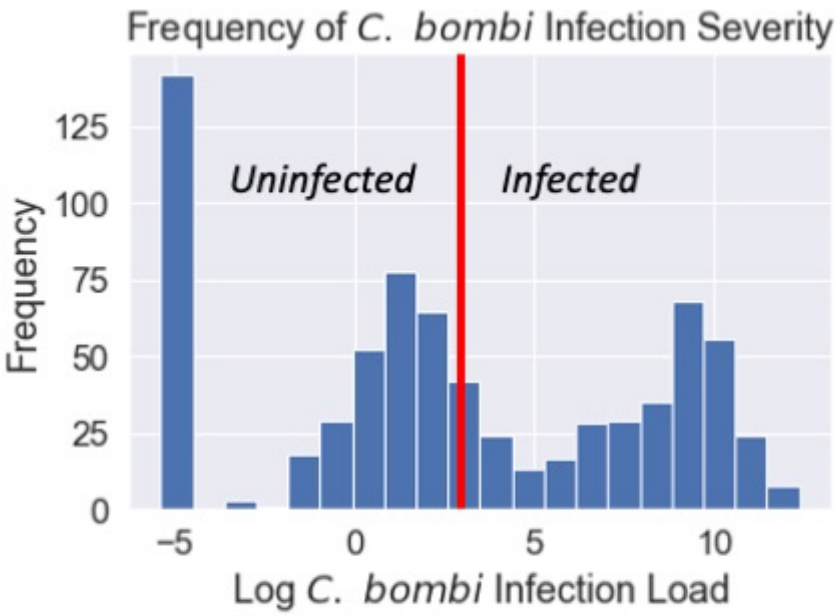
Due to the high volume of samples, I adopted a batched protocol based on the assumption that most samples would be negative. All samples from a pooled batch that tested negative were flagged as negative with imputed *C. bombi* loads of 0. A batch of size *b* and infection rate *p* tests positive with probability $1-(1-p)^b$. For *s* samples, the expected number of positive batches, *x,* is $\frac{s}{b}(1-(1-p)^b)$. I chose to run batches in duplicate and singles in triplicate. Thus, the total number of expected reactions was $\frac{s}{b}\cdot 2 + x\cdot b\cdot 3$. For samples run in single, I calculated infection load as the estimated starting quantity of *C. bombi* GAPDH per *ng* of gut DNA supplied to the reaction.

I used binomial regression for quality control. I found no significant relationship between the concentration of a batch and its positivity (Wald Test, $z=-0.556, p=0.578$) nor the

concentration of a sample and its positivity (Wald Test, $z = -0.712, p = 0.477$), indicating that sample quality was not confounding.

*C. bombi* infection severity shows a strongly bimodal distribution on a log transformed scale (Figure 4). The distribution is indicative of a high number of mild or extreme cases, with few in between. This is in agreement with prior observations of context dependent virulence in *C. bombi* (Brown et al. 2003). The mound shaped distribution of samples with low detected *C. bombi* loads may partially be due to instrument error near the detection limit.

**Figure 4:** *Frequency of C. bombi Infection Severity*. *C. bombi* infection severity was calculated as the estimated starting quantity of *C. bombi* GAPDH per *ng* of gut DNA supplied to the qPCR reaction. On the x-axis is the log transformed *C. bombi* infection load. Samples that passed the batch screen were imputed as 0. A pseudocount of 0.001 was added to all samples for the log transformation. The samples eliminated through the batch screens fill the bin on the far left. The flagged samples run in single show a strongly bimodal distribution on the log transformed scale. The red line demarcates samples with infections greater than $10^3$ $template\ molecule/\mu l$ from those with less.

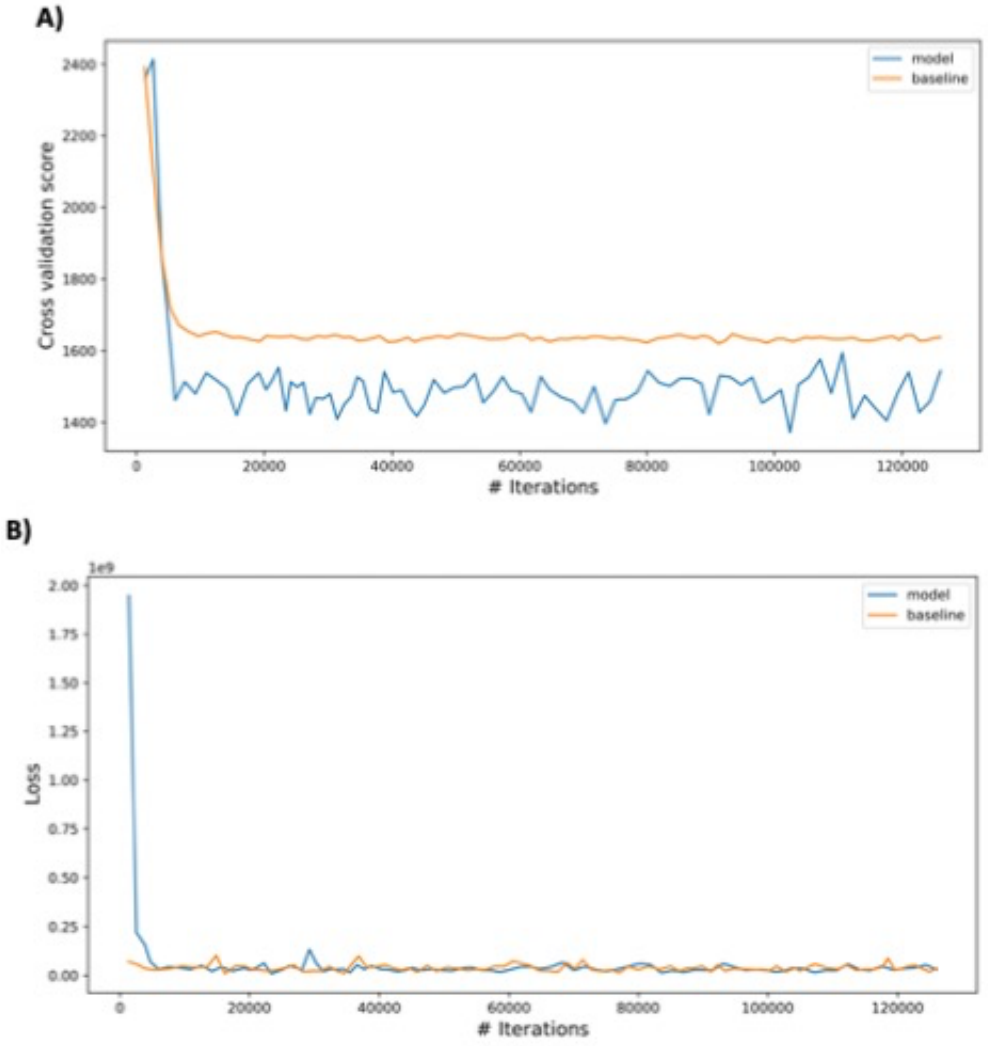


### 2.10 Differential ranking with Songbird

I used Songbird (v. 1.0.4) to calculate relative log fold change of microbial taxa with respect to change in *Crithidia bombi* infection load (Morton et al. 2019). Songbird assumes that observed samples are an unknown fraction of the total microbial load and calculates the "relative differential," or log fold change in relative abundance of taxa between conditions. Songbird uses the differential abundances of taxa as reference frames for each other and produces a ranking of how much taxa change relative to each other.

I included collection year and date in my *Crithidia bombi* infection load model to correct for their observed relationships with infection severity and microbiome composition (Table 5). The model loss and cross validation exponentially decrease before approaching asymptotes, indicating predictive error reduces with fit to the training data (Figure 5). $Q^2$ is defined as one minus the ratio of absolute model error to null model error. Values for $Q^2$ are bounded by negative infinity (more error than the null model) and one (no error). The observed $Q^2$ of 0.0963 indicates that relative abundances of microbiota change with *Crithidia bombi* load and collection date, but that there is significant additional variation in microbiome composition not explained by the model.

**Figure 5:** ***Differential ranking with Songbird. A)*** *Cross validation.* Model cross validation in comparison to the null. The cross validation exponentially decreases before stabilizing, indicating that it fit to the data. The model explains more of the observed variation in relative abundance than the null ($Q^2 = 0.0963$). ***B)*** *Loss.* The model loss decreases exponentially with training iterations before approaching an asymptote, indicating fit to the data.



# CHAPTER THREE

# RESULTS

## 3.1 Maine's bumblebee species share a common microbiome community

Bumblebee microbiomes are composed of high abundances of specific, core microbiota. *Kwong et al.* previously identified *Snodgrassella, Gilliamella, Lactobacillus* Firm-4, *Lactobacillus* Firm-5, and *Bifidobacterium* as coevolved endosymbionts of the corbiculate bees through analysis of strain and host phylogenies (Kwong et al. 2017). They further identified *Bombiscardovia* and *Schmidhempelia* as *Bombus*-specific, and *Apibacter, Lactobacillus* Firm-3, and *Parasaccaribacter* as corbiculate associated microbiota found in bumblebees (Kwong et al. 2017).

ASVs assigned to the core phylotypes *Snodgrassella, Gilliamella, Lactobacillus, Schmidhempelia, Apibacter, Bombiscardovia,* and *Bifidobacterium* have been detected in our dataset of wild Maine bumblebees. These seven core phylotypes are detected in all eleven species included in the field survey (Figure 6B). *Snodgrassella, Gilliamella,* and *Schmidhempelia* are the most common, at dataset relative abundances of greater than 20% (Figure 6A). *Snodgrassella* and *Gilliamella* are the most frequently observed, at rates greater than 95 and 93% of samples per species, respectively. *Apibacter* is also detected in the majority of samples per species, but *Bifidobacterium, Bombiscardovia,* and *Lactobacillus* are more variable (Figure 6B).

The bumblebee microbiome is shared between host species and robust to variation in environmental influence. No correlation between the pairwise weighted UniFrac dissimilarities and host divergence was found, indicating that bumblebee species harbor compositionally similar

microbiomes (Table 5; Mantel test, $\rho$ = -0.0175, p =.49). Similarly, no correlation was found with the geographic distance between collection sites (Table 5; Mantel test, $\rho$ = -5.4e-4, p = 0.98). These results support the narrative of a coevolved, bumblebee-specific microbiome over the null hypothesis of passive diffusion of microbes through host populations.

Despite its robustness to geography, microbiome composition varies over the course of the summer. For each year of the field survey, microbiome compositional differences increase with difference in collection date, even when geography is controlled for (Figure 6C). For 2018 and 2019, the spearman correlation is highly significant (Table 5; Partial Mantel Tests, p < 0.001). The subset of samples from 2017 shows an insignificant correlation but is also smaller than the other years (Table 5; Partial Mantel Tests, p = 0.367). In addition to within-summer trends, partial Mantel tests reveal significant compositional variation between collection years. Binary-coded year dissimilarities show a greater correlation with weighted UniFrac than the raw integer values, indicating that the yearly trend is not linear. Instead, the collection years are unique in their own way (Table 5).

**Figure 6:** ***Maine's bumblebee species share a coevolved microbiome****. A) Dataset abundance of core microbes.* Seven coevolved phylotypes were detected in the dataset. *Snodgrassella, Gilliamella,* and *Schmidhempelia* are the most common, with abundances greater than 20%. **B)** *Phylotype observation by host species.* The seven core phylotypes were detected in samples of all species included in the bumblebee field survey. *Snodgrassella* and *Gilliamella* are found in greater than 95% and 93% of samples per species, respectively. Annotations give the raw number of samples; shading is by proportion. **C)** *Microbiome composition varies over the summer.* For 2018 and 2019, there are significant positive trends between difference in sample collection dates and weighted UniFrac distance.

**Table 5:** ***Correlates of Weighted UniFrac Distance***

| | **Control (for Partial Mantel)** | **Spearman $\rho$** | **p-value** | **n** |
|---|---|---|---|---|
| **Host phylogeny** | n/a | -0.0175 | 0.49 | 638 |
| **Geography** | n/a | -5.43e-4 | 0.976 | 601 |
| **Collection date (2017)** | Geography | 0.0093 | 0.367 | 134 |
| **Collection date (2018)** | Geography | 0.1747 | **< 0.001** | 195 |
| **Collection date (2019)** | Geography | 0.0935 | **< 0.001** | 272 |
| **Year (binary)** | Geography | 0.0747 | **< 0.001** | 601 |
| **Year (numeric)** | Geography | 0.0416 | **< 0.001** | 601 |

### 3.2 Rare taxa vary with time, space, and host phylogeny

In the wild, bumblebees exchange microbes through social contact and pollination. Therefore, it is expected that bumblebees sharing pollination routes will share a greater proportion of microbial taxa. Unweighted UniFrac distance is sensitive to changes in the distribution of all taxa, rather than overall composition. In the field survey, unweighted UniFrac dissimilarities increase with difference in capture date for all years, and with geographic distance for 2017 and 2019 (Table 6; Partial Mantel Tests, $p < 0.05$). Together, these results indicate that samples from the same time and place host more similar taxa. Host phylogeny is also a significant correlate of unweighted UniFrac distance (Table 6; Mantel Test, $\rho = 0.088$, $p < 0.001$). Bumblebee species have unique floral preferences and may reinforce species-specific taxa through shared pollination routes.

Each year of the field survey is unique for its distribution of non-core microbiota (Figure 7A). By Mantel tests, collection year is the greatest correlate of unweighted UniFrac distance, with the binary coded year dissimilarities outperforming the linear encoding (Table 6; Partial Mantel Tests, $p < 0.001$). The collection years partially separate on a PCoA plot of the unweighted UniFrac dissimilarities (Figure 7B). The PCoA shows an axis of symmetry along the first component, with three stacked clusters on either side corresponding to years of the field survey. Samples from 2019 are found in the top cluster, samples from 2017 are in the middle, and samples from 2018 are found at either end.

**Figure 7:** ***Rare taxa vary by time, space, and host phylogeny****. **A)** Taxa Venn diagram.* The majority of ASVs in the PERFect-filtered dataset are not detected in multiple years of the field survey. ***B)*** *Samples separate by year on unweighted UniFrac.* The first two components of a PCoA of the unweighted UniFrac dissimilarities capture 34% of total variation. The plot shows an axis of symmetry along the first component with three stacked clusters on either side. The clusters along the second component correspond to years of the field survey.

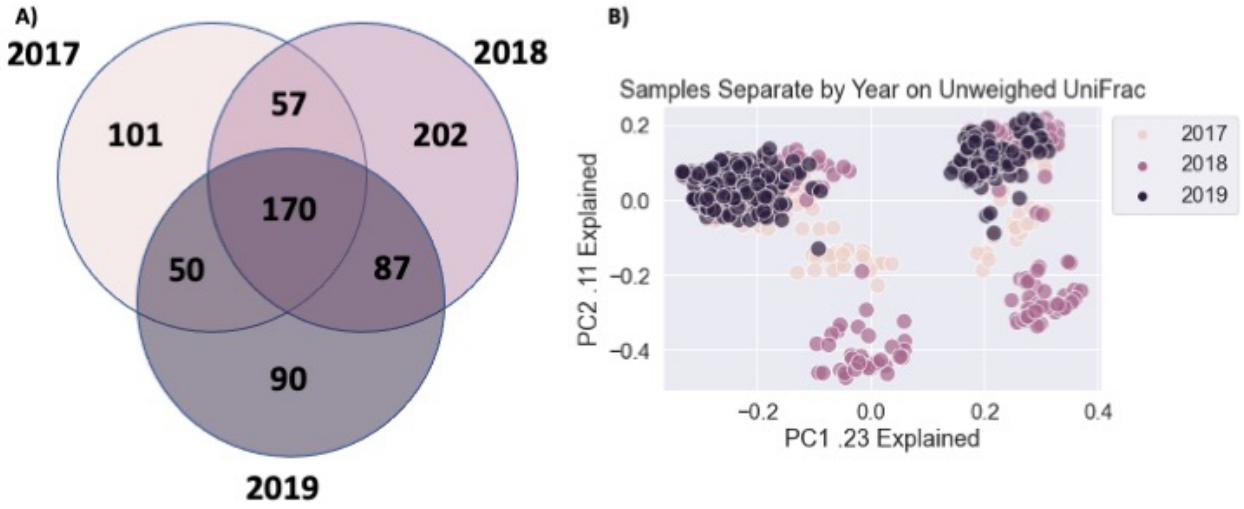



**Table 6:** ***Correlates of Unweighted UniFrac Distance***

| | **Control (for Partial Mantel)** | **Spearman *ρ*** | **p-value** | **n** |
|---|---|---|---|---|
| **Host phylogeny** | n/a | 0.0881 | **< 0.001** | 630 |
| **Geography (2017)** | Collection date | 0.0917 | **0.016** | 132 |
| **Geography (2018)** | Collection date | -0.0130 | 0.751 | 193 |
| **Geography (2019)** | Collection date | 0.0455 | **< 0.001** | 268 |
| **Collection date (2017)** | Geography | 0.0455 | **0.041** | 132 |
| **Collection date (2018)** | Geography | 0.1347 | **< 0.001** | 193 |
| **Collection date (2019)** | Geography | 0.1152 | **< 0.001** | 268 |
| **Year (binary)** | Geography | 0.2158 | **< 0.001** | 593 |
| **Year (numeric)** | Geography | 0.1457 | **< 0.001** | 593 |

### 3.3 Sympatric long and short faced bees harbor clade-specific microbiota

Bumblebee species divide into monophyletic long and short-faced clades. Long faced species are longer tongued, predisposing them to specialist pollination of flowers with deep corollas (Prys-Jones 1982; Goulson 2010). Sympatric bumblebee species have been observed to partition resources by tongue length (Pyke 1982). From this observation, it is expected bees with more similar tongue lengths will more frequently exchange microbes through pollination and harbor more similar microbiomes.

Long-faced species are rare in mainland Maine, making interclade comparison difficult. However, offshore island bumblebee populations are compositionally unique. The majority of all long-faced specimens included in the field survey were collected from Allen Island, a small, partially forested island of just 450 acres (Figure 8A). Due to the island's small size and geographic isolation, analyses of diversity are not confounded by influx of microbes from neighboring sites. Allen island was most deeply sampled during the summer of 2018, with 20 long-faced and 38 short-faced specimens.

Consistent with the full field survey, compositional diversity varies significantly with collection date but not phylogeny within the 2018 Allen Island samples (Table 7; Partial Mantel Tests, $p<0.001$). The long and short-faced clades are monophyletic, so interclade comparisons correspond to greater phylogenetic distances. The absence of a significant association between compositional diversity and host phylogenetic distance indicates that clade membership is not a driver of compositional variation.

In contrast to the compositional results, the distribution of rare taxa shows a relationship with host phylogeny. Unweighted UniFrac is driven by both collection date and host phylogeny, with host phylogeny having a larger spearman correlation (Figure 8B; Table 7; Partial Mantel

Tests, $p<0.001$). As indicated by these results, short and long-faced bees host unique low-abundance microbiota, even when constrained to a small and isolated geographic area. Of the 395 ASVs detected in the 2018 Allen Island samples, 184 and 38 are unique to the short and long-faced populations, respectively. Mostly, these correspond to non-bumblebee associated phylotypes. However, 20 ASVs corresponding to core microbiota are unique to the short-faced population, and 9 are unique to the long-faced (Figure 8C). This observed differential abundance may be an artifact of the low abundances of individual ASVs, or it could be an indicator of emerging host specificity between long and short-faced endosymbionts.

**Figure 8:** ***Sympatric long and short faced bees harbor clade-specific microbiota****. **A)** Bumblebee species compositional variation.* The offshore islands included in the field survey hosted compositionally unique populations of bumblebee species. Allen Island was the most frequently sampled site and hosted the highest proportion of long-faced species. The bar chart includes samples from all years. ***B)*** *Rare taxa vary with phylogeny and time.* Allen Island is small, geographically isolated, and hosts the highest proportion of long-faced species of all sites included in the field survey. Unweighted UniFrac distance increases with host phylogenetic variation and difference in collection date. The 58 samples collected during the summer of 2018 were used for analysis. ***C)*** *Clade specific core ASVs.* 20 ASVs corresponding to core microbiota are unique to the 2018 Allen Island short-faced population and 9 are unique to the short faced.

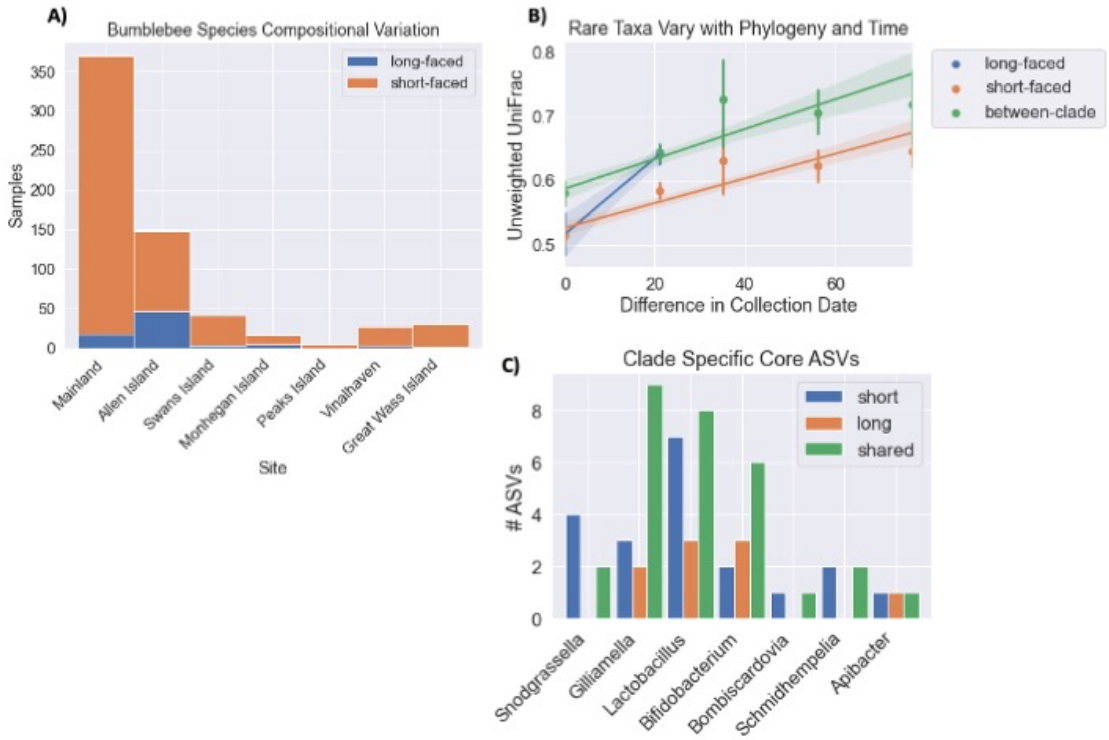


**Table 7:** ***Allen Island 2018 Diversity Analysis***

| | **Control** | **UniFrac** | **Spearman $\rho$** | **p-value** | **n** |
|---|---|---|---|---|---|
| **Phylogeny** | Time | unweighted | 0.2228 | **<0.001** | 58 |
| **Time** | Phylogeny | unweighted | 0.2135 | **<0.001** | 58 |
| **Phylogeny** | time | weighted | -0.02677 | 0.628 | 58 |
| **Time** | phylogeny | weighted | 0.2561 | **<0.001** | 58 |

### 3.4 Microbiome communities cluster by composition

Even though our recorded covariates are not strong predictors of community composition, there is still significant variation. A UMAP projection of the weighted UniFrac distance matrix shows support for nine robust clusters (Figure 3; mean silhouette = 0.59). The UMAP projection has a major axis of variation along the upper diagonal with divergent points falling to the right and below (Figure 9A). Cluster 0, at the base of the diagonal, is characterized by near complete colonization by *Snodgrassella,* a core endosymbiont of the corbiculate bees. Cluster 1, on the upper right end, is characterized by similarly high abundances of *Schmidhempelia,* an endosymbiont of just the bumblebees (Kwong et al. 2017). Clusters 4, 7, 8, and 3 are on the middle of the diagonal and have higher diversities of core microbiota. Samples from clusters 2, 5, and 6 are detached from the main diagonal and have high compositions of non-bumblebee associated microbes (Figure 9B).

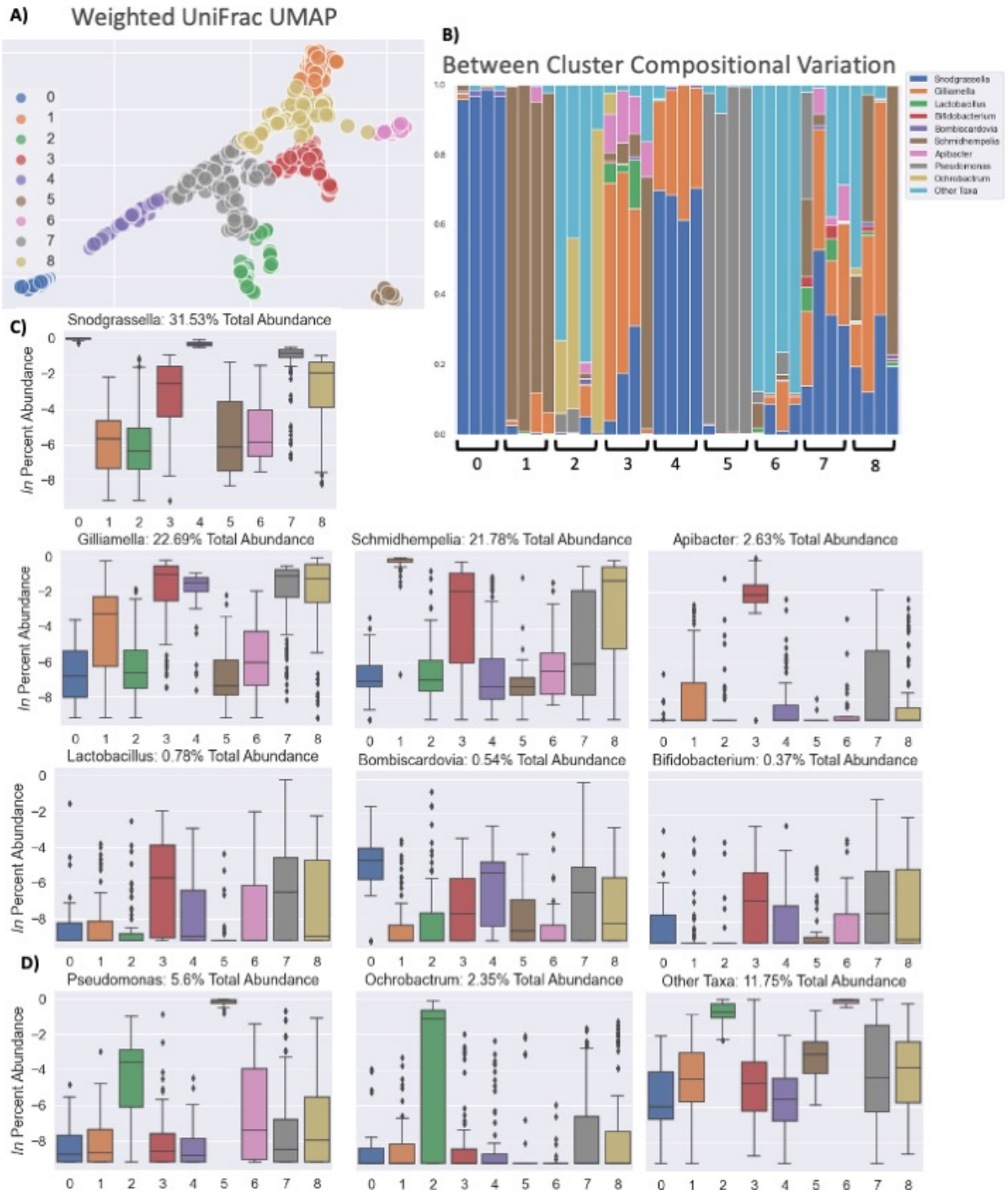
Differential abundance testing on the genus-level phylotypes returned 32 significantly differentially abundant taxa between clusters, including all seven core-microbiota detected in the dataset (Figure 9C; ANCOM-BC, holm adj. $q < 0.05$). Clusters along the main axis of variation on the UMAP plot have greater diversities of core microbiota. Samples from cluster 3 are particularly diverse, with high abundances of *Snodgrassella, Gilliamella, Schmidhempelia, Apibacter,* and *Lactobacillus.* Cluster 3 also hosts high abundances of *Bombiscardovia* and *Bifidobacterium*, but these phylotypes are rarer and have less dramatic inter-cluster variation.

Clusters 7 and 8 are similar to cluster 3 in distribution of core microbiota but lack high abundances of *Apibacter.*

Twenty-five non-core phylotypes are also differentially abundant between clusters (Figure 9D). Largely, they are of lower dataset abundances, with 21 falling below 1%. Their low abundances and lack of prior association with bumblebees indicate that they may be environmentally associated, rather than adapted to the bumblebee microbiome. *Pseudomonas* and *Ochrobactrum,* the two highest abundance non-core microbiota, show distinct differential abundance patterns (Figure 9D). Samples from cluster 5 are nearly completely colonized by *Pseudomonas* and cluster 2 supports a large and variable range of *Ochrobactrum* colonization. Other non-core taxa are very highly abundant in clusters 6 and 2. Of the clusters with primary colonization by core microbiota, 7 and 8 host the highest mean abundances of non-core microbes.

**Figure 9:** ***Microbiome communities cluster by composition. A)*** *Weighted UniFrac UMAP*. The UMAP projection of the weighted UniFrac distance matrix shows strong support for 9 clusters (Figure 2). The plot shows significant variation along the upper diagonal with divergent points falling to the right and below. ***B)*** *Between cluster compositional variation.* A taxa bar plot highlighting core microbiota and the two highest abundance non-core microbes (*Pseudomonas, Ochrobactrum*) shows significant compositional variation between clusters. ***C)*** *Core microbe percent abundance boxplots.* Percent abundances are shown on a natural log scale. A pseudocount of 0.001 was added prior to the log transformation. Due to the transformation, values of 1 correspond to 100% relative abundance and values of -7 correspond to 0%. ***D)*** *Non-core microbe percent abundance boxplots.* The two highest abundance non-core taxa are

included, as well as other pooled taxa. A natural log transformation was applied in the same manner as *C.*

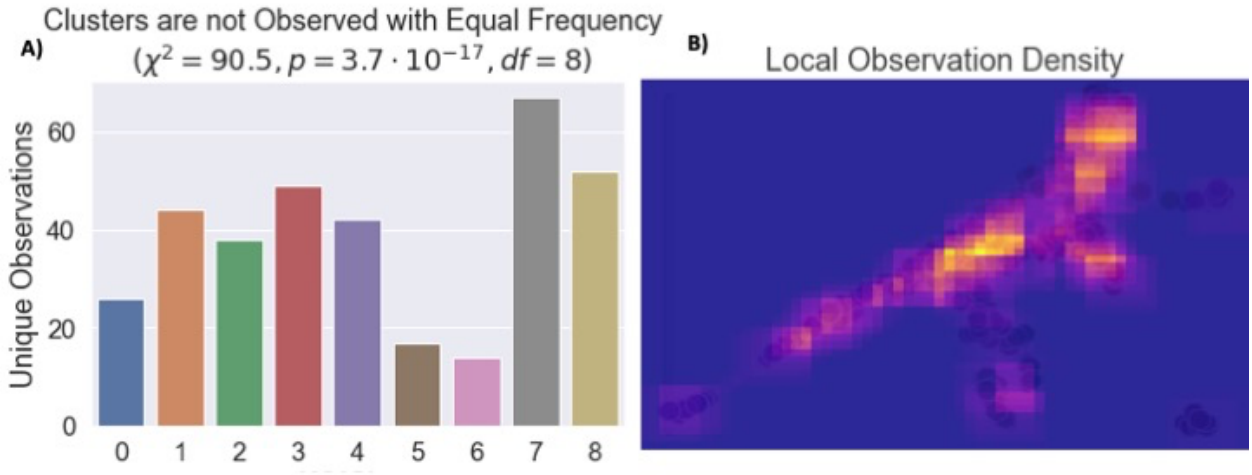


### 3.5 Clusters defined by core microbiota are more frequently observed in wild populations

The compositional microbiome clusters are not evenly represented in the dataset. Visually, they differ on the number of samples assigned to each (Figure 9A). Presumably, some reoccur throughout the field survey, while others are restricted to specific communities of bumblebees. Analyses based directly upon the number of samples per cluster may be confounded by heavy sampling of sites with atypical microbiota. Scoring clusters by the number of unique collection trips they were observed on corrects for bias in unequal sampling depth. Clusters 7 and 8 (diverse core and non-core) are the most frequently observed in the field survey, followed by 1 (*Schmidhempelia*), 2 (*Ochrobactrum*), 3 (diverse core), and 4 (*Snodgrassella* and *Gilliamella*). Clusters 0 (*Snodgrassella*), 5 (*Pseudomonas*), and 6 (diverse non-core) are relatively rare. A $\chi^2$ test confirms the presence of significant variation (Figure 10A; $p < 0.001$).

Generally, the most frequently observed clusters are all categorized by high abundances of diverse core microbes. Cluster 2 is an exception to the rule for its high abundances of *Ochrobactrum*, a non-bumblebee-associated microbe. Cluster 1 is also a partial exception for its extreme relative abundances of *Schmidhempelia*. When the UMAP projection is recolored by local observational density, divergent points falling below and to the right of the main axis of variation fade (Figure 10B). The preserved axis corresponds to samples with high abundances of core microbiota, indicating that wild bumblebees are primarily colonized by their coevolved endosymbionts, and that dysbiosis is a rare and variable phenomenon.

**Figure 10:** ***Compositional clusters defined by core microbiota are longer lived in the data. A)*** *Clusters are not observed with equal frequency.* Clusters were ranked by the number of unique collection trips they were observed on. A $\chi^2$ test confirms significant variation between clusters ($\chi^2 = 90.5, df = 8, p = 3.7 \times 10^{-17}$). Clusters 0, 5, and 6 are infrequently observed. ***B)*** *Local observation density.* The UMAP scatter plot is overlayed with a density plot colored by local frequency of unique collection trips. The main axis of variation corresponding to high abundances of core microbiota is preserved. Divergent points falling to the right and below have faded.



### 3.6 *Crithidia bombi* is abundant in Maine

*Crithidia bombi* infections were quantified with qPCR targeting the GAPDH gene. Due to the high volume of samples, pooled batches were used for initial screens. Samples flagged from the batched screens were run in single, and those with $C_T$ values less than the $10^3\ template\ molecule/\mu l$ standard and were classified as infected. *Crithidia bombi* infection loads were estimated from the starting quantities of individual runs and imputed as 0 for samples removed through batching. In total, 290 of the 638 samples were classified as infected. Infection severity displays a bimodal distribution on a log transformed scale, indicating high frequencies of uninfected and highly infected samples, with few in between (Figure 11A).

*Crithidia bombi* infection severity and frequency increase over the duration of the summer (Figure 11B). In each year of the field survey, both the mean *C. bombi* load and positivity rate show significant Spearman correlations with the number of days since May first (Table 8; t-test, $p< 0.05$). The trend is weakest in 2017, but this may be due to the lower number of samples collected in the first year of the field survey. The increase in frequency over the summer is in line with the expectation that colonies horizontally exchange microbiota through overlapping pollination networks. A subset of colonies are founded by infected queens each spring, and infections spread as the summer progresses. Geographically, the coastal region appears to be most affected by *C. bombi* (Figure 11C). Infections are detected throughout Maine, but coastal sites host the most severely infected samples. It is not immediately clear what causes this trend, but it is visually apparent.

**Figure 11:** ***Crithidia bombi is abundant throughout Maine A)*** *Frequency of infection severity. C. bombi* infection severity was quantified with qPCR. Frequency of infection load has a bimodal distribution on a log transformed axis. Samples with $C_T$ values less than the $10^3$ $template\ molecule/\mu l$ standard were classified as infected. The vertical red line gives the division uninfected and infected samples. In total, 290 samples were classified as infected. ***B)*** *Infection severity increases over the summer.* For each year of the field survey, there is a significant positive spearman correlation between infection load and collection date, quantified as the number of days since May first (t-test, $p < 0.05$). ***C)*** *Geographic distribution of infections.* Points correspond to individual samples and are sized by infection load.

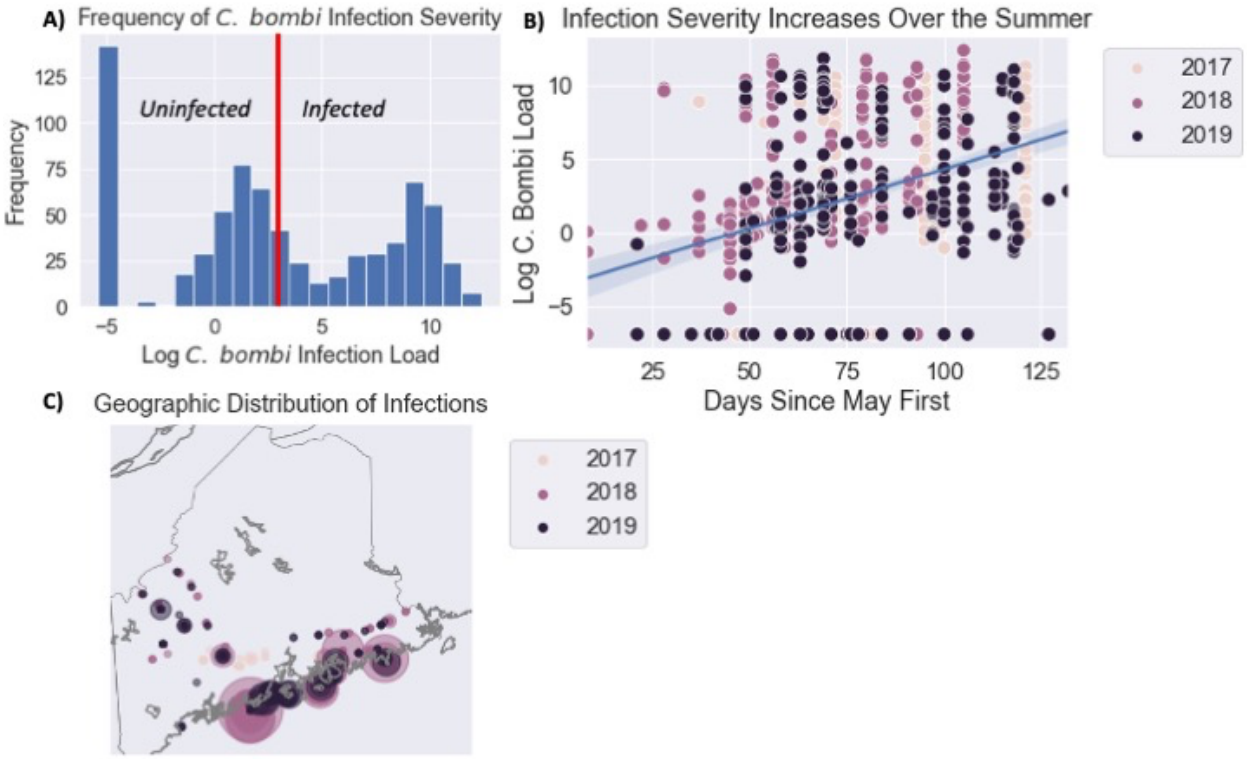


**Table 8: *Crithidia bombi* trends with collection date**

| | **Infection Frequency** | **Infection Load** |
|---|---|---|
| **2017** | $(\boldsymbol{\rho = 0.194, p = 0.0159})$ | $(\boldsymbol{\rho = 0.200, p = 0.0123})$ |
| **2018** | $\boldsymbol{\rho = 0.392, p = 5.51 \times 10^{-9}})$ | $(\boldsymbol{\rho = 0.484, p = 1.70 \times 10^{-13}})$ |
| **2019** | $(\boldsymbol{\rho = 0.175, p = 3.39 \times 10^{-3}})$ | $(\boldsymbol{\rho = 0.293, p = 6.31 \times 10^{-7}})$ |

### 3.7 Core microbiota are negatively associated with *Crithidia* infection severity

Microbiome compositional clusters differ in their average *Crithidia bombi* infection severity (Kruskal Wallis Ranked Sum Test, H= 100.99, $p = 6.74 \times 10^{-19}$). By Mann-Whitney U tests, samples from cluster 5 (*Pseudomonas*) have the greatest infection load, cluster 6 (diverse non-core) has the second highest, and clusters 3 (diverse core) and 1 (*Schmidhempelia*) have the lowest infection loads (Figure 12A). The other clusters with high abundances of core microbiota have relatively low infection loads, but not as significantly as clusters 3 and 1.

When compared directly, there is a highly significant relationship between the summed relative abundance of all core microbiota and infection severity (Figure 12B; $\rho = -.31$; t-test $p = 1.5 \times 10^{-15}$). Although this correlation affirms the relationship between core microbiota and host health, it is not informative of the relationships between individual taxa and infection severity. Due to the compositional nature of microbiome sequencing data, traditional correlations that assume independence are not applicable for testing relationships between taxa (Friedman and Alm 2012). Instead, analyses focused on how the ratios of taxa relative abundances change with a study condition are more accurate.

Songbird (v. 1.0.4) calculates log fold change of taxa relative abundance, or differentials, between sample conditions (Morton et al. 2019). A model for collection year, date, and *Crithidia*

infection severity fit to the field survey data successfully predicts more microbial variation than the null (Figure 5; $Q^2 = 0.0963$). Of all taxa observed in the field survey, *Pseudomonas* has the greatest differential with *Crithidia* infection severity, indicating that relative abundance of *Pseudomonas* increases with infection load (Figure 12C). Except for *Bombiscardovia,* all core microbiota have negative differentials, indicating their relative abundances decrease with increased infection load. *Lactobacillus* has the greatest negative association and *Snodgrassella* has the weakest. The rankings of differentials indicate that although *Snodgrassella, Gilliamella,* and *Schmidhempelia* are the most frequently observed core microbiota, decreased abundance of *Apibacter* and *Lactobacillus* may be better predictors of host resistance to *Crithidia bombi* infection.

**Figure 12:** ***Core microbiota are negatively associated with Crithidia infection severity. A)*** *Between cluster variation in infection load.* There is significant between cluster variation in infection load (Kruskal Wallis Ranked Sum Test, $H = 100.99,\ p = 6.74 \times 10^{-19}$). Clusters characterized by high abundances of non-core microbiota have greater infection loads. ***B)*** *Infection load by abundance of microbes.* The relative abundance of core microbiota shows a significant spearman correlation with *C. bombi* infection load ($\rho = -.31$, t-test, $p = 1.5 \times 10^{-15}$). Points are colored by cluster. ***C)*** *Songbird compositional associations.* Songbird was used calculate differential rankings of associations of microbiota with *C. bombi* infection severity. All core microbiota, save *Bombiscardovia,* have negative log differentials, indicating negative relationships with *C. bombi* load. *Pseudomonas* has the greatest association with *C. bombi* load of all microbiota.

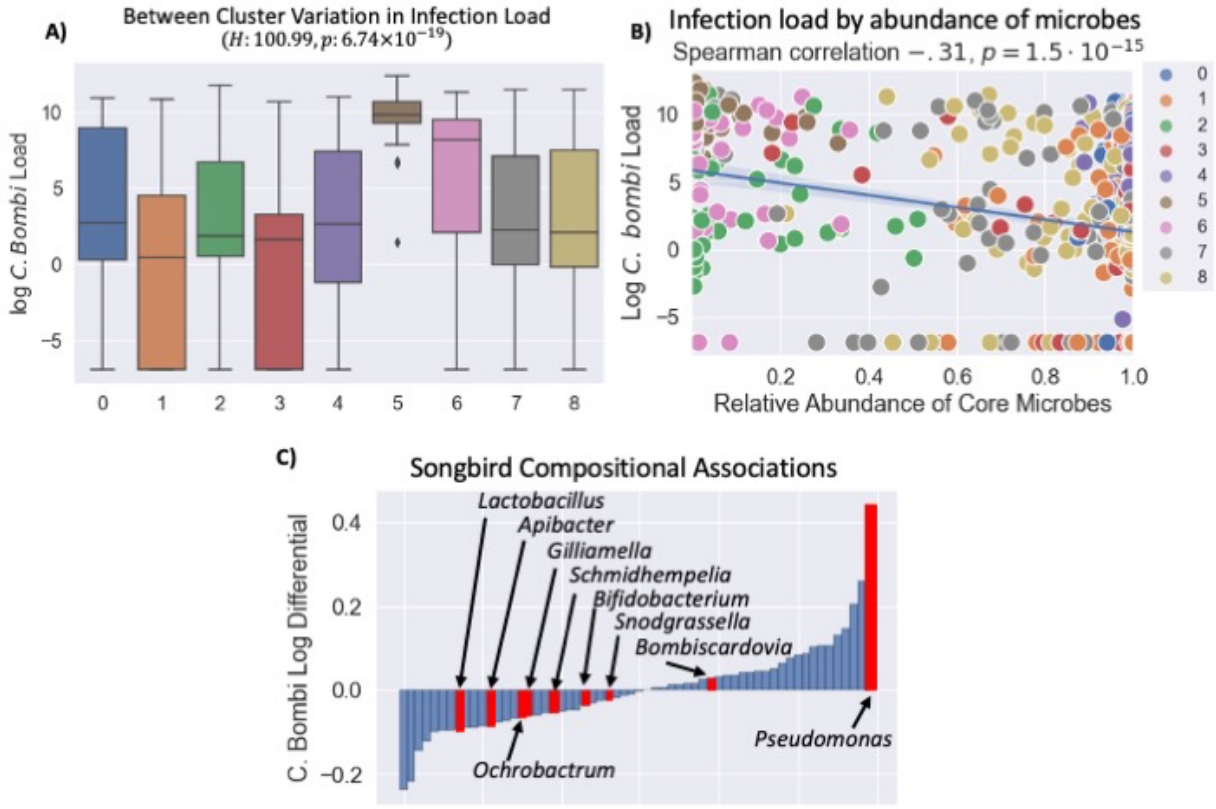


**Table 9: Cluster Infection Load Mann-Whitney U Tests**

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| **1** | $\boldsymbol{U = 1015, p = 2.6 \times 10^{-3}}$ | | | | | | | |
| **2** | $U = 1203, p = 2.9 \times 10^{-1}$ | $\boldsymbol{U = 1756, p = 1.0 \times 10^{-3}}$ | | | | | | |
| **3** | $\boldsymbol{U = 1057, p = 2.0 \times 10^{-2}}$ | $U = 2480, p = 2.0 \times 10^{-1}$ | $\boldsymbol{U = 1895, p = 2.7 \times 10^{-2}}$ | | | | | |
| **4** | $U = 1376, p = 1.7 \times 10^{-1}$ | $\boldsymbol{U = 2381, p = 1.2 \times 10^{-2}}$ | $U = 2476, p = 3.0 \times 10^{-1}$ | $\boldsymbol{U = 2321, p = 3.3 \times 10^{-2}}$ | | | | |
| **5** | $\boldsymbol{U = 248, p = 3.5 \times 10^{-8}}$ | $\boldsymbol{U = 157, p = 2.1 \times 10^{-19}}$ | $\boldsymbol{U = 323, p = 1.1 \times 10^{-11}}$ | $\boldsymbol{U = 131, p = 2.8 \times 10^{-16}}$ | $\boldsymbol{U = 248, p = 7.2 \times 10^{-15}}$ | | | |
| **6** | $\boldsymbol{U = 533, p = 3.7 \times 10^{-2}}$ | $\boldsymbol{U = 612, p = 9.9 \times 10^{-7}}$ | $\boldsymbol{U = 763, p = 1.5 \times 10^{-3}}$ | $\boldsymbol{U = 595, p = 3.1 \times 10^{-6}}$ | $\boldsymbol{U = 871, p = 4.3 \times 10^{-4}}$ | $\boldsymbol{U = 324, p = 7.6 \times 10^{-6}}$ | | |
| **7** | $U = 2543, p = 2.4 \times 10^{-1}$ | $\boldsymbol{U = 4122, p = 2.4 \times 10^{-3}}$ | $U = 4543, p = 3.9 \times 10^{-1}$ | $\boldsymbol{U = 4253, p = 3.6 \times 10^{-2}}$ | $U = 5529, p = 4.6 \times 10^{-1}$ | $\boldsymbol{U = 747, p = 9.4 \times 10^{-14}}$ | $\boldsymbol{U = 1674, p = 8.0 \times 10^{-4}}$ | |
| **8** | $U = 1545, p = 1.9 \times 10^{-1}$ | $\boldsymbol{U = 2619, p = 7.9 \times 10^{-3}}$ | $U = 2784, p = 3.3 \times 10^{-1}$ | $\boldsymbol{U = 2631, p = 4.3 \times 10^{-2}}$ | $U = 3472, p = 5.0 \times 10^{-1}$ | $\boldsymbol{U = 393, p = 2.1 \times 10^{-14}}$ | $\boldsymbol{U = 989, p = 5.2 \times 10^{-4}}$ | $U = 6144, p = 4.5 \times 10^{-1}$ |

# CHAPTER FOUR

# DISCUSSION

## 4.1 Evidence for a core microbiome

The findings of the Maine bumblebee field survey support the narrative of a robust and coevolved bumblebee microbiome. Overwhelmingly, samples from all *Bombus spp.* are primarily colonized by seven of the eight core phylotypes identified by *Kwong et al.* (Figure 6) (Kwong et al. 2017). The missing phylotype, *Parasaccaribacter*, is not included in the Silva 16S database (release 132.99) and was thus undetectable by my methods.

*Snodgrassella* and *Gilliamella* were the most frequently observed endosymbionts, colonizing just over 98% and 96% of samples, respectively (Figure 1). These high infection rates are not unexpected; in other studies involving field-caught bees, *Snodgrassella* and *Gilliamella* colonize samples to a similar extent (Cariveau et al. 2014; Powell et al. 2016; Mockler et al. 2018). Both enjoy a cosmopolitan range, inhabiting gut microbiomes of phylogenetically diverse *Bombus, Meliponini,* and *Apis* species. Their specificity to the corbiculate tribe is indicative of diversification alongside social bees (Koch et al. 2013; Powell et al. 2016; Kwong et al. 2017). Both additionally show host genus specificity; although *Bombus* and *Apis* strains can colonize sterile, non-native hosts, they are rapidly outcompeted by native strains (Kwong et al. 2014).

Besides *Snodgrassella* and *Gilliamella*, *Schmidhempelia* is the next most abundant phylotype. All three exist at dataset relative abundances of greater than 20%, while the next most abundant microbe is at just over 5% relative abundance (Figure 9). Unlike *Snodgrassella* and *Gilliamella, Schmidhempelia* is an endosymbiont of the bumblebees only (Martinson et al. 2014). *Gilliamella* and *Schmidhempelia* are members of the family *Orbaceae* and show conservation of genes involved in fermentation, likely an adaptation to the anaerobic bumblebee

hindgut (Kwong et al. 2014; Martinson et al. 2014). In contrast, *Snodgrassella* is an obligate anaerobe possessing genes for the complete TCA cycle, (Kwong et al. 2014). All three show significant genomic reduction, and likely partition niches along the aerobic gradient within the gut (Kwong et al. 2014; Martinson et al. 2014).

Bees appear to horizontally exchange microbes through local pollination networks. When all endosymbionts are treated equally, we observe a significant relationship between the geographic distance between samples and their $\beta$ diversity in two of the three years of the field survey (Table 6; Partial Mantel Tests, $p < 0.05$). When bees visit flowers, they deposit and pick up microbiota (Keller et al. 2021). The composition of pollinator species is predictive of the floral microbiome, and bee adapted microbiota are able to persist in floral environments (McFrederick et al. 2017; Zemenick et al. 2021). Thus, it is likely that flowers act as a vector of inter-colony microbial transmission. The results of the Partial Mantel tests are in support of this model, as bees from the same geographic area should forage the same flowers, reinforcing their local microbiome.

Although microbes are exchanged locally, the relative abundances of core microbiota are robust to geographic influence. In contrast to the results of the unweighted $\beta$ diversity analysis, the weighted metric shows no relationship with the geographic distance between samples (Table 5; Mantel Test, $p = 0.976$). Weighted metrics are indicators of community composition and reflect the differential abundance of major taxa (Figure 2). From these results, it appears that core microbiota colonize samples to the same extent throughout the field survey. Minor taxa are exchanged through local pollination networks, but the central features of the bumblebee microbiome are robust to external pressure. These results indicate host specificity and support previous findings of bee-endosymbiont coevolution.

Additionally, the eleven species included in the field survey appear to host compositionally similar microbiomes. By Mantel testing, host phylogeny has no relationship with weighted UniFrac distance (Table 5; p=0.49). This result is unexpected from prior investigations into the relationship between microbiome community and host phylogeny. Sympatric populations of *Apis* and *Bombus* have been found to host distinct microbiota by the unweighted Sørensen–Dice and weighted Bray-Curtis indices (Powell et al. 2016; Kwong et al. 2017). Although there is significant inter-tribe variation, it appears that individual *Bombus* species we have sampled have not compositionally diverged.

Despite the lack of a relationship between composition and host phylogeny, there is evidence for emerging variation between *Bombus* species. By unweighted UniFrac distance, the abundances of rare taxa significantly correlate with host phylogeny (Table 6; Mantel Test, p <0.001). This variation is likely driven by pollinator preference. Sympatric species partition resources through variation in tongue length, emergence time, flight efficiency, and other morphological factors (Ranta and Vepsäläinen 1981; Goulson 2010). Tongue length is an especially strong covariate of preference for shallow or deep corollas (Pyke 1982). Analogous to how bees from the same area are more likely to exchange microbes through floral vectors, foragers of the same species are more likely to choose the same flowers. In support of this model, sympatric long and short-faced species isolated to a small island host clade specific microbiota (Figure 8). Of the 395 ASVs detected in the 2018 Allen Island samples, 184 and 38 are specific to just the short and long-faced populations, respectively. Twenty ASVs corresponding to core microbiota are short-faced unique, and nine are unique to just the long-faced species.

Besides separation in qualitative community makeup, these results indicate that short-faced populations may host more phylogenetically diverse microbiota. Pollination networks are largely nested, with specialist pollinators, like long-faced bees, interacting with a subset of the plant species generalist pollinators interact with (Olesen et al. 2007). Assuming shorter tongue lengths facilitate more generalist pollination, we would expect shorter tongued species to pollinate more diverse floral resources and access greater microbial diversity in turn (Pyke 1982). The greater ASV diversity observed in short-faced bees may contribute to current trends in bumblebee conservation. Largely, rare and declining species are long-faced with narrower diets (Goulson and Darvill 2004).

### 4.2 Gut flora change with the seasons

The single significant predictor of microbiome composition is date of collection. All years show correlations between weighted UniFrac distance and difference in sample collection date, with significant Partial Mantel tests for 2018 and 2019 (Figure 6; Table 5; $p < 0.001$). The bumblebee lifecycle and the changing availability of floral resources may be responsible for this observed relationship.

In honeybees, resource availability is a driver of compositional variation, indicating that the relative abundances of core microbiota adapts to diet (Jones et al. 2018). Over the winter, honeybee microbiomes further change, with increased bacterial load but lower $\alpha$-diversity (Kešnerová et al. 2020). The bumblebee lifecycle is more extreme. Individual queens hibernate over the winter and form new colonies each spring. This is likely a major bottleneck on bumblebee microbial diversity. Population size drops dramatically and endosymbionts vary in overwintering success (Koch et al. 2013). In the field survey data, each year has a unique set of

low abundance ASVs, possibly due to the hibernation bottleneck (Figure 7). Samples caught close together in time have more similar microbiota, even when geography is controlled for (Figure 6). The colony lifecycle and distribution of floral resources both correlate with collection date and may be the causal agents of this association.

### 4.3 Wild populations have high compositional variation

Besides the moderate relationship with collection date, there is evidence for unknown drivers of compositional variation between samples. A Uniform Manifold Approximation and Projection of the weighted UniFrac distance matrix shows support for 9 distinct clusters (Figure 9). I chose to use UMAP for dimensionality reduction because it outperforms comparable algorithms on sensitivity to classification problems involving biological count data (Dorrity et al. 2020).

I used the mean silhouette coefficient to assess support for clustering (Rousseeuw 1987). Silhouette coefficients are bounded between 0 and 1. Our observed mean coefficient of 0.59 indicates high support for clusters. For comparison, in a study applying similar methods to cluster human microbiomes, an average coefficient of just over 0.2 was taken as strong support for clustering (Arumugam et al. 2011). It is not immediately clear why bumblebee microbiomes cluster with greater success. The inclusion of multiple species should theoretically introduce interspecies variation, but as previously shown, phylogeny is not a driver of community composition. The low diversity of high abundance phylotypes may lend bumblebees to microbiome clustering, as they have fewer axes of major compositional variation (Figure 6). Additionally, the relatively large sample size of the bumblebee field survey (638 samples) facilitates the specification of more numerous and tightly defined clusters.

In total, 32 genus-level phylotypes are differentially abundant between clusters, including the seven core microbiota detected in the dataset (Figure 9; ANCOM-BC, holm adj. $q < 0.05$). I selected ANCOM-BC for differential abundance testing for its high sensitivity and control for false discovery through awareness of variation in sampling fraction (Lin and Peddada 2020). The large sample size further limited false discovery, as indicated by the moderate proportion of differentially abundant phylotypes detected (32 of 151). However, the high number of clusters introduced comparisons that may not be entirely informative. Of the 25 non-core phylotypes found to be differentially abundant, 21 have dataset abundances less than 1%. The observed differential abundance of these fringe taxa may be influenced by the stochastic randomness of low abundance features.

The UMAP plot shows a major axis of variation along the upper diagonal, with divergent points falling to the right and below (Figure 9). The lack of a "horseshoe" pattern to the distribution of points indicates that the inclusion of phylogenetic information to the weighted UniFrac dissimilarities prevented oversaturation of the metric (Morton et al. 2017).

The upper diagonal of the plot primarily corresponds to variation in relative abundances of core microbiota. On the bottom left, samples are characterized by near 100% abundances of *Snodgrassella.* On the top right, samples are colonized by *Schmidhempelia* to relative abundances greater than 80%. From other studies, extreme colonization by a single taxa appears to be rare in wild bumblebees (Cariveau et al. 2014; Li et al. 2015; Powell et al. 2016; Mockler et al. 2018). Samples along the middle of the axis harbor the more canonical bumblebee microbiome community and samples detached from the main diagonal are characterized by high abundances of non-bumblebee-associated phylotypes.

The clusters along the main diagonal are all more frequently observed than the detached clusters (Figure 10). From this, dysbiosis, or an imbalance of the typical microbiome, appears to be relatively uncommon in the dataset. Additionally, bees in states of dysbiosis have highly variable microbiome communities. Some are characterized by extreme colonization by single taxa, while others are more diverse. There is high compositional variation among bees with typical gut microbiota, as well as among those with atypical.

### 4.4 The relative abundance of core microbiota is negatively associated with *C. bombi* infection

From experimental results, it is apparent that core microbiota have a functional role in *Crithidia bombi* immunity (Mockler et al. 2018). However, the exact mechanisms remain unknown. In the field survey data, microbiome compositional clusters differ significantly in their average *C. bombi* infection severity (Figure 12, Kruskal Wallis Ranked Sum Test, $p = 6.74 \times 10^{-19}$). Largely, this variation appears to be driven by differences in proportion of core microbiota, a finding in agreement with prior results (Figure 12; $\rho = -.31$, t-test, $p = 1.5 \times 10^{-15}$) (Koch and Schmid-Hempel 2011; Cariveau et al. 2014; Mockler et al. 2018).

Differentials calculated by Songbird (v.1.0.4) provide a ranking of microbe compositional associations with *Crithidia* infection severity. All core microbiota show negative relationships with *Crithidia* infections, save *Bombiscardovia* (Figure 12). The rankings show that high abundances of *Lactobacillus* and *Apibacter* are strong indicators of a healthy bee, and that *Pseudomonas* is a marker of *Crithidia* infections. By differentials, relative abundance of *Snodgrassella* has a very minor association with *Crithidia* infection severity.

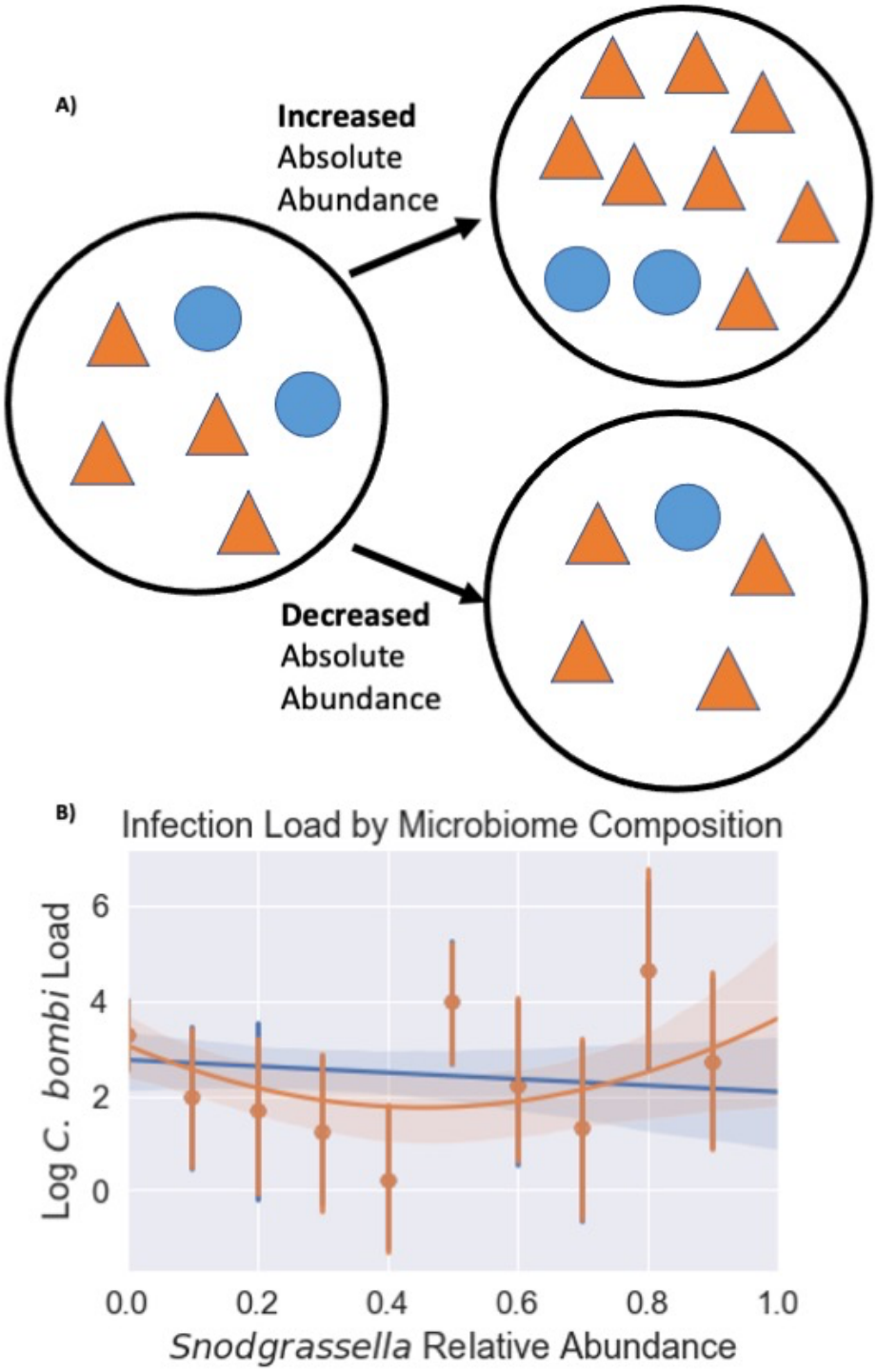
The low differential ranking of *Snodgrassella* is counterintuitive and may be an artifact of the non-independence of compositional data. The log differential rankings capture change in relative, not absolute abundance (Morton et al. 2019). From purely compositional data, it is impossible to tell which taxa are changing in abundance (Figure 13A). However, through prior results, we are aware of *Snodgrassella's* metabolic niche in the bumblebee gut. *Snodgrassella* is an obligate aerobe that produces biofilm in direct contact with host epithelial cells (Kwong et al. 2014). Although *Snodgrassella* has the greatest dataset abundance of all phylotypes in the field survey, its niche within the gut is restricted along the aerobic gradient. By this logic, it appears highly likely that samples with apparent extreme relative abundances of *Snodgrassella* actually suffer from decimated gut microbiota.

To test this hypothesis, we can look directly at the relationship between *Snodgrassella* abundance and infection severity. If the relative abundances of *Snodgrassella* consistently reflect absolute abundances, we would expect a weak, negative linear trend. If instead, extreme relative abundances of *Snodgrassella* indicate a decimated anaerobic microbiome, we would expect a quadratic relationship; infection severity would be high for samples lacking *Snodgrassella* and samples with only *Snodgrassella*. This is exactly what is seen in the data (Figure 13). By likelihood ratio testing, the quadratic model explains significantly more variance than a model with just a linear term ($p = 1.11 \times 10^{-19}$).

Due to the compositional nature of 16S sequencing data, naïve interpretations of abundance analyses can lead to conclusions at odds with true biological phenomena. From the results of the field survey, it appears that the bumblebee core microbiota do not change in abundance at equal rates. *Snodgrassella* shows significant affinity for the bumblebee gut, remaining at high abundances even when much of the rest of the gut community has disappeared.

Relative abundances of individual core phylotypes are coarse and sometimes misleading metrics for community health. Overall community composition, rather than abundance of individual taxa, appears to be the strongest predictor of host health.

**Figure 13:** ***The relative abundance of core microbiota is associated with C. bombi infection. A)*** *Relative differentials are not absolute differentials.* In this mock example, circles and triangles represent two bumblebee endosymbionts. An increase in the absolute abundance of triangles and decrease in abundance of circles cause the same change in relative abundance. Through compositional data alone, it is impossible to differentiate between these two scenarios. ***B)*** *Infection load by microbiome composition.* The relative abundance of *Snodgrassella* has a quadratic trend with infection severity. The GLM including the quadratic term explains significantly more variation than a model with just a linear term ($Likelihood\ Ratio, p = 1.11 \times 10^{-19}$).


A)
Increased
Absolute
Abundance
Decreased
Absolute
Abundance
B)
Infection Load by Microbiome Composition
Log C. bombi Load
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0
Snodgrassella Relative Abundance

ACKNOWLEDGEMENTS

My thesis has been many years in the making. I gave my first bumblebee poster at the 2018 Colby Undergraduate Summer Research Retreat. After a long summer of field work, another research assistant and I presented an investigation into the $\alpha$-diversity of samples with *Crithidia* infections detected through amplicon sequencing. Despite our lack of experience and unawareness of "significance" (neither of us had taken any statistics), we were encouraged to present by Professor Angelini and other members of the lab group. In the years between then and now, I have been repeatedly encouraged by my professors and members of the Angelini Lab to return to the bumblebee project. Without them, I am sure that I would not be the scientist that I am today. This draft benefits from my education in genomics, machine learning, evolutionary biology, and statistical analysis.

I would like to acknowledge the contributions of a few specific collaborators. First, Professor Angelini deserves acknowledgement for getting me started with bumblebee microbiomes and enabling me as an undergraduate researcher ever since. Second, my readers Professor Noh and Professor Moore are acknowledged for providing invaluable feedback and guidance, without going so far as to direct the project for me. Third, I would like to recognize the contributions of the other research assistants involved with the bumblebee project, most significantly Josefine Just, Jane Lee, Thomas McMahon, and James Gonzalez. My project benefits directly from their hard work.

Research reported here was supported by the Colby College Division of Natural Sciences, by the McVey Data Science Initiative, by the Up East Foundation, and by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM0103423.

# REFERENCES


Althoff DM, Segraves KA, Johnson MTJ (2014) Testing for coevolutionary diversification: Linking pattern with process. Trends Ecol Evol 29:82–89. https://doi.org/10.1016/j.tree.2013.11.003

Arumugam M, Raes J, Pelletier E, et al (2011) Enterotypes of the human gut microbiome. Nature 473:174–180. https://doi.org/10.1038/nature09944

Bolyen E, Rideout JR, Dillon MR, et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9

Brown MJF, Loosli R, Schmid-Hempel P (2000) Condition-dependent expression of virulence in a trypanosome infecting bumblebees. Oikos 91:421–427. https://doi.org/10.1034/j.1600-0706.2000.910302.x

Brown MJF, Schmid-Hempel R, Schmid-Hempel P (2003) Strong context-dependent virulence in a host-parasite system: Reconciling genetic evidence with theory. J Anim Ecol 72:994–1002. https://doi.org/10.1046/j.1365-2656.2003.00770.x

Caliński T, Harabasz J (1974) A Dendrite Method For Cluster Analysis. Commun Stat 3:1–27. https://doi.org/10.1080/03610927408827101

Callahan BJ, McMurdie PJ, Rosen MJ, et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869

Cameron SA, Hines HM, Williams PH (2007) A comprehensive phylogeny of the bumble bees ( Bombus ). The Linnean Society of London

Cameron SA, Lozier JD, Strange JP, et al (2011) Patterns of widespread decline in North American bumble bees. Proc Natl Acad Sci U S A 108:662–667. https://doi.org/10.1073/pnas.1014743108

Cariveau DP, Elijah Powell J, Koch H, et al (2014) Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (Bombus). ISME J 8:2369–2379. https://doi.org/10.1038/ismej.2014.68

Caughley G (1994) Directions in Conservation Biology. J Anim Ecol 63:215–244

Chittka L, Gumbert A, Kunze J (1997) Foraging dynamics of bumble bees: Correlates of movements within and between plant species. Behav Ecol 8:239–249.

https://doi.org/10.1093/beheco/8.3.239

Colla SR, Packer L (2008) Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on Bombus affinis Cresson. Biodivers Conserv 17:1379–1391. https://doi.org/10.1007/s10531-008-9340-5

Cresswell JE, Osborne JL, Goulson D (2000) An economic model of the limits to foraging range in central place foragers with numerical solutions for bumblebees. Ecol. Entomol. 25:249–255

Cresswell JE, Robertson AW (1994) Discrimination by pollen-collecting bumblebees among differentially rewarding flowers of an alpine wildflower, Campanula rotundifolia (Campanulaceae). Oikos 304–308

Cumber RA (1953) Some aspects of the biology and ecology of bumble-bees bearing upon the yields of red-clover seed in New Zealand. New Zeal J Sci Technol 11:227–240

Darvill B, Ellis JS, Lye GC, Goulson D (2006) Population structure and inbreeding in a rare and declining bumblebee, Bombus muscorum (Hymenoptera: Apidae). Mol Ecol 15:601–611. https://doi.org/10.1111/j.1365-294X.2006.02797.x

Desneux N, Decourtye A, Delpuech JM (2007) The sublethal effects of pesticides on beneficial arthropods. Annu Rev Entomol 52:81–106. https://doi.org/10.1146/annurev.ento.52.110405.091440

Dively GP, Kamel A (2012) Insecticide residues in pollen and nectar of a cucurbit crop and their potential exposure to pollinators. J Agric Food Chem 60:4449–4456. https://doi.org/10.1021/jf205393x

Donovan BJ, Wier SS (1984) Development of hives for field population increase, and studies on the life cycles of the four species of introduced bumble bees in New Zealand. New Zeal J Agric Res 21:733–756. https://doi.org/10.1080/00288233.1978.10427476

Dornhaus A, Chittka L (2001) Food alert in bumblebees (Bombus terrestris): Possible mechanisms and evolutionary implications. Behav Ecol Sociobiol 50:570–576. https://doi.org/10.1007/s002650100395

Dorrity MW, Saunders LM, Queitsch C, et al (2020) Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nat Commun 11:. https://doi.org/10.1038/s41467-020-15351-4

Duchateau MJ, Hoshiba H, Velthuis HHW (1994) Diploid males in the bumble bee Bombus

terrestris Sex determination, sex alleles and viability. Entomol Exp Appl 71:263–269. https://doi.org/10.1111/j.1570-7458.1994.tb01793.x

Durrer S, Schmid-Hempel P (1994) Shared use of flowers leads to horizontal pathogen transmission. Proc R Soc B Biol Sci 258:299–302. https://doi.org/10.1098/rspb.1994.0176

Ellegaard KM, Tamarit D, Javelind E, et al (2015) Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. BMC Genomics 16:1–22. https://doi.org/10.1186/s12864-015-1476-6

Ellington CP, Machin KE, Casey TM (1990) Oxygen consumption of bumblebees in forward flight. Nature 472–473

Engel P, Martinson VG, Moran NA (2012) Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A 109:11002–11007. https://doi.org/10.1073/pnas.1202970109

Ferrari MJ, Perkins SE, Pomeroy LW, Bjrnstad ON (2011) Pathogens, social networks, and the paradox of transmission scaling. Interdiscip Perspect Infect Dis 2011:. https://doi.org/10.1155/2011/267049

Friedman J, Alm EJ (2012) Inferring Correlation Networks from Genomic Survey Data. PLoS Comput Biol 8:1–11. https://doi.org/10.1371/journal.pcbi.1002687

Gerloff CU, Schmid-Hempel P (2005) Inbreeding depression and family variation in a social insect, Bombus terrestris (Hymenoptera: Apidae). Oikos 111:67–80. https://doi.org/10.1111/j.0030-1299.2005.13980.x

Goulson D (2010) bumblebees behaviour, ecology, and conservation, Second. Oxford University Press

Goulson D, Darvill B (2004) Niche overlap and diet breadth in bumblebees; are rare species more specialized in their choice of flowers? Apidologie 35:55–63. https://doi.org/10.1051/apido

Goulson D, Hanley ME, Darvill B, et al (2004) Causes of rarity in bumblebees. Biol Conserv 122:1–8. https://doi.org/10.1016/j.biocon.2004.06.017

Goulson D, Hanley ME, Darvill B, Ellis JS (2006) Biotope associations and the decline of bumblebees (Bombus spp.). J Insect Conserv 10:95–103. https://doi.org/10.1007/s10841-006-6286-3

Goulson D, Lye GC, Darvill B (2008) Decline and conservation of bumble bees. Annu Rev

Entomol 53:191–208. https://doi.org/10.1146/annurev.ento.53.103106.093454

Goulson D, Nicholls E, Botías C, Rotheray EL (2015) Bee declines driven by combined Stress from parasites, pesticides, and lack of flowers. Science (80- ) 347:. https://doi.org/10.1126/science.1255957

Goulson D, Peat J, Stout JC, et al (2002) Can alloethism in workers of the bumblebee, bombus terrestris, be explained in terms of foraging efficiency? Anim Behav 64:123–130. https://doi.org/10.1006/anbe.2002.3041

Goulson D, Stout JC (2001) Homing ability of the bumblebee Bombus terrestris (Hymenoptera: Apidae). Apidologie 32:105–111. https://doi.org/10.1051/apido:2001115

Greenleaf SS, Kremen C (2006) Wild bee species increase tomato production and respond differently to surrounding land use in Northern California. Biol Conserv 133:81–87. https://doi.org/10.1016/j.biocon.2006.05.025

Groussin M, Mazel F, Alm EJ (2020) Co-evolution and Co-speciation of Host-Gut Bacteria Systems. Cell Host Microbe 28:12–22. https://doi.org/10.1016/j.chom.2020.06.013

Heinrich B (1996) The Thermal Warriors. Harvard University Press, Cambridge

Heinrich B (1979) Bumblebee Economics. Harvard University Press

Henry M, Béguin M, Requier F, et al (2012) A Common Pesticide Decreases Foraging Success and Survival in Honey Bees. Science (80- ) 336:3–5

Hines HM (2008) Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: Bombus). Syst Biol 57:58–75. https://doi.org/10.1080/10635150801898912

Imhoof B, Schmid-Hempel P (1999) Colony success of the bumble bee, Bombus terrestris, in relation to infections by two protozoan parasites, Crithidia bombi and Nosema bombi. Insectes Soc 46:233–238. https://doi.org/10.1007/s000400050139

Imhoof B, Schmid-Hempel P (1998) Single-clone and mixed-clone infections versus host environment in Crithidia bombi infecting bumblebees. Parasitology 117:331–336. https://doi.org/10.1017/S0031182098003138

Jones JC, Fruciano C, Hildebrand F, et al (2018) Gut microbiota composition is associated with environmental landscape in honey bees. Ecol Evol 8:441–451. https://doi.org/10.1002/ece3.3597

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:

Improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Keller A, McFrederick QS, Dharampal P, et al (2021) (More than) Hitchhikers through the network: The shared microbiome of bees and flowers. Curr Opin Insect Sci 44:8–15. https://doi.org/10.1016/j.cois.2020.09.007

Kešnerová L, Emery O, Troilo M, et al (2020) Gut microbiota structure differs between honeybees in winter and summer. ISME J 14:801–814. https://doi.org/10.1038/s41396-019-0568-8

Koch H, Abrol DP, Li J, Schmid-Hempel P (2013) Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. Mol Ecol 22:2028–2044. https://doi.org/10.1111/mec.12209

Koch H, Schmid-Hempel P (2011) Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. Proc Natl Acad Sci U S A 108:19288–19292. https://doi.org/10.1073/pnas.1110474108

Kosior A, Celary W, Olejniczak P, et al (2007) The decline of the bumble bees and cuckoo bees (Hymenoptera: Apidae: Bombini) of Western and Central Europe. Oryx 41:79–88. https://doi.org/10.1017/S0030605307001597

Kwong WK, Engel P, Koch H, Moran NA (2014) Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc Natl Acad Sci U S A 111:11509–11514. https://doi.org/10.1073/pnas.1405838111

Kwong WK, Medina LA, Koch H, et al (2017) Dynamic microbiome evolution in social bees. Sci Adv 3:1–17. https://doi.org/10.1126/sciadv.1600513

Kwong WK, Steele MI, Moran NA (2018) Genome sequences of Apibacter spp., gut symbionts of Asian honey bees. Genome Biol Evol 10:1174–1179. https://doi.org/10.1093/gbe/evy076

Li J, Powell JE, Guo J, et al (2015) Two gut community enterotypes recur in diverse bumblebee species. Curr Biol 25:R652–R653. https://doi.org/10.1016/j.cub.2015.06.031

Lin H, Peddada S Das (2020) Analysis of compositions of microbiomes with bias correction. Nat Commun 11:1–11. https://doi.org/10.1038/s41467-020-17041-7

Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 73:1576–1585. https://doi.org/10.1128/AEM.01996-06

Martinson VG, Mago T, Koch H, et al (2014) Genomic features of a bumble bee symbiont reflect its host environment. Appl Environ Microbiol 80:3793–3803. https://doi.org/10.1128/AEM.00322-14

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

McFrederick QS, Thomas JM, Neff JL, et al (2017) Flowers and Wild Megachilid Bees Share Microbes. Microb Ecol 73:188–200. https://doi.org/10.1007/s00248-016-0838-1

McInnes L, Healy J, Melville J (2018) UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv

Michener CD (1974) The Social Behavior of the Bees: A Comparative Study, Second. Harvard University Press, Cambridge

Mockler BK, Kwong WK, Moran NA, Koch H (2018) Microbiome structure influences infection by the parasite Crithidia bombi in bumble bees. Appl Environ Microbiol 84:1–11. https://doi.org/10.1128/AEM.02335-17

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190. https://doi.org/10.1146/annurev.genet.41.110306.130119

Moret Y, Schmid-Hempel P (2000) Survival for immunity: The price of immune system activation for bumblebee workers. Science (80- ) 290:1166–1168. https://doi.org/10.1126/science.290.5494.1166

Morton JT, Marotz C, Washburne A, et al (2019) Establishing microbial composition measurement standards with reference frames. Nat Commun 10:. https://doi.org/10.1038/s41467-019-10656-5

Morton JT, Toran L, Edlund A, et al (2017) Uncovering the Horseshoe Effect in Microbial Analyses. mSystems 2:1–8

Müller CB, Schmid-Hempel P (1993) Correlates of reproductive success among field colonies of Bombus lucorum L.: the importance of growth and parasites. Ecol Entomol 17:343–353

Müller H (1883) The effect of the change of colour in the flowers of Pulmonaria officinalis upon its fertilizers. Nature 81

Nowak MA, Tarnita CE, Wilson EO (2010) The evolution of eusociality. Nature 466:1057–1062. https://doi.org/10.1038/nature09205

O'Donnell S, Reichardt M, Foster R (2000) Individual and colony factors in bumble bee division of labor (Bombus bifarius nearcticus Handl; Hymenoptera, Apidae). Insectes Soc 47:164–170. https://doi.org/10.1007/PL00001696

Olesen JM, Bascompte J, Dupont YL, Jordano P (2007) The modularity of pollination networks. Proc Natl Acad Sci U S A 104:19891–19896. https://doi.org/10.1073/pnas.0706375104

Otterstatter MC, Thomson JD (2007) Contact networks and transmission of an intestinal pathogen in bumble bee (Bombus impatiens) colonies. Oecologia 154:411–421. https://doi.org/10.1007/s00442-007-0834-8

Packer L, Owen R (2001) Population Genetic Aspects of Pollinator Decline Conservation Ecology. Ecol Soc 5:1–24

Park HS, Jun CH (2009) A simple and fast algorithm for K-medoids clustering. Expert Syst Appl 36:3336–3341. https://doi.org/10.1016/j.eswa.2008.01.039

Pascall DJ, Tinsley MC, Clark BL, et al (2021) Predictors of virus prevalence and diversity across a wild bumblebee community. bioRxiv

Peat J, Goulson D (2005) Effects of experience and weather on foraging rate and pollen versus nectar collection in the bumblebee, Bombus terrestris. Behav Ecol Sociobiol 58:152–156. https://doi.org/10.1007/s00265-005-0916-8

Powell E, Ratnayeke N, Moran NA (2016) Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. Mol Ecol 25:4461–4471. https://doi.org/10.1111/mec.13787

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5:. https://doi.org/10.1371/journal.pone.0009490

Prys-Jones OE (1982) Ecological studies of foraging and life history in bumblebees. University of Cambridge

Pyke GH (1982) Local geographic distributions of bumblebees near Crested Butte, Colorado: competition and community structure. Ecology 555–573

Quast C, Pruesse E, Yilmaz P, et al (2013) The SILVA ribosomal RNA gene database project:

Improved data processing and web-based tools. Nucleic Acids Res 41:590–596. https://doi.org/10.1093/nar/gks1219

Ranta E, Vepsäläinen K (1981) Why are there so many species? Spatio-temporal heterogeneity and northern bumblebee communities. Oikos 28–34

Röseler P-F (1991) Roles of morphogenetic hormones in caste polymorphism in bumble- bees. In: Gupta AP (ed) Morphogenetic Hormones in Arthropods: Roles in Histogenesis, Organogenesis, and Morphogenesis. Rutgers University Press, New Brunswick, NJ, pp 384–399

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Satchell KJF (2011) Structure and function of MARTX toxins and other large repetitive RTX proteins. Annu Rev Microbiol 65:71–90. https://doi.org/10.1146/annurev-micro-090110-102943

Schmid-Hempel P, Schmid-Hempel R (1993) Transmission of a pathogen in Bombus terrestris, with a note on division of labour in social insects. Behav Ecol Sociobiol 33:319–327

Schmid-Hempel R, Schmid-Hempel P (2000) Female mating frequencies in Bombus spp. from Central Europe. Insectes Soc 47:36–41. https://doi.org/10.1007/s000400050006

Shykoff JA, Schmid-Hempel P (1991) Incidence and effects of four parasites in natural populations in bumble bees in Switzerland. Apidologie 22:117–126

Simmons WR, Angelini DR (2017) Chronic exposure to a neonicotinoid increases expression of antimicrobial peptide genes in the bumblebee Bombus impatiens. Sci Rep 7:1–9. https://doi.org/10.1038/srep44773

Simon-Delso N, Amaral-Rogers V, Belzunces LP, et al (2015) Systemic insecticides (Neonicotinoids and fipronil): Trends, uses, mode of action and metabolites. Environ Sci Pollut Res 22:5–34. https://doi.org/10.1007/s11356-014-3470-y

Smirnova E, Huzurbazar S, Jafari F (2019) PERFect: PERmutation filtering test for microbiome data. Biostatistics 20:615–631. https://doi.org/10.1093/biostatistics/kxy020

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of

progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680. https://doi.org/10.1093/nar/22.22.4673

Thompson JN (1989) Concepts of coevolution. Trends Ecol Evol 4:179–183. https://doi.org/10.1016/0169-5347(89)90125-0

Tung J, Barreiro LB, Burns MB, et al (2015) Social networks predict gut microbiome composition in wild baboons. Elife 2015:1–18. https://doi.org/10.7554/eLife.05224

Weiss S, Xu ZZ, Peddada S, et al (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5:1–18. https://doi.org/10.1186/s40168-017-0237-y

Whitehorn PR, O'Connor S, Wackers FL, Goulson D (2012) Neonicotinoid pesticide reduces bumble bee colony growth and queen production. Science (80- ) 336:351–352. https://doi.org/10.1126/science.1215025

Wilcove DS, Rothstein D, Dubow J, et al (1998) Quantifying Threats to Imperiled Species in the United States. Bioscience 48:607–615. https://doi.org/10.2307/1313420

Williams PH, Osborne JL (2009) Bumblebee vulnerability and conservation world-wide. Apidologie 40:367–387. https://doi.org/10.1051/apido/2009025

Yourth CP, Schmid-Hempel P (2006) Serial passage of the parasite Crithidia bombi within a colony of its host, Bombus terrestris, reduces success in unrelated hosts. Proc R Soc B Biol Sci 273:655–659. https://doi.org/10.1098/rspb.2005.3371

Zemenick AT, Vanette RL, Rosenheim JA (2021) Linked networks reveal dual roles of insect dispersal and species sorting for bacterial communities in flowers. Oikos 1–11. https://doi.org/10.1111/oik.06818