ABSTRACT

STOTT, CAROLINE MARIE. Gene Transfers and Speciation Patterns in the Corbiculate Bee Microbiome. (Under the direction of Dr. Louis-Marie Bobay, Dr. Kasie Raymann).

Microbial communities that maintain symbiotic relationships with animals evolve by adapting to the specific environmental niche provided by their host, yet understanding their patterns of speciation remains challenging. Whether bacterial speciation occurs primarily through allopatric or sympatric processes remains an open question. In addition, patterns of DNA transfers, which are pervasive in bacteria, are more constrained in a closed host-gut system. Eusocial bees have co-evolved with their specialized microbiota for over 85 million years, constituting a simple and valuable system to study the complex dynamics of host-associated microbial interactions. Here we studied the patterns of speciation and evolution of seven specialized gut bacteria from three clades of eusocial bee species: western honey bees, eastern honey bees and bumblebees. We conducted genomic analyses to infer species delineation relative to the patterns of homologous recombination (HR), and horizontal gene transfer (HGT). The studied bacteria presented various modes of evolution and speciation relative to their hosts, but some trends were consistent across all of them. We observed a clear interruption of homologous recombination between bacteria inhabiting different bee hosts, which is consistent with a mechanism of allopatric speciation, but we also identified interruptions of homologous recombination within hosts, suggesting recent or ongoing sympatric speciation. In contrast to HR, we observed that HGT events were not constrained by species borders. Overall, our findings show that in host-associated bacterial populations, patterns of HR and HGT have different impacts on speciation patterns, which are driven by both allopatric and sympatric speciation processes.

Gene Transfers and Speciation Patterns within the Corbiculate Bee Microbiome

by
Caroline Stott

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Microbiology

Raleigh, North Carolina
2025

APPROVED BY:

Dr. Louis-Marie Bobay Committee Co-Chair	Dr. Kasie Raymann Committee Co-Chair
Dr. Michael Hyman	Dr. Carlos C. Goller

BIOGRAPHY

Caroline was born in Greensboro, NC and attended the University of North Carolina at Greensboro before transferring to Pratt Institute in Brooklyn, NY to pursue a career in the arts. She received a BFA in Graphic Design and worked as an interactive designer creating web and game interfaces. Driven by her passion for biological sciences, she made the decision to pivot her career, returning to UNCG to earn her bachelor's degree in biology. She then joined the labs of Drs Kasie Raymann and Louis-Marie Bobay to pursue a graduate degree researching microbial evolution.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: Introduction and background

Part 1: Prokaryotic Evolution

1.1.1: The species problem

Prokaryotes are the most abundant organisms on the planet (Whitman et al. 1998), yet it is only in recent decades that we have begun unravelling the dynamics behind their rapid and efficient adaptive capabilities under selective pressures. Recent advances in the speed and availability of DNA sequencing technology have allowed for the close examination of the diversity of bacterial populations and their genomic structures. It is now known that bacteria: (i) are asexual, but less than 10% of those currently sequenced are found to be actually clonal (Shapiro 2016; Diop et al. 2022), (ii) evolve very quickly via mutations and genetic transfers, even if no environmental changes occur (Good et al. 2017; Lenski 2017), and (iii) are capable of laterally exchanging DNA with one another, even if distantly related (Van Der Meer et al. 1992; Ochman et al. 2000; Daubin and Szöllősi 2016). Attempts to reconcile these patterns in the context of Darwinian evolution, which aims to narrow groups of organisms into species who have inherited advantageous traits from their ancestors, have proved problematic (Cohan 2002; Shapiro et al. 2016). As such, there is currently not a universally accepted set of criteria for classifying prokaryotic organisms into distinct species. There are those that argue that microbial species cannot actually exist given that the evolutionary history of an individual bacterial genome depends on the section of genetic material being observed (Lawrence 2002; Doolittle 2019) and that the seemingly conflicting observations of bacterial processes "…highlight the ontological vacuity of that vexed word." (Doolittle 2012). Others argue that while species borders might be a bit "fuzzy" due to genetic exchange, prokaryotic species are a biological reality that exist as distinct genomic groups of individuals (Konstantinidis et al. 2006; Hanage

2013; Hanage 2016; Conrad et al. 2024). Either way, a more structured taxonomic framework has important implications for assessing plant and animal health, advancing agricultural and industrial technology, and is essential for proper communication within and between various fields of research. For eukaryotes, the most common theoretical concept for delineating organisms into species is the Biological Species Concept (BSC) (Mayr 1942). This framework aids in our understanding of what constitutes a species, stating that populations form cohesive clusters based on phenotypic and genomic similarities that are maintained by genetic recombination during mating. Speciation and the borders between species are defined in terms of barriers to sexual reproduction which disrupts genetic cohesion between populations. Endeavors to develop species concepts tailored for microbial populations have been ongoing for the last several decades and are rife with controversy, ambiguity, and philosophical disagreements.

1.1.2 Methods for identifying bacterial taxa

For microbial populations, methods of detecting phenotypic and genomic relatedness has allowed for the designation of "species" groups based on DNA identity. DNA-DNA hybridization (DDH) was the first method used to detect nucleic acid sequence homology. DNA from two organisms' genomes sharing >70% identity were considered to be the same "species". (Yasuda et al. 1987). Later, the identification of regions of DNA that represent allelic variants of highly conserved "housekeeping" genes allowed for the development of PCR primers for targeting and amplifying these genes for downstream sequencing and classification. Historically, the 16s rRNA housekeeping gene subunit has been the most widely used genetic marker for the identification of bacteria due to its universality and slow evolution (Woese 1990). Using this method, operational taxonomic units (OTUs) can be classified sharing a 97-99% sequence identity. Such results are generally reliable at the genus level, but resolution at the species or strain level remains debatable (Mignard and Flandrois 2006; Bartoš et al. 2024). The variable regions of the 16s gene are also used to identify taxa and their relative abundance from deep sequenced metagenomic datasets of entire communities without the need to culture bacterial isolates. Multilocus sequence typing (MLST) builds allelic profiles for strains by detecting variations in multiple (up to seven) housekeeping genes from whole genome sequence (WGS) or metagenomic datasets. The development of a catalogued database of ribosomal subunit reference sequences that are more diverse than the 16s region aims to classify bacteria 'from domain to strain' (Jolley et al. 2012; Larsen et al. 2012; Maiden et al. 2013). Similarly, metagenomic amplicon strain typing (MAST) has been developed to capture strain level diversity of taxa in natural communities, but instead uses custom primers designed to target unique housekeeping genes identified from specific strains (Bobay et al. 2020).

Following the improvement of sequencing technologies, a more recent and widely accepted means to measure sequence homology is the computational pairwise calculation of the Average Nucleotide Identity (ANI) between pairs of fully sequenced genomes (Jain et al. 2018). Using this method, bacterial genomes are classified into "species" via a cutoff of ≥95% or ≥94% using the set of genes shared by each genome pair (Rosselló-Mora 2001; Konstantinidis and Tiedje 2005). The results of this sequence-derived cutoff point correspond to the previously used 70% DDH criterion (Goris et al. 2007), and offers a nominal method to classify bacterial species based on genetic identity; however, a single threshold may not be applicable across all prokaryotic taxa given the wide array of ways that they differ in their ecology and modes of evolution (Bobay 2020).

All of these methods are profoundly useful for identifying bacterial taxa as clusters of individuals with similar genetic makeup that can be used to create "species" groups. However, they do not provide us with a usable species concept beyond arbitrary or empirical numerical assignments of nucleotide identity thresholds and evolutionary divergence.

1.1.3 Species concepts

The Phylogenetic Species Concept aims to classify species based on their patterns of ancestry that cluster individuals into monophyletic groups (Cracraft 1983; de Queiroz and Gauthier 1994). Trees can be computationally built using a selected housekeeping gene, or a set of genes, and these alignments can be used to measure evolutionary distance/divergence, reveal common ancestry relative to each other, and create databases of taxonomic groups of bacteria (Parks et al. 2018). This concept is useful, as it can be applied to both eukaryotes and prokaryotes, but it is argued that it is not enough on its own because it does not "take into adequate account established population genetic principles" (Avise and Wollenberg 1997).

A bacterial population's ecology plays a crucial role in its genetic functionality and has inspired the Stable Ecotype Model (SEM) (Cohan and Perry 2007) which posits that a species' homogeneity is maintained by recurrent genome-wide selective sweeps for an environmental niche which purges diversity that might have accumulated within a given "ecotype". Under this mode, constant competition and the periodical genome sweeps maintain the genetic diversity of the species. Proponents of the SEM argue that ecology is the cohesive force defining an ecotype regardless of the possibility for genetic exchange (Polz et al. 2006; Ward et al. 2006). Occasionally, one advantageous mutant can colonize a new (micro)niche, thus creating a new ecotype, and, without competing with the previous population, this new ecotype diverges genetically from the original population. This concept likely only applies to fairly clonal populations; indeed, a population seemingly demonstrating the SEM was found to exhibit contradictory behavior when two divergent "ecotypes" were found exchanging genes at rates that decelerated as the populations differentiated (Shapiro et al. 2012). Gene-specific sweeps rather than genome wide sweeps frequently occur in natural communities (Bendall et al. 2016; Porter et al. 2017; Woods et al. 2020), and these observations could represent a scenario not unlike an interruption to gene flow (or genetic recombination) that results in the segregation of genetic pools and thus is potentially driving speciation.

This ability for prokaryotes to exchange genes with one another has inspired species concepts similar to the Biological Species Concept (BSC). Bacterial strains involved could be clustered into ecologically and genetically meaningful populations by inferring recent gene flow events (Shapiro and Polz 2014; Arevalo et al. 2019; Diop et al. 2022) Evidence of reductions in gene flow comparable to eukaryotic speciation patterns might provide a framework for species delineations (Fraser et al. 2009; Shapiro and Polz 2015; Bobay and Ochman 2017a), however defining these barriers remains a challenge since genetic loci within a bacterial population reach "fixation" at different points in time (Retchless and Lawrence 2007). Further, individuals belonging to the same species group might contain genes not found in the genomes of their siblings. The term "pangenome" (Tettelin et al. 2005) encompasses the entire genetic repertoire/diversity within a given set of genomes belonging to the same clade. Depending on their promiscuousness, the number of genomes available, and the parameters used, a pangenome's size can vary drastically. For example, a study has shown that the core genome of just 20 strains of *E. coli* accounts for only ten percent of their combined genomic repertoire (Touchon et al. 2009). A pangenome can be considered closed (finite) or open (possibly infinite)

(Tettelin et al. 2008) – closed pangenomes might therefore be clear representatives of a purely clonal "species"; however, for most bacteria it is necessary to identify overlaps in genetic repertoires between strains and classify the pangenome into two distinct regions – the core genome and the accessory (or flexible) genome.

1.1.4 Gene flow within the core genome

A set of "conspecific" bacterial strains will have a shared set of conserved genes referred to as the "core genome". These genes are generally necessary for basic survival, encoding proteins used for such processes as cell maintenance, cell division, and metabolism (Rousset et al. 2021). Core genomes are constructed computationally, using robust software that classifies homologous genes by performing comparisons of whole genome sequences—e.g., *CoreCruncher* (Harris et al. 2021)—and are commonly used for inferring species cohesion and phylogeny, gene flow, and evidence of selection (Rokas et al. 2003; Bobay and Ochman 2018). The genes that make up the core genome are inherited either vertically from ancestral bacterial cells, or they can be impacted by a process called homologous recombination (HR) which takes place typically between closely related strains (Thomas and Nielsen 2005). HR can occur when a bacterial chromosome experiences double-stranded breakage, invoking highly conserved DNA repair enzymes that are able to integrate foreign homologous segments (Figure 1.1a) into the repaired region (Rocha et al. 2005; Buton and Bobay 2021). The result is gene flow – the replacement of gene sequences within the core genome and can lead to allelic exchange – but which ultimately homogenizes genomic diversity across strains and helps maintain a cohesive bacterial population. Interruptions to and cessation of gene flow between bacteria can therefore provide a BSC-like framework for the classification of populations into species. Rates of genetic recombination can vary drastically for different bacterial populations, with their effect on a population highly influenced concurrently with the frequency of selection, genomic defense mechanisms (restriction-modification systems, CRISPR-Cas-9), and ecological lifestyles (Shapiro and Polz 2015; Shapiro 2016; González-Torres et al. 2019). Early periods of speciation could be triggered by the selection for genes in the core genome required for adaptation to a new ecological niche. On the other hand, selection leading to speciation could also occur outside the core genome by the gain of novel and advantageous accessory genes.

1.1.5 Horizontal gene transfer of accessory genes

The accessory (or flexible) genome consists of genes that are not part of the core genome, and whose presence is not usually crucial for basic cell survival and functionality. It is often characterized by its drastic variability which contributes to functional plasticity, either by high mutation rates (when compared to core genes) or the acquisition of new accessory genes. The lateral transfer of genes between the accessory genomes of bacteria is referred to as horizontal gene transfer (HGT), and these events can introduce new genetic functionality to a bacteria or population without replacing an existing homologous sequence (Figure 1.1b) (Ochman et al. 2000). If a bacterium receives a deleterious (or neutral) gene, it will most likely be purged from the population much like an individual with a deleterious mutation would be (Nguyen et al. 2022). However, the acquisition of novel and advantageous genetic material increases the phenotypic heterogeneity of a population which can aid in ecological adaptations such as virulence, antibiotic resistance, or the ability to fix nitrogen (Hacker and Carniel 2001; Raymond et al. 2004). A population experiencing frequent HGT events cold not exist within the SEM which is reliant on genome-wide sweeps to maintain genetic homogeneity.

Instances of HGT can be inferred computationally by several different methods: phylogenetic analyses of specific genes, comparative genomics that search for sequence anomalies, and locating nearly identical accessory gene orthologues between bacterial strains (Ravenhall et al. 2015). The mobility of genetic information between genes in the accessory genomes of bacteria, and the fact that HGT can easily occur without the need for a homologous sequence means that even distantly related organisms can exchange genes with one another.

While reductions to gene flow within core genomes can eventually lead to speciation, processes of HGT of accessory genes can still occur and perhaps might even increase. It is therefore imperative to view bacterial speciation through the lens of the evolution of the entire pangenome of a population – the genetic shifts occurring in the core and accessory genomes must therefore be distinguished since HR can lead to the genetic cohesiveness of the core genome whereas HGT can lead to the genetic differentiation of the accessory genome. Both mechanisms have a profound effect on the evolutionary trajectory of microbial populations, and these events have obfuscated our ability to clearly define species borders using the previous methods described above.

1.1.6 Allopatry and sympatry

From a biogeographical perspective, barriers to gene flow can be described as allopatric or sympatric. Allopatric speciation refers to population's genetic divergence due to geographical or physical barriers (Figure 1.2a). While widely recognized for sexual organisms, whether these barriers exist for bacteria remains in question. On account of their microscopic size, ease of dispersal, and shear abundance, some argue that geographic borders do not exist at all for prokaryotes (Finlay 2002). Yet others assert that geographic isolation alone can be the initial cause for divergence within a population of microbes who seem to form genetic clusters based on sample location (Whitaker et al. 2003). Sympatric speciation, on the other hand (Figure 1.2b), refers to the divergence of a population that is not caused by geographical isolation, but by the disruptive selection of traits that may confer ecological specialization, altering mating patterns. In the case of bacteria, theoretical analyses indicate that sympatric speciation tends to be initiated by directional selection for niche adaptations at a few specific alleles (Friedman et al. 2013).

Previously defined allopatric and sympatric speciation concepts for eukaryotes are somewhat useful in this context; however, bacteria exhibit conflicting abilities to follow these rules. For example, one recent study revealed that microbial populations living in geographically isolated lakes thousands of kilometers apart still showed evidence of gene flow, preventing them from diverging (Hoetzinger et al. 2021). This could be due to their populations' similarity in ecology as well as the ability to be relocated by unknown vectors such as birds or winds which would allow them to genetically recombine with one another. Alternatively, coexisting strains of bacterioplankton have seemingly ecologically diverged, showing separate preferences for microhabitats despite their shared environment and nearly identical genomes (Hunt et al. 2008). Many of these studies are performed on free-living bacteria, but these dynamics are likely easier to study in host-symbiont models, where microbial habitats are isolated yet contain many microniches. When a host population has undergone allopatric speciation, coevolution of the symbiotic microbial community in tandem with its host has been observed (Moran 2001; Youngblut et al. 2019). In such cases, host divergence could be considered geographical isolation, creating an allopatric barrier to gene flow for microbiota. On the other hand, a shift in ecology, could provide a new micro-niche for its symbionts, instigating sympatric speciation.

Part 2: Host-associated microbial populations

It remains unclear how new bacterial species emerge and yet others maintain their genomic similarities amidst such pervasive exchanges of genetic material. A major limitation in answering these questions studying free-living bacteria is our inability to unambiguously identify where each bacterial population is living. Host-associated microbes offer an ideal system to study bacterial speciation. They maintain an intimate relationship with their hosts, resulting in a coevolved symbiotic relationship between host and microbiome (Koch et al. 2013; Bennett and Moran 2015; Moeller et al. 2016; Groussin et al. 2020). Although these populations have established host specificity via coevolution with their hosts over long time periods, they have also been shown to evolve within their hosts over time periods as short as a few days (Crook et al. 2019; Garud et al. 2019; Groussin et al. 2021), as a result of mutations and gene transfers. The interactions and adaptation of bacterial communities within the mammalian gut are of growing interest but typically consist of hundreds or thousands of species. Therefore it is more ideal to utilize a simpler host-gut system to decipher how such pervasive genetic exchange facilitates speciation and maintains species borders, and how they help facilitate the health of their host (Bobay and Raymann 2019; Douglas 2019).

1.2.1 The corbiculate bee microbiome

Corbiculate bees (Hymenoptera: Apidae) are well-studied and offer a simpler and conserved system for examining bacterial genetic processes. They play a crucial role as pollinators in nature and agriculture practices and consist of four main tribes – *Apini* (honey bees), *Bombini* (bumble bees), *Meliponini* (stingless bees), and *Euglossini* (orchid bees). ‘Corbiculate’ is derived from the Latin ‘corbicula’, and refers to the concave ‘pollen basket’ located on the hind legs of bees that is capable of carrying up to 35% of their body weight (Capinera 2008; Anon). Corbiculate bees are characterized by their highly evolved social behavior (reproductive division of labor, cooperative brood care, and overlap of generations) that is estimated to have evolved over 87 Mya (Cardinal and Danforth 2011).

Particularly, *Apis mellifera* (the western honey bee) and *Bombus spp.* (the bumble bee) have been instrumental in various types of research due to their many intriguing traits such as advanced social behavior (Danforth et al. 2003), importance as agricultural pollinators (Klein et al. 2007), and transitional lifestyle plasticity (Elekonich and Roberts 2005). In recent years, corbiculate bees have also proven to be an ideal model for studying the gut microbiome (Wang et al. 2018). Their microbiota exhibit: (i) highly conserved biological pathways (Ihle et al. 2014), (ii) gene homology with humans and other insects (Foret et al. 2012), (iii) pathogen defense mechanisms (Cariveau et al. 2014), and (iv) an influence on their complex behavioral capabilities (Lewis and Lizé 2015). The closed, hive-dwelling and social lifestyles of these bee species have resulted in a highly conserved microbiome for adult workers with some taxa that have not been found anywhere else on Earth (Engel and Moran 2013).

There are five core taxa (*Snodgrassella*, *Gilliamella*, *Bifidobacterium*, *Lactobacillus*, and *Bombilactobacillus*) that are prevalent in nearly every individual of two of these tribes: *Apini* and *Bombini* (Koch and Schmid-Hempel 2011; Moran et al. 2012; Kwong and Moran 2016). The ability of *Apini* and *Bombini* bees to maintain a consistent yet distinctive microbiome is due to the vertical transmission of microbiota from adult worker bees to newly emerged workers (NEWs). Some of these core species have been lost in the *Euglossini* and *Meliponini* tribes, likely due to their less advanced eusocial behavior, causing their microbiomes to be acquired by NEWs both vertically from hive mates and horizontally from the environment (Cerqueira et al.

2021; Kueneman et al. 2023), whereas, most other insects (including solitary bees) obtain their microbiome mainly from environmental interactions, and their gut communities show high levels of variation between individuals (Hammer et al. 2017; Kapheim et al. 2021).

The digestive tract within adult bee abdomens is made up of the crop, midgut, and hindgut (ileum and rectum), however the vast majority of bacteria reside in the hindgut (Martinson et al. 2012) (Figure 1.3). In the *Apini* and *Bombini* tribes, the gut ileum is dominated by the gram-negative microbial taxa *Snodgrassella* and *Gilliamella* (Koch and Schmid-Hempel 2011; Martinson et al. 2012; Moran 2015). *Snodgrassella* forms a biofilm along the epithelial lining of the ileum, protecting it from pathogen colonization and plays an important role in host immune system functionality both by deploying type VI defensive secretions against pathogens, and by inducing antimicrobial gene expression in the host (Steele et al. 2017; Horak et al. 2020). *Gilliamella* forms a biofilm and colonizes on top of *Snodgrassella* and functions primarily as a sugar metabolizer, but also has been found to possess similar antitoxin genes to those found in *Snodgrassella*, indicating horizontal transfer between the two genera (Kwong et al. 2014). Gram-positive taxa *Lactobacillus*, *Bombilactobacillus* (formerly called *Lactobacillus* Firm-4) and *Bifidobacterium* occupy the rectum, and aid in host health by breaking down pollens and providing essential nutrients (Engel et al. 2012; Brochet et al. 2021).

Apart from the core species found in adult workers, two other taxa of bacteria have been found to play crucial roles in bee health and development –*Bombella apis* and *Apilactobacillus kunkeei*. The queen harbors a distinctive gut community from her progeny (Zumkhawala-Cook et al. 2024); the acetic acid bacterium *B. apis* (previously named *Parasaccharibacter apium*) is the dominant member of the microbiome of queen gut as well as the final larval stage of a worker (Smith et al. 2020). *B. apis* provides antifungal properties (Miller et al. 2021), synthesis of essential amino acids for its host (Parish et al. 2022), and defense against other host pathogens (Corby-Harris et al. 2016). The lactic acid bacterium *A. kunkeei* is found both in the environment and within the bee gut and protects honeybee health by providing probiotic (Olofsson et al. 2016) and anti-pathogenic support within the gut (Berríos et al. 2018). These two taxa have been less studied in *Bombus* hosts, but likely provide the same benefits.

1.2.2 *Apini* lifestyle: Impacts on microbiome

A honeybee colony is initially founded by a swarm of between 3,700 and 21,800 individual workers (Lee and Winston 1985) when a queen departs an overpopulated hive with a large proportion of its workers – therefore the transmission of microbiota never experiences a large bottleneck. Western (*Apis mellifera*) and eastern (*Apis cerana*) honeybees are the only domesticated species out of the eight species that make up the *Apini* tribe. Beekeeping practices since ancient times have likely had a homogenizing effect on their microbiome compositions (Ellegaard et al. 2020). Their irreplaceable importance in agriculture has led to current industrial beekeeping practices that involve the shipment of hundreds of hives for thousands of miles (Today 2019). These practices can negatively affect bee health by exposing them to herbicides, pesticides, and fungicides which alter their gut microbial communities (Pettis et al. 2013; Colwell et al. 2017; Motta et al. 2018). Further, this industry maintains crowded apiaries and transportation vehicles which increase exposure to other bees and pathogens (Seeley and Smith 2015). The need for beekeepers' to use antibiotics as an attempt to combat these issues is of equal concern(Raymann et al. 2017; Aljedani 2022), and has led to the acquisition of antibiotic resistance genes.

1.2.3 *Bombini* lifestyle: Impacts on microbiome

A key difference between *Bombini* and *Apini* ecology is that an initial bumble bee colony is founded by one solitary queen in early Spring. She alone inoculates her first round of NEWs with their microbiome, thus creating a severe bottleneck for the microbiota. Subsequent NEWs obtain their gut microbiota from their sisters. Bumble bee hives are much smaller than honey bees – their colonies grow to about 50-500 individuals, depending on habitat and species. (Williams et al. 2014), and it is suspected that this "primitive eusocial" tribe obtain elements of their microbiome horizontally as well (Hammer et al. 2021). In the Fall, the queen will lay only drones (males) and new queens, both of whom will leave the hive to mate. Hive workers and drones will die off and the newly mated queens will dig underground and enter diapause, an intricately regulated dormant phase characterized by a drastic reduction in metabolic activity that greatly alters the composition of her microbiome (Hotchkiss et al. 2024). In lab-reared queens, a reduction in *Gilliamella* and *Snodgrassella* and an increase in other microbial taxa was observed (Bosmans et al. 2018), further intensifying the bottleneck for a hive's initial gut community.

There are at least 250 species of bumble bees worldwide, with only two species that have been raised commercially, and only since the last century (Lee-Mäder et al. 2010). As with honeybees, the effects of commercial rearing have had a negative effect on their gut health, causing in increase in the deadly parasites *Crithidia spp*. and *Nosema bombi* (Mockler et al. 2018; Fernandez De Landa et al. 2024). Exposure of commercially reared bees to feral bees is of increasing concern (Murray et al. 2013).

1.2.4 The corbiculate bee microbiome as a model to study bacterial speciation

As corbiculate bees have speciated from their common ancestor and from one another, so too have their microbiota, indicating an intimate and coevolved symbiotic relationship between host and microbiome (Koch et al. 2013), and provides an ideal opportunity to examine whether host divergence acts as an allopatric barrier to gene flow for their gut microbiota. Indeed, competition assays performed using strains of *Snodgrassella alvi* from bumblebees and honeybees have demonstrated that strains specialized for their host species are not able to colonize the other host species (Kwong and Moran 2015), indicating that host divergence and specialization in this case might have disrupted the ability for gene flow between allopatric species of symbionts.

Within a host species, the gut harbors distinct niches that could create phenotypic divergences, and sympatric speciation could occur when selection for an alternate substrate occurs (Zheng et al. 2019). One recent study has showed that separate species of Gilliamella occupy distal regions of the hindgut (Li et al. 2022), and could be representative of such a case. The bee microbiome is permeated with genetic diversity (Engel and Moran 2013; Kwong et al. 2014). Acquisition of new genes can be highly adaptive both for host and microbiota, gut microbes must quickly adapt to changing environments, and higher rates of HGT of genes of certain functionality could be representative of environmental pressures. Studies have found genes responsible for antibiotic resistance and the degradation of toxic sugar molecules in the genomes of honeybee gut microbiota (Zheng et al. 2016; Zheng et al. 2019; Sun et al. 2023).

Increasing focus on the impact bacterial communities have on plants, animals, and the environment has shed light on the need to better understand their modes of adaptation, evolution, and speciation. Of particular importance is properly defining bacterial species borders and disentangling the driving forces behind bacterial speciation. Detecting and identifying core genes that have recombined and accessory genes that have been transferred horizontally within and among the microbiota of honeybees and bumblebees allows for the investigation of allopatric and sympatric speciation in relation to patterns of HR and HGT for prokaryotes.

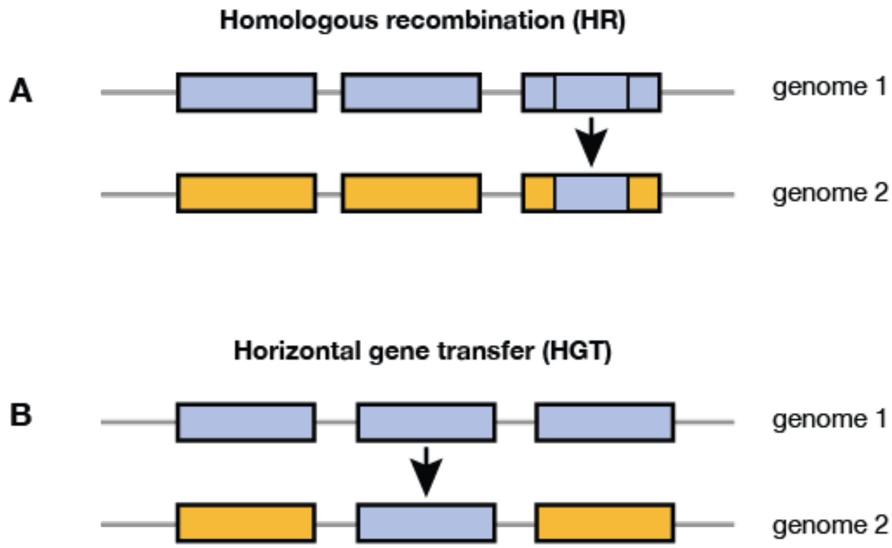


Figure 1.1. Type of gene transfers: (A) Homologous recombination (HR) is the allelic replacement of a gene by a homologous sequence between closely related bacterial species. This mechanism helps maintain genetic cohesion, allowing for gene flow within a population. (B) Horizontal gene transfer results in the recipient cell receiving a new gene either from a closely or distantly related species without the replacement of a homologous sequence.

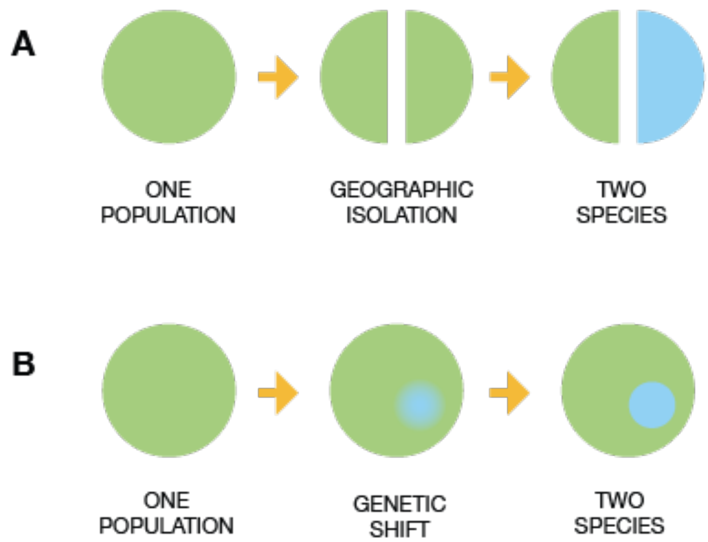


Figure 1.2. Allopatric speciation (A) results from a barrier to gene flow caused by the physical separation of populations. Sympatric speciation (B) is driven by ecological mechanisms of isolation where populations adapt to a specific niche which interrupts gene flow.

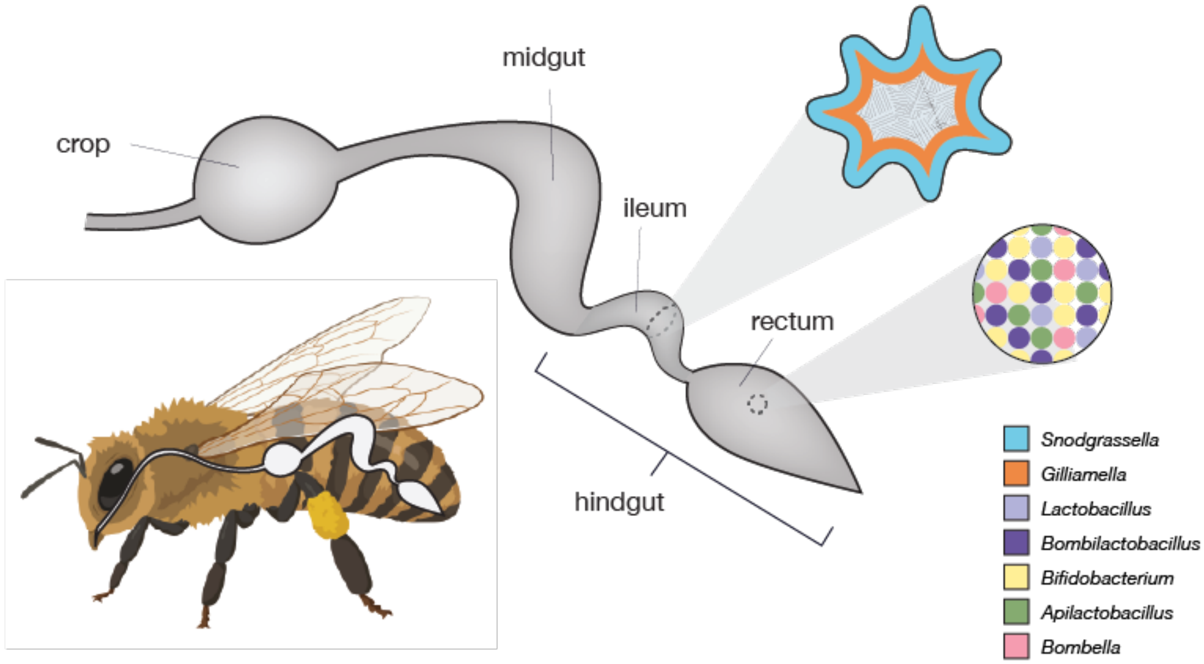


Figure 1.3. The digestive tract of corbiculate bees. The microbiome consists of 5–8 core taxa. The following 7 are analyzed in this study: *Snodgrassella*, *Gilliamella*, *Bifidobacterium*, *Lactobacillus*, and *Bombilactobacillus, Apilactobacillus,* and *Bombella*. There are few to no bacteria in the crop and midgut of the corbiculate bee. *Snodgrassella* forms a biofilm and lines the inner epithelial tissue of the ileum. *Gilliamella* forms a biofilm and colonizes on top of Snodgrassella. *Bifidobacterium*, *Lactobacillus*, and *Bombilactobacillus, Apilactobacillus,* and *Bombella* are found in the rectum.

Chapter 2: Co-evolution and Gene Transfers Drive Speciation Patterns in Host-Associated Bacteria

2.1 Introduction

Host-associated microbial communities perform key functions that impact the health and development of their hosts. Understanding the drivers of their evolution and speciation have important implications for human and environmental health, but these mechanisms remain elusive. Bacterial communities within animal hosts are often extremely complex and are driven by both host ecology and internal microbial interactions (Ochman et al. 2010; Smillie et al. 2011; Polz et al. 2013; Stein et al. 2013; Faust et al. 2018). Co-evolution between host and their associated microbes can lead to tight symbiotic relationships where microbes become highly specialized to the host as its niche (Ochman et al. 2010; Moeller et al. 2016; Groussin et al. 2020). Symbionts that are dependent on their host therefore represent a very informative system to study microbial speciation. Indeed, our understanding of speciation in microbes has been hindered by the difficulty of characterizing the ecological and geographic distribution of these organisms. As a result, most fundamental aspects of microbial speciation remain poorly understood. For instance, little is known about the relative impact of sympatric and allopatric speciation on microbial evolution. Allopatric speciation occurs when a barrier to gene flow disrupts the genetic cohesiveness of a species, whereas sympatric speciation is not the direct result of an interruption of gene flow between two populations. Although models based on reproductive isolation have been developed for sexual organisms, their core concepts are also relevant to most microbial organisms since gene exchange has been shown to play a key role in the evolution of the vast majority of bacterial species (Shapiro et al. 2012; Garud et al. 2019; Groussin et al. 2020).

The role that genetic transfers play in microbial speciation is a pressing question. Although species concepts have been debated for prokaryotes, a species is broadly defined as a population with a certain degree of genetic homogeneity that usually occupy a similar ecological niche. For sexually reproducing organisms, gene flow represents an important force that helps maintain the genetic cohesion of a species. Speciation is therefore associated with a reduction or cessation of gene flow, where divergent populations can no longer exchange substantial amounts of genetic information. Here we define genetic recombination (gene flow) among bacteria as homologous recombination within the core genome (i.e., the set of genes shared by all members of the population) of closely related individuals or strains. In bacteria, homologous recombination typically involves the replacement of short pieces of related DNA (i.e., <1kb) (Mau et al. 2006; Bobay et al. 2015); this process closely resembles gene conversion in animals and plants where repair mechanisms prompted by double-stranded DNA breakage facilitate the replacement of allelic variants of a related gene sequence. The exchange of allelic variants via gene flow requires sequence homology with high relatedness, which is more likely to be present in the core genomes of related strains rather than divergent strains. Additionally, bacteria are capable of acquiring novel accessory genes via recombination, transduction, or conjugation. Accessory genes are not conserved across all members of a species, and, in contrast to gene flow, their addition can more readily provide novel functions to the recipient bacterium. The difference between both processes is critical – gene flow requires the presence of homologous sequences shared between the donor and the recipient, whereas HGT can lead to the introduction of new accessory genes and does not necessarily require homology of sequences. This distinction is relevant in the context of speciation because gene flow almost exclusively occurs between highly genetically related bacteria and can help maintain a cohesive bacterial population or species

(Fraser et al. 2009; Friedman et al. 2013; Bobay and Ochman 2017b; González-Torres et al. 2019; Bobay 2020). Conversely, HGT can occur across more divergent populations and could effectively aid in plasticity within a population or in ecological adaptation that could trigger divergence between two populations. Examining the patterns of gene flow and HGT within and between host-associated communities and how they shape evolutionary processes can provide key insights regarding the mechanisms of speciation at play between bacterial populations.

The eusocial bee gut microbiome is a highly conserved host-associated microbial community comprised of few taxa, offering a unique opportunity to examine the drivers of allopatric and sympatric speciation within bacterial populations. Eusocial corbiculate bees, such as honey bees (*Apini* tribe) and bumble bees (*Bombini* tribe), have evolved a complex social structure within a closed living environment; one key result being the "vertical" transmission of the microbiota from the mother or siblings to newly emerged bees, thus isolating the gut community and facilitating co-diversification with its host over 85 million years (Grimaldi and Engel 2005; Cardinal and Danforth 2011). Since this time, the multiple speciation events of eusocial bees may have acted as allopatric barriers to gene flow for their gut symbionts. Indeed, studies have shown that some strains isolated from bumble bee hosts are unable to colonize the honey bee gut (Kwong et al. 2014; Kwong and Moran 2015). Alternatively, divergent selection for shifts in micro-ecological niches within the gut lumen could outweigh the constraints maintained by gene flow, thereby instigating sympatric speciation. Several studies have recently characterized the presence of ecologically and spatially distinct—but closely related—bacterial populations cohabiting within an individual bee, indicating that niche differentiation exists within the bee gut (Zhang et al. 2021; Li et al. 2022). Therefore, the relative simplicity and stability of the eusocial bee microbiome is an attractive system for studying the complex processes of genetic exchange and their effect on bacterial speciation.

Here we examined the population structures of seven highly specialized clades of gut bacteria unique to eusocial bees. We analyzed the patterns of gene flow (i.e., genetic recombination of homologous sequences) and HGT among populations within and between these genera. Interestingly, several genera presented patterns of speciation that were tightly associated with their host distribution, whereas others did not. Our results support a model where bacterial populations within at least two genera have undergone allopatric speciation associated with a sharp interruption of gene flow (i.e., *Snodgrassella* and *Gilliamella*). Moreover, we show that *Gilliamella's* evolution likely reflects additional events of sympatric speciation within the host. Interestingly, we found that both gene flow and HGT were interrupted between allopatric species; however, sympatric species displayed an interruption of gene flow, but not an interruption of HGT. Overall, these findings suggest that both allopatric and sympatric speciation occur in host-associated bacteria. Allopatric speciation seems to be the direct result of prolonged physical separation, which prohibits both gene flow and HGT. In contrast, sympatric speciation is likely the result of ecological specialization, which results in a strong reduction of gene flow, but not HGT.

2.2 Results and Discussion

2.2.1 Species Diversity Within Each Bacterial Clade

We analyzed the whole genomes of bacterial strains isolated from either *Apini* or *Bombini* hosts. These strains are currently classified across seven genera: *Bombella* (n=10), *Bombilactobacillus* (formerly *Lactobacillus* Firm-4, n=10), *Apilactobacillus* (formerly

Lactobacillus Firm-5, n=36), *Lactobacillus* (n=40), *Bifidobacterium* (n=40), *Snodgrassella* (n=57) and *Gilliamella* (n=104). All these taxa have been described as core members of the bee gut, except for *Bombella*, but this taxon is frequently found in the bee gut. In order to better classify species within each genus by examining gene flow, strains were first grouped into "species" by establishing the core genome for each genus, computing the average nucleotide identity (ANI) of orthologous core genes between whole genome pairs, and then clustering individuals into ANI-defined species based on the previously established identity score of $\geq$94% (Rosselló-Mora 2001; Konstantinidis and Tiedje 2005; Bobay 2020). While ANI scores reveal genomic relatedness, they may not necessarily be representative of exclusive groups of bacterial strains with shared common ancestry (i.e., monophyly). Therefore, phylogenetic trees of each taxon were constructed using concatenate of all core gene alignments (Fig. 2.1, Fig. 2.5) and revealed that all ANI-defined species in our analysis were also phylogenetic species based on their monophyly. For *Gilliamella*, *Lactobacillus* and *Snodgrassella*, phylogenetic species trees and the classification into ANI-defined species yielded a clear genetic divergence into either Eastern *Apis*, Western *Apis*, or *Bombus* hosts (Fig. 2.1a-c). Moreover, these results suggest that there are more species of both *Snodgrassella* and *Gilliamella* than currently defined. In contrast, for *Bifidobacterium*, *Apilactobacillus*, *Bombilactobacillus* and *Bombella*, phylogenetic species trees and the classification into ANI-defined species yielded patterns of mixed strains from either Eastern *Apis*, Western *Apis*, or Bombus hosts (Fig. 2.1d, Fig. 2.5a-c). These findings support low levels of specialization for their hosts where the processes of speciation are likely independent from their host. We further tested the robustness of the trees by re-building each tree using the consensus of core gene trees. Across most genera, all ANI-species were monophyletic in both the phylogenetic tree constructed from the consensus of individual gene trees and the tree constructed from the core genome concatenate (Fig. 2.1, Fig. 2.5) with the exception of two species (one in *Apilactobacillus* and one in *Bifidobacterium*). Overall, the topologies of the trees built with the two approaches were highly congruent and unambiguously confirmed the specialization of each ANI-species to their host in *Gilliamella*, *Lactobacillus* and *Snodgrassella*.

Snodgrassella strains in our dataset were classified into a total of five ANI-species by our analyses but there are currently only three formally characterized species – one *Apis*-derived species (*S. alvi*) (Kwong and Moran 2013), and two species recently described derived from *Bombus* hosts (*S. gandavensis* and *S. communis*) (Cornet et al. 2022). Supporting previous findings (Powell et al. 2016; Cornet et al. 2022; Li et al. 2022), we identified a single ANI-species belonging to the Western *Apis* host group, one ANI-species composed of three strains from three different host species (*A. cerana, A. florea,* and *A. andreniformis*) representing the Eastern *Apis* host group, and three *Bombus*-derived ANI-species (Fig. 2.1a).

Within the *Gilliamella* genus, we identified 17 ANI-species (Fig. 2.1b) although there are currently only six species formally described – two originating from Western *Apis* hosts (*G. apicola* (Kwong and Moran 2013) and *G. apis* (Ludvigsen et al. 2018), and four named species derived from *Bombus* hosts (*G. intestine, G. bombicola, G. bombi*, and *G. mensalis* (Praet et al. 2017)). While our results do not conflict with previous studies (Ludvigsen et al. 2018; Ellegaard and Engel 2019; Li et al. 2022), we found extensive diversity within *Gilliamella*. Although six of the ANI-defined *Gilliamella* species are composed of a single strain, our results indicate that there are a total of six ANI-species within the Eastern *Apis* host group, four species from (Western) *Apis mellifera*, and seven species associated to *Bombus* hosts. A previous study reported more ANI-species for *Snodgrassella* and *Gilliamella* than suggested from our results (Zhang et al. 2022); however, in their study, pairwise genomic comparisons appear to have included accessory genes which would result in lower nucleotide identity scores between strains and an increase in the number of clusters below the 94% cutoff mark.

We identified five ANI-species across the *Lactobacillus* strains (Fig. 2.1c) although there are currently six species from honeybee formally described – five *Apis*-derived species (*L. apis, L. helsingborgensis, L. kullabergensis, L. kimbladi* and, *L. melliventris*) and one derived from *Bombus* hosts (*L. bombicola*). Four of our ANI-species were isolated from Western *Apis* and one from *Bombus*.

Bifidobacterium strains in our dataset were classified into three ANI-species (Fig. 2.1d) while there are currently six formally described species – all derived from *Apis* hosts (*B. asteroides, B. coryneforme, B. indicum, B. apousia, B. polysaccharolyticum* and *B. choladohabitans*). In our analysis, two ANI-species were composed of strains from the Western *Apis* host group (Powell et al. 2016; Cornet et al. 2022; Li et al. 2022) and one ANI-species was composed of strains from different hosts: Western *Apis* (23 strains) and one strain from *A. cerana* representing the Eastern *Apis* host group.

In addition, strains within each of the genera *Apilactobacillus* and *Bombilactobacillus* were classified into two ANI-species, respectively, which is consistent with their current taxonomic description, and they are all associated with *Apis* hosts (Fig. 2.5a-b). One ANI-species within each *Apilactobacillus* and *Bombilactobacillus* is composed of a single strain and four strains, respectively, and have been isolated from Western *Apis*. The remaining ANI-species within both genera are composed of a mix of strains from Western *Apis* and Eastern *Apis*. Speciation patterns of *Apilactobacillus* and *Bombilactobacillus* appear therefore unrelated to their host distribution, as opposed to the patterns observed in *Snodgrassella* and *Gilliamella*, Within the genus *Bombella*, we inferred two ANI-species from our dataset with one species that has been formally characterized – derived from *Apis* and *Bombus* (*B. apis*) and one species isolated from *Apis* that has not formally been characterized (*B. sp*.) (Fig. 2.5c).

The criterion for defining a bacterial species currently lacks an overall consensus, complicating bacterial taxonomy. Using our above methods, we found that for the two clades residing in the bee's ileum (*Snodgrassella* and *Gilliamella*), there are more species than currently characterized. Our findings are largely in agreement with the recently reclassified *Lactobacillus* clade (Bradford et al. 2022) with the exception of *L. kimbladii* and *L. kullabergensis* which could be considered the same species. We found that multiple named species of *Bifidobacterium* could also be the same species, indicating a possible need for reclassification. Overall, these results yielded clusters of putative species (ANI-species) within the seven genera for further analyses in this study.

2.2.2 Patterns of Gene Flow Support a Model of Allopatric and Sympatric Speciation

As mentioned above, gene flow is thought to play an essential role in maintaining the genomic cohesion in sexual populations, and patterns of gene flow have been proposed as a new framework to classify bacteria into biological species (BSC-species) (Dykhuizen and Green 1991; Ochman et al. 2005; Shapiro and Polz 2015; Bobay and Ochman 2017c; Bobay 2020; Diop et al. 2022). Host-associated microbial systems provide a unique opportunity to examine how host divergence and within-host compartmentalization may affect the impact of gene flow between bacteria from related hosts. Co-evolution with differentiated hosts could lead to an interruption of gene flow that would initiate allopatric speciation. In contrast, adaptation to specific host-derived metabolites within the gut could lead to ecological specialization that could initiate sympatric speciation.

First, we tested for evidence of gene flow across the different ANI-defined species characterized above within each genus. Gene flow was measured by inferring the ratio of homoplasic alleles (*h*) to non-homoplasic alleles (*m*) within the core genes of each species group as previously described (Bobay and Ochman 2017b). Homoplasies are alleles whose phylogenies are inconsistent with vertical inheritance from a shared ancestor, and they are predominantly the result of homologous recombination over short evolutionary time scales (Bobay and Ochman 2017b).

Within the Western *Apis* host group, we found a single *Snodgrassella* ANI-species, whereas we inferred four *Gilliamella* ANI-species, four *Lactobacillus* ANI-species and three *Bifidobacterium* ANI-species residing within this one host species, providing the possibility to examine gene flow between ANI-species within the same host for those genera. Our results revealed two interesting dynamics between ANI-species within this host species (*Apis mellifera*). First, we found that two ANI-species of *Gilliamella* (ANI 17 and ANI 7) within the Western *Apis* host group were found to be engaging in gene flow based on the patterns of homoplasies (Fig. 2.1b). These two populations are under the typical threshold of 94% identity to classify species but are still closely related with an average ANI of 92.9% between the two populations. In addition, two ANI-species of *Bifidobacterium* (ANI 1 and ANI 2) within the Western *Apis* host group (ANI 2 includes one strain from *A. cerana* (Eastern *Apis*)) were defined into a single biological species based on the patterns of homoplasies (Fig. 2.1d). These two populations are slightly under the typical threshold of 94% identity to classify species but are still closely related with an ANI of 93.3%. For further confirmation, we compared the topology of individual gene trees against the topology of the entire species tree for each genus to identify individual core genes that have undergone homologous recombination. We found that the vast majority of the phylogenic trees of individual core genes were congruent with the core genome phylogenies for all genera, with the exception—again— of *Gilliamella* ANI 17 and ANI 7, and *Bifidobacterium* ANI 1 and ANI 2 which we found were engaging in high rates of gene flow (~65% and 82% of core genes presented clear evidence of recombination between the two ANI-species within *Gilliamella* and *Bifidobacterium*, respectively) (Fig. 2.1b, 2.1d). Therefore, the datasets generated using these two methods to infer BSC-species based on gene flow showed a strong agreement with species inference based on ANI and phylogeny with the exception of *Gilliamella* ANI 17 and ANI 7, and *Bifidobacterium* ANI 1 and ANI 2 which form a single BSC-species but two highly related ANI-species, respectively. In the previous study by Li et al. (2022), only a few regions of identical DNA were found to have experienced recent recombination events between strains W8131 and M1-2G. Our results are in agreement with this observation: our characterizations of gene flow suggest that the three strains that make up the ANI-7 group are also biological species with the strains belonging to *G. Apis* (ANI-17), despite nucleotide identity scores below the typical cutoff used to define ANI-species.

Secondly, our results further identified two divergent *Gilliamella* species within western honey bees (*G. apis*, ANI 17 and *G. apicola*, ANI 12) exhibiting almost no evidence of gene flow with one another (Fig. 2.1b). Several studies have noted the marked variation between *Gilliamella* strains for specialization in plant polysaccharide degradation, carbohydrate usage, and antibiotic resistance (Zheng et al. 2016; Zheng et al. 2019; Zhang et al. 2021). Interestingly, a recent study has demonstrated that strains from these two species (strain M6-3G (ANI 17) and strain wkB7 (ANI 12), respectively) are able to co-colonize antipodal regions of the ileum of western honey bees, likely due to ecological specialization for urea or pectin degradation (Li et al. 2022). Our results revealed that these two ecologically specialized groups of sympatric bacteria have ceased to exchange DNA through gene flow and may be considered different biological species and ecological species under this criterion. This scenario is compatible with a mechanism of sympatric speciation resulting from niche partitioning within the same host and supports previous findings that these are two distinct species (Ludvigsen et al. 2018; Ellegaard et al. 2020; Li et al. 2022).

2.2.3 Asymmetrical Patterns of HGT and Gene Flow

Horizontal transfers of accessory genes often confer new phenotypes and likely play a central role in ecological divergence and speciation, yet their extensive genetic variation and exchange makes detecting these events challenging. We applied a method to infer transfers of accessory genes (HGT) between strains based on the sequence similarity of accessory genes relative to the overall similarity of the core genome (see Methods). Due to the action of purifying selection, core genes are expected to evolve more slowly than accessory genes; therefore, the presence of highly similar accessory genes within strains with significantly more divergent core genomes strongly suggests that these genes have been acquired horizontally. HGT events were detected in all genera except *Apilactobacillus*, where all strains were too closely related to infer HGT events with our method. For all other clades, HGT was higher between species living in the same host group and lower between species living in different host groups (Fig. 2.2a): *Snodgrassella* and *Gilliamella* (Wilcoxon test, $P<10^{-15}$), *Lactobacillus* (Wilcoxon test, $P< 10^{-14}$), *Bifidobacterium* (Wilcoxon test, $P<10^{-4}$), *Bombilactobacillus* (Wilcoxon test, P=0.44) and *Bombella* (P=0.07), although it was not significant in the last two clades, which may be due to their small sampling data (10 strains each). These results suggest that allopatry resulting from inhabiting different hosts constitutes an efficient barrier to HGT. Within the genus

Bifidobacterium, strains belonging to the same ANI-species presented higher numbers of HGT events (Wilcoxon test, $P<10^{-15}$). In contrast, within the genus *Snodgrassella*, strains belonging to the same ANI-species presented significantly lower amounts of HGTs (Wilcoxon test, $P<10^{-15}$). Furthermore, no significant difference was observed in the amounts of HGTs between strains from the same or from different ANI-species within the genera *Gilliamella* and *Lactobacillus* (Wilcoxon tests, P=0.47 and P=0.07, respectively). This observation is consistent with the fact that HGT does not primarily rely on sequence identity for DNA exchange (Fig. 2.2b). Within the *Apis mellifera Gilliamella* clade, we found elevated transfers of accessory genes between *Gilliamella* ANI-species 17 and 12 and ANI-species 13 and 12, although they show no evidence of gene flow in their core genomes (Fig. 2.2b). In addition, *Lactobacillus* ANI-species 1 and 4 and ANI-species 1 and 5 presented elevated amounts of HGT but moderate amounts of gene flow, respectively, between Western *Apis* derived ANI-species (Fig. 2.1c). However, the ANI-species of *Snodgrasella* that displayed higher amounts of gene flow also had higher HGT rates between *Bombus* derived ANI-species (Fig. 2.1a). Due to the limited number of available genomes, we could not analyze the patterns of HGT in *Bombilactobacillus* and *Bombella* of the same host group (Fig. 2.2b). Moreover, HGTs were not inferred within the genus *Apilactobacillus* due to their high sequence identity between pairs of genomes (~100% core gene sequence identity), which did not allow us to distinguish horizontally acquired genes from vertically inherited ones. Although our results may seem different between *Snodgrassella, Gilliamella,* and *Lactobacillus* they are not contradictory. Indeed, just as we found almost no gene flow between host groups, we also found little evidence of HGT between host groups for all the clades, which supports previous findings (Kwong et al. 2014). In addition, *Snodgrassella*, *Lactobacillus*, *Bifidobacterium*, *Bombilactobacillus* and *Bombella* do not present multiple ANI-species inhabiting the same host species, respectively and it is therefore impossible to conclude whether distinct species within those clades living in the same host would be able to engage in HGT as we observed for *Gilliamella*. In summary, our results are consistent with a scenario where host-driven allopatry imposes an efficient barrier to both gene flow and HGT. In contrast, species living in sympatry exhibit a clear interruption of gene flow but no clear interruption of HGT, indicating that both mechanisms likely have different impacts on bacterial evolution.

2.2.4 Species Borders within Bombus are Fuzzy

We established above that, based on the patterns of gene flow, BSC-species and ANI-species are largely congruent except for *Gilliamella* ANI 17 and ANI 7 and *Bifidobacterium* ANI 1 and ANI 2 that engage in gene flow despite being part of different ANI-species (Fig. 2.1b, 2.1d). The overall interruption of gene flow does not preclude occasional exchange of DNA in the core genome through recombination akin to the introgression process in sexual organisms and which was observed between some prokaryotic species (Eppley et al. 2007; Richards et al. 2019; Diop et al. 2022). Using a phylogenetic approach, we compared individual gene trees to the species phylogeny tree of *Snodgrassella*, *Gilliamella*, *Lactobacillus*, *Bifidobacterium*, *Apilactobacillus Bombilactobacillus* and *Bombella*, respectively (Fig. 1.1, 1.5). We found that the phylogenies of individual core genes were almost always congruent with the species tree phylogeny of bacterial species living in *Apis* hosts. The only notable exception, as reported above, was for *Gilliamella* ANI 17 and ANI 7 and *Bifidobacterium* ANI 1 and ANI 2, which we inferred as the same BSC-species, respectively. This result indicates that introgression is rare between host-associated bacterial species, including between those living within the same host species. However, we observed that the patterns of introgression were more complex between bacteria residing within *Bombus* hosts whose divergence is much more recent (Santos Júnior et al. 2022). In our study, the *Bombus* host group is comprised of ten distinct species of bumble bees representing six subgenera. Our results yielded five ANI-species of *Gilliamella*, three ANI-species of *Snodgrassella* and one ANI-species of *Lactobacillus* within the *Bombus* host group, with each ANI-species consisting of strains isolated from more than one bumble bee species. In the *Snodgrassella* clade, we found high amounts of introgression between *Bombus*-derived ANI-species – in ANI 5, 40% of core genes appear to have recombined with several other ANI-species, and most of those (61%) were introgressed from ANI 3 (Fig. 2.1a). Similarly, in *Gilliamella*, ANI 6 and ANI 2 exhibit ~50% of introgressed core genes with each other despite their divergence based on phylogeny and ANI (Fig. 2.1b). In the previous study by Li et al. (2022), regions of identical DNA were found as evidence of recent gene flow between strains W8131 (our ANI-7) and M1-2G (our ANI-17) even though their gANI score was <95% (93.1%). In addition, their study found recent gene flow between strain Choc5-1 (our ANI-2) with both GilliExp13 (our ANI-6) and Gris1-4 (our ANI-6), although their gANI scores were as low as 86%. Our results are in agreement with these observations: our characterizations of gene flow suggest that ANI-7 species group are also biological species with the strains belonging to *G. Apis* (ANI-17) based on extensive historical recombination of core genes, and, between ANI-2 and ANI-6, we found ~50% of core genes introgressed despite nucleotide identity scores below the cutoff used to define ANI-species.

2.2.5 Functional Analysis of Transferred Genes

In order to compare the functional distributions of genes transferred between species, we reconstructed the pangenome for each genus and grouped orthologous genes into gene families.

While *Gilliamella's* core genome is smaller than *Snodgrassella*'s and *Bombella's* (1,130 vs 1,403 and 1,130 vs 1,213 genes, respectively), the pangenome for *Gilliamella* was double and three times the size of *Snodgrassella's* and *Bombella's* (10,024 vs 4,860 and 10,024 vs 2783 gene families, respectively), which reflects the fact that more genomes were available for *Gilliamella* (104 vs 57 and 104 vs 10 genomes, respectively). Similarly, *Apilactobacillus* presented a small pangenome compared to *Bombilactobacillus* (2718 vs 3484 gene families) with 36 versus 10 genomes available, respectively. Gene families were annotated and categorized by COG (Categories of Orthologous Groups) designation with EggNog (Fig. 2.3). A large proportion of the pangenomes of *Snodgrassella*, *Gilliamella*, *Lactobacillus*, *Bifidobacterium*, *Apilactobacillus* *Bombilactobacillus* and *Bombella* were uncharacterized. Proportions of gene function categories for cellular processes and signaling, information storage and processing, and metabolism were similar for all genera, however *Gilliamella*, *Lactobacillus*, *Bifidobacterium* and *Bombilactobacillus* (*Lactobacillus*-firm4) displayed—as previously reported—a marked increase in carbohydrate transport and metabolism genes; the majority of which are housed in the accessory genome and likely play a key role in the ability for populations to adapt to shifts in host diet (Engel et al. 2012; Kwong et al. 2014; Zheng et al. 2016; Zhang et al. 2022).

Within *Apis spp*. hosts, our analysis revealed that less than 1% of core genes were introgressed except for the two sympatric BSC-species of *Gilliamella* (ANI 13 and ANI 17/7) where at least 16% of core genes were inferred as introgressed and the sympatric BSC-species of *Lactobacillus* (ANI 1, ANI 3, ANI 4 and ANI 5) where introgression events ranged from 9 to 23% of the core genome. BSC-species within the *Bombus* host group exhibited many introgression events for *Snodgrassella* and *Gilliamella*, ranging from 9 – 40% and 4 – 50%, respectively. Overall, no significant difference was observed in the composition of the functions of the introgressed genes relative to the entire core genome suggesting than introgression is not particularly biased towards specific gene functions. Many genes categorized as “translation and ribosomal structure and biogenesis”, “amino acid transport and metabolism” and “defensive mechanisms” were found to be more likely introgressed than expected for *Gilliamella*, but these trends are not statistically significant (Chi-squared test, P=0.99, Fig. 2.3-2.4). In addition, in *Lactobacillus*, a slight trend was observed where introgressed genes were more frequently involved in “amino acid transport and metabolism” and “Cell cycle control, cell division, chromosome partitioning” (non-significant, Chi-Squared test with Bonferroni correction, P=0.31, Fig. 2.3-2.4). We found ~900 introgressed core genes between *Snodgrassella* ANI-species primarily from the *Bombus* host group. Genes involved in "inorganic ion transport and metabolism", "energy production and conversion", and "translation and ribosomal structure and biogenesis" were slightly more likely to be introgressed relative to their distribution within the core genome (non-significant, Chi-Square test with Bonferroni correction, P=0.83, Fig. 2.3-2.4). Most of these introgression events occurred between Bombus-derived groups with ANI 2 as the donor and ANI 5 as the recipient. The lack of significant enrichment of functional categories in introgressed genes suggest that introgression of core genes is rarely adaptive.

In contrast to introgressed genes, the functions of accessory genes acquired by HGT were strongly functionally dependent compared to the functional composition of the whole accessory genome. This suggests that HGT events are much more function-dependent than introgression events (Fig. 2.4), which may reflect stronger selective pressures on HGT events and the fact some accessory genes belong to mobile elements.

Within the accessory genome of *Gilliamella*, we found that genes involved in transcriptional regulation were overrepresented among horizontally transferred genes (Fig. 2.4b, Chi-squared test, $P<10^{-12}$), and those genes may confer plasticity and adaptability to their recipients. We also found an increase in the transfer of genes involved in recombination and repair, transposases, integrases, helicases, CRISPRs, and phage genes, especially between *Gilliamella* strains. In addition, we identified five genes involved in antibiotic resistance that have been transferred between *Gilliamella* strains within Western *Apis* hosts mostly between ANI 12 and ANI 17, or between *Apis* and *Bombus* host groups. Two of those cases of transferred genes are involved in resistance to Tetracycline (Tet genes), the antibiotic most frequently used by beekeepers (Aljedani 2022). Genes of unknown function were the most frequently transferred by HGT in *Snodgrassella*, as well as replication, recombination and repair (Fig. 2.4f, $P<10^{-5}$). Moreover, genes of unknown function, defense mechanisms, replication, recombination, and repair were more frequently transferred within the accessory genome of *Lactobacillus* (Fig. 2.4d, P=0.002). Interestingly, in *Bifidobacterium* most of the genes transferred via HGT are involved in defense mechanisms (Fig. 2.4h, $P<10^{-15}$) including genes that encode type IV toxin/antitoxin systems, ABC and efflux transporters, and Type I restriction modification systems (Ochman et al. 2000; Vlková and Silander 2022). Overall, these results support the view that patterns of HGT are significantly biased towards certain functions whereas introgression events are not. This pattern likely reflects the fact that HGT events are most likely adaptive or correspond to mobile elements being exchanged.

2.2.6 HGT Between Genera

The genera studied here are mostly predominant in the hindgut of *Apis* and *Bombus* hosts. *Snodgrassella* and *Gilliamella* both colonize the ileum, and are in direct contact with one another, and complement each other's distinctive metabolic capabilities (Kwong et al. 2014).

Snodgrassella colonizes the inner epithelium of the ileum and *Gilliamella* forms a biofilm and colonizes directly atop *Snodgrassella* (Martinson et al. 2012; Moran 2015). However, other bacteria such as *Lactobacillus*, *Bifidobacterium Apilactobacillus*, *Bombilactobacillus* colonize diverse hindgut compartments like lumen, ileum and the rectum with direct or indirect contact with *Gilliamella* and *Snodgrassella* depending on the niche specialization. Bacteria living in close proximity to one another within a host for long periods of time are likely to have engaged in HGT even if they are distant species (Ochman et al. 2000). We therefore used the previous method for inferring accessory gene transfers to identify HGT events between each pair of these genera. Although all genera share different levels of overall average core gene amino acid identity (AAI) ranging from 38% (between *Bombella* and *Bombilactobacillus*) to 50.2% (between *Lactobacillus* and *Bombilactobacillus*), our analysis revealed multiple pairs of accessory genes sharing much higher identity. To infer HGT events between genera, we defined a threshold of nucleotide identity which corresponds to the top 1 percentile of the distribution of core gene identity for each pair of genera (the most stringent threshold, see Methods). Our results revealed many (>11,000) pairs of accessory genes with identity scores above the defined threshold in *Snodgrassella* and *Gilliamella*, and >3,000 accessory gene pairs with a nucleotide sequence identity over 95%. Among all pairs of genera, besides the elevated number of genes with unknown function transferred, the genes involved in carbohydrate transport and metabolism, replication, recombination and repair and defense mechanism were frequently transferred between genera. A notable number of genes encoded for functions involved in metabolism, defense mechanisms, and replication, recombination and repair and transcription were transferred between *Snodgrassella* and *Gilliamella*. In contrast, most genes transferred by HGT between *Bifidobacterium* and *Bombella* were involved in cell wall/membrane/envelope biogenesis and in translation, ribosomal structure and biogenesis. In addition, among the transferred genes between *Snodgrassella* and *Gilliamella*, three genes were involved in antibiotic resistance including two genes conferring resistance to tetracycline (Tet genes). Similarly, one tetracycline resistance gene was transferred between *Bifidobacterium* and *Bombella*. The relatively recent transfer of these resistance genes in the accessory genome of those species is likely due to the prolonged and pervasive practice of using tetracycline prophylactically in the beekeeping industry. In *Snodgrassella* and *Apilactobacillus*, most of the HGT genes were found involved in carbohydrate transport and metabolism and replication, recombination, and repair. However, in *Snodgrassella* and *Bombella*, genes involved in information storage and processing and cellular processing and signaling were the most commonly transferred. Accessory genes sharing lower identity scores—but above the top percentile of the distribution of identity scores of the core genome—could represent older gene transfers while orthologous pairs with higher identity scores likely correspond to more recently transferred genes. These results indicate that, as previously suggested (Kwong et al. 2014), many HGT events have occurred between the genera and particularly between *Snodgrassella* and *Gilliamella* throughout the long period of time during which these two genera have been closely interacting within the bee ileum.

2.3 Conclusion

Defining bacterial species using established frameworks developed for eukaryotic biological species creates challenges, as speciation is a result of physical barriers and ecological adaptations that are more difficult to pinpoint on a microscopic scale. Here we showed that, for bacteria that are highly host-specific, host speciation can act as an efficient allopatric barrier to gene flow, while adaptations to ecological niches that confer a fitness benefit to host and symbiont might be the drivers of sympatric speciation events. We cannot rule out that more complex scenarios are at play. For instance, allopatric speciation followed by the reintroduction of a new species into a different host would strongly resemble sympatric speciation. However, bee gut bacteria appear host-specific: it has been shown that *Snodgrassella* strains from the honeybee are not capable of effectively colonizing bumble bee hosts (Kwong and Moran 2015). Ecological specialization can initiate sympatric speciation where two populations become specialized to a specific niche (i.e., source of metabolite). As the populations become increasingly specialized, their overall genome progressively becomes fine-tuned towards their respective niche. Although this mechanism may impact both gene flow and HGT, HGT is likely not as deeply impacted due to the large number of accessory genes that encode mobile elements and other elements that are not niche specific (i,e., warfare systems, restriction enzymes) (Bobay et al. 2013; Oliveira et al. 2017). However, it remains unclear whether host speciation is the direct driver of bacterial speciation through allopatry. Indeed, episodes of sympatric speciation followed by the loss of one or multiple species may resemble allopatric speciation. Nevertheless, our results clearly revealed that allopatric barriers established by host specialization appear to strongly limit DNA transfers. The high number of introgression events observed between bacteria living in different *Bombus* hosts strongly suggest that these species have only recently lost their ability to freely recombine with one another and this may be the direct result of ongoing allopatric separation through host speciation.

Species delineation inferred via phylogenetic, ANI, and BSC methods were highly congruent and support the idea that gene flow represents an efficient theoretical basis for defining bacterial species capable of exchanging DNA. Our findings are in general agreement with the study conducted by Li et al. (2022), highlighting gene flow as a crucial factor in bacterial species delineation. In some cases, we found extensive evidence of prolonged gene flow within the core genomes of populations, while their study focused on very recent events of gene flow across entire genomes (core and accessory) of the same genomes. Importantly, we observed that patterns of homologous recombination in the core genome (gene flow) differ from the patterns of HGT of accessory genes. Given that proximity to one another is required for gene exchange, the question remains as to why populations engaging in gene flow would not also promiscuously exchange accessory genes. Likewise, HGT was found to be prevalent between sympatric species which no longer engage in gene flow. These results support the view that, while homologous recombination maintains the genetic cohesiveness of a population, HGT events of accessory genes play an important role in niche adaptation and specialization. In agreement with this hypothesis, we observed that gene flow was not particularly biased across gene functions, whereas HGT events were strongly biased towards some gene functions. This further emphasizes the need of differentiating these two types of DNA transfers as they likely have opposite impacts on bacterial genomes.

2.4 Methods and Materials

2.4.1 Genomic Datasets and Core Genome Phylogeny

Previously published whole genome sequences for all strains analyzed were downloaded from the NCBI GenBank database (ftp://ftp.ncbi.nlm.nih.gov/genomes/ in October 2019), resulting in 104, 57, 10, 10, 36, 40 and 40 genomes for *Gilliamella, Snodgrassella, Bombella, Bombilactobacillus, Apilactobacillus, Lactobacillus* and *Bifidobacterium*, respectively. Protein sequences were extracted from the annotation files. Core genome protein families for each genus were built with CoreCruncher using default parameters (which exclude paralog/xenologs) (Harris et al. 2021) and orthologs were identified with Usearch (Edgar 2010). Protein sequences were aligned using MUSCLE v3.8.31 (Edgar 2004) and the alignments were backtranslated into their original nucleotide sequences using a custom script. Orthologs were classified as core genes if they exhibited ≥70% identity, a minimum of 80% alignment length and were present in at least 90% of the genomes of each genus. Two independent methods were used to generate the phylogenetic tree of each genus. (i) Phylogenetic trees were built for each genus using the consensus of core gene trees with ASTRID version 2.2.1 + FastME under NNI and SPR distance-based methods and 100 bootstrap replicates (Fig. S1) (Vachaspati and Warnow 2015). (ii) The aligned core genes were concatenated and used for constructing a phylogenetic tree (Fig. 2.1) for each genus using RAxML v 8.2.12 (Stamatakis 2014) with a GTR+GAMMA model and 100 bootstrap replicates.

2.4.2 Species Delineation by Core Genome Nucleotide Sequence Identity

Genomes were clustered into ANI-species by calculating the ANI of the previously generated core gene concatenates (i.e., core ANI). Distance matrices were calculated for each genus using the distmat feature of the Emboss software suite (https://www.bioinformatics.nl/cgi-bin/emboss/distmat). Using R (v4.2.2, Core Team 2022), all genomes with a core genome identity score of ≥94% were clustered into de novo ANI-species as described in (Wittouck et al. 2019). Six ANI clusters within *Gilliamella* comprised of only one representative genome and were not used for within-species analyses, resulting in 11 *Gilliamella*, five *Snodgrassella* and *Lactobacillus*, respectively, two *Bombella*, *Bifidobacterium* and *Bombilactobacillus*, respectively and one *Apilactobacillus* ANI-species for downstream analyses.

2.4.3 Species Delineation by Inferring Gene Flow

To infer instances of homologous recombination of core genes between ANI-species, distance matrices of the core genome concatenate were obtained from the trees computed with RAxML (see above). The matrices of distances and concatenates were used to infer the ratio of homoplasic and non-homoplasic alleles (*h/m*) using the ConSpeciFix framework as descried in (Bobay and Ochman 2017b; Bobay et al. 2018). Ratios of homoplasic to non-homoplasic alleles *h/m* were generated for each ANI-species alone and for each pair of ANI-species when enough genomes (≥15) were available. Two ANI-species were considered as engaging in gene flow when their *h/m* value remained similar when calculated alone or when included with each other (i.e., not statistically different). Additionally, individual trees for each core gene family were built using RAxML v8.0.0. Tree topologies for each gene were compared against the genus tree and rooted using one of our previously defined ANI-species. Non-monophyletic genes trees were inferred as being introgressed, and the source ANI-species was identified when possible. We further tested whether the introgressed genes presented tree topologies that were significantly different from the topology of the corresponding core genome phylogeny. For *Bifidobacterium*, *Gilliamella*, *Snodgrassella* and *Lactobacillus*, topologies of individual gene trees were compared to the topology of the corresponding genome tree using the Approximately Unbiased (AU) test implemented in IQ-TREE (Shimodaira 2002; Minh et al. 2020). Because the trees needed to present the extract same set taxa between gene trees and genome trees, we could not analyze all the individual genes trees. We tested 54% of the genes that we predicted as introgressed (n=2,169) and found that 99.7% of them (n=2,163) displayed a tree topology that was significantly different from the topology of the corresponding genome tree ($P < 0.05$ after Benjamini-Hochberg adjustment for multiple testing. These analyses strongly support the view that genes that we predicted as introgressed have evolved through a path that is incompatible with vertical evolution.

2.4.4 Inference of Horizontal Gene Transfers (HGT) Between ANI-species

For each genome pair, Usearch Global was used to identify homologous accessory genes using a bidirectional best hit of >70% nucleotide identity. For each pair of genomes, we built the distribution of the identity scores of the genes of the core genome. Core genes encode housekeeping functions that evolve slowly due to purifying selection, whereas accessory genes are less constrained by selection and typically evolve faster than the core genome (Touchon et al. 2009). We used the divergence of core genes to estimate the typical rate of divergence through vertical inheritance for each pair of genomes. For each pair of genomes, we used the distribution of the identity scores of their core genes as a reference to define accessory genes that have been exchanged by HGT. To infer HGT events, we tested three thresholds of nucleotide identity which corresponds to the top 1st, 3rd and 5th percentiles of the distribution of core gene identity for each pair of genomes within each genus. The differences between these thresholds and the identity of the accessory gene scores are significantly different for all three thresholds tested (*t*-test and Wilcoxon tests with Bonferroni correction). We selected the most stringent threshold of nucleotide identity which corresponds to the top 1% of the distribution of core gene identity. Accessory genes were then defined as HGT events when their identity score was above the 99th percentile of the identity scores of the core genes. This is a stringent threshold for inferring accessory gene transfers, as they would not be expected to be evolving more slowly than core genes. Pairs of genes with highly similar core genomes (when the 99th percentile corresponds to 100% identity) cannot be used to distinguish horizontally acquired genes from vertically inherited ones and HGT events were therefore not inferred between these pairs of genomes (it is the case of all the genomes within the *Apilactobacillus* genus). The number of HGT events that we estimated was not biased by the different sequence thresholds used across strains. HGT transfers between ANI-species were adjusted for the number of strains involved; strain number within each ANI-species ranged from two to 35.

2.4.5 HGT Events Between Genera

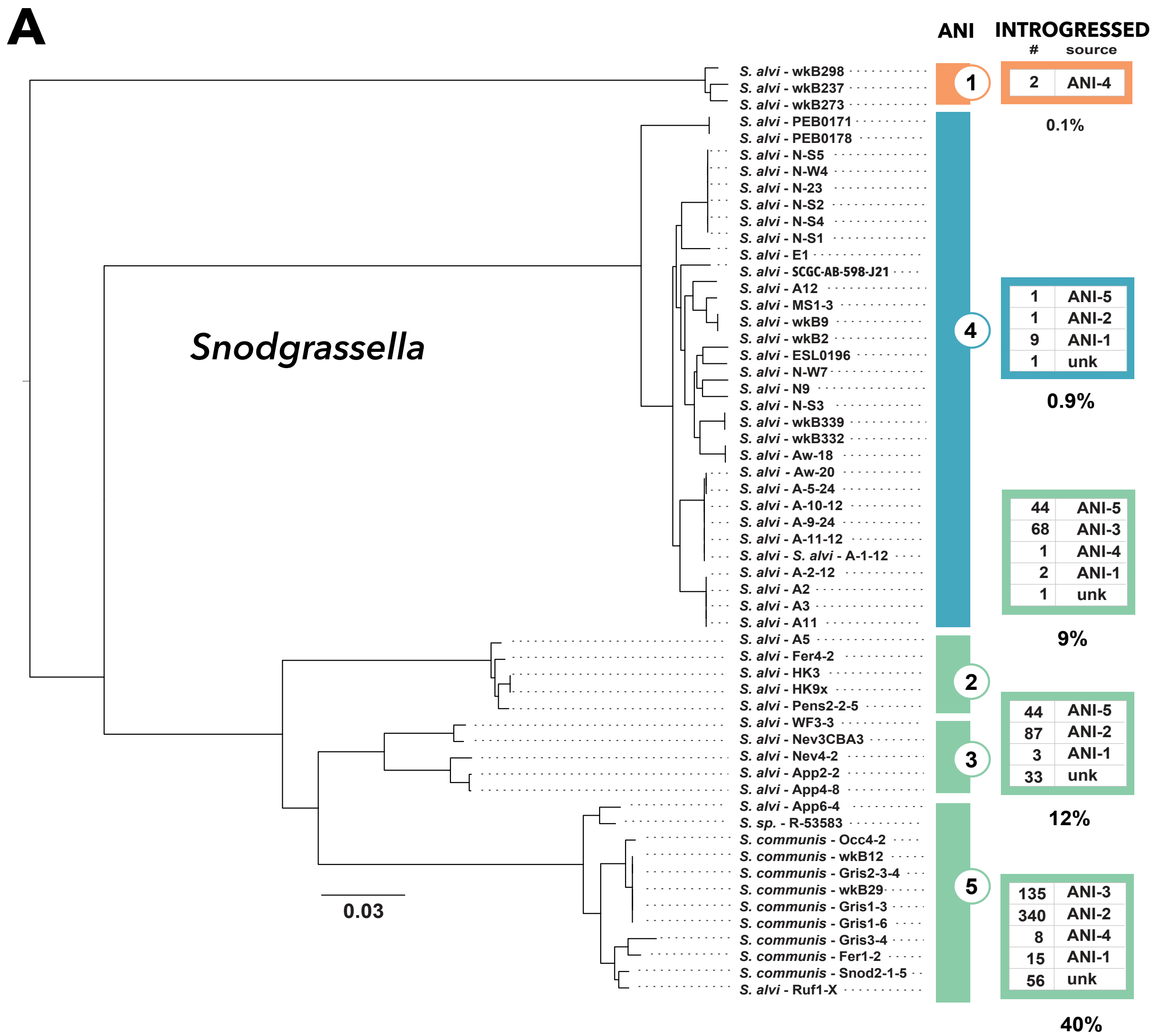
To determine the amino acid identity between the two clades, whole genome protein sequence alignments from each pair of genera included in this study were compared using AAI-calculator FastAAI (Konstantinidis et al. 2022). Pairwise whole genome nucleotide comparisons between each pair of genomes (for all genera) were computed using Usearch Global with a 50% identity cutoff. Using the previous method for inferring HGT based on the upper 99th percentile of the distribution of their core gene identity as a threshold, HGT events were inferred for the accessory genes between the two genera.

2.4.6 Pangenome Construction and Gene Annotation

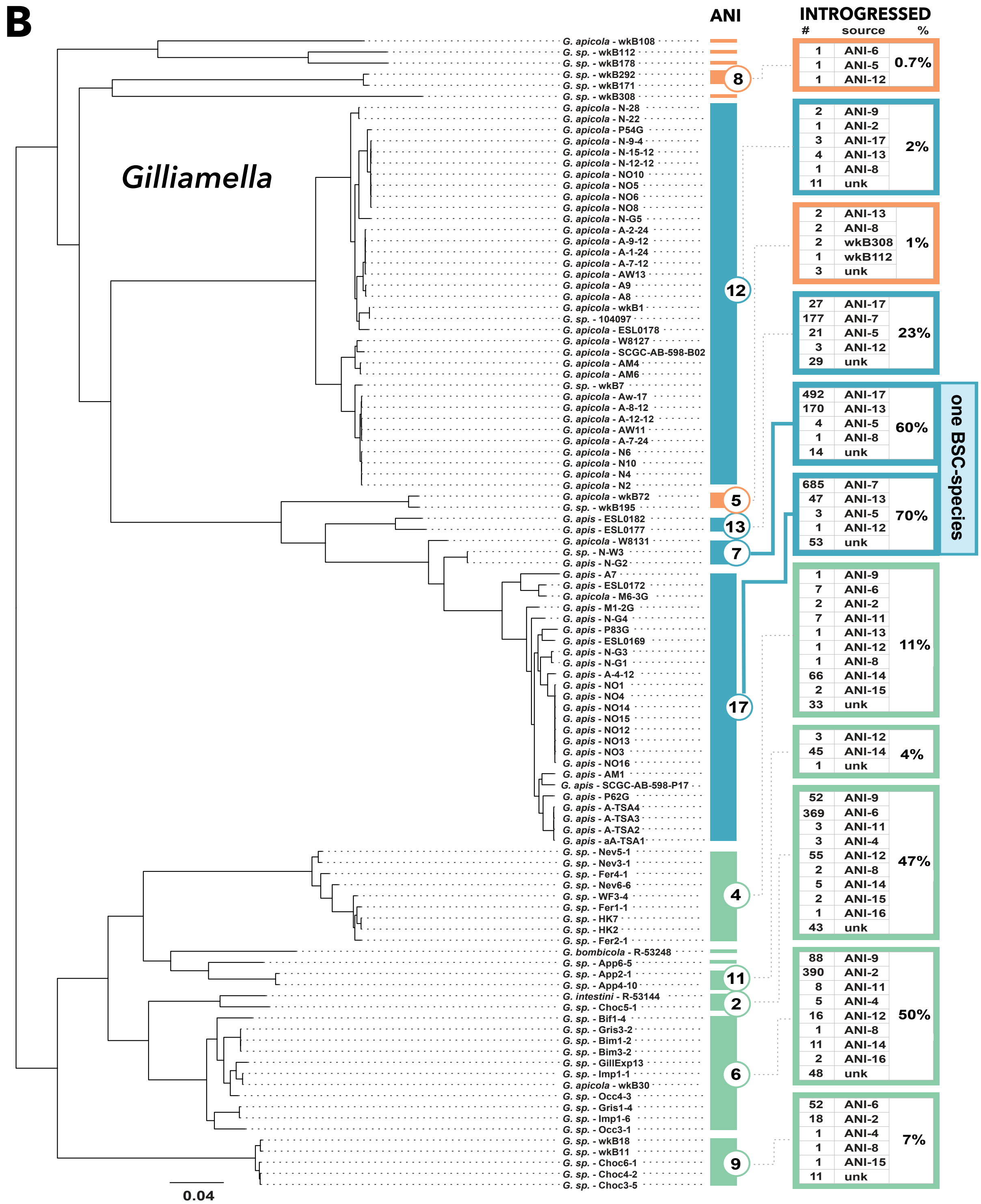
From the pairwise comparisons computed above with Usearch Global, homologous gene families were built by transitivity for each genus, i.e., two homologous sequences were systematically classified into the same gene family. The pangenome of each genus was defined as the entire set of homologous gene families. Orphan genes were counted as their own gene family. Gene annotation was performed using eggNOG-mapper v2.1.4 and the DIAMOND (Buchfink et al. 2021) database for identifying the function of each pangenome family from the

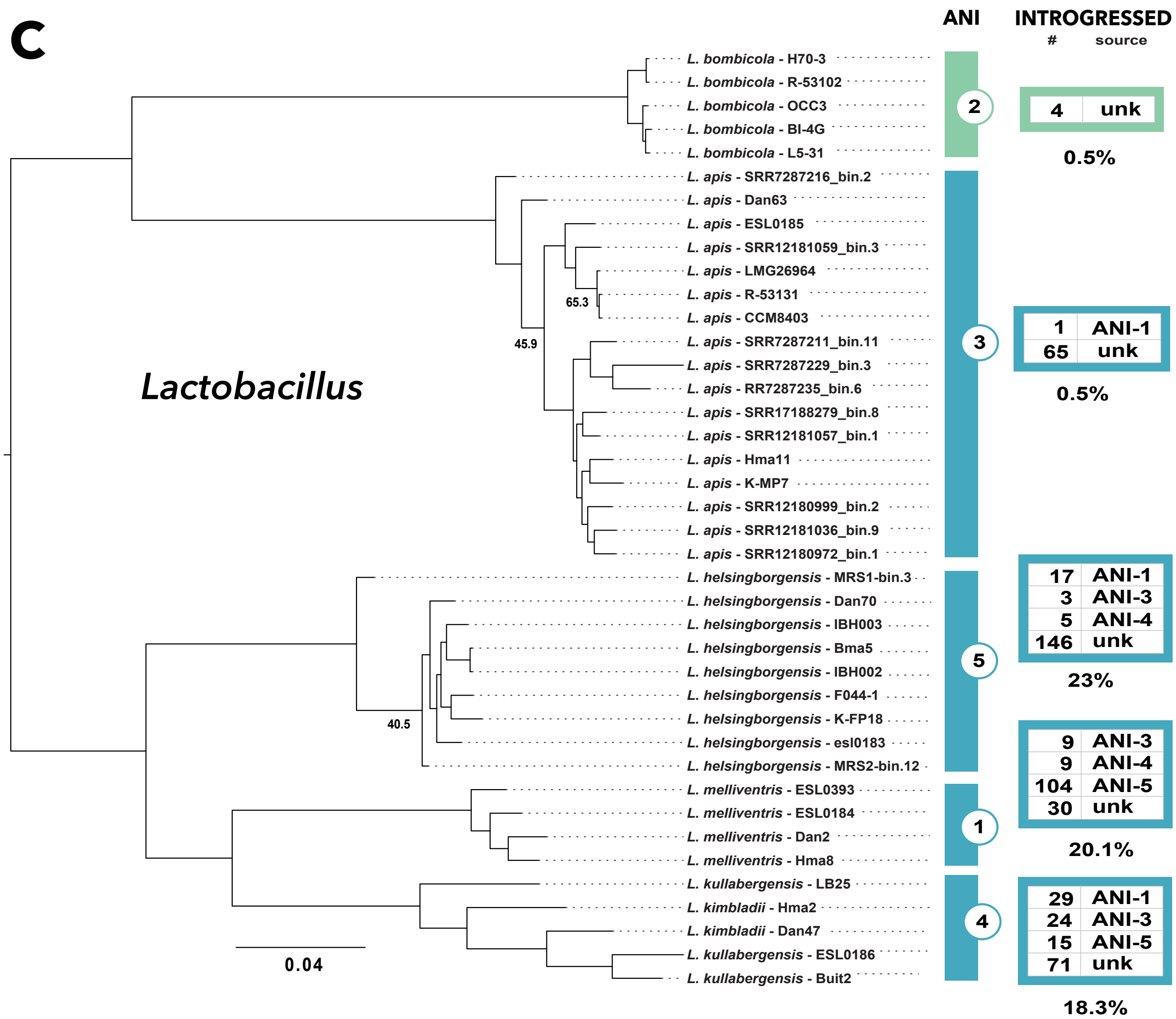
Clusters of Orthologous Groups (COGs) database. In addition, antibiotic resistance genes were interpedently characterized using the RGI web server (https://card.mcmaster.ca/analyze/rgi).

Figure 2.1. Phylogeny and species definition of the four major genera. a) *Snodgrassella*. b) *Gilliamella*. c) *Lactobacillus*. d) *Bifidobacterium*. Species borders were defined using average nucleotide identity (ANI-species) and patterns of gene flow (BSC-species), which were highly congruent. For each strain, the bee host is indicated in blue for Western honeybees, orange for Eastern honeybees and green for bumblebees. Each phylogenetic tree was built using the core genome concatenate of each genus with a maximum likelihood method with 100 bootstrap replicates. Internal nodes without values indicate bootstrap values >90. Introgression events were inferred using the species tree and the phylogenetic tree of each core gene built independently. When identified, the source of the introgressed sequence was indicated in the right column.

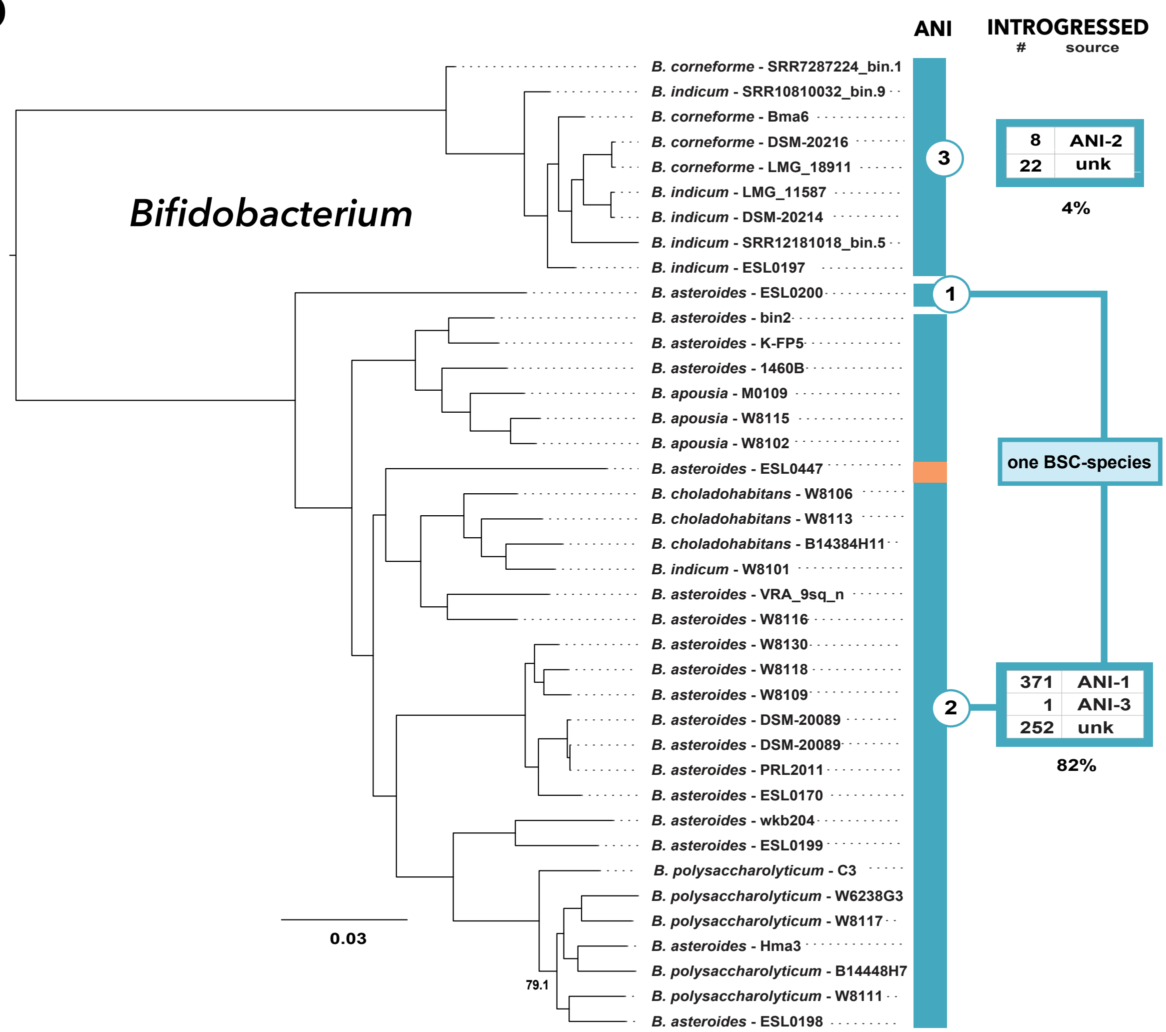


A
ANI
INTRODUCED
#
source
Snodgrassella
S. alvi - wkB298
S. alvi - wkB237
S. alvi - wkB273
S. alvi - PEB0171
S. alvi - PEB0178
S. alvi - N-S5
S. alvi - N-W4
S. alvi - N-23
S. alvi - N-S2
S. alvi - N-S4
S. alvi - N-S1
S. alvi - E1
S. alvi - SCGC-AB-598-J21
S. alvi - A12
S. alvi - MS1-3
S. alvi - wkB9
S. alvi - wkB2
S. alvi - ESL0196
S. alvi - N-W7
S. alvi - N9
S. alvi - N-S3
S. alvi - wkB339
S. alvi - wkB332
S. alvi - Aw-18
S. alvi - Aw-20
S. alvi - A-5-24
S. alvi - A-10-12
S. alvi - A-9-24
S. alvi - A-11-12
S. alvi - S. alvi - A-1-12
S. alvi - A-2-12
S. alvi - A2
S. alvi - A3
S. alvi - A11
S. alvi - A5
S. alvi - Fer4-2
S. alvi - HK3
S. alvi - HK9x
S. alvi - Pens2-2-5
S. alvi - WF3-3
S. alvi - Nev3CBA3
S. alvi - Nev4-2
S. alvi - App2-2
S. alvi - App4-8
S. alvi - App6-4
S. sp. - R-53583
S. communis - Occ4-2
S. communis - wkB12
S. communis - Gris2-3-4
S. communis - wkB29
S. communis - Gris1-3
S. communis - Gris1-6
S. communis - Gris3-4
S. communis - Fer1-2
S. communis - Snod2-1-5
S. alvi - Ruf1-X
0.03
1
2 ANI-4
0.1%
4
1 ANI-5
1 ANI-2
9 ANI-1
1 unk
0.9%
44 ANI-5
68 ANI-3
1 ANI-4
2 ANI-1
1 unk
9%
2
3
44 ANI-5
87 ANI-2
3 ANI-1
33 unk
12%
5
135 ANI-3
340 ANI-2
8 ANI-4
15 ANI-1
56 unk
40%


B
Gilliamella
ANI
INTROGRESSED
source %
one BSC-species
0.04


C
ANI
INTROGRESSED
#
source
L. bombicola - H70-3
L. bombicola - R-53102
L. bombicola - OCC3
L. bombicola - BI-4G
L. bombicola - L5-31
L. apis - SRR7287216_bin.2
L. apis - Dan63
L. apis - ESL0185
L. apis - SRR12181059_bin.3
L. apis - LMG26964
L. apis - R-53131
L. apis - CCM8403
L. apis - SRR7287211_bin.11
L. apis - SRR7287229_bin.3
L. apis - RR7287235_bin.6
L. apis - SRR17188279_bin.8
L. apis - SRR12181057_bin.1
L. apis - Hma11
L. apis - K-MP7
L. apis - SRR12180999_bin.2
L. apis - SRR12181036_bin.9
L. apis - SRR12180972_bin.1
L. helsingborgensis - MRS1-bin.3
L. helsingborgensis - Dan70
L. helsingborgensis - IBH003
L. helsingborgensis - Bma5
L. helsingborgensis - IBH002
L. helsingborgensis - F044-1
L. helsingborgensis - K-FP18
L. helsingborgensis - esl0183
L. helsingborgensis - MRS2-bin.12
L. melliventris - ESL0393
L. melliventris - ESL0184
L. melliventris - Dan2
L. melliventris - Hma8
L. kullabergensis - LB25
L. kimbladii - Hma2
L. kimbladii - Dan47
L. kullabergensis - ESL0186
L. kullabergensis - Buit2
65.3
45.9
40.5
Lactobacillus
0.04
2
3
5
1
4
4 unk
0.5%
1 ANI-1
65 unk
0.5%
17 ANI-1
3 ANI-3
5 ANI-4
146 unk
23%
9 ANI-3
9 ANI-4
104 ANI-5
30 unk
20.1%
29 ANI-1
24 ANI-3
15 ANI-5
71 unk
18.3%

D


ANI
INTROGRESSED
source
Bifidobacterium
B. corneforme - SRR7287224_bin.1
B. indicum - SRR10810032_bin.9
B. corneforme - Bma6
B. corneforme - DSM-20216
B. corneforme - LMG_18911
B. indicum - LMG_11587
B. indicum - DSM-20214
B. indicum - SRR12181018_bin.5
B. indicum - ESL0197
B. asteroides - ESL0200
B. asteroides - bin2
B. asteroides - K-FP5
B. asteroides - 1460B
B. apousia - M0109
B. apousia - W8115
B. apousia - W8102
B. asteroides - ESL0447
B. choladohabitans - W8106
B. choladohabitans - W8113
B. choladohabitans - B14384H11
B. indicum - W8101
B. asteroides - VRA_9sq_n
B. asteroides - W8116
B. asteroides - W8130
B. asteroides - W8118
B. asteroides - W8109
B. asteroides - DSM-20089
B. asteroides - DSM-20089
B. asteroides - PRL2011
B. asteroides - ESL0170
B. asteroides - wkb204
B. asteroides - ESL0199
B. polysaccharolyticum - C3
B. polysaccharolyticum - W6238G3
B. polysaccharolyticum - W8117
B. asteroides - Hma3
B. polysaccharolyticum - B14448H7
B. polysaccharolyticum - W8111
B. asteroides - ESL0198
79.1
0.03
3
8 ANI-2
22 unk
4%
1
one BSC-species
2
371 ANI-1
1 ANI-3
252 unk
82%

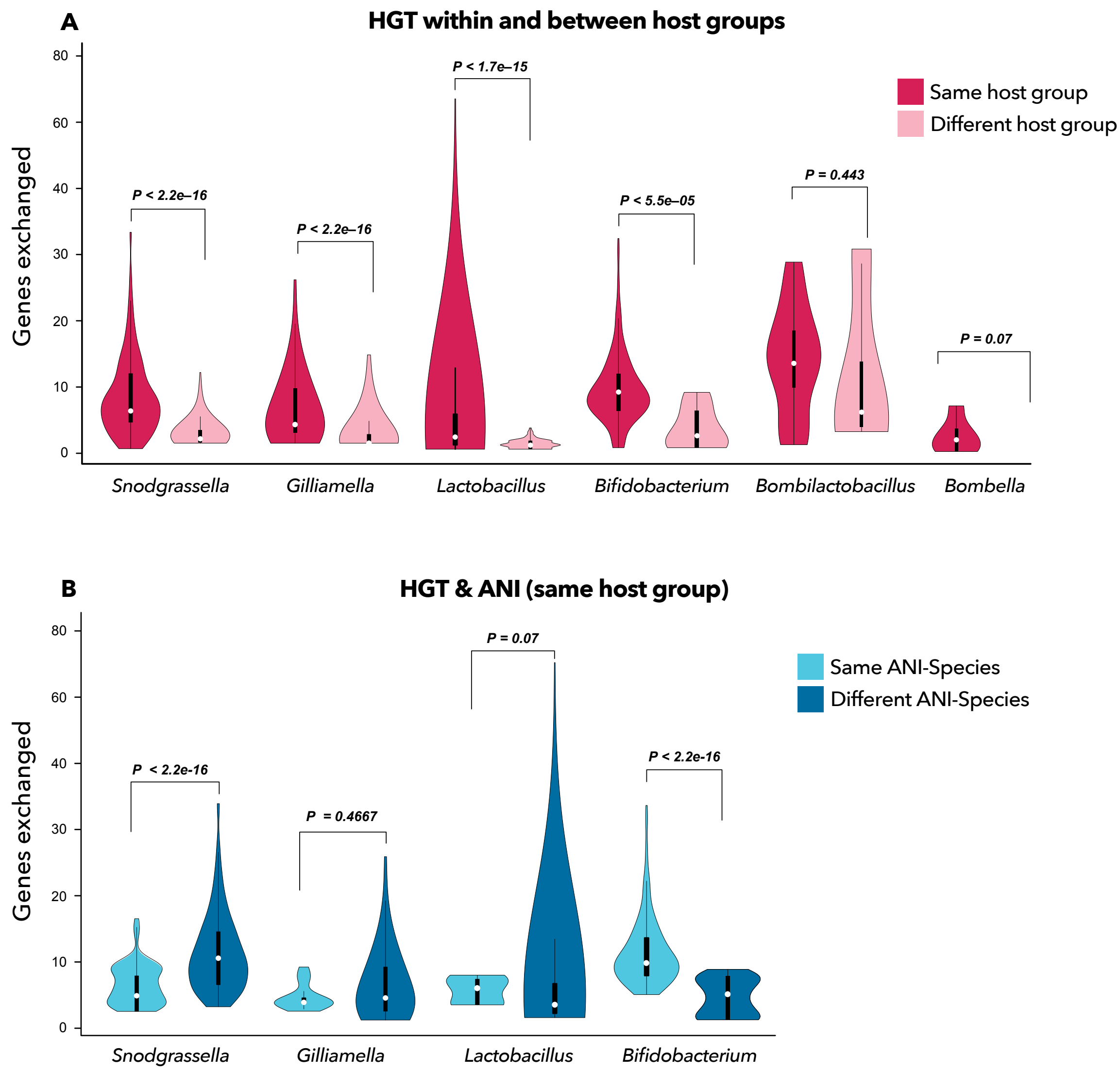


Figure 2.2. Patterns of HGT across species within seven genera of the honeybee gut. a) Average number of accessory genes exchanged between species found within the same bee host group or between different bee host groups. b) Average number of accessory genes exchanged between bacteria classified as part of the same ANI-species or classified as different ANI-species. Statistics were conducted with a Wilcoxon test.

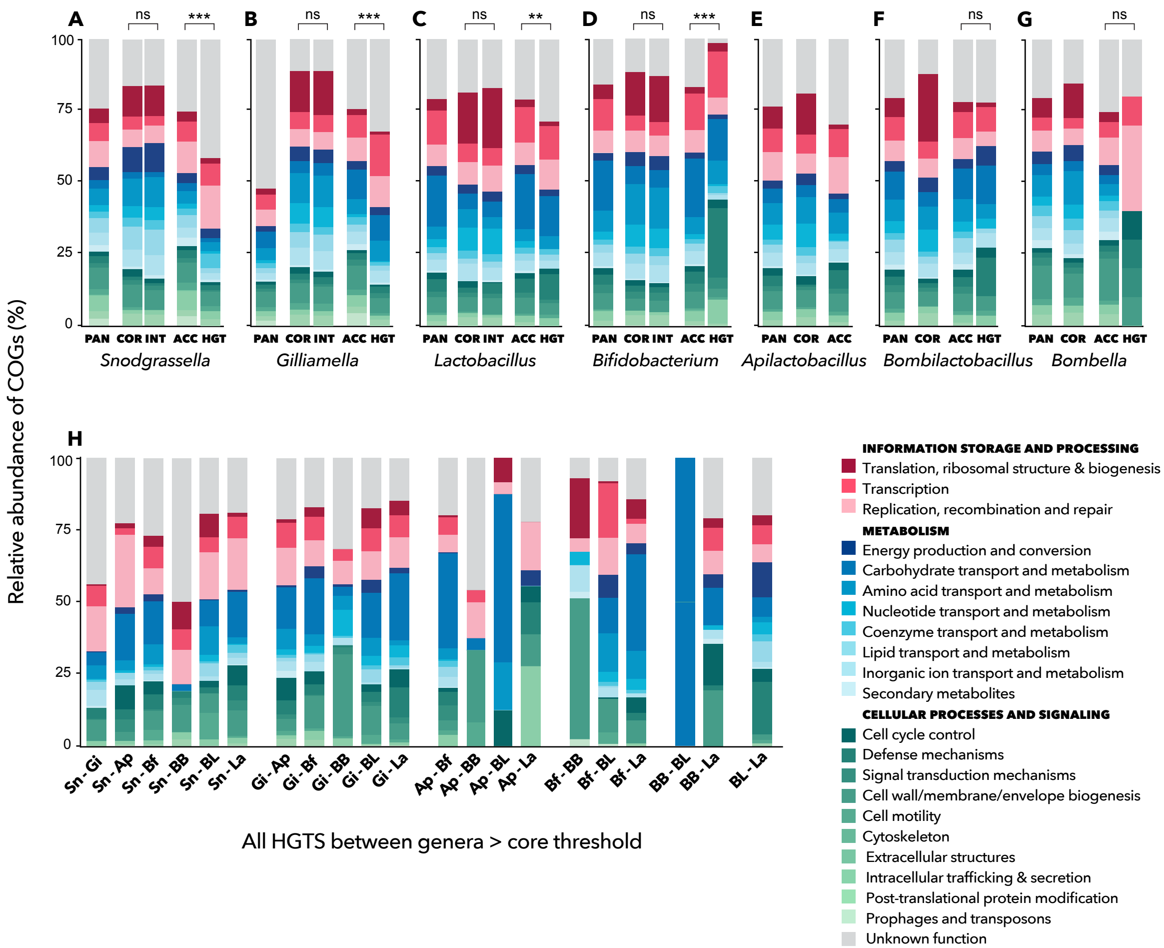


Figure 2.3. Functional characterization of DNA transfers within and between seven genera. Gene functions shown for pangenomes. Comparisons between core and introgressed genes as well as accessory genome HGT genes. Results for: a) *Snodgrassella*, b) *Gilliamella*, c) *Lactobacillus*, d) *Bifidobacterium*, e) *Apilactobacillus*, f) *Bombilactobacillus* and g) *Bombella* respectively. h) HGT events inferred for functionality between each pair of genera, respectively, where *Snodgrassella*=Sn, *Gilliamella*=Gi, *Lactobacillus*=La, *Bifidobacterium*=Bf, *Apilactobacillus*=Ap, *Bombilactobacillus*=BL and *Bombella*=BB. Statistics were performed with a Chi-square test (*: P<0.05, **: P<0.01, ***: P<0.001).

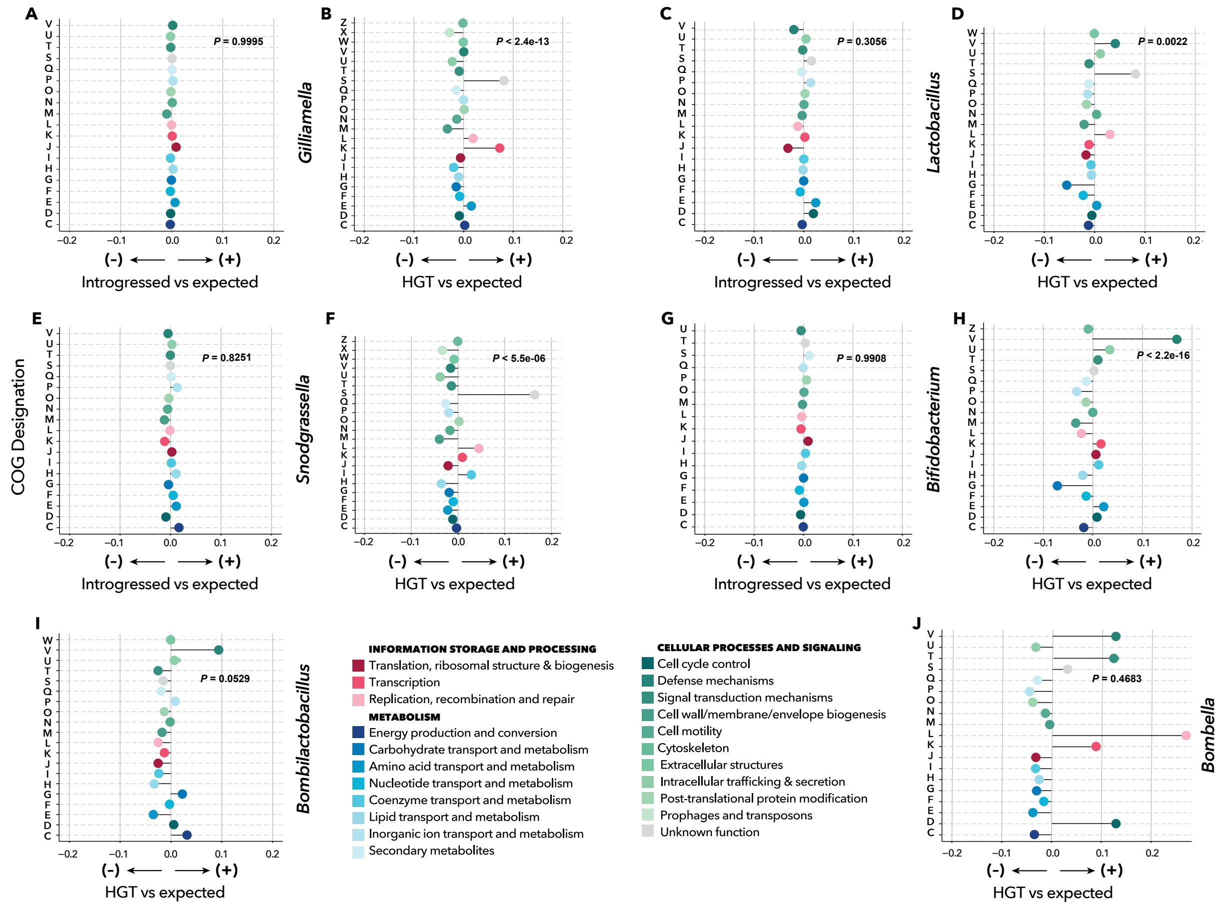
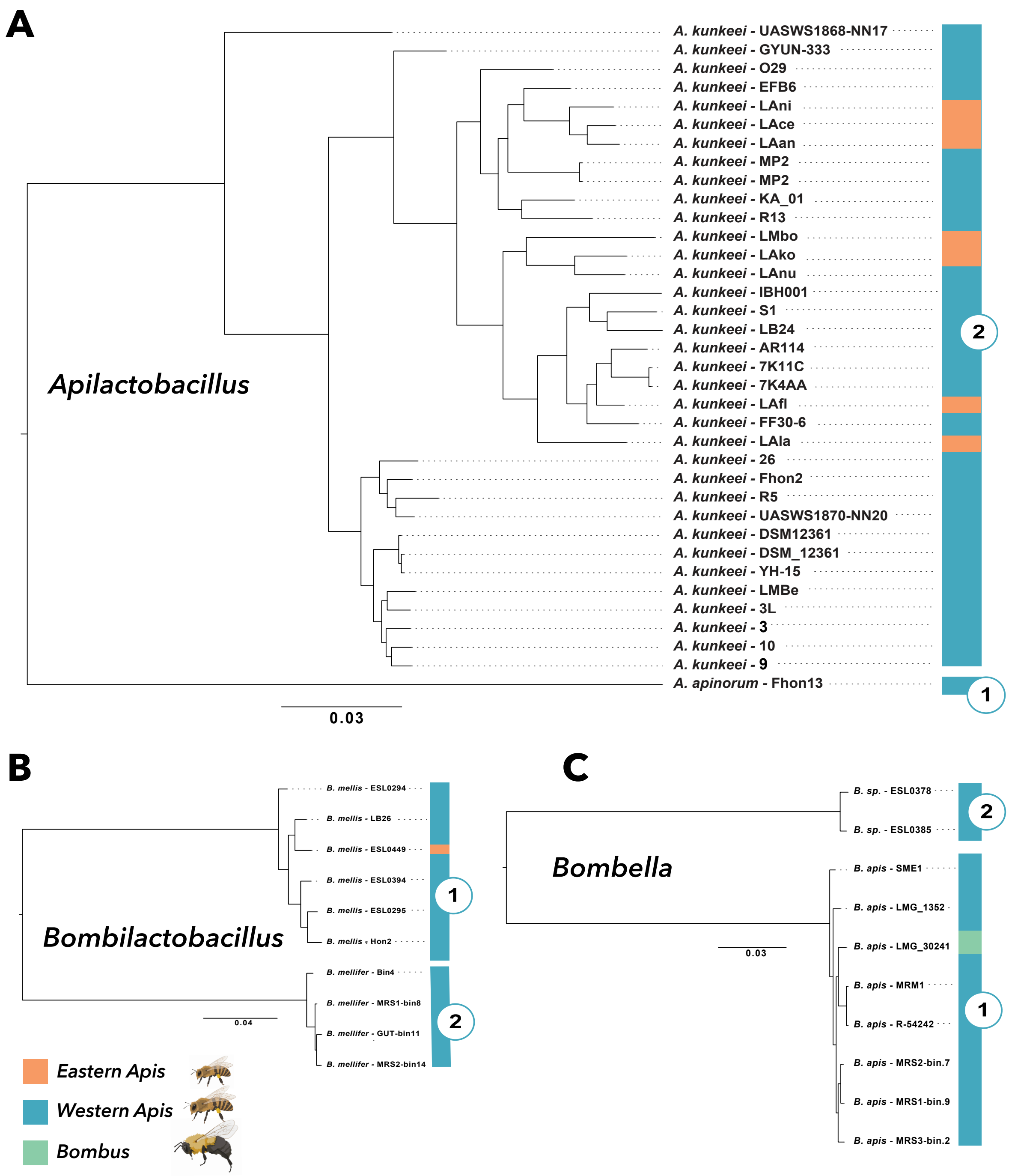


Figure 2.4. Enrichment of functional categories exchanged by introgression and HGT. Functional categories enriched (+) or depleted (–) for introgressed core genes as expected based on the overall gene functions of the core genome and HGT events as expected based on the overall gene functions of the accessory genome for *Gilliamella* (a & b), *Lactobacillus* (c & d), *Snodgrassella* (e & f), *Bifidobacterium* (g & h), HGT events in *Bombilactobacillus* (i), and HGT events in *Bombella* (j). Statistics were conducted using Chi-square tests with Bonferroni corrections.

Figure S1 (maybe 2.5). Phylogeny and species definition of the three minor genera: a) *Apilactobacillus*, b) *Bombilactobacillus* and c) *Bombella*. Species borders were defined using average nucleotide identity (ANI-species) and patterns of gene flow (BSC-species), which were highly congruent. For each strain, the bee host is indicated in blue for Western honeybees, orange for Eastern honeybees and green for bumblebees. Each phylogenetic tree was built using the core genome of each genus with a maximum likelihood method with 100 bootstrap replicates. Internal nodes without values indicate bootstrap values >90. Introgression events were inferred using the species tree and the phylogenetic tree of each core gene built independently. When identified, the source of the introgressed sequence was indicated in the right column.


A
Apilactobacillus
A. kunkeei - UASWS1868-NN17
A. kunkeei - GYUN-333
A. kunkeei - O29
A. kunkeei - EFB6
A. kunkeei - LAni
A. kunkeei - LAce
A. kunkeei - LAan
A. kunkeei - MP2
A. kunkeei - MP2
A. kunkeei - KA_01
A. kunkeei - R13
A. kunkeei - LMbo
A. kunkeei - LAko
A. kunkeei - LAnu
A. kunkeei - IBH001
A. kunkeei - S1
A. kunkeei - LB24
A. kunkeei - AR114
A. kunkeei - 7K11C
A. kunkeei - 7K4AA
A. kunkeei - LAfl
A. kunkeei - FF30-6
A. kunkeei - LAla
A. kunkeei - 26
A. kunkeei - Fhon2
A. kunkeei - R5
A. kunkeei - UASWS1870-NN20
A. kunkeei - DSM12361
A. kunkeei - DSM_12361
A. kunkeei - YH-15
A. kunkeei - LMBe
A. kunkeei - 3L
A. kunkeei - 3
A. kunkeei - 10
A. kunkeei - 9
A. apinorum - Fhon13
2
1
0.03
B
Bombilactobacillus
B. mellis - ESL0294
B. mellis - LB26
B. mellis - ESL0449
B. mellis - ESL0394
B. mellis - ESL0295
B. mellis - Hon2
B. mellifer - Bin4
B. mellifer - MRS1-bin8
B. mellifer - GUT-bin11
B. mellifer - MRS2-bin14
1
2
0.04
Eastern Apis
Western Apis
Bombus
C
Bombella
B. sp. - ESL0378
B. sp. - ESL0385
B. apis - SME1
B. apis - LMG_1352
B. apis - LMG_30241
B. apis - MRM1
B. apis - R-54242
B. apis - MRS2-bin.7
B. apis - MRS1-bin.9
B. apis - MRS3-bin.2
2
1
0.03

REFERENCES

Aljedani DM. 2022. Antibiotic treatment (Tetracycline) effect on bio-efficiency of the larvae honey bee (Apis mellifera jemenatica). *Saudi J. Biol. Sci.* 29:1477–1486.

Anon. ENY152/IN868: The Benefits of Pollen to Honey Bees. *Ask IFAS - Powered EDIS* [Internet]. Available from: https://edis.ifas.ufl.edu/publication/IN868

Arevalo P, VanInsberghe D, Elsherbini J, Gore J, Polz MF. 2019. A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations. *Cell* 178:820-834.e14.

Avise JC, Wollenberg K. 1997. Phylogenetics and the origin of species. *Proc. Natl. Acad. Sci.* 94:7748–7755.

Bartoš O, Chmel M, Swierczková I. 2024. The overlooked evolutionary dynamics of 16S rRNA revises its role as the “gold standard” for bacterial species identification. *Sci. Rep.* 14:9067.

Bendall ML, Stevens SLR, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, et al. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* 10:1589–1601.

Bennett GM, Moran NA. 2015. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci.* 112:10169–10176.

Berríos P, Fuentes JA, Salas D, Carreño A, Aldea P, Fernández F, Trombert AN. 2018. Inhibitory effect of biofilm-forming Lactobacillus kunkeei strains against virulent Pseudomonas aeruginosa in vitro and in honeycomb moth (Galleria mellonella) infection model. *Benef. Microbes* 9:257–268.

Bobay L-M. 2020. The Prokaryotic Species Concept and Challenges. In: Tettelin H, Medini D, editors. The Pangenome. Cham: Springer International Publishing. p. 21–49. Available from: https://link.springer.com/10.1007/978-3-030-38281-0_2

Bobay L-M, Ellis BS-H, Ochman H. 2018. ConSpeciFix: classifying prokaryotic species based on gene flow.Hancock J, editor. *Bioinformatics* 34:3738–3740.

Bobay L-M, Ochman H. 2017a. Biological Species Are Universal across Life’s Domains. *Genome Biol. Evol.* 9:491–501.

Bobay L-M, Ochman H. 2017b. Impact of Recombination on the Base Composition of Bacteria and Archaea. *Mol. Biol. Evol.* 34:2627–2636.

Bobay L-M, Ochman H. 2017c. The Evolution of Bacterial Genome Architecture. *Front. Genet.* 8:72.

Bobay L-M, Ochman H. 2018. Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evol. Biol.* 18:153.

Bobay L-M, Raymann K. 2019. Population Genetics of Host-Associated Microbiomes. *Curr. Mol. Biol. Rep.* 5:128–139.

Bobay L-M, Rocha EPC, Touchon M. 2013. The Adaptation of Temperate Bacteriophages to Their Host Genomes. *Mol. Biol. Evol.* 30:737–751.

Bobay L-M, Traverse CC, Ochman H. 2015. Impermanence of bacterial clones. *Proc. Natl. Acad. Sci.* 112:8893–8900.

Bobay L-M, Wissel EF, Raymann K. 2020. Strain Structure and Dynamics Revealed by Targeted Deep Sequencing of the Honey Bee Gut Microbiome.Campbell BJ, editor. *mSphere* [Internet] 5. Available from: https://journals.asm.org/doi/10.1128/mSphere.00694-20

Bosmans L, Pozo MI, Verreth C, Crauwels S, Wäckers F, Jacquemyn H, Lievens B. 2018. Hibernation Leads to Altered Gut Communities in Bumblebee Queens (Bombus terrestris). *Insects* 9:188.

Bradford EL, Wax N, Bueren EK, Walke JB, Fell R, Belden LK, Haak DC. 2022. Comparative genomics of Lactobacillaceae from the gut of honey bees, *Apis mellifera* , from the Eastern United States.Engel P, editor. *G3 GenesGenomesGenetics* 12:jkac286.

Brochet S, Quinn A, Mars RA, Neuschwander N, Sauer U, Engel P. 2021. Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife* 10:e68583.

Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18:366–368.

Capinera JL. 2008. Encyclopedia of entomology. Second ed. Dordrecht: Springer

Cardinal S, Danforth BN. 2011. The Antiquity and Evolutionary History of Social Behavior in Bees.Moreau CS, editor. *PLoS ONE* 6:e21086.

Cariveau DP, Elijah Powell J, Koch H, Winfree R, Moran NA. 2014. Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (Bombus). *ISME J.* 8:2369–2379.

Cerqueira AES, Hammer TJ, Moran NA, Santana WC, Kasuya MCM, Da Silva CC. 2021. Extinction of anciently associated gut bacterial symbionts in a clade of stingless bees. *ISME J.* 15:2813–2816.

Cohan FM. 2002. What are Bacterial Species? *Annu. Rev. Microbiol.* 56:457–487.

Cohan FM, Perry EB. 2007. A Systematics for Discovering the Fundamental Units of Bacterial Diversity. *Curr. Biol.* 17:R373–R386.

Colwell MJ, Williams GR, Evans RC, Shutler D. 2017. Honey bee-collected pollen in agro-ecosystems reveals diet diversity, diet quality, and pesticide exposure. *Ecol. Evol.* 7:7243–7253.

Conrad RE, Brink CE, Viver T, Rodriguez-R LM, Aldeguer-Riquelme B, Hatt JK, Venter SN, Rossello-Mora R, Amann R, Konstantinidis KT. 2024. Microbial species and intraspecies units exist and are maintained by ecological cohesiveness coupled to high homologous recombination. *Nat. Commun.* 15:9906.

Corby-Harris V, Snyder L, Meador CAD, Naldo R, Mott B, Anderson KE. 2016. *Parasaccharibacter apium* , gen. nov., sp. nov., Improves Honey Bee (Hymenoptera: Apidae) Resistance to *Nosema*. *J. Econ. Entomol.* 109:537–543.

Cornet L, Cleenwerck I, Praet J, Leonard RR, Vereecken NJ, Michez D, Smagghe G, Baurain D, Vandamme P. 2022. Phylogenomic Analyses of *Snodgrassella* Isolates from Honeybees and Bumblebees Reveal Taxonomic and Functional Diversity.Pupo MT, editor. *mSystems* 7:e01500-21.

Cracraft J. 1983. Species Concepts and Speciation Analysis. In: Johnston RF, editor. Current Ornithology. New York, NY: Springer US. p. 159–187. Available from: https://doi.org/10.1007/978-1-4615-6781-3_6

Crook N, Ferreiro A, Gasparrini AJ, Pesesky MW, Gibson MK, Wang B, Sun X, Condiotte Z, Dobrowolski S, Peterson D, et al. 2019. Adaptive Strategies of the Candidate Probiotic E. coli Nissle in the Mammalian Gut. *Cell Host Microbe* 25:499-512.e8.

Danforth BN, Conway L, Ji SH. 2003. Phylogeny of Eusocial Lasioglossum Reveals Multiple Losses of Eusociality within a Primitively Eusocial Clade of Bees (Hymenoptera: Halictidae). *Syst. Biol.* 52:14.

Daubin V, Szöllősi GJ. 2016. Horizontal Gene Transfer and the History of Life. *Cold Spring Harb. Perspect. Biol.* 8:a018036.

Diop A, Torrance EL, Stott CM, Bobay L-M. 2022. Gene flow and introgression are pervasive forces shaping the evolution of bacterial species. *Genome Biol.* 23:239.

Doolittle WF. 2012. Population Genomics: How Bacterial Species Form and Why They Don't Exist. *Curr. Biol.* 22:R451–R453.

Doolittle WF. 2019. Speciation without Species: A Final Word. *Philos. Theory Pract. Biol.* [Internet] 11. Available from: http://hdl.handle.net/2027/spo.16039257.0011.014

Douglas AE. 2019. Simple animal models for microbiome research. *Nat. Rev. Microbiol.* 17:764–775.

Dykhuizen DE, Green L. 1991. Recombination in Escherichia coli and the definition of biological species. 173.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.

Elekonich MM, Roberts SP. 2005. Honey bees as a model for understanding mechanisms of life history transitions. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* 141:362–371.

Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota. *Nat. Commun.* 10:446.

Ellegaard KM, Suenami S, Miyazaki R, Engel P. 2020. Vast Differences in Strain-Level Diversity in the Gut Microbiota of Two Closely Related Honey Bee Species. *Curr. Biol.* 30:2520-2531.e7.

Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci.* 109:11002–11007.

Engel P, Moran NA. 2013. Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. *Gut Microbes* 4:60–65.

Eppley JM, Tyson GW, Getz WM, Banfield JF. 2007. Genetic Exchange Across a Species Boundary in the Archaeal Genus Ferroplasma. *Genetics* 177:407–416.

Faust K, Bauchinger F, Laroche B, de Buyl S, Lahti L, Washburne AD, Gonze D, Widder S. 2018. Signatures of ecological processes in microbial community time series. *Microbiome* 6:120.

Fernandez De Landa G, Alberoni D, Braglia C, Baffoni L, Fernandez De Landa M, Revainera PD, Quintana S, Zumpano F, Maggi MD, Di Gioia D. 2024. The Gut Microbiome of Two Wild Bumble Bee Species Native of South America: Bombus pauloensis and Bombus bellicosus. *Microb. Ecol.* 87:121.

Finlay BJ. 2002. Global Dispersal of Free-Living Microbial Eukaryote Species. *Science* 296:1061–1063.

Foret S, Kucharski R, Pellegrini M, Feng S, Jacobsen SE, Robinson GE, Maleszka R. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc. Natl. Acad. Sci.* 109:4968–4973.

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity. *Science* 323:741–746.

Friedman J, Alm EJ, Shapiro BJ. 2013. Sympatric Speciation: When Is It Possible in Bacteria?Immler S, editor. *PLoS ONE* 8:e53539.

Garud NR, Good BH, Hallatschek O, Pollard KS. 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts.Gordo I, editor. *PLOS Biol.* 17:e3000102.

González-Torres P, Rodríguez-Mateos F, Antón J, Gabaldón T. 2019. Impact of Homologous Recombination on the Evolution of Prokaryotic Core Genomes.Heitman J, editor. *mBio* 10:e02494-18, /mbio/10/1/mBio.02494-18.atom.

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* 551:45–50.

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57:81–91.

Grimaldi D, Engel MS. 2005. Evolution of the Insects. Cambridge University Press

Groussin M, Mazel F, Alm EJ. 2020. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host Microbe* 28:12–22.

Groussin M, Poyet M, Sistiaga A, Kearney SM, Moniz K, Noel M, Hooker J, Gibbons SM, Segurel L, Froment A, et al. 2021. Elevated rates of horizontal gene transfer in the industrialized human microbiome. *Cell* 184:2053-2067.e18.

Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity: A Darwinian view of the evolution of microbes. *EMBO Rep.* 2:376–381.

Hammer TJ, Janzen DH, Hallwachs W, Jaffe SP, Fierer N. 2017. Caterpillars lack a resident gut microbiome. *Proc. Natl. Acad. Sci.* 114:9641–9646.

Hammer TJ, Le E, Martin AN, Moran NA. 2021. The gut microbiota of bumblebees. *Insectes Sociaux* 68:287–301.

Hanage WP. 2013. Fuzzy species revisited. *BMC Biol.* 11:41.

Hanage WP. 2016. Not So Simple After All: Bacteria, Their Population Genetics, and Recombination. *Cold Spring Harb. Perspect. Biol.* 8:a018069.

Harris CD, Torrance EL, Raymann K, Bobay L-M. 2021. CoreCruncher : Fast and Robust Construction of Core Genomes in Large Prokaryotic Data Sets.Ouangraoua A, editor. *Mol. Biol. Evol.* 38:727–734.

Hoetzinger M, Pitt A, Huemer A, Hahn MW. 2021. Continental-Scale Gene Flow Prevents Allopatric Divergence of Pelagic Freshwater Bacteria.Ochman H, editor. *Genome Biol. Evol.* 13:evab019.

Horak RD, Leonard SP, Moran NA. 2020. Symbionts shape host innate immunity in honeybees. *Proc. R. Soc. B Biol. Sci.* 287:20201184.

Hotchkiss MZ, Forrest JRK, Poulain AJ. 2024. Changes in bumblebee queen gut microbiotas during and after overwintering diapause. *Insect Mol. Biol.*:imb.12957.

Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. 2008. Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science* 320:1081–1085.

Ihle KE, Baker NA, Amdam GV. 2014. Insulin-like peptide response to nutritional input in honey bee workers. *J. Insect Physiol.* 69:49–55.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.

Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, et al. 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 158:1005–1015.

Kapheim KM, Johnson MM, Jolley M. 2021. Composition and acquisition of the microbiome in solitary, ground-nesting alkali bees. *Sci. Rep.* 11:2993.

Klein A-M, Vaissière BE, Cane JH, Steffan-Dewenter I, Cunningham SA, Kremen C, Tscharntke T. 2007. Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. B Biol. Sci.* 274:303–313.

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.

Koch H, Schmid-Hempel P. 2011. Bacterial Communities in Central European Bumblebees: Low Diversity and High Specificity. *Microb. Ecol.* 62:121–133.

Konstantinidis K, Ruiz-Perez C, Gerhardt K, Rodriguez-R L, Jain C, Tiedje J, Cole J. 2022. FastAAI: Efficient Estimation of Genome Average Amino Acid Identity and Phylum-level relationships using Tetramers of Universal Proteins. Available from: https://www.researchsquare.com/article/rs-1459378/v1

Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. B Biol. Sci.* 361:1929–1940.

Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci.* 102:2567–2572.

Kueneman JG, Bonadies E, Thomas D, Roubik DW, Wcislo WT. 2023. Neotropical bee microbiomes point to a fragmented social core and strong species-level effects. *Microbiome* 11:150.

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.* 111:11509–11514.

Kwong WK, Moran NA. 2013. Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of Snodgrassella alvi gen. nov., sp. nov., a member of the family Neisseriaceae of the Betaproteobacteria , and Gilliamella apicola gen. nov., sp. nov., a member of Orbaceae fam. nov., Orbales ord. nov., a sister taxon to the order ' Enterobacteriales ' of the Gammaproteobacteria. *Int. J. Syst. Evol. Microbiol.* 63:2008–2018.

Kwong WK, Moran NA. 2015. Evolution of host specialization in gut microbes: the bee gut as a model. *Gut Microbes* 6:214–220.

Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. *Nat. Rev. Microbiol.* 14:374–384.

Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, et al. 2012. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J. Clin. Microbiol.* 50:1355–1361.

Lawrence JG. 2002. Gene Transfer in Bacteria: Speciation without Species? *Theor. Popul. Biol.* 61:449–460.

Lee PC, Winston ML. 1985. The influence of swarm size on brood production and emergent worker weight in newly founded honey bee colonies (Apis mellifera L.). *Insectes Sociaux* 32:96–103.

Lee-Mäder E, Spivak M, Evans E. 2010. Managing Alternative Pollinators: A Handbook for Beekeepers, Growers, and Conservationists. Sustainable Agriculture Research and Education

Lenski RE. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J.* 11:2181–2194.

Lewis Z, Lizé A. 2015. Insect behaviour and the microbiome. *Curr. Opin. Insect Sci.* 9:86–90.

Li Y, Leonard SP, Powell JE, Moran NA. 2022. Species divergence in gut-restricted bacteria of social bees. *Proc. Natl. Acad. Sci.* 119:e2115013119.

Ludvigsen J, Porcellato D, Amdam GV, Rudi K. 2018. Addressing the diversity of the honeybee gut symbiont Gilliamella: description of Gilliamella apis sp. nov., isolated from the gut of honeybees (Apis mellifera). *Int. J. Syst. Evol. Microbiol.* 68:1762–1770.

Maiden MCJ, Van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11:728–736.

Martinson VG, Moy J, Moran NA. 2012. Establishment of Characteristic Gut Bacteria during Development of the Honeybee Worker. *Appl. Environ. Microbiol.* 78:2830–2840.

Mau B, Glasner JD, Darling AE, Perna NT. 2006. Genome-wide detection and analysis of homologous recombination among sequenced strains of Escherichia coli. *Genome Biol.* 7:R44.

Mignard S, Flandrois JP. 2006. 16S rRNA sequencing in routine bacterial identification: A 30-month experiment. *J. Microbiol. Methods* 67:574–581.

Miller DL, Smith EA, Newton ILG. 2021. A Bacterial Symbiont Protects Honey Bees from Fungal Disease.Graf J, editor. *mBio* 12:e00503-21.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.Teeling E, editor. *Mol. Biol. Evol.* 37:1530–1534.

Mockler BK, Kwong WK, Moran NA, Koch H. 2018. Microbiome Structure Influences Infection by the Parasite Crithidia bombi in Bumble Bees.Drake HL, editor. *Appl. Environ. Microbiol.* 84:e02335-17.

Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. 2016. Cospeciation of gut microbiota with hominids. *Science* 353:380–382.

Moran NA. 2001. The Coevolution of Bacterial Endosymbionts and Phloem-Feeding Insects. *Ann. Mo. Bot. Gard.* 88:35.

Moran NA. 2015. Genomics of the honey bee microbiome. *Curr. Opin. Insect Sci.* 10:22–28.

Moran NA, Hansen AK, Powell JE, Sabree ZL. 2012. Distinctive Gut Microbiota of Honey Bees Assessed Using Deep Sampling from Individual Worker Bees.Smagghe G, editor. *PLoS ONE* 7:e36393.

Motta EVS, Raymann K, Moran NA. 2018. Glyphosate perturbs the gut microbiota of honey bees. *Proc. Natl. Acad. Sci.* 115:10305–10310.

Murray TE, Coffey MF, Kehoe E, Horgan FG. 2013. Pathogen prevalence in commercially reared bumble bees and evidence of spillover in conspecific populations. *Biol. Conserv.* 159:269–276.

Nguyen ANT, Woods LC, Gorrell R, Ramanan S, Kwok T, McDonald MJ. 2022. Recombination resolves the cost of horizontal gene transfer in experimental populations of *Helicobacter pylori*. *Proc. Natl. Acad. Sci.* 119:e2119010119.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. 405:6.

Ochman H, Lerat E, Daubin V. 2005. Examining bacterial species under the specter of gene transfer and exchange. *Proc. Natl. Acad. Sci.* 102:6595–6599.

Ochman H, Worobey M, Kuo C-H, Ndjango J-BN, Peeters M, Hahn BH, Hugenholtz P. 2010. Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities.Achtman M, editor. *PLoS Biol.* 8:e1000546.

Oliveira PH, Touchon M, Cury J, Rocha EPC. 2017. The chromosomal organization of horizontal gene transfer in bacteria. *Nat. Commun.* 8:841.

Olofsson TC, Butler È, Markowicz P, Lindholm C, Larsson L, Vásquez A. 2016. Lactic acid bacterial symbionts in honeybees – an unknown key to honey's antimicrobial and therapeutic activities. *Int. Wound J.* 13:668–679.

Parish AJ, Rice DW, Tanquary VM, Tennessen JM, Newton ILG. 2022. Honey bee symbiont buffers larvae against nutritional stress and supplements lysine. *ISME J.* 16:2160–2168.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36:996–1004.

Pettis JS, Lichtenberg EM, Andree M, Stitzinger J, Rose R, vanEngelsdorp D. 2013. Crop Pollination Exposes Honey Bees to Pesticides Which Alters Their Susceptibility to the Gut Pathogen Nosema ceranae.Nascimento FS, editor. *PLoS ONE* 8:e70182.

Polz MF, Alm EJ, Hanage WP. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* 29:170–175.

Polz MF, Hunt DE, Preheim SP, Weinreich DM. 2006. Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos. Trans. R. Soc. B Biol. Sci.* 361:2009–2021.

Porter SS, Chang PL, Conow CA, Dunham JP, Friesen ML. 2017. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J.* 11:248–262.

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.

Praet J, Cnockaert M, Meeus I, Smagghe G, Vandamme P. 2017. Gilliamella intestini sp. nov., Gilliamella bombicola sp. nov., Gilliamella bombi sp. nov. and Gilliamella mensalis sp. nov.: Four novel Gilliamella species isolated from the bumblebee gut. *Syst. Appl. Microbiol.* 40:199–204.

de Queiroz K, Gauthier J. 1994. Toward a phylogenetic system of biological nomenclature. *Trends Ecol. Evol.* 9:27–31.

Ravenhall M, Škunca N, Lassalle F, Dessimoz C. 2015. Inferring Horizontal Gene Transfer.Wodak S, editor. *PLOS Comput. Biol.* 11:e1004095.

Raymann K, Shaffer Z, Moran NA. 2017. Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees.Gore J, editor. *PLOS Biol.* 15:e2001861.

Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The Natural History of Nitrogen Fixation. *Mol. Biol. Evol.* 21:541–554.

Retchless AC, Lawrence JG. 2007. Temporal Fragmentation of Speciation in Bacteria. *Science* 317:1093–1096.

Richards VP, Velsko IM, Alam MT, Zadoks RN, Manning SD, Pavinski Bitar PD, Hassler HB, Crestani C, Springer GH, Probert BM, et al. 2019. Population Gene Introgression and High Genome Plasticity for the Zoonotic Pathogen Streptococcus agalactiae. *Mol. Biol. Evol.* 36:2572–2590.

Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804.

Rosselló-Mora R. 2001. The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25:39–67.

Rousset F, Cabezas-Caballero J, Piastra-Facon F, Fernández-Rodríguez J, Clermont O, Denamur E, Rocha EPC, Bikard D. 2021. The impact of genetic diversity on gene essentiality within the Escherichia coli species. *Nat. Microbiol.* 6:301–312.

Santos Júnior JE, Williams PH, Dias CAR, Silveira FA, Faux P, Coimbra RTF, Campos DP, Santos FR. 2022. Biogeography and Diversification of Bumblebees (Hymenoptera: Apidae), with Emphasis on Neotropical Species. *Diversity* 14:238.

Seeley TD, Smith ML. 2015. Crowding honeybee colonies in apiaries can increase their vulnerability to the deadly ectoparasite Varroa destructor. *Apidologie* 46:716–727.

Shapiro BJ. 2016. How clonal are bacteria over time? *Curr. Opin. Microbiol.* 31:116–123.

Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ. 2012. Population Genomics of Early Events in the Ecological Differentiation of Bacteria. *Science* 336:48–51.

Shapiro BJ, Leducq J-B, Mallet J. 2016. What Is Speciation?Matic I, editor. *PLOS Genet.* 12:e1005860.

Shapiro BJ, Polz MF. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol.* 22:235–247.

Shapiro BJ, Polz MF. 2015. Microbial Speciation. *Cold Spring Harb. Perspect. Biol.* 7:a018143.

Shimodaira H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection.Goldman N, editor. *Syst. Biol.* 51:492–508.

Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480:241–244.

Smith EA, Anderson KE, Corby-Harris V, McFrederick QS, Parish AJ, Rice DW, Newton ILG. 2020. Reclassification of seven honey bee symbiont strains as *Bombella apis*. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.05.06.081802

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Steele MI, Kwong WK, Whiteley M, Moran NA. 2017. Diversification of Type VI Secretion System Toxins Reveals Ancient Antagonism among Bee Gut Microbes.Lindow SE, editor. *mBio* [Internet] 8. Available from: https://journals.asm.org/doi/10.1128/mBio.01630-17

Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB. 2013. Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota.von Mering C, editor. *PLoS Comput. Biol.* 9:e1003388.

Sun H, Li H, Zhang X, Liu Y, Chen H, Zheng L, Zhai Y, Zheng H. 2023. The honeybee gut resistome and its role in antibiotic resistance dissemination. *Integr. Zool.* 18:1014–1026.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, et al. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci.* 102:13950–13955.

Tettelin H, Riley D, Cattuto C, Medini D. 2008. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11:472–477.

Thomas CM, Nielsen KM. 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat. Rev. Microbiol.* 3:711–721.

Today E. 2019. Road Trip: How Hive Transportation Puts Stress on Honey Bees. *Entomol. Today* [Internet]. Available from: http://entomologytoday.org/2019/04/01/road-trip-hive-transportation-stress-honey-bees/

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, et al. 2009. Organised Genome Dynamics in the Escherichia coli Species Results in Highly Diverse Adaptive Paths.Casadesús J, editor. *PLoS Genet.* 5:e1000344.

Vachaspati P, Warnow T. 2015. ASTRID: Accurate Species TRees from Internode Distances.

Van Der Meer J, De Vos W, Harayama S, Zehnder AJB. 1992. Molecular Mechanisms of Genetic Adaptation to Xenobiotic Compounds. *ASM Microbiol. Rev.*:677–694.

Vlková M, Silander OK. 2022. Gene regulation in Escherichia coli is commonly selected for both high plasticity and low noise. *Nat. Ecol. Evol.* 6:1165–1179.

Wang X, Zhang X, Zhang Z, Lang H, Zheng H. 2018. Honey bee as a model organism to study gut microbiota and diseases. *Drug Discov. Today Dis. Models* 28:35–42.

Ward DM, Bateson MM, Ferris MJ, Kühl M, Wieland A, Koeppel A, Cohan FM. 2006. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos. Trans. R. Soc. B Biol. Sci.* 361:1997–2008.

Whitaker RJ, Grogan DW, Taylor JW. 2003. Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science* 301:976–978.

Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci.* 95:6578–6583.

Williams PH, Thorp RW, Richardson LL, Colla SR. 2014. Bumble Bees of North America: An Identification Guide. Princeton University Press

Wittouck S, Wuyts S, Meehan CJ, van Noort V, Lebeer S. 2019. A Genome-Based Species Taxonomy of the *Lactobacillus* Genus Complex.Gibbons SM, editor. *mSystems* [Internet] 4. Available from: https://journals.asm.org/doi/10.1128/mSystems.00264-19

Woese CR. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.

Woods LC, Gorrell RJ, Taylor F, Connallon T, Kwok T, McDonald MJ. 2020. Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *Proc. Natl. Acad. Sci.* 117:26868–26875.

Yasuda PH, Steigerwalt AG, Sulzer KR, Kaufmann AF, Rogers F, Brenner DJ. 1987. Deoxyribonucleic Acid Relatedness between Serogroups and Serovars in the Family Leptospiraceae with Proposals for Seven New Leptospira Species. *Int. J. Syst. Bacteriol.* 37:407–415.

Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, Ley RE, Farnleitner AH. 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat. Commun.* 10:2200.

Zhang Z, Guo Y, Yang F, Li J. 2022. Pan-Genome Analysis Reveals Functional Divergences in Gut-Restricted Gilliamella and Snodgrassella. *Bioengineering* 9:544.

Zhang Zi-Jing, Huang M, Qiu L, Song R, Zhang Zi-Xuan, Ding Y, Zhou X, Zhang X, Zheng H. 2021. Diversity and functional analysis of Chinese bumblebee gut microbiota reveal the metabolic niche and antibiotic resistance variation of *Gilliamella*. *Insect Sci.* 28:302–314.

Zheng H, Nishida A, Kwong WK, Koch H, Engel P, Steele MI, Moran NA. 2016. Metabolism of Toxic Sugars by Strains of the Bee Gut Symbiont Gilliamella apicola.Maleszka R, McFall-Ngai MJ, editors. *mBio* [Internet] 7. Available from: https://journals.asm.org/doi/10.1128/mBio.01326-16

Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, Tringe SG, Moran NA. 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc. Natl. Acad. Sci.* 116:25909–25916.

Zumkhawala-Cook A, Gallagher P, Raymann K. 2024. Diet affects reproductive development and microbiota composition in honey bees. *Anim. Microbiome* 6:64.