



Faculté de biologie et de médecine

Département de Microbiologie Fondamentale

Gut microbiome evolution and ecology in honeybees

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Aiswarya PRASAD

Master of Science, Indian Institute of Science, Bengaluru

Jury

Prof Tadeusz Kawecki, Président

Prof Philipp Engel, Directeur de thèse

Prof Sara Mitri, Experte

Prof Shinichi Sunagawa, Expert

Lausanne (2025)

"Equipped with his five senses, man explores the universe around him and calls the adventure Science."

– Edwin Hubble

ACKNOWLEDGEMENTS

I thank my supervisor, *Philipp*, for his consistent support and patient guidance over these five years. I am grateful for his unending energy and enthusiasm for science, which have been instrumental in my progress and scientific growth.

I thank my thesis committee – *Prof. Tadeusz Kawecki, Prof. Sara Mitri, and Prof. Shinichi Sunagawa* – for reading this thesis (and the mid-thesis report) and for their valuable feedback.

I am grateful to all our collaborators, especially *Sze Hue Yek*, whose efforts ensured we had honeybees to analyse. Special thanks to *Asha Devi* for all the challenges she braved, especially during the early field trips and to *John Chan*. I thank *Axel Brockmann* (for sharing inspiring literature that deepened my fascination with honeybees) and his lab members for hosting me in India during the intense weeks of sampling and sample preparation.

I am grateful to everyone at the Engel lab for the camaraderie and collaboration. I especially want to thank *Florent* for being a wonderful mentor, a perpetual source of encouragement and positivity, and for the deep dives (vortex) into concepts and interpretations of results. I will strive to embody his approach to mentorship in the future. Thanks also to *Garance* for sharing the majority of this journey with me; *Malick*, for deep discussions of concepts and scientific and methodological decisions; *Gonçalo*, for detailed feedback on experiment design and for guiding me through my first Syncrom assembly and bee experiment; *Alexandra* for her support in the lab; *Aurélian*, for always being kind and helpful in the lab; and all the other current and former members of the Engel lab for their contributions and support both in the lab/office and outside. I am grateful to my former lab members who shaped my early journey into this field and whose work has enabled mine. I am thankful to *German* for shaping my scientific approach and appreciation for bioinformatics (and for sharing the most insightful and entertaining stories and experiences), *Kirsten*, whose work laid the foundation for my work. Thanks to my department colleagues for fostering a supportive and scientifically rich environment. Special thanks to *Bevika* and *Roberto*, who began and shared this PhD journey with me.

I want to thank my IISc classmates, with whom I began my scientific journey and whose influence continues to shape my approach to science and life. I am especially thankful to *Jyotsna* for the co-working sessions from halfway across the globe, which helped me through the final leg of thesis writing. I am grateful to my family in Germany – *Katie, Holger, and Melanie* – for making me feel at home on this side of the globe. I thank my parents, *Amma* and *Appa*, for their infinite love, support, and belief in me. Most of all, I am deeply grateful to my husband, *Nicolas*, for being a constant source of love, joy, and unwavering support.

USE OF AI TOOLS

During the preparation of this work, the author(s) used AI tools, including DeepL, OpenAI's ChatGPT and Grammarly, to assist with the grammatical refinement of the papers. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication/manuscript.

SUMMARY

Animals often form long-term associations with microbial partners, and the gut is one of the most densely colonized and ecologically complex environments where such relationships unfold. Yet, the processes shaping gut microbial community composition and how they operate across evolutionary and ecological timescales are not entirely understood. A central challenge has been our limited ability, until recently, to survey and study the structure and function of complex microbial communities, particularly across host species. In this thesis, I investigate the patterns and processes underlying gut microbiota evolution and community assembly, leveraging recent developments in genome sequencing, bioinformatic approaches, and the tractable honeybee model system.

I examined the gut microbiomes of five honeybee species with shared evolutionary history and comparable ecology and behavior using high-throughput metagenomic analysis. Honeybee gut microbiomes were largely host-specific in structure, with functional consequences. Host-microbe phylogenies suggest a dynamic evolutionary history involving symbiont gains, losses, and host switching more than co-diversification. Further, even for very closely related bacteria, the host specificity of individual species and constituent strains was not always the same. These findings highlight the importance of moving beyond community-level patterns to investigate strain-level structure and function. We find that co-diversification, which has been reported in other animals, is less prominent in the honeybee gut microbiota. This study underscores the value of studying microbial evolutionary dynamics through high-throughput cross-species comparative analyses.

In a complementary line of inquiry, I examined ecological processes governing microbiome assembly over shorter timescales by manipulating synthetic microbial communities in microbiota-deprived honeybees. Focusing on priority effects at the strain level, I found that early-arriving strains frequently outcompeted later ones, helping explain why even similar individuals, such as nestmates acquiring microbiota from the same source, can harbor distinct communities at the strain level. Dropout experiments revealed that late-arriving strains could establish more effectively in the absence of competitors, especially of the same genus. This highlights an important role for priority effects through niche pre-emption and sets the stage to investigate interspecies interactions in the context of niche overlap. These results demonstrate that priority effects are a key ecological process shaping the honeybee gut microbiome at the strain level. Such ecological processes shaping community assembly early in life could play an essential role in maintaining microbial diversity and prevalent patterns such as host-specificity over evolutionary time.

This thesis advances our understanding of the eco-evolutionary processes shaping gut microbiomes by combining comparative and experimental approaches. Integrating evolutionary perspectives with mechanistic insights into microbial community assembly provides a foundation for future efforts to predict, manipulate, or engineer gut communities.

RÉSUMÉ

Les animaux forment souvent des associations à long terme avec des partenaires microbiens. L'intestin est l'un des environnements les plus densément colonisés et écologiquement complexes où de telles relations se développent. Pourtant, les processus qui déterminent la composition des communautés microbiennes intestinales et la manière dont ils opèrent à différentes échelles de temps évolutives et écologiques restent mal compris, en partie à cause de notre capacité limitée à recenser et à étudier la structure et la fonction de ces communautés complexes entre espèces hôtes. Dans cette thèse, j'étudie les modèles et les processus qui régissent l'évolution et l'assemblage du microbiote intestinal, grâce aux récentes avancées du séquençage génomique, de la bioinformatique et du modèle de l'abeille mellifère.

J'ai examiné les microbiomes intestinaux de cinq espèces d'abeilles ayant une histoire évolutive commune, une écologie et un comportement comparables, en utilisant une analyse métagénomique à haut débit. La structure de leurs microbiomes intestinaux était largement spécifique à l'hôte, avec des conséquences fonctionnelles. Les phylogénies hôte-microbe suggèrent une évolution dynamique, marquée par des acquisitions et pertes de symbiotes ainsi que par des changements d'hôte, davantage que par co-diversification. De plus, la spécificité d'hôte varie même entre souches bactériennes proches et montre l'importance d'aller au-delà des communautés pour étudier la structure et la fonction au niveau des souches. Nous constatons aussi que la co-diversification observée chez d'autres animaux est moins importante dans le microbiote intestinal de l'abeille. Cette étude souligne l'intérêt d'étudier la dynamique évolutive microbienne grâce à des analyses comparatives interspécifiques à haut débit.

En complément, j'ai examiné les processus écologiques régissant l'assemblage du microbiome sur des échelles de temps plus courtes, en manipulant des communautés microbiennes synthétiques chez des abeilles mellifères privées de microbiote. En me concentrant sur les effets de priorité au niveau des souches, j'ai constaté que les souches arrivant en premier surpassaient fréquemment celles introduites plus tard. Cela contribue à expliquer pourquoi des congénères d'une même colonie acquérant leur microbiote de la même source peuvent abriter des souches distinctes.

Les expériences d'abandon ont révélé que l'établissement des souches tardives est facilité en l'absence de compétiteurs, en particulier du même genre. Cela met en avant l'importance des effets de priorité par la préemption de niche et ouvre la voie à l'étude des interactions inter-espèces lors du chevauchement des niches. Ces résultats démontrent que les effets de priorité sont un processus écologique clé qui façonne le microbiome intestinal de l'abeille au niveau des souches. De tels processus écologiques façonnant l'assemblage de la communauté en début de vie pourraient jouer un rôle important dans le maintien de la diversité microbienne et de la spécificité de l'hôte au cours de l'évolution.

Cette thèse améliore notre compréhension des processus éco-évolutifs qui façonnent les microbiomes intestinaux en combinant des approches comparatives et expérimentales. L'intégration de perspectives évolutives et de connaissances mécanistiques sur l'assemblage des communautés microbiennes fournit une base pour les efforts futurs visant à prédire, manipuler ou concevoir des communautés intestinales.

SUMMARY FOR THE GENERAL PUBLIC

Our guts are home to trillions of microbes influencing our health and well-being. The same is true for honeybees. But how do these tiny microbes assemble in our guts into microbial communities or microbiomes? How do microbiomes change over time and evolve with their hosts? In this thesis, I studied the gut microbiomes of different honeybee species to understand how these communities form and what shapes them.

First, I looked at gut bacteria in honeybee species that evolved separately from a common ancestor millions of years ago. By comparing how their gut microbiomes look today, we can understand what differences must have been accrued through the million years of their evolution. I found that honeybee species tended to have their unique microbial community. Not only that, the set of functions that the community performed based on the genes they encode was also often specific to each host. However, the relationships between bees and their microbes were not always one-to-one. Some bacteria were shared between species, while others had been gained by a honeybee species or lost over time. This means gut microbes evolve in parallel with their host, but not strictly so. Instead, gut microbiota must have had opportunities to be gained or exchanged between host species lineages over evolutionary time.

Next, I explored how these communities assemble over short timescales. I worked with bees reared without any gut bacteria and introduced specific gut microbes. I wanted to know whether the order in which bacteria arrived would change the final community composition. We found that the bacteria that arrived first often took over the gut. This “first come, first served” effect, called priority effects, helps explain why gut communities can be different even in similar honeybees because the order of arrival of various bacterial strains is a matter of chance.

My research shows that gut microbes are shaped both by long-term evolution and by short-term, chance-based events. Understanding these patterns brings us closer to predicting and even designing healthy microbiomes, not just for bees. People often ask why I study honeybees and not human gut microbes. My favorite answer is that bees and humans are very different, but the microbes in our guts and the processes that shape them are remarkably similar!

RÉSUMÉ POUR LE GRAND PUBLIC

Nos intestins abritent des billions de microbes qui influencent notre santé et notre bien-être. Il en va de même pour les abeilles. Mais comment ces minuscules microbes s'assemblent-ils dans nos intestins pour former des communautés microbiennes ou des microbiomes ? Comment les microbiomes changent-ils au fil du temps et évoluent-ils avec leurs hôtes ? Dans cette thèse, j'ai étudié les microbiomes intestinaux de différentes espèces d'abeilles afin de comprendre comment ces communautés se forment et ce qui les façonne.

Tout d'abord, j'ai étudié les bactéries intestinales d'espèces d'abeilles qui ont évolué séparément à partir d'un ancêtre commun il y a des millions d'années. En comparant l'aspect actuel de leurs microbiomes intestinaux, nous pouvons comprendre les différences qui se sont accumulées au cours des millions d'années de leur évolution. J'ai découvert que les espèces d'abeilles mellifères avaient tendance à avoir leur propre communauté microbienne. De plus, l'ensemble des fonctions que la communauté remplit, sur la base des gènes qu'elle code, est souvent spécifique à chaque hôte. Cependant, les relations entre les abeilles et leurs microbes ne sont pas toujours exclusives. Certaines bactéries étaient partagées entre les espèces, tandis que d'autres ont été gagnées par une espèce d'abeille ou perdues au fil du temps. Cela signifie que les microbes intestinaux évoluent parallèlement à leur hôte, mais pas de manière stricte. Au contraire, le microbiote intestinal doit avoir eu l'occasion d'être acquis ou échangé entre les lignées d'espèces hôtes au cours de l'évolution.

Ensuite, j'ai étudié la manière dont ces communautés s'assemblent sur de courtes périodes. J'ai travaillé avec des abeilles élevées sans bactéries intestinales, dans lesquelles j'ai introduit des microbes intestinaux spécifiques. Je voulais savoir si l'ordre d'arrivée des bactéries modifierait la composition finale de la communauté. J'ai constaté que les bactéries arrivées en premier prenaient souvent le contrôle de l'intestin. Cet effet « premier arrivé, premier servi », appelé effet de priorité, permet d'expliquer pourquoi les communautés intestinales peuvent être différentes même chez des abeilles semblables, car l'ordre d'arrivée des différentes souches bactériennes relève du hasard.

Mes recherches montrent que les microbes intestinaux sont façonnés à la fois par l'évolution à long terme et par des événements aléatoires à court terme. La compréhension de ces modèles nous rapproche de la prédiction, voire de la conception de microbiomes sains, et pas seulement pour les abeilles. On me demande souvent pourquoi j'étudie les abeilles et non les microbes intestinaux humains. Ma réponse préférée est que les abeilles et les humains sont très différents, mais que les microbes présents dans nos intestins et les processus qui les façonnent sont remarquablement similaires !

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	IV
SUMMARY	V
RÉSUMÉ.....	VI
SUMMARY FOR THE GENERAL PUBLIC	VII
RÉSUMÉ POUR LE GRAND PUBLIC	VIII
TABLE OF CONTENTS.....	IX
CHAPTER-1.....	12
INTRODUCTION	12
1 PATTERNS IN GUT MICROBIAL COMMUNITIES	15
1.1 COMMUNITY LEVEL PATTERNS WITHIN INDIVIDUALS.....	15
1.2 COMMUNITY LEVEL PATTERNS ACROSS INDIVIDUALS	16
1.3 COMMUNITY LEVEL PATTERNS ACROSS SPECIES	17
1.4 BEYOND COMMUNITY LEVEL PATTERNS	18
2 PROCESSES SHAPING GUT MICROBIAL COMMUNITIES.....	20
2.1 ECO-EVOLUTIONARY PROCESSES SHAPING GUT MICROBIAL COMMUNITIES	20
2.2 EVOLUTIONARY PROCESSES AND THEIR CONSEQUENCES	24
3 THE HONEYBEE GUT MICROBIAL COMMUNITY	28
3.1 HONEYBEE EVOLUTION AND ECOLOGY.....	28
3.2 PATTERNS AND PROCESSES IN THE HONEYBEE GUT MICROBIOME	32
4 THESIS OUTLINE.....	36
REFERENCES.....	38
CHAPTER-2.....	45
SYMBIOT LOSS AND GAIN, RATHER THAN CO-DIVERSIFICATION, SHAPES HONEYBEE GUT MICROBIOTA DIVERSITY AND FUNCTION	45
SUMMARY.....	46
CONTRIBUTION	46
PUBLICATION STAGE.....	46
INTRODUCTION.....	47
RESULTS	50
AN EXPANDED CATALOG OF BACTERIAL GENOMES FROM THE GUT OF FIVE HONEYBEE SPECIES	50
MOST HONEYBEE SPECIES HARBOR HOST-SPECIFIC MICROBIAL COMMUNITIES IN THEIR GUT	51
HOST SPECIFICITY OF INDIVIDUAL COMMUNITY MEMBERS VARIES ACROSS BACTERIA AND HOST SPECIES	52
HOST SWITCHES AND SPECIES GAINS AND LOSSES INTERRUPTED THE HOST-SPECIFIC EVOLUTION OF GUT MICROBIOTA MEMBERS OF HONEYBEES	54
FUNCTIONAL DIFFERENCES IN THE GUT MICROBIOTA ACROSS HONEYBEES	56
DISCUSSION.....	59
METHODS.....	62
SAMPLING AND STORAGE.....	62
DISSECTION AND DNA EXTRACTION.....	62

METAGENOMIC SEQUENCING	62
MAPPING TO DATABASES	63
ASSEMBLY AND BINNING OF MAGS.....	63
TAXONOMIC PROFILING	64
HOST SPECIFICITY AND STRAIN-LEVEL PROFILES	65
PHYLOGENY CONSTRUCTION AND CO-DIVERSIFICATION TEST	65
GENE FUNCTIONAL ANNOTATION AND PROFILING	67
STATISTICAL TESTS AND BIOINFORMATIC ANALYSES.....	67
ACKNOWLEDGMENTS	68
AUTHOR CONTRIBUTIONS	68
COMPETING INTERESTS	68
DATA AVAILABILITY	69
CODE AVAILABILITY.....	69
SUPPLEMENTARY MATERIAL.....	70
SUPPLEMENTARY FIGURES	70
SUPPLEMENTARY TABLES.....	99
REFERENCES	100
CHAPTER-3.....	107
PRIORITY EFFECTS AT THE STRAIN LEVEL IN HONEYBEE GUT MICROBIOTA.....	107
SUMMARY.....	108
CONTRIBUTION	108
PUBLICATION STAGE.....	108
INTRODUCTION.....	109
RESULTS	111
COMMUNITY SELECTION, EXPERIMENTAL DESIGN, AND VALIDATION APPROACH	111
PRIORITY EFFECTS DETERMINE COMMUNITY COMPOSITION AT THE STRAIN LEVEL.....	112
PRIORITY EFFECTS ARE CONSISTENTLY OBSERVED FOR ALL STRAINS, BUT VARY IN STRENGTH	113
CONSPECIFICS CONTRIBUTE TO PRIORITY EFFECTS IN ADDITION TO OTHER CLOSE RELATIVES.....	116
DISCUSSION.....	119
METHODS.....	122
STRAIN SELECTION AND SEQUENCE DATABASE ESTABLISHMENT	122
VALIDATION OF THE CUSTOM DATABASE AND STRAIN DETECTION APPROACH.....	123
SYNTHETIC COMMUNITY ASSEMBLY	123
EXPERIMENT SETUP AND PROCEDURE	124
VALIDATION OF EXPERIMENTAL PARAMETERS (PILOT EXPERIMENT)	124
DNA EXTRACTION	124
LIBRARY PREPARATION AND AMPLICON SEQUENCING	125
QUANTIFICATION OF TOTAL BACTERIAL ABUNDANCE	125
PREPROCESSING OF SEQUENCING DATA	126
ESTIMATION OF STRAIN ABSOLUTE ABUNDANCE	126
ACKNOWLEDGMENTS	127
AUTHOR CONTRIBUTIONS	127
COMPETING INTERESTS	127
DATA AVAILABILITY	127
CODE AVAILABILITY.....	127
SUPPLEMENTARY MATERIAL.....	128
SUPPLEMENTARY FIGURES	128
SUPPLEMENTARY TABLES	142
REFERENCES	143

CHAPTER-4.....	146
GENERAL DISCUSSION	146
REASSESSING CO-DIVERSIFICATION.....	146
DRIVERS AND CONSEQUENCES OF HOST SPECIFICITY	147
PROCESSES SHAPING GUT MICROBIAL COMMUNITY ASSEMBLY.....	149
THE INTERDISCIPLINARY NATURE OF GUT MICROBIOME RESEARCH	149
REFERENCES.....	151
CURRICULUM VITAE.....	153

CHAPTER-1

"The cell is basically a historical document, and gaining the capacity to read it (by the sequencing of genes) cannot but drastically alter the way we look at all of biology."

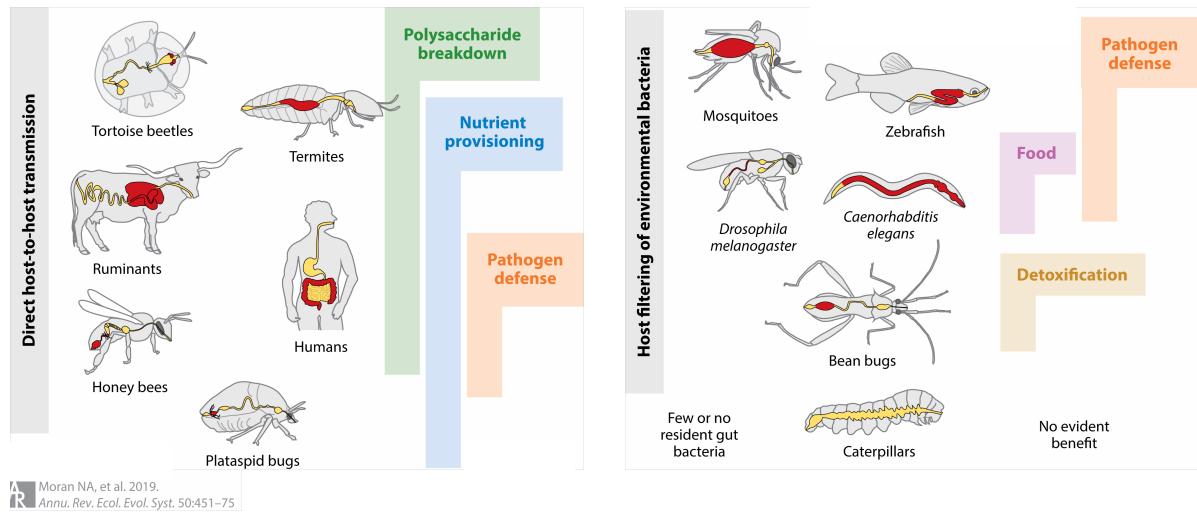
– Carl Woese, Bacterial Evolution
(Woese, 1987)

INTRODUCTION

Over the centuries since humanity's first peek into the microbial world through the lens of Anton van Leeuwenhoek's microscope about four centuries ago, our perspective of the microscopic world of bacteria has grown immensely in depth and breadth. Today, we no longer need to "see to believe" that there are bacteria on our skin surface or all over our gut lining. Without a single "look" we can tell in high resolution which bacteria are present and even what they are doing. With this immense power to study microbes, we have witnessed a profound shift from viewing bacteria as rogue agents of disease to recognizing them as versatile constituents of microbial communities, significantly shaping their environment and even as essential partners in symbioses (Drew et al., 2021).

This shift in understanding is evidenced by the rapid surge of interest and excitement over the past two decades about the gut microbiome, where bacteria play an important role in the health and behavior of their host (Cani, 2018; Engel & Moran, 2013; Shreiner et al., 2015). Thanks to advances in our understanding and modern molecular and computational technology, we can now study bacteria on a scale orders of magnitude greater than just a decade ago (Van Rossum et al., 2020). We are now uniquely positioned to explore these complex communities and understand their structure and function. With a detailed understanding of these patterns of bacterial diversity across communities, we can uncover the ecological and evolutionary processes that shape them (Mazel et al., 2023). This, in turn, can ultimately enable us to understand, predict and control microbiomes and how they interact with their host (Vrancken et al., 2019).

Though all animals have evolved in a "bacterial world," (McFall-Ngai et al., 2013) "Not all animals need a microbiome." (Hammer et al., 2019) Animals have evolved a myriad of strategies over millions of years to prevent the growth of bacteria within, yet several animals are associated with bacteria, albeit to varying degrees (Hammer et al., 2019; Lim & Bordenstein, 2020). This can range from animals that cannot survive without their bacterial associates or derive key benefits from them to those that tolerate random microbes from their environment to a limited degree (**Fig. 1**). Gut microbial communities of several animals across the kingdom are consistently found associated with their host species (Mallott & Amato, 2021), influential in their health and behavior but not typically essential for their survival. Exactly how prevalent are such host-microbiome associations, and how are they formed and maintained? To this end, it is important to study how bacteria are distributed across host species and how they assemble into communities to understand how such specific gut microbial communities can emerge and be maintained over ecological and evolutionary timescales.



A

Moran NA, et al. 2019.
Annu. Rev. Ecol. Evol. Syst. 50:451–75

Figure 1. Sources of acquisition and significance of host-microbiome relationship. Animals acquire their gut (marked in yellow) microbiomes from different sources and by different means. Some animals host a high abundance of bacteria in particular regions of their gut (marked in red). These animals are represented as examples of various scenarios observed in nature, ranging from those with no substantial gut microbiota to those that acquire microbiota through host filtering from the environment or directly from conspecifics, some even hosting bacteria in specialized organs (e.g., Tortoise beetles). Gut microbial communities may provide a range of benefits to their host. Adapted from: (Moran et al., 2019).

Advances in sequencing technologies have led to a large upsurge of datasets on gut microbial community composition. Much of the data so far has come from studies using 16S rRNA amplicon sequencing, which is limited in resolution. Further, it provides little information about the functional potential of the microbiota, which is essential to understanding the nature of the host microbiome relationship. Shotgun metagenomics, especially when combined with a genome-resolved approach, can provide unprecedented insight into the diversity and function of even previously uncharacterized microbiomes (Levin et al., 2021). To uncover eco-evolutionary processes, it is essential to compare several host species whose ecology and evolutionary history are well known. Despite the large number of studies on microbiomes, there are few comparative studies across host species (Fig. 2) (Degregori et al., 2025) and even fewer using shotgun metagenomics.

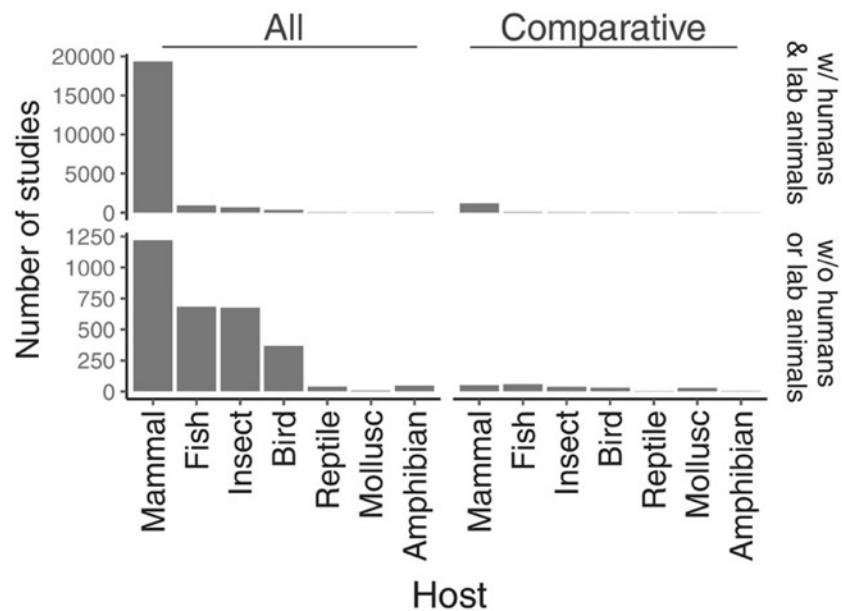


Figure 2. Summary of gut microbiome studies across host species. Published journal articles on gut microbiome collected as of September 2023 using the Scopus document search tool. The results include studies focusing on a single host species (All) or comparing several host species (Comparative). Adapted from: (Degregori et al., 2025).

Comparative studies of natural microbiomes by high-resolution characterization are needed to reveal community-level patterns and, more importantly, the underlying distribution of individual microbes to understand how these communities might be assembled and maintained. Another powerful approach to further understanding the processes behind those patterns is studying the microbiota in simple synthetic communities in a controlled setting. In this thesis, I address these questions by leveraging both approaches using the honeybee gut microbiome model.

The following sections of this general introduction present an overview of the patterns in gut microbiome diversity maintained across evolutionary timescales revealed by comparisons across host species and the underlying eco-evolutionary processes involved in their assembly and maintenance, concluded by an overview of the study system, the honeybee gut microbiome and previous work on these topics.

1 PATTERNS IN GUT MICROBIAL COMMUNITIES

The key questions about gut microbial community patterns are: (1) How does gut microbiome composition look at different taxonomic levels? (2) How does gut microbiome composition look across similar individuals? (3) How does gut microbiome composition vary across distinct individuals, e.g., different host species? Such patterns at the community level are a reflection of the distribution of individual taxa. Understanding community-level patterns across distinct individuals and the underlying distribution of individual taxa at high taxonomic resolution, such as at the strain level, is essential to uncover the evolutionary and ecological processes shaping gut microbial community structure and, consequently, function.

1.1 COMMUNITY LEVEL PATTERNS WITHIN INDIVIDUALS

Gut microbial communities are diverse across taxonomic levels (Fig. 3). The level at which we characterize microbial communities can affect our conclusions, especially in comparative studies. For example, the genus *Lactobacillus* may be found in several environments, ranging from fermented food to the honeybee and the human gut. However, at the species level, those microbes found in each environment are divergent clades often found explicitly in their respective environments and can encode several different functions (Ellegaard et al., 2019). Hence, a fundamental challenge in characterizing the diversity of microbial communities is defining a unit of diversity. This endeavor faces several challenges, beginning with the debate on whether bacterial species exist (Doolittle, 2012). This question has been revisited by several studies using modern sequencing methodologies and massive datasets (Jain et al., 2018; Olm et al., 2020; Varghese et al., 2015; Viver et al., 2024), identifying approaches to demarcate bacterial species.

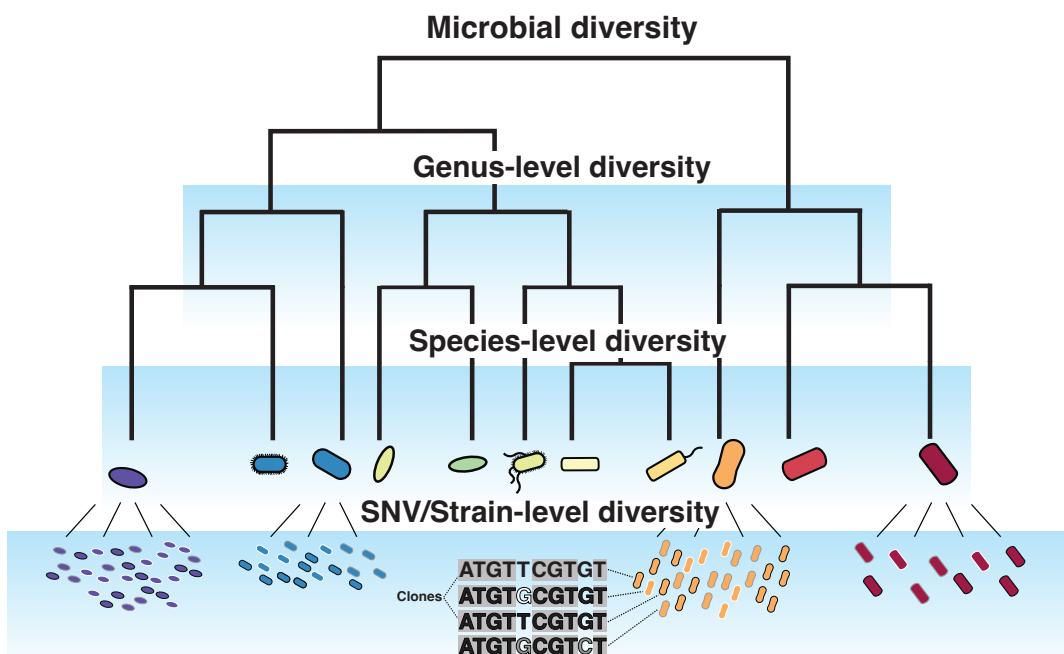


Figure 3: Diversity across taxonomic levels in a microbial community. Microbial communities typically consist of bacteria from several taxa within which more diversity is found. The most basic unit is a single cell, which is accompanied by several clones as well as other strains of the same species, comprising several SNVs (single-

nucleotide variants) of various genes. These strains collectively constitute the population corresponding to each species in the community. The number of taxonomic groups increases as the resolution increases (genera > species > set of SNVs).

Regardless of whether we can agree on how to define a bacterial species or the lack thereof, our immediate need is to identify a unit of diversity representative of species or strains, enabling the meaningful characterization and comparison of communities. A practical unit widely used today and in this thesis is ASVs (amplicon sequence variants), which are unique sequence variants of a highly conserved genomic region, such as the 16S rRNA gene, and can represent strains within a bacterial species when considering the full-length of the gene (this unit is used and further explained in Chapter 3). Another unit is ANI-based (average nucleotide identity) clusters of genomes with 95% or higher genomic identity, which correspond well to bacterial species (this unit, referred to as magOTUs, is used and further explained in Chapter 2). Another essential but often ignored consideration when characterizing microbial communities is the absolute abundance of bacteria in the community, measured by methods such as qPCR. This can differentiate resident microbiota from transients and contaminants and reveal ecological and evolutionary patterns that would be difficult to delineate with relative abundances alone (Hammer et al., 2019).

1.2 COMMUNITY LEVEL PATTERNS ACROSS INDIVIDUALS

No two gut microbial communities can be expected to be identical in composition. Even within the same individual, the gut microbiome can be highly dynamic due to factors such as diet or disturbances due to pathogen infection (Gibbons et al., 2017). However, individual variations over time are minor compared to differences between individuals (Schloissnig et al., 2013; Faith et al., 2013). Despite the highly dynamic nature of gut microbiomes, the most abundant members of individual microbiomes remain stably associated over decades (Faith et al., 2013). A general pattern in gut microbial community structure is that there are several highly abundant members but many more at low abundance (Nemergut et al., 2013). Usually, the most abundant members are also the most prevalent (found across many similar samples) and are often referred to as “core” members. They are typically the focus of analyses and comparisons, as they are the most consistently associated with their respective hosts and likely to be the most relevant and functionally important to the gut ecosystem. Besides, rare members, while they could still play a role, are more likely to be transiently seen in specific individuals or even false positives. While “core microbiome” is a practical term for such salient members, it is crucial to consider the caveats in relying on its definition (Custer et al., 2023; Neu et al., 2021; Shade & Handelsman, 2012).

To describe diversity within a sample (α -diversity) or an individual, taxa or species richness is a simple and commonly used measure. It refers to the number of taxa in the community. Several animals that “have a microbiome” or consistently harbor a microbial community (Hammer et al., 2019) contain a stable number of abundant taxa. In some cases, this pattern may be consistent with a species-area relationship, which postulates that larger habitats harbor more diversity (Connor & McCoy, 1979). In the case of gut microbiomes, if and how this principle may be applied has been explored (Ellegaard et al., 2020; Kwong et al., 2017). For example, in the case of social insects, habitat size might refer to gut size or colony size, and the diversity supported might be considered as the number of species in the community or the number of strains of each species.

A pair of communities might contain several species of the same genus with similar metabolic input and output, or different genera with very different functions. These differences in community composition between samples (β -diversity) can have significant functional consequences and address key questions about the variation in microbiome diversity across individuals. β -diversity can be captured by several metrics that consider presence-absence, relative abundances of taxa, or even their phylogenetic relatedness (Lozupone & Knight, 2008), which is primarily correlated with functional potential. Although taxonomic structure can be a proxy for functional potential, looking directly at gene functions is essential to truly capture functional differences.

1.3 COMMUNITY LEVEL PATTERNS ACROSS SPECIES

To understand how patterns in gut microbiomes have risen and been maintained over evolutionary time, we must compare communities from related host species. Such comparative analyses, though limited to some well-known systems (Degregori et al., 2025), have revealed a pattern known as phylosymbiosis (Lim & Bordenstein, 2020; Mallott & Amato, 2021) whereby more closely related host species harbour more similar communities. Phylosymbiosis has been observed across several animals, especially vertebrates (Mallott & Amato, 2021), with a stronger signal within some clades such as ungulates (Song et al., 2020). Such a pattern can emerge when the members of the microbiome all diversify in parallel with their host (Fig. 4, left panel), but such a pattern does not necessarily imply that these have been such a parallel evolution (see Section 2.2). Phylosymbiosis could also result from more similar hosts filtering more of the same microbes (*i.e.*, selection or host-filtering) from their environment (Fig. 4, right panel). Since phylogenetically similar hosts usually exhibit more similar traits and consequently present more similar environments to colonize for microbes, this could also result in a correlation between microbiome similarity and host phylogeny (Mazel et al., 2018). Especially across highly divergent hosts, we can expect this to be important given the drastic differences in host physiology, for example, between humans, deer and bats. To understand the patterns that arise from evolutionary processes, we must ask what patterns occur among closely related hosts with less drastic differences in physiology and host ecology. Hence, comparisons across host species with shared evolutionary history and similar ecological and physiological conditions are essential. More diverse sampling is needed, across various systems, to better understand the extent and strength of these patterns across animals (Degregori et al., 2025).

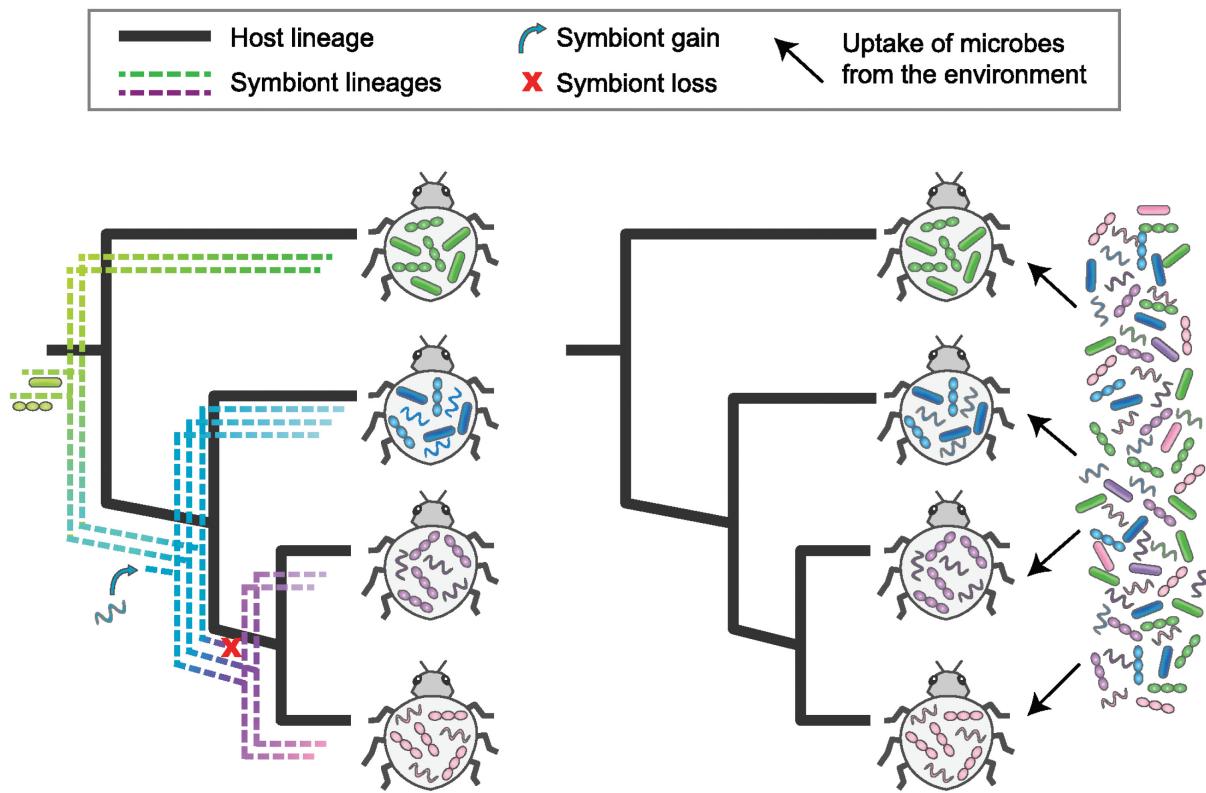


Figure 4. Alternative processes that could underlie the pattern of phylosymbiosis. (A) Each community members lineage is parallel to host lineage with some gain and loss over evolutionary time arising from evolutionary processes such as co-diversification or co-evolution (B) Host species have diverged in their selectivity of microbes from their environment for example due to differences in host physiology regardless of the evolutionary history of the microbes maintained by the ecological processes of selection or host-filtering. These evolutionary and ecological processes are further explained in the section outlining processes shaping gut microbial communities. Adapted from: (Moran & Sloan, 2015)

1.4 BEYOND COMMUNITY LEVEL PATTERNS

Community level patterns are a reflection of the underlying distribution of individual members of the community. To understand the processes underlying those patterns, it is imperative that we delineate patterns at the level of individual microbes. Some microbes may be highly prevalent (among all individuals) across several environments, while others might be restricted to specific environments. For example, a versatile microbe able to tolerate a high pH range may be found in a broader range of environments than one that can grow in a more limited pH range. Similarly, across host species, some microbes might be prevalent in several hosts (generalist) while others might be restricted to one host species (specialist or host-specific). As seen, for example, in rodents, (Mazel et al., 2024; Raulo et al., 2024) this might be explained, albeit only in part, by some microbes being better at moving around due to phenotypes such as aerotolerance and spore-forming ability. Specialists might rely on specific metabolites provided only by their host species, such as in the Rhizobium-legume symbiosis (Sharma et al., 1993) in some well-known cases. The extent to which gut microbiota are host-specific across different host-microbiota systems and the mechanisms underlying host-specificity are intriguing and warrant further exploration (Mazel et al., 2025).

Some specialist clades exhibit a pattern referred to as co-phylogeny, characterized by the congruence of the host and bacterial phylogenies (Groussin et al., 2020). When co-phylogeny is observed, closer relatives also host more closely related bacteria from their clade. It is essential to note the distinction between this pattern and phylosymbiosis. While phylosymbiosis describes a community-level pattern where more closely related hosts harbor more similar communities, it does not indicate that the community members are also more closely related. Co-phylogeny may be quantified by reconciling tree topologies of the host and microbes or by correlating the phylogenetic distances of hosts and their respective microbes (Groussin et al., 2020). This pattern is well-described in host-microbe systems that are intimately linked and are likely under strong selection, such as pathogens (de Vienne et al., 2013) and the nutritional endosymbiosis of aphids and *Buchnera* (Moran, 2001); however, it is unclear how prevalent this pattern is in gut microbes, which are much less intimately associated with their hosts. Some gut microbes of primates display a strong pattern of co-phylogeny (Moeller et al., 2016; Suzuki et al., 2021), but the rest of the members of their microbiome do not. Broad sampling and comparison of microbes from various host-microbiota systems is essential to understand the prevalence and importance of these patterns. The implications of co-phylogeny on the processes responsible for shaping the microbiome are discussed in the following section.

2 PROCESSES SHAPING GUT MICROBIAL COMMUNITIES

In this section, I introduce the concepts underlying the processes that shape microbiomes and how they can be elucidated from patterns observed in nature, then further explored in controlled experiments. First, I discuss the eco-evolutionary processes that shape microbial communities, their role in shaping community assembly and how we study evolutionary processes that operate over millions of years and can explain why gut microbiomes appear as they do across host species.

A vast body of work in macroecology has been built over the past several decades. However, applying macroecological concepts to microbial ecology comes with certain unique challenges (Nemergut et al., 2013) and promises (Orr et al., 2025). For example, microbes are much smaller than animals and consequently show very different dispersal abilities; they have much shorter doubling times and can drastically alter their metabolic function in response to environmental cues (dormancy). In addition to these biological differences, our methods to survey microbial community structure and diversity, as well as their limitations and biases, are very different from those for studying populations of large animals. On the other hand, microbial systems offer distinct advantages for studying eco-evolutionary processes. For example, their short generation times and rapid dynamics allow for direct observation of phenomena that would take much longer to observe in plant or animal systems. Additionally, the ability to precisely manipulate microbial populations can provide deeper insights into the mechanisms shaping species interactions. It is essential to carefully conceptualize and develop frameworks informed by ecological theory that can be applied to microbial communities.

2.1 ECO-EVOLUTIONARY PROCESSES SHAPING GUT MICROBIAL COMMUNITIES

Eco-evolutionary processes are reflected in the distribution of individual microbiota and the evolutionary relationship between them and their hosts. These processes influence which microbes colonize a host and how and whether they persist and interact in the gut microbial community. Subsequently, over a span of millions of years, these associations are maintained or lost as the microbes and their host evolve.

Over the past decades, ecologists have strived to outline the mechanisms shaping community diversity. By the end of the 20th century, Palmer cataloged no fewer than 120 different hypotheses attempting to account for variation in species richness (Palmer, 1994). This overwhelming range of theories led Vellend to propose a unified framework distilling these concepts into four fundamental processes: speciation, drift, selection and dispersal (Vellend & Agrawal, 2010). The excellence of this framework lies in its integration of niche-based and neutral theories, despite their seemingly opposing stands (Zhou & Ning, 2017). Niche-based theory attributes biodiversity patterns to deterministic factors such as adaptation to environmental conditions (niches). Neutral theory (Bell, 2001; Hubbell, 2001) explains patterns of biodiversity using stochastic processes like random birth, death, colonization, and extinction, assuming functional equivalence among species (neutralism), without invoking adaptation (Wennekes et al., 2012; J. Whitfield, 2002). Although these views were once considered mutually exclusive, recent perspectives acknowledge that both deterministic (niche-based) and stochastic (neutral) processes simultaneously shape

ecological communities. Hence, the framework integrates the inevitable evolutionary changes that microbes undergo (speciation) alongside ecological processes such as dispersal and drift. Vellend's synthesis has provided a strong starting point for building a unified conceptual framework for microbial community assembly (Nemergut et al., 2013). The fundamental four forces operate simultaneously, and their relative importance across various contexts is yet to be entirely understood.

Here, I present a brief overview of Vellend's framework's four processes, which have been thoroughly reviewed before (Nemergut et al., 2013; Zhou & Ning, 2017) and how they can be applied to gut microbiomes (Kohl, 2020) within and across host species. Speciation, in the case of microbial communities, is better framed as **diversification** or the generation of new genetic variation. Differences in community structure and function can arise from genetic changes that do not necessarily result in new species, and within-species diversity is highly relevant for the structure and function of microbial communities (Van Rossum et al., 2020). Diversification is an essential process shaping microbial communities not only over shorter periods but also through millions of years of evolution of their respective hosts.

While diversification could result in adding new species, this is more often a result of the following process. **Dispersal** refers to the movement of microbes across space and, in the case of gut microbiota, within and outside the host gut or between individuals and across host species. It is often discussed in the context of "dispersal limitation", which is particularly relevant for gut microbial communities, as several physical and immune bottlenecks can limit host gut colonization. The host can also affect microbiota dispersal ability; for example, vertically transmitted microbiota may have fewer opportunities for dispersal than horizontally or environmentally acquired ones. This invites reconsideration of the applicability (to gut microbiomes) of the widely invoked Bass-Becking hypothesis (Becking, 1931; De Wit & Bouvier, 2006), "everything is everywhere, but the environment selects". In contrast to more open ecosystems such as soil or ocean microbiomes, dispersal may be more prominent in shaping gut microbial communities. That said, I turn to selection, which remains an essential key driver across ecosystems.

Stochastic changes in the relative abundance of different community members are referred to as ecological **drift**. Random births, deaths, and colonization events can lead to divergence in community structure even in identical environments. Theory and empirical data indicate that drift is more important when community diversity is low, and selection is weak. In gut microbial communities, drift is likely to be most relevant during early stages of colonization or following perturbations such as antibiotic treatments.

Selection refers to changes in community structure caused by deterministic fitness differences between community members. In gut microbiomes, selection can be seen in the influence of host diet (Brochet et al., 2021; David et al., 2014) and, even more so, immune response (Martino et al., 2022). Factors underlying selection are often grouped as abiotic (environmental conditions such as temperature, pH etc.) and biotic (interactions with other community members). In the context of gut microbiota, the host gut is the environment. Interactions with the host can be viewed in the same context as interactions with the environment, typically referred to as abiotic interactions. Across host species, physiological or ecological differences could result in the selection of different microbes best adapted to

the gut conditions of each host in a process akin to selection called host-filtering. Biotic factors can be viewed more closely through the framework of co-existence theory as elaborated subsequently.

Gut microbes are usually acquired early in life. In mammals, newborns acquire their microbiota from their mother upon birth (Martino et al., 2022), and in bees, pupae acquire their microbiota from nestmates and hive material upon emergence as adults (Powell et al., 2014). Hence, the eco-ecological processes involved in community assembly are also essential drivers of microbiome diversity within and across host species. Deterministic processes can explain how divergent individuals or distinct environments (e.g., different host species) select for different microbiota even when drawing from a shared species pool (**Fig. 5**, top panel). For example, phylosymbiosis can arise from ecological processes such as host-filtering (or more generally, selection) even in the absence of a consistent shared evolutionary history between the hosts and their microbiota (**Fig. 4**, right panel). Further, dispersal limitation can result in few opportunities for microbes of one host species to reach the gut of another, especially given that individuals of the same host species interact and exchange microbes frequently. On the other hand, when a microbe does get dispersed from one host species to another, crossing all barriers to colonization, over an extended period, diversification and adaptation to the new host could result in the hosts having two different microbial species that selection ensures can now only exist in their respective hosts. Within the same host species, substantial variation exists as well, even though (e.g., individuals within the same host species), selective pressures are expected to be similar (Zhou & Ning, 2017). A combination of deterministic and stochastic processes (drift) may explain such variation. For example, diversification may lead to new variation within individuals, and drift can lead to the extinction of some species. Diversification can shape microbiomes in the short but also, in the long term. The latter is reflected in the variation of gut microbiome diversity across host species. This is an area of key focus in this thesis (*Chapter 2*) and further explained in *Section 2.2*. Hence, the four fundamental processes act in combination to shape community diversity, although their relative importance is unclear, and each of their effects is challenging to disentangle in practice (Nemergut et al., 2013).

Within-host diversification has been observed in gut microbial communities, but strains that must have diverged much earlier than when the community was assembled are also found (Wolff et al., 2021). Some strains are also shared across individuals far away from each other, indicating that they may not have stochastically emerged in each individual (Schloissnig et al., 2013). Incorporating the temporal scale can provide an alternate scenario of community assembly where individuals offering the same selective environment can end up with different microbial communities contingent on the history of arrival of their members (**Fig. 5**, bottom panel). Here, dispersal history can modulate how the other processes influence community structure, further entangling the processes.

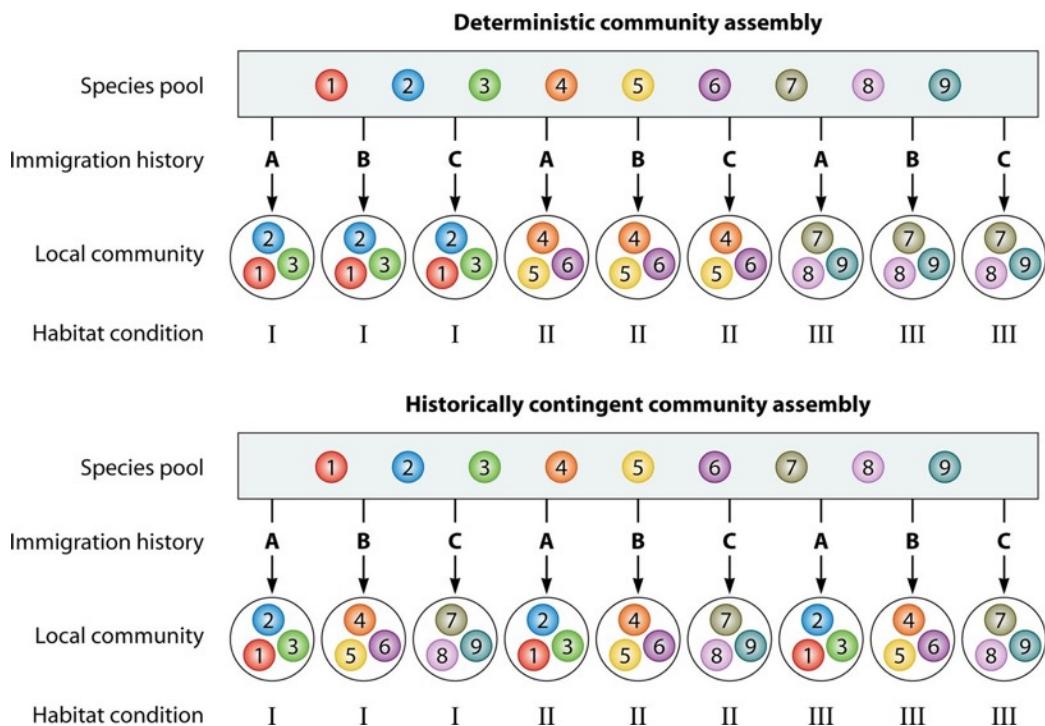


Figure 5. Alternative scenarios of community assembly. Colored circles represent hypothetical bacterial species (or strains), and arrows represent species immigration, with the letter indicating different arrival orders. Circles with black outlines represent different individual hosts, and the Roman numerals below indicate different host species. In the top panel, each host species harbors a specific set of bacteria (selection or host-filtering) regardless of immigration order. In the bottom panel, immigration order dictates species composition (*i.e.*, priority effects), leading to divergent communities even between individuals with similar conditions for selection. Adapted from: (Nemergut et al., 2013)

Priority effects, where the outcome of microbial interactions depends on the order in which species arrive, are the basis of historically contingent assembly. These effects can lead to divergent community compositions, even under identical environmental conditions, simply due to variation in colonization sequences. Priority effects are typically categorized into two types: niche pre-emption, where early-arriving species occupy available niches and exclude later arrivals through competition, and niche modification, where early colonists alter the environment in ways that affect the success of later species (Debray et al., 2022). Of the two, niche pre-emption is likely the more widespread mechanism in gut microbiomes, as it does not require specialized functional capabilities and can be enacted by a wide range of taxa. The role of priority effects in gut microbial communities, especially early life, has been reviewed (Sprockett et al., 2018), highlighting their potential to create persistent individual differences despite similar host environments. The conditions under which priority effects are most influential, and the extent to which they explain variation in gut microbial community structure, remain poorly understood. Clarifying these dynamics will require the integration of conceptual frameworks with carefully designed experiments in tractable microbial systems (Orr et al., 2025).

Decades of research in coexistence theory can inform us about understanding and predicting priority effects (Grainger et al., 2019). It provides a basis to predict the outcome of species interactions (or, for bacteria, also strain interactions) as conceptualized over the past few years (Fukami et al., 2016). Modern coexistence theory or MCT, like Vellend's framework for

community ecology, also successfully integrates niche-based and neutral concepts (Adler et al., 2007). The framework is founded on the ideas of niche and fitness difference between a pair of competing species and their response to environmental fluctuations in terms of an increase or decrease in niche and fitness difference (Orr et al., 2025). One species would dominate the other (competitive exclusion) if they occupied the same niche and had a significant fitness difference. On the other hand, if they had no difference in fitness (neutralism) or occupied different niches with little overlap, they could coexist despite significant fitness differences, as demonstrated for *Lactobacillus* species (Brochet et al., 2021). Priority effects will occur when a pair with a slight fitness difference competes in the same niche, but with an external factor that can provide an advantage to the stronger competitor (positive frequency-dependent growth rate). This situation can arise when a species depletes more of a resource that a competitor depends on, for example, by arriving earlier or having a larger population, hence consuming at a faster rate (niche pre-emption). Priority effects have been observed between gut microbiota (Laursen & Roager, 2023; Ojima et al., 2022; Segura Munoz et al., 2022), and their importance in microbiome assembly has been recognized (Martínez et al., 2018; Sprockett et al., 2018); however, a compelling avenue that is only beginning to be investigated is the role of priority effects in strain-level interactions in the gut microbial community. I use this framework to investigate the prevalence and role of priority effects in shaping the strain-level diversity in the honeybee gut microbiome (Chapter 3).

2.2 EVOLUTIONARY PROCESSES AND THEIR CONSEQUENCES

“...evolutionary aspects of host–parasite associations cannot be observed within the lifespan of a researcher. Methods for inferring the effects of interactions have thus been developed based on comparisons of the phylogenies of the interacting species.”

– D. M. de Vienne *et al* (de Vienne et al., 2013)

The consistent close association of gut microbiota and their hosts across generations, along with apparently widespread observations of phylosymbiosis, has led to hypotheses that gut microbiota co-evolve or co-diversify with their hosts (Groussin et al., 2020). The result of this process would be congruent or parallel host-microbe phylogenies, a pattern referred to as co-phylogeny. Event-based and Distance-based methods of detecting and comparing congruence between host and microbe phylogeny (co-phylogenetic signal) have been thoroughly reviewed and their strengths and shortcomings discussed (de Vienne et al., 2013; Dismukes et al., 2022; Groussin et al., 2020; Moeller et al., 2023). Careful inspection and comparison of host and microbe trees are essential to correctly identify the processes underlying co-phylogeny patterns. Several events, such as losses, gains, host switches etc., can occur in the course of host-microbe evolution and lead to different phylogenetic patterns (**Fig. 6**).

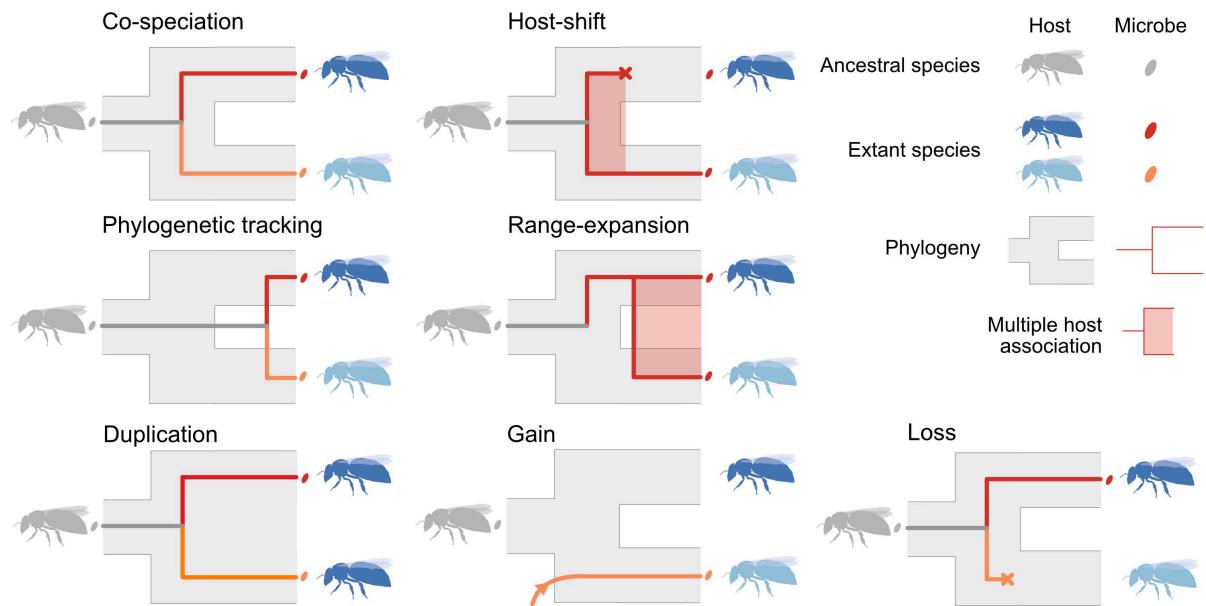


Figure 6: Evolutionary events reflected in host and microbe phylogenies. Co-speciation is reflected by congruent host and microbial phylogenies. Further testing can establish if the events might have occurred at the same time period. Phylogenetic tracking sometimes referred to as pseudocospeciation also involves congruent trees but the speciation events did not happen in the same time period. Host-shift can be either complete or incomplete. The latter case is referred to as range-expansion. Finally, duplications, gains and loss can also be inferred from host and microbial phylogenies.

Congruent phylogenies or high co-phylogenetic signal can result from: co-evolution, phylogenetic tracking or vicariance (Althoff et al., 2014; Blasco-Costa et al., 2021). Vicariance represents the scenario where congruent phylogenies emerge from biogeographic factors rather than any biological interaction. Co-evolution is on the other end of the spectrum, wherein the host and microbiota acquire reciprocal genetic change due to the selective pressure imposed on one another. This process operates when the host and microbiota strongly affect each other's fitness, for example, as in the case of pea aphids and their endosymbiont *Buchnera* (Moran, 2001), where they supply each other with amino acids that the other does not encode the biosynthetic pathway to make (Baumann, 2005). In a co-evolving system, reciprocal genetic changes result in reproductive isolation of the host and ensure that host-microbe pairs cannot survive without each other. As a result, host-specific relationships are maintained over evolutionary timescales.

For stable co-evolution, several criteria must be met, such as strong fitness dependencies and faithful transmission across generations. Unlike in obligate symbioses, most host-gut microbiota systems do not strictly satisfy these requirements. This explains why we do not see strong evidence for co-evolution among gut microbiota and their hosts despite their significant influence on each other (Groussin et al., 2020; Moran & Sloan, 2015). Instead, co-diversification a scenario less strict than co-evolution, has been reported in studies, ranging from mammals to insects (Arora et al., 2023; Koch et al., 2013; Moeller et al., 2016; Rühlemann et al., 2024; Suzuki et al., 2021). A recent study of primates observed that only 16 – 36% of gut microbial community members co-diversify with their host (Sanders et al., 2023), indicating that other important processes shaping microbiome evolution are also at play. Co-diversification does not imply reciprocal genetic change between the host and microbes; it only means that there were simultaneous co-speciation events in their shared history (**Fig. 6**). An important question yet unanswered is how prevalent co-diversification is among gut microbiota and what other processes can lead to such patterns of congruence in host-microbe phylogenies.

While congruence and strong co-phylogenetic signals are taken to be evidence for co-diversification, it is essential to consider several alternate scenarios that can contribute to this pattern before confirming co-diversification (**Fig. 6**). When there is a speciation event in the microbial phylogeny, it results from co-diversification if it coincides with a speciation in the host phylogeny. If the event happens much later, it contributes to the congruence between host and microbial phylogenies even though there is no co-diversification (phylogenetic tracking) (**Fig. 6**). The resultant asynchrony (when small enough) may not degrade the strength of co-phylogenetic signal or congruence (**Fig. 7B**) and has often led to incorrect observations of co-diversification (de Vienne et al., 2013). Further, a duplication event in an ancestor would result in hosts emerging from subsequent speciation events associated with both microbial species. If each host loses the microbial species that are less competitive in their gut environment (selection), congruent phylogenies will result. These scenarios can be distinguished if the timing of the host and microbial speciation events can be inferred and compared. For example, if available, fossil records can enable the estimation of host speciation events. For microbes, this is complicated, but an interesting approach could be to estimate the rate at which the microbe must have been evolving had the host and microbe speciation events coincided and assess whether the estimate falls within expected ranges for bacteria (Sanders et al., 2023). Finally, other events such as loss, gain, host range-expansion, or host-shift (also referred to as switches) can contribute in tandem with the different events to increase or degrade the co-phylogenetic signal (**Fig. 6**). The relative contribution and frequency of these events across different host-gut microbiome systems are still under active investigation, and more high-throughput comparative studies of closely related host species is essential.

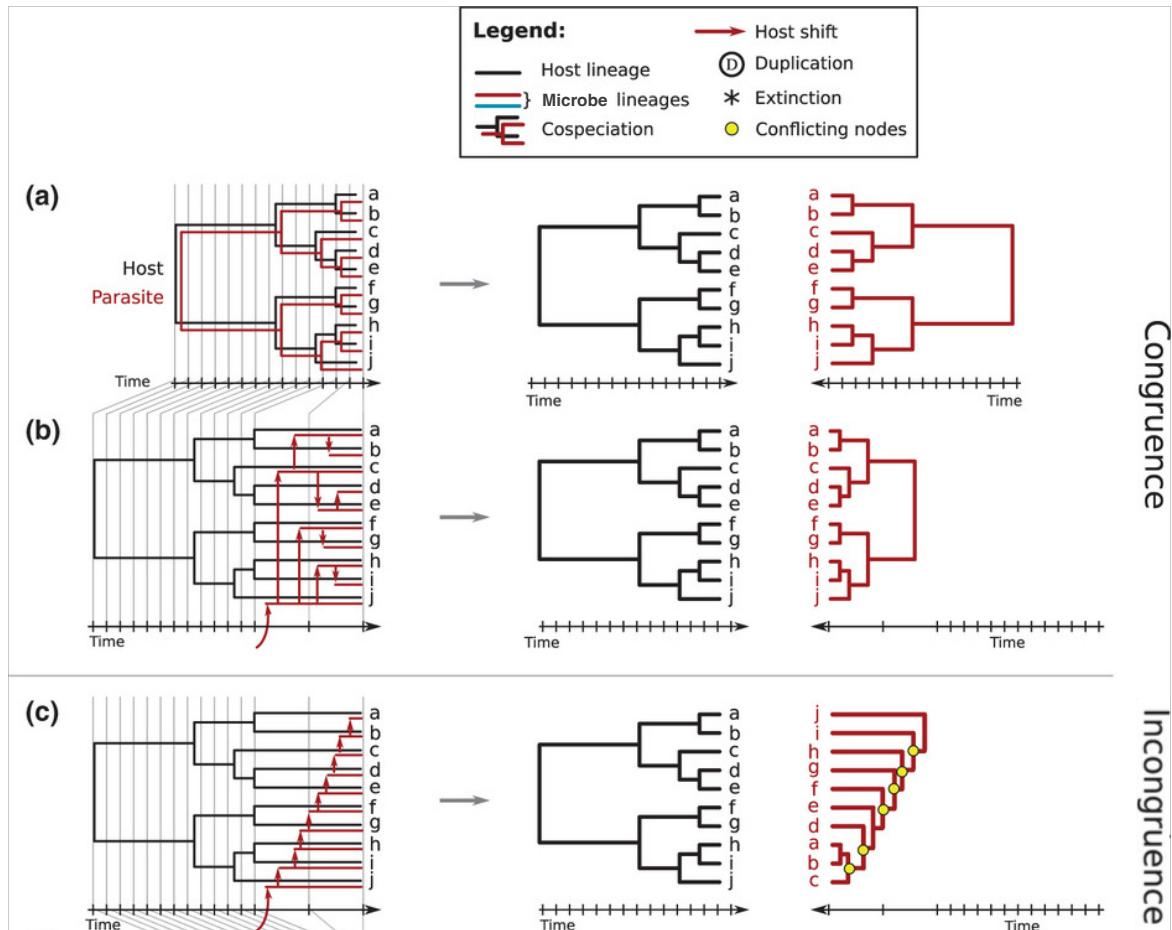


Figure 7. Host-microbe phylogenetic patterns resulting from different evolutionary scenarios. The phylogenies in black represent that of the host, and the ones in blue and red represent those of the microbes. (a) Co-diversification with the strongest co-phylogenetic signal. Congruent phylogenies due to co-diversification. (b) No co-diversification with a strong co-phylogenetic signal (by topology or distance). Congruent phylogenies, but with shorter branches in the microbe lineages due to host-shift speciation events or phylogenetic tracking. (c) Host-shift speciation events leading to incongruent phylogenies. Adapted from: (de Vienne et al., 2013)

Finally, the concepts and methods for testing the congruence of host-microbe phylogenies have been developed for host-pathogen systems or one-to-one symbioses between a host and a microbial clade (genus or species). In gut microbiota, hundreds of clades (species) are typically associated with each host species. The taxonomic level at which the microbes' phylogenies are inferred is a critical consideration. When comparing hosts that diverged 10 million years ago (mya), microbes of different phyla (e.g., *Lactobacillus reuteri* and *Bifidobacterium bifidum*), which diverged more than 10^3 mya, (Kumar et al., 2022) must have already speciated well before the hosts. Associations may instead be tested between family or genus-level trees and host genera or species within a genus. A clear framework and appropriate methods considering the complexity of host microbiota evolution across host species are still needed, and work in this area is only beginning (Mazel et al., 2025; Moeller et al., 2023).

3 THE HONEYBEE GUT MICROBIAL COMMUNITY

The honeybee gut microbiome is a tractable system apt to study patterns and processes driving both the long-term evolution and short-term ecology of gut microbiota (Douglas, 2019; Engel et al., 2016). In contrast to the diversity of vertebrate gut microbiomes, which harbor hundreds of species spanning over a hundred genera (Schloissnig et al., 2013), honeybees harbor relatively simpler communities still high in within-species diversity (Ellegaard et al., 2019). For example, *Apis mellifera* comprises fewer than 50 species across ~10 genera (Prasad et al., 2025). The members of the honeybee gut microbiome are primarily associated with honeybees, specialized and are consistently found across individuals (Martinson et al., 2011). This simplicity and consistency make it feasible to resolve extensive strain-level diversity using shotgun metagenomic sequencing in large-scale studies.

The species *Apis mellifera*, or the western honeybee, is well-studied given its economic and agricultural relevance. However, less is known about the gut microbiome of the rest of the clade, despite their important roles in their ecosystems (Hepburn & Radloff, 2011). Their prevalence and well-known ecology and evolutionary history provide an excellent opportunity to survey gut microbial communities across related host species in their natural ecosystem. Finally, *Apis mellifera* can be reared in apiaries and controlled laboratory conditions, also free of gut microbiota, allowing precise manipulation and observation of microbial colonization and assembly. This enables the investigation of ecological processes shaping the community and the mechanisms behind them in the short term. In this section, I present an overview of the evolution and ecology of the *Apis* clade with notes on how these characteristics can influence their gut microbiome, followed by gut microbiome diversity patterns observed within and across *Apis* species and finally, how the honeybee microbiome can be a powerful model to investigate eco-evolutionary processes.

3.1 HONEYBEE EVOLUTION AND ECOLOGY

"Despite the fact that they are systematically so far removed from man, forager worker bees have several features in common with mature humans: they are social, have a highly developed communication system and learning capacity, are diurnally active, spend their resting phases in sheltered surroundings and have one main bout of sleep, at night, every 24 h."

– Walter Kaiser and Jana Steiner-Kaiser
(Kaiser & Steiner-Kaiser, 1987)

Honeybees (genus *Apis*) are a group of eusocial insects. They live in large colonies comprising thousands to tens of thousands of individuals, including sterile female worker bees who attend to most duties of foraging and nursing, a fertile queen solely responsible for laying eggs, and a few hundred drones or male bees. Honeybees are part of the Apidae family, comprising bumble and stingless bees. The *Apis* clade is estimated to have originated 30-40

million years ago (Arias & Sheppard, 2005). It comprises several species that, while sharing key traits such as social behavior, colony structure, and nectar foraging, differ markedly in their morphology, ecology, and geographic distribution (Hepburn & Radloff, 2011). These differences make the *Apis* clade an informative system for comparative studies across closely related taxa with divergent life histories and environmental niches. Despite their shared evolutionary origin and broadly similar ecological roles as pollinators, the variation across *Apis* species provides a unique opportunity to understand how host traits and behaviour can lead to differences in gut microbiome diversity.

The genus *Apis* is traditionally divided into three main lineages based on morphology and nesting behavior: the dwarf honeybees (subgenus *Micrapis*), the giant honeybees (subgenus *Megapis*), and the cavity-nesting honeybees (subgenus *Apis*) (**Fig. 8A**). This classification, originally based on morphological and behavioral traits (Hepburn & Radloff, 2011), is also supported by modern molecular data and phylogenies (Carr, 2023).

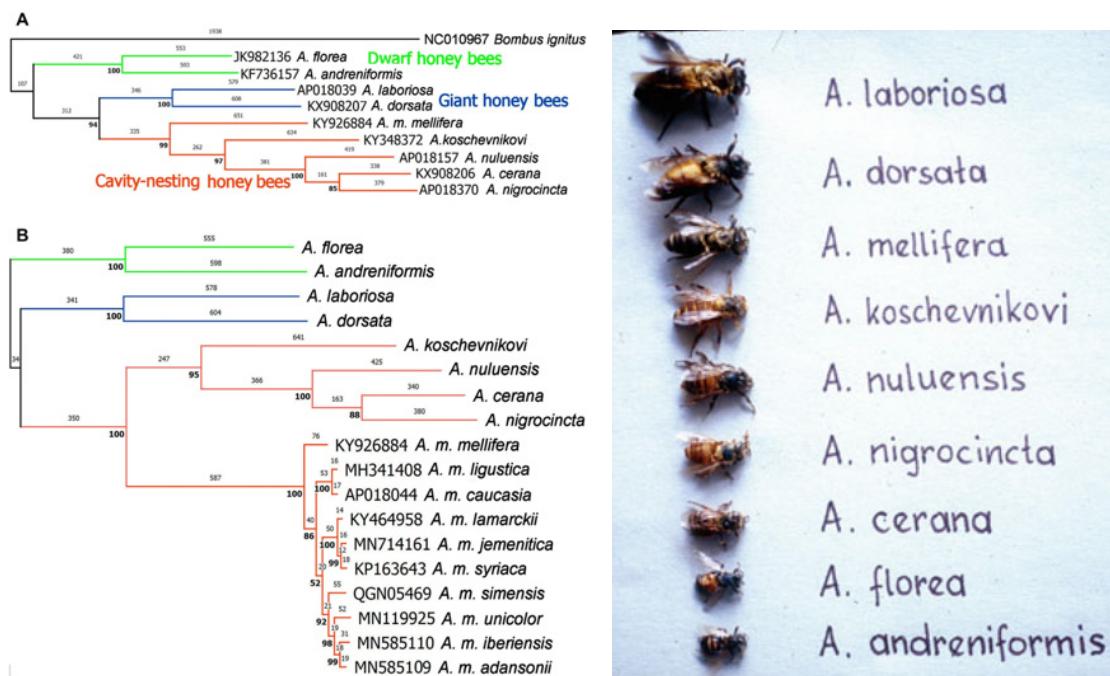


Figure 8. Honeybee phylogeny and picture of worker bees of nine honeybee species. Maximum parsimony analysis of phylogenetic relationships among mitochondrial DNA sequences of nine (A) honeybee species rooted with a bumblebee *Bombus ignitus* as outgroup. (B) Maximum parsimony analysis of phylogenetic relationships among mitochondrial DNA sequences of honeybee species with several extant subspecies of *A. mellifera* included. Reproduced from: (Carr, 2023) (C) Worker bees of different *Apis* species are arranged in order to decrease worker size Reproduced from: (Hepburn & Radloff, 2011).

The cavity-nesting lineage includes *Apis mellifera*, the western honeybee, which is the most widely distributed and economically significant species, native to Europe, Africa, and parts of Asia. Its evolutionary history, especially among recently diverged subspecies (**Fig. 8B**), has been a subject of hot debate, with earlier studies suggesting an “Out of Africa” model (C. W. Whitfield et al., 2006) and subsequent ones suggesting an Asian origin followed by multiple expansions into Africa and Europe, (Dogantzis et al., 2021) neither of which is entirely consistent with recent data (Carr, 2023). *Apis cerana*, the eastern or asian honeybee, is another cavity-nesting species prevalent across Asia. It is often studied alongside *Apis mellifera* due to their similar nesting habits and importance in regional apiculture (Hepburn

& Radloff, 2011). However, the modern *Apis mellifera* populations commonly used in research and agriculture are shaped by a complex history of domestication, breeding, and human-mediated movement. These include distinct subspecies selectively bred for traits like docility or honey productivity. This anthropogenic mixing has led to highly admixed populations with divergent evolutionary and ecological histories from their wild ancestors (Wallberg et al., 2014). It is important to take this into consideration when comparing *Apis mellifera* to other *Apis* species, particularly *Apis cerana*. The closer relatives of *Apis cerana*, such as *Apis koschevnikovi*, *Apis nigrocincta*, *Apis nuluensis* which only found in the island of Borneo are much less studied and present interesting avenues for further exploration of cross-species gut microbiome variation especially among sympatric host species.

Dwarf honeybees comprise *Apis florea* and the black dwarf honeybee *Apis andreniformis*. These bees are primarily found in tropical and subtropical regions of Asia. *Apis florea* has a wide distribution ranging from eastern Oman and southern Iran, across the Indian subcontinent, to Southeast Asia, while *Apis andreniformis* is more restricted, found from the eastern Himalayan foothills through Southeast Asia and into parts of Malaysia. Although their ranges overlap in some regions, they are behaviorally and morphologically distinct. Notably, *Apis florea* has shown signs of recent range expansion into new areas, potentially aided by human activities. Both species exhibit high swarming and migratory tendencies, likely adaptations to seasonal resource availability. Their small colony sizes (Table 1) and limited nest defensibility contribute to a life history strategy characterized by frequent relocation and reproduction (Hepburn & Radloff, 2011).

Table 1. Features of honeybee species (Dyer & Seeley, 1991; Hepburn & Radloff, 2011; Smith, 2020)

SPECIES	COLONY SIZE	WORKER SIZE (LENGTH)	NESTING (HABITAT)	DISTRIBUTION
A. MELLIFERA	10,000–50,000	10–15 mm	Cavity (hollow trees, managed hives)	Europe, Africa, western Asia; now global via domestication
A. CERANA	5,000–8,000	10–11 mm	Cavity (trees, rock crevices)	South, Southeast, and East Asia
A. DORSATA	50,000–100,000	17–20 mm	Open (exposed single combs on cliffs or branches)	South and Southeast Asia
A. FLOREA	6,000	7–10 mm	Open (small single combs on low shrubs)	South and Southeast Asia
A. ANDRENIFORMIS	6,000	8–9 mm	Open (single combs in low vegetation)	Overlaps with A. florea in tropical Asia

The giant honeybees, represented by the best-known species of the clade, *Apis dorsata*, are recognized by their large size and open nesting behaviour (notoriety as the most dangerous

honeybees to sample based on the trying experience of my close collaborator Asha). They typically construct open combs on remarkably tall trees or vertical rock faces (and on tall buildings in urban environments) (**Fig. 9A**). Giant bees possess some unique characteristics among honeybees besides large worker size and colony size (**Table 1**). *Apis dorsata* are known to form massive clusters of colonies, sometimes hundreds of hives on one tree (Hepburn & Radloff, 2011) (**Fig. 9B**). This means that there might be greater contact and potential opportunities for microbiota exchange between honeybees of different colonies. All the giant honeybee species except the Himalayan giant honeybee, *Apis laboriosa*, are capable of nocturnal foraging (Young et al., 2021), possibly leading to exposure to a different set of floral microbiomes, as other plants are in bloom during the day than at night. While absconding (much to the dismay of wild bee beekeepers and honey harvesters) and migration are characteristic of Asian and African species of honeybees, the frequency of these behaviours and distances travelled can vary. Colonies of *Apis dorsata* can migrate large distances (up to 200 km away). This could mean that giant honeybees encounter a larger pool of microbial species than do the other honeybee species.



Figure 9. Picture of selected open hives sampled in India and Malaysia. (A) *Apis dorsata* sampled in India. The hive was located on the highest ledge of a four-story building. (B) Large tree with several colonies of *Apis dorsata* in Temenggor Forest Reserve, Malaysia. Wooden scaffolds were built by local honey harvesters from the Orang Asli community. (Picture credit: John Chan, Nature inspired, Malaysia) (C) Close-up of *Apis andreniformis* hive sampled in Malaysia from a low tree branch. (D) *Apis florea* hive on a small tree, sampled in Malaysia.

3.2 PATTERNS AND PROCESSES IN THE HONEYBEE GUT MICROBIOME

The honeybee gut microbiome and its influence on the host are best studied in the western honeybee *Apis mellifera*. Gut microbes perform essential functions for the host, including regulating gut pH, metabolizing complex plant-derived polysaccharides via carbohydrate-active enzymes (CAZymes), detoxification of xenobiotics, and modulating the host immune system (Motta & Moran, 2023). *A. mellifera* harbors a highly conserved, specialized microbial community composed of a small number of genera that are nearly universally present across

individuals (Engel et al., 2012; Moran et al., 2012; Sarton-Lohéac et al., 2023). They include *Lactobacillus*, *Bifidobacterium* (genera also found in other gut microbiomes, including that of primates), *Bombilactobacillus*, *Snodgrassella*, *Gilliamella*, *Frischella* (less prevalent) and *Bartonella* (especially prevalent over winter) (Kešnerová et al., 2020), several of which are also found in other distant clades of bees, including several species of stingless bees and bumble bees (Kwong et al., 2017). Though the genera prevalent in *Apis mellifera* have been regarded as honeybee core microbiota, some are not as prevalent in other honeybee species (e.g., *Snodgrassella* is not prevalent or abundant in *Apis dorsata*). Further, some genera absent in *Apis mellifera* are highly prevalent in other honeybee species, such as *Apibacter* in *Apis dorsata* and *Apis cerana*.

Among shared genera of social bees are several host-specific clades only found in their respective host species (Koch et al., 2013). These host-specific clades were, however, not congruent with host phylogeny, challenging the long-held idea that gut microbiota have been associated with the common ancestor of social bees and have since co-diversified with their hosts (Sarton-Lohéac et al., 2023). To further understand the relevance and prevalence of co-diversification to gut microbiome evolution, genomic information of the entire community across several closely related host species must be compared (Chapter 2).

The structure in the honeybee gut microbial community, as revealed by 16S rRNA amplicon sequencing, suggests essential variations across species at the genus level (Kwong et al., 2017). The patterns in species and strain-level diversity of the entire gut microbial community across individuals within as well as between host species have since been revealed by shotgun metagenomics. In the *Apis mellifera* gut microbiome, several sequence-discrete populations (SDPs, similar to the taxonomic level of species) co-exist within individuals (Ellegaard & Engel, 2019). As demonstrated for four *Lactobacillus* species in the honeybee gut, this could be enabled by niche partitioning of complex nutrients from pollen, a key component of the honeybee diet (Brochet et al., 2021). Many variant strains of an SDP are found among individuals of a colony. However, strain-level diversity found within a colony is segregated among individuals (Ellegaard & Engel, 2019). This segregation may be enabled by priority effects (at the strain level). Further analysis across colonies of honeybees at the strain level are needed to further clarify these patterns, and subsequent laboratory experiments can help us understand the processes (e.g., priority effects) maintaining them.

Comparison between the *Apis mellifera* and *Apis cerana* gut microbiomes shows significant differences in genomic diversity within microbial species or SDPs with important functional consequences (Ellegaard et al., 2020). One explanation for this difference is human influence on *Apis mellifera*. Most of the accessible colonies of *Apis mellifera* are human-managed. Concerted efforts to survey the microbiome of wild populations of *Apis mellifera* (found across Africa) are needed to establish the effect of human management on honeybee gut microbiome diversity. Another explanation is differences in host ecological conditions (**Table 1**). For example, the differences in diversity may be explained by a species-area relationship. Habitat size in this context may be interpreted as the gut size of each individual or the colony size. Such a relationship has been observed in a large study of several social bees including honeybees (Kwong et al., 2017).

Evolutionary processes may explain the host-specific nature of the *Apis cerana* and *Apis mellifera* gut microbiota. A comparison of host and microbial phylogenies across host species is required to clarify this further. Neutral or stochastic processes may also suffice to explain these differences. For example, honeybees acquire their microbiota from conspecifics, so dispersal limitation between host species can maintain host specificity. On the other hand, selection (host-filtering) can also explain these patterns. For example, the microbial species found in *Apis cerana* may be better adapted to its host after a prolonged period of association and hence unlikely to switch. However, such an explanation might better suit more divergent hosts with very different physiologies than species of the same genus. It has been shown that while *Snodgrassella* species from bumble bees were unsuccessful in colonizing the *Apis mellifera* gut, species isolated from other *Apis* bees were successful. Further work is needed to test the patterns and extent of host specificity across the members of the honeybee gut microbiome. A detailed understanding of the patterns in gut microbiome diversity across closely related host species at a finer scale, followed by experiments to test the prevalence and importance of various ecological processes, is essential.

Honeybees are particularly well-suited as experimental models for studying patterns in gut microbiome diversity and the processes shaping them. They can be reared free of gut microbiota in a laboratory setting, allowing controlled colonization experiments (Douglas, 2019; Zheng et al., 2018). Individual bacterial species or defined communities can be introduced through feeding or direct inoculation (**Fig. 10**). Most of the community members of *Apis mellifera* have been isolated and characterized. Several genetic tools have also been developed to study the mechanisms underlying microbe-microbe and host-microbe interactions in the gut. As described later, I utilize this tractable system (Chapter 3) to study ecological processes shaping strain-level diversity in the gut microbiome.



Figure 10. Bacterial community inoculation. A picture of one video frame showing the inoculation procedure used for honeybee gut microbiome experiments. A honeybee (*Apis mellifera*) raised free of gut microbiota was fed a bacterial community diluted in PBS and Sugar water. The honeybee was placed inside a sterile 1.5 mL microcentrifuge tube with the bottom removed. A sterile pipette tip with 5 μ L of inoculum was fed to the

honeybee. In this frame, the honeybee extends its proboscis in response to the pipette tip containing the inoculum. Video: <https://doi.org/10.5281/zenodo.15447400>.

4 THESIS OUTLINE

In this thesis, I sought to elucidate the patterns in gut microbial community composition and the processes shaping their assembly across closely related host species and using the honeybee gut microbiome as a model. Two complementary approaches might be employed in this pursuit. One is that of keen observation and comparison of diverse natural microbiomes across comparable but different host species. By examining how microbial communities look and function in systems separated by millions of years of evolution, we can gain valuable insights into how host-microbiome associations are shaped over evolutionary timescales. This is the approach I employed in **Chapter 2** of this thesis. The other approach is manipulating isolated microbes and simplified synthetic communities in controlled microbiota-free honeybees. This approach allows for targeted investigation of microbial traits and interactions, opening up opportunities to elucidate the mechanistic underpinnings of community assembly in shorter, ecological timescales. This is the approach I leveraged in **Chapter 3** of this thesis.

In **Chapter 2**, I examine gut microbiome composition across five species of honeybees at high resolution, down to strain-level profiles, and assess evolutionary processes that have shaped these communities. While microbiomes were largely host-specific, underlying this pattern was a mix of many host-specific bacterial species and several closely related bacterial species that were less host-specific or shared by several hosts. Closer analysis of various shared bacterial genera revealed that host-specificity was not maintained by strict co-diversification. Microbiome structure and function were shaped by symbiont gains, losses and host switches over evolutionary time. This chapter highlights the value of modern sequencing technology and bioinformatic approaches that enable going beyond community-level patterns in microbiome diversity to understand the evolution and ecology of microbiota.

Chapter 3 shifts direction from patterns to ecological processes, focusing on priority effects in community assembly at the strain level. Bacterial species can co-exist by niche partitioning in the nutritionally complex environment of the honeybee gut. I hypothesize that within bacterial species, priority effects by way of niche-preemption would dictate that the strain arriving first dominates the community. Using synthetic communities inoculated into microbiota-deprived honeybees, I observed strong priority effects across all tested species, where the strains that arrived first dominated over strains that came later. Dropout experiments further revealed that in several cases, the hypothesis holds that second arrivers had better colonization success when the other strain from the species was dropped out of the community inoculated first. However, in all cases, interspecies interactions or other effects might play a role, indicating the niches may be more flexible in their overlap between related species. This study provides valuable insight into an important ecological process driving gut microbial communities at the strain level.

Finally, **Chapter 4** synthesizes these findings, discussing the effect and relevance of various eco-evolutionary factors driving microbiome diversity and function. I outline future research

directions toward achieving a comprehensive understanding of microbiome diversity and dynamics that will enable us to precisely manipulate and engineer microbiomes.

REFERENCES

- Adler, P. B., HilleRisLambers, J., & Levine, J. M. (2007). A niche for neutrality. *Ecology Letters*, 10(2), 95–104. <https://doi.org/10.1111/j.1461-0248.2006.00996.x>
- Althoff, D. M., Segraves, K. A., & Johnson, M. T. J. (2014). Testing for coevolutionary diversification: Linking pattern with process. *Trends in Ecology & Evolution*, 29(2), 82–89. <https://doi.org/10.1016/j.tree.2013.11.003>
- Arias, M. C., & Sheppard, W. S. (2005). Phylogenetic relationships of honey bees (Hymenoptera:Apinae:Apini) inferred from nuclear and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, 37(1), 25–35. <https://doi.org/10.1016/j.ympev.2005.02.017>
- Arora, J., Buček, A., Hellemans, S., Beránková, T., Romero Arias, J., Fisher, B. L., Clitheroe, C., Brune, A., Kinjo, Y., Šobotník, J., & Bourguignon, T. (2023). Evidence of cospeciation between termites and their gut bacteria on a geological time scale. *Proceedings of the Royal Society B: Biological Sciences*, 290(2001), 20230619. <https://doi.org/10.1098/rspb.2023.0619>
- Baumann, P. (2005). BIOLOGY OF BACTERIOCYTE-ASSOCIATED ENDOSYMBIONTS OF PLANT SAP-SUCKING INSECTS. *Annual Review of Microbiology*, 59(Volume 59, 2005), 155–189. <https://doi.org/10.1146/annurev.micro.59.030804.121041>
- Becking, L. B. (1931). *Gaia of leven en aarde: Rede uitgesproken bij het aanvaarden van het ambt van gewoon hoogleeraar aan de Rijksuniversiteit te Leiden op woendssdag den 28 januari 1931*. Nijhoff.
- Bell, G. (2001). Neutral Macroecology. *Science*, 293(5539), 2413–2418. <https://doi.org/10.1126/science.293.5539.2413>
- Blasco-Costa, I., Hayward, A., Poulin, R., & Balbuena, J. A. (2021). Next-generation cophylogeny: Unravelling eco-evolutionary processes. *Trends in Ecology & Evolution*, 36(10), 907–918. <https://doi.org/10.1016/j.tree.2021.06.006>
- Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., & Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife*, 10, e68583. <https://doi.org/10.7554/eLife.68583>
- Cani, P. D. (2018). Human gut microbiome: Hopes, threats and promises. *Gut*, 67(9), 1716–1725. <https://doi.org/10.1136/gutjnl-2018-316723>
- Carr, S. M. (2023). Multiple mitogenomes indicate Things Fall Apart with Out of Africa or Asia hypotheses for the phylogeographic evolution of Honey Bees (*Apis mellifera*). *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-35937-4>
- Connor, E. F., & McCoy, E. D. (1979). The Statistics and Biology of the Species-Area Relationship. *The American Naturalist*, 113(6), 791–833.
- Custer, G. F., Gans, M., van Diepen, L. T. A., Dini-Andreote, F., & Buerkle, C. A. (2023). Comparative Analysis of Core Microbiome Assignments: Implications for Ecological Synthesis. *mSystems*, 8(1), e01066-22. <https://doi.org/10.1128/msystems.01066-22>
- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484), 559–563. <https://doi.org/10.1038/nature12820>
- de Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: Methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, 198(2), 347–385. <https://doi.org/10.1111/nph.12150>

De Wit, R., & Bouvier, T. (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology*, 8(4), 755–758. <https://doi.org/10.1111/j.1462-2920.2006.01017.x>

Degregori, S., Wang, X., Kommala, A., Schulhof, N., Moradi, S., MacDonald, A., Eblen, K., Jukovich, S., Smith, E., Kelleher, E., Suzuki, K., Hall, Z., Knight, R., & Amato, K. R. (2025). Comparative gut microbiome research through the lens of ecology: Theoretical considerations and best practices. *Biological Reviews*, 100(2), 748–763. <https://doi.org/10.1111/brv.13161>

Dismukes, W., Braga, M. P., Hembry, D. H., Heath, T. A., & Landis, M. J. (2022). Cophylogenetic Methods to Untangle the Evolutionary History of Ecological Interactions. *Annual Review of Ecology, Evolution, and Systematics*, 53(Volume 53, 2022), 275–298. <https://doi.org/10.1146/annurev-ecolsys-102320-112823>

Dogantzs, K. A., Tiwari, T., Conflitti, I. M., Dey, A., Patch, H. M., Muli, E. M., Garnery, L., Whitfield, C. W., Stolle, E., Alqarni, A. S., Allsopp, M. H., & Zayed, A. (2021). Thrice out of Asia and the adaptive radiation of the western honey bee. *Science Advances*, 7(49), eabj2151. <https://doi.org/10.1126/sciadv.abj2151>

Doolittle, W. F. (2012). Population Genomics: How Bacterial Species Form and Why They Don't Exist. *Current Biology*, 22(11), R451–R453. <https://doi.org/10.1016/j.cub.2012.04.034>

Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, 17(12), 764–775. <https://doi.org/10.1038/s41579-019-0242-1>

Drew, G. C., Stevens, E. J., & King, K. C. (2021). Microbial evolution and transitions along the parasite–mutualist continuum. *Nature Reviews Microbiology*, 19(10), 623–638. <https://doi.org/10.1038/s41579-021-00550-7>

Dyer, F. C., & Seeley, T. D. (1991). Nesting Behavior and the Evolution of Worker Tempo in Four Honey Bee Species. *Ecology*, 72(1), 156–170. <https://doi.org/10.2307/1938911>

Ellegaard, K. M., Brochet, S., Bonilla-Rosso, G., Emery, O., Glover, N., Hadadi, N., Jaron, K. S., Meer, J. R., Robinson-Rechavi, M., Sentchilo, V., Tagini, F., SAGE class 2016-17, & Engel, P. (2019). Genomic changes underlying host specialization in the bee gut symbiont *Lactobacillus Firm5*. *Molecular Ecology*, 28(9), 2224–2237. <https://doi.org/10.1111/mec.15075>

Ellegaard, K. M., & Engel, P. (2019). Genomic diversity landscape of the honey bee gut microbiota. *Nature Communications*, 10(1), 446. <https://doi.org/10.1038/s41467-019-08303-0>

Ellegaard, K. M., Suenami, S., Miyazaki, R., & Engel, P. (2020). Vast Differences in Strain-Level Diversity in the Gut Microbiota of Two Closely Related Honey Bee Species. *Current Biology*, 30(13), 2520-2531.e7. <https://doi.org/10.1016/j.cub.2020.04.070>

Engel, P., Kwong, W. K., McFrederick, Q., Anderson, K. E., Barribeau, S. M., Chandler, J. A., Cornman, R. S., Dainat, J., de Miranda, J. R., Doublet, V., Emery, O., Evans, J. D., Farinelli, L., Flenniken, M. L., Granberg, F., Grasis, J. A., Gauthier, L., Hayer, J., Koch, H., ... Dainat, B. (2016). The Bee Microbiome: Impact on Bee Health and Model for Evolution and Ecology of Host-Microbe Interactions. *mBio*, 7(2), e02164-15. <https://doi.org/10.1128/mBio.02164-15>

Engel, P., Martinson, V. G., & Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences*, 109(27), 11002–11007. <https://doi.org/10.1073/pnas.1202970109>

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, 37(5), 699–735. <https://doi.org/10.1111/1574-6976.12025>

Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., Clemente, J. C., Knight, R., Heath, A. C., Leibl, R. L., Rosenbaum, M., & Gordon, J. I. (2013). The Long-Term Stability of the Human Gut Microbiota. *Science*, 341(6141), 1237439. <https://doi.org/10.1126/science.1237439>

- Fukami, T., Mordecai, E. A., & Ostling, A. (2016). A framework for priority effects. *Journal of Vegetation Science*, 27(4), 655–657. <https://doi.org/10.1111/jvs.12434>
- Gibbons, S. M., Kearney, S. M., Smillie, C. S., & Alm, E. J. (2017). Two dynamic regimes in the human gut microbiome. *PLOS Computational Biology*, 13(2), e1005364. <https://doi.org/10.1371/journal.pcbi.1005364>
- Grainger, T. N., Letten, A. D., Gilbert, B., & Fukami, T. (2019). Applying modern coexistence theory to priority effects. *Proceedings of the National Academy of Sciences*, 116(13), 6205–6210. <https://doi.org/10.1073/pnas.1803122116>
- Groussin, M., Mazel, F., & Alm, E. J. (2020). Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host & Microbe*, 28(1), 12–22. <https://doi.org/10.1016/j.chom.2020.06.013>
- Hammer, T. J., Sanders, J. G., & Fierer, N. (2019). Not all animals need a microbiome. *FEMS Microbiology Letters*, 366(10), fnz117. <https://doi.org/10.1093/femsle/fnz117>
- Hepburn, H. R., & Radloff, S. E. (Eds.). (2011). *Honeybees of Asia*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-16422-4>
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton University Press. <https://www.jstor.org/stable/j.ctt7rj8w>
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1), 5114. <https://doi.org/10.1038/s41467-018-07641-9>
- Kaiser, W., & Steiner-Kaiser, J. (1987). Sleep Research on Honeybees: Neurophysiology and Behavior. In R. Menzel & A. Mercer (Eds.), *Neurobiology and Behavior of Honeybees* (pp. 112–120). Springer. https://doi.org/10.1007/978-3-642-71496-2_10
- Kešnerová, L., Emery, O., Troilo, M., Liberti, J., Erkosar, B., & Engel, P. (2020). Gut microbiota structure differs between honeybees in winter and summer. *The ISME Journal*, 14(3), Article 3. <https://doi.org/10.1038/s41396-019-0568-8>
- Koch, H., Abrol, D. P., Li, J., & Schmid-Hempel, P. (2013). Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Molecular Ecology*, 22(7), 2028–2044. <https://doi.org/10.1111/mec.12209>
- Kohl, K. D. (2020). Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1798), 20190251. <https://doi.org/10.1098/rstb.2019.0251>
- Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, 39(8), msac174. <https://doi.org/10.1093/molbev/msac174>
- Kwong, W. K., Medina, L. A., Koch, H., Sing, K.-W., Soh, E. J. Y., Ascher, J. S., Jaffé, R., & Moran, N. A. (2017). Dynamic microbiome evolution in social bees. *Science Advances*, 3(3), e1600513. <https://doi.org/10.1126/sciadv.1600513>
- Laursen, M. F., & Roager, H. M. (2023). Human milk oligosaccharides modify the strength of priority effects in the *Bifidobacterium* community assembly during infancy. *The ISME Journal*, 17(12), 2452–2457. <https://doi.org/10.1038/s41396-023-01525-7>
- Levin, D., Raab, N., Pinto, Y., Rothschild, D., Zanir, G., Godneva, A., Mellul, N., Futorian, D., Gal, D., Leviatan, S., Zeevi, D., Bachelet, I., & Segal, E. (2021). Diversity and functional landscapes in the microbiota of animals in the wild. *Science*, 372(6539), eabb5352. <https://doi.org/10.1126/science.abb5352>
- Lim, S. J., & Bordenstein, S. R. (2020). An introduction to phylosymbiosis. *Proceedings of the Royal Society B*. <https://doi.org/10.1098/rspb.2019.2900>

- Lozupone, C. A., & Knight, R. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiology Reviews*, 32(4), 557–578. <https://doi.org/10.1111/j.1574-6976.2008.00111.x>
- Mallott, E. K., & Amato, K. R. (2021). Host specificity of the gut microbiome. *Nature Reviews Microbiology*, 19(10), 639–653. <https://doi.org/10.1038/s41579-021-00562-3>
- Martínez, I., Maldonado-Gómez, M. X., Gomes-Neto, J. C., Kittana, H., Ding, H., Schmaltz, R., Joglekar, P., Cardona, R. J., Marsteller, N. L., Kembel, S. W., Benson, A. K., Peterson, D. A., Ramer-Tait, A. E., & Walter, J. (2018). Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly. *eLife*, 7, e36521. <https://doi.org/10.7554/eLife.36521>
- Martino, C., Dilmore, A. H., Burcham, Z. M., Metcalf, J. L., Jeste, D., & Knight, R. (2022). Microbiota succession throughout life from the cradle to the grave. *Nature Reviews Microbiology*, 20(12), 707–720. <https://doi.org/10.1038/s41579-022-00768-z>
- Martinson, V. G., Danforth, B. N., Minckley, R. L., Rueppell, O., Tingek, S., & Moran, N. A. (2011). A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, 20(3), 619–628. <https://doi.org/10.1111/j.1365-294X.2010.04959.x>
- Mazel, F., Davis, K. M., Loudon, A., Kwong, W. K., Groussin, M., & Parfrey, L. W. (2018). Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? *mSystems*, 3(5), e00097-18. <https://doi.org/10.1128/mSystems.00097-18>
- Mazel, F., Guisan, A., & Parfrey, L. W. (2024). Transmission mode and dispersal traits correlate with host specificity in mammalian gut microbes. *Molecular Ecology*, 33(1), e16862. <https://doi.org/10.1111/mec.16862>
- Mazel, F., Knowles, S. C. L., Videvall, E., & Sweeny, A. R. (2023). Evolutionary patterns and processes in animal microbiomes. *Journal of Evolutionary Biology*, 36(12), 1653–1658. <https://doi.org/10.1111/jeb.14248>
- Mazel, F., Prasad, A., & Engel, P. (2025). Host specificity of gut microbiota associated with social bees: Patterns and processes. *Microbiology and Molecular Biology Reviews*, 0(0), e00080-23. <https://doi.org/10.1128/mmbr.00080-23>
- McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealon, K., Pierce, N. E., ... Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, 110(9), 3229–3236. <https://doi.org/10.1073/pnas.1218525110>
- Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., Pusey, A. E., Peeters, M., Hahn, B. H., & Ochman, H. (2016). Cospeciation of gut microbiota with hominids. *Science*, 353(6297), 380–382. <https://doi.org/10.1126/science.aaf3951>
- Moeller, A. H., Sanders, J. G., Sprockett, D. D., & Landers, A. (2023). Assessing co-diversification in host-associated microbiomes. *Journal of Evolutionary Biology*, n/a(n/a). <https://doi.org/10.1111/jeb.14221>
- Moran, N. A. (2001). The Coevolution of Bacterial Endosymbionts and Phloem-Feeding Insects. *Annals of the Missouri Botanical Garden*, 88(1), 35–44. <https://doi.org/10.2307/2666130>
- Moran, N. A., Hansen, A. K., Powell, J. E., & Sabree, Z. L. (2012). Distinctive Gut Microbiota of Honey Bees Assessed Using Deep Sampling from Individual Worker Bees. *PLoS ONE*, 7(4), e36393. <https://doi.org/10.1371/journal.pone.0036393>
- Moran, N. A., Ochman, H., & Hammer, T. J. (2019). Evolutionary and Ecological Consequences of Gut Microbial Communities. *Annual Review of Ecology, Evolution, and Systematics*, 50(1), 451–475. <https://doi.org/10.1146/annurev-ecolsys-110617-062453>
- Moran, N. A., & Sloan, D. B. (2015). The Hologenome Concept: Helpful or Hollow? *PLOS Biology*, 13(12), e1002311. <https://doi.org/10.1371/journal.pbio.1002311>

Motta, E. V. S., & Moran, N. A. (2023). The honeybee microbiota and its impact on health and disease. *Nature Reviews Microbiology*, 1–16. <https://doi.org/10.1038/s41579-023-00990-3>

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P., & Ferrenberg, S. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews : MMBR*, 77(3), 342–356. <https://doi.org/10.1128/MMBR.00051-12>

Neu, A. T., Allen, E. E., & Roy, K. (2021). Defining and quantifying the core microbiome: Challenges and prospects. *Proceedings of the National Academy of Sciences*, 118(51), e2104429118. <https://doi.org/10.1073/pnas.2104429118>

Ojima, M. N., Jiang, L., Arzamasov, A. A., Yoshida, K., Odamaki, T., Xiao, J., Nakajima, A., Kitaoka, M., Hirose, J., Urashima, T., Katoh, T., Gotoh, A., van Sinderen, D., Rodionov, D. A., Osterman, A. L., Sakanaka, M., & Katayama, T. (2022). Priority effects shape the structure of infant-type *Bifidobacterium* communities on human milk oligosaccharides. *The ISME Journal*, 16(9), 2265–2279. <https://doi.org/10.1038/s41396-022-01270-3>

Olm, M. R., Crits-Christoph, A., Diamond, S., Lavy, A., Matheus Carnevali, P. B., & Banfield, J. F. (2020). Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems*, 5(1), e00731-19. <https://doi.org/10.1128/mSystems.00731-19>

Orr, J. A., Armitage, D. W., & Letten, A. D. (2025). Coexistence Theory for Microbial Ecology, and Vice Versa. *Environmental Microbiology*, 27(3), e70072. <https://doi.org/10.1111/1462-2920.70072>

Palmer, M. W. (1994). Variation in species richness: Towards a unification of hypotheses. *Folia Geobotanica et Phytotaxonomica*, 29(4), 511–530. <https://doi.org/10.1007/BF02883148>

Powell, J. E., Martinson, V. G., Urban-Mead, K., & Moran, N. A. (2014). Routes of Acquisition of the Gut Microbiota of the Honey Bee *Apis mellifera*. *Applied and Environmental Microbiology*, 80(23), 7378–7387. <https://doi.org/10.1128/AEM.01861-14>

Prasad, A., Pallujam, A. D., Siddaganga, R., Suryanarayanan, A., Mazel, F., Brockmann, A., Yek, S. H., & Engel, P. (2025). *Symbiont loss and gain, rather than co-diversification shapes honeybee gut microbiota diversity and function* (p. 2024.09.11.612390). bioRxiv. <https://doi.org/10.1101/2024.09.11.612390>

Raulo, A., Bürkner, P.-C., Finerty, G. E., Dale, J., Hanski, E., English, H. M., Lamberth, C., Firth, J. A., Coulson, T., & Knowles, S. C. L. (2024). Social and environmental transmission spread different sets of gut microbes in wild mice. *Nature Ecology & Evolution*, 8(5), 972–985. <https://doi.org/10.1038/s41559-024-02381-0>

Rühlemann, M. C., Bang, C., Gogarten, J. F., Hermes, B. M., Groussin, M., Waschyna, S., Poyet, M., Ulrich, M., Akoua-Koffi, C., Deschner, T., Muyembe-Tamfum, J. J., Robbins, M. M., Surbeck, M., Wittig, R. M., Zuberbühler, K., Baines, J. F., Leendertz, F. H., & Franke, A. (2024). Functional host-specific adaptation of the intestinal microbiome in hominids. *Nature Communications*, 15(1), Article 1. <https://doi.org/10.1038/s41467-023-44636-7>

Sanders, J. G., Sprockett, D. D., Li, Y., Mjungu, D., Lonsdorf, E. V., Ndjango, J.-B. N., Georgiev, A. V., Hart, J. A., Sanz, C. M., Morgan, D. B., Peeters, M., Hahn, B. H., & Moeller, A. H. (2023). Widespread extinctions of co-diversified primate gut bacterial symbionts from humans. *Nature Microbiology*, 8(6), Article 6. <https://doi.org/10.1038/s41564-023-01388-w>

Sarton-Lohéac, G., Nunes da Silva, C. G., Mazel, F., Baud, G., de Bakker, V., Das, S., El Chazli, Y., Ellegaard, K., Garcia-Garcera, M., Glover, N., Liberti, J., Nacif Marçal, L., Prasad, A., Somerville, V., SAGE class 2019-2020 and 2020-2021, Bonilla-Rosso, G., & Engel, P. (2023). Deep Divergence and Genomic Diversification of Gut Symbionts of Neotropical Stingless Bees. *mBio*, 0(0), e03538-22. <https://doi.org/10.1128/mbio.03538-22>

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D. R., Kultima, J. R., Martin, J., Kota, K., Sunyaev, S. R., Weinstock, G. M., & Bork, P. (2013). Genomic variation landscape of the human gut microbiome. *Nature*, 493(7430), 45–50. <https://doi.org/10.1038/nature11711>

Segura Munoz, R. R., Mantz, S., Martínez, I., Li, F., Schmaltz, R. J., Pudlo, N. A., Urs, K., Martens, E. C., Walter, J., & Ramer-Tait, A. E. (2022). Experimental evaluation of ecological principles to understand and modulate the

outcome of bacterial strain competition in gut microbiomes. *The ISME Journal*, 16(6), 1594–1604. <https://doi.org/10.1038/s41396-022-01208-9>

Shade, A., & Handelsman, J. (2012). Beyond the Venn diagram: The hunt for a core microbiome. *Environmental Microbiology*, 14(1), 4–12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x>

Sharma, P. K., Kundu, B. S., & Dogra, R. C. (1993). Molecular mechanism of host specificity in legume-rhizobium symbiosis. *Biotechnology Advances*, 11(4), 741–779. [https://doi.org/10.1016/0734-9750\(93\)90002-5](https://doi.org/10.1016/0734-9750(93)90002-5)

Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease: *Current Opinion in Gastroenterology*, 31(1), 69–75. <https://doi.org/10.1097/MOG.0000000000000139>

Smith, D. R. (2020). Biogeography of Honey Bees. In C. K. Starr (Ed.), *Encyclopedia of Social Insects* (pp. 1–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-90306-4_60-1

Song, S. J., Sanders, J. G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., Mazel, F., Lutz, H. L., Winker, K., Graves, G. R., Humphrey, G., Gilbert, J. A., Hackett, S. J., White, K. P., Skeen, H. R., Kurtis, S. M., Withrow, J., Braile, T., Miller, M., ... Knight, R. (2020). Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. *mBio*, 11(1), 10.1128/mbio.02901-19. <https://doi.org/10.1128/mbio.02901-19>

Sprockett, D., Fukami, T., & Relman, D. A. (2018). Role of priority effects in the early-life assembly of the gut microbiota. *Nature Reviews Gastroenterology & Hepatology*, 15(4), 197–205. <https://doi.org/10.1038/nrgastro.2017.173>

Suzuki, T. A., Fitzstevens, L., Schmidt, V. T., Enav, H., Huus, K., Mbong, M., Adegbite, B. R., Zinsou, J. F., Esen, M., Velavan, P., Adegnika, A. A., Song, L. H., Spector, T. D., Marchi, N., Blekhman, R., Ségurel, L., Youngblut, N. D., Kremsner, P., & Ley, R. E. (2021). *Codiversification of gut microbiota with humans*.

Van Rossum, T., Ferretti, P., Maistrenko, O. M., & Bork, P. (2020). Diversity within species: Interpreting strains in microbiomes. *Nature Reviews Microbiology*, 18(9), 491–506. <https://doi.org/10.1038/s41579-020-0368-1>

Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyriides, N. C., & Pati, A. (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Research*, 43(14), 6761–6771. <https://doi.org/10.1093/nar/gkv657>

Vellend, M., & Agrawal, A. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85(2), 183–206. <https://doi.org/10.1086/652373>

Viver, T., Conrad, R. E., Rodriguez-R, L. M., Ramírez, A. S., Venter, S. N., Rocha-Cárdenas, J., Llabrés, M., Amann, R., Konstantinidis, K. T., & Rossello-Mora, R. (2024). Towards estimating the number of strains that make up a natural bacterial population. *Nature Communications*, 15(1), 544. <https://doi.org/10.1038/s41467-023-44622-z>

Vrancken, G., Gregory, A. C., Huys, G. R. B., Faust, K., & Raes, J. (2019). Synthetic ecology of the human gut microbiota. *Nature Reviews Microbiology*, 17(12), 754–763. <https://doi.org/10.1038/s41579-019-0264-8>

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

Wennekes, P. L., Rosindell, J., & Etienne, R. S. (2012). The Neutral—Niche Debate: A Philosophical Perspective. *Acta Biotheoretica*, 60(3), 257–271. <https://doi.org/10.1007/s10441-012-9144-6>

Whitfield, C. W., Behura, S. K., Berlocher, S. H., Clark, A. G., Johnston, J. S., Sheppard, W. S., Smith, D. R., Suarez, A. V., Weaver, D., & Tsutsui, N. D. (2006). Thrice Out of Africa: Ancient and Recent Expansions of the Honey Bee, *Apis mellifera*. *Science*, 314(5799), 642–645. <https://doi.org/10.1126/science.1132772>

Whitfield, J. (2002). Neutrality versus the niche. *Nature*, 417(6888), 480–481. <https://doi.org/10.1038/417480a>

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, 51(2), 221–271. <https://doi.org/10.1128/mr.51.2.221-271.1987>

Wolff, R., Shoemaker, W., & Garud, N. (2021). Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome [Preprint]. *Ecology*. <https://doi.org/10.1101/2021.09.30.462616>

Young, A. M., Kodabalagi, S., Brockmann, A., & Dyer, F. C. (2021). A hard day's night: Patterns in the diurnal and nocturnal foraging behavior of *Apis dorsata* across lunar cycles and seasons. *PLOS ONE*, 16(10), e0258604. <https://doi.org/10.1371/journal.pone.0258604>

Zheng, H., Steele, M. I., Leonard, S. P., Motta, E. V. S., & Moran, N. A. (2018). Honey bees as models for gut microbiota research. *Lab Animal*, 47(11), 317–325. <https://doi.org/10.1038/s41684-018-0173-x>

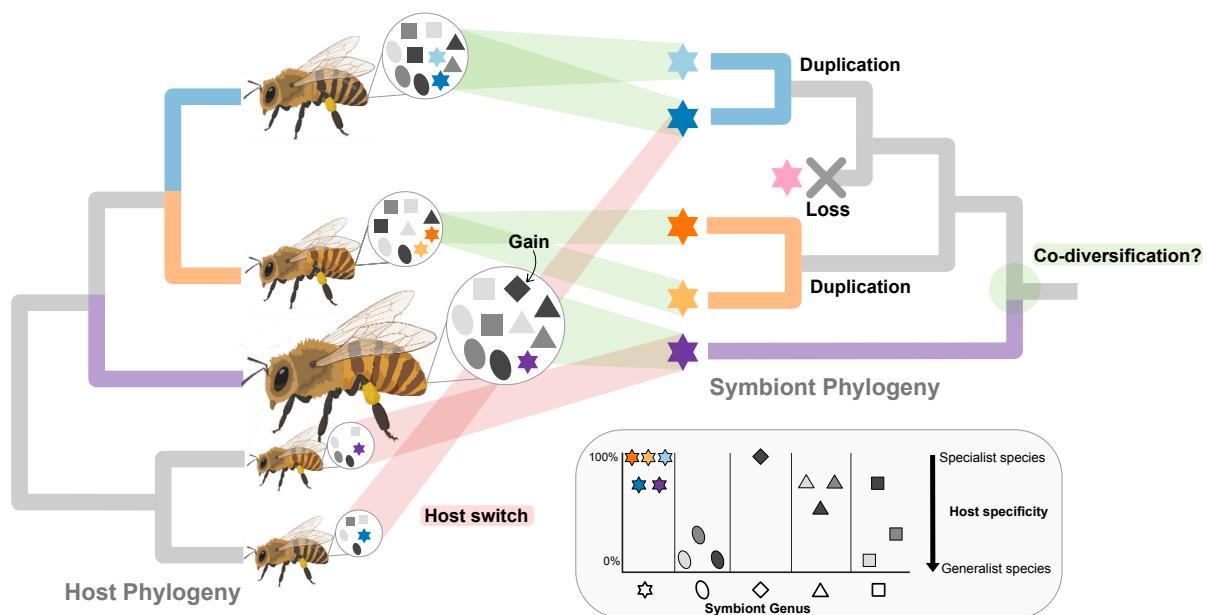
Zhou, J., & Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews*, 81(4), 10.1128/mmbr.00002-17. <https://doi.org/10.1128/mmbr.00002-17>

CHAPTER-2

"... a genome-centered approach to metagenomic data is extraordinarily more appealing than a read-based approach. Genes are expressed within cells, not in a homogenized cytoplasmic soup."

- Katherine McMahon, Metagenomics 2.0
(McMahon, 2015)

SYMBIONT LOSS AND GAIN, RATHER THAN CO-DIVERSIFICATION, SHAPES HONEYBEE GUT MICROBIOTA DIVERSITY AND FUNCTION



EVOLUTION OF GUT MICROBIOTA ACROSS HONEYBEE SPECIES REVEALED BY COMPARATIVE METAGENOMICS

Aiswarya Prasad¹, Asha D Pallujam², Rajath Siddaganga³, Ashwin Suryanarayanan³, Florent Mazel¹, Axel Brockmann³, Sze H Yek^{2,4} and Philipp Engel¹

¹Department of Fundamental Microbiology, University of Lausanne, Switzerland, ²School of Science, Monash University Malaysia, Malaysia, ³National Centre for Biological Sciences-Tata Institute of Fundamental Research, Bengaluru, Karnataka, India, ⁴Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Malaysia

SUMMARY

Studying gut microbiota evolution across animals is crucial for understanding symbiotic interactions but is hampered by the lack of high-resolution genomic data. Honeybees, with their specialized gut microbiota and well-known ecology, offer an ideal system to study this evolution. Using shotgun metagenomics on 200 worker bees from five honeybee species, we recovered thousands of metagenome-assembled genomes, identifying several novel bacterial species. While microbial communities were mostly host-specific, we found both specialists and generalists, even among closely related bacterial species, with notable variation between honeybee hosts. Some bacterial generalists emerged host-specific only at the strain level, suggesting recent host switches. While we found some evidence of co-diversification between hosts and symbionts, this was much less pronounced than what has been observed for gut bacteria of hominids and small mammals. Instead, symbiont gains, losses, and replacements emerged as important factors for honeybees. This highly dynamic evolution of the specialized honeybee gut microbiota has led to taxonomic and functional differences across hosts, such as the ability to degrade pollen-derived pectin. Our results provide new insights into the evolutionary processes govern gut microbiota diversity across closely related hosts and uncover the functional potential of the previously underexplored gut microbiota of these important pollinators.

CONTRIBUTION

I contributed to refining the research questions, co-developed the strategy for sampling honeybees, and created SOPs (Standard Operating Procedures for sampling, storage, transport and DNA extractions) to maintain consistency across locations. I led a month-long field trip for sample collection, processing, and sequencing in India. Subsequently, I carried out the processing and metagenomic sequencing of all the honeybees sampled in Malaysia by my collaborators. Finally, set up the bioinformatic pipeline and wrote scripts to analyze and visualize the data. I led the drafting of the manuscript, which has since been developed based on the advice and contribution of my supervisors and co-authors.

PUBLICATION STAGE

This chapter is presented in the form of a manuscript posted on BioRxiv and submitted to the peer-reviewed journal, Nature Communications, where, after following rounds of revisions, it is currently under review

INTRODUCTION

Animals harbor specialized microbial communities in their gut, which can influence health and disease susceptibility through diverse functional capabilities (Levin et al., 2021; McFall-Ngai et al., 2013). These communities are often host-specific, where the gut microbiota of individuals of the same host species are more similar in composition to each other than to those of different host species (Mallott & Amato, 2021). In some animal groups, it has also been observed that more closely related host species harbor more similar gut microbiota in composition, a pattern referred to as phylosymbiosis (Lim & Bordenstein, 2020). This host specificity at the community level arises from the host-restricted distribution of individual community members, which in turn is driven by microbial dispersal limitation of microbes and/or host-filtering processes (Mallott & Amato, 2021; Mazel et al., 2018).

Some gut bacterial clades – for example in termites (Arora et al., 2023) and hominids (Good, 2023; Moeller et al., 2016; Rühlemann et al., 2024; Sanders et al., 2023; Suzuki et al., 2022) – exhibit strong host specificity, with phylogenies that closely mirror those of their hosts. Such phylogenetic congruence (or co-phylogeny) can result from co-diversification (i.e., repeated co-speciation), in which hosts and symbionts undergo parallel speciation events, often driven by vertical inheritance from parent to offspring. However, current studies indicate that only a small fraction of the analyzed bacterial lineages show evidence of co-diversification. In contrast, numerous examples of generalist gut symbionts have also been documented (Groussin et al., 2017; Song et al., 2020). This variation in the distribution of different gut bacteria suggests that these microbes do not follow the same evolutionary trajectory across various animal species. Indeed, a recent study (Mazel et al., 2024) showed that the degree of host specificity differs across bacterial lineages in the animal gut, with both generalists and specialists present. Notably, this variation cannot be fully explained by microbial phenotype alone (Raulo et al., 2024), highlighting a possible role of the host's ecology and environment in shaping these associations (Mallott & Amato, 2021). This raises the question of why generalist and specialist gut bacteria exist and whether specialist bacteria primarily evolved via repeated switches between hosts or through co-diversification with hosts.

Our current understanding of the evolution and diversification of gut microbial communities across hosts is limited by at least four technical and conceptual factors. (1) Previous studies have focused on highly diverse and variable systems across host individuals (Douglas, 2019), making it challenging to understand whether gut communities change by acquisition of new members or by *in situ* evolution. (2) Significant differences in the relatedness and ecology of the compared hosts hinder the identification of the drivers of community change over evolutionary time scales and the prediction of their functional consequences. (3) Samples from different animals often come from different geographic regions, making geography an important confounding factor that could explain the host-restricted distribution observed for some gut symbionts (Moeller, 2025; Rühlemann et al., 2024). (4) Most comparative studies are based on 16S rRNA gene analysis, which lacks the resolution to determine the distribution of closely related bacterial species, strains and functional gene content across hosts (Nayfach & Pollard, 2016).

Comparative studies of the microbiota of closely related animals with similar phenotypes and geographic distribution at high taxonomic resolution are needed to overcome these limitations. The gut microbiota of honeybee workers is a promising model offering the combined advantage of a simple yet specialized microbial community shared among related host species with similar ecology, social behavior, and overlapping geographic ranges (Engel et al., 2012; Hepburn & Radloff, 2011; Mazel et al., 2025). Prominent honeybee gut microbiota members include *Lactobacillus*, *Bombilactobacillus*, *Bifidobacterium*, and several proteobacteria. These bacterial lineages from honeybees are usually monophyletic but have diversified into host-specific species, several of which can coexist within the same host individual (Ellegaard et al., 2015, 2019; Sarton-Lohéac et al., 2023). Thus, the predominant members of the honeybee gut microbiota have likely been acquired in a common ancestor of social bees and subsequently diversified in honeybees by separation into different host species and distinct ecological niches in the gut (Kwong et al., 2017; Y. Li et al., 2022; Zhang et al., 2022). Co-diversification has been suggested as an important mode of evolution (Engel et al., 2012; Koch et al., 2013; Kwong et al., 2017; Luo et al., 2024; J. Wu et al., 2021), but formal tests across the major community members have not yet been conducted (Mazel et al., 2025). So far, most comparative studies are based on either 16S rRNA gene amplicon sequencing with low genetic resolution or relatively few isolate genomes biased towards isolates from the managed Western honeybee *Apis mellifera* (Ganeshprasad et al., 2021; Ge et al., 2021; Gruneck et al., 2021, 2022; Khan et al., 2023; Kwong et al., 2017, p. 201; Niode et al., 2021; J. Wu et al., 2021). Both approaches provide limited quantitative insights into the distribution of divergent strains and species across hosts. Deep shotgun metagenomics allows the sequencing of most genomes in a given sample, characterization of the distribution of community members at high resolution (down to the strain level) across hosts, and assessment of the functional consequences of differences among hosts. However, this approach has only been used to study the Western honeybee (*Apis mellifera*) (Ellegaard & Engel, 2019; Engel et al., 2012; Regan et al., 2018; J. Wu et al., 2021) and the Asian honeybee (*Apis cerana*) (Ellegaard et al., 2020; Y. Wu et al., 2022).

In this study, we applied shotgun metagenomics to individual worker bees of the five most prevalent species of honeybees described so far, and spanning all three major clades of honeybees, namely *Micrapis* (dwarf honeybee): *A. florea* and *A. andreniformis*, *Megapis* (giant honeybee): *A. dorsata*, *Apis*: *A. mellifera* and *A. cerana*. While *A. mellifera* is globally distributed due to human introduction, the other honeybee species are mainly found across Asia (Hepburn & Radloff, 2011). All five honeybee species share several traits, such as a complex social lifestyle and a diet mainly consisting of processed pollen and nectar. However, they differ in certain key aspects, including body and colony size, nesting habit (open versus cavity nesting) and migrating ability (Fouks et al., 2021).

From a total of 200 shotgun metagenomes from the gut of individual worker bees, sampled across Peninsular Malaysia and South India, we generated a comprehensive dataset of metagenome-assembled genomes (MAGs), expanding the catalog of the bacterial genomes available from these important pollinators and model for microbiome research. We reveal striking differences in host specificity across host and bacterial species and identify previously unrecognized but prevalent members of the bee gut microbiota that expand the functional repertoire of these communities in specific host species. Overall, our results show that the

evolution of specialized gut microbiota-host interactions in honeybees has been shaped by symbiont loss and gain, and co-diversification.

RESULTS

AN EXPANDED CATALOG OF BACTERIAL GENOMES FROM THE GUT OF FIVE HONEYBEE SPECIES

We shotgun sequenced ($35M \pm 19M$ Illumina paired-end reads per sample) the hindguts of 200 individual worker bees from 40 different colonies of five honeybee species sampled in South India and various locations across peninsular Malaysia (Fig. 1A, Supplementary Table S1, Supplementary Fig. S1).

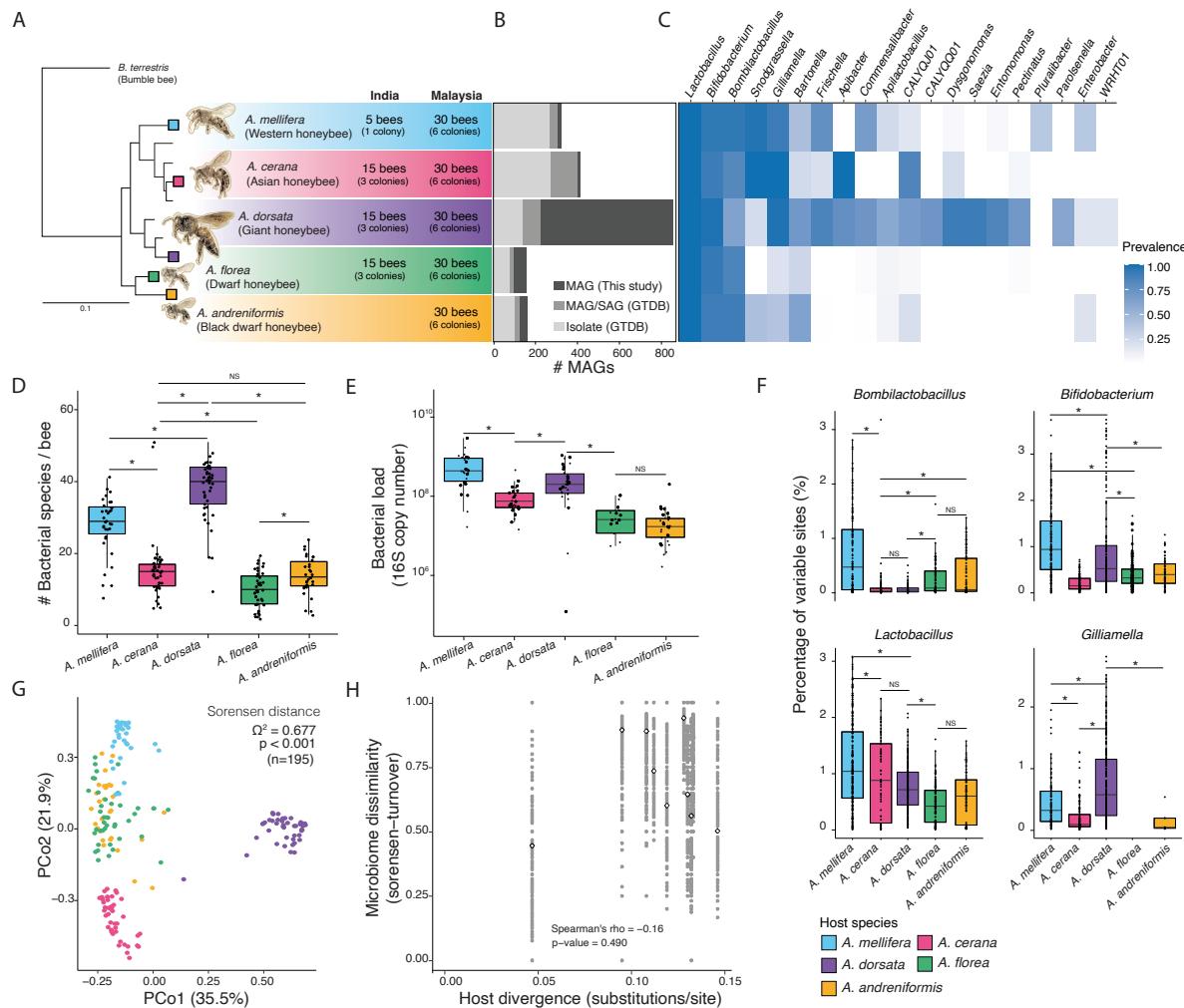


Figure 1. Community-level patterns in the honeybee gut microbiome across taxonomic scales. (A) Apis bee phylogeny with tips indicating extant species and the colored tips species included in this study. Pictures of bees (Kotthoff et al., 2011) are to-scale relative to each other. Numbers and text next to species names indicate the sampling design. (B) Barplot showing the number of MAGs recovered from this study for which no isolate genomes of the same species (>95% ANI) were found in GTDB, those that were represented by MAGs or SAGs in GTDB and those that are similar to published genomes from isolated bacteria represented in GTDB. (C) Heatmap with the prevalence of each genus within each host species showing genera with prevalence > 10% excluding the following three: *Klebsiella* (*K. pneumonia*) – 10% in *A. mellifera*; *Pantoea* (*P. dispersa*) – 17% in *A. mellifera*; *Melissococcus* (*M. plutonius*) – 13% in *A. andreniformis*). (D) Boxplot of the number of bacterial species in each honeybee gut across host species. (E) Boxplot of the number of 16S rRNA gene copies per $0.1\mu\text{L}$ of individual bee gut homogenate across host species. (F) Boxplot of the percentage of variable sites per individual of each host species shown for four most prevalent genera. Each point represents the percentage of variable sites found

in each bacteria species of the genus within each individual of that host species. (G) PCoA plot based on the Sorenson distance across host species. (H) Scatter plot of microbiota dissimilarity (as the turnover component of Sorenson distance) as a function of host species divergence based on 16S+12S rRNA (Ghonche-Golan et al., 2022) with white points with thick black outlines marking the median of values for the comparisons of all pairs of samples from a given pair of host species.

We recovered a total of 1,959 high- and medium-quality metagenome-assembled genomes (MAGs, i.e. completeness score > 50% and contamination <5%) (**Supplementary Fig. S2**), including 156 from *A. florea*, 163 from *A. andreniformis*, 869 from *A. dorsata*, 324 from *A. mellifera* and 413 from *A. cerana*) (**Fig. 1B, Supplementary Table S2**). The MAGs were discriminated into 150 species spanning 23 genera (prevalence > 10%) using 95% ANI (average nucleotide identity) clustering (**Supplementary Fig. S3, Supplementary Fig. S4**). A majority of these species belonged to genera previously reported to be predominant in the gut microbiota of honeybees, including *Lactobacillus* (formerly Firm5), *Bombilactobacillus* (formerly Firm4), *Bifidobacterium*, *Gilliamella*, *Snodgrassella*, *Frischella*, *Bartonella* and *Apibacter* (constituted by 91 / 150 species clusters), but also lesser known genera such as *Dysgonomonas* (family Dysgonomonadaceae), *Saezia* (family Burkholderiaceae), *Pectinatus* (family Selenomonadaceae), *Entomomonas* (family Pseudomonadacea), *Pluralibacter* (family Enterobacteriaceae), WRHT01 (Desulfovibrionaceae family), CALYQQ01 (family Enterobacteriaceae) and CALYQJ01 (family Lactobacillaceae) (constituted by 21 / 150 species). Around 47% of the identified bacterial species (71 / 150 species) have not been cultured to date, with many belonging to the predominant genera (41 / 71 species) and with genomic information currently lacking in public databases (**Fig. 1B**). The established collection of MAGs massively expands the catalog of bacterial genomes from the gut environment of these underexplored honeybee species.

MOST HONEYBEE SPECIES HARBOR HOST-SPECIFIC MICROBIAL COMMUNITIES IN THEIR GUT

We assessed the taxonomic diversity of the gut microbiota across the five honeybee species using our collection of MAGs. Between 7 - 56M reads from each sample mapped to these MAGs (**Supplementary Table S3**), providing sufficient depth of bacterial reads for MAG recovery and bacterial species detection from individual guts (**Supplementary Fig. S5**). Rarefaction curves show that our sampling effort from different colonies of each honeybee species (**Fig. 1A**) was sufficient to cover most of the species-level diversity in each honeybee species (**Supplementary Fig. S6**).

Prevalence of the identified genera varied greatly across host species (**Fig. 1C**). *Lactobacillus*, *Bombilactobacillus*, and *Bifidobacterium* were consistently present across all five honeybee species (**Fig. 1B**). All other genera were only found in a subset of the five host species or occurred sporadically throughout. Notably, samples of *A. dorsata* contained several additional genera in high prevalence, such as *Dysgonomas* (90.9%), *Pectinatus* (70.5%), and *Saezia* (90.9%), which were rarely found in the other host species.

Comparison of the species-level bacterial diversity revealed that *A. dorsata* contained the most diverse microbiota, followed by *A. mellifera*, *A. cerana* and then the two dwarf honeybees *A. florea* and *A. andreniformis* (**Fig. 1D**). This pattern holds when considering only species of the shared genera (i.e., genera with a prevalence of at least 10% in each host),

except that *A. dorsata* and *A. mellifera* were not significantly different from each other (**Supplementary Fig. S7**). Interestingly, the bacterial biomass per individual gut was the highest in *A. mellifera*, followed by *A. dorsata*, *A. cerana* and the dwarf bees, and hence, body size cannot necessarily explain the differences in alpha-diversity (**Fig. 1E**). Most bacterial species displayed significant strain level diversity as measured by the percentage of single nucleotide variants (SNVs) per bacterial species within each bee gut with the possible co-occurrence of several closely related strains (**Supplementary Fig. S8**). The Western honeybee *A. mellifera* hosted significantly more strains than other hosts for most bacterial species, while the dwarf honeybees harbored the least strain level diversity (**Fig. 1F**).

Taxonomic composition (**Supplementary Fig. S9-10**) was host-specific at the species level as suggested by β -diversity analyses (PERMANOVA Sorenson distance $\Omega^2 = 0.6777$, $p=0.001$, $N=195$), but with no distinction between the two dwarf honeybee species (**Fig. 1G**) and no influence of geography ($\Omega^2=0.009$, $p=0.086$, $N=131$ excluding two host species - *A. andreniformis* and *A. mellifera* - sampled unequally across locations) (**Supplementary Fig. S11**). We found no evidence for phylosymbiosis, as there was no correlation between bacterial community distance and host divergence (Spearman's p -value = 0.49, mean q -value = 0.619 ± 0.204 across 1000 Mantel tests, **Fig. 1H**). For example, *A. andreniformis* and *A. mellifera* are the most distant host species, but have much smaller distances in bacterial community composition than other pairs of host species.

HOST SPECIFICITY OF INDIVIDUAL COMMUNITY MEMBERS VARIES ACROSS BACTERIA AND HOST SPECIES

Specificity at the community level emerges because of the host-specific distribution of individual community members (Mazel et al., 2024). We calculated the specificity per bacterial species based on Rohde's Index (Rohde & Rohde, 2008), which considers the prevalence of each community member across samples from different hosts (**Fig. 2A, Supplementary Table S4**).

Overall, we found large differences in the specificity of individual bacterial species both within and between genera. For example, the genus *Bombilactobacillus* mainly consisted of generalist species found across all five honeybee species, while the genera *Lactobacillus* or *Bifidobacterium* contained specialists and generalists (**Fig. 2A**). Some bacterial species showed an intermediate level of specificity as they were shared only among a subset of the five host species and were more prevalent in one host than another (e.g., one of the *Snodgrassella* species). Few shared genera consisted of only specialist species (*Apibacter* and *Commensalibacter*). Further, there was a clear difference in the specificity of bacterial species between hosts: the microbiota of *A. dorsata* and *A. mellifera* were composed of both generalist and specialist species, while the microbiota of the other smaller honeybee species (*A. cerana*, *A. florea*, and *A. andreniformis*) was primarily composed of generalist species (**Fig. 2A**).

Bacterial species may be shared between host species but exhibit host specificity at the strain level. To look for such strain-level host specificity, we measured the extent of shared SNVs between samples for each bacterial species shared between two or more host species using popANI (Olm et al., 2021). We found both bacterial species that segregate and species that

do not segregate by host species in terms of strain-level composition (**Fig. 2B**, **Fig. 2C**). This pattern was observed across all analyzed genera (**Fig. 2B**, **Fig. 2C**). Moreover, in some cases, we also found that samples segregated by sampling location (**Fig. 2B**, **Fig. 2C**). Overall, 75% of the species were host-specific at either strain or species level.

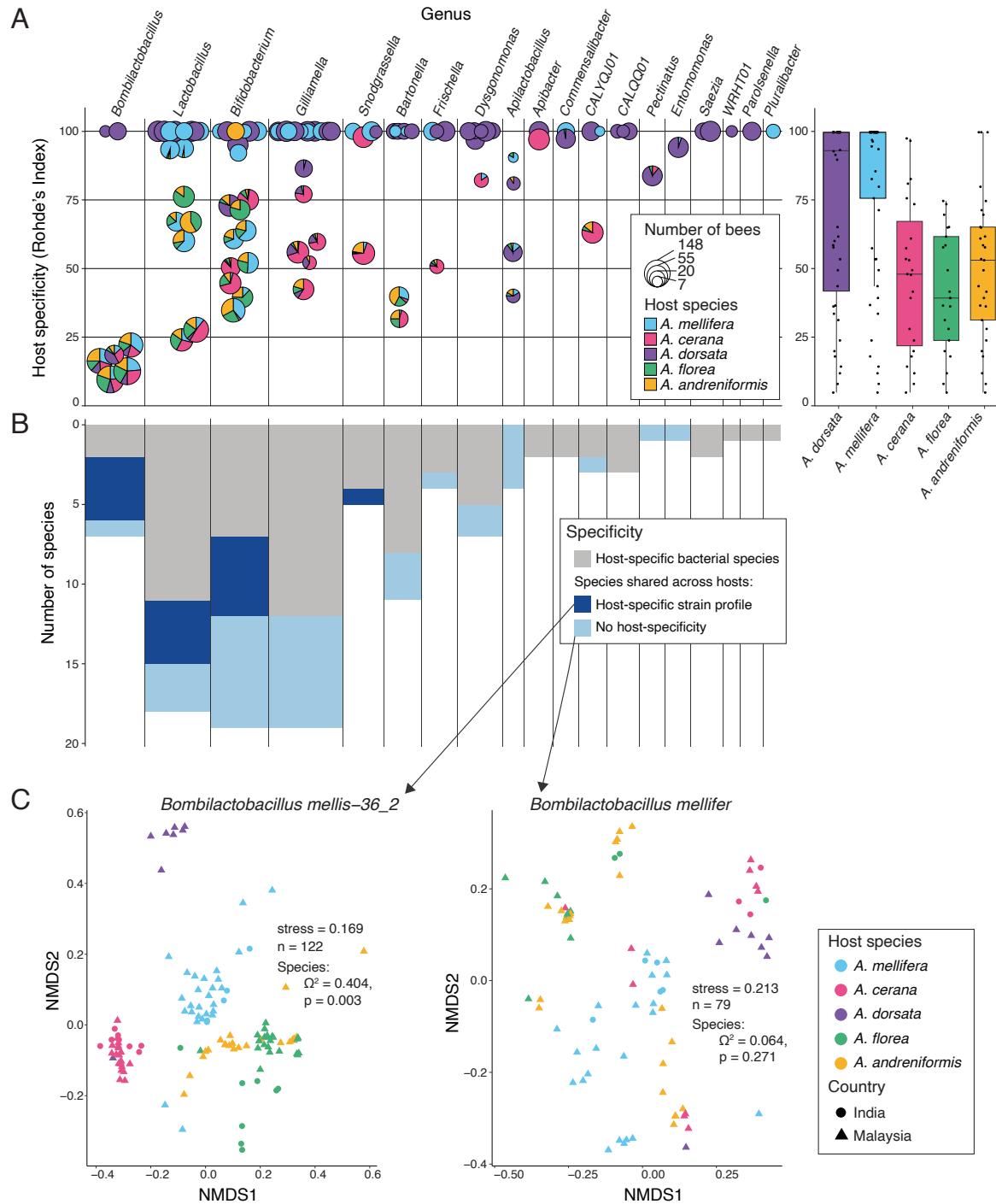


Figure 2. Host-specificity of individual gut microbial lineages across honeybee hosts. (A) Scatter-pie plot of Rohde's index of specificity with each circle representing a bacterial species with the size representing the number of bees in which the species was found, and the pie chart showing the proportion of samples belonging to each honeybee host. The adjacent boxplot shows Rohde's specificity index for each species found in each honeybee host. (B) Distribution of species richness per genus across three categories of specificity levels: host-specific at the species level (grey), host-specific at the strain level, but the species is detected across hosts (dark blue) or not

specific even at the strain level (light blue). (C) NMDS plots from a strain level generalist and specialist bacterial species are included as an example. The Jaccard dissimilarity was calculated in terms of strain level composition (popANI) tested using PERMANOVA (significant if the p-value is < 0.05). Only bacterial species with enough coverage to recover SNPs in at least five samples spanning at least two host species were included.

HOST SWITCHES AND SPECIES GAINS AND LOSSES INTERRUPTED THE HOST-SPECIFIC EVOLUTION OF GUT MICROBIOTA MEMBERS OF HONEYBEES

Co-diversification has been described in other host-associated microbiomes as an important driver of host specificity patterns (Good, 2023; Sanders et al., 2023; Suzuki et al., 2022). To test for evidence of co-diversification in the honeybee gut microbiota, we measured congruency between host and bacterial phylogenies (or co-phylogenetic patterns). In case of strict co-diversification one would expect the bacterial and host phylogeny to mirror each other at all nodes (see Fig. 3A for a scheme).

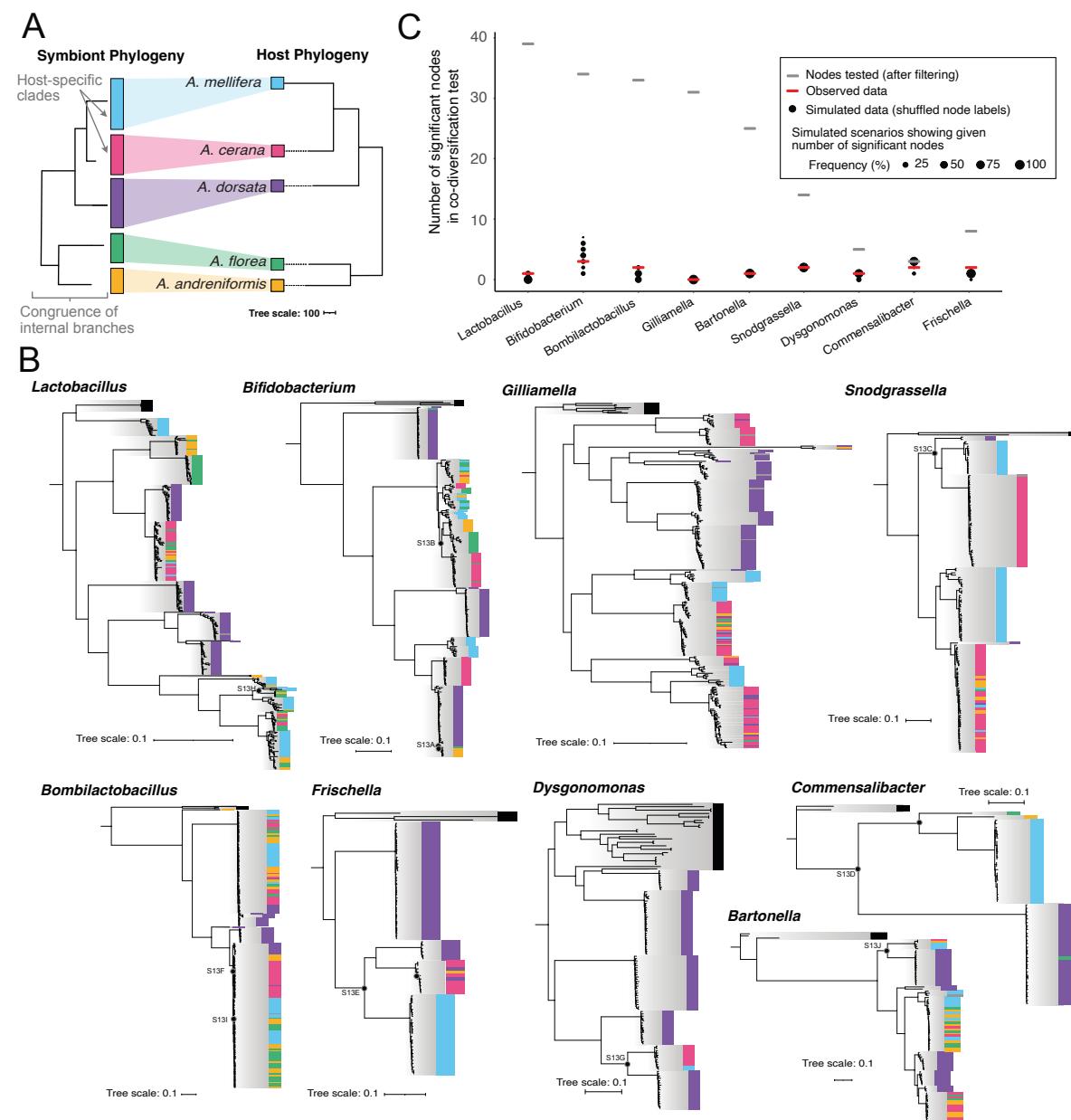


Figure 3. Bacterial genus-level core genome phylogenies of MAGs. (A) A tanglegram of a theoretical phylogeny of symbionts (left) showing perfect congruence with the host tree of *Apis* (right). The host tree is based on actual data and to scale according to source (adapted from: (Carr, 2023)). The scale bar is the inferred number of nucleotide substitutions. Signal for topological congruence between symbiont and host phylogeny in the Hommola test of co-phylogeny can result from two effects, as indicated: (i) congruence at (internal) branches separating clades of symbionts from different hosts, and (ii) clustering of symbionts into host-specific clades (without congruence at internal nodes separating them). (B) Core genome phylogenies of MAGs from the nine most prevalent genera found in more than one host species. Colors at the leaf tip indicate the host species from which the MAG was recovered and solid grey boxes depict the set of genomes included to represent outgroup species of the genus from environments other than honeybee guts. The scale bar represents 0.1 amino-acid substitutions per site. The phylogenies are maximum likelihood trees inferred by IQ-TREE (see Methods). Black filled circles indicate subtrees of nodes that were significant in the Hommola test for co-phylogeny. These subtrees are further illustrated in Supplementary Fig. S12-S13. (C) Scatter plot representing the number of nodes significant (under strict threshold) for co-diversification in the actual node-by-node comparison (red line) compared with that in the second-order permutation test comparisons ($N=100$) with host tree node labels shuffled. The other short horizontal line (grey) denotes the number of nodes compared after excluding those that were filtered (did not have at least 7 MAGs from at least 3 different host species).

We constructed core genome phylogenies for the nine most prevalent genera using the collection of all high- and medium-quality MAGs assembled from the five honeybee species, including several MAGs from each bacterial species within the genus. We used genus-level phylogenies to test for co-diversification because divergence time within genera was roughly equivalent to divergence times in the *Apis* genus (see methods). The resulting trees are depicted in **Fig. 3B**, but can also be inspected interactively online (<https://itol.embl.de/shared/Aiswarya>). Most honeybee-derived MAGs from a given bacterial genus formed clades closely related to those from bumble bees and stingless bees, but segregated into distinct species-level clusters based on ANI (**Fig. 3B** and **Supplementary Fig. S3**). In agreement with the observed differences in host specificity, 24%–47% of these species-level clades were exclusive to a single host species, while the remainder included MAGs from multiple hosts. To test for congruency between host and bacterial phylogenies, we used the Hommola cospeciation test, a generalization of the Mantel test for host-parasite interaction graphs (Hommola et al., 2009). We applied this test to all nodes containing at least 7 different MAGs from three or more host species. The Hommola test detects phylogenetic congruence driven by two main factors: (i) clustering of bacterial taxa (i.e., tree tips) by host species, and (ii) topological congruence of internal branches between bacterial and host phylogeny (see scheme in **Fig. 3A**). Only the latter provides evidence for true co-diversification across host species. However, the former can also produce low p-values and elevated R values, as noted in previous studies (Moeller et al., 2023; Sanders et al., 2023) and demonstrated by our evaluation of various thresholds (see **Supplementary Fig. S12**, **Supplementary Tables S5–S6**, and Methods). To minimize false positives, we adopted stringent significance criteria ($R > 0.75$ and $p < 0.01$), consistent with the thresholds used in prior work (Sanders et al., 2023). We then carried out second-order permutation tests ($n=100$) with shuffled tip labels to account for the rate of false discoveries as described in previously published work (Moeller et al., 2023) (see Methods for more details and an illustration of the approach). Overall, this analysis identified significant phylogenetic congruence between hosts and symbionts in only 13 of the 192 tested nodes/subtrees (**Fig. 3B** and <https://itol.embl.de/shared/Aiswarya>), which was no more than expected by chance in any of the nine tested genera as determined by our permutation tests (**Fig. 3C**, **Supplementary Tables S5–S6**). Visual inspection confirmed that a subset of the significant subtrees exhibited

topological congruence with the host phylogeny, consistent with co-diversification – for example, in subtrees of *Bifidobacterium* (**Supplementary Fig. S13A** and **S13B**), *Commensalibacter* (**Supplementary Fig. S13C**), or *Snodgrassella* (**Supplementary Fig. S13D**). In contrast, other significant subtrees, lacked topological congruence with the host tree (e.g. *Frischella* subtree in **Supplementary Fig. S13E**, or *Bombilactobacillus* subtree in **Supplementary Fig. S13F**), or showed internal branch lengths inconsistent with host divergence times (e.g. *Lactobacillus* subtree in **Supplementary Fig. S13G**, *Bombilactobacillus* subtree in **Supplementary Fig. S13H**, or *Bartonella* subtree in **Supplementary Fig. S13I**). Notably, none of the significant subtrees included symbiont clades from all five honeybee species, suggesting that these co-diversifying lineages were either lost in some hosts or not acquired in the common ancestor of the five honeybee species but in later stages.

In summary, we find relatively few nodes with clear evidence of co-diversification. Instead, as seen in **Fig. 3B**, deep-branching host-specific clades are frequently interrupted by generalist clades, indicating recent host switches. Likewise, deep-branching generalist clades – some of which exhibit host specificity at the strain level – often neighbor host-specific clades indicating recent host specialization that likely occurred after host divergence. Together, these patterns reflect a dynamic evolutionary history shaped by host switching and secondary specialization.

FUNCTIONAL DIFFERENCES IN THE GUT MICROBIOTA ACROSS HONEYBEES

To test whether the host specificity observed at the taxonomic level translates into functional host specificity of the honeybee gut microbiota, we performed a functional analysis of the MAGs. Of 4,759,622 prokaryotic ORFs in our MAG database, more than half (2,685,895 ORFs) were assigned a KO (KEGG Ortholog) identifier. The rarefaction curve analysis of these KOs indicated that most of the functional diversity across the analyzed samples was recovered with our MAG database (**Supplementary Fig. S14**). Overall, the number of KOs per sample followed a similar trend as the taxonomic diversity, with *A. dorsata* having the most and *A. florea* having the fewest KOs (**Fig. 4A**).

Multivariate analysis of normalized KO counts (and the Robust Aitchison distance) showed that the functional composition of the gut microbiota of five honeybee species was distinct, even between *A. florea* and *A. andreniformis*, which overlapped in composition at the taxonomic level ($\Omega^2=0.148$, $p=0.001$) (**Fig. 4B**). To understand which bacteria drive these differences, we performed a similar analysis using counts of all KO families in each high- and medium-quality MAG across all samples. As expected, MAGs of bacterial species from the same genus clustered across all host species, suggesting that they occupy similar functional niches and that the overall functional capacity of the bee microbiota is conserved across hosts (**Fig. 4C**). Yet, there were some important differences. Some of the genera only found in *A. dorsata* formed distinct clusters (e.g., *Dysgonomonas*), suggesting that these bacteria occupy unique functional niches in the microbiota of this host species. Other genera (e.g., *Saezia* and *Entomomonas*) clustered with community members prevalent in other honeybee species (e.g., *Snodgrassella*) but absent from *A. dorsata* (**Fig. 4C**). Closer inspection of their metabolic profiles indicated that both *Saezia* and *Entomomonas* like *Snodgrassella* lack major carbohydrate metabolism pathways (**Supplementary Fig. S15**), suggesting potential

functional replacements. Finally, although MAGs from the same genus mostly overlapped in terms of their functional profile, we found some cases where MAGs of the same bacterial genus clustered by honeybee species from which they were recovered, suggesting host-specific differences in the functional repertoire at the species-level (and potentially, adaptation) within these genera (**Supplementary Fig. S16**).

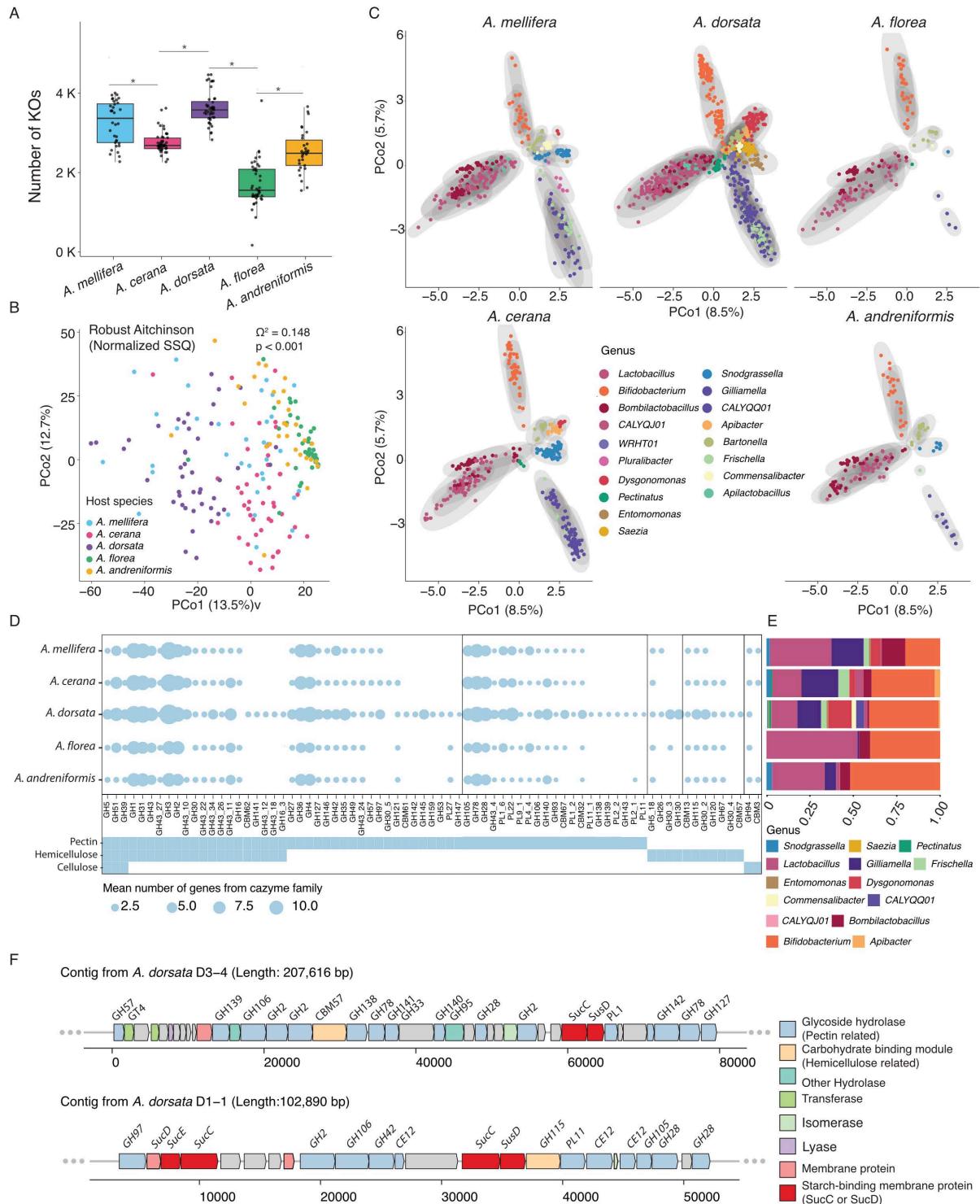


Figure 4. Functional potential of honeybee gut microbiota across host species. (A) Boxplot of the number of KEGG ortholog (KO) families detected per individual of each host species Wilcoxon test ($p < 0.05$) indicated on the plot. (B) PCoA of KO family composition across all

samples, visualized using the Robust Aitchison distance on a matrix of reads per kilobase per million (RPKM) summed across genes belonging to the respective KO family. Each point on the plot represents an individual honeybee sample and the color the host species identity of the sample. PERMANOVA test indicated on the plot tests the effect of host species identity. (C) PCoA of all KO families from each MAG using Robust-Aitchison distance on the RPKM matrix of KOs. Each point represents a bacterial MAG colored by the bacterial genus it belongs to. MAGs obtained from different host species were separated across panels for clearer visualization. The axes across the five panels have been maintained to the same limits. So, the panels can be superposed.. (D) Dotplot of the mean number of genes from each CAZyme family with a potential role in pectin, hemicellulose, or cellulose degradation per individual of each host species. (E) Barplot of the proportion of genes from CAZyme families contributed by each bacterial genus. (F) Loci within *Dysgonomonas* MAGs encoding putative polysaccharide utilization loci (PULs).

Since pollen-derived glycan (*i.e.*, pectin and hemicellulose) breakdown has been described as one of the characteristic catabolic functions of the gut microbiota of *A. mellifera* (Engel et al., 2012; Zheng et al., 2019), we specifically looked at the Carbohydrate-active enzyme (CAZyme) repertoire of the microbiota across the five host species. We detected 19,562 CAZyme families among the prokaryotic ORFs in our MAG database. Overall, the composition of these gene families was host-specific (**Supplementary Fig. S17**). While all five honeybee species harbored a similar set of CAZymes for pectin, hemicellulose, and cellulose degradation, *A. dorsata* contained a more extensive and diverse set of such gene families, especially many of which are involved in pectin degradation (**Fig. 4D**). To understand which bacteria drove these differences, we considered the number of genes belonging to CAZyme families from each bacterial species. Previously recognized members contributed most families to degrade pollen-derived glycans, specifically *Gilliamella* and *Bifidobacterium* (Engel et al., 2012; Zheng et al., 2019) (**Fig. 4E**), except in the giant honeybee *A. dorsata*, where the genus *Dysgonomonas* also contributed key enzyme families involved in pectin degradation such as PL1_2, PL11_1, GH28, GH78, GH138, or CE12 (**Fig. 4F, Supplementary Fig. S18**). Interestingly, the CAZyme genes of *Dysgonomonas* were often organized in large genomic islands reminiscent of specialized PULs (polysaccharide utilization loci) found in gut symbionts of mammals (Wardman et al., 2022), suggesting that this community member contributes to the capability of this particular honeybee species to degrade plant-derived fibers (**Fig. 4F**). Moreover, the fact that the MAGs of *Dysgonomonas* found in *Apis dorsata* form deep-branching clades, only distantly related to species isolated from cockroaches (Vera-Ponce de León et al., 2020) and human stool (Bilen et al., 2019), suggests host specialization.

DISCUSSION

Many animals harbor host-specific bacterial communities in their gut (Lim & Bordenstein, 2020; Mallott & Amato, 2021; Mazel et al., 2018, 2024). However, it is still unclear whether and why host specificity varies among closely related host species or bacterial lineages, and which evolutionary processes underlie the observed patterns of specificity and their functional consequences. Our comparative shotgun metagenomic analysis of the gut microbiota of five honeybee species provides novel insights into these questions. It advances our understanding of the evolutionary processes that govern gut microbiota composition across closely related animal species.

We find that the community composition of the gut microbiota of three (*A. mellifera*, *A. cerana*, *A. dorsata*) of the five analyzed honeybee species is host-specific. In contrast, the gut microbiota of the two dwarf honeybee species, *A. florea* and *A. andreniformis*, were indistinguishable from each other, suggesting that host specificity can vary between closely related host species.

Our genome-level data, coupled with our broad sampling design, allowed us to move beyond community-level patterns and to identify host specificity of individual community members. We found that the vast majority of the analyzed community members were host-specific, but also that host specificity emerges at different taxonomic levels, for example, at the genus (such as *Saezia* and *Dysgonomonas* in *A. dorsata*), species (such as many *Gilliamella*, *Snodgrassella* and *Lactobacillus* species), and strain level (such as *Bombilactobacillus mellis-36_2*). Considering that all of these bacteria belong to clades specific to social bees, this suggests that the host-specific association between different gut symbionts and their honeybee hosts has been established across different time points in their evolutionary history. These findings also highlight that patterns of host-specific associations can be overlooked without strain-level/genomic resolution.

Another key finding of our analysis is that the degree of host specificity varies across bacteria (i.e., there are generalist and specialist species). This has also been observed in other systems (Gile, 2024; Mazel et al., 2024; Raulo et al., 2024) and linked to bacterial traits, such as aerotolerance and spore-formation, that affect dispersal ability (Kearney et al., 2018; Mazel et al., 2024; Moeller et al., 2018; Tung et al., 2015; Valles-Colomer et al., 2023). However, unlike previous studies, we found unexpectedly high variability in host specificity even among closely related bacteria, i.e., species belonging to the same genus. This suggests that conserved microbial traits alone cannot fully account for variation in specificity patterns. Instead, our results point to the identity of the host species as a key determinant of bacterial host specificity. Most of the highly host-specific bacterial species were found in *A. dorsata* and *A. mellifera*, while the communities of *A. cerana* and the two dwarf honeybees were dominated by shared bacterial species. *A. mellifera* and *A. dorsata* guts may harbor more specific ecological niches (Engel & Moran, 2013) (a narrow range of conditions or resources that an organism can exploit) than the other bee species, facilitating the colonization and maintenance of additional ‘specialist’ bacteria. This is consistent with the observation that *A. cerana* and the two dwarf honeybees harbor significantly fewer bacterial species per individual bee than *A. dorsata* and *A. mellifera*. Several different factors could contribute to

these patterns. Ecological niches in the host gut can vary even among closely related host species due to differences in host physiology, diet, or gut structure. For example, the host-specific colonization of *S. alvi* of *A. mellifera* may be supported by specific organic acids secreted by the host into the gut lumen and utilized by this bacterium (Quinn et al., 2024). Slight differences in dietary preferences, e.g. the amount or type of pollen, may allow different combinations of species to co-exist in one host species compared to another (Brochet et al., 2021). Variation in body and colony size between *Apis* species may determine the carrying capacity and hence bacterial diversity in the gut, as proposed previously (Ellegaard et al., 2020; Kwong et al., 2017). Another possibility is that host species differ in the degree of control they exert over their gut microbiota - potentially through mechanisms such as immune regulation, as suggested in other animals (Wilde et al., 2024). In general, the molecular and ecological drivers of host specificity remain poorly understood. However, the host-specific clades and those showing evidence of recent host switches identified in this study offer promising targets for future research into these processes.

A central objective of our study was to uncover the evolutionary processes shaping the above described structure and distribution of gut symbiont diversity in honeybees. While we found limited evidence for strict co-diversification between gut bacteria and their hosts across the five analyzed *Apis* species, our findings point to an important role for symbiont gain and loss, as well as host-switching, in shaping these specialized gut communities. Several lines of evidence support this conclusion. First, only a small proportion of nodes in the genus-level phylogenies showed significant congruence with the host phylogeny - a pattern that would be expected in case of co-diversification. Second, some bacterial genera that are abundant in certain honeybee species are completely or nearly absent in others, suggesting the gain or loss of entire genera during host evolution - an observation also reported for the microbiota of closely related stingless bees (Cerdeira et al., 2021). Third, even among genera shared across all five honey bee species, the subtrees displaying strong signals of co-diversification never encompassed all hosts, implying lineage-specific acquisition or loss of these symbionts. Fourth, several species-level clades (24%–47% of all species-level clades) include MAGs from multiple host species, consistent with recent host switching or ongoing interspecies transmission.

These results appear to contradict previous studies that reported evidence for co-diversification between bee gut symbionts and their hosts (Koch et al., 2013; Kwong et al., 2017; Y. Li et al., 2022; Powell et al., 2016). This discrepancy may stem from the more comprehensive approach of our study, which incorporates improved design and analytical methods, allowing us to detect patterns that may have been overlooked in previous work. However, even the earlier studies acknowledged that host switches, symbiont gains, and losses must have occurred, as evidenced by the imperfect congruence between host and bacterial phylogenies. For example, one study that applied statistical methods and found evidence for co-diversification still estimated that the rate of co-speciation was about half the rate of symbiont switches (Koch et al., 2013). Furthermore, this study focused primarily on bumble bee bacteria from the genera *Snodgrassella* and *Gilliamella*. It is possible that the signal for co-diversification in gut symbionts of bumble bees may be stronger compared to those from honeybees. This hypothesis should be tested in future studies using more comparable approaches, including a larger number of host species, and complementing analysis based on metagenomes with isolate genome data. Our results also markedly contrast

with co-diversification findings in mammals, specifically hominids (Arora et al., 2023; Good, 2023; Moeller et al., 2016; Rühlemann et al., 2024; Sanders et al., 2023; Suzuki et al., 2022) and rodents (Sprockett et al., 2025). Those studies, using a similar approach, found that the ratio of the number of significantly congruent clades observed and expected only by chance was one order of magnitude higher than in our dataset (~12 vs. ~1.2). However, even in hominids, only 16%-36% (Rühlemann et al., 2024; Sanders et al., 2023; Suzuki et al., 2022) of all tested bacterial clades showed statistically significant signals of co-diversification, suggesting that the evolution of only a minority of the microbiota is driven by co-diversification.

While the social lifestyle of bees likely facilitates within-host transmission and could, over evolutionary timescales, promote co-diversification, ecological opportunities for microbial spillover may be relatively common - perhaps due to the sympatric distribution of *Apis* species in Asia. As a result, long-standing associations between specific hosts and their gut symbionts may occasionally be disrupted by the invasion of competing microbes from related host species, thereby limiting the potential for symbionts to consistently co-diversify with their hosts across species boundaries. Furthermore, since all five honeybee species have similar dietary preferences, gut structure, and physiology and lack an adaptive immune system, the barrier to colonization between native and non-native bacteria may be low. This is supported by experiments with microbiota-free bees, which showed that bacterial strains isolated from honeybees can colonize a non-native honeybee species and sometimes even outcompete native bacterial strains (Kwong et al., 2017; Zhang et al., 2022).

Finally, our study highlights that the host-specific composition of the microbiota in these honeybee species translates into meaningful functional differences. In particular, both the presence of different variants of conserved members and the gain of novel taxa with distinct functional capabilities significantly influenced microbiome function. For example, *Apis dorsata* harbors honey bee-specific lineages of *Dysgonomonas* - a genus of polysaccharolytic Bacteroidetes also found in the guts of cockroaches, termites, and humans (Bilen et al., 2019; Vera-Ponce de León et al., 2020; Yang et al., 2014), which substantially expanded the repertoire of pectin and hemicellulose degradation genes in the gut microbiota of *A. dorsata* without necessarily replacing other gut symbionts. *A. dorsata* may be distinct in its broad exposure to more diverse microbes compared to other honeybees, given its highly migratory nature (Dyer & Seeley, 1994). Interestingly, differences in diet may not explain differences in microbiome diversity, considering that *A. dorsata* tends to forage on fewer mass-flowering plants (Raffiudin et al., 2024), and plant-pollinator network analysis suggests that *A. florea*, *A. cerana*, and *A. dorsata* show substantial overlap in the flowers they visit (Stewart et al., 2018). In summary, the honeybee gut microbial community shows a highly dynamic evolution despite the presence of a few conserved lineages.

Our results lay the foundation for more detailed functional characterization of diverse bacterial gut symbionts that were previously unrecognized as important members of the honeybee gut microbiota. This work substantially broadens our understanding of honeybee-associated microbiota, which has so far been based primarily on studies of the managed Western honeybee. More broadly, it provides new insights into host specificity in gut microbiota-host interactions, demonstrating that a dynamic evolutionary history can critically shape microbiome functions across divergent animal hosts.

METHODS

SAMPLING AND STORAGE

All honeybee workers (10-50 individuals per colony) were collected from colonies across locations described in the metadata sheet (**Supplementary Table S1**). Most colonies were wild except those *A. mellifera* samples labeled “Langstroth hive” under the hive description column. *A. cerana* samples from Langstroth hives were local bees that were temporarily housed in these hives with minimal intervention but not managed as *A. mellifera* normally are. Bees were stored in small, ventilated boxes upon collection and transferred to absolute ethanol within a day. Bees were then stored at -20°C until dissection and DNA isolation, sufficiently preserving the sample for sequencing the microbiome (Hammer et al., 2015, p. 015). Our sampling strategy prioritized covering many colonies across locations to capture the diversity of each host species rather than many individuals from the same colony, which was done before (Ellegaard et al., 2020; Ellegaard & Engel, 2019).

DISSECTION AND DNA EXTRACTION

To decrease the ratio of DNA from the host cells to bacterial cells, we only used the hindgut which contains most of the bacterial biomass. To do so, we carefully opened the cuticle, removed the stinger and venom glands, and, finally, cut the tissue near the pylorus, below the malpighian tubules removing the midgut and crop. Then, the hindgut tissue was transferred to a bead-beating tube containing 478 µL of sterile PBS and 2.0 mm sterile glass beads and subjected to bead-beating for 30s at 6 m/s in FastPrep-24 5G (MP Biomedicals). Next, 0.1 mm zirconia beads along with 250 µL of 2x CTAB, 2 µL of β-mercaptoethanol and 20 µL of proteinase K (20 mg/mL) was added and bead-beating was carried out for 30s at 6 m/s in FastPrep-24 5G (MP Biomedicals). The homogenate was incubated for 1h at 56°C. Next, 750 µL of Phenol-chloroform isoamyl alcohol (25:24:1, v/v) at pH 8.0 was added and mixed thoroughly, followed by centrifugation for 10 min at 16,000 xg. The upper aqueous layer was transferred to a new tube, and 500 µL of chloroform was added and mixed thoroughly. The mixture was centrifuged for 10 min at 16,000 xg, and the 300 µL of the upper phase was mixed with 900 µL of chilled absolute ethanol and incubated at -20°C overnight. Next, the precipitated DNA was pelleted by centrifugation for 30 mins at 16,000 xg at 4°C. After two washes with 70% ethanol, it was resuspended in 50 µL of nuclease-free water. Next, it was treated with RNaseA and purified using magnetic beads (Clean NGS). DNA concentration was measured using Qubit™ (Thermo Fischer Scientific). Bacterial absolute abundance was measured using qPCR with primers targeting the 16S rRNA region as described earlier (Kešnerová et al., 2017) with 0.1x dilution of gut homogenate.

METAGENOMIC SEQUENCING

Library preparation was carried out using the Nextera XT library kit (Illumina), according to manufacturer instructions, with the number of PCR cycles chosen to account for the amount of DNA in samples. Illumina sequencing was carried out using Novaseq (2x 150 bp) with a target of 10-50 million microbial reads. Since there was a significant difference between host species in the total gut microbial content as determined by qPCR using universal primers,

first, a run of sequencing was carried out without normalization of the amount of samples added to the library to estimate the percentage of host DNA and microbial DNA in the library made from each sample. This was useful in ensuring that there was enough DNA from the bacterial community in all the samples, including those where the library might have been dominated by host DNA due to lower bacterial biomass. Further sequencing was then carried out after pool correction, where the library was adjusted in such a way when possible that at least 15 million microbial reads were obtained (based on expectation from the first shallow sequencing), even for the samples with the lowest percentage of microbial DNA. All library preparation and sequencing was done at the Lausanne Genomic Technologies facility (GTF), UNIL. Details about all further bioinformatic analysis, including the code used, can be found on GitHub (<https://github.com/Aiswarya-prasad/honeybee-cross-species-metagenomics>) and Zenodo (<https://zenodo.org/doi/10.5281/zenodo.13732977>) in the repositories.

MAPPING TO DATABASES

To assess the amount of bacterial, host, and other reads in each sample, reads were mapped to host and MAG databases. The host database comprised draft or complete genomes of all the four host species (GCA_000184785.2, GCA_000469605.1, GCA_001442555.1, GCF_003254395.2) except *Apis andreniformis*, since no genome was available in public databases for this host species. A database was compiled using representative MAGs identified for each 95% ANI cluster or bacterial species with the highest quality score (calculated as described by dRep (Olm et al., 2017) using various MAG characteristics, including checkM contamination and completeness). Trimmed and quality-filtered reads were mapped to the host database using BWA-mem (H. Li & Durbin, 2009) followed by the MAG database (and vice versa to account for non-specific mapping). Three samples were removed from the downstream analysis as their metagenomic reads did not map the host that they were labelled as (**Supplementary Table S3**) and might have faced cross-contamination during sample processing. Samples for which relatively few reads mapped to the MAG database were also often low in bacterial biomass as measured by qPCR (**Supplementary Fig. S16**). Indicating that the sequencing and workflow to detect microbial diversity in each sample was otherwise robust. Initial taxonomic profiling of the reads using mOTUs (Ruscheweyh et al., 2022) detected the expected core marker genes from various honeybee gut microbiota members.

ASSEMBLY AND BINNING OF MAGS

Raw reads were trimmed and quality-filtered before assembly using Trimmomatic (Bolger et al., 2014) and bbtools. We did not carry out host filtering at this stage as we have no host genome for *A. andreniformis* and did not want to bias our samples due to non-specific mapping. For some samples that had a large number of reads (19/200 samples) due to which assemblies were not successful, bbnorm was used to down-sample regions of high depth. For samples with too many host reads due to which assemblies failed (3/200 samples), host-filtered reads were used. The detailed workflow used for assembly and binning into MAGs is described in GitHub (<https://github.com/Aiswarya-prasad/honeybee-cross-species-metagenomics>). Briefly, contigs were assembled using metaSpades (Nurk et al., 2017). Quality-filtered reads were mapped against assembled scaffolds (all against all samples), and then scaffolds were binned into MAGs using coverage information as implemented in

MetaBAT2 (Kang et al., 2019). The quality and completeness of the resulting MAGs were assessed using CheckM (Parks et al., 2015) and low (< 50% completeness or >5% contamination), medium (50-90% completeness and <5% contamination), high (> 90% completeness and <5% contamination) MAGs were then clustered into groups averaging 95% average nucleotide identity (ANI) using dRep (Olm et al., 2017). To further ensure that our MAGs represent true species and are not assembly and binning artifacts, we inferred phylogenies, including isolate genomes along with our MAGs. The rationale is that if MAGs are artifacts, they would rarely cluster together in the phylogeny and would instead be placed apart from isolate genomes. This new analysis shows that our MAGs form discrete phylogenetic clades, which include MAGs assembled from several different samples. Further, these clades often include closely related isolate genomes in cases where several strains were included from the same species. MAGs belonging to clusters without isolate genomes represent novel species not isolated before (**Supplementary Fig S4**). We are, therefore, confident that the MAGs represent true species and can be used to infer meaningful phylogenies. Other studies in the field, which have used similar approaches to ours, are similarly able to establish meaningful phylogenies from MAGs (Sanders et al., 2023).

Taxonomic annotation was carried out using GTDB-tk (Chaumeil et al., 2020). No manual curation of MAGs was performed; however, Kaiju (Menzel et al., 2016) (nr database) and Kraken2 (Wood et al., 2019) (custom database including bacteria, viruses and honeybee) classification of scaffolds screened by random sampling further confirmed that scaffolds from the same MAG matched each other and the taxonomic assignment of the MAG. Each cluster of MAGs was named to reflect the GTDB species name or, in cases where none was assigned, a unique identifier, including the genus name was used. Combined with the inspection of ANI heatmaps (**Supplementary Fig. S3**), a final set of species names was assigned to each of the 95% ANI clusters.

TAXONOMIC PROFILING

InStrain (Olm et al., 2021) was used according to its recommended workflow for taxonomic profiling at the bacterial species and strain level. Trimmed and quality-filtered reads were competitively mapped to the database of representative MAGs using the tool bowtie2 (Langmead & Salzberg, 2012), which presents mapping scores in a format compatible with InStrain. Among many approaches to detecting microbial species in a dataset, this stands out in its appropriate filters applied to mapped reads, including read ANI, which helps differentiate genomes of closely related species (<95% ANI) and breadth of the genome mapped to avoid falsely identifying species as present due to high coverage of a region shared with others or horizontally transferred. Moreover, genome coverage of a given species can be misleading when comparing communities of varying diversity, as an equally abundant species would have higher coverage in a community with low diversity. A species was considered present if at least 50% of its genome (comprising all scaffolds) was covered by the reads. The justification for the choice of this threshold is exemplified for several *Snodgrassella* species in **Supplementary Fig. S5**: briefly: there is a clear gap between samples where a given species is detected (in large green circles) or not (in small red circles). We inspected such plots for all the species represented in the database and visually verified that the breadth coverage we used (50%) was valid across species. These plots are included with the dataset found in the Zenodo repository:

(<https://zenodo.org/doi/10.5281/zenodo.13732977>).

Relative coverage, when used, was obtained by dividing the coverage of each species by that of the sum of all and qPCR copy numbers of the bacterial 16S rRNA region were used to estimate absolute abundance for comparison between samples where mentioned. For samples from India (n=50), qPCR values were interpolated based on the median of the host species. The underlying data (presence-absence, relative abundance etc.) and normalization approach are explicitly indicated for each downstream analysis section. After filtering samples as mentioned earlier and removing samples with fewer than 1 million mapping quality filtered reads mapping to the representative database of MAGs, a total of 195 samples were included in the analysis (*A. mellifera* 34, *A. cerana* 45, *A. dorsata* 44, *A. florea* 42 and *A. andreniformis* 30).

HOST SPECIFICITY AND STRAIN-LEVEL PROFILES

Host specificity was calculated for each bacterial species as the Rohde's index of specificity (Rohde & Rohde, 2008) using the prevalence of the species across hosts. Strain-level composition profiles for each bacterial species represented in the MAG database were generated per sample using InStrain in database mode. The percentage of variant sites was calculated as the percentage ratio of the number of SNVs to the total length of the MAG covered. A site is counted as an SNV if the site is covered at least 5x and the allele is present at a frequency of at least 0.05. Regions mapping with a read ANI of <0.92 (might be cross-mapping of reads from a different species) or coverage of <1x were not considered. To compare the strain profile of samples, the popANI measure, as comprehensively described by InStrain (Olm et al., 2021), was used. Briefly, a value from 0 to 1 was calculated for each species for each pair of samples in which the genome was covered at least by 5x and at least half of the genome was covered in both samples. A position of the genome was considered if there was more than one allele with a frequency of at least 0.05. If both samples share no alleles in that position, it is counted as a popANI substitution. The value for that pair of samples was then calculated as the ratio of positions containing popANI substitutions and the number of positions compared. This value was used to create a matrix of Jaccard distances between all pairs of samples. Finally, each of the bacterial species was either considered host-specific, if they were detected in only one host species, or shared if not. Shared species were considered host-specific at the strain level if samples showed significant clustering by host species ($p < 0.01$) according to a PERMANOVA test.

PHYLOGENY CONSTRUCTION AND CO-DIVERSIFICATION TEST

To gain a rough estimate of the taxonomic level at which to expect co-diversification, we computed a core genome phylogeny using the nucleotide sequence of 120 bacterial marker genes for all medium and high-quality MAGs identifying core genes using GTDB-tk, aligning in a codon-aware manner using MACSE (Ranwez et al., 2011) and constructed using IQ-TREE (Nguyen et al., 2015). From this tree, the median substitutions per site were 0.014 for MAGs of the same species, 0.39 for different species of the same genus, and 2.65 for different genera (**Supplementary Table S7**). Bacteria evolve at a rate in the order of magnitude of 10^{-3} substitutions per site per million years (Ochman et al., 1999). Since the *Apis* clade is about 30 million years old (Su et al., 2023), gut microbes that diverged during this timeline could

be as divergent as different species within a genus or closer, confirming that genus-level trees are appropriate for these tests.

OrthoFinder (Emms & Kelly, 2019) was used to infer orthologous gene families of all medium and high-quality MAGs for each genus along with previously published genomes of the same genus from other environments as outgroup (details of genomes used in metadata files included in the Github (<https://github.com/Aiswarya-prasad/honeybee-cross-species-metagenomics>) and Zenodo (<https://zenodo.org/doi/10.5281/zenodo.13732977>) repository. Phylogenies were constructed using IQ-TREE using the amino acid sequences of genus-level single-copy core genes aligned by MAFFT (Katoh et al., 2002). All medium and high-quality MAGs (including several per bacterial species) were included in the phylogenies in order to preserve maximum information and power of resolution. Since medium-quality MAGs were included in phylogenies, core genes can be randomly missing in some MAGs. Naively defining core genes as those present in all MAGs would result in the exclusion of most orthologs. Hence, as implemented in OrthoFinder, when <1000 single-copy orthologs are present in all MAGs, it iteratively identifies the orthologs that are found once in N-1 MAGs until either considering one less MAG does not double the number of orthologs or more than half the MAGs are dropped. Core single-copy orthologs identified were in agreement with those orthologs identified as core by mOTUpa (Buck et al., 2022), which uses a Bayesian approach informed by the completeness score of MAGs. The number of orthologous gene families (OG) used for core genome phylogenies and the minimum percentage of MAGs of the genus in which each OG was detected (e.g., 100% means all OGs were detected in all MAGs and 50% means every OG was found in 50% or more of the MAGs): Apibacter - 1096 (94.5%), Apilactobacillus - 704 (81.8%), Bartonella - 705 (81.5%), Bifidobacterium - 118 (84.0%), Bombilactobacillus - 194 (89.4%), CALYQJ01 - 883 (89.5%), CALYQQ01 - 1221 (84.6%), Commensalibacter - 1120 (96.6%), Dysgomononas - 223 (86.5%), Entomomonas - 1144 (91.7%), Frischella - 984 (85.9%), Gilliamella - 189 (88.0%), Pectinatus - 1143 (97.3%), Pluralibacter - 1419 (100.0%), Saezia - 1000 (92.3%), Snodgrassella - 799 (91.5%), WRHT01 - 1897 (100.0%). Further information about the parameters and best-fit model chosen to compute the phylogenies can be found in the log files included in the corresponding directory containing IQ-TREE results in the Zenodo (<https://zenodo.org/doi/10.5281/zenodo.13732977>) repository.

A host phylogeny inferred based on complete mtDNA sequences of honeybee species rooted by a bumble species published before (Carr, 2023) was used for co-diversification tests. Tests were carried out using the Homola cospeciation test (The scikit-bio development team, 2020) (implemented in Python) on all nodes containing at least 7 MAGs from at least three host species and with a maximum tip-tip distance not exceeding 1 for 100 permutations. This node-by-node comparison was then carried out on trees with tip labels artificially permuted randomly across (N=100) second-order permutation test to estimate the number of nodes displaying significant results by chance, as described in an earlier study (Moeller et al., 2023). See (**Supplementary Fig. S20**) for a detailed schematic of this approach. Thresholds were set at various levels for comparison of the number of significant nodes in the original tree and the median number of significant nodes in tip label permuted trees: relaxed ($p<0.05$), medium ($p<0.05$ and $r>0.75$) and strict ($p<0.01$ and $r>0.75$). Finally, if nodes including outgroups were picked up as significant (2 out of 15), they were excluded because the outgroup genomes come from sources other than the honeybee species being

compared. The remaining significant subtrees were visually inspected for congruence between symbionts and their hosts. Specifically, we examined whether the branch lengths separating symbionts from different host species were consistent with the estimated divergence times of their hosts. For instance, given that all host species diverged several million years ago, we expected symbionts associated with different hosts to exhibit clear signs of sequence divergence in our phylogenetic trees.

GENE FUNCTIONAL ANNOTATION AND PROFILING

Open reading frames (ORFs) were identified on all assembled scaffolds of each sample using Prodigal (Hyatt et al., 2010). ORFs marked partial or shorter than 300 bp were removed. Further, those ORFs detected on scaffolds identified as being of eukaryotic origin by Whokaryote (Pronk & Medema, 2022) or unclassified as bacteria by Kaiju (nr database) were also excluded. The remaining ORFs were then annotated using DRAM (Shaffer et al., 2020) to obtain KEGG ortholog (KO) identities. Finally, the gene catalog comprising all these ORFs was assigned to CAZyme families as implemented in Cayman (Ducarmon et al., 2024).

Trimmed and quality-filtered reads from each sample were mapped back to their respective assemblies using bowtie2. Low-quality ($Q<20$) alignments were removed, and coverage per gene/ORF was calculated using BEDtools (Quinlan & Hall, 2010). Genes were considered detected if >5 reads were mapped to it and $> 90\%$ of the bases in the gene were covered by at least a read. The abundance of each gene was then calculated in terms of reads per kilobase per million (RPKM) calculated as follows:

$$RPKM = \frac{\# Mapped\ reads}{\left(\frac{Gene\ length}{1000}\right) * \left(\frac{\# Total\ reads}{10^6}\right)}$$

For KOs and CAZymes, summed RPKMs of constituent genes were used. To estimate the average count of CAZymes of interest (all families other than GT and AA families) in each genus per individual, the RPKM of all CAZymes in each genus was summed per sample, and then the average of this sum across samples was converted to a relative proportion to represent the relative contribution of each genus on average in each host species and plotted. Only genera with a prevalence >0.05 in the respective host were considered.

To infer biological pathways represented within each MAG, we collected the KOs detected within each MAG and used MinPath (Ye & Doak, 2009), a parsimony approach which aims to explain the set of protein (KO) families using the minimum number of pathways. Further, we used MaAsLin 2 (Mallick et al., 2021), which was developed to assess multivariable associations of microbial community features with complex metadata. With the presence-absence matrix of pathways as the dataset and 'genus' as the explanatory variable we identified the most likely functional similarities driving certain genera to cluster together in the PCoA plot of all KOs per MAG (**Fig. 4C**).

STATISTICAL TESTS AND BIOINFORMATIC ANALYSES

All details about the specific parameters used for each tool and their versions, along with conda environment specifications, are included in the GitHub repository ([\(Baselga and Orme 2012\)](#)) and Zenodo (<https://zenodo.org/doi/10.5281/zenodo.13732977>) repositories. In

general, bioinformatic workflows were implemented using Snakemake (v7.28.1) to parallelize jobs on a computing cluster at the University of Lausanne, preprocessing and filtering using Python scripts, visualization using ggplot with several related packages in R and iTOL for phylogenetic trees and arrangement and decoration of plots using Adobe Illustrator. Unless otherwise specified, statistical tests of the results of each section were carried out as follows. Significance in boxplots was tested using the Wilcoxon Rank Sum test (significant if $p < 0.05$). Distance matrices were calculated using the method indicated in their respective plots. The package Betapart (Baselga & Orme, 2012) was used to estimate the turnover component of the Jaccard and Sorensen distances. Multivariate analyses were visualized using PCoA or NMDS and tested using the adonis2 function implemented in the R package Vegan. The effect size was adjusted for sample size to get Ω^2 using the adonis_OmegaSq function from the package Russel88/MicEco (<https://zenodo.org/badge/latestdoi/83547545>). To test for phyllosymbiosis, we minimized bias due to pseudo-replication by performing 1000 permutations. Each permutation included five randomly selected samples, one of each host species, and tested for correlation between bacterial and host distances using a Mantel test (as implemented in the Vegan package). The p-values were adjusted using the FDR method.

ACKNOWLEDGMENTS

We thank the following people and organizations for supporting honeybee sampling and access. The Nature Inspired team (John Chan, Steven Wong, & JC Tan), Mr. HoneyBee Farm, The Orang Asli community (Temengor, Perak), Sherwani Honey Bee, Mr. Wee, Mr. Steve, and MY Bee Saviour. The Cayman team (Quinten Ducarmon, Nicolai Karcher, and Christian Schudoma) for their support with early access to their software and custom annotation of our dataset. The Lausanne Genomics Technology Facility (GTF) team at the University of Lausanne. This work was supported by a Spirit grant (grant number 189496), an SNSF Consolidator grant ('GLOBEE,' grant number 213860), and the NCCR Microbiomes, a National Centre of Competence in Research (grant number 180575 and 225148), funded by the Swiss National Science Foundation. We thank Charlotte Huyghe and Malick Ndaye for their feedback on the revised manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization – PE, SHY, AP, FM; Methodology – AP, FM, PE; Software – AP; Validation – AP; Formal analysis – AP; Investigation – ADP, AP, RS, AS; Resources – PE, SHY, AB; Data curation – AP; Writing - Original Draft – AP; Review and Editing – PE, FM, AP, AB, SHY; Visualization – AP; Supervision – PE, SHY, FM, AB; Project administration – PE, SHY, AB; Funding acquisition – PE, SHY

COMPETING INTERESTS

All authors declare they have no competing interests.

DATA AVAILABILITY

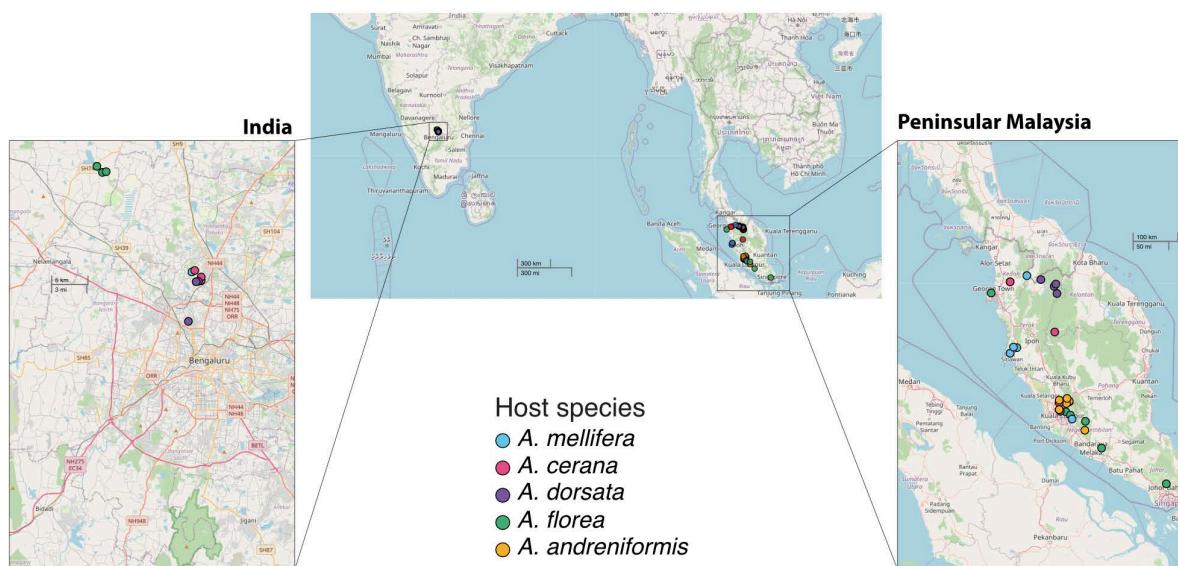
Raw metagenomic data has been deposited to the NCBI Sequence Read Archive (SRA) under the Project ID PRJNA1157353. The collection of MAGs, some important intermediate files, and tables can be found in the Zenodo repository:
[\(https://zenodo.org/doi/10.5281/zenodo.13732977\)](https://zenodo.org/doi/10.5281/zenodo.13732977).

CODE AVAILABILITY

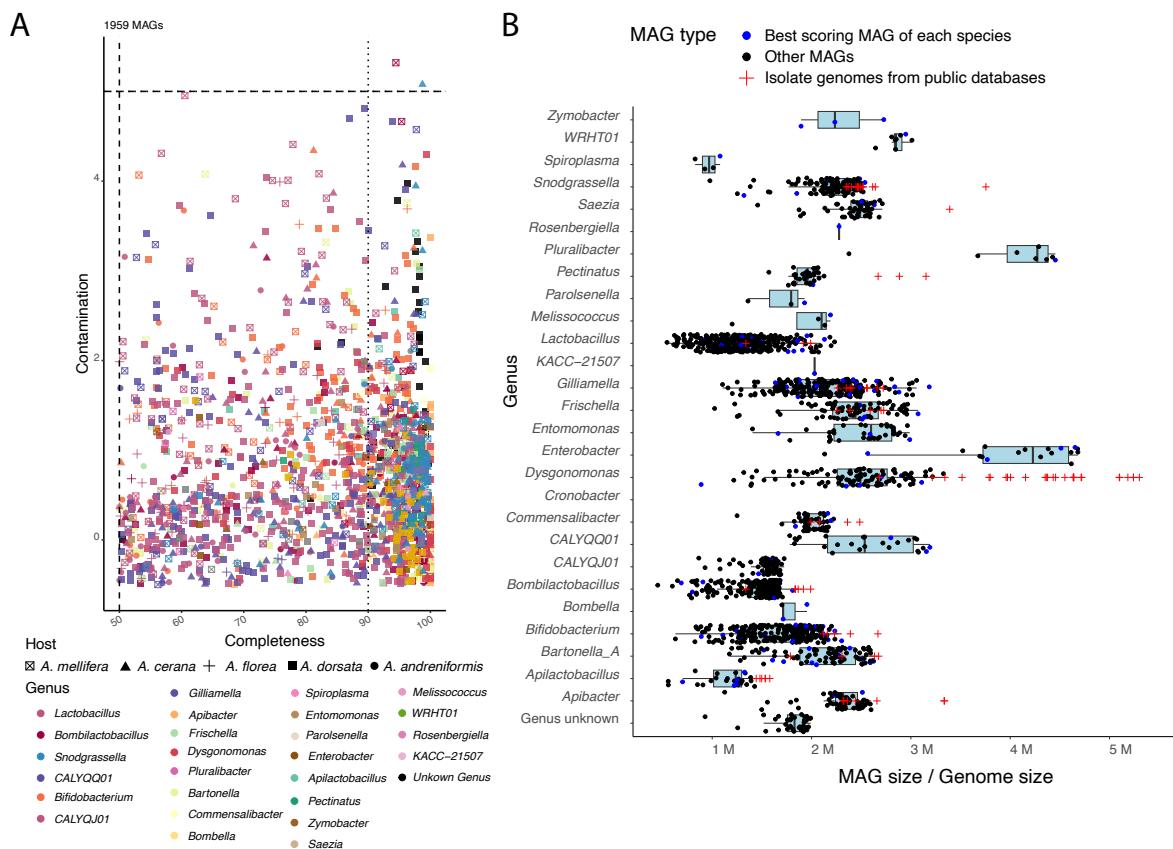
Code and details of parameters and software used are available at:
<https://github.com/Aiswarya-prasad/honeybee-cross-species-metagenomics>
archived at:
<https://zenodo.org/doi/10.5281/zenodo.13732977>

SUPPLEMENTARY MATERIAL

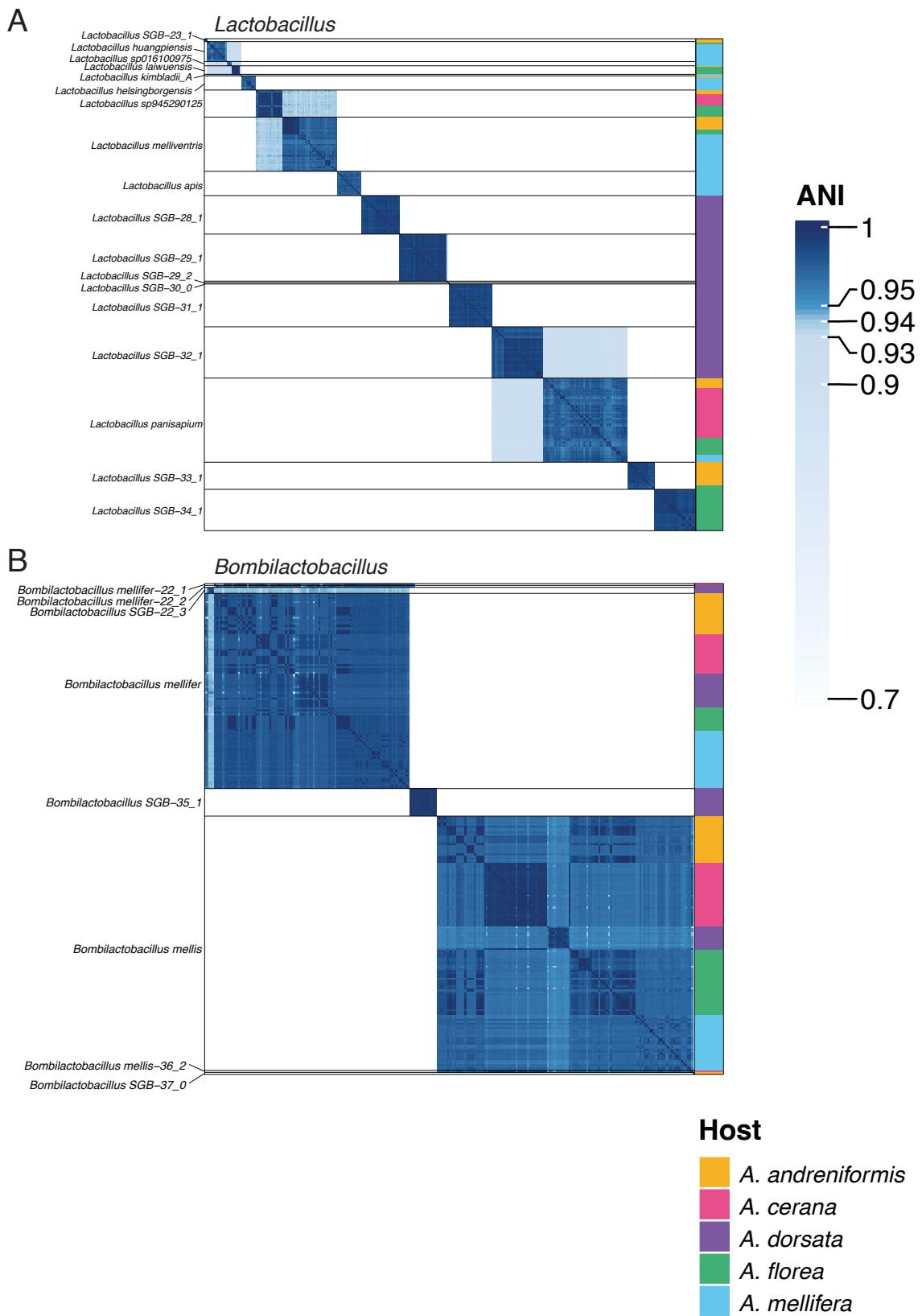
SUPPLEMENTARY FIGURES

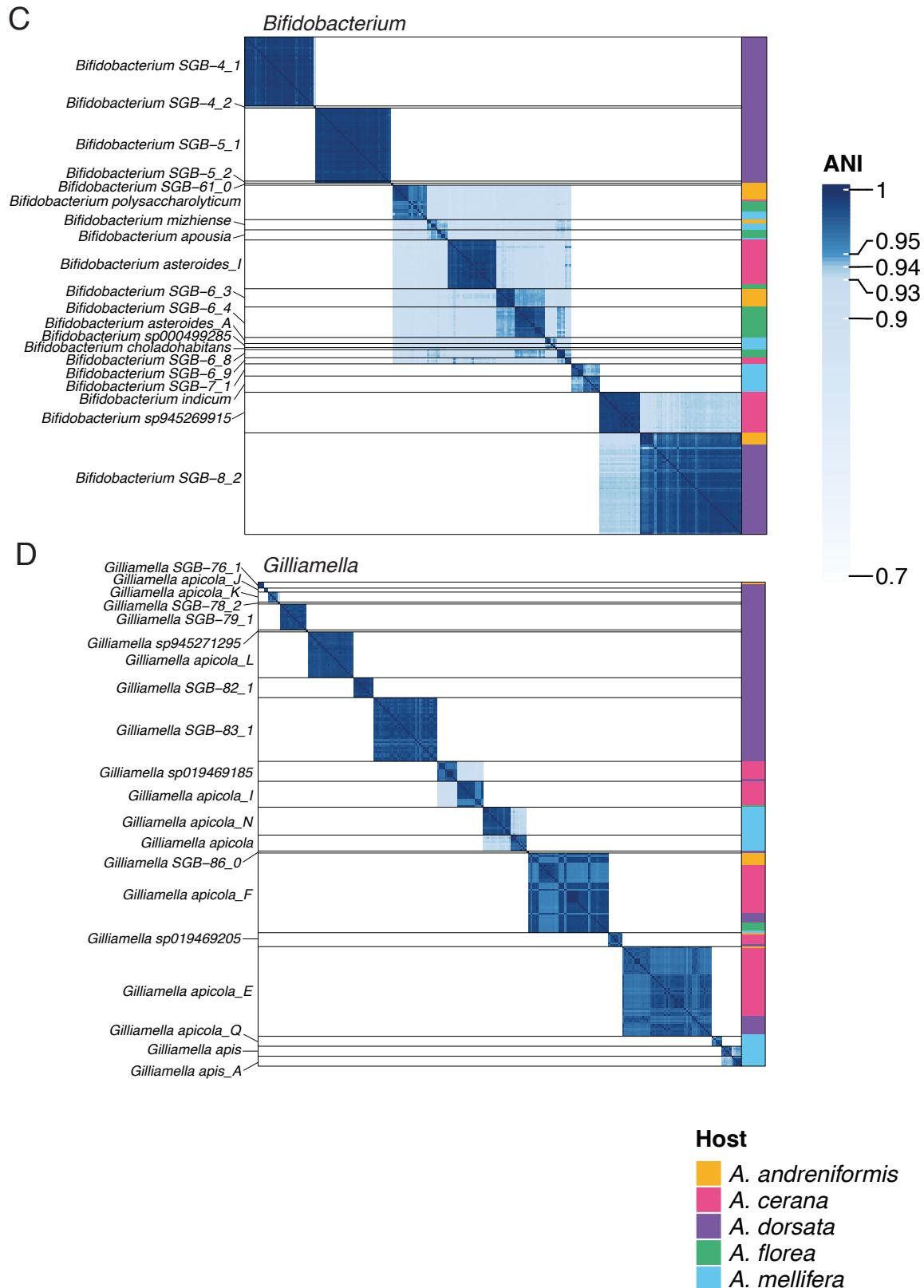


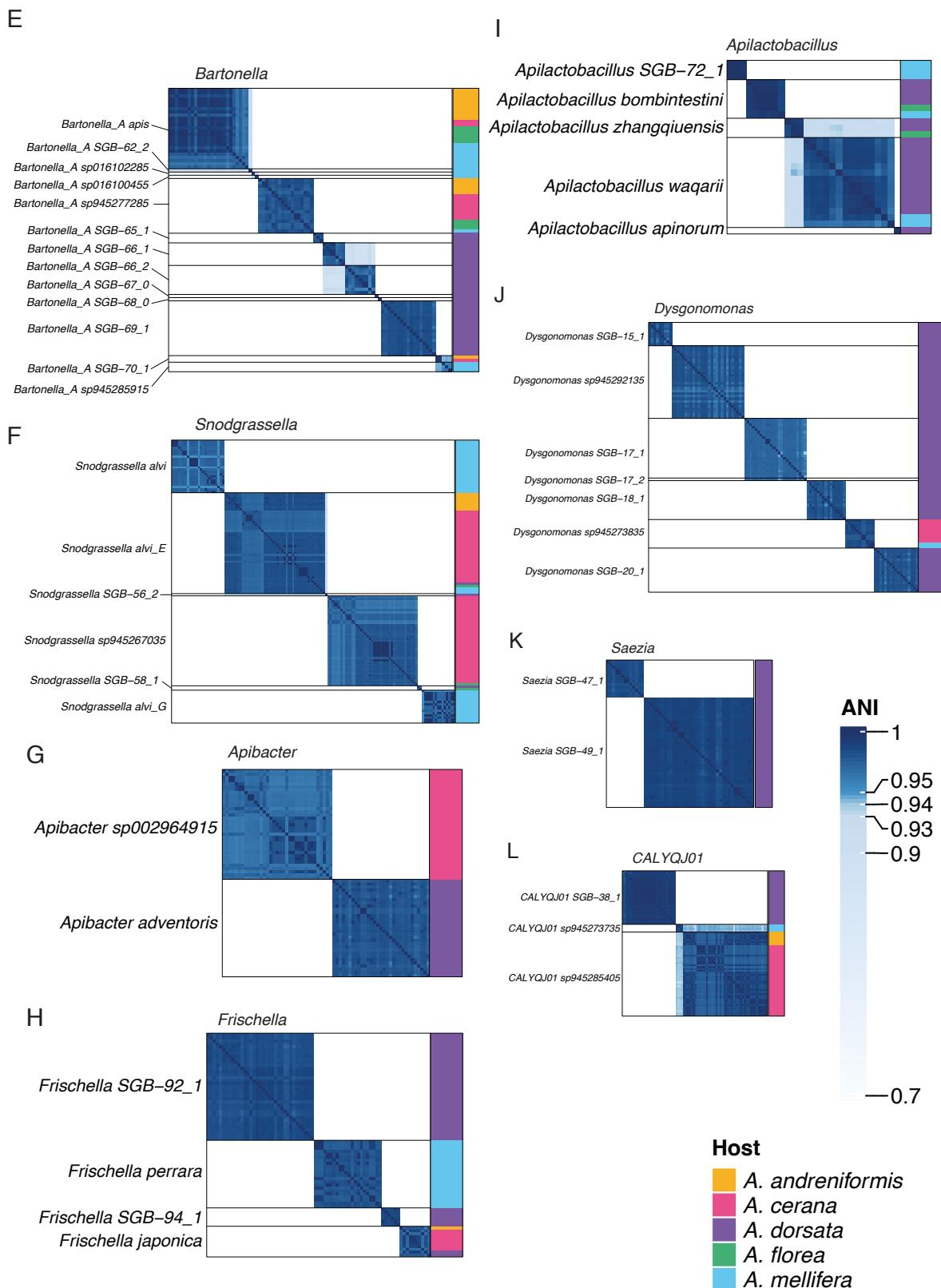
Supplementary Figure S1. Map of sampling locations. Colored points indicate the location of the colonies of each host species sampled (size of points are exaggerated for visibility).



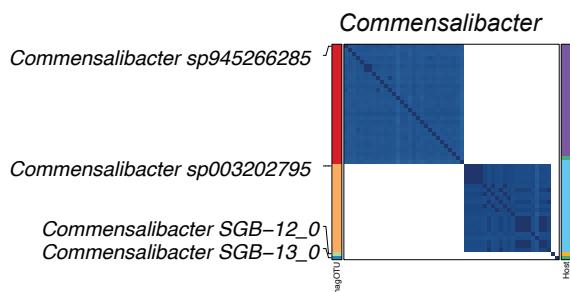
Supplementary Figure S2. Characteristics of recovered MAGs. (A) Scatter plot showing the checkM completeness and contamination of each MAG of medium or high quality. Dotted and dashed lines represent the thresholds used to separate high and medium quality MAGs. (B) Boxplot showing the total size of MAGs (sum of length of all contigs binned into the MAG) compared to known sizes of isolated reference genomes of other species from the genus found in NCBI. The best-scoring MAG of each species within the genus is highlighted in blue.



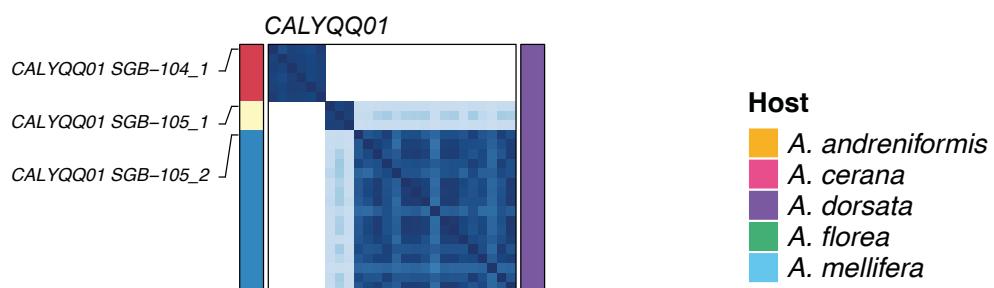




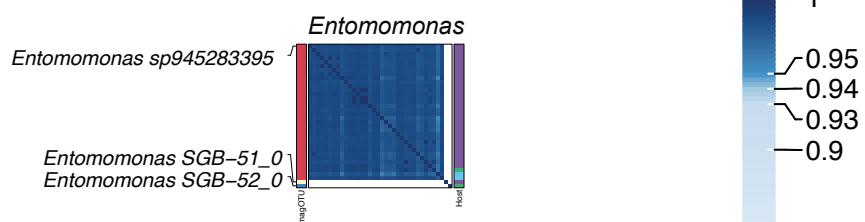
M



N



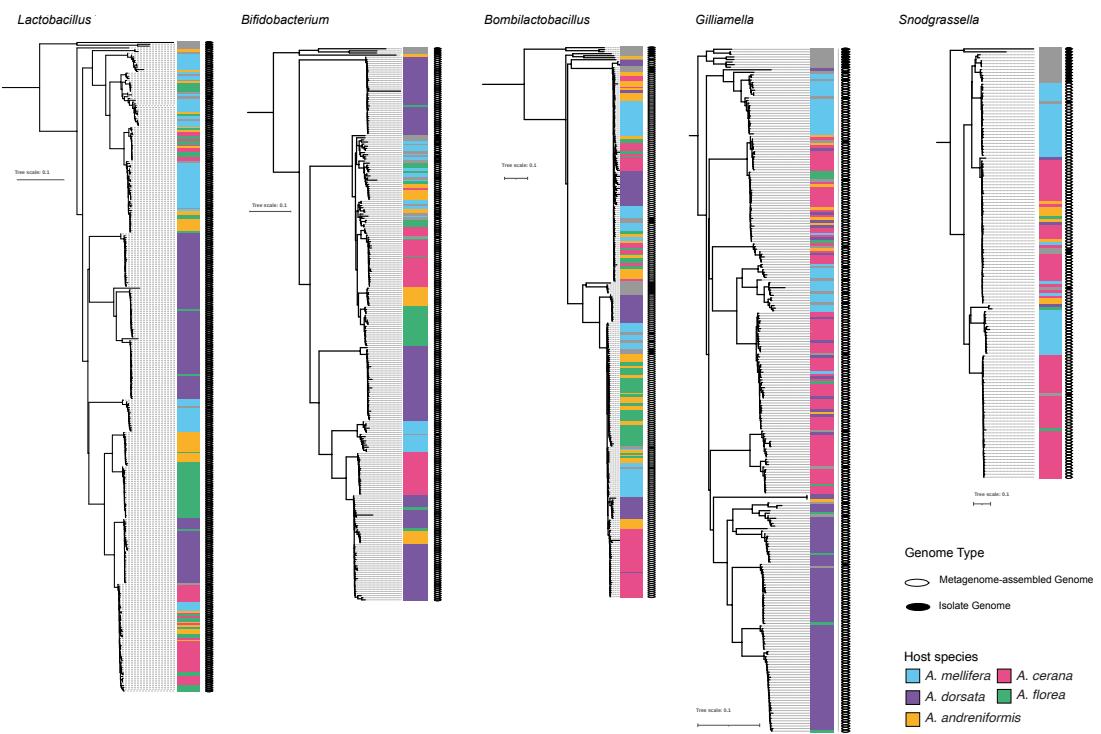
O



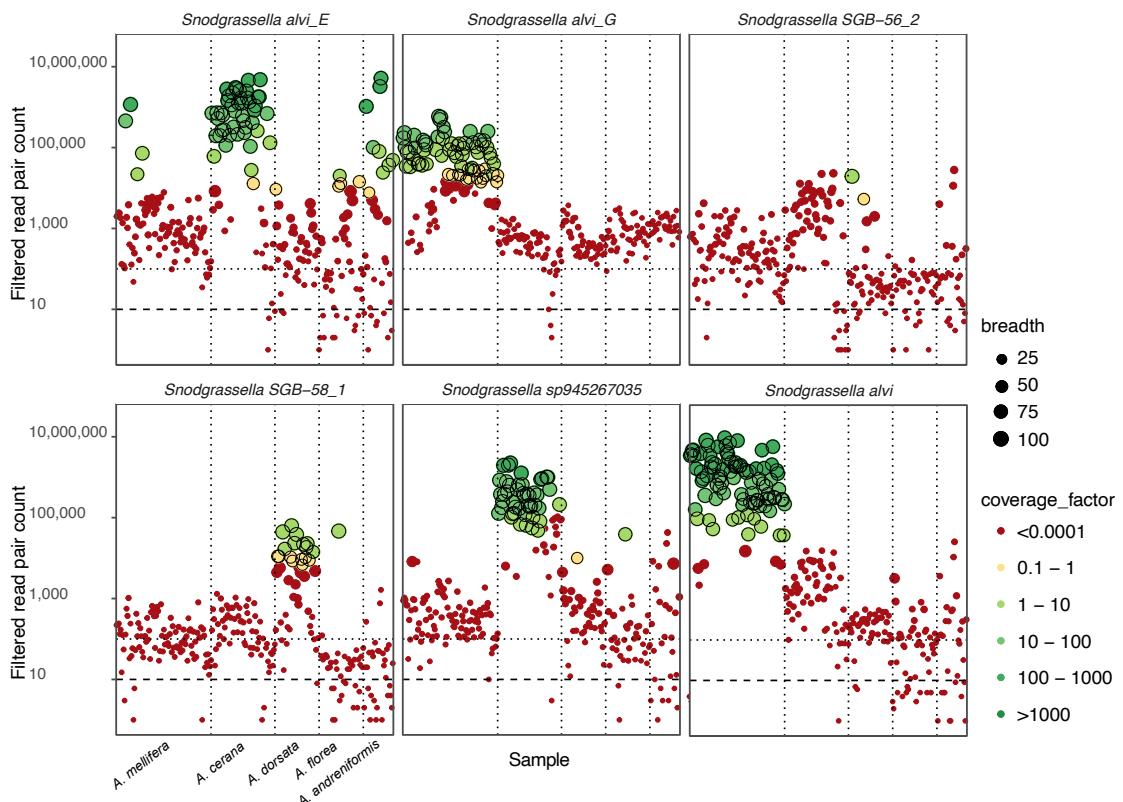
P



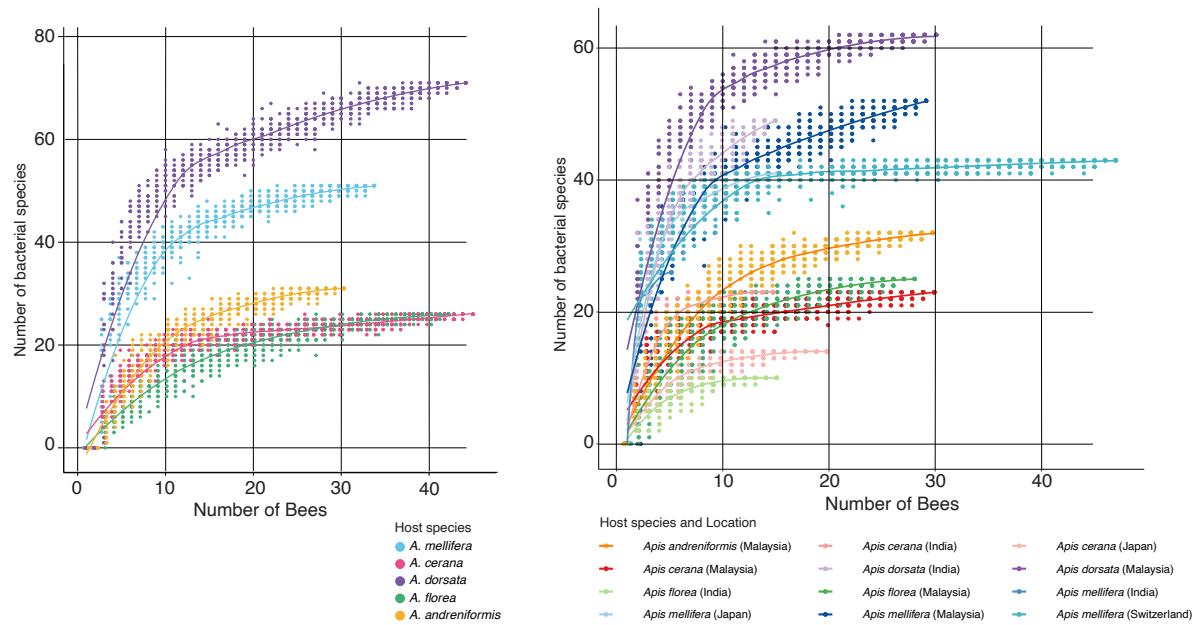
Supplementary Figure S3. ANI heatmaps of MAGs. ANI heatmaps of all medium- and high-quality MAGs per genus with labels on the left indicating the bacterial species assigned and the colors on the right indicating the host species from which the MAG was recovered.



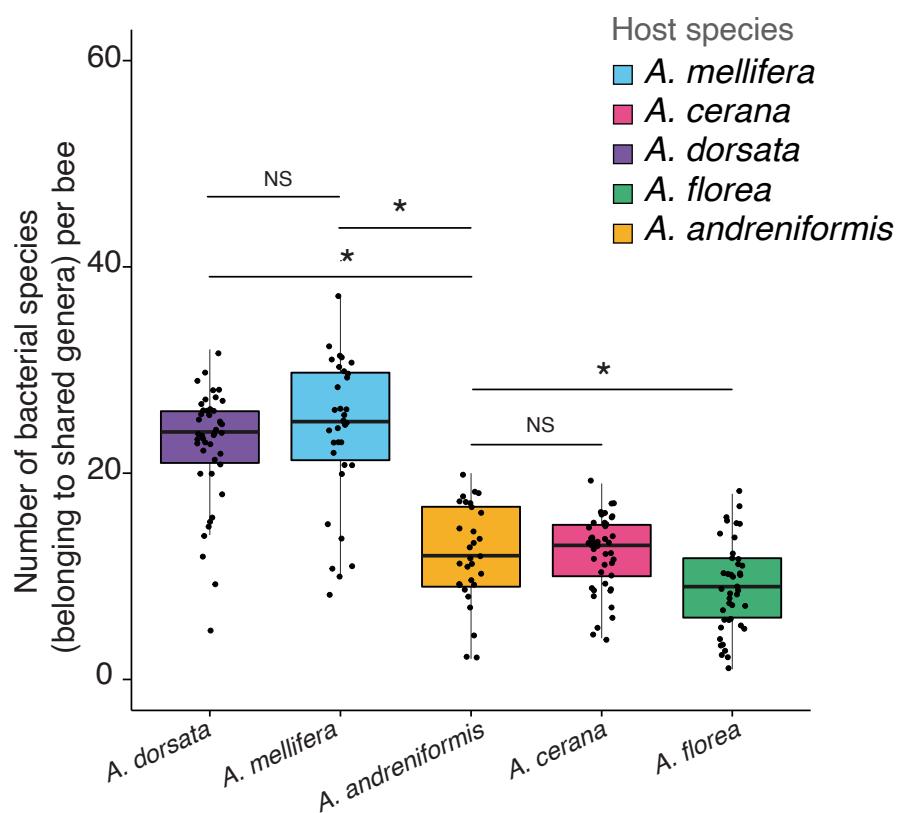
Supplementary Figure S4. Phylogenies of MAGs and isolate genomes. Core genome phylogenies, including all high- and medium-quality MAGs of the bacterial genus and selected isolate genomes from the GTDB database for several bacterial species. Tree leaves corresponding to isolate genomes are marked by a black circle. Bacterial species represented by clades of MAGs corresponding to a given harbor isolate genomes of that species within.



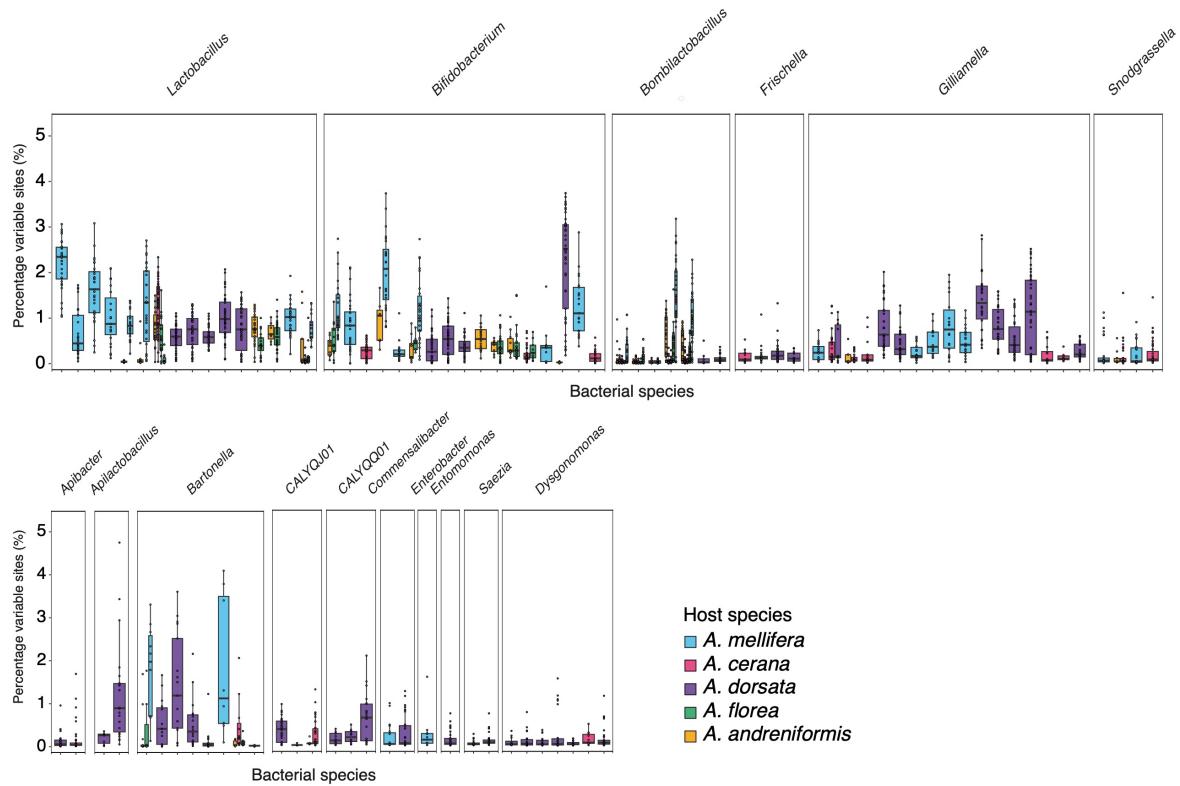
Supplementary Figure S5. Read mapping to species-representative MAGs. The panels show an example of six *Snodgrassella* species across samples as an example of read mapping plots for closely related species. Different *Snodgrassella* species are represented in each panel. The x-axis in each panel represents samples grouped by host species (separated by vertical dotted lines), and the y-axis is the number of read-pairs mapping (after filtering to keep only high-quality mapping) to the focal species in each sample. Horizontal dashed and dotted lines represent different depth cut-offs for visualization. The size of the points represents the breadth of the species-representative MAG covered in the sample and the color of the range into which its coverage depth falls. Points outlined with a black line correspond to samples in which the respective species was considered present according to thresholds used in the pipeline.



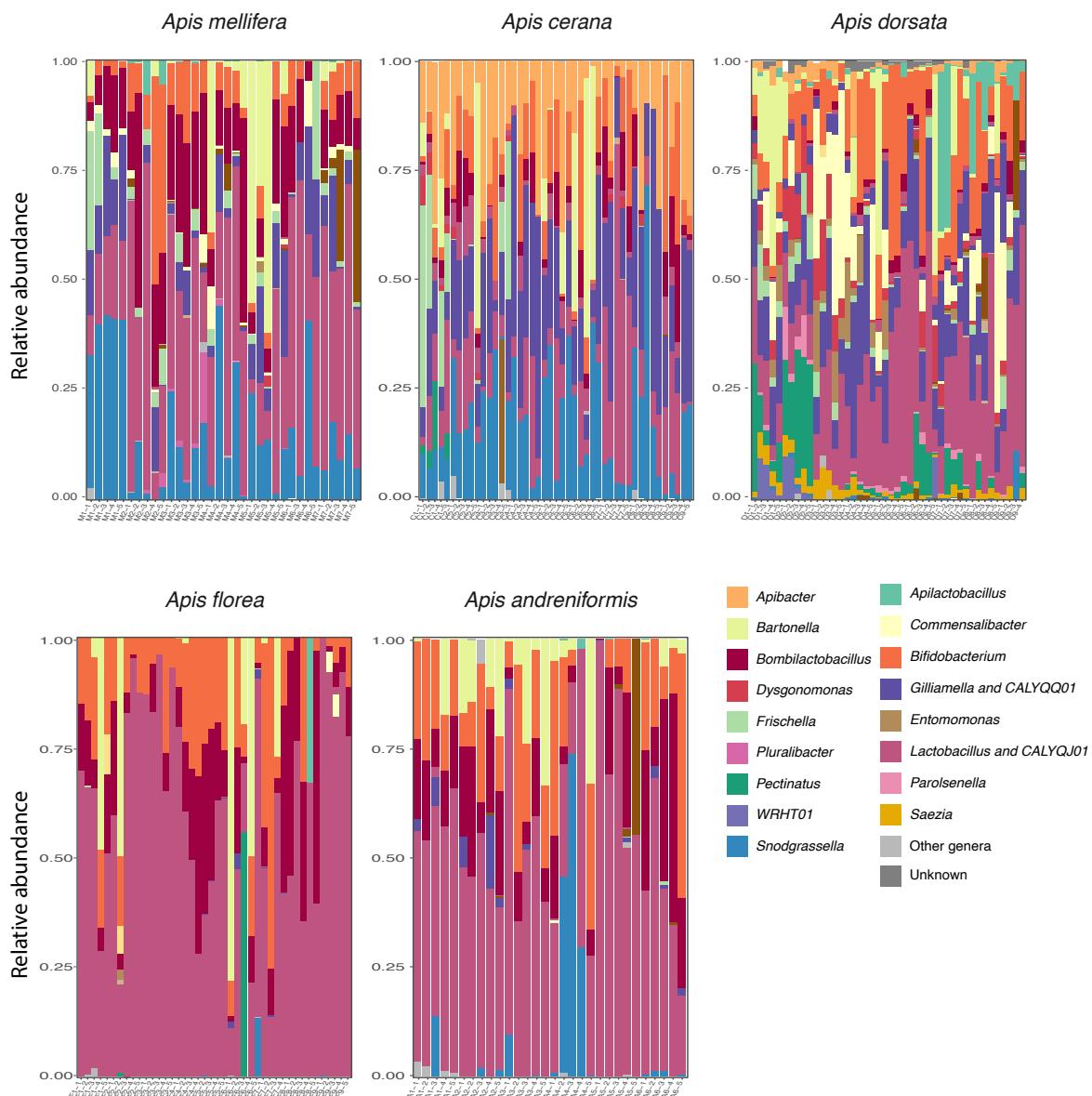
Supplementary Figure S6. Cumulative curves of the number of bacterial species. Number of bacterial species as a function of sample size by host species and sampling location.



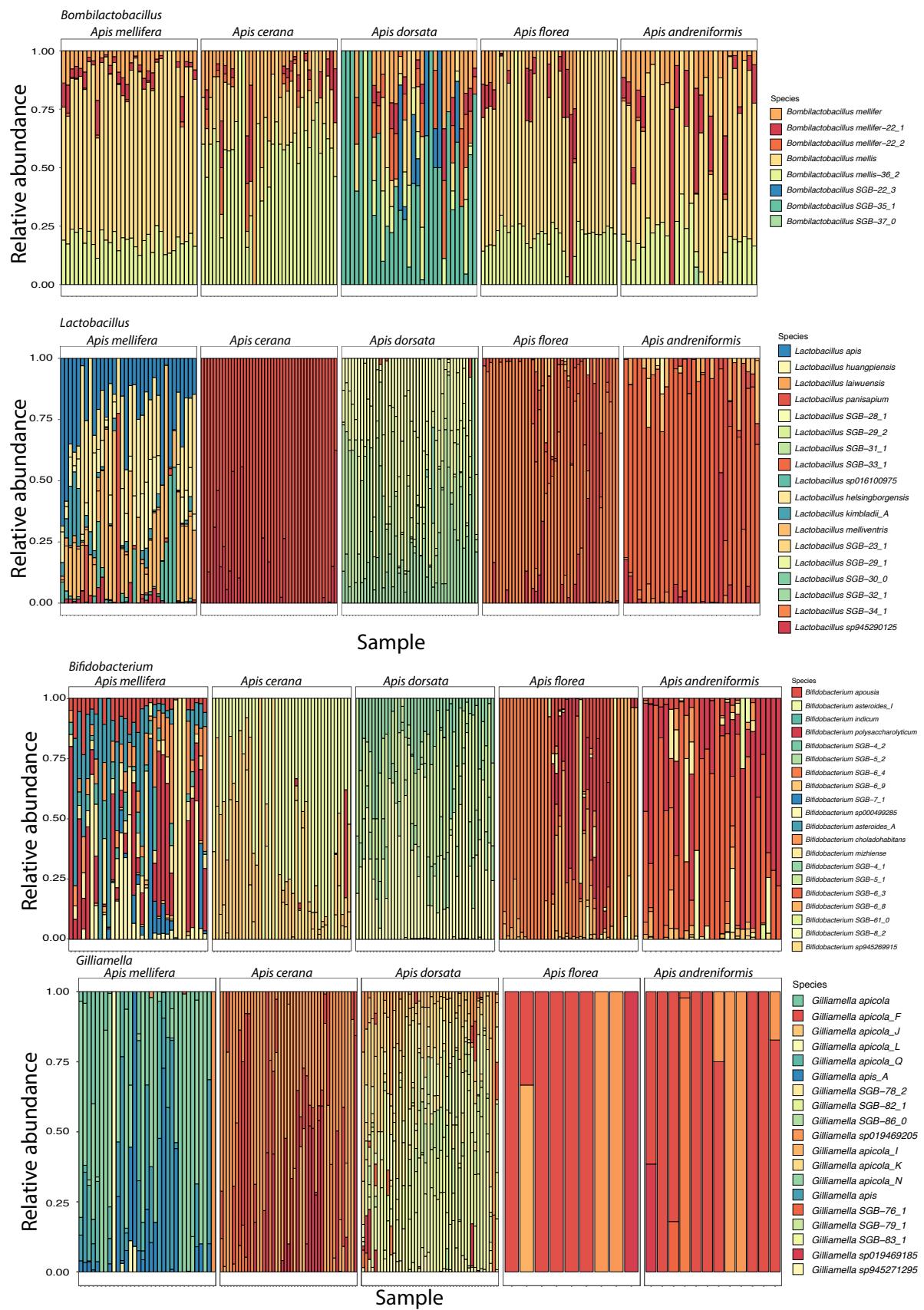
Supplementary Figure S7. Number of species from shared genera per individual. Boxplot showing the number of species only from genera that are detected in at least one bee of all host species (namely, *Bombilactobacillus*, *Lactobacillus*, *Bifidobacterium*, *CALYQJ01*, *Gilliamella*, *Snodgrassella*, *Frischella* and *Bartonella*).

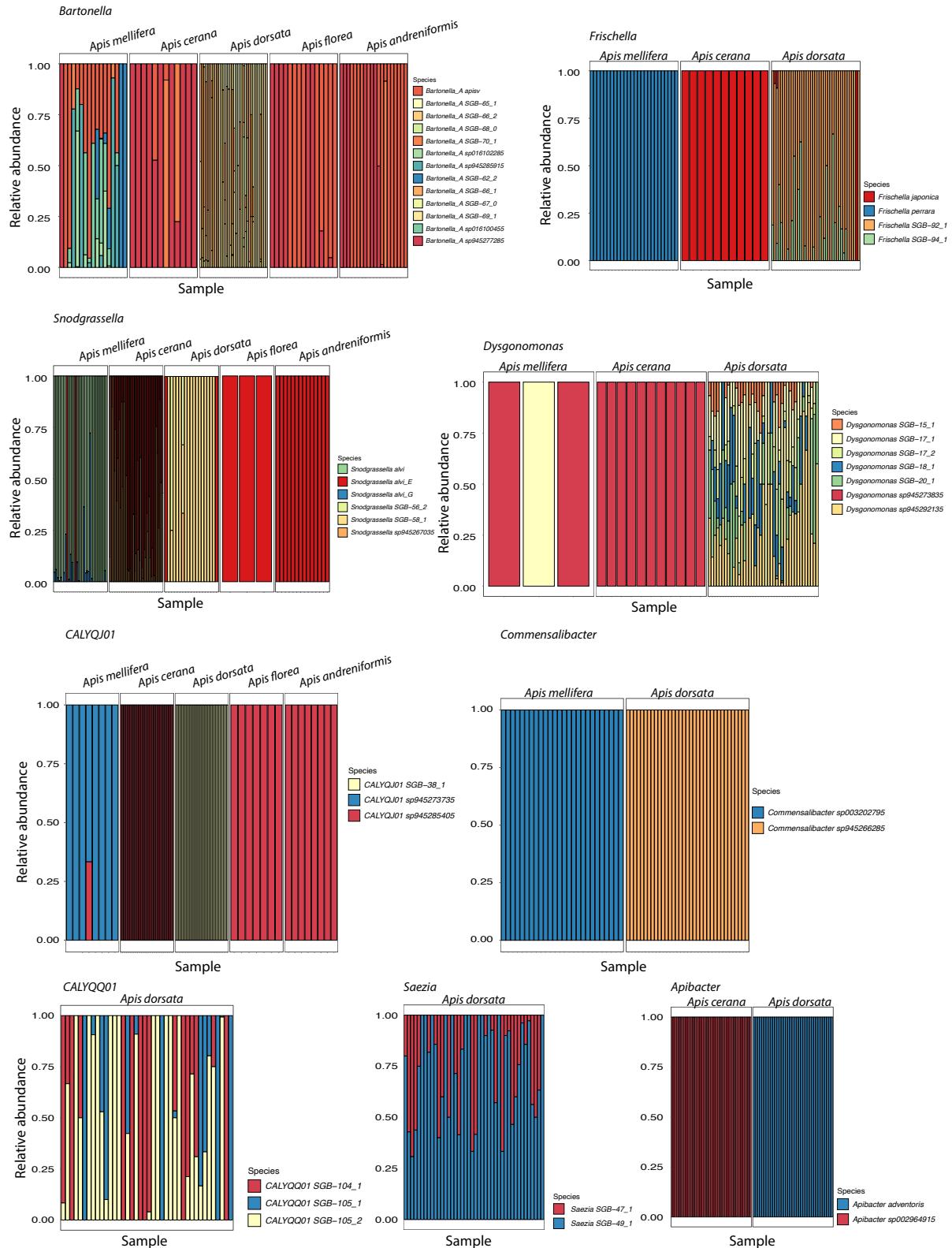


Supplementary Figure S8. Strain level diversity of each bacterial species across samples. Boxplots showing the percentage of variable sites (y-axis) marked by SNPs including samples grouped by hosts and for each bacterial species (x-axis)

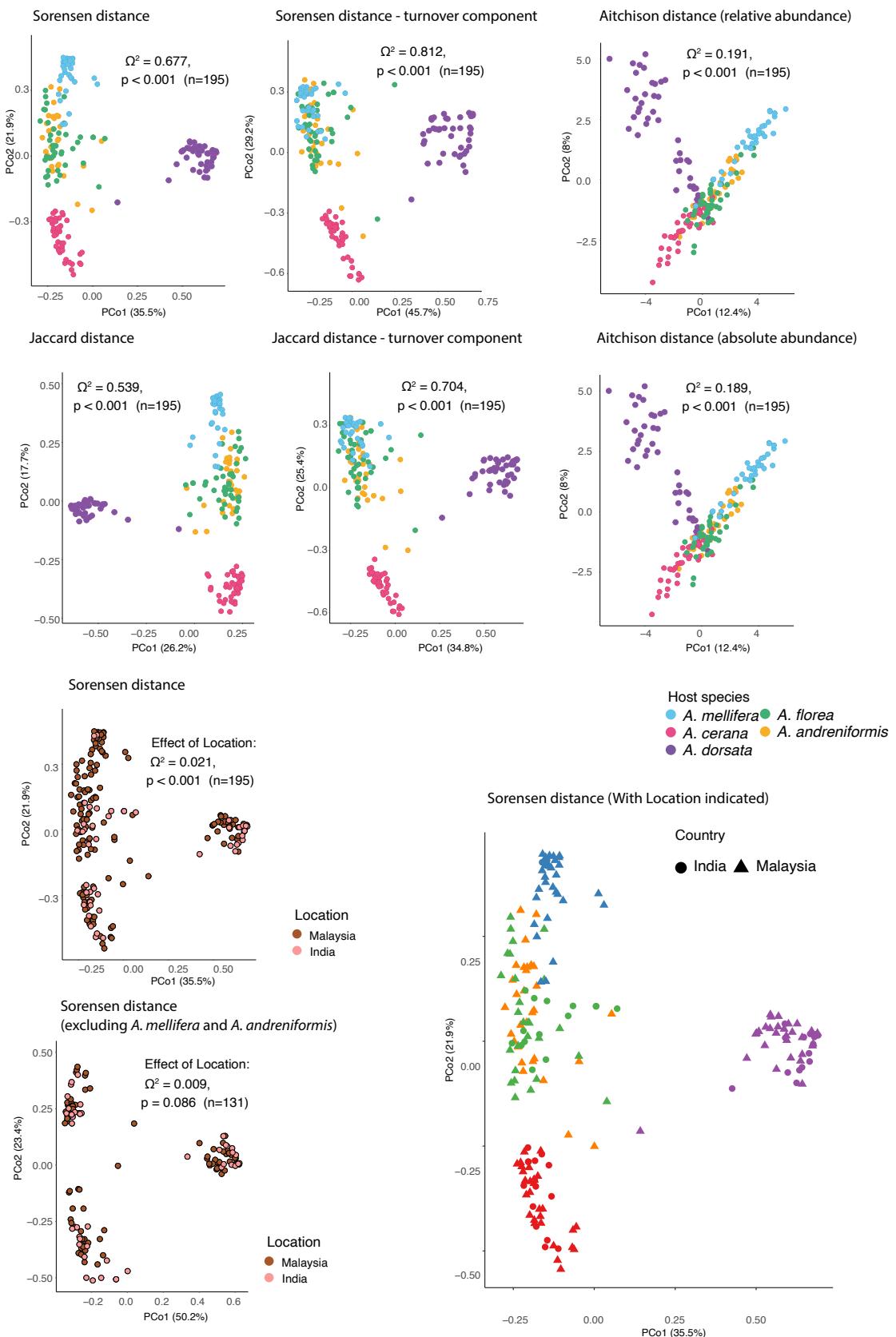


Supplementary Figure S9. Relative abundance of bacterial genera across samples. Barplots showing the relative abundance of bacterial genera across samples grouped by host species with white lines within each genus indicating different bacterial species within the genus.





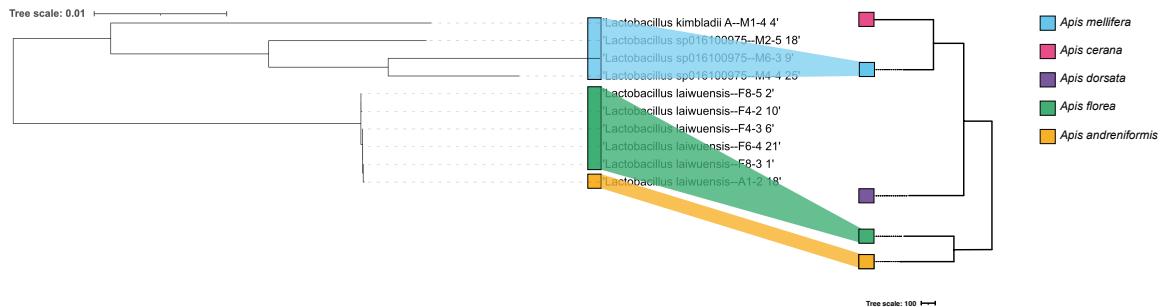
Supplementary Figure S10. Barplots of the relative abundance of bacterial species by genus. Barplots summarizing the relative abundance of bacterial species by genus across samples from different host species. Only species found in at least two samples of a given host species and samples that contain at least one bacterial species of the respective genus are included.



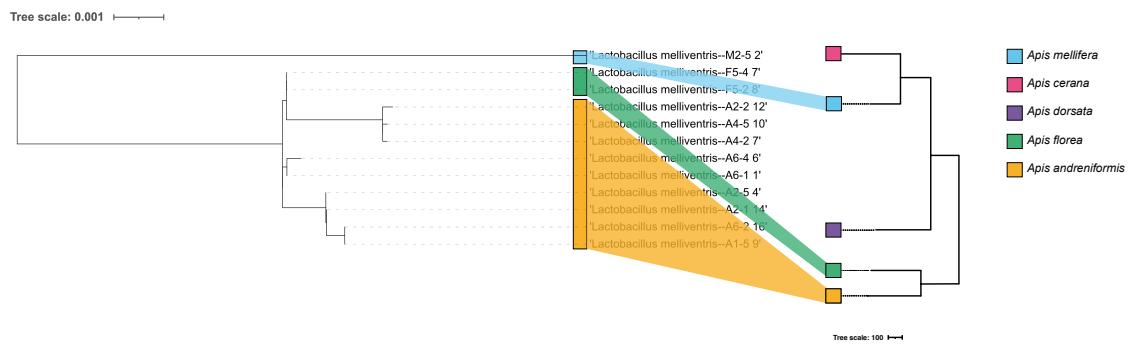
Supplementary Figure S11. Beta-diversity of bacterial community across host species. PCoA of taxonomic diversity using various distance metrics as indicated above each plot. Results of the PERMANOVA test for the effect of host species are shown on each panel (unless otherwise indicated). For one of the plots by location, two

host species that were not sampled equally across locations were excluded to test for the significance of sampling location on the bacterial species diversity more accurately.

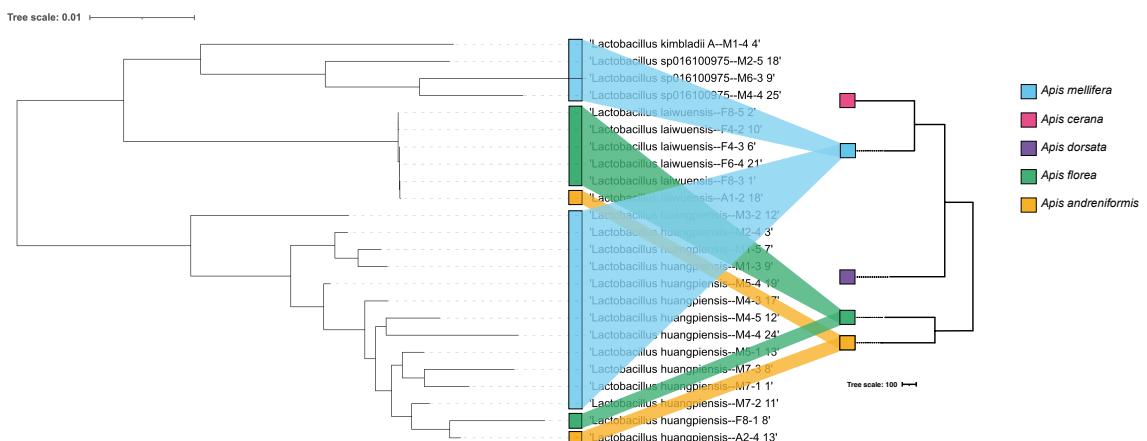
A Strict threshold ($r=0.77, p=0.002$)



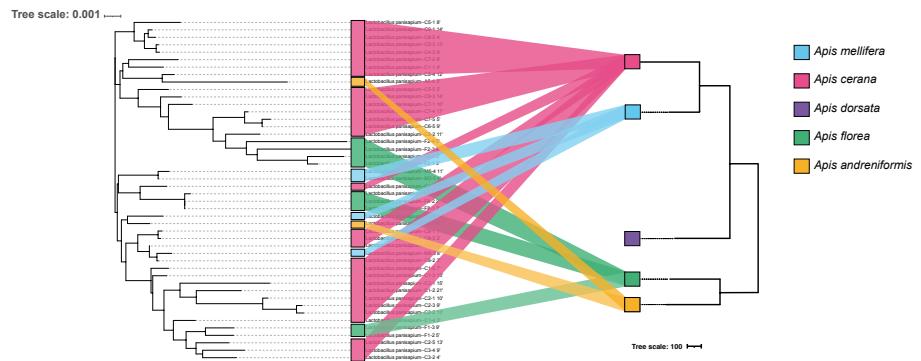
B Medium threshold ($r=0.79, p=0.021$)



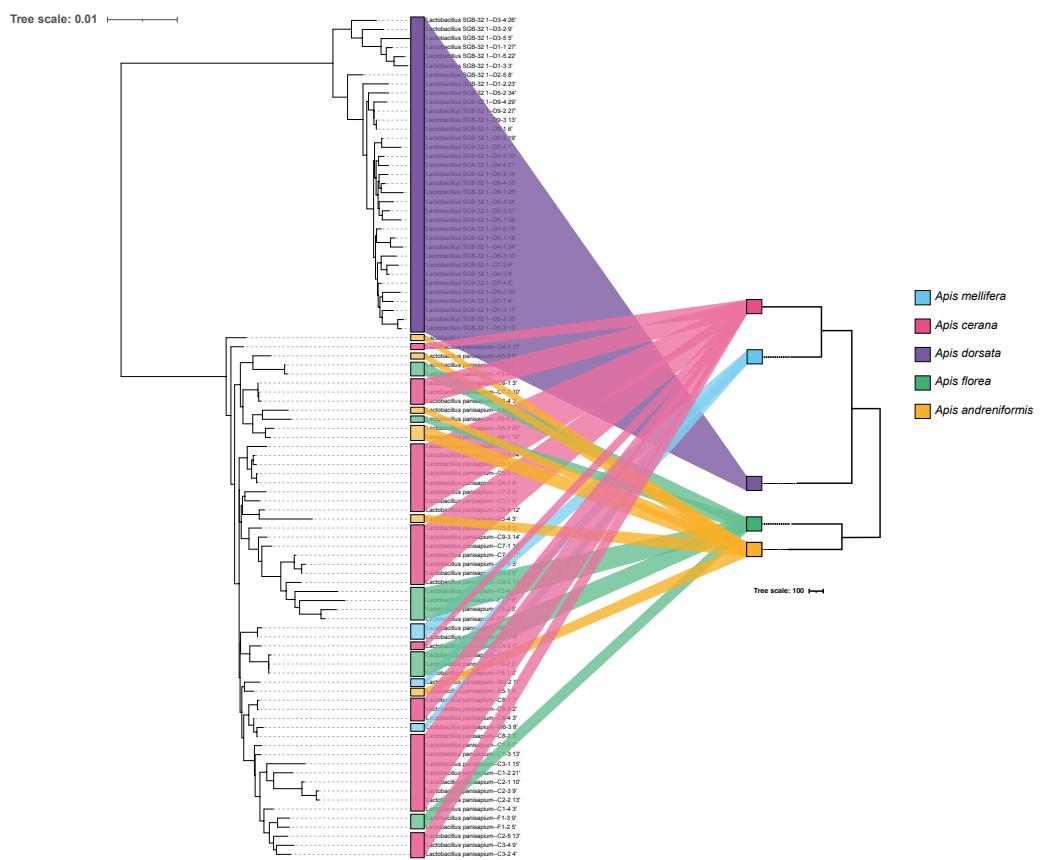
C Relaxed threshold ($r=0.29, p=0.006$)



D Relaxed threshold ($r=0.261, p=0.007$)



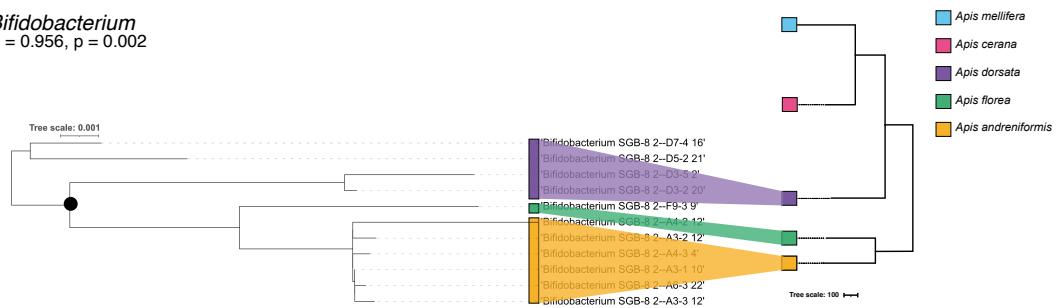
E Relaxed threshold ($r=0.496, p=0.001$)



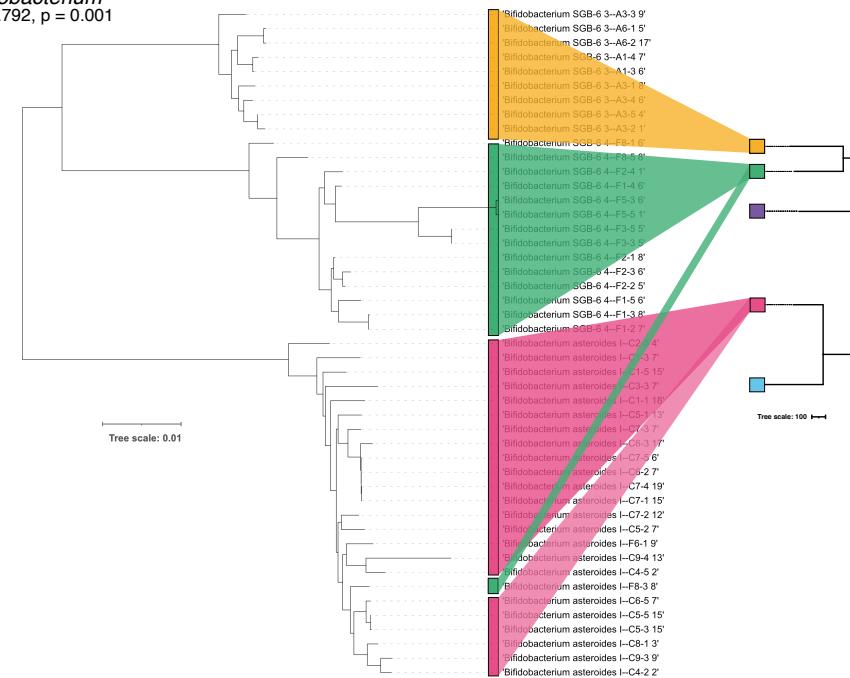
Supplementary Figure S12. Tanglegrams of host (right) and symbiont (left) phylogenies for symbiont nodes showing significant signal for topological congruence at different thresholds (strict, medium, relaxed). Each tip of the tree is labeled with the corresponding species and MAG name. The colors indicate the host species from which it was recovered. The host tree is based on actual data and to scale according to source (adapted(Carr, 2023)). The host tree scale bar is the inferred number of nucleotide substitutions. The symbiont scale bar represents 0.1 amino-acid substitutions per site.

A

Bifidobacterium
R = 0.956, p = 0.002

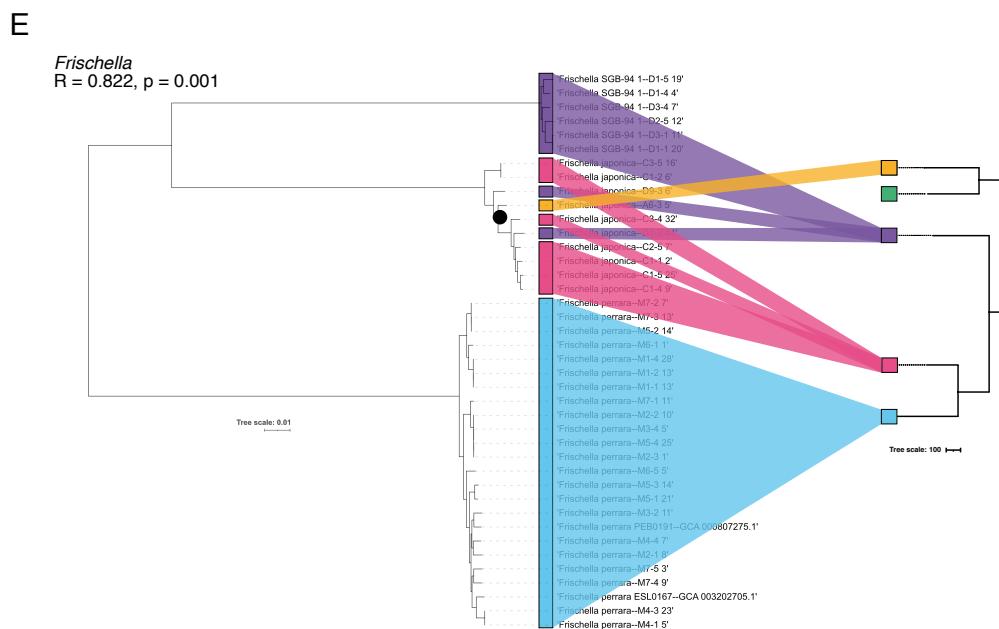
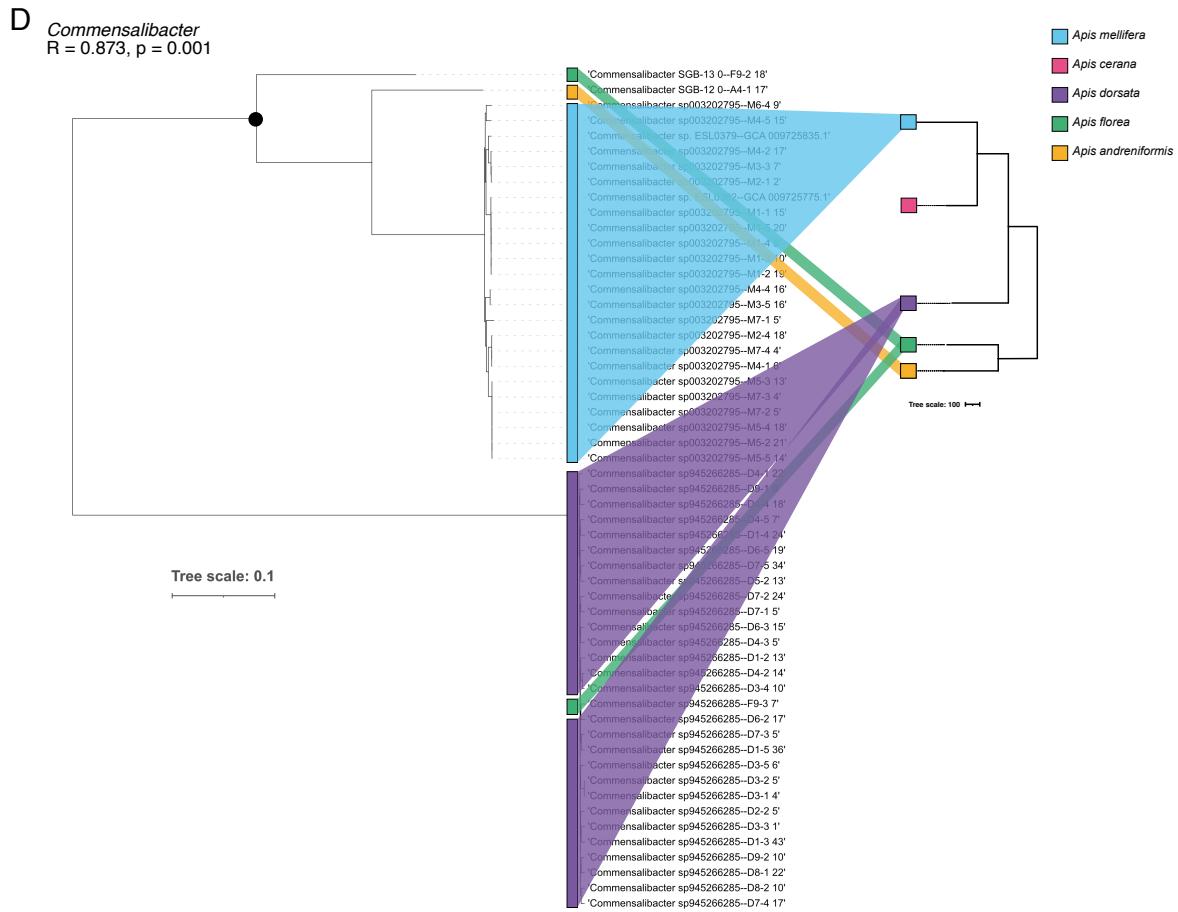
**B**

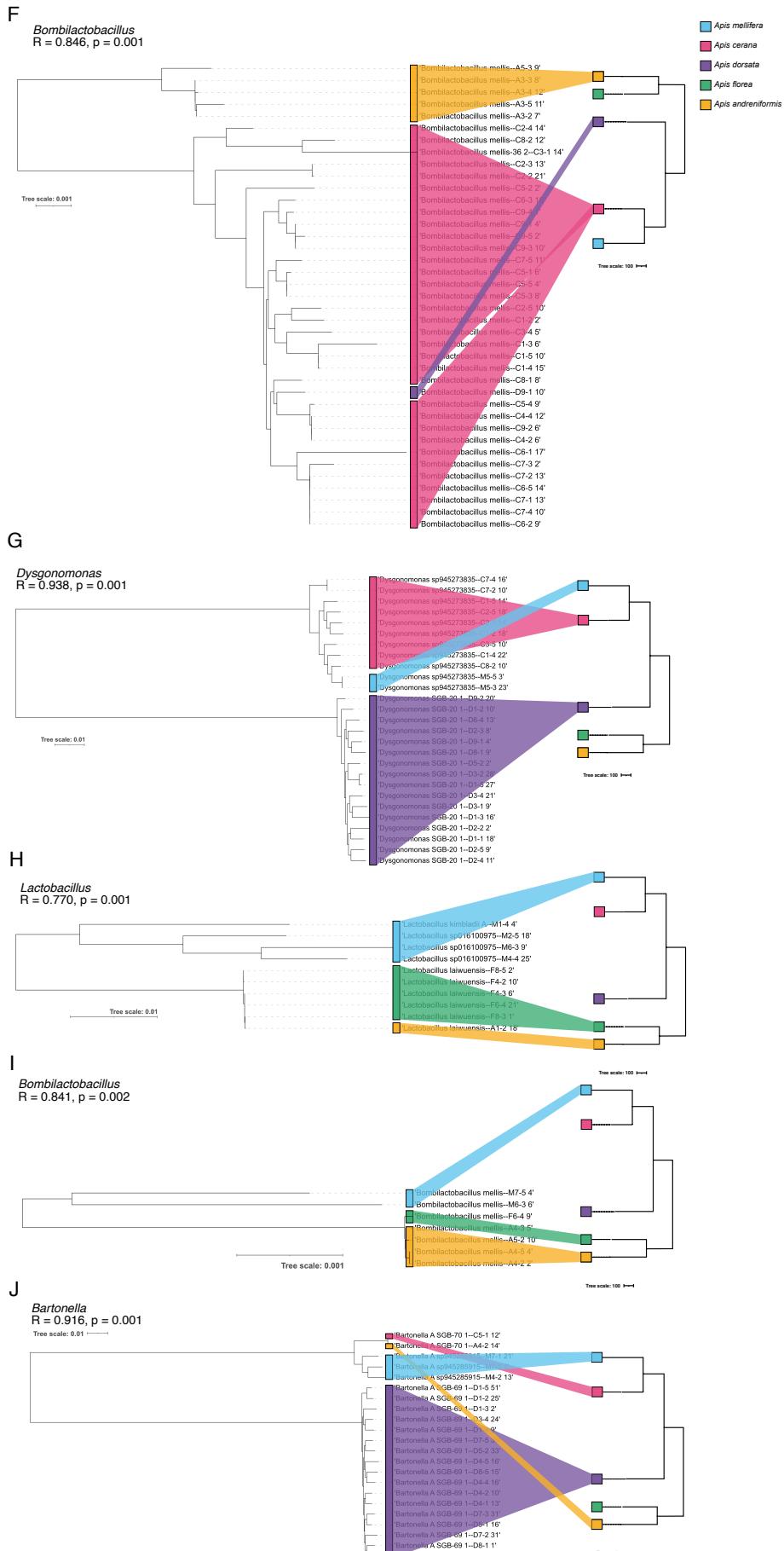
Bifidobacterium
R = 0.792, p = 0.001

**C**

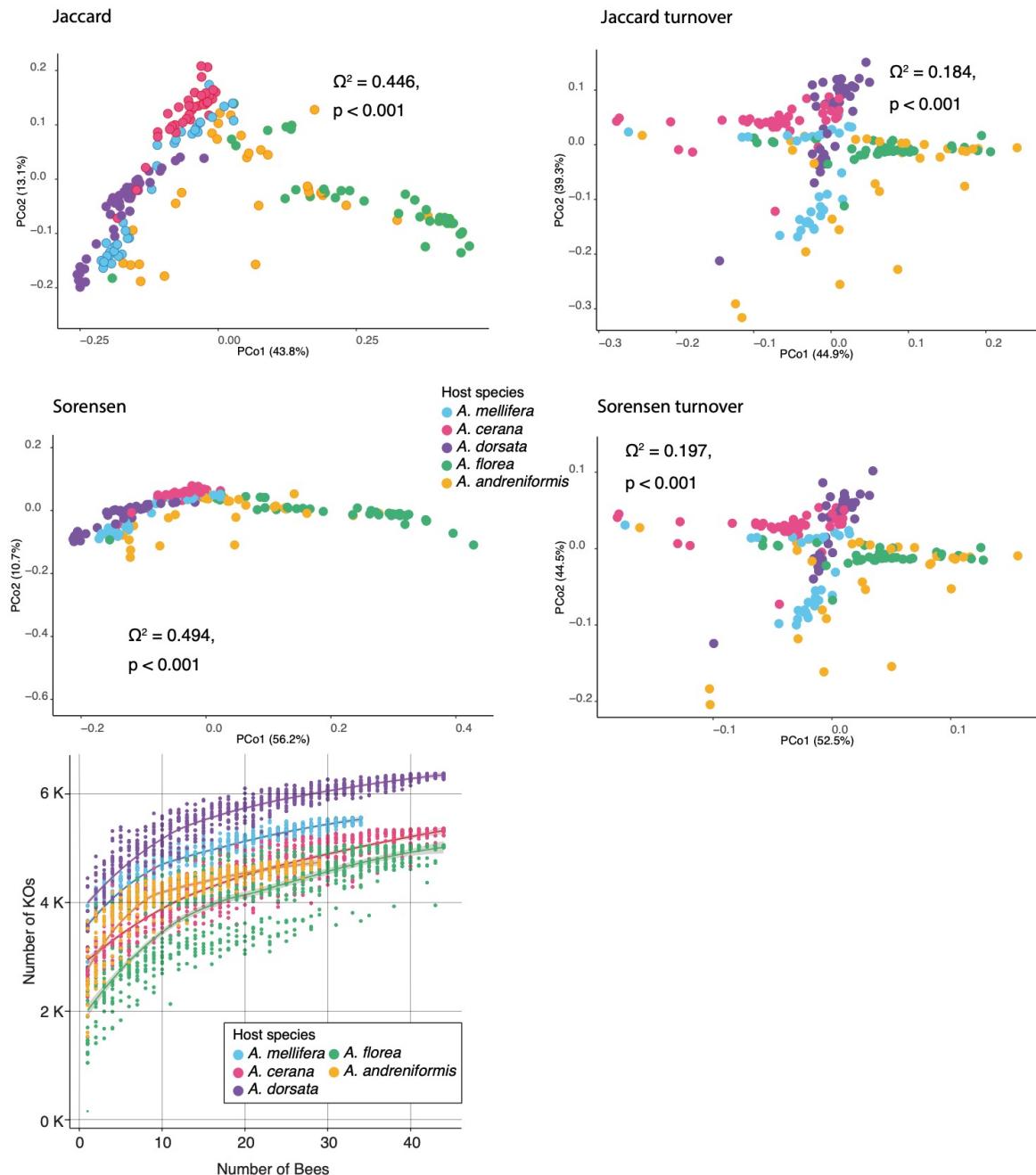
Snodgrassella
R = 0.961, p = 0.001



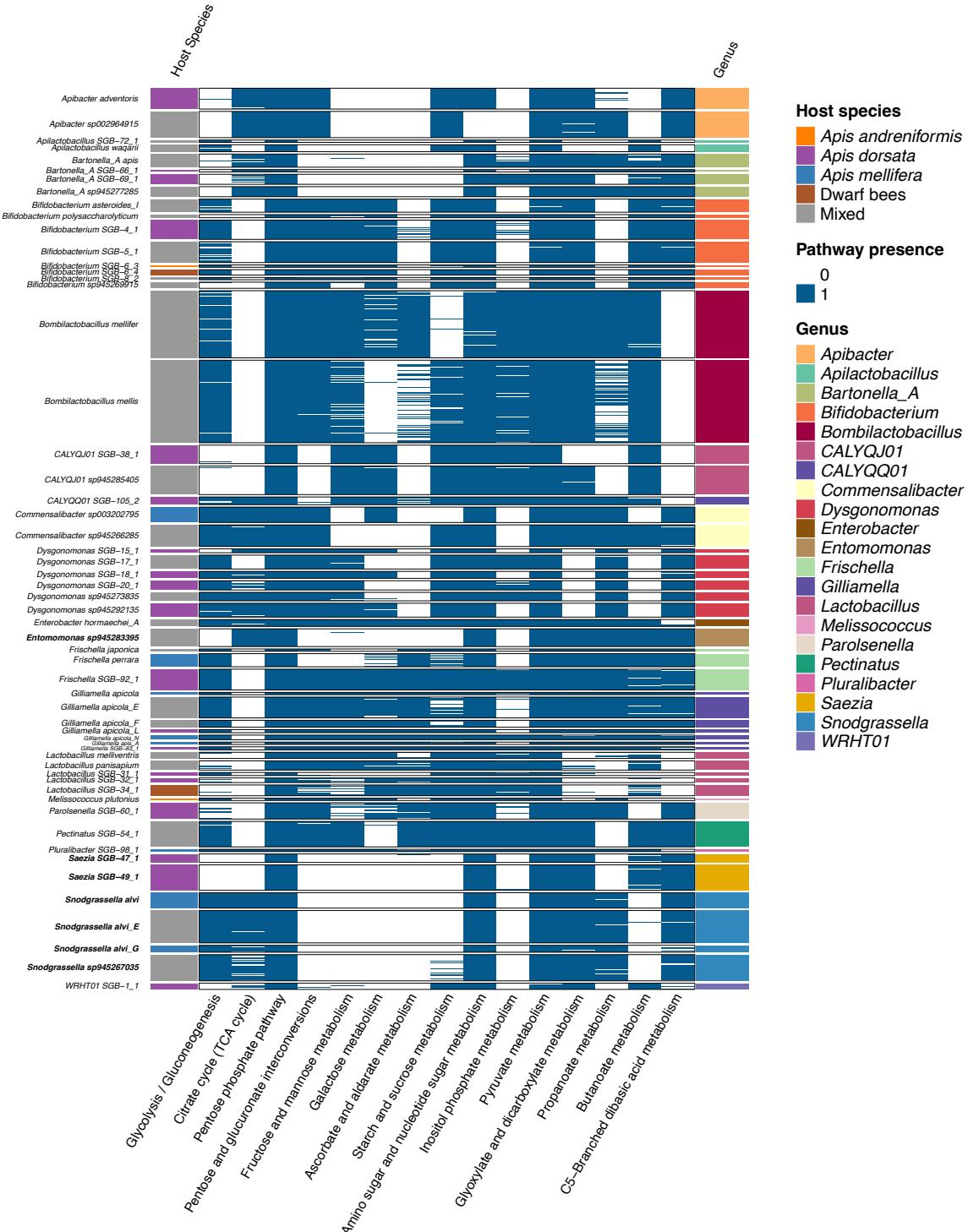




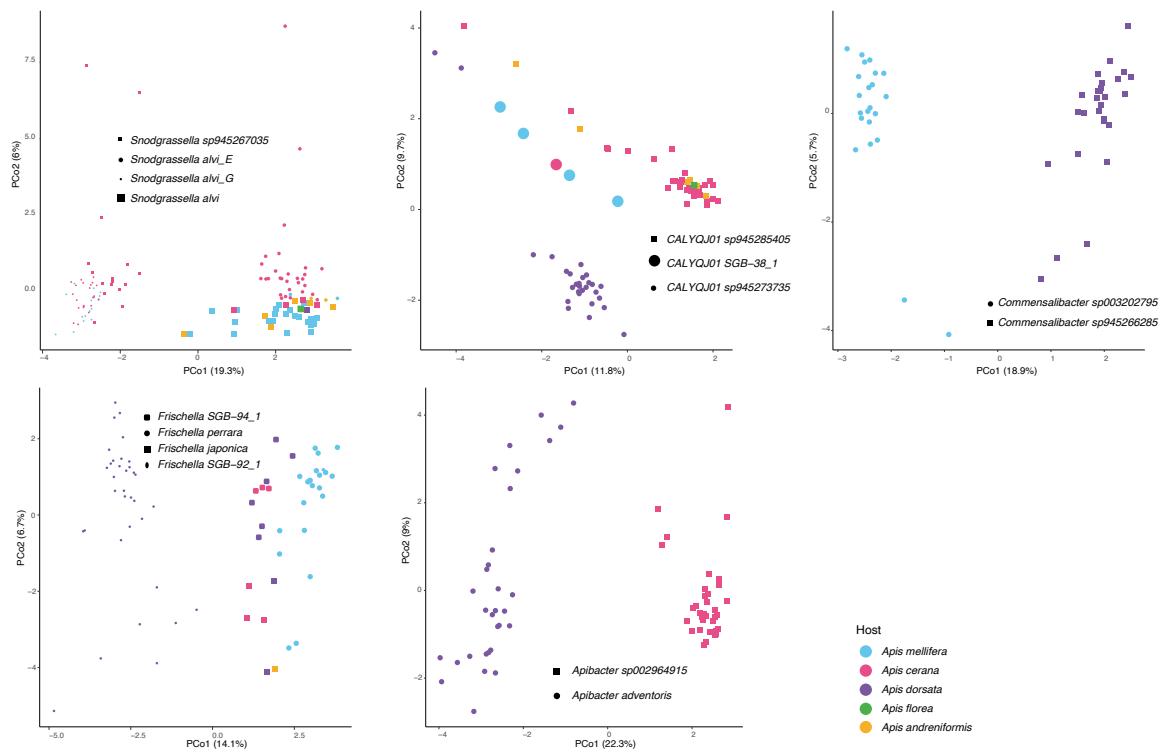
Supplementary Figure S13. Tanglegrams of host (right) and symbiont (left) phylogenies for symbiont nodes showing significant signal for topological congruence ($R > 0.75$, $p < 0.01$). These trees correspond to the subtrees of the nodes indicated with a black filled circle in Fig. 3 in the main text. In three cases, a significant node was nested within the depicted subtrees (see black circle in panel A, D and E). Each tip of the tree is labelled with the corresponding species and MAG name. The colors indicate the host species from which it was recovered. The host tree is based on actual data and to scale according to source (adapted(Carr, 2023)). The host tree scale bar is the inferred number of nucleotide substitutions. The symbiont scale bar represents 0.1 amino-acid substitutions per site. In contrast to the trees shown in panel A-D, which show clear evidence of co-diversification, subtrees in panel E and F are not congruent with the host tree topology, and in the subtrees in panel G-J, some branches between symbionts from different hosts are not consistent with host divergence. For example, there is little to no divergence between MAGs from *A. florea* and *A. andreniformis* in the subtrees of panel H and I, despite the fact that these two bee species have diverged 6 mya. Also there is little divergence between MAGs from *A. mellifera* and *A. cerana* in subtrees in panel G and J relatively to the very deeply branching sister clade of *A. dorsata* MAGs.



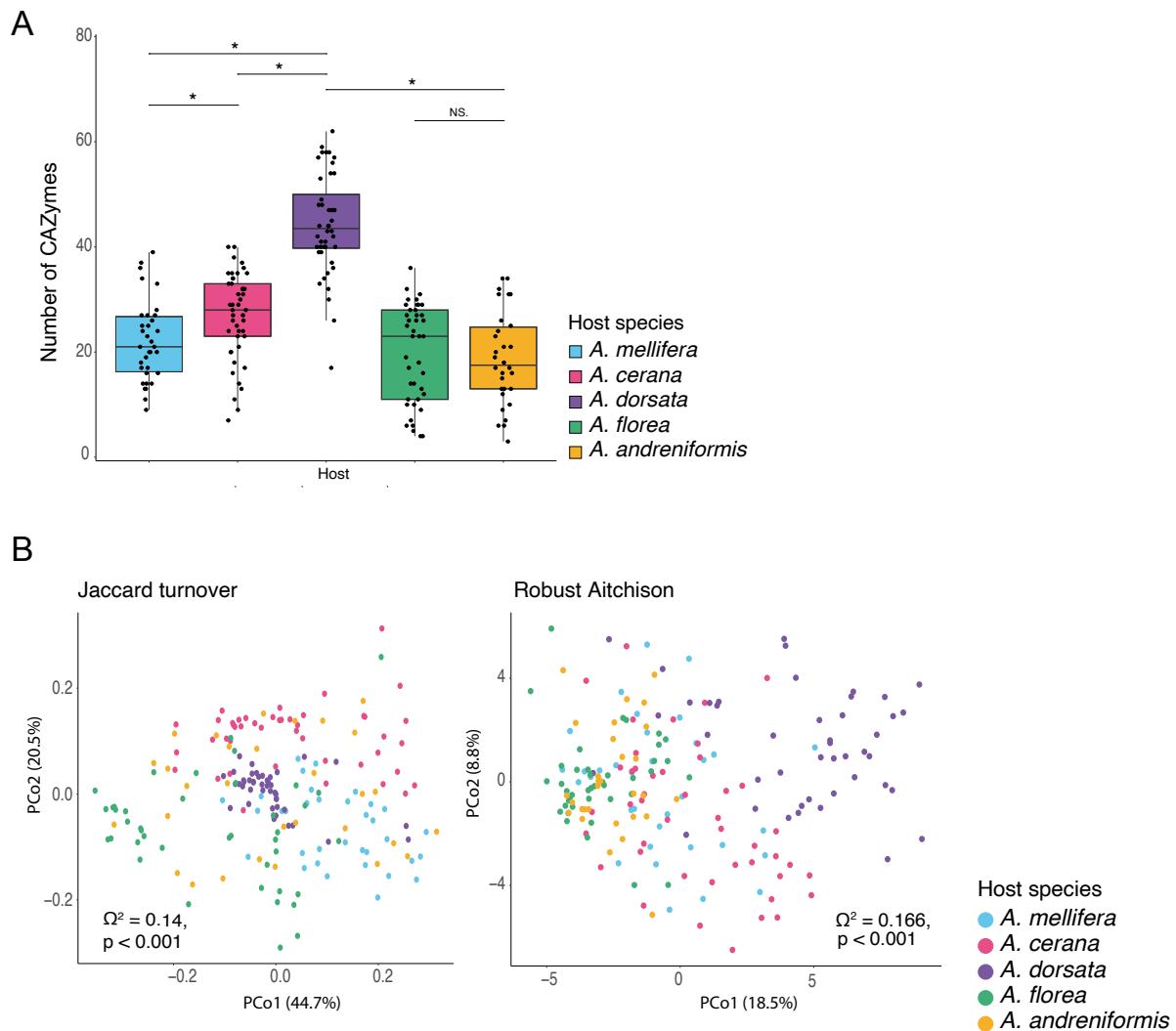
Supplementary Figure S14. Community-level KO diversity. (A) PCoA plots using various distance metrics on the KO presence-absence matrix with the result of PERMANOVA for the effect of host species shown on each panel. (B) Cumulative curves of number KOs vs individuals sampled for each host species.



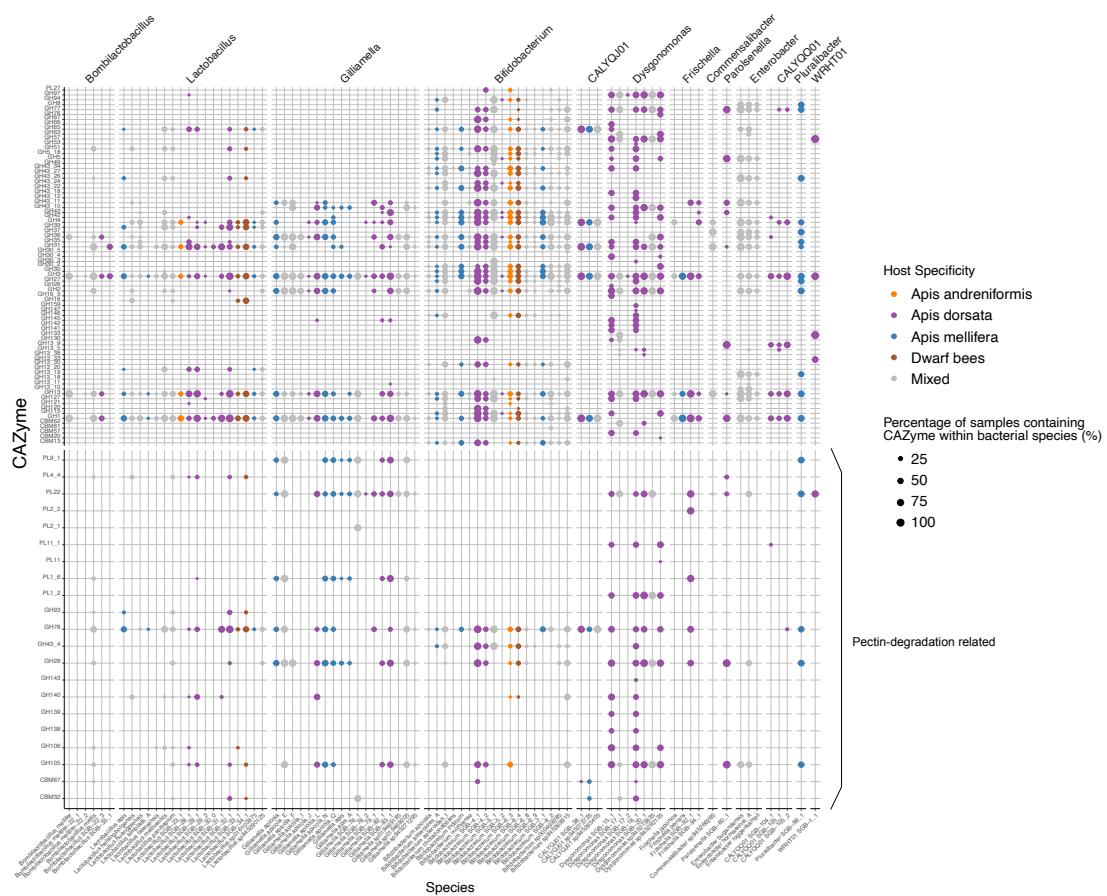
Supplementary Figure S15. Presence of KEGG Pathways in MAGs. The heatmap shows high-quality MAGs of different bacterial species as rows and the presence of KEGG pathways shown as columns inferred from the KOs encoded by the respective MAGs using MinPath. Similar heatmaps including the most prevalent genera and several other pathways are available in the Zenodo (<https://zenodo.org/doi/10.5281/zenodo.13732977>) repository.



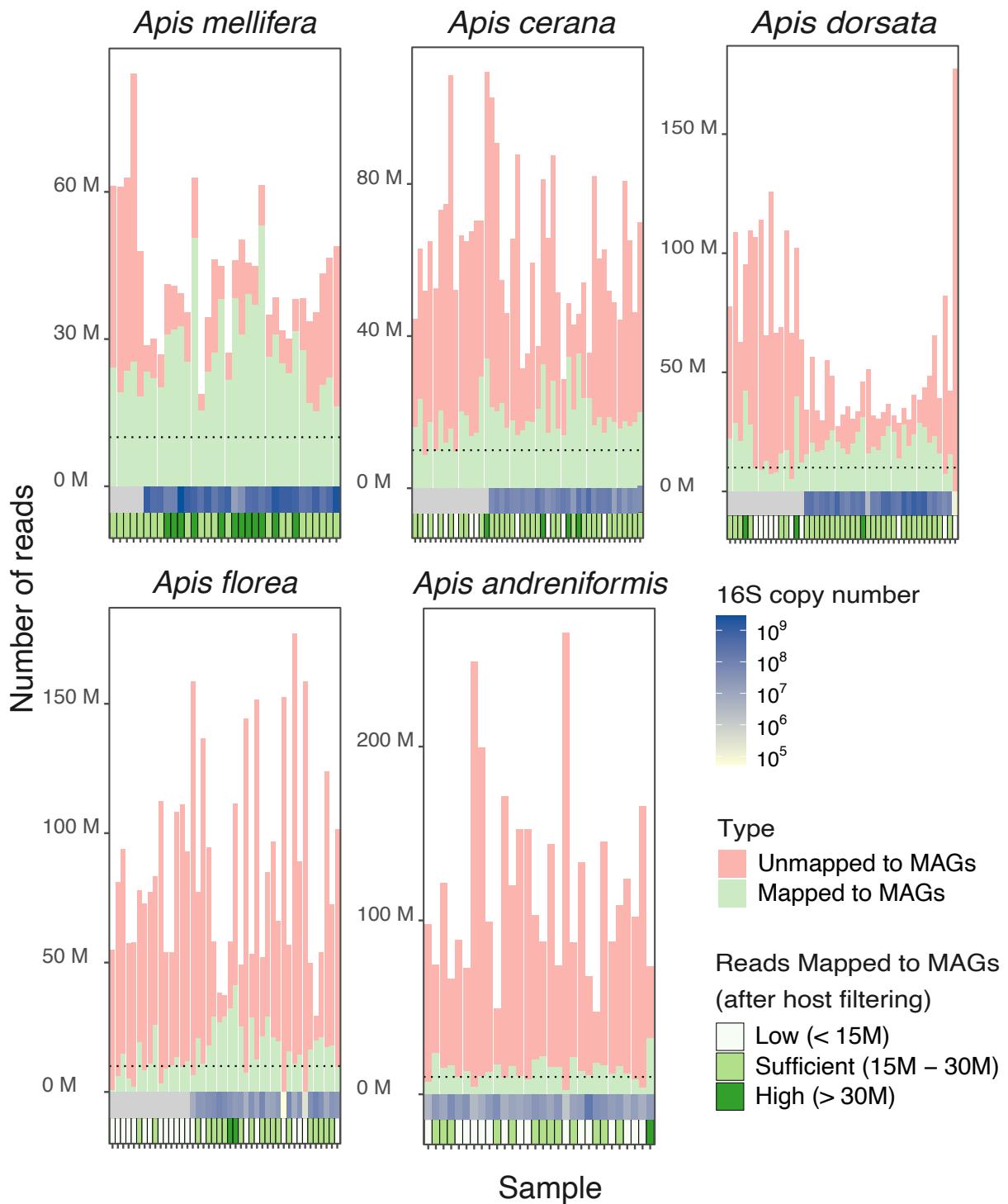
Supplementary Figure S16. Functional potential of MAGs across bacterial species within the same genus. PCoA plots of all KOs from MAGs using the Robust-Aitchison distance on the RPKM matrix of bacterial species within a genus with color and shape indicating the host species from which it was recovered and the bacterial species to which it was assigned. Some randomly selected genera and species showing the patterns discussed are displayed.



Supplementary Figure S17. Diversity of CAZymes across honeybee species. (A) Boxplot showing the number of CAZymes detected per individual bee gut across host species. (B) PCoA plots using various distance metrics on the CAZyme RPKM matrix with the result of PERMANOVA for the effect of host species shown on each panel.



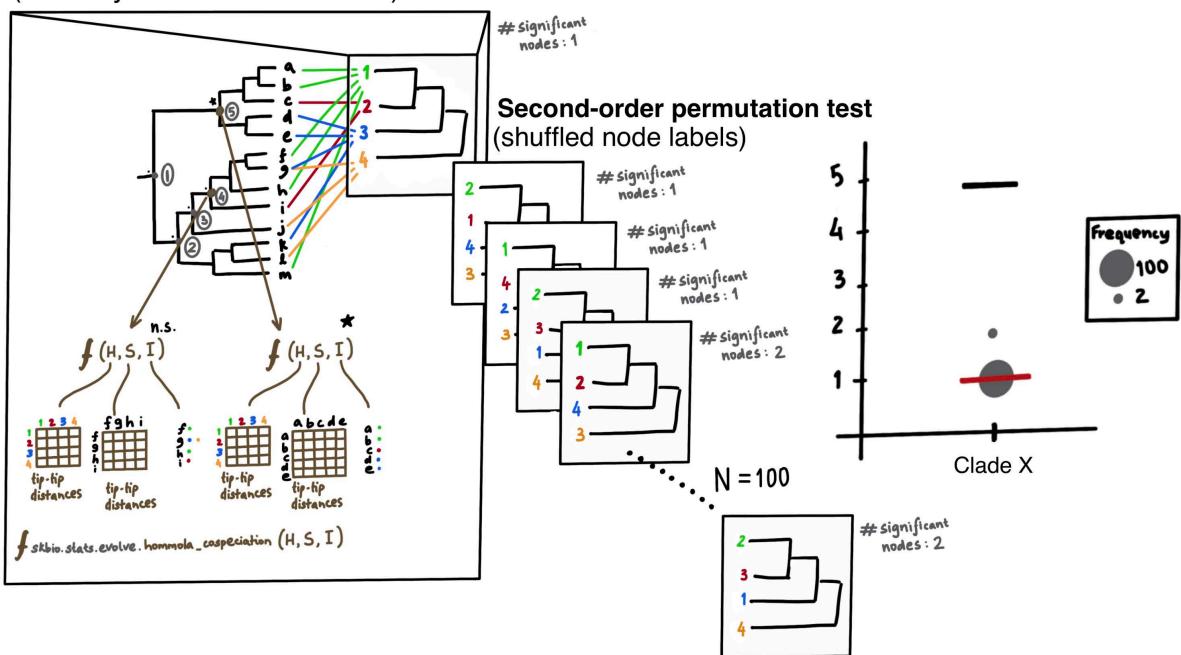
Supplementary Figure S18. CAZyme families detected in each bacterial species. Dotplot showing the prevalence (percentage) of CAZymes in each bacterial species. Calculated as the ratio between the number of samples in which the CAZyme is detected on MAGs from the focal bacterial species and the number of samples in which the species was detected. Since this includes several medium-quality MAGs with completeness of only over 50%, CAZymes may not be detected in all MAGs of the species, even if present. The color represents the host species in which the respective bacterial species are found. If it is not a host-specific species, it is marked as mixed. Genera containing at least 5 CAZymes and found in at least 5 samples are included in the plot above (excluding *Escherichia*, *Klebsiella* and *Pantoea*).



Supplementary Figure S19. Read mapping to recovered MAGs. The total number of reads ($R_1 + R_2$) from each sample, with the colors showing the number of reads mapped to the MAGs database upon directly mapping trimmed and quality-filtered reads. The dotted line in each panel is drawn at 10M. The tiles below indicate the qPCR copy number using 16S rRNA universal primers (qPCR data not available for samples from India), and the bottom-most tiles show the extent of reads mapped to MAGs after excluding reads mapped to the host genome database.

Clade X

Actual comparison
(node-by-node Hommola test)



Supplementary Figure S20. Schematic of co-diversification testing approach. This schematic describes the approach used to test for co-diversification. For a given clade (in the case of our trees, genus), the nodes that pass filtering thresholds (e.g., here, at least 4 MAGs and at least 3 host species represented) are identified. The Hommola cospeciation test (as implemented in the Python package skbio) is applied to each identified node in the tree, considering, in each case, the host matrix (H) and symbiont matrix (S) containing tip-tip distances from phylogeny and the host-symbiont interactions list (I). Subsequently, the number of significant nodes in the tree can be found. Next, to establish a false-discovery (FDR) rate, an expectation for the number of significant nodes can be computed using second-order permutation tests with tip labels of the host tree shuffled in different permutations. Using this, an FDR can be established to indicate whether the significant node(s) identified in the actual comparison is(are) likely to be false positive(s). The black horizontal line on the final plot on the right represents the total number of nodes tested, the grey circles are the number of significant nodes across second-order permutation tests with their size indicating the frequency, and the horizontal red line is the number of significant nodes in the actual comparison.

SUPPLEMENTARY TABLES

Supplementary tables can be found in the file at this link:

https://unils-my.sharepoint.com/:g/personal/aiswarya_prasad_unil_ch/EWOEbD_rcP5NvtOOAUy-0J0Bbbw48I8kCxZewn9HK9Wjqw?e=2y2D4B

Supplementary Table S1. Sample collection information and metadata.

Supplementary Table S2. Metadata about MAGs including fields (in bold) required by the MIMAG standards.

Supplementary Table S3. Sequencing depth and Host and MAG database mapping information.

Supplementary Table S4. Result of multivariate analyses (PERMANOVA test) applied on strain-level profiles for each bacterial species. Column n displays the number of samples in which the species was detected with enough coverage to test for the effect of host species or country on strain-level in a PERMANOVA analysis and the p-value and omega squared values for the respective tests are shown. Host specificity is based on the Rohde's index if strain-level comparison is not available. Values < 1 for which there were not enough samples with good enough coverage for strain-level comparison are considered generalist if there is enough coverage, they are labelled specialist if detected in only one host species, and shared (strain-level specialist) if found in more than one host species but the PERMANOVA test is significant for effect of host species.

Supplementary Table S5. All nodes in original trees identified by hommola test as significant ($p < 0.05$) with the line separating those that are detected with strict thresholds (above the line). sym_subtree displays a Newick string of the subtree at the node being considered. host_tip span is the number of host species represented in the subtree, sym_tips_count the number of MAGs in the subtree.

Supplementary Table S6. Number of nodes tested and found significant under various thresholds. "# significant" is the number of nodes under the respective threshold, "median # significant" is the median number of significant nodes in (N=100) second-order permutation test comparisons with randomly shuffled host tip labels and "SD # significant" is the standard deviation.

Supplementary Table S7. Tip-tip distance between MAGs of the genus in nucleotide tree constructed using 120 core genes (bac120 as identified in GTDB-tk).

REFERENCES

- Arora, J., Buček, A., Hellemans, S., Beránková, T., Romero Arias, J., Fisher, B. L., Clitheroe, C., Brune, A., Kinjo, Y., Šobotník, J., & Bourguignon, T. (2023). Evidence of cospeciation between termites and their gut bacteria on a geological time scale. *Proceedings of the Royal Society B: Biological Sciences*, 290(2001), 20230619. <https://doi.org/10.1098/rspb.2023.0619>
- Baselga, A., & Orme, C. D. L. (2012). betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution*, 3(5), 808–812. <https://doi.org/10.1111/j.2041-210X.2012.00224.x>
- Bilen, M., Fonkou, M. D. M., Dubourg, G., Tomei, E., Richez, M., Delerce, J., Levasseur, A., Daoud, Z., Raoult, D., & Cadoret, F. (2019). *Dysgonomonas massiliensis* sp. Nov., a new species isolated from the human gut and its taxonogenomic description. *Antonie van Leeuwenhoek*, 112(6), 935–945. <https://doi.org/10.1007/s10482-019-01227-1>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., & Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife*, 10, e68583. <https://doi.org/10.7554/eLife.68583>
- Buck, M., Mehrshad, M., & Bertilsson, S. (2022). mOTUpa: A robust Bayesian approach to leverage metagenome-assembled genomes for core-genome estimation. *NAR Genomics and Bioinformatics*, 4(3), lqac060. <https://doi.org/10.1093/nargab/lqac060>
- Carr, S. M. (2023). Multiple mitogenomes indicate Things Fall Apart with Out of Africa or Asia hypotheses for the phylogeographic evolution of Honey Bees (*Apis mellifera*). *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-35937-4>
- Cerdeira, A. E. S., Hammer, T. J., Moran, N. A., Santana, W. C., Kasuya, M. C. M., & da Silva, C. C. (2021). Extinction of anciently associated gut bacterial symbionts in a clade of stingless bees. *The ISME Journal*, 15(9), 2813–2816. <https://doi.org/10.1038/s41396-021-01000-1>
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6), 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
- Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, 17(12), 764–775. <https://doi.org/10.1038/s41579-019-0242-1>
- Ducarmon, Q. R., Karcher, N., Tytgat, H. L. P., Delannoy-Bruno, O., Pekel, S., Springer, F., Schudoma, C., & Zeller, G. (2024). Large-scale computational analyses of gut microbial CAZyme repertoires enabled by Cayman (p. 2024.01.08.574624). bioRxiv. <https://doi.org/10.1101/2024.01.08.574624>
- Dyer, F. C., & Seeley, Th. D. (1994). Colony migration in the tropical honey bee *Apis dorsata* F. (Hymenoptera: Apidae). *Insectes Sociaux*, 41(2), 129–140. <https://doi.org/10.1007/BF01240473>
- Ellegaard, K. M., Brochet, S., Bonilla-Rosso, G., Emery, O., Glover, N., Hadadi, N., Jaron, K. S., Meer, J. R., Robinson-Rechavi, M., Sentchilo, V., Tagini, F., SAGE class 2016-17, & Engel, P. (2019). Genomic changes underlying host specialization in the bee gut symbiont *Lactobacillus Firm5*. *Molecular Ecology*, 28(9), 2224–2237. <https://doi.org/10.1111/mec.15075>
- Ellegaard, K. M., & Engel, P. (2019). Genomic diversity landscape of the honey bee gut microbiota. *Nature Communications*, 10(1), 446. <https://doi.org/10.1038/s41467-019-08303-0>

Ellegaard, K. M., Suenami, S., Miyazaki, R., & Engel, P. (2020). Vast Differences in Strain-Level Diversity in the Gut Microbiota of Two Closely Related Honey Bee Species. *Current Biology*, 30(13), 2520-2531.e7. <https://doi.org/10.1016/j.cub.2020.04.070>

Ellegaard, K. M., Tamarit, D., Javelind, E., Olofsson, T. C., Andersson, S. G., & Vásquez, A. (2015). Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics*, 16(1), 284. <https://doi.org/10.1186/s12864-015-1476-6>

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>

Engel, P., Martinson, V. G., & Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences*, 109(27), 11002–11007. <https://doi.org/10.1073/pnas.1202970109>

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, 37(5), 699–735. <https://doi.org/10.1111/1574-6976.12025>

Fouks, B., Brand, P., Nguyen, H. N., Herman, J., Camara, F., Ence, D., Hagen, D. E., Hoff, K. J., Nachweide, S., Romoth, L., Walden, K. K. O., Guigo, R., Stanke, M., Narzisi, G., Yandell, M., Robertson, H. M., Koeniger, N., Chantawannakul, P., Schatz, M. C., ... Rueppell, O. (2021). The genomic basis of evolutionary differentiation among honey bees. *Genome Research*, 31(7), 1203–1215. <https://doi.org/10.1101/gr.272310.120>

Ganeshprasad, D. N., Jani, K., Shouche, Y. S., & Sneharani, A. H. (2021). *Gut Bacterial Inhabitants of Open Nested Honey Bee, Apis Florea* [Preprint]. In Review. <https://doi.org/10.21203/rs.3.rs-225332/v1>

Ge, Y., Jing, Z., Diao, Q., He, J.-Z., & Liu, Y.-J. (2021). Host Species and Geography Differentiate Honeybee Gut Bacterial Communities by Changing the Relative Contribution of Community Assembly Processes. *mBio*, 12(3), e00751-21. <https://doi.org/10.1128/mBio.00751-21>

Ghonche-Golan, S., Nazemi-Rafie, J., & Rezapanaah, M. (2022). The relationship study among *Apis* spp. Using mitochondrial markers, Procrustes coordinates and residuals of geometric morphometric method. *Biologia*, 77(9), 2519–2530. <https://doi.org/10.1007/s11756-022-01123-6>

Gile, G. H. (2024). Protist symbionts of termites: Diversity, distribution, and coevolution. *Biological Reviews*, 99(2), 622–652. <https://doi.org/10.1111/brv.13038>

Good, B. H. (2023). *Limited codiversification of the gut microbiota with humans* (p. 2022.10.27.514143). bioRxiv. <https://doi.org/10.1101/2022.10.27.514143>

Groussin, M., Mazel, F., Sanders, J. G., Smillie, C. S., Lavergne, S., Thuiller, W., & Alm, E. J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications*, 8(1), 14319. <https://doi.org/10.1038/ncomms14319>

Gruneck, L., Gentekaki, E., Khongphinitbunjong, K., & Popluechai, S. (2022). Distinct gut microbiota profiles of Asian honey bee (*Apis cerana*) foragers. *Archives of Microbiology*, 204(3), 187. <https://doi.org/10.1007/s00203-022-02800-5>

Gruneck, L., Khongphinitbunjong, K., & Popluechai, S. (2021). Gut microbiota associated with two species of domesticated honey bees from Thailand. *Symbiosis*, 83(3), 335–345. <https://doi.org/10.1007/s13199-021-00754-8>

Hammer, T. J., Dickerson, J. C., & Fierer, N. (2015). Evidence-based recommendations on storing and handling specimens for analyses of insect microbiota. *PeerJ*, 3, e1190. <https://doi.org/10.7717/peerj.1190>

Hepburn, H. R., & Radloff, S. E. (Eds.). (2011). *Honeybees of Asia*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-16422-4>

- Hommola, K., Smith, J. E., Qiu, Y., & Gilks, W. R. (2009). A Permutation Test of Host–Parasite Cospeciation. *Molecular Biology and Evolution*, 26(7), 1457–1468. <https://doi.org/10.1093/molbev/msp062>
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359. <https://doi.org/10.7717/peerj.7359>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kearney, S. M., Gibbons, S. M., Poyet, M., Gurry, T., Bullock, K., Allegretti, J. R., Clish, C. B., & Alm, E. J. (2018). Endospores and other lysis-resistant bacteria comprise a widely shared core community within the human microbiota. *The ISME Journal*, 12(10), 2403–2416. <https://doi.org/10.1038/s41396-018-0192-z>
- Kešnerová, L., Mars, R. A. T., Ellegaard, K. M., Troilo, M., Sauer, U., & Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLOS Biology*, 15(12), e2003467. <https://doi.org/10.1371/journal.pbio.2003467>
- Khan, K. A., Ganeshprasad, D. N., Sachin, H. R., Shouche, Y. S., Ghramh, H. A., & Sneharani, A. H. (2023). Gut microbial diversity in Apis cerana indica and Apis florea colonies: A comparative study. *Frontiers in Veterinary Science*, 10. <https://doi.org/10.3389/fvets.2023.1149876>
- Koch, H., Abrol, D. P., Li, J., & Schmid-Hempel, P. (2013). Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Molecular Ecology*, 22(7), 2028–2044. <https://doi.org/10.1111/mec.12209>
- Kotthoff, U., Wappler, T., & Engel, M. (2011). Miocene honey bees from the Randeck Maar of southwestern Germany (Hymenoptera, Apidae). *ZooKeys*, 96, 11–37. <https://doi.org/10.3897/zookeys.96.752>
- Kwong, W. K., Medina, L. A., Koch, H., Sing, K.-W., Soh, E. J. Y., Ascher, J. S., Jaffé, R., & Moran, N. A. (2017). Dynamic microbiome evolution in social bees. *Science Advances*, 3(3), e1600513. <https://doi.org/10.1126/sciadv.1600513>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Levin, D., Raab, N., Pinto, Y., Rothschild, D., Zanir, G., Godneva, A., Mellul, N., Futorian, D., Gal, D., Leviatan, S., Zeevi, D., Bachelet, I., & Segal, E. (2021). Diversity and functional landscapes in the microbiota of animals in the wild. *Science*, 372(6539), eabb5352. <https://doi.org/10.1126/science.abb5352>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, Y., Leonard, S. P., Powell, J. E., & Moran, N. A. (2022). Species divergence in gut-restricted bacteria of social bees. *Proceedings of the National Academy of Sciences*, 119(18), e2115013119. <https://doi.org/10.1073/pnas.2115013119>
- Lim, S. J., & Bordenstein, S. R. (2020). An introduction to phylosymbiosis. *Proceedings of the Royal Society B*. <https://doi.org/10.1098/rspb.2019.2900>
- Luo, S., Zhang, X., & Zhou, X. (2024). Temporospatial dynamics and host specificity of honeybee gut bacteria. *Cell Reports*, 43(7), 114408. <https://doi.org/10.1016/j.celrep.2024.114408>
- Mallick, H., Rahnavařd, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson,

J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology*, 17(11), e1009442. <https://doi.org/10.1371/journal.pcbi.1009442>

Mallott, E. K., & Amato, K. R. (2021). Host specificity of the gut microbiome. *Nature Reviews Microbiology*, 19(10), 639–653. <https://doi.org/10.1038/s41579-021-00562-3>

Mazel, F., Davis, K. M., Loudon, A., Kwong, W. K., Groussin, M., & Parfrey, L. W. (2018). Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? *mSystems*, 3(5), e00097-18. <https://doi.org/10.1128/mSystems.00097-18>

Mazel, F., Guisan, A., & Parfrey, L. W. (2024). Transmission mode and dispersal traits correlate with host specificity in mammalian gut microbes. *Molecular Ecology*, 33(1), e16862. <https://doi.org/10.1111/mec.16862>

Mazel, F., Prasad, A., & Engel, P. (2025). Host specificity of gut microbiota associated with social bees: Patterns and processes. *Microbiology and Molecular Biology Reviews*.

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealon, K., Pierce, N. E., ... Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, 110(9), 3229–3236. <https://doi.org/10.1073/pnas.1218525110>

McMahon, K. (2015). 'Metagenomics 2.0'. *Environmental Microbiology Reports*, 7(1), 38–39. <https://doi.org/10.1111/1758-2229.12253>

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, 7(1), 11257. <https://doi.org/10.1038/ncomms11257>

Moeller, A. H. (2025). Partner fidelity, not geography, drives co-diversification of gut microbiota with hominids. *Biology Letters*, 21(1), 20240454. <https://doi.org/10.1098/rsbl.2024.0454>

Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., Pusey, A. E., Peeters, M., Hahn, B. H., & Ochman, H. (2016). Cospeciation of gut microbiota with hominids. *Science*, 353(6297), 380–382. <https://doi.org/10.1126/science.aaf3951>

Moeller, A. H., Sanders, J. G., Sprockett, D. D., & Landers, A. (2023). Assessing co-diversification in host-associated microbiomes. *Journal of Evolutionary Biology*, n/a(n/a). <https://doi.org/10.1111/jeb.14221>

Moeller, A. H., Suzuki, T. A., Phifer-Rixey, M., & Nachman, M. W. (2018). Transmission modes of the mammalian gut microbiota. *Science*, 362(6413), 453–457. <https://doi.org/10.1126/science.aat7164>

Nayfach, S., & Pollard, K. S. (2016). Toward Accurate and Quantitative Comparative Metagenomics. *Cell*, 166(5), 1103–1116. <https://doi.org/10.1016/j.cell.2016.08.007>

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>

Niode, N. J., Adji, A., Rimbing, J., Tulung, M., & Tallei, T. E. (2021). Composition and diversity of bacteria from giant Asian honeybee *Apis dorsata* gut. *Biodiversitas Journal of Biological Diversity*, 22(2). <https://doi.org/10.13057/biodiv/d220246>

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, 27(5), 824–834. <https://doi.org/10.1101/gr.213959.116>

Ochman, H., Elwyn, S., & Moran, N. A. (1999). Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences*, 96(22), 12638–12643. <https://doi.org/10.1073/pnas.96.22.12638>

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), 2864–2868. <https://doi.org/10.1038/ismej.2017.126>

Olm, M. R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B. A., Morowitz, M. J., & Banfield, J. F. (2021). inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nature Biotechnology*, 39(6), 727–736. <https://doi.org/10.1038/s41587-020-00797-0>

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>

Powell, E., Ratnayake, N., & Moran, N. A. (2016). Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. *Molecular Ecology*, 25(18), 4461–4471. <https://doi.org/10.1111/mec.13787>

Pronk, L. J. U., & Medema, M. H. (2022). Whokaryote: Distinguishing eukaryotic and prokaryotic contigs in metagenomes based on gene structure. *Microbial Genomics*, 8(5), 000823. <https://doi.org/10.1099/mgen.0.000823>

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>

Quinn, A., El Chazli, Y., Escrig, S., Daraspe, J., Neuschwander, N., McNally, A., Genoud, C., Meibom, A., & Engel, P. (2024). Host-derived organic acids enable gut colonization of the honey bee symbiont *Snodgrassella alvi*. *Nature Microbiology*, 9(2), 477–489. <https://doi.org/10.1038/s41564-023-01572-y>

Raffiudin, R., Dyahastuti, M., Nugraha, R., Sayasti, T., Djuita, N. R., Suwananda, E., Allvioningrum, V., Mardhony, R., Biagioni, S., Setyaningsih, C. A., Prasetyo, L. B., Priawandiputra, W., Atmowidi, T., Saad, A., & Behling, H. (2024). The effect of land cover on the foraging behavior and pollen in the honey of the giant bee *Apis dorsata* in Sumatra. *Frontiers in Bee Science*, 2. <https://doi.org/10.3389/frbee.2024.1366287>

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLOS ONE*, 6(9), e22594. <https://doi.org/10.1371/journal.pone.0022594>

Raulo, A., Bürkner, P.-C., Finerty, G. E., Dale, J., Hanski, E., English, H. M., Lamberth, C., Firth, J. A., Coulson, T., & Knowles, S. C. L. (2024). Social and environmental transmission spread different sets of gut microbes in wild mice. *Nature Ecology & Evolution*, 8(5), 972–985. <https://doi.org/10.1038/s41559-024-02381-0>

Regan, T., Barnett, M. W., Laetsch, D. R., Bush, S. J., Wragg, D., Budge, G. E., Highet, F., Dainat, B., de Miranda, J. R., Watson, M., Blaxter, M., & Freeman, T. C. (2018). Characterisation of the British honey bee metagenome. *Nature Communications*, 9(1), 4995. <https://doi.org/10.1038/s41467-018-07426-0>

Rohde, K., & Rohde, P. P. (2008). How to measure ecological host specificity. *Vie Milieu*.

Rühlemann, M. C., Bang, C., Gogarten, J. F., Hermes, B. M., Groussin, M., Waschina, S., Poyet, M., Ulrich, M., Akoua-Koffi, C., Deschner, T., Muyembe-Tamfum, J. J., Robbins, M. M., Surbeck, M., Wittig, R. M., Zuberbühler, K., Baines, J. F., Leendertz, F. H., & Franke, A. (2024). Functional host-specific adaptation of the intestinal microbiome in hominids. *Nature Communications*, 15(1), Article 1. <https://doi.org/10.1038/s41467-023-44636-7>

Ruscheweyh, H.-J., Milanese, A., Paoli, L., Karcher, N., Clayssen, O., Keller, M. I., Wirbel, J., Bork, P., Mende, D. R., Zeller, G., & Sunagawa, S. (2022). Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome*, 10(1), 212. <https://doi.org/10.1186/s40168-022-01410-z>

Sanders, J. G., Sprockett, D. D., Li, Y., Mjungu, D., Lonsdorf, E. V., Ndjango, J.-B. N., Georgiev, A. V., Hart, J. A., Sanz, C. M., Morgan, D. B., Peeters, M., Hahn, B. H., & Moeller, A. H. (2023). Widespread extinctions of co-diversified primate gut bacterial symbionts from humans. *Nature Microbiology*, 8(6), Article 6. <https://doi.org/10.1038/s41564-023-01388-w>

Sarton-Lohéac, G., Nunes da Silva, C. G., Mazel, F., Baud, G., de Bakker, V., Das, S., El Chazli, Y., Ellegaard, K., Garcia-Garcera, M., Glover, N., Liberti, J., Nacif Marçal, L., Prasad, A., Somerville, V., SAGE class 2019-2020 and 2020-2021, Bonilla-Rosso, G., & Engel, P. (2023). Deep Divergence and Genomic Diversification of Gut Symbionts of Neotropical Stingless Bees. *mBio*, 0(0), e03538-22. <https://doi.org/10.1128/mbio.03538-22>

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Soden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, 48(16), 8883–8900. <https://doi.org/10.1093/nar/gkaa621>

Song, S. J., Sanders, J. G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., Mazel, F., Lutz, H. L., Winker, K., Graves, G. R., Humphrey, G., Gilbert, J. A., Hackett, S. J., White, K. P., Skeen, H. R., Kurtis, S. M., Withrow, J., Braile, T., Miller, M., ... Knight, R. (2020). Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. *mBio*, 11(1), 10.1128/mbio.02901-19. <https://doi.org/10.1128/mbio.02901-19>

Sprockett, D. D., Dillard, B. A., Landers, A. A., Sanders, J. G., & Moeller, A. H. (2025). Recent genetic drift in the co-diversified gut bacterial symbionts of laboratory mice. *Nature Communications*, 16(1), 2218. <https://doi.org/10.1038/s41467-025-57435-z>

Stewart, A. B., Sritongchuay, T., Teartisup, P., Kaewsomboon, S., & Bumrungsri, S. (2018). Habitat and landscape factors influence pollinators in a tropical megacity, Bangkok, Thailand. *PeerJ*, 6, e5335. <https://doi.org/10.7717/peerj.5335>

Su, Y.-C., Chiu, Y.-F., Warrit, N., Otis, G. W., & Smith, D. R. (2023). Phylogeography and species delimitation of the Asian cavity-nesting honeybees. *Insect Systematics and Diversity*, 7(4), 5. <https://doi.org/10.1093/isd/ixad015>

Suzuki, T. A., Fitzstevens, J. L., Schmidt, V. T., Enav, H., Huus, K. E., Mbong Ngwese, M., Grießhammer, A., Pfleiderer, A., Adegbite, B. R., Zinsou, J. F., Esen, M., Velavan, T. P., Adegnika, A. A., Song, L. H., Spector, T. D., Muehlbauer, A. L., Marchi, N., Kang, H., Maier, L., ... Ley, R. E. (2022). Codiversification of gut microbiota with humans. *Science*, 377(6612), 1328–1332. <https://doi.org/10.1126/science.abm7759>

The scikit-bio development team. (2020). *scikit-bio: A Bioinformatics Library for Data Scientists, Students, and Developers* [Computer software]. <http://scikit-bio.org>

Tung, J., Barreiro, L. B., Burns, M. B., Grenier, J.-C., Lynch, J., Grieneisen, L. E., Altmann, J., Alberts, S. C., Blekhman, R., & Archie, E. A. (2015). Social networks predict gut microbiome composition in wild baboons. *eLife*, 4, e05224. <https://doi.org/10.7554/eLife.05224>

Valles-Colomer, M., Blanco-Míguez, A., Manghi, P., Asnicar, F., Dubois, L., Golzato, D., Armanini, F., Cumbo, F., Huang, K. D., Manara, S., Masetti, G., Pinto, F., Piperni, E., Punčochář, M., Ricci, L., Zolfo, M., Farrant, O., Goncalves, A., Selma-Royo, M., ... Segata, N. (2023). The person-to-person transmission landscape of the gut and oral microbiomes. *Nature*, 614(7946), 125–135. <https://doi.org/10.1038/s41586-022-05620-1>

Vera-Ponce de León, A., Jahnes, B. C., Duan, J., Camuy-Vélez, L. A., & Sabree, Z. L. (2020). Cultivable, Host-Specific Bacteroidetes Symbionts Exhibit Diverse Polysaccharolytic Strategies. *Applied and Environmental Microbiology*, 86(8), e00091-20. <https://doi.org/10.1128/AEM.00091-20>

Wardman, J. F., Bains, R. K., Rahfeld, P., & Withers, S. G. (2022). Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nature Reviews Microbiology*, 20(9), Article 9. <https://doi.org/10.1038/s41579-022-00712-1>

Wilde, J., Slack, E., & Foster, K. R. (2024). Host control of the microbiome: Mechanisms, evolution, and disease. *Science*, 385(6706), eadi3338. <https://doi.org/10.1126/science.adia3338>

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1), 257. <https://doi.org/10.1186/s13059-019-1891-0>

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

Wu, Y., Zheng, Y., Wang, S., Chen, Y., Tao, J., Chen, Y., Chen, G., Zhao, H., Wang, K., Dong, K., Hu, F., Feng, Y., & Zheng, H. (2022). Genetic divergence and functional convergence of gut bacteria between the Eastern honey bee *Apis cerana* and the Western honey bee *Apis mellifera*. *Journal of Advanced Research*, 37, 19–31. <https://doi.org/10.1016/j.jare.2021.08.002>

Yang, Y., Zhang, N., Ji, S., Lan, X., Zhang, K., Shen, Y., Li, F., & Ni, J. (2014). *Dysgonomonas macrotermitis* sp. Nov., isolated from the hindgut of a fungus-growing termite. *International Journal of Systematic and Evolutionary Microbiology*, 64(Pt_9), 2956–2961. <https://doi.org/10.1099/ijss.0.061739-0>

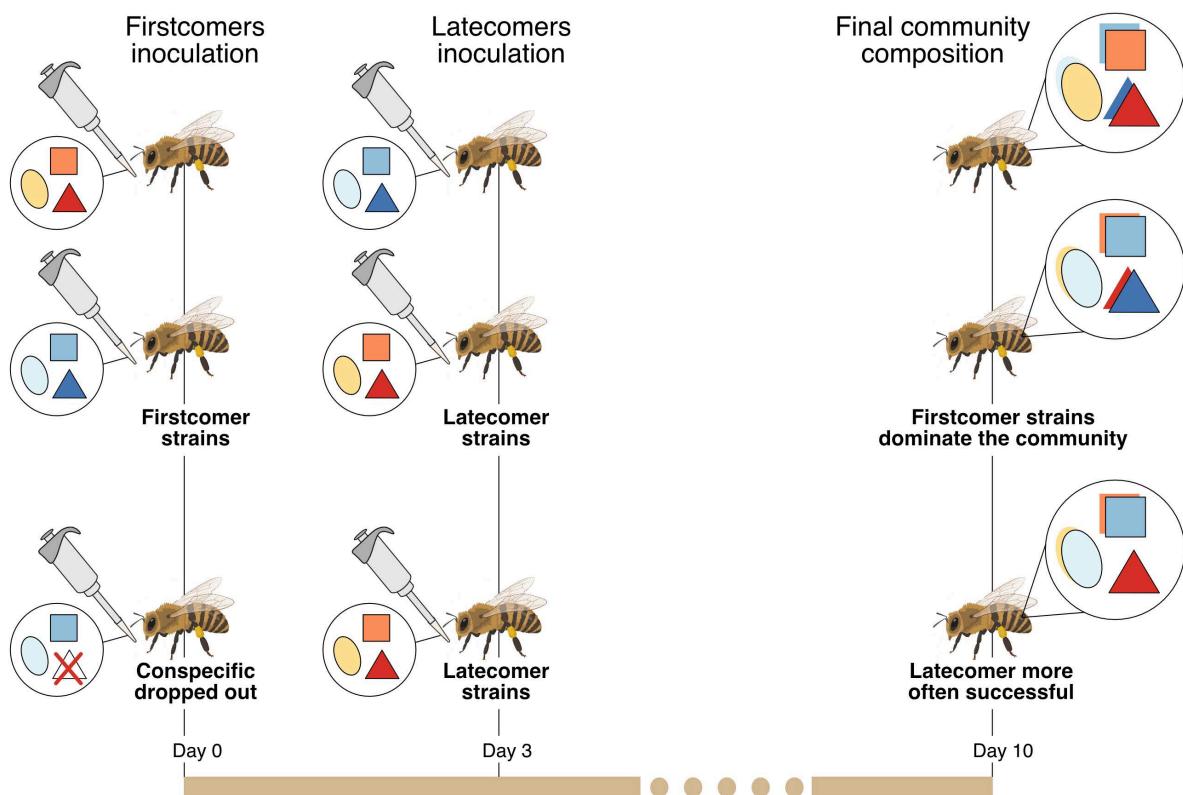
Ye, Y., & Doak, T. G. (2009). A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLOS Computational Biology*, 5(8), e1000465. <https://doi.org/10.1371/journal.pcbi.1000465>

Zhang, W., Zhang, X., Su, Q., Tang, M., Zheng, H., & Zhou, X. (2022). Genomic features underlying the evolutionary transitions of *Apibacter* to honey bee gut symbionts. *Insect Science*, 29(1), 259–275. <https://doi.org/10.1111/1744-7917.12912>

Zheng, H., Perreau, J., Powell, J. E., Han, B., Zhang, Z., Kwong, W. K., Tringe, S. G., & Moran, N. A. (2019). Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proceedings of the National Academy of Sciences*, 116(51), 25909–25916. <https://doi.org/10.1073/pnas.1916224116>

CHAPTER-3

PRIORITY EFFECTS AT THE STRAIN LEVEL IN HONEYBEE GUT MICROBIOTA



PRIORITY EFFECTS AT THE STRAIN LEVEL IN HONEYBEE GUT MICROBIOTA

Aiswarya Prasad¹, Gonçalo Santos Matos^{1,2}, Alexandra Szigeti-Genoud¹, Florent Mazel^{1,2}, and Philipp Engel^{1,2}

¹Department of Fundamental Microbiology, University of Lausanne, Switzerland, ²NCCR Microbiomes, Lausanne, Switzerland

SUMMARY

Gut microbial communities often differ at the strain level even among closely related individuals, but the ecological mechanisms driving this variation remain poorly understood. One potential driver is priority effects, differences in the timing and order of microbial colonization, which can lead to the assembly of distinct communities, even under similar environmental conditions. Priority effects may specifically play an important role in shaping microbial communities at the strain level, given that strains of the same species typically occupy similar ecological niches. To test this, we examined gut microbiota assembly in honeybees, in which age-matched nestmates are known to host similar microbial communities at the species level but vary in strain composition. We sequentially colonized microbiota-depleted honeybees with two distinct microbial communities, each composed of the same twelve species but different strains. We found that firstcomer strains consistently dominated the resulting communities, though the strength of these priority effects varied among closely related strains and species. Dropping out individual strains from the firstcomer community only partially improved the colonization success of latecomer conspecifics, suggesting that priority effects also act across species boundaries. Our results underscore the importance of priority effects for gut microbial community assembly at the strain level and in shaping the specialized gut microbiota of bees.

CONTRIBUTION

I planned and executed the pilot experiment with the help of one of the co-authors (GSM) and subsequently carried out the main experiment. I established a sample processing and library preparation method for the pilot experiment, which was carried out by another co-author (AS) for the main experiment. Then, I set up and validated the bioinformatic approach and carried out the data analysis and visualization. Finally, I led the drafting of the manuscript with inputs from my co-authors (PE, FM).

PUBLICATION STAGE

This chapter is a manuscript that will be submitted to a peer-reviewed journal upon receiving further comments from the co-authors.

INTRODUCTION

Gut microbial communities exhibit substantial bacterial diversity, particularly at the strain level (Ellegaard & Engel, 2019; Ley et al., 2008; Schloissnig et al., 2013). Closely related strains can co-exist stably within the same host over long periods of time. But, different hosts typically harbour distinct strain profiles, even when closely related or exposed to the same environment (Garud et al., 2019; Wolff et al., 2023). These differences in strain-level composition can have significant implications for the functional potential of the microbiota and its interactions with the host (Yan et al., 2020). Therefore, understanding which factors contribute to community assembly at the strain level is key.

An important driver of microbial community assembly is environmental filtering, where differences in physicochemical conditions such as nutrient availability, pH, transit time, oxygen levels, and immune system activity determine which microbes can survive, grow, and persist in a given environment. These deterministic factors have been relatively well studied in the gut microbiota and specifically act on microbes with functionally distinct traits. Other processes influencing community assembly have been studied less. These include dispersal limitation, or the ability of microbes to reach and colonize specific hosts (due to spatial barriers such as geography), and stochastic processes, random colonization events that occur independently of microbial traits or environmental conditions (Nemergut et al. 2013; Zhou and Ning 2017). One such stochastic phenomenon is priority effects (Debray et al., 2022), where the timing and order of microbial arrival shape the resulting community composition (Gleason, 1927; Nemergut et al., 2013). Microbes that arrive earlier can occupy (preempt) or alter the environment, affecting the colonization success of subsequent arrivals. While microbiome studies have investigated priority effects between bacterial species (Carlström et al., 2019; Martínez et al., 2018; Ojima et al., 2022), few studies have studied priority effects at the strain level and in the community context (Chen et al., 2024; Segura Munoz et al., 2022).

Contemporary coexistence theory (Adler et al., 2007; Chesson, 2000; Fukami et al., 2016; Mayfield & Levine, 2010; Orr et al., 2025) predicts that priority effects should be strong between competitors with a small fitness difference and overlapping niches. Hence, we hypothesize that priority effects should, in particular, be an important factor in shaping gut microbiome composition across individuals at the strain level.

Western honeybees (*Apis mellifera*) present an excellent model for studying processes underlying community assembly, particularly for testing the role of priority effects. They harbor a relatively simple and stable gut microbiota, composed of a few specialized bacterial genera that include multiple closely related species and strains (Ellegaard & Engel, 2019; Engel et al., 2012). While most genera and species are consistently present and often co-occur in individual bees, strain-level variation is more individualized, i.e., closely related strains typically segregate among individual bees, even among nestmates within the same colony (Ellegaard and Engel 2019). Among others, priority effects early in life, when newly emerged adult bees acquire the microbiota from their nestmates, have been proposed as one possible mechanism for the observed patterns.

In this study, we took advantage of the ability to generate and colonize microbiota-deprived (MD) bees under laboratory conditions with defined microbial communities, allowing us to experimentally test the role of priority effects in the assembly of gut communities at the strain level. We built distinct microbial communities from our strain collection, each composed of the same twelve species but divergent strains, and sequentially colonized microbiota-deprived honeybees with two communities three days apart. To quantify priority effects, we measured the ratio of the absolute abundance in the hindgut of each strain when it arrived with the firstcomer community versus the latecomer community. Further, we measured the colonization success of these strains in dropout community combinations where a randomly selected species was dropped out from the firstcomer community to observe its effect on the strains of the latecomer community. Our results show that priority effects are an important ecological process in shaping gut microbiome diversity at the strain level.

RESULTS

COMMUNITY SELECTION, EXPERIMENTAL DESIGN, AND VALIDATION APPROACH

To test whether priority effects operate in the honeybee gut bacterial community among closely related strains, we designed multi-species synthetic communities comprising the same species but different strains. We included twelve species of bacteria found prevalently in the rectum of western honeybees (*Apis mellifera*) (Callegari et al., 2021; Ellegaard & Engel, 2019), spanning the genera *Bifidobacterium*, *Lactobacillus*, and *Bombilactobacillus* (**Fig. 1A**, **Supplementary Table S1**). We focused on these three genera because they constitute the predominant native community members in the rectum of honeybee workers (full details on strain selection in **Methods**).

Microbiota-depleted bees were inoculated at Day 0 with community A (firstcomer strains) and then challenged with community B (latecomer strains) at Day 3 (**Fig. 1D**). The priority effect was measured for each strain in a given community as the ratio of its absolute abundance when inoculated with the community on Day 0 and on Day 3 (**Fig. 1B**). To replicate the measurement of priority effect for each strain under different backgrounds, we made six community combinations by shuffling strains between A and B such that in each combination, a given strain co-occurred with a different set of strains of the other 11 species. (**Supplementary Table S2**, **Fig. 1C**).

For each community combination, we included four treatments (A, B, AB and BA) (**Fig. 1D**). In treatments A and B, microbiota-deprived bees were only inoculated with one community at Day 0. In AB and BA, the two communities were inoculated one after the other on Day 0 and Day 3, respectively (**Fig. 1D**). Each strain appeared in either treatment A or B of each community combination. Consequently, if a given strain was in community A, it was a firstcomer in treatment AB and a latecomer in treatment BA. For each treatment, ~10 bees were analyzed. This way, each strain was measured ~60 times as a firstcomer and a latecomer across the six different community combinations. Subsequently, we conducted a drop-out experiment to assess the effect of dropping out one strain from the firstcomer community on the latecomer strains. Each dropout treatment included $A\Delta B$ and $B\Delta A$, where the two strains of one species were reciprocally dropped out in the community inoculated on Day 0. To assess the effect of the drop-out of the firstcomer strain on the colonization of the latercomer strain, the absolute abundances of the latecomer strain in the dropout treatment and the full community were compared.

We used the findings from a pilot experiment (see **Methods**) to inform our choice of experimental parameters and validate our assumptions.

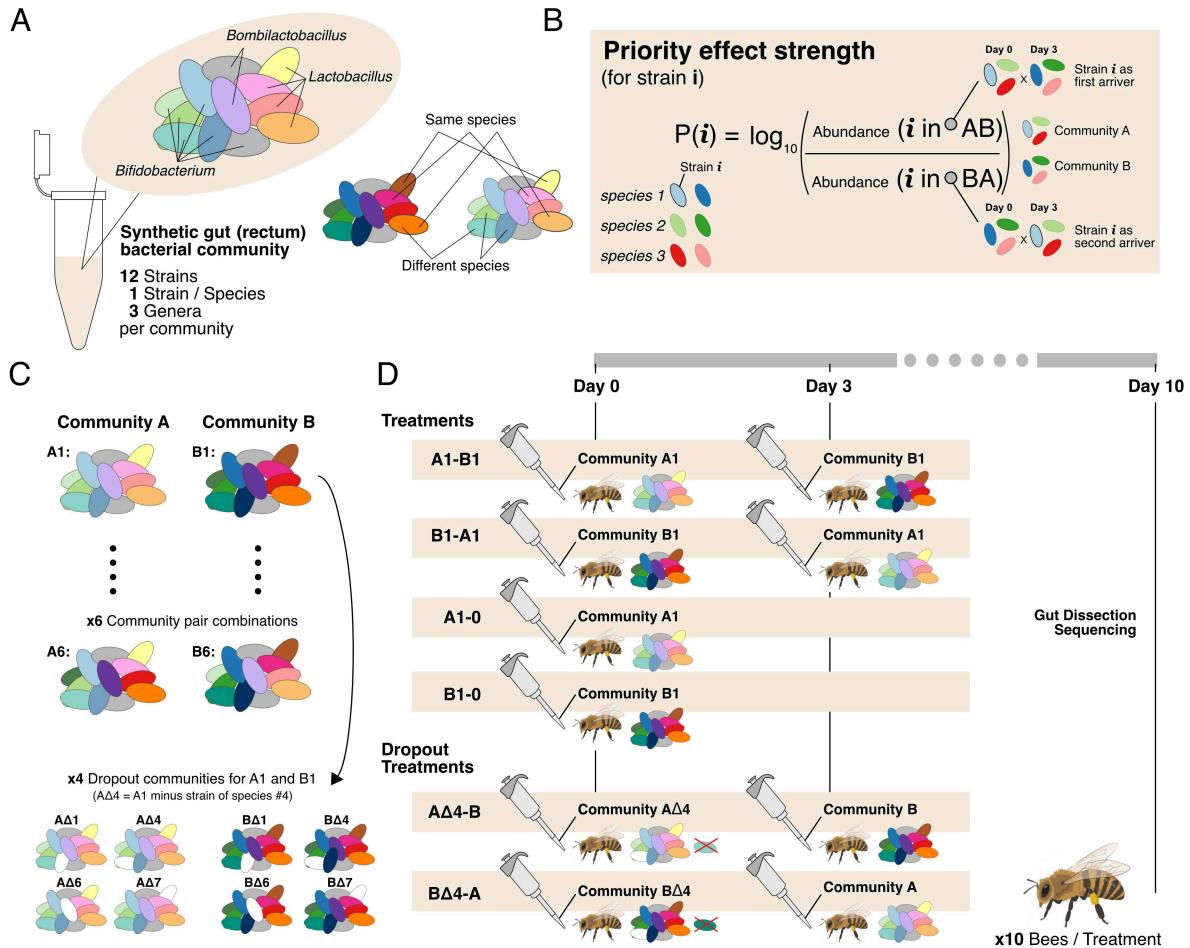


Figure 1. Experimental design and community selection. (A) Schematic of the multi-species bacterial communities used for the experiments. (B) Illustration of the approach used to measure the priority effect for a given strain i . (C) Schematic representing community pairs A and B made in 6 different combinations (1 – 6) by shuffling strains between them, and dropout treatments made by dropping out the strain of one species 1, 4, 6 or 7 from community A1 and B1. (D) Illustration of the experimental design to test each community combination (here, 1) and additional dropout treatments for the four selected species (here, 4).

PRIORITY EFFECTS DETERMINE COMMUNITY COMPOSITION AT THE STRAIN LEVEL

Almost all honeybees across treatments were successfully colonized by the inoculated community, as demonstrated by the total bacterial abundances (**Supplementary Fig. S1**) established for each sample using qPCR and full-length amplicon sequencing of the 16S rRNA gene. The abundance of bacteria in the gut of honeybees at the sampling day (i.e. Day 10) was orders of magnitude higher than in the inocula (fed on Day 0 and Day 3) and in non-inoculated honeybees only fed sterile sugar water (MD bees), indicating that the strains in the inoculum proliferated within the honeybee gut resulting in the assembled communities (**Supplementary Fig. S2**).

All twelve species were detected in most honeybee samples and were represented by one or several strains, depending on the treatment (**Supplementary Fig. S3**). We only detected strains present in the inocula used for each treatment, confirming both the success of the experimental setup and the gnotobiotic status of our bees (**Supplementary Fig. S4**). Total bacterial loads in each sample were not significantly correlated with the number of strains

present (Spearman correlation $\rho = 0.02$, $p = 0.219$) (**Supplementary Fig. S5**), and the summed abundance for each species did not differ significantly whether one or more strains of the species were inoculated in the treatment (**Supplementary Fig. S6**). This indicates that the strains in each community reach carrying capacity and that latecomer strains compete in the niches occupied by the firstcomers rather than occupying additional empty niches.

To determine whether priority effects act at the community level, we used the absolute abundance of strains to estimate the Bray-Curtis distance between the samples of the four treatments in each community combination (~10 bees per treatment). PERMANOVA tests confirmed a significant ($p < 0.05$) effect of treatment across all community combinations. Communities AB closely resembled communities A, while communities BA resembled communities B across all replicates. These results indicate that the firstcomer strains dominated the communities and that the order of arrival of strains determines the overall community composition (**Fig. 2**). These differences were also visible across all replicates when using relative abundances or the absence/presence of strains to compare the communities (**Supplementary Fig. S7**).

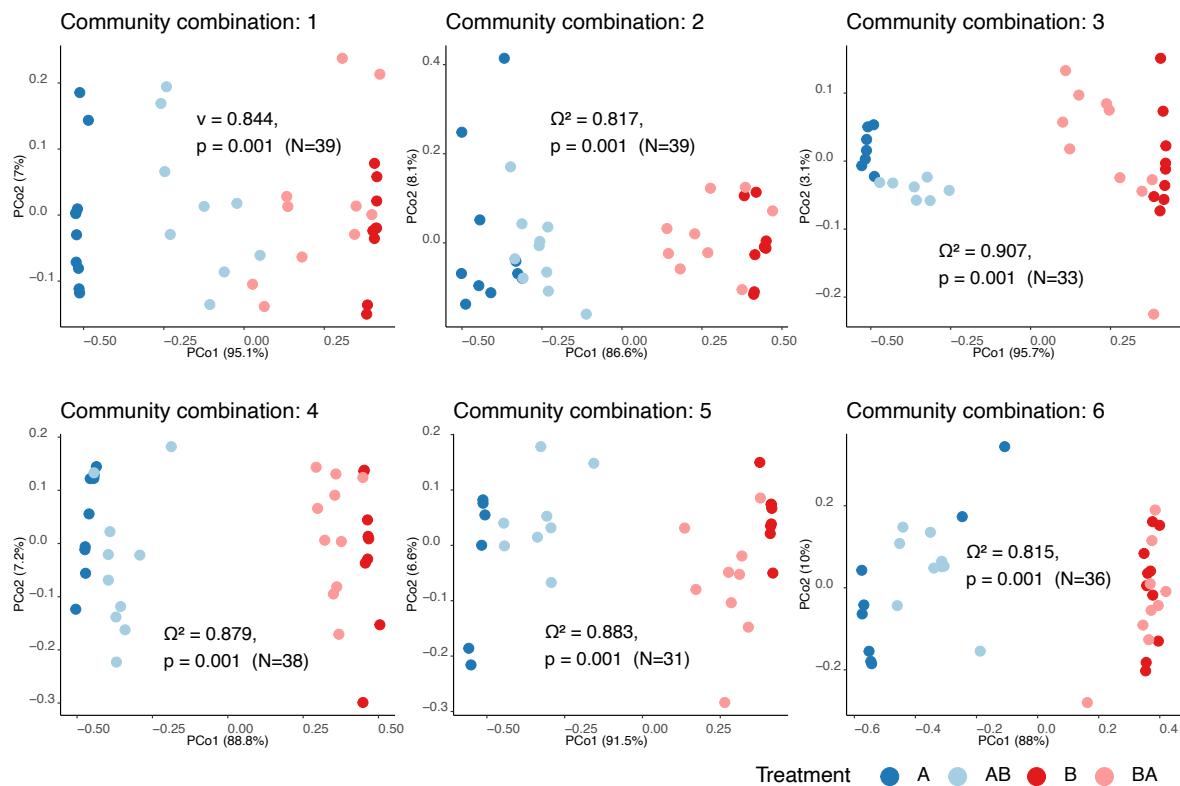


Figure 2. Community composition across samples. PCoA plots to visualize Bray-Curtis distance estimated from the matrix of log-transformed absolute abundance of bacterial strains across community combinations, with color indicating the treatment (A/B – only community A/B on Day 0 respectively, AB – community A on Day 0 and community B on Day3 and vice versa for BA).

PRIORITY EFFECTS ARE CONSISTENTLY OBSERVED FOR ALL STRAINS, BUT VARY IN STRENGTH

Strains of the same species may vary in the extent to which their niches overlap or in their ability to interfere with already established strains - both factors that can influence the

strength of priority effects. Therefore, we next aimed to assess whether there are differences in priority effects among individual strains within the tested communities. All strains grew to significantly lower abundances (Wilcox rank sum test, $p < 0.05$) when they were inoculated on Day 3 as latecomers (A in BA, or B in AB) than when inoculated on Day 0, either as the only inoculation (treatments A and B) or as firstcomers (i.e. A in AB, or B in BA) (**Fig. 3**) suggesting that priority effects act on all tested strains. Notably, while several strains went nearly undetected when inoculated as latecomers (**Fig. 3, Supplementary Fig. S8**, e.g., *Bifidobacterium asteroides* ESL017 and ESL0822, or *Bifidobacterium* sp2. ESL0200 and ESL0819), others were barely affected (e.g. *Lactobacillus apis* strain ESL0263). To quantify these differences, we calculated the priority effect strength per community combination for each strain as the log ratio of the median abundance of each strain when it was a firstcomer and when it was a latecomer (**Supplementary Table S4**). As expected, priority effect strength was higher than zero for all strains (**Fig. 4**), although we observed a considerable amount of variation across replicates (i.e. different community combinations). A generalized linear mixed model confirmed that arrival (Only – as a reference – compared to first and late comers) as a fixed effect and community combination as a random effect (not pronounced) explains the absolute abundance of each strain ($p < 0.05$ for second and $p > 0.05$ for first) except for the *Lactobacillus apis* ESL0263 (**Supplementary Table S5**).

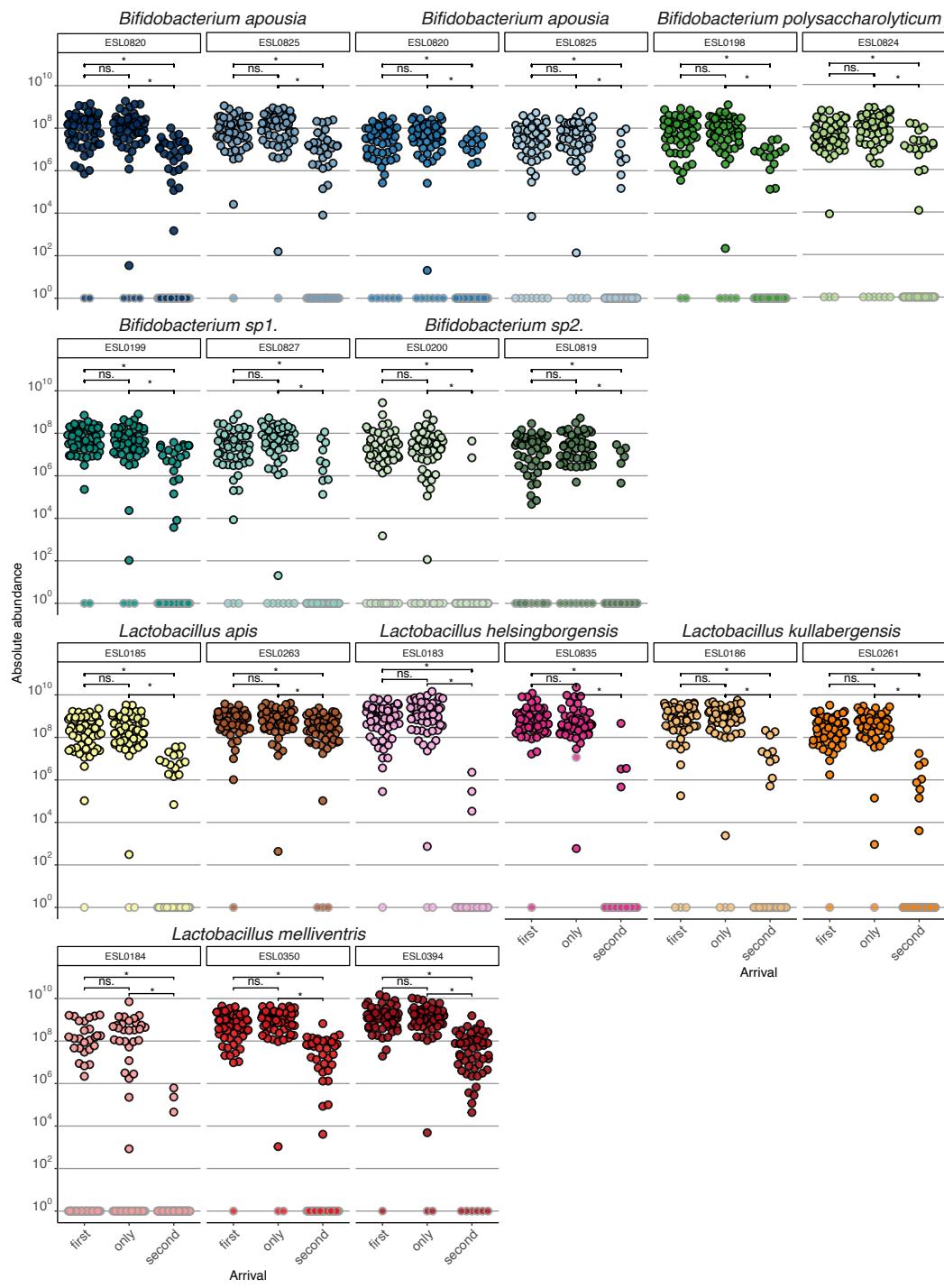


Figure 3. Abundance of each strain by arrival order. Absolute abundance of each strain (shown by facet) by arrival order (Only- inoculated with Day 0 Community, no further inoculations, First- inoculated with Day 0 Community and then on Day 3 by another community including its conspecific strain, Second – inoculated on Day 3 when there was already another community including its conspecific strain inoculated on Day 0). Each plot includes all the treatments across the six community combinations. Result of Wilcoxon rank sum test (two-sided) are annotated (* - $p < 0.05$), complete results of statistical test, including sample size per, are included separately (Supplementary Table S3). Points with grey outlines represent samples where the strain was not detected (below the detection threshold for that sample) and hence set to a value of 1.

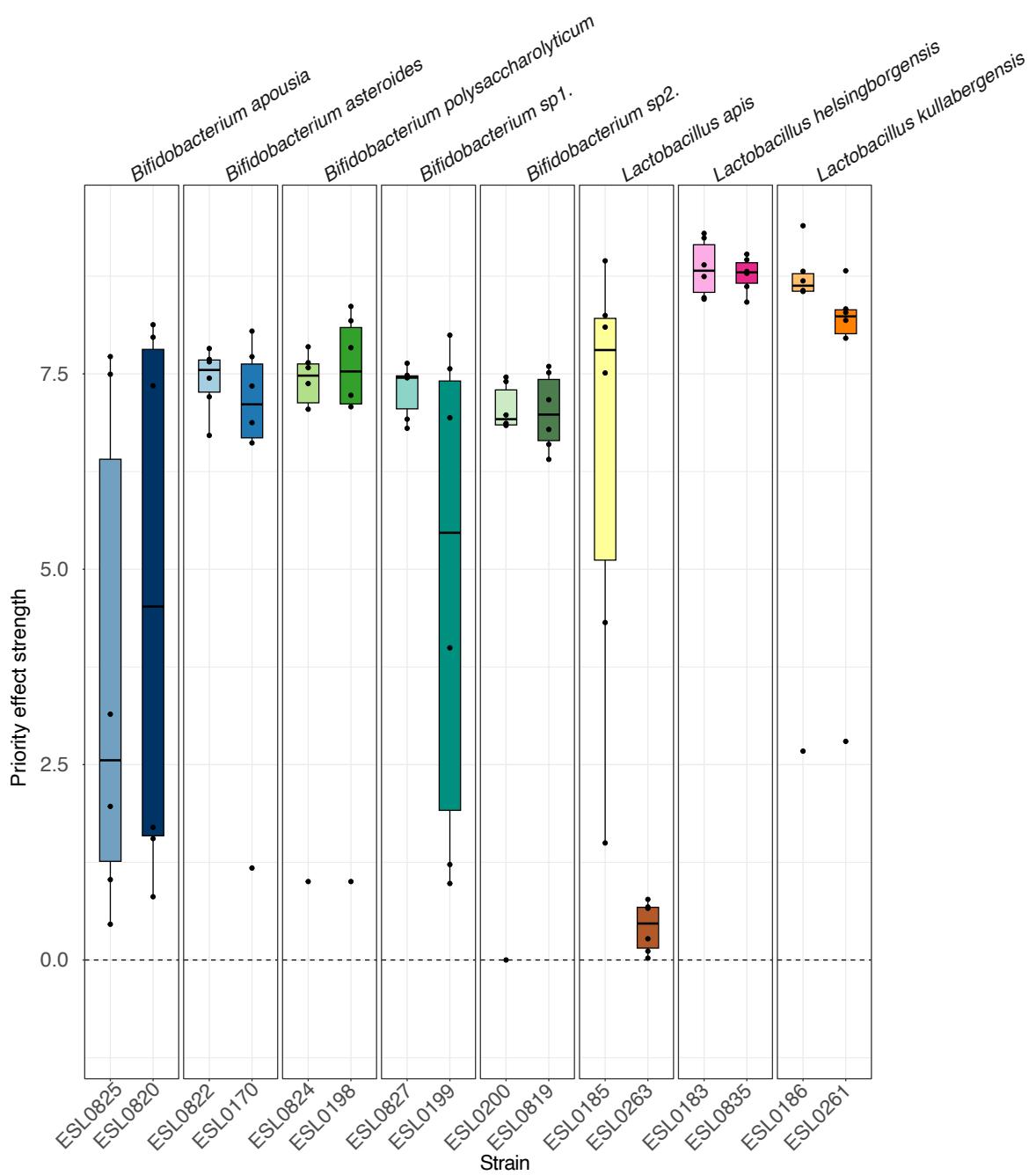


Figure 4. Strength of priority effect by strain across community combinations. The strength of priority effects for each strain in all six community combinations. The priority effect strength was calculated as the log ratio of the median absolute abundance of the strain in ~10 bees where they were in the community inoculated first vs. ~10 bees where they were inoculated second (AB vs. BA where the strain was in the community A) for each community combination of A* and B* where * goes from 1 to 6.

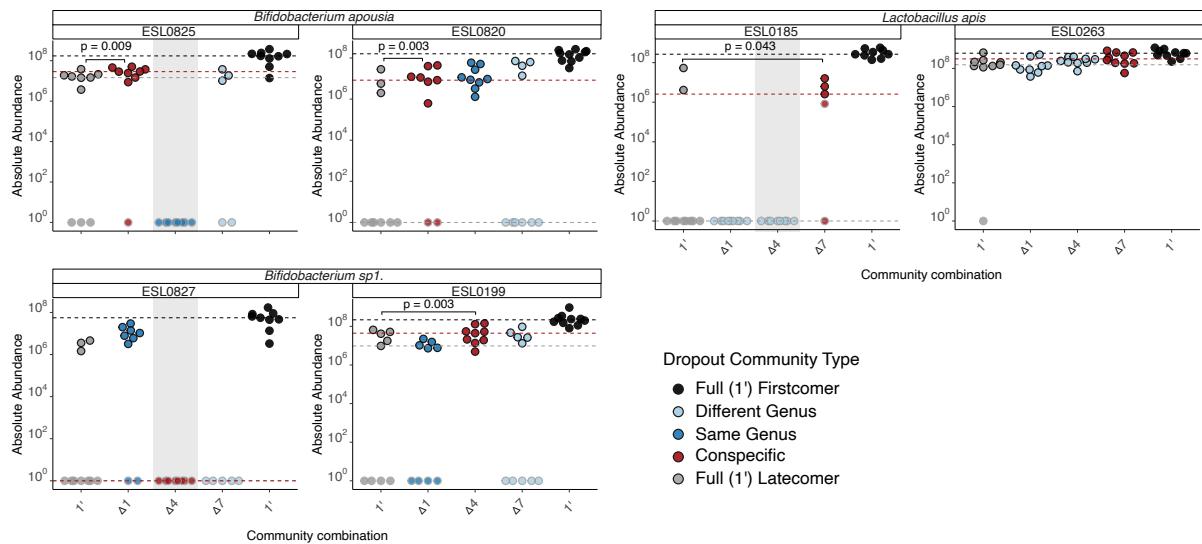
CONSPECIFICS CONTRIBUTE TO PRIORITY EFFECTS IN ADDITION TO OTHER CLOSE RELATIVES

We hypothesized that the priority effects are mediated by within-species competition rather than between-species competition, under the assumption that the niche overlap is higher between conspecific than allo-specific strains. Therefore, we dropped out the strain of one species in the firstcomer community at Day 0 and challenged with the full latecomer

community at Day 3 (**Fig. 1D**). We tested this for three random species (1: *Bifidobacterium apousia*, 4: *Bifidobacterium sp1.*, and 7: *Lactobacillus apis*) and both strains of these species reciprocally (i.e. dropping out the firstcomer strain in AB and then in BA). To assess the effect of the presence of a conspecific strain in the firstcomer community, we compared the abundance of the respective latecomer strains in dropout treatments (1,4 and 7) to their abundance in the full community (1'). Our expectation was that the dropout would increase the abundance of the latecomer strain of that species.

Among the conspecifics of the dropped-out strains in each treatment, all but one strain, *Lactobacillus apis* (ESL0263), showed some improvement in colonization success compared to the full community treatment (**Fig. 5**). The effect of dropping out *Bifidobacterium sp1.* Strain ESL0199 could not be determined (no data for latecomers in Δ4 in shaded panels) due to low sequencing depth (**Supplementary Fig. S1**). Some latecomer strains belonging to other species than one dropped out were also affected. In particular, several *Bifidobacterium* species in treatments where the dropouts were *Bifidobacterium* strains (**Fig. 5, Supplementary Table S6**), were higher in frequency or quantity in those two treatments than in the full community or other dropout treatments (**Fig. 5, Supplementary Table S6**). Based on these results, we conclude that within-species and between-species competition, specifically among closely related species, contribute to the priority effects for each strain in the community.

Conspecifics of dropped out strains:



Other strains:

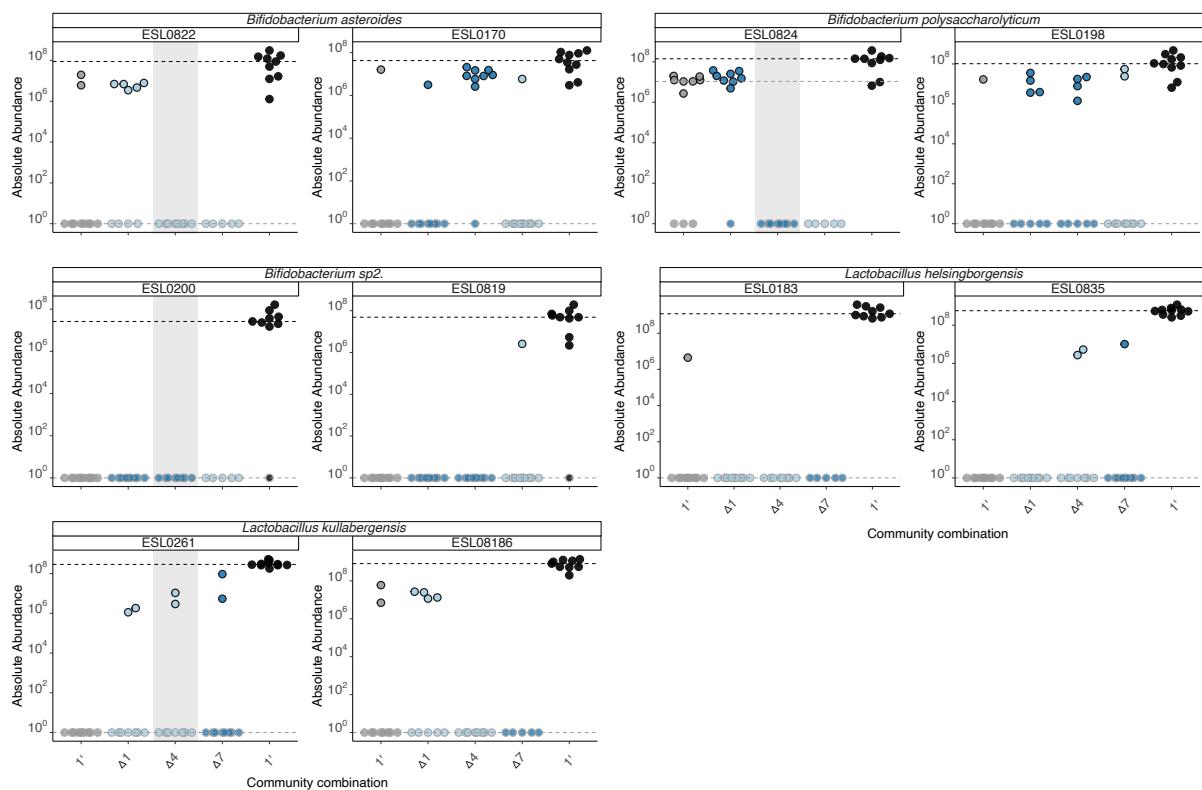


Figure 5. Colonization success of strains as latecomers in full and dropout treatments. Absolute abundance of each latecomer strain in full and dropout treatments, with community combination shown on the x-axis and colors indicating the relationship between each strain and the drop-out strain in the respective community combination. Result of permutation test ($n = 1000$) between the abundance of latecomers in full and dropout treatments with conspecifics are noted where significant ($p < 0.05$). Shaded panels indicate missing data due to insufficient sequencing depth. Dashed horizontal lines mark the median value of the group indicated by color. Points with grey outlines represent samples where the strain was not detected (below the detection threshold for that sample) and hence set to a value of 1.

DISCUSSION

The influence of order of arrival (*i.e.*, priority effects) on the outcome of community assembly in microbial communities has been demonstrated before (Boyle et al., 2021; Carlström et al., 2019; Garrido-Sanz & Keel, 2024; Peay et al., 2011), including gut microbiomes (Chen et al., 2024; Gurung et al., 2024; Laursen & Roager, 2023; Ojima et al., 2022; Segura Munoz et al., 2022). However, few studies have focused on tracking individual strain-level patterns *in vivo* and in the community context. Leveraging full-length 16S rRNA amplicon sequencing to discriminate between conspecific strains (of the same species), we demonstrate the importance of priority effects in a multi-species community in the outcome of strain-level interactions *in vivo* using the honeybee gut microbiome model.

We assessed the importance of priority effects by using a synthetic microbial community composed of several related species rather than individual strains. This approach ensured that the observed effects were relevant in the presence of inter-species interactions, such as competition for space and nutrients, mimicking the natural environment in the bee gut. Further, our approach enabled us to simultaneously evaluate the impact of arrival order on several strains (2x10 strains for each community pair) inoculated reciprocally as firstcomers and latecomers (three days apart). To account for unexpected effects of interactions between strains in the same community, we tested several community combinations, *i.e.* the priority effect of each strain was assessed in the background of different strains.

Our findings demonstrate that priority effects play a significant role in shaping the strain-level composition of the bee gut microbiota. When bees were sequentially inoculated with the two communities, their gut microbiota consistently resembled that of bees exposed only to the first community. This indicates that strains from the firstcomer community effectively dominated, hindering the establishment of later-arriving strains. Such priority effects may help explain - or at least contribute to - the natural variation in strain-level composition observed among individual worker bees within a colony (Ellegard et al. NatComm). Similar early-life effects have been shown to drive microbiota variation in laboratory mouse models, even under tightly controlled host and environmental conditions (Martínez et al., 2018). Like in other mammals, the gut microbiota of adult bees is acquired after pupal eclosion through social interactions with nestmates. Therefore, the specific strains a newly emerged bee acquires likely depend on which nestmates it encounters during its early post-emergence period. These initial interactions can thus result in persistent individual differences in microbiota composition and could even contribute to the host-specific nature of the bee gut microbiota (Mazel et al., 2025).

Priority effects may be particularly strong in honeybees, as newly emerged adults spend the first few days inside the hive and only sporadically embark on defecation flights. As a result, food residence time in the gut may be relatively long, potentially leading to low strain-level

turnover, and ecological opportunities for exposure to the microbiota of other honeybee species are low. However, as adult worker bees age, they transition from nurses to foragers and undergo significant physiological and behavioural changes. Notably, they usually empty their rectum of pollen and shift to a nectar/honey-based diet. Recent findings show that foragers harbour a distinct strain-level community compared to nurses (Baud et al., 2023). Therefore, even though early arriving strains might dominate the community in nurse bees, other processes and mechanisms than priority effects are likely at play, such as selection in the host gut environment or dietary differences, and may affect the community composition, e.g., when bees transition to foragers.

Strains of the same species in the bee gut can differ substantially in their accessory gene content, leading to variation in functional potential (Van Rossum et al., 2020). As a result, they may influence the host in distinct ways—for example, through differences in carbohydrate degradation or the production of metabolites with neuroactive properties. Consequently, the order in which strains colonize the gut early in life, and the resulting differences in community composition at the strain-level, can have lasting effects on how the microbiota interact with and impact the host. Such host effects of arrival order have been shown to occur in legume-rhizobium mutualism (Boyle et al., 2021) and may likely also play an important role in more complex microbial communities, such as those in the animal gut. Our results also suggest that probiotic strains, which are already used in bee management, need to be administered early in the bee's life to ensure effective and persistent gut colonization.

While we found that arrival order influenced colonization outcomes for all tested community members, the strength of priority effects varied notably across strains. For instance, all strains of *Bifidobacterium asteroides*, *Bifidobacterium* sp. 2, *Lactobacillus helsingborgensis*, and *Lactobacillus kullabergensis* were almost entirely undetectable when introduced as part of the latecomer community. In contrast, strains from other species were still able to establish as latecomers, albeit at much lower abundances than when introduced first. The strain ES0263 of *Lactobacillus apis* stood out for its exceptional ability to colonize successfully even as a latecomer. The underlying reasons for these differences among strains remain to be elucidated, but may involve variation in niche overlap or the ability to interfere with established strains. Further investigation into the mechanisms behind these differences could reveal valuable traits for probiotic development or microbial engineering. These approaches are also of growing interest for promoting the health of managed honey bee populations (Motta et al., 2022).

A previous study demonstrated that phylogenetic relatedness predicts the extent of priority effects among microbes (Peay et al., 2011). Since we tested priority effects of strains in multi-species communities, it was not possible to disentangle the effect of inter- and intra-species interactions on a given strain. To mitigate this, we used dropout communities where one firstcomer strain was dropped at a time. Our expectation was that the latecomer strain of the same species would increase in colonization if the priority effect was mediated through intra-

specific competition. While this was the case for at least four of the six tested strains, we noted that the effect was relatively small. In other words, the latecomer strain was far from reaching the same abundance as when present in the firstcomer community. Moreover, several strains of other species in the latecomer communities increased in abundance when a strain dropped out. These results suggest that priority effects are also mediated through interspecific interactions, in particular between species belonging to the same genus as the one dropped out. We hypothesize that not only strains of the same species but also those of closely related species have considerable niche overlap, and hence contribute to priority effects via niche pre-emption. Measuring the metabolic capabilities of these strains *in silico* and *in vitro* could be used to determine the niche overlap and subsequently predict community assembly and invasion. This was recently done in a study on testing the capability of *Salmonella* to invade defined synthetic communities *in vitro* and *in vivo* in the mouse gut (Spragge et al., 2023).

Our findings underscore the importance of arrival order in shaping gut microbial communities, particularly at the strain level and for the assembly of the gut microbiota of honeybees. While natural systems may buffer or modulate these effects through increased diversity and environmental complexity, the foundational role of early colonizers in determining community trajectories remains evident. These insights contribute to a broader understanding of how host-associated microbial communities are structured and maintained, with implications for both microbiome engineering in the context of bee health and ecological theory.

METHODS

STRAIN SELECTION AND SEQUENCE DATABASE ESTABLISHMENT

The bacterial strains used in this experiment were isolated from the honeybee gut and stored as 20% Glycerol in PBS at -80°C. More details about each strain can be found in **Supplementary Table S1**. Illumina genomes proved to be suboptimal in obtaining the 16S rRNA gene sequences for each strain. We used Barrnap (Seemann, 2013/2025) to extract the gene sequence and found that several Illumina genomes only provided partial sequences of some copies. Hence, we re-sequenced most of these strains using long-read sequencing, either PacBio or Nanopore sequencing, and inferred 16S rRNA gene sequences and copy number from these genomes (**Supplementary Table S1, Supplementary Table S7**). We found that the V4 region (commonly used in previous studies) was identical for almost all our strains of the same species. However, outside this region, there were two or more distinct SNPs (sometimes >300 base pairs apart) in each 16S copy that could differentiate most of them. Hence, we used full-length 16S rRNA sequencing using PacBio long reads to uniquely identify each strain. Non-redundant sequences obtained from the genomes using Barrnap were assigned unique IDs (uid) and collected into a custom database (**Supplementary Table S7**).

For the experiment, we selected the closest relatives within each species based on ANI (**Supplementary Fig. S10**), which were also distinguished by at least one unique 16S rRNA gene sequence from each other (**Supplementary Fig. S11**). In the pilot, ESL0262 and ESL0183 were included but not in the experiment, as we subsequently determined that they are indistinguishable, as they contain the same four copies of the 16S rRNA gene. Finally, For the *Bombilactobacillus* species and for *Bifidobacterium coryneforme*, only one strain was included in all the experiments, and hence, the same strain was present in all the respective treatments across community combinations. This was done to avoid leaving niches available to others that these prevalent species might otherwise occupy. For *Lactobacillus melliventris*, one community of each pair included two strains, as one of its strains was mis-annotated as a *Bombilactobacillus mellis*, and in those communities, there was hence no representative of this species. Specifically, ESL0295 and ESL0394 were intended to be two representatives of *Bombilactobacillus mellis*, but re-sequencing and comparison of the genomes after the experiments were complete revealed that ESL0394 is a *Lactobacillus melliventris* genome. This mis-annotation in our internal database has been corrected, and all figures reflect the correct annotation. Hence, we exclude ESL0295 from the priority effect calculations since it was only present in one of each community pair. However, ESL0394 and ESL0350 were in the same community (except in combination 4), with ESL0184 being their conspecific counterpart. The dropout treatments AΔ6B and BΔ6A which were intended to be *Bombilactobacillus mellis* dropouts, were also excluded from comparisons.

VALIDATION OF THE CUSTOM DATABASE AND STRAIN DETECTION APPROACH

To validate our database (**Supplementary Table S7**) and strain detection approach, we amplified and sequenced each isolated strain and matched the most abundant amplicon sequence variants (ASVs) to our database. Other than a few strains with low sequencing depth (<100 reads), we detected the expected number of sequences in most (**Supplementary Fig. S12**). There were several false positives (since samples were extremely low in diversity, one strain, and sequenced very deeply), but the false positives were orders of magnitude less abundant than the expected sequences. We further compared the sequences that did not match any strains to all the sequences inferred across all experiments (four different sequencing runs) to confirm that they were false-positives. Hence, the ASVs from the isolate genomes resulted in exact matches to sequences in our database and those that did not but matched ASVs from the experiments were investigated further.

We included two sequences (uid 61 and 62) *a posteriori*. The strain ESL0820 went undetected because neither of the two sequences identified from its genome matched any ASVs from the experiment. However, one ASV occurred exactly in the samples where ESL0820 was added and, upon further inspection, matched both uid 2 and 3 closely. It was not identical to either because it had both the unique SNPs of those two sequences. Since this ASV was consistently detected across several sequencing runs, we conclude that it is not an artifact (chimeric) but rather that there was an error in Nanopore genome assembly, resulting in two copies assembled, one with each SNP, instead of two identical copies, each containing both the SNPs. Hence, we added this sequence to the database as uid 61 expected to be present in two copies in ESL0820 instead of one copy each of uid 2 and 3. Similarly, among the two copies expected to be found in ESL0827, uid 10 was detected consistently, but never uid 11. Instead, another ASV sequence matching uid 11 closely, but for one SNP, was consistently detected. We added that ASV sequence to our database as uid 62, expected in one copy in ESL0827 instead of uid 11. Hence, we established and validated our database of distinct full-length 16S rRNA gene sequences from each of the strains used in this study.

SYNTHETIC COMMUNITY ASSEMBLY

Selected strains were revived from stocks stored in 20% Glycerol in PBS at -80°C. All strains were grown on solid De Man – Rogosa – Sharpe agar (MRSA) (supplemented with 2% w/v fructose and 0.2% w/v L-cysteine-HCl) in petri plates at 34°C under anaerobic conditions. For assembling into communities, each strain was re-grown in two plates from single colonies. After ~48h each strain was harvested using a sterile loop by scraping the entire contents of its plates and resuspended in 300µL in a sterile solution of 20% Glycerol in PBS. The solution was diluted in 20% Glycerol in PBS, to an OD₆₀₀ value of 2. Each strain was mixed in equal volumes (10µL) into a sterile tube. This resulted in about 120µL for each community combination (for dropout communities, sterile 20% Glycerol in PBS solution was added instead of the dropped strain), made up to 200µL using 20% Glycerol in PBS solution. In this resulting mixture, each strain would be present in a quantity equivalent to an OD₆₀₀ value of 0.1. The final mixture for each community was thoroughly mixed and separated into 50µL aliquots, which were flash frozen using liquid Nitrogen and stored at -80°C until used.

EXPERIMENT SETUP AND PROCEDURE

Microbiome-deprived (MD) honeybees were obtained from *Apis mellifera carnica* colonies maintained in the University of Lausanne as described before (Kešnerová et al., 2017). The day on which the pupae emerged as adults was considered Day 0. On this day, honeybees were fed 5µL of inoculum from glycerol stocks of community aliquots diluted with 450µL of a 1:1 sterile sugar water (50% sucrose solution, w/v) and 1x PBS mixture at room temperature. The bees were then transferred to their respective sterilized cages (one per treatment), containing sterile pollen and sugar water tubes, which were replaced upon exhaustion through the course of the experiment. In each treatment, 10 – 12 bees were included per cage. For treatments involving a second feeding on Day 3, feeding was repeated as done on Day 0. The experiment was concluded on Day 10, which allows several days for the community to stabilize (achieved ~ Day 5) based on prior studies (Brochet et al., 2021; Kešnerová et al., 2017). Finally, on Day 10, all surviving honeybees (8 – 10) were sacrificed and their guts were dissected, flash frozen using liquid Nitrogen, and stored at -80°C until DNA extraction.

VALIDATION OF EXPERIMENTAL PARAMETERS (PILOT EXPERIMENT)

We conducted a pilot experiment with a community combination similar to A1 and B1 (**Supplementary Table S2**) to inform our choice of experimental parameters and validate our assumptions. We included twelve treatments with one cage per treatment containing 10 honeybees, of which five were sequenced. To confirm that all our strains could establish themselves in the honeybee when inoculated simultaneously, we included Treatment AB-0, where both communities A and B were inoculated on Day 0 (**Supplementary Fig. S13**). All strains were able to establish themselves and grew to similar abundances. Next, to verify that a given community was able to establish itself in the honeybee gut just as well on Day 3 as on Day 0, we compared B-0 and B-3 (there was a technical issue with the inoculum A-0). This validated our assumption that reduced colonization success of the Day 3 community strains was due to the community fed on Day 0 and not from being inoculated in an older honeybee (**Supplementary Fig. S13**). Treatments AB-1, AB-3, BA-1 and BA-3 demonstrated that the effect of challenging the Day 0 community on Day 1 and Day 3 was visible, with Day 3 showing a clearer difference for some strains (**Supplementary Fig. S13**). Hence, we chose to challenge on Day 3 in subsequent experiments. Finally, we also confirmed that the microbiota-deprived (MD) bees that were only inoculated with sterile sugar water and fed sugar water and pollen in their cage were largely free of gut microbes (**Supplementary Fig. S13**) given our experimental procedure and setup.

DNA EXTRACTION

DNA was extracted from dissected bee guts using a magnetic bead-based protocol optimized for high-throughput processing on the Opentrons OT-2 liquid-handling robot. Samples were homogenized in bead-beating tubes containing a mix of 1 mm glass beads and 0.1 mm zirconia beads, then lysed in 1x G2 buffer supplemented with lysozyme (100 mg/mL). Lysis was performed at 37 °C for 30 min with shaking (900 rpm), followed by the

addition of Proteinase K (20 mg/mL) and further incubation at 56 °C for 1 h. For aliquotes of communities fed to bees, 165 µL was used for beadbeating. Lysates were centrifuged, and 80 µL was transferred to a 96-well PCR plate. Extraction blanks were included at this stage as wells only containing Nuclease-free water instead of gut homogenates. DNA purification was performed using CleanNGS magnetic beads (Clean NA #CNGS-0050) on an Opentrons® OT-2 robot. Automated steps included magnetic bead binding, two 80% ethanol washes, drying, and elution in nuclease-free water. Eluates (30 µL) were collected in a new plate and stored at –20 °C or –80 °C. DNA concentrations were quantified using a Qubit™ fluorometer and the 1X dsDNA HS assay. 0.1x dilutions were made and stored in an additional plate and used for further steps.

For the pilot experiment, DNA extraction was carried out using the CTAB and Phenol-Chloroform-Isoamyl alcohol-based extraction method, as described before (Prasad et al., 2025). High molecular weight DNA used for long-read genome sequencing of isolates was obtained using the Maxwell® RSC instrument and Maxwell® RSC PureFood GMO and Authentication kit, using an additional bead-beating step in Qiagen Power Pro tubes for 10 minutes on a bench-top Vortex mixer.

LIBRARY PREPARATION AND AMPLICON SEQUENCING

For amplicon sequencing, multiplexed amplicon libraries were prepared using the Amplification of Full-Length 16S Gene with Barcoded Primers for Multiplexed SMRTbell® Library Preparation and Sequencing protocol (Version 05). PCR amplification was performed using the KAPA HiFi HotStart ReadyMix (KAPA Biosystems) and a pre-prepared barcoded primer plate containing the forward and reverse primers

5' GCATC/barcode/AGRGT_TYGATYMTGGCTCAG 3' and

5' GCATC/barcode/RGYTACCTTGTTACGACTT 3'

in each well combined to result 96 unique pairs as recommended in the protocol. Amplicons were diluted to 0.01x and quantified using qPCR as described in the following section. Using the results of the quantification, amplicons were pooled in equimolar quantities for library preparation using the SMRTbell Prep Kit 3.0. Library preparation and sequencing were carried out in the Next Generation Sequencing Platform of the University of Bern. Samples from two plates at a time were further multiplexed into the same sequencing run carried out in a SMRT cell 25M on a PacBio Revio system. Sequenced reads in fastq format were obtained using the SMRT Link (Version 25.1.0.257715) software.

QUANTIFICATION OF TOTAL BACTERIAL ABUNDANCE

qPCR was performed on DNA extracts diluted to 0.1x to minimize the effect of contaminants. Primers targeting a 162 bp stretch of the V4 region of the 16S rRNA gene were used to quantify the entire bacterial community as described before (Kešnerová et al., 2017). Briefly, the qPCR reactions were conducted in triplicate with a 10 µL total volume containing 0.2 µM of each forward and reverse primer, 1x SYBR® Select Master Mix (Applied Biosystems), and 1 µL of the sample. Thermal cycling conditions included a denaturation stage at 50°C for 2

minutes, followed by 95°C for 2 minutes, 40 cycles of 95°C for 15 seconds, and 60°C for 1 minute. Standard curves were prepared with serial dilutions of plasmid DNA as described previously (Kešnerová et al., 2017) to determine primer efficiency. Copies of bacterial targets in gut samples were hence calculated by applying these efficiencies to the qPCR results as E^{i-C_T} (E – primer efficiency calculated from the slope of the standard curve as $10^{\frac{-1}{slope}}$, i – intercept from standard curve and C_T – cycle threshold obtained from the qPCR run). C_T values above 26.53 could not be distinguished from the blank in standard curves and were hence considered below the detection limit. Copy number was adjusted for the DNA extraction fraction volume, elution volume, and dilution factor to estimate the number of copies per bee using the code (qpcr_data_parse.R) provided in the associated GitHub repository.

PREPROCESSING OF SEQUENCING DATA

Raw reads from sequencing were processed using DADA2 v1.30.0 (Callahan et al., 2016) to infer amplicon sequence variants (ASV) and their counts per sample. The detailed code used for this analysis is provided in the associated GitHub repository. The strategy to be used for full-length 16S amplicon sequencing using PacBio reads and new functions and parameters defined for this purpose in DADA2 has been described before (Callahan et al., 2019). Briefly, the removePrimers function was used for primer removal and automatically re-orienting all reads in the forward direction. Reads were then filtered to only keep reads in the range of 1000 – 1600 bp. Finally reads were dereplicated and denoised using a PacBio sequencing specific error model.

ESTIMATION OF STRAIN ABSOLUTE ABUNDANCE

The ASVs hence obtained were matched with unique sequences in our custom database and used to infer the counts per strain that would result in the observed ASV counts table generated by DADA2. This was done using the code (infer_species_counts.py) provided in the associated GitHub repository. To resolve strain-level composition from ASV counts, we implemented a matrix-based least-squares approach, leveraging known relationships between strains and their 16S rRNA gene copy profiles. The observed ASV counts per sample (A) were modeled as a linear combination of strain copy profiles (C) and the unknown strain abundances (S), such that $A = C \cdot S$. Because C is typically not square (i.e., the system is over-determined due to more than one unique ASV per strain), S was estimated by minimizing $\|CS - A\|^2$ using the `np.linalg.lstsq` solver in Python. This was done for each sample such that A and S were the counts per ASV and counts per strain respectively for that sample. Matrix S from each sample was combined into the strain/species table used for further analysis. Reads from ASVs not matched to any sequences in the database were aggregated as unknown reads (6.25% or less of the total reads in 90% of the inoculated bee samples).

Samples with low 16S rRNA gene copies ($C_T > 26.53$ or undetermined) or sequencing depth (<100 known reads) were removed from further analyses ($n = 15$ of 365). The relative abundance of each strain was calculated the ratio of number of reads assigned to that strain and the total number of reads in that sample. The absolute abundance of each strain in a

sample was estimated as the product of relative abundance and total copy number for that sample. The detection limit of each sample was the number of copies yielding 1 read in that sample, *i.e.*, the ratio of total 16S copies and the number of reads sequenced. Any values of abundance below this were considered below the limit and their value was set to 1 (for plotting on a log scale without errors).

ACKNOWLEDGMENTS

We thank the Lausanne Genomics Technology Facility (GTF) team at the University of Lausanne and the Next Generation Sequencing Platform of the University of Bern for performing the high-throughput sequencing for the experiments. This work was supported by a Spirit grant (grant number 189496), an SNSF Consolidator grant ('GLOBEE,' grant number 213860), and the NCCR Microbiomes, a National Centre of Competence in Research (grant number 180575), funded by the Swiss National Science Foundation.

AUTHOR CONTRIBUTIONS

Conceptualization – AP, GM, FM, PE; Methodology – AP, GM, FM, PE; Software – AP; Validation – AP; Formal analysis – AP; Investigation – AP, AS, GM; Resources – PE; Data curation – AP; Writing - Original Draft – AP; Review and Editing – AP, FM, PE; Visualization – AP; Supervision – PE, FM; Project administration – PE; Funding acquisition – PE

COMPETING INTERESTS

All authors declare they have no competing interests.

DATA AVAILABILITY

Raw reads will be deposited in the NCBI SRA database (TBD). Sequenced genomes from Isolates will be deposited to NCBI (TBD). Several intermediate files, including RDS objects, are found in the GitHub repository:

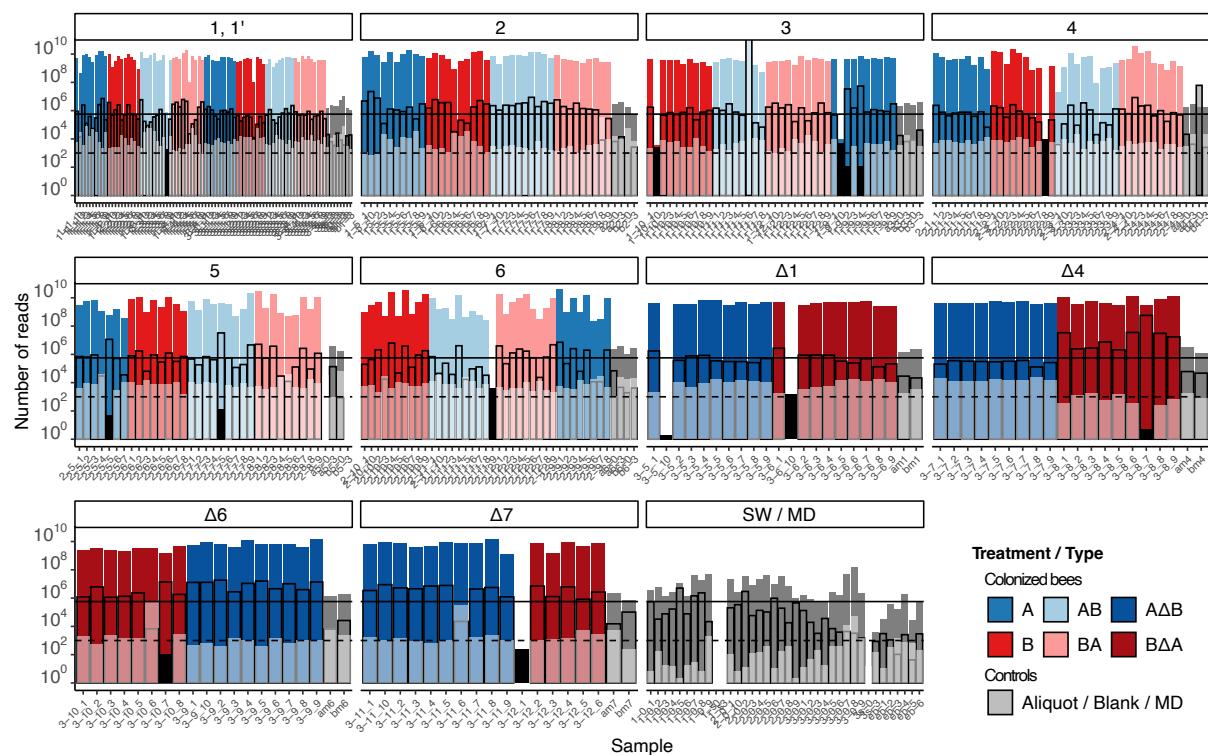
https://github.com/Aiswarya-prasad/20240399_aprasad_PriorityEffects

CODE AVAILABILITY

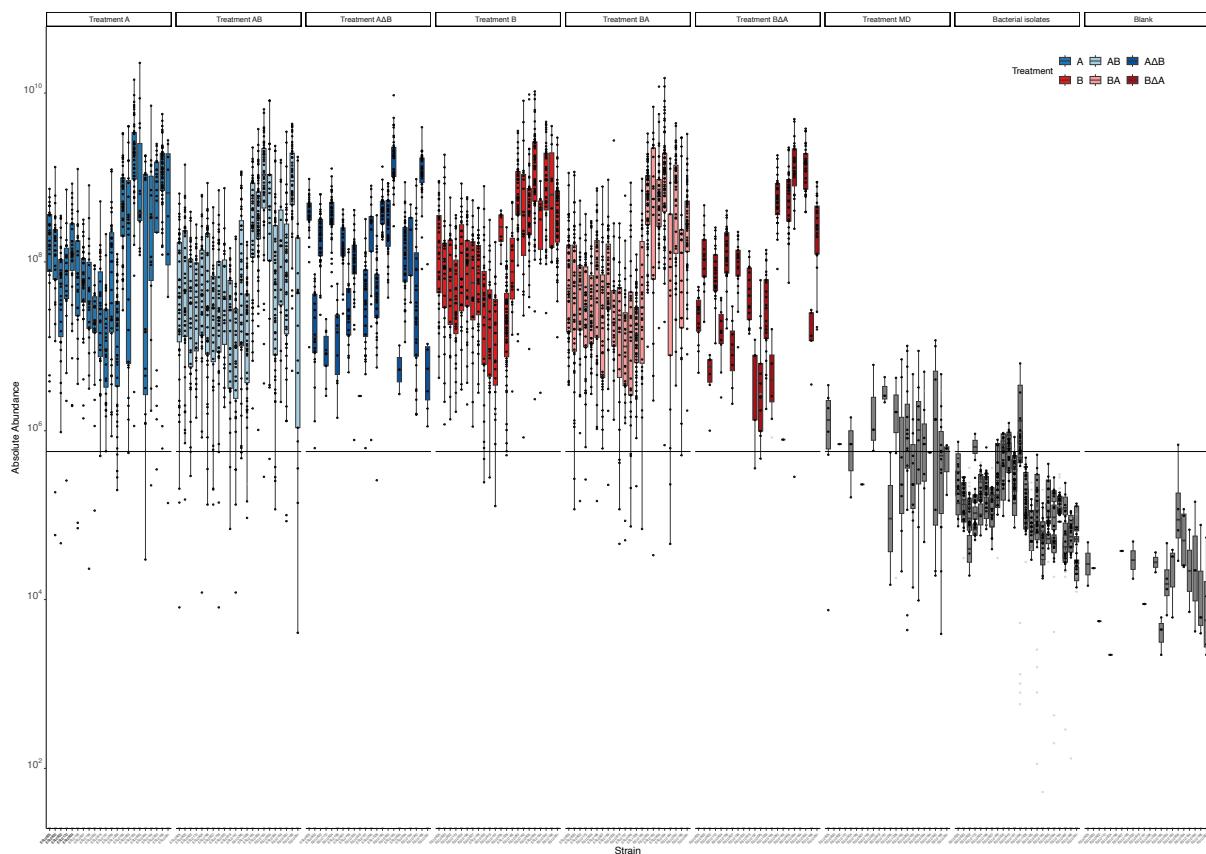
All the code and intermediate files, including RDS objects, are found in the GitHub repository:
https://github.com/Aiswarya-prasad/20240399_aprasad_PriorityEffects

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY FIGURES



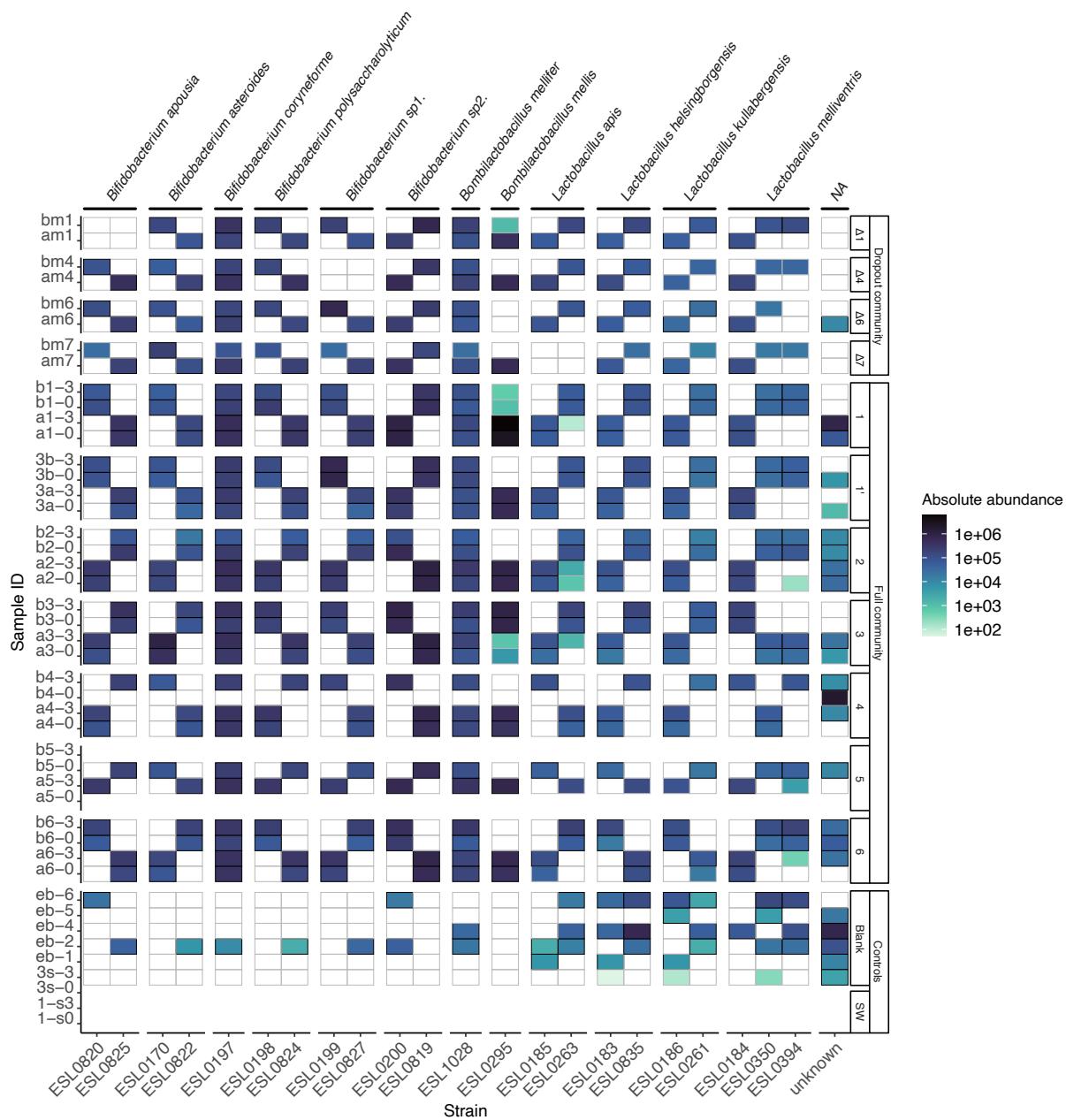
Supplementary Figure S1. Total bacterial abundance across samples. Total abundance of all strains per sample colored by treatment and facetted by the community combination ($m=\Delta$). White shaded bars indicate the sampling depth in terms of the number of filtered reads. Solid black bars denote samples with low depth that were excluded from further analysis. The bars outlined in black represent the detection threshold for each sample established as the bacterial abundance corresponding to one sequencing read (i.e., the total bacterial 16S copies from qPCR divided by the total number of reads sequenced per sample). The solid horizontal line denotes the median detection threshold across all samples, and the dotted line the median depth per sample.



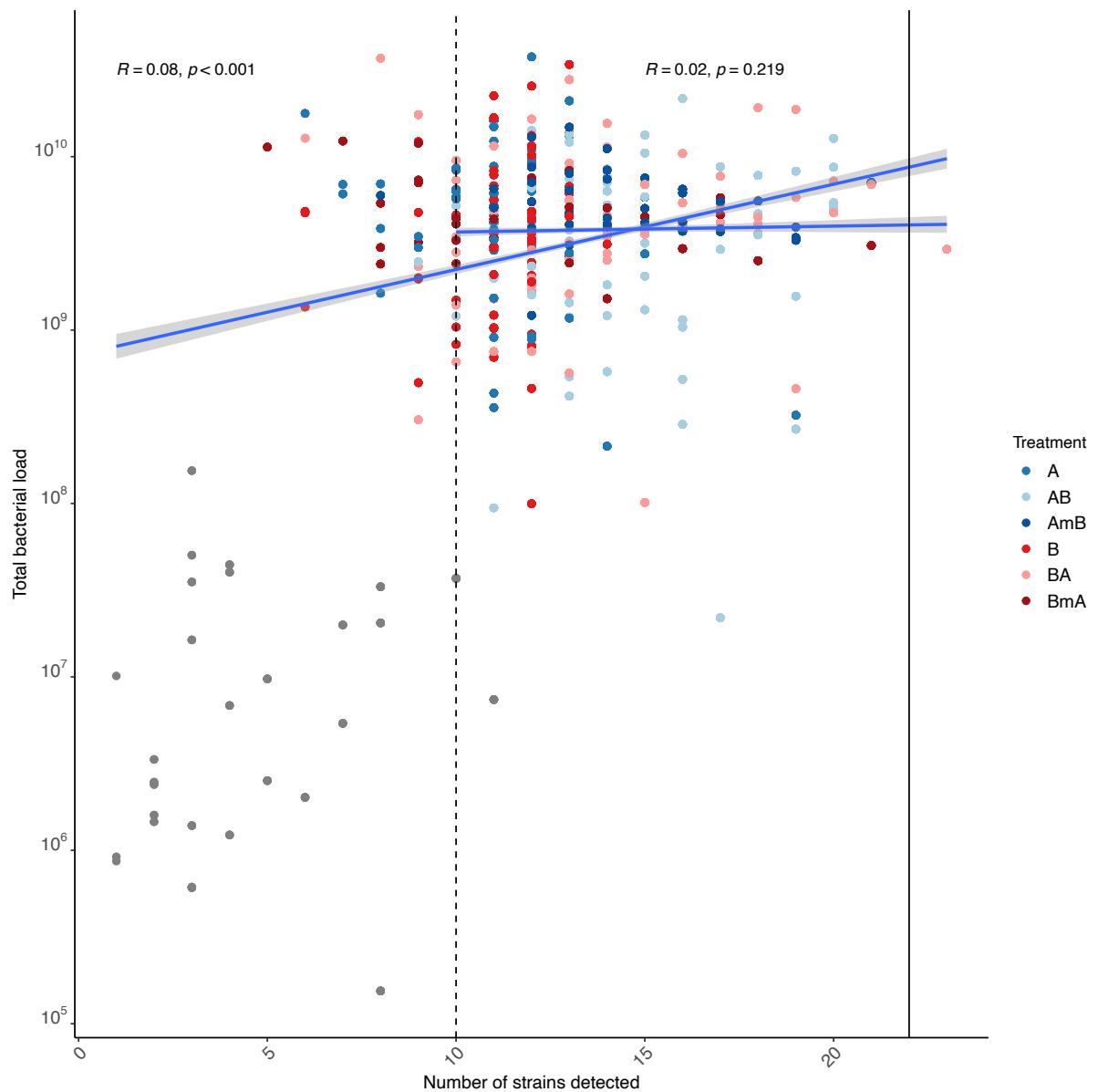
Supplementary Figure S2. Abundance of strains in honeybee gut and control treatments. Boxplots are based on the absolute abundance of each strain (on the x-axis) across different samples of each treatment, if it was above the detection threshold. Facets are made based on the treatment group for bee gut samples (MD treatment in grey), bacteria samples representing aliquots fed during the experiment, and blanks included during sample processing. Grey translucent points represent strains that were below the detection threshold in their respective samples. The horizontal line denotes the median detection threshold across all samples.



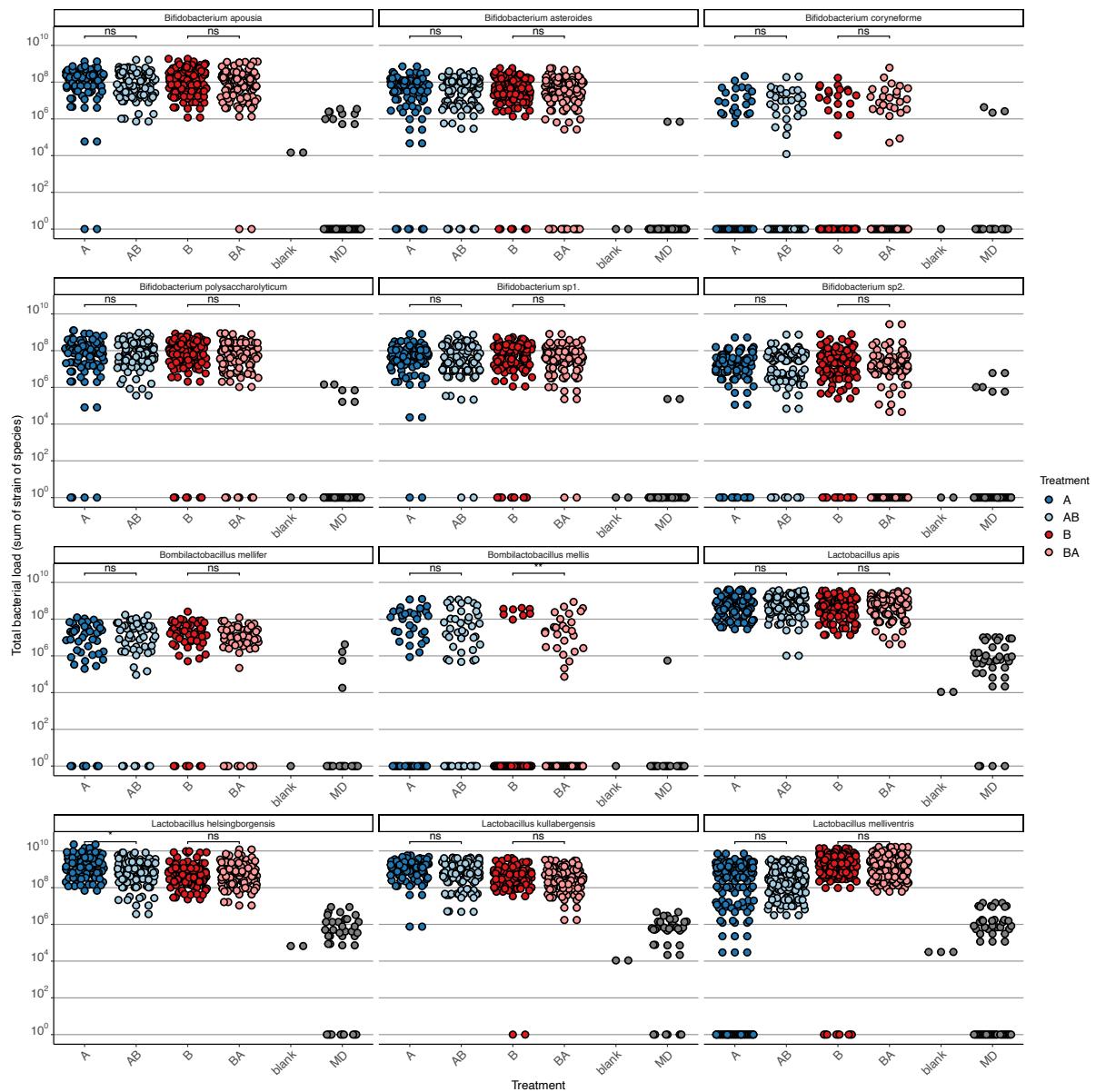
Supplementary Figure S3. Heatmap of the number of strains detected in each sample. Each row corresponds to a species, and each column represents a different honeybee sample. The heatmaps are faceted by experiment, community combination and treatments. The color on the heatmap indicates the number of strains of the species detected (absolute abundance above the detection threshold of the respective sample).



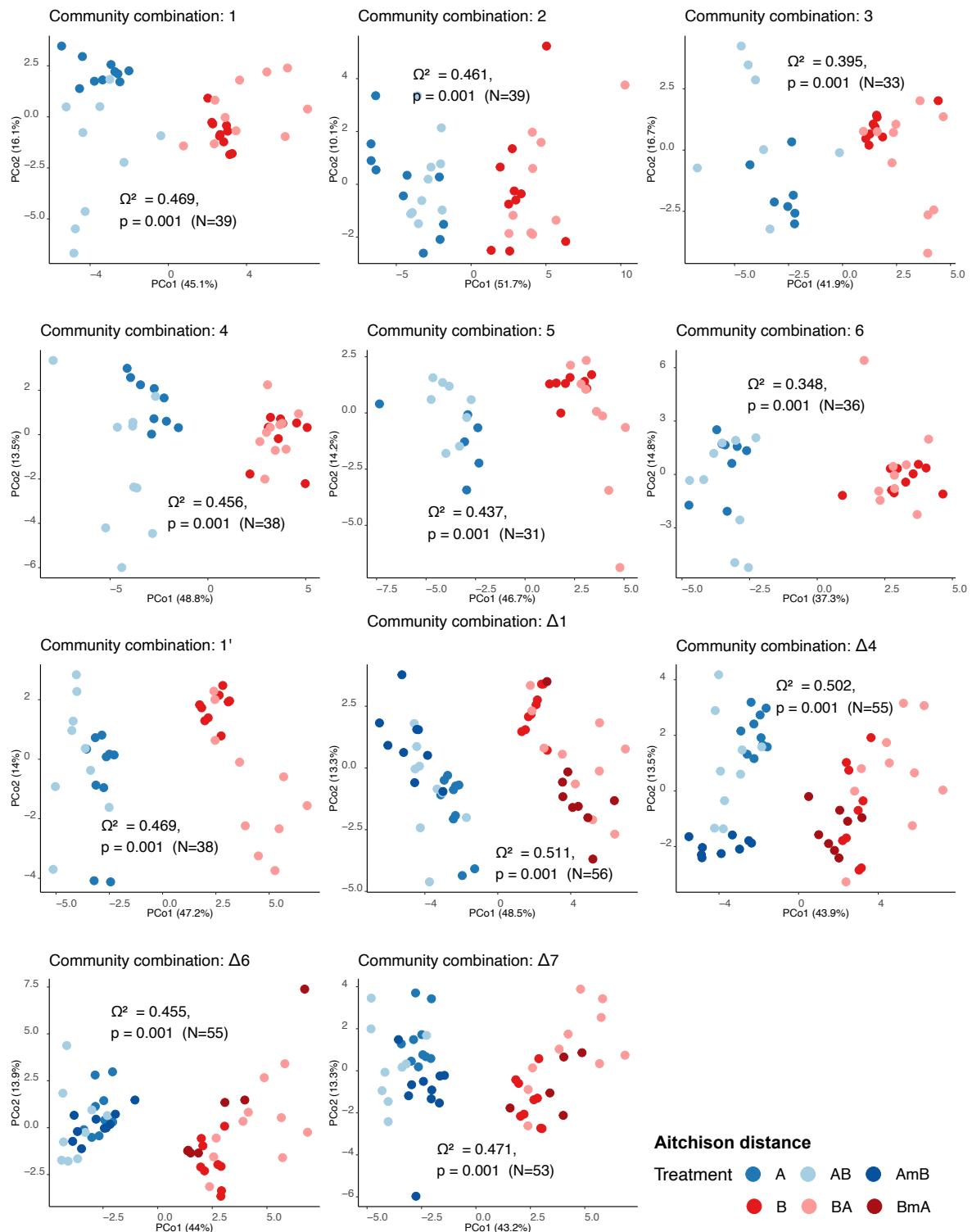
Supplementary Figure S4. Heatmap showing strains detected in aliquots of inoculum. Each column represents a strain, and each row is a different sample faceted by community type. For dropout communities, the aliquots for sequencing were collected from the tube used on Day 0, for the full community samples, an aliquot was taken from the tube used for inoculation on Day 0 and on Day 3 (sample IDs suffixed with '-0' or '-3'). The controls comprised six extraction blanks ('eb-'), which were samples of nuclease-free water added to the plates for DNA extraction and further processing. The others are samples of sterile sugar water from Day 0 and Day 3.

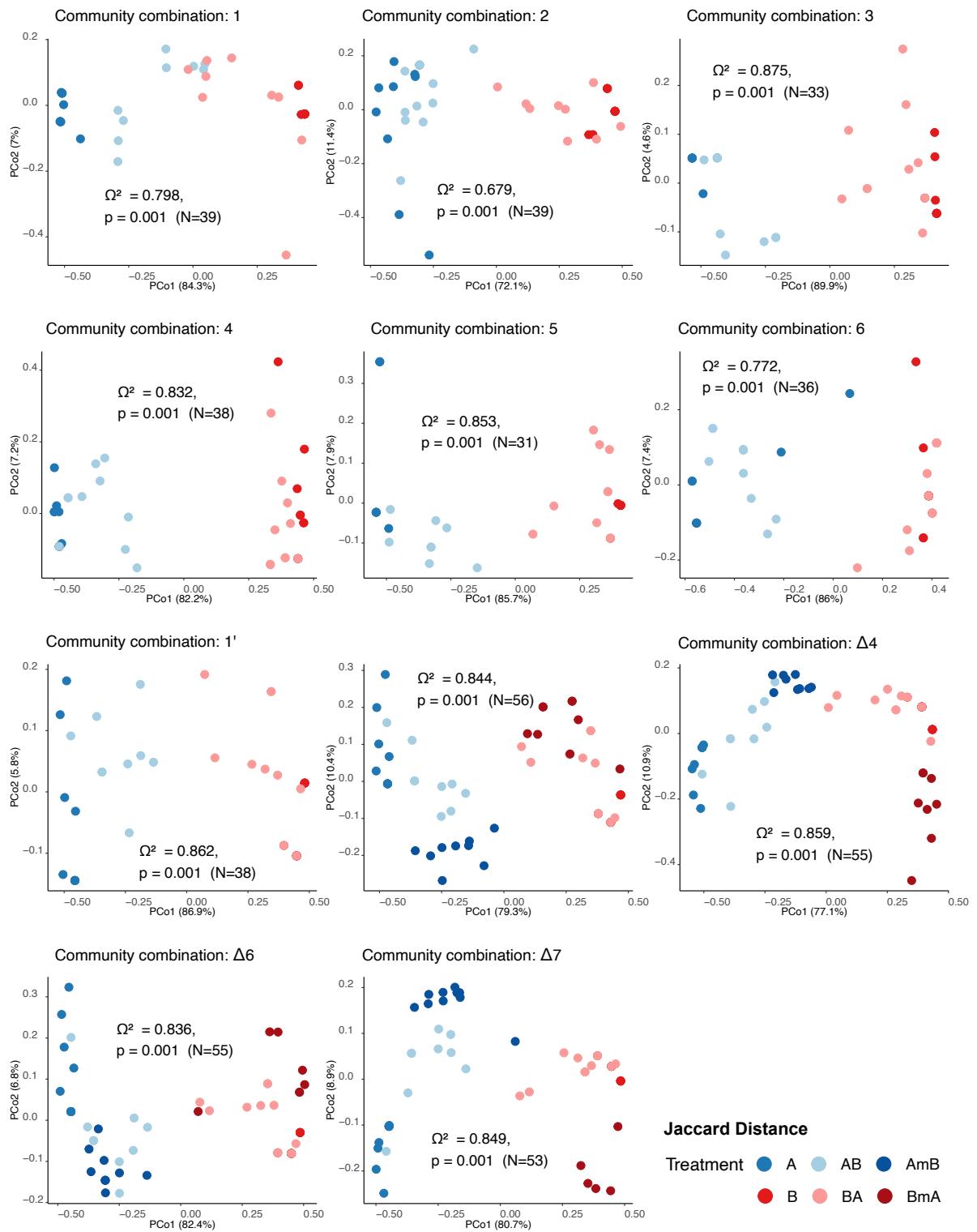


Supplementary Figure S5. Correlation between the number of strains and bacterial load. A scatter plot of the total 16S rRNA copies estimated by qPCR for each sample against the number of strains detected in the sample. The colors of the points indicate the treatment group to which the sample belongs. Spearman's ρ and p -value are estimated for including and excluding control and low bacterial load samples (due to lack of successful colonization of > 1 strain) and displayed on the plot.

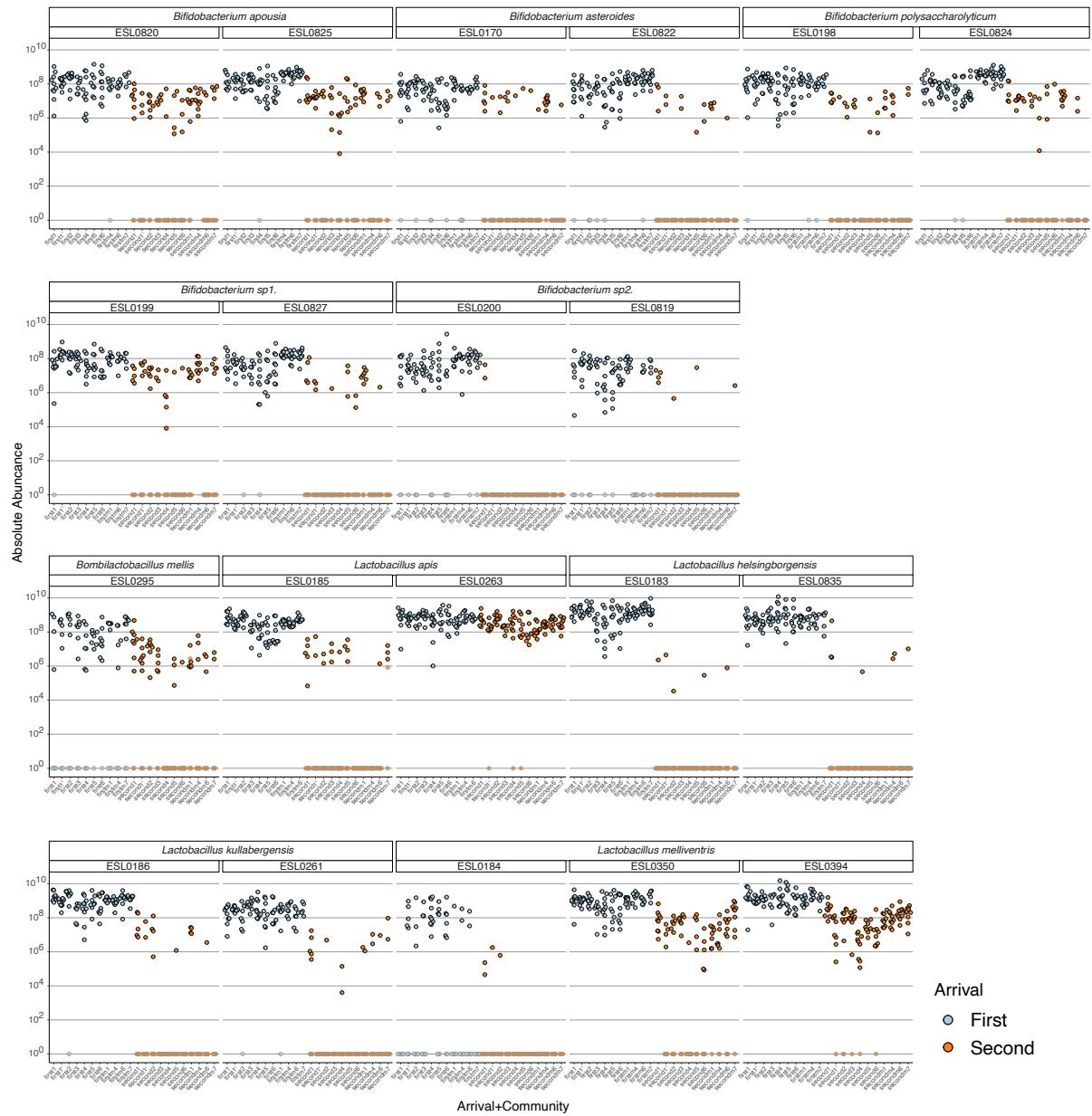


Supplementary Figure S6. Total bacterial load of each species. Absolute abundance of strains summed by species for each treatment group A and B, where only one strain per species was inoculated, and AB and BA, where two strains, one first comer (Day 0) and one late comer (Day 3) were inoculated. Only points crossing the detection threshold for their respective sample are included. Horizontal lines represent the Benjamini-Hochberg adjusted p-value ($p < 0.05$) between treatments with one or two strains inoculated (A vs. AB and B vs. BA). Points where none of the strains of the species were detected were set to 1.

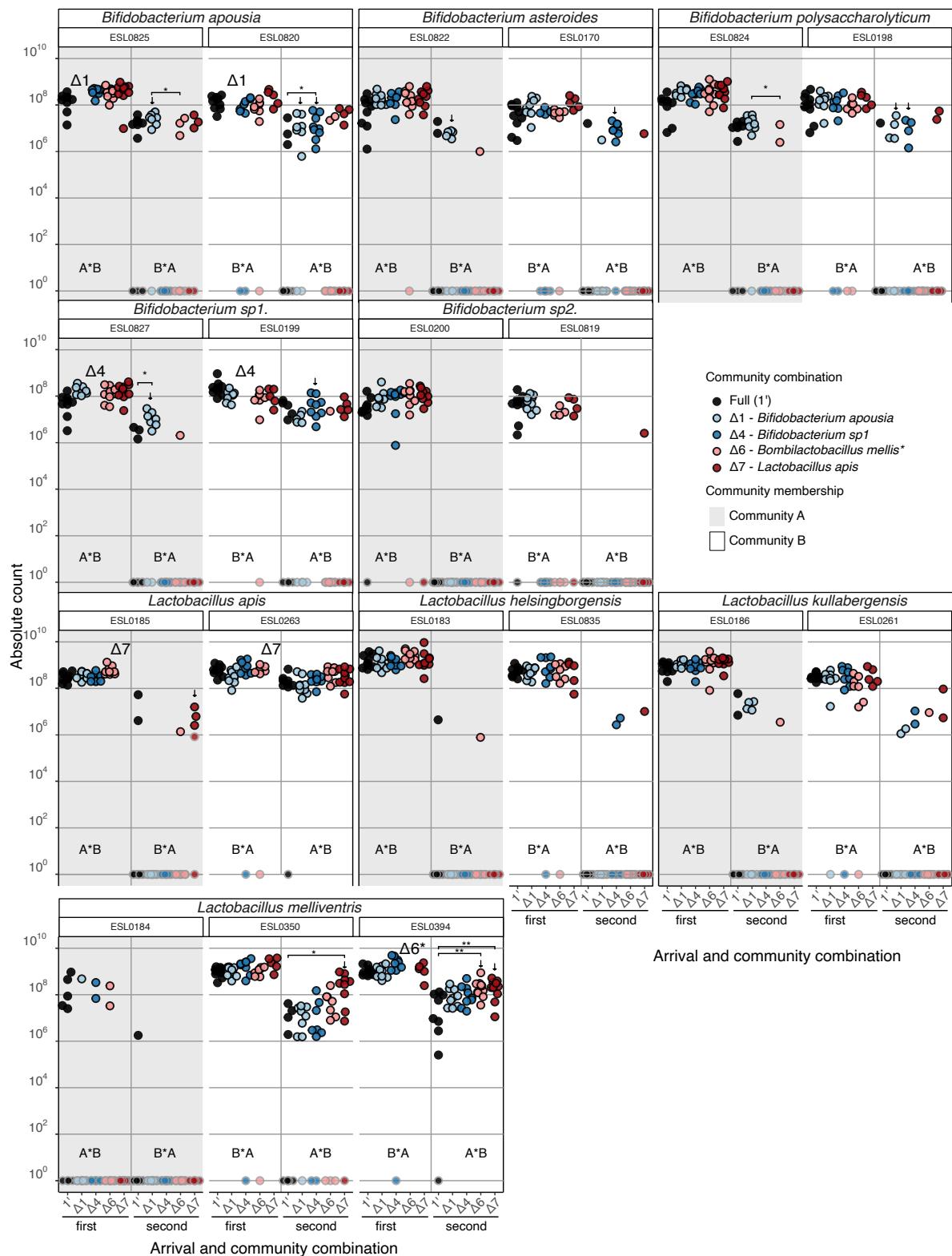




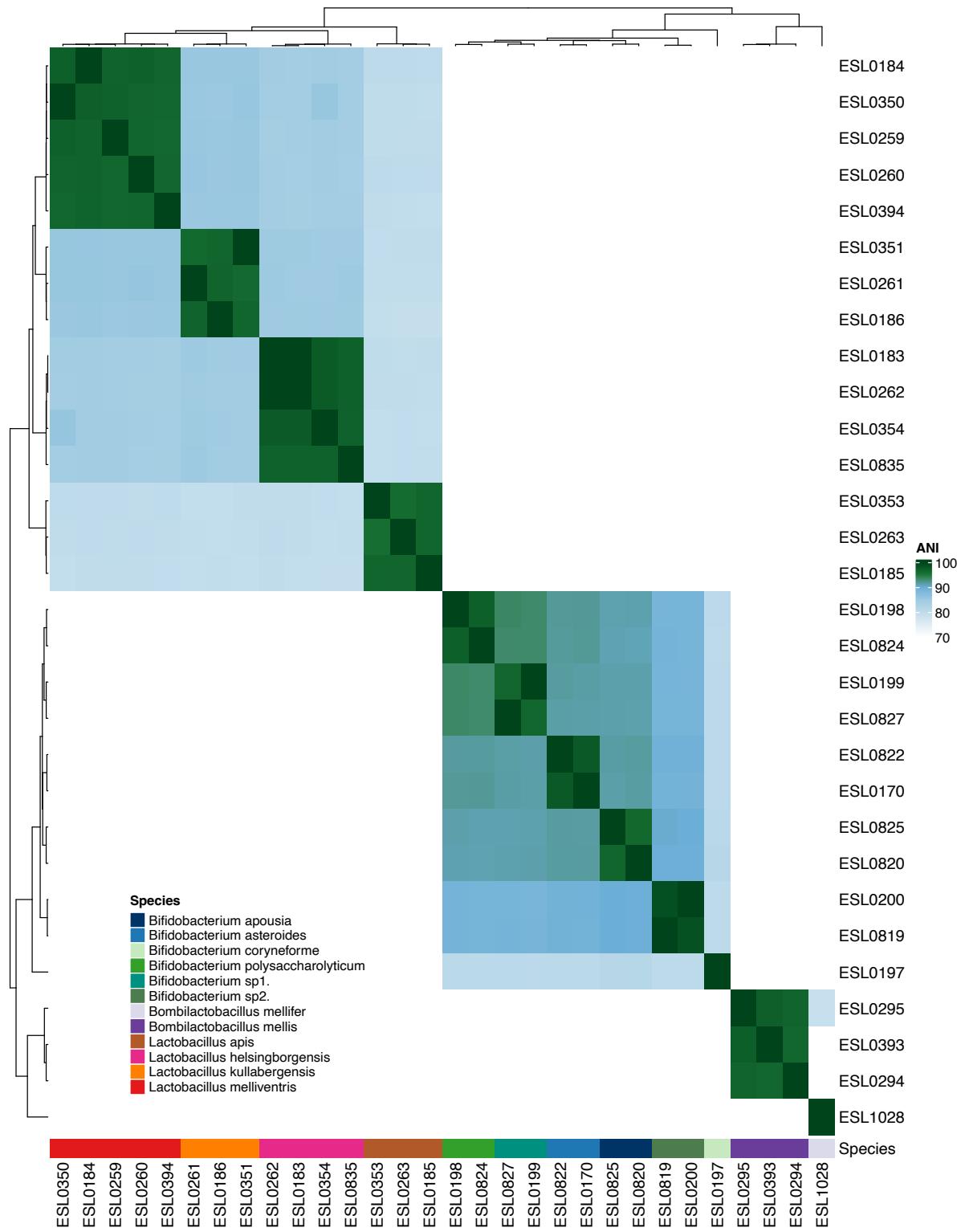
Supplementary Figure S7. Community composition across samples. PCoA plots to visualize the distance estimated by two different methods (Jaccard based on presence-absence and Aitchison based on relative abundance of strains) across samples inoculated with different community combinations across treatments A, B, AB, BA, and/or $A\Delta^*B$ and $B\Delta^*A$.



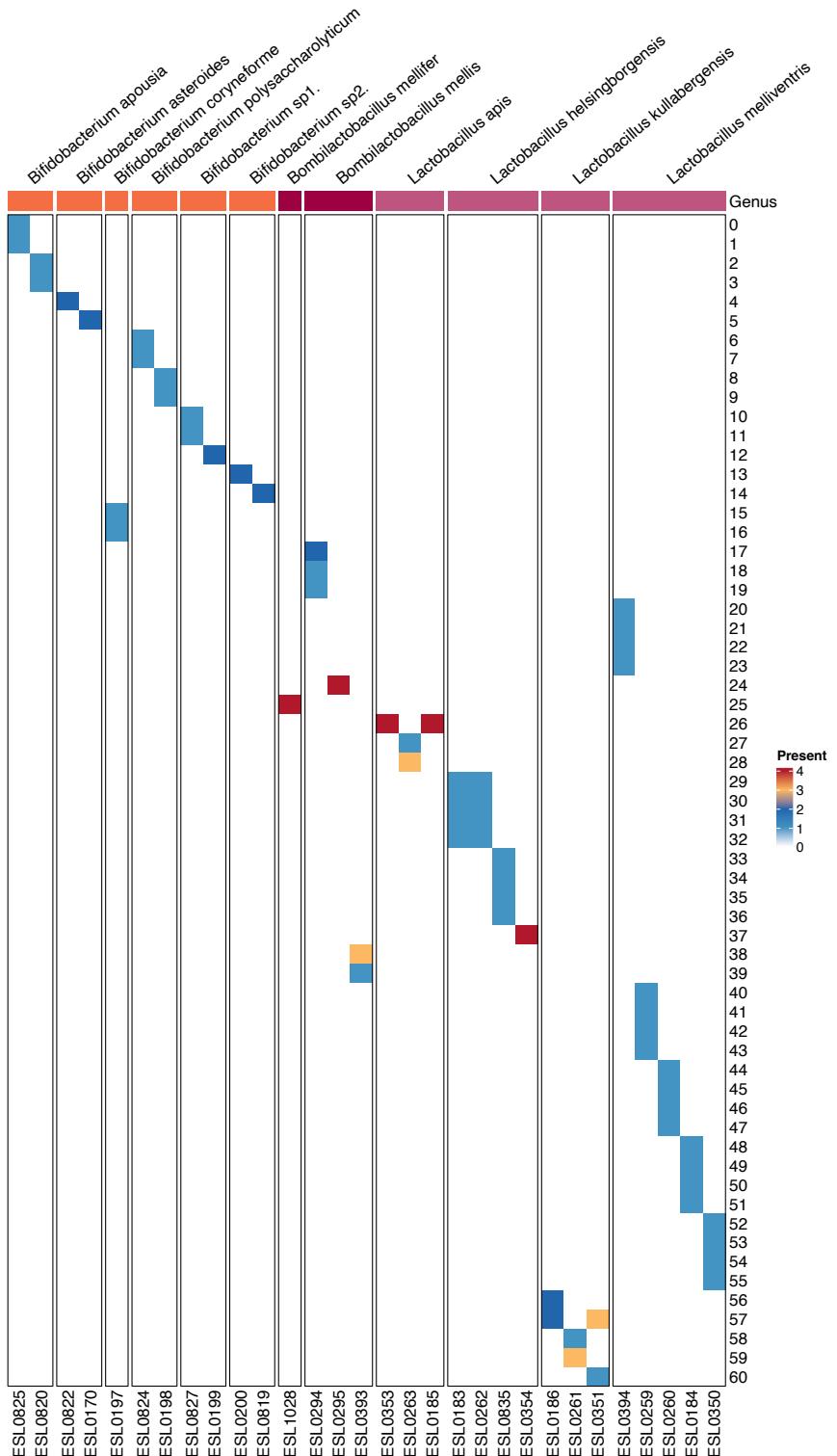
Supplementary Figure S8. Colonization success of strains across treatments. Absolute abundance of each strain across treatments where the strain was either in the first or second inoculated community, also indicated by the color of the points. Points with grey outlines represent samples where the strain was not detected (below the detection threshold for that sample) and hence set to a value of 1.



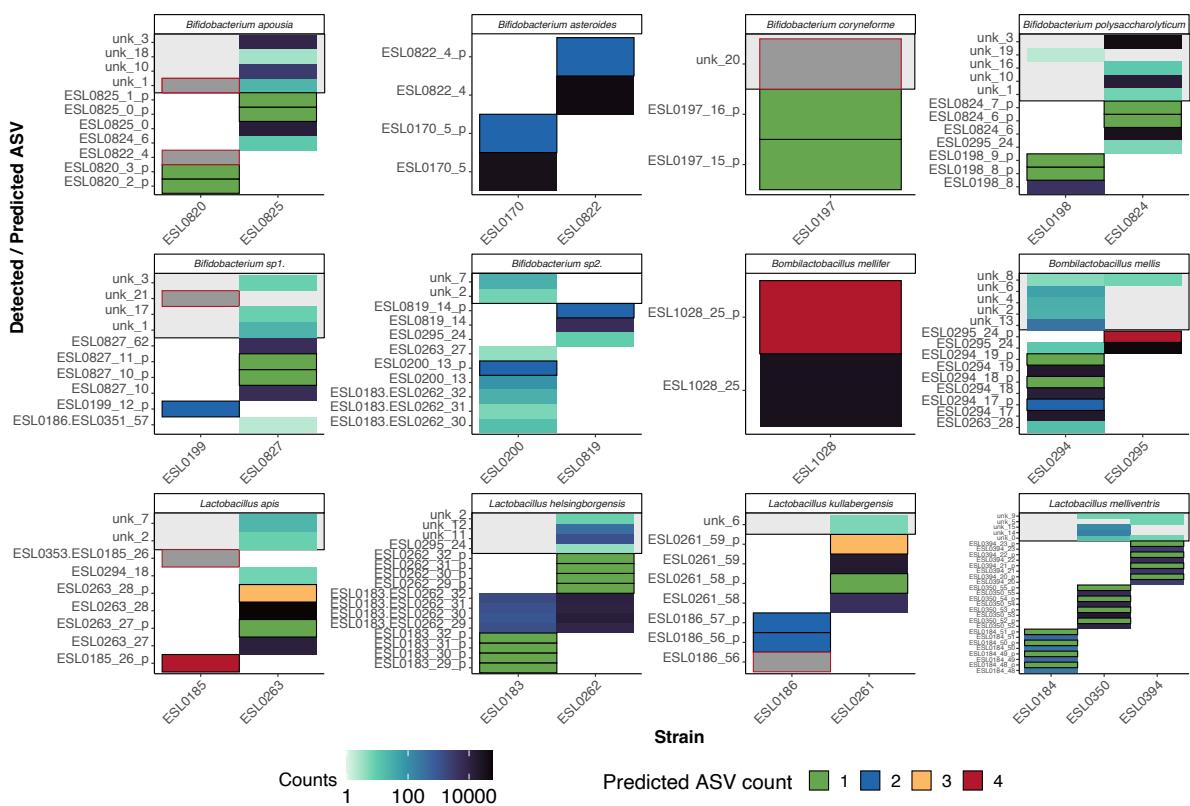
Supplementary Figure S9. Effect of dropping out species from firstcomer community. Abundance of each strain across full and dropout treatments. Shaded panels indicate strains that were present in community A and its dropout variations and hence, were firstcomers in A*B and second in B*A. Result of Wilcoxon rank sum test (two-sided) between abundance of latecomers in full and dropout treatments are annotated (* - $p < 0.05$), complete results of statistical test including sample size per group are included separately (Supplementary Table S6). Small black arrows highlight dropout treatments where the latecomer showed better colonization success in terms of quantity or frequency. Points with grey outlines represent samples where the strain was not detected (below the detection threshold for that sample) and hence set to a value of 1.



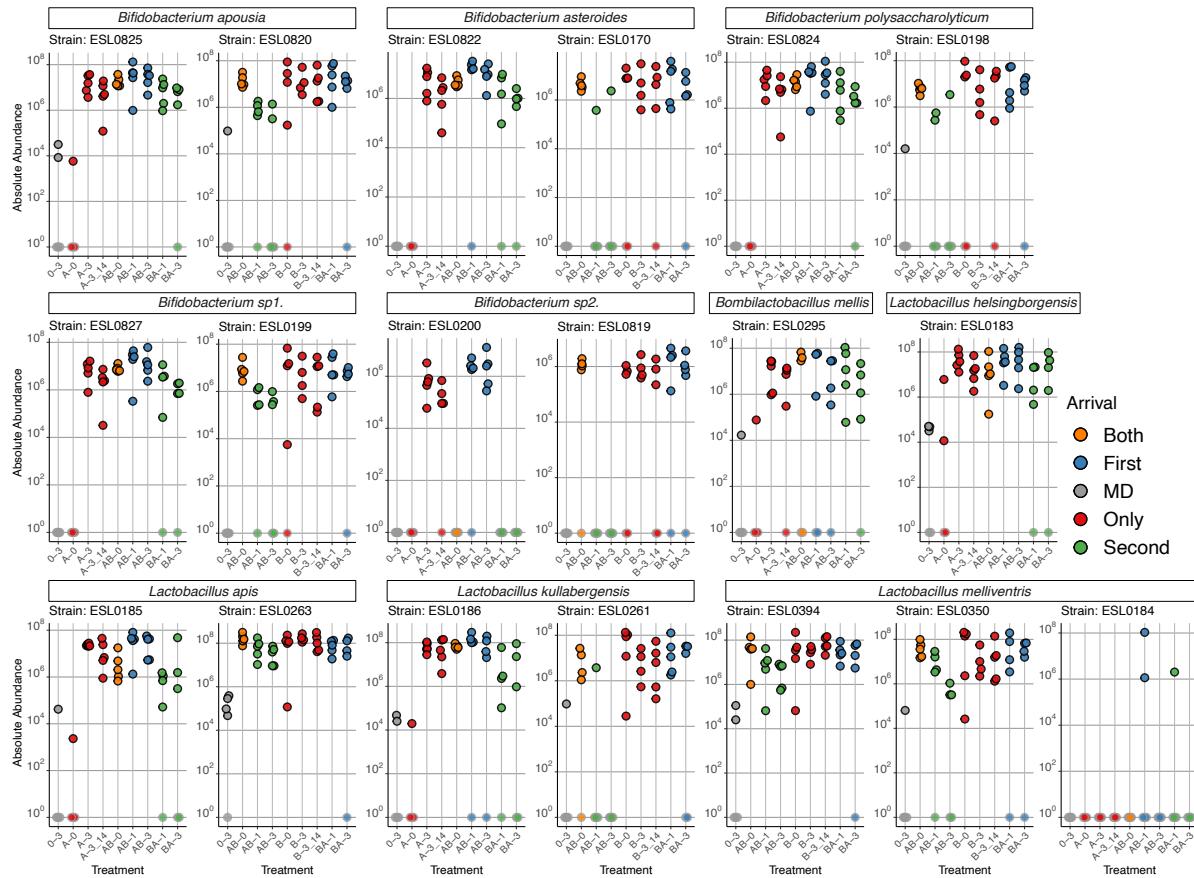
Supplementary Figure S10. Heatmap of average nucleotide identity. ANI heatmap of all strains clustered and annotated by species.



Supplementary Figure S11. Heatmap of 16S rRNA gene sequence variants predicted across strains. Each row represents a unique ASV of the 16S rRNA gene and each column a strain. The color of the boxes indicates the number of copies of that variant found in each respective strain as inferred based on Barrnap using the re-sequenced long-read genome assemblies of these strains. Strains are faceted by the species they belong to and the bar on top colors panels based on the Genus. ASVs are named by the unique IDs (uid) assigned to them in our custom database.



Supplementary Figure S12 Predicted and detected ASV sequences and counts. For each strain, copy number of predicted ASVs (marked by thick black boxes), and counts of detected sequences are shown faceted by species. Samples where the sequencing depth was too low for detection of ASVs (total reads <100) are indicated by a grey box (with red outline). On the vertical axis, ASV sequences predicted from the genome are suffixed with “_p”, their exact matches are named with the strain name and suffixed with their unique ID (_uid) in the custom database, and unknown sequences are prefixed with “unk_”.



Supplementary Figure S13. Pilot experiment. Absolute abundance of each strain across treatments in the pilot experiment, colored and faceted by arrival order of the strain in the respective treatments. Points with grey outlines represent samples where the strain was not detected (below the detection threshold for that sample) and hence set to a value of 1.

SUPPLEMENTARY TABLES

Supplementary tables can be found in the file at this link:

https://unils-my.sharepoint.com/:x/g/personal/aiswarya_prasad_unil_ch/EQV11r4ypXNDmOSIQLP6L8wBMSsi-VdLxeIYxJpM80FyyQ?e=aOogT6

Supplementary Table S1. Information about each strain and its genomes. Associated ASVs provides a list of ASVs by Unique ID that are found in the respective genome. Repeated IDs in the list represent multiple copies of the same sequence in the genome. Sequences associated with ASV IDs are provided separately in Supplementary Table S7.

Supplementary Table S2. Composition of communities with examples of community combination pairs in treatments.

Supplementary Table S3. Result of Wilcoxon rank sum test (one-sided) carried out to compare latecomer strains across dropout and full community combinations in Experiment 3.

Supplementary Table S4. List of absolute abundance of each strain in all the bees of given treatment. Eg. the list for ESL0825 under "only" is the list of the abundance of all the bees in which for community combination 1 the community in which ESL0825 was present (A) was fed as the only community and "first" lists its abundance when community A was fed first (AB). Finally, the set of values in 1' to be compared with the strains in the dropout treatment are all highlighted in bold.

Supplementary Table S5. Results of GLMM using the function `glmer` model `log_10(AbsoluteAbundance) ~ ArrivalOrder + (1 | CommunityCombo)`, with a log link specified by `family = Gamma(link = "log")`

Supplementary Table S6. Result of wilcoxon rank sum test (two-sided) carried out to compare latecomer strains across full community combinations in In Figure 3.

Supplementary Table S7. Table of all copies of 16S full length Amplicon sequence either predicted from full genome sequences or obtained from sequencing of 16S rRNA amplicons of isolate strains.

Supplementary Table S8. Complete table of raw data per strain across all samples and treatments

REFERENCES

- Adler, P. B., HilleRisLambers, J., & Levine, J. M. (2007). A niche for neutrality. *Ecology Letters*, 10(2), 95–104. <https://doi.org/10.1111/j.1461-0248.2006.00996.x>
- Baud, G. L. C., Prasad, A., Ellegaard, K. M., & Engel, P. (2023). Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers. *Genome Biology*, 24(1), 283. <https://doi.org/10.1186/s13059-023-03131-4>
- Boyle, J. A., Simonsen, A. K., Frederickson, M. E., & Stinchcombe, J. R. (2021). Priority effects alter interaction outcomes in a legume–rhizobium mutualism. *Proceedings of the Royal Society B: Biological Sciences*, 288(1946), 20202753. <https://doi.org/10.1098/rspb.2020.2753>
- Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., & Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife*, 10, e68583. <https://doi.org/10.7554/eLife.68583>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., McGill, S. K., & Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, 47(18), e103. <https://doi.org/10.1093/nar/gkz569>
- Callegari, M., Crotti, E., Fusi, M., Marasco, R., Gonella, E., De Noni, I., Romano, D., Borin, S., Tsiamis, G., Cherif, A., Alma, A., & Daffonchio, D. (2021). Compartmentalization of bacterial and fungal microbiomes in the gut of adult honeybees. *Npj Biofilms and Microbiomes*, 7(1), 42. <https://doi.org/10.1038/s41522-021-00212-9>
- Carlström, C. I., Field, C. M., Bortfeld-Miller, M., Müller, B., Sunagawa, S., & Vorholt, J. A. (2019). Synthetic microbiota reveal priority effects and keystone strains in the *Arabidopsis* phyllosphere. *Nature Ecology & Evolution*, 3(10), Article 10. <https://doi.org/10.1038/s41559-019-0994-z>
- Chen, J. Z., Junker, A., Zheng, I., Gerardo, N. M., & Vega, N. M. (2024). A strong priority effect in the assembly of a specialized insect-microbe symbiosis. *Applied and Environmental Microbiology*, 90(10), e00818-24. <https://doi.org/10.1128/aem.00818-24>
- Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, 31(1), 343–366. <https://doi.org/10.1146/annurev.ecolsys.31.1.343>
- Debray, R., Herbert, R. A., Jaffe, A. L., Crits-Christoph, A., Power, M. E., & Koskella, B. (2022). Priority effects in microbiome assembly. *Nature Reviews Microbiology*, 20(2), 109–121. <https://doi.org/10.1038/s41579-021-00604-w>
- Ellegaard, K. M., & Engel, P. (2019). Genomic diversity landscape of the honey bee gut microbiota. *Nature Communications*, 10(1), 446. <https://doi.org/10.1038/s41467-019-08303-0>
- Engel, P., Martinson, V. G., & Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences*, 109(27), 11002–11007. <https://doi.org/10.1073/pnas.1202970109>
- Fukami, T., Mordecai, E. A., & Ostling, A. (2016). A framework for priority effects. *Journal of Vegetation Science*, 27(4), 655–657. <https://doi.org/10.1111/jvs.12434>
- Garrido-Sanz, D., & Keel, C. (2024). Priority effects of heritable seed-borne bacteria drive early assembly of the wheat rhizosphere microbiome (p. 2024.10.21.619384). bioRxiv. <https://doi.org/10.1101/2024.10.21.619384>

Garud, N. R., Good, B. H., Hallatschek, O., & Pollard, K. S. (2019). Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLOS Biology*, 17(1), e3000102. <https://doi.org/10.1371/journal.pbio.3000102>

Gleason, H. A. (1927). Further Views on the Succession-Concept. *Ecology*, 8(3), 299–326. <https://doi.org/10.2307/1929332>

Gurung, A., Mukherjee, S., Declercq, M., Souffreau, C., & De Meester, L. (2024). Strain-dependent and host genotype-dependent priority effects in gut microbiome assembly affect host fitness in. *Limnology and Oceanography*, 69(8), 1782–1796. <https://doi.org/10.1002/limo.12614>

Kešnerová, L., Mars, R. A. T., Ellegaard, K. M., Troilo, M., Sauer, U., & Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLOS Biology*, 15(12), e2003467. <https://doi.org/10.1371/journal.pbio.2003467>

Laursen, M. F., & Roager, H. M. (2023). Human milk oligosaccharides modify the strength of priority effects in the *Bifidobacterium* community assembly during infancy. *The ISME Journal*, 17(12), 2452–2457. <https://doi.org/10.1038/s41396-023-01525-7>

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., & Gordon, J. I. (2008). Worlds within worlds: Evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, 6(10), 776–788. <https://doi.org/10.1038/nrmicro1978>

Martínez, I., Maldonado-Gómez, M. X., Gomes-Neto, J. C., Kittana, H., Ding, H., Schmaltz, R., Joglekar, P., Cardona, R. J., Marsteller, N. L., Kembel, S. W., Benson, A. K., Peterson, D. A., Ramer-Tait, A. E., & Walter, J. (2018). Experimental evaluation of the importance of colonization history in early-life gut microbiota assembly. *eLife*, 7, e36521. <https://doi.org/10.7554/eLife.36521>

Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities: Phylogeny and coexistence. *Ecology Letters*, 13(9), 1085–1093. <https://doi.org/10.1111/j.1461-0248.2010.01509.x>

Mazel, F., Prasad, A., & Engel, P. (2025). Host specificity of gut microbiota associated with social bees: Patterns and processes. *Microbiology and Molecular Biology Reviews*, 0(0), e00080-23. <https://doi.org/10.1128/mmbr.00080-23>

Motta, E. V. S., Powell, J. E., Leonard, S. P., & Moran, N. A. (2022). Prospects for probiotics in social bees. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1853), 20210156. <https://doi.org/10.1098/rstb.2021.0156>

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P., & Ferrenberg, S. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews : MMBR*, 77(3), 342–356. <https://doi.org/10.1128/MMBR.00051-12>

Ojima, M. N., Jiang, L., Arzamasov, A. A., Yoshida, K., Odamaki, T., Xiao, J., Nakajima, A., Kitaoka, M., Hirose, J., Urashima, T., Katoh, T., Gotoh, A., van Sinderen, D., Rodionov, D. A., Osterman, A. L., Sakanaka, M., & Katayama, T. (2022). Priority effects shape the structure of infant-type *Bifidobacterium* communities on human milk oligosaccharides. *The ISME Journal*, 16(9), 2265–2279. <https://doi.org/10.1038/s41396-022-01270-3>

Orr, J. A., Armitage, D. W., & Letten, A. D. (2025). Coexistence Theory for Microbial Ecology, and Vice Versa. *Environmental Microbiology*, 27(3), e70072. <https://doi.org/10.1111/1462-2920.70072>

Peay, K. G., Belisle, M., & Fukami, T. (2011). Phylogenetic relatedness predicts priority effects in nectar yeast communities. *Proceedings of the Royal Society B: Biological Sciences*, 279(1729), 749–758. <https://doi.org/10.1098/rspb.2011.1230>

Prasad, A., Pallujam, A. D., Siddaganga, R., Suryanarayanan, A., Mazel, F., Brockmann, A., Yek, S. H., & Engel, P. (2025). *Symbiont loss and gain, rather than co-diversification shapes honeybee gut microbiota diversity and function* (p. 2024.09.11.612390). bioRxiv. <https://doi.org/10.1101/2024.09.11.612390>

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D. R., Kultima, J. R., Martin, J., Kota, K., Sunyaev, S. R., Weinstock, G. M., & Bork, P. (2013). Genomic variation landscape of the human gut microbiome. *Nature*, 493(7430), 45–50. <https://doi.org/10.1038/nature11711>

Seemann, T. (2025). *Tseemann/barrnap* [Perl]. <https://github.com/tseemann/barrnap> (Original work published 2013)

Segura Munoz, R. R., Mantz, S., Martínez, I., Li, F., Schmaltz, R. J., Pudlo, N. A., Urs, K., Martens, E. C., Walter, J., & Ramer-Tait, A. E. (2022). Experimental evaluation of ecological principles to understand and modulate the outcome of bacterial strain competition in gut microbiomes. *The ISME Journal*, 16(6), 1594–1604. <https://doi.org/10.1038/s41396-022-01208-9>

Spragge, F., Bakkeren, E., Jahn, M. T., Araujo, E. B. N., Pearson, C. F., Wang, X., Pankhurst, L., Cunrath, O., & Foster, K. R. (2023). Microbiome diversity protects against pathogens by nutrient blocking. *Science*. <https://doi.org/10.1126/science.adj3502>

Van Rossum, T., Ferretti, P., Maistrenko, O. M., & Bork, P. (2020). Diversity within species: Interpreting strains in microbiomes. *Nature Reviews Microbiology*, 18(9), 491–506. <https://doi.org/10.1038/s41579-020-0368-1>

Wolff, R., Shoemaker, W., & Garud, N. (2023). Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. *mBio*, 14(2), e02502-22. <https://doi.org/10.1128/mbio.02502-22>

Yan, Y., Nguyen, L. H., Franzosa, E. A., & Huttenhower, C. (2020). Strain-level epidemiology of microbial communities and the human microbiome. *Genome Medicine*, 12(1), 71. <https://doi.org/10.1186/s13073-020-00765-y>

CHAPTER-4

GENERAL DISCUSSION

In this thesis, I set out to understand how gut microbial communities are shaped in the long term and patterns resulting from short-term processes shaping community assembly. In the first study (*Chapter 2*), I addressed the question of how host-associated gut microbiomes are shaped over evolutionary time by comparing the gut microbiota of five closely related honeybee species. In the second study (*Chapter 3*), I study priority effects, an important factor influencing gut microbiota assembly across individuals at the strain level and could play a role in maintaining community composition over longer periods of time.

REASSESSING CO-DIVERSIFICATION

A key finding of the study presented in *Chapter 2* was that, despite a small number of clades showing a signal of co-diversification between honeybee species and their microbiota, this was not higher than expected by chance. Co-diversification has already been widely stated to operate in several host-microbiome systems, although rigorous quantitative verification across systems is still needed. When and why one might expect co-diversification between gut microbes and their host warrants more profound thought (Yoder, 2024). The conceptual foundation underlying co-diversification was derived from ideas put forth more than a century ago by Fahrenholz, a parasitologist studying lice. He wrote that the evolutionary history and taxonomy of parasites should closely reflect those of their host, based on morphological observations of hosts and their parasites (Fahrenholz, 1913). Such observations across systems ranging from plants to birds led to the belief that co-speciation (or co-diversification) was very common. However, conclusions of co-diversification were lacking sufficient evidence (de Vienne et al., 2013). Only in the last two decades of the 20th century was co-speciation tested more rigorously with robust independent phylogenies of hosts and parasites. Likewise, several reports of co-diversifying lineages, particularly in primate gut microbiomes, have led to a widely held belief that co-diversification must be prevalent among gut microbiota and their hosts; however, this has been challenged before in social bees (Sarton-Lohéac et al., 2023).

Extensive studies across systems are required to establish the prevalence and importance of co-diversification in shaping gut microbiome diversity. With the advent of high-throughput sequencing technologies, large-scale studies characterizing gut microbiomes in high resolution and breadth, it is now possible to infer and compare phylogenies of hundreds of microbes and their hosts. Further, shared evolutionary history need not be the explanation for the patterns of host specificity and phyllosymbiosis seen in several host-associated microbial communities. For example, host-filtering, a process wherein each host species selects different microbiota from the same environment, can also lead to the same pattern (Mazel et al., 2018). Further, once host lineages have diversified, dispersal limitation can ensure that their microbiota remain associated with their respective host across generations,

with limited opportunities for cross-host exchange. This and the fact that co-diversifying lineages are not prevalent (<36%) within gut microbiomes (Rühlemann et al., 2024; Sanders et al., 2023; Suzuki et al., 2021) suggests that other processes than co-diversification are likely to be important drivers of gut microbiome diversity across host species. Over-emphasis on the importance of shared evolutionary history risks the exploration of such essential ecological processes.

Comparison of patterns in gut microbiota composition across *Apis* bees has been informative about the processes shaping finer-scale diversity, e.g., strain-level composition, as there are several shared species (rather than shared genera or phyla between more divergent hosts). More in-depth comparisons of closely related hosts with similar ecology and behaviour are essential to further understand the processes that shape strain diversity over evolutionary time. This can be achieved through comparative analysis of microbiome diversity between the lesser-known relatives of *Apis cerana* and the several subspecies of *Apis mellifera* distributed across Africa (see *Introduction Section 3.1*). I am compiling the results of comparing several wild *Apis mellifera* subspecies in an ongoing study (not included in this thesis). Further, I compare their gut microbiome to that of managed honeybees from different parts of the world. This analysis will reveal to what extent human management, mixing of colonies and populations, have shaped gut microbiome diversity in the honeybee gut.

DRIVERS AND CONSEQUENCES OF HOST SPECIFICITY

Analysis of patterns in gut microbiome composition can offer helpful insight into the nature of host-specificity of microbiota. Though not all animals “have a microbiome” (Hammer et al., 2019), those that do often harbor microbial communities that are distinct from those of other species. However, the extent and nature of host specificity, particularly at the level of individual microbial species, is still not fully understood. Host-specific microbiota do not necessarily imply that closely related hosts harbor closely related bacteria (i.e., co-phylogeny), nor does it imply that each microbial lineage is strictly restricted to a single host. Instead, patterns of host specificity can emerge through a combination of ecological, evolutionary, and stochastic processes (Mazel et al., 2025).

Our study of the five honeybee species of the genus *Apis* (*Chapter 2*) found that gut microbial communities consist of specialist and generalist bacteria. Some microbes are host-restricted and consistently found in only one *Apis* species, while others are shared between two or more species, though often at different prevalences or in distinct strain compositions. This pattern suggests that host specificity operates at multiple taxonomic levels and might be driven by various underlying mechanisms.

Honeybee gut microbiota rely on host-mediated transmission for persistence across generations. In honey bees, transmission occurs through both vertical and horizontal routes; vertically during colony fission events like swarming, and horizontally via social interactions such as trophallaxis, grooming, and contact with hive surfaces (Moran & Sloan, 2015; Powell et al., 2014). Experimental inoculation studies have shown that microbes like *Snodgrassella alvi* and *Gilliamella apicola* are efficiently transmitted between individual *Apis mellifera* bees, and that colonization success likely depends on social contact (Powell et al., 2014). Cross-

species colonization only between *Apis* species is possible, at least in some cases. For example, a *Snodgrassella* species from *A. cerana* could colonize *A. mellifera*, but not the species from bumble bees (Kwong et al., 2017). Further work in isolating strains from the understudied Asian honeybees and experimental work for *in vitro* characterization, as well as *in vivo* experiments, are needed to understand the determinants of host-specificity and the ecological processes maintaining it (Mazel et al., 2025).

Differences in foraging behavior, colony structure, gut morphology, or physiology can create distinct ecological niches within each host species. These niches may favor particular microbial taxa, either through direct filtering by the host or through competition and niche partitioning among microbes, which has been demonstrated to promote species coexistence in the honeybee gut (Brochet et al., 2021). Experimental approaches are essential for disentangling these mechanisms. For example, colonization and competition experiments with strains isolated from different *Apis* species can help identify whether microbes are physiologically adapted to their native host (Mazel et al., 2025). Such studies can reveal to what extent microbial competition within the gut environment limits the establishment of foreign strains. Further insight can be gained by characterizing the structure of ecological niches within the gut and comparing them across host species. How similar are the gut environments of different *Apis* species? To what extent do these differences explain observed microbial specificity? Addressing these questions will require combining detailed characterization of the host gut environments and metabolic characterization of the gut microbial community members.

Beyond adaptation to host-specific niches, neutral processes such as dispersal limitation can also maintain host-specificity. Since honeybee gut microbes are primarily acquired through social transmission, limited contact between host species may restrict opportunities for microbial exchange. This could explain some of the variation, particularly at higher taxonomic levels (Custer et al., 2022; Hildebrand et al., 2021). For example, aerotolerant bacteria were more likely to be transmitted between hosts (Moeller et al., 2018). However, even within a single bacterial genus, different species or strains show varying degrees of host specificity. This subverts expectations that host-specificity may be driven by the dispersal ability of bacteria (Mazel et al., 2024) because traits influencing dispersal ability across host guts, including sporulation, aerotolerance etc., are typically conserved among closely related bacteria of the same species. On the other hand, host behaviour can also influence the dispersal of their microbiota. For example, microbiota composition was associated with social structure in wild baboons and enriched in anaerobic and non-spore forming species (Tung et al., 2015). Further, priority effects (investigated in Chapter 3) could also play a role in ensuring that early arriving strains (acquired from nestmates) dominate the gut, preventing other strains from related host species from colonizing the gut and establishing themselves in this host species. Further discussion of this topic can be found in our recent review (Mazel et al., 2025) on host-specificity in social bees.

Another key finding from our analysis (Chapter 2) was that host-specificity has important functional consequences, particularly in the diversity of carbohydrate-active enzymes (CAZymes) involved in pollen digestion. To further understand the importance of host-specificity, it is essential to examine a broader range of microbial functions relevant to honeybees. Gut microbiota can influence the host in various ways, ranging from offering

protection against pathogens (Steele et al., 2021) to cognitive improvements (Cabirol et al., 2023). However, it remains unclear to what extent such benefits may be provided by the native microbiota of the host species. Honeybees provide a promising model for studying this question because non-native microbes can sometimes colonize different honeybee species, and, by defining native and non-native communities, they can be recreated. Our extensive dataset, along with the collection of gut homogenates (that I collected in Malaysia with the help of our collaborators in a subsequent field trip), offers a valuable resource to experimentally test how host specificity influences the functional benefits provided by gut microbiota.

PROCESSES SHAPING GUT MICROBIAL COMMUNITY ASSEMBLY

Understanding how gut microbial communities assemble and maintain diversity at fine taxonomic scales, such as the strain level, remains a key challenge in microbiome research (Van Rossum et al., 2020). Our findings from *Chapter 3* demonstrate that priority effects, the influence of colonization order and timing on community composition, play an important role in shaping strain-level variation within the honeybee gut microbiota. By sequentially colonizing microbiota-deprived bees with two distinct but species-matched communities, we showed that first-arriving strains consistently dominated, though the strength of this effect varied among species and strains. We found that removing individual early strains partially improved the success of conspecifics arriving later, but also some other species of the same genus, but not others. The extent to which strains affect each other in the context of priority effects by niche pre-emption could depend on their niche overlap. Strains with overlapping niches may compete strongly, but differences in resource use, spatial distribution, or metabolic functions can allow them to coexist. For example, species can co-exist by spatial (Li et al., 2022) or nutritional niche partitioning (Brochet et al., 2021), but the mechanisms that allow strains to co-exist and shape their variation within and across individuals remain to be fully understood. Metabolic niche-mapping of the community, taking into account strain-level differences, could help understand the extent to which conspecific and allo-specific strains affect each other. Finally, longitudinal studies that track microbial composition changes over time within individual hosts can reveal how initial colonization events interact with subsequent eco-evolutionary processes such as selection, drift, and diversification to stabilize or alter community structure upon assembly. These insights can enable the development of accurate models of microbiome community assembly and maintenance.

THE INTERDISCIPLINARY NATURE OF GUT MICROBIOME RESEARCH

Gut microbiome research has been shaped by decades of work across multiple areas, bringing together essential knowledge, perspectives and conceptual frameworks. Microbiomes comprise bacteria, and understanding their functions, traits, and interactions is essential so microbiology is at its core. Such is the interdisciplinary nature of microbiome research that following numerous discoveries and developments over the past two decades, the importance of integrating classical microbiology with microbiome science has been reiterated recently (Radlinski & Bäumler, 2025). Where there are bacteria, there are phages. The study of phages, informed by virology, and their relationship with the microbial

community is a significant area of interest. Ecology contributes theories of community assembly and dynamics, though microbial communities often challenge assumptions made in macroecology, prompting the development of microbial ecology frameworks (Nemergut et al., 2013; Zhou & Ning, 2017). Evolutionary biology and bacterial genetics offer approaches and tools to study microbial diversification and adaptation through phylogenetics, pangenomics, and now metapangenomics. Bioinformatics, powered by metagenomics and other 'omics tools, allows researchers to generate high-dimensional data and track microbial populations and functions with growing precision, often incorporating computational advances from other fields like AI and machine learning. Organismal biology and symbiosis research also inform microbiome science by anchoring microbial studies within the biology of host organisms, including humans. Work in model organisms has expanded our ability to probe microbial mechanisms and establish causality in host-microbe relationships (Douglas, 2019).

As exemplified above (non-exhaustively), gut microbiome research is built on so many concepts that are core to it that no one field of research can claim it for its own. This has been one of the most exciting aspects of working in this area for me. The challenge of working in such a field, though, is to avoid the pitfall of becoming the scientific equivalent of being acquainted with many crafts but a master of none. I think a successful microbiome researcher distinguishes themselves as a "Swiss army knife" equipped with concepts and tools from a range of disciplines that they can apply aptly to the study of gut microbiota. The key to this is integrating ideas through collaboration and conversation with researchers from a variety of areas, which I have had the pleasure of doing throughout my work on the projects outlined in this thesis and beyond.

REFERENCES

- Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., & Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife*, 10, e68583. <https://doi.org/10.7554/eLife.68583>
- Cabirol, A., Schafer, J., Neuschwander, N., Kesner, L., Liberti, J., & Engel, P. (2023). *A defined community of core gut microbiota members promotes cognitive performance in honey bees* (p. 2023.01.03.522593). bioRxiv. <https://doi.org/10.1101/2023.01.03.522593>
- Custer, G. F., Bresciani, L., & Dini-Andreote, F. (2022). Ecological and Evolutionary Implications of Microbial Dispersal. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.855859>
- Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, 17(12), 764–775. <https://doi.org/10.1038/s41579-019-0242-1>
- Hammer, T. J., Sanders, J. G., & Fierer, N. (2019). Not all animals need a microbiome. *FEMS Microbiology Letters*, 366(10), fnz117. <https://doi.org/10.1093/femsle/fnz117>
- Hildebrand, F., Gossmann, T. I., Frioux, C., Özkurt, E., Myers, P. N., Ferretti, P., Kuhn, M., Bahram, M., Nielsen, H. B., & Bork, P. (2021). Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host & Microbe*, 29(7), 1167-1176.e9. <https://doi.org/10.1016/j.chom.2021.05.008>
- Kwong, W. K., Medina, L. A., Koch, H., Sing, K.-W., Soh, E. J. Y., Ascher, J. S., Jaffé, R., & Moran, N. A. (2017). Dynamic microbiome evolution in social bees. *Science Advances*, 3(3), e1600513. <https://doi.org/10.1126/sciadv.1600513>
- Li, Y., Leonard, S. P., Powell, J. E., & Moran, N. A. (2022). Species divergence in gut-restricted bacteria of social bees. *Proceedings of the National Academy of Sciences*, 119(18), e2115013119. <https://doi.org/10.1073/pnas.2115013119>
- Mazel, F., Guisan, A., & Parfrey, L. W. (2024). Transmission mode and dispersal traits correlate with host specificity in mammalian gut microbes. *Molecular Ecology*, 33(1), e16862. <https://doi.org/10.1111/mec.16862>
- Mazel, F., Prasad, A., & Engel, P. (2025). Host specificity of gut microbiota associated with social bees: Patterns and processes. *Microbiology and Molecular Biology Reviews*, 0(0), e00080-23. <https://doi.org/10.1128/mmbr.00080-23>
- Moeller, A. H., Suzuki, T. A., Phifer-Rixey, M., & Nachman, M. W. (2018). Transmission modes of the mammalian gut microbiota. *Science*, 362(6413), 453–457. <https://doi.org/10.1126/science.aat7164>
- Moran, N. A., & Sloan, D. B. (2015). The Hologenome Concept: Helpful or Hollow? *PLOS Biology*, 13(12), e1002311. <https://doi.org/10.1371/journal.pbio.1002311>
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P., & Ferrenberg, S. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews : MMBR*, 77(3), 342–356. <https://doi.org/10.1128/MMBR.00051-12>
- Powell, J. E., Martinson, V. G., Urban-Mead, K., & Moran, N. A. (2014). Routes of Acquisition of the Gut Microbiota of the Honey Bee *Apis mellifera*. *Applied and Environmental Microbiology*, 80(23), 7378–7387. <https://doi.org/10.1128/AEM.01861-14>
- Radlinski, L. C., & Bäumler, A. J. (2025). Microbiome science needs more microbiologists. *Nature Microbiology*, 10(2), 263–264. <https://doi.org/10.1038/s41564-024-01922-4>
- Rühlemann, M. C., Bang, C., Gogarten, J. F., Hermes, B. M., Groussin, M., Waschyna, S., Poyet, M., Ulrich, M., Akoua-Koffi, C., Deschner, T., Muyembe-Tamfum, J. J., Robbins, M. M., Surbeck, M., Wittig, R. M., Zuberbühler, K., Baines, J. F., Leendertz, F. H., & Franke, A. (2024). Functional host-specific adaptation of the intestinal microbiome in hominids. *Nature Communications*, 15(1), Article 1. <https://doi.org/10.1038/s41467-023-44636-7>

Sanders, J. G., Sprockett, D. D., Li, Y., Mjungu, D., Lonsdorf, E. V., Ndjango, J.-B. N., Georgiev, A. V., Hart, J. A., Sanz, C. M., Morgan, D. B., Peeters, M., Hahn, B. H., & Moeller, A. H. (2023). Widespread extinctions of co-diversified primate gut bacterial symbionts from humans. *Nature Microbiology*, 8(6), Article 6. <https://doi.org/10.1038/s41564-023-01388-w>

Sarton-Lohéac, G., Nunes da Silva, C. G., Mazel, F., Baud, G., de Bakker, V., Das, S., El Chazli, Y., Ellegaard, K., Garcia-Garcera, M., Glover, N., Liberti, J., Nacif Marçal, L., Prasad, A., Somerville, V., SAGE class 2019-2020 and 2020-2021, Bonilla-Rosso, G., & Engel, P. (2023). Deep Divergence and Genomic Diversification of Gut Symbionts of Neotropical Stingless Bees. *mBio*, 0(0), e03538-22. <https://doi.org/10.1128/mbio.03538-22>

Steele, M. I., Motta, E. V. S., Gattu, T., Martinez, D., & Moran, N. A. (2021). The Gut Microbiota Protects Bees from Invasion by a Bacterial Pathogen. *Microbiology Spectrum*, 9(2), e00394-21. <https://doi.org/10.1128/Spectrum.00394-21>

Suzuki, T. A., Fitzstevens, L., Schmidt, V. T., Enav, H., Huus, K., Mbong, M., Adegbite, B. R., Zinsou, J. F., Esen, M., Velavan, P., Adegnika, A. A., Song, L. H., Spector, T. D., Marchi, N., Blekhman, R., Ségurel, L., Youngblut, N. D., Kremsner, P., & Ley, R. E. (2021). Codiversification of gut microbiota with humans.

Tung, J., Barreiro, L. B., Burns, M. B., Grenier, J.-C., Lynch, J., Grieneisen, L. E., Altmann, J., Alberts, S. C., Blekhman, R., & Archie, E. A. (2015). Social networks predict gut microbiome composition in wild baboons. *eLife*, 4, e05224. <https://doi.org/10.7554/eLife.05224>

Van Rossum, T., Ferretti, P., Maistrenko, O. M., & Bork, P. (2020). Diversity within species: Interpreting strains in microbiomes. *Nature Reviews Microbiology*, 18(9), 491–506. <https://doi.org/10.1038/s41579-020-0368-1>

Zhou, J., & Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiology and Molecular Biology Reviews*, 81(4), 10.1128/mmbr.00002-17. <https://doi.org/10.1128/mmbr.00002-17>

CURRICULUM VITAE

AISWARYA PRASAD

Lausanne, Switzerland | aiswarya.prasad@unil.ch

EDUCATION

<i>Ph.D. in Quantitative Biology</i> University of Lausanne, Switzerland Thesis title: Gut microbiome ecology and evolution in honeybees (Supervisor: Prof. Philipp Engel)	2020 – Present
<i>BS and MS in Biology (4+1 yrs)</i> Indian Institute of Science, Bangalore Master thesis title: Analysis of the human gut microbiome by nanopore-based metagenome sequencing (Supervisor: Prof. Utpal Tatu)	2015 – 2020

FELLOWSHIPS AND AWARDS

<i>Faculty of Biology and Medicine (FBM) PhD Fellowship</i> International PhD fellowship award granted by the Faculty of Biology and Medicine (FBM), University of Lausanne	2020 – 2023
<i>Mathilde Agassiz scholarship</i> A scholarship of CHF 70k awarded to one chosen candidate recommended by the FBM.	2020 – 2021
<i>KVPY Fellowship</i> Stipend and contingency to fund BSc and MS awarded by the Indian government (All-India Rank: 335)	2015 – 2020
<i>Bronze Medal in iGEM Competition</i> Boston, USA Co-founded the first team from the institute and contributed to securing 3 grants for research and travel from the Wellcome Trust Indian Alliance, the Govt. of India, DBT and the State government of Karnataka.	2016

PUBLICATIONS

Mazel, F., [Prasad, A.](#), & Engel, P. (2025). Host specificity of gut microbiota associated with social bees: patterns and processes. [Microbiology and Molecular Biology Reviews](#), e00080-23.

Somerville, V., [Prasad, A.](#), Maurice, C. F., Garrido-Sanz, D., Ulrich, E., Mazel, F., ... & Engel, P. (2025). Strain-level and phenotypic stability contrasts with plasmid and phage variability in water kefir communities. [bioRxiv](#), 2025-02. (In Review)

Prasad, A., Pallujam, A. D., et al (2024). *Symbiont loss and gain, rather than co-diversification shapes honeybee gut microbiota diversity and function*. [bioRxiv](#), 2024.09.11.612390 (In Review)

Baud, G. L., Prasad, A., Ellegaard, K. M., & Engel, P. (2023). *Turnover of strain-level diversity modulates functional traits in the honeybee gut microbiome between nurses and foragers*. [Genome Biology](#), 24(1), 283.

Sarton-Lohéac, ... , Prasad, A., Sommerville, V., SAGE class 2019-2020 and 2020-2021, Bonilla-Rosso, G., & Engel, P. (2023). *Deep divergence and genomic diversification of gut symbionts of neotropical stingless bees*. [Mbio](#), 14(2), e03538-22.

Beal, J., Haddock-Angelli, T., Baldwin, G., Gershater, M., Dwijayanti, A., Storch, M., De Mora, K., Lizarazo, M., Rettberg, R. and with the iGEM Interlab Study Contributors, 2018. *Quantification of bacterial fluorescence using independent calibrants*. [PloS one](#), 13(6), p.e0199432.

INVITED TALKS

<i>Microbiome Virtual International Forum (Best Selected Talk award)</i> Online (video link)	Nov 2024
<i>The Local Pangenome</i> Alicante, Spain	Oct 2023
<i>Academic Seminar Series</i> School of Science, Monash University Malaysia	Apr 2023

CONFERENCE PARTICIPATION

<i>Gordon Research Conference and Seminar (Animal Microbe Symbiosis)</i> Lucca, Italy Poster: Evolutionary and functional insight into gut microbial community diversity across honeybee species using metagenomics	Jun 2023
<i>CSHL Meeting: Microbiome</i> Cold Spring Harbor Laboratories, USA Poster: Whole-genome shotgun metagenomics reveals differences in gut microbiomes among honeybee species	Oct 2023
<i>SymbNET PhD Summer School on Host-Microbe Symbiosis</i> Instituto Gulbenkian de Ciência, Lisbon	Jul 2022

Poster: Understanding the evolution of the gut microbiome across honeybee species

PROFESSIONAL SERVICE

<i>Peer review of journal publications</i> <i>Ecological Monographs (1) and Insect Molecular Biology (1)</i>	2023 – Present
<i>Workshop Evaluation Committee</i> <i>Workshop and tutorial selection for the BC2 (Basel Computational Biology Conference)</i>	Jan 2023
<i>Local volunteer</i> <i>ISME-18 (International Society of Microbial Ecology) conference</i>	Aug 2022

DEPARTMENT SERVICE

<i>Organizing committee</i> Annual (DMF) Department Science Retreat	Sep 2023
<i>DMF Faculty Hiring Committee</i> Postdoc/PhD Representative	May 2023

OUTREACH AND EXTRACURRICULAR ENGAGEMENT

Nucleate Switzerland Co-Director of strategy at the regional leadership of Nucleate, a student-led organization dedicated to accelerating impactful academic bioventures.	Oct 2024 – Present
The Consulting Society, EPFL Senior member and Consulting project manager for a microbiology diagnostics startup	Mar 2024 – Present
<u>Contributor</u> - Methods in Microbiomics Set of guidelines and best practices for robust and reproducible pipelines	Jan 2024
Open Science Meeting National center for competence in Research (NCCR) Microbiomes, Switzerland	Nov 2023
Decolonizing North-South Research Collaboration Centre for Development and Environment, University of Bern	Aug 2022
<u>Moderator</u> – Health and Data Privacy Committee iGEM India BioSummit (Online)	Jul 2020

Updated, May 2025

Delegate at United Nations Convention on Biological Diversity
Sharm El-Sheikh, Egypt

Nov 2018

TEACHING

Teaching Assistant – SAGE (Sequence A GEnome) course
Masters-level course teaching bioinformatics through research

2021 - 2024