UNIL | Université de Lausanne
Faculté de biologie et de médecine

Département de Microbiologie Fondamentale

Mechanisms underlying the evolution and coexistence of closely related gut bacteria in the honey bee, *Apis mellifera*

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

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

par

Silvia BROCHET

Master en biologie, orientation Microbiologie,
Université de Lausanne, Lausanne, Suisse

Jury

Prof. John Pannell, Président
Prof. Philipp Engel, Directeur de thèse
Prof. Sara Mitri, Experte
Prof. Martin Ackermann, Expert

Lausanne
(2021)

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	John	**Pannell**
Directeur·trice de thèse	Monsieur	Prof.	Philipp	**Engel**
Expert·e·s	Madame	Prof.	Sara	**Mitri**
	Monsieur	Prof.	Martin	**Ackermann**

le Conseil de Faculté autorise l'impression de la thèse de

Silvia Brochet

Maîtrise universitaire ès Sciences en sciences moléculaires du vivant, Université de Lausanne

intitulée

Mechanisms underlying the evolution and coexistence of closely related gut bacteria in the honey bee, *Apis mellifera*

Lausanne, le 18 mars 2022

pour le Doyen
de la Faculté de biologie et de médecine

Prof. John Pannell

Acknowledgments

First of all I would like to thank **Philipp**. You thought me so much during these five years! Thank you for this, for always being available to discuss about my experiments and my writing, for transmitting to me your infinite passion about science and ultimately for being my friend.

I would like to thank the members of my thesis committee (Prof. **John Pannell**, Prof. **Sara Mitri** and Prof. **Martin Ackermann**) for taking the time for reading my thesis (and my mid-thesis!) and discussing it with me.

I had to chance to do my PhD in a group full of amazing people to work (and party!) with. I need to thank deeply **Germán** (who thought me how cool ecology is), **Sudip** (who, among other things, thought me how to extract bacterial RNA of the finest quality) and **Andrew** (who thought me about the intriguing world of metabolomics). A big thanks goes to the other lab-members as well: **Joanito**, **Amélie**, **Silvi**, **Yassine**, **Gilles**, **Vincent**, **Garance**, **Aiswarya**, **Malick**, **Nicolas** (without whom there would be no metabolomics in this thesis), **Lucie** (Lucinka!), **CG**, **Kirsten**, **Konstantin**, **Olivier** and my dear squirrel-friend **Théo**. Thanks also to the stagiaires and master students that I worked with: **Pauline**, **Simon** and **Katharina**.

I need to thank all the technicians that work in the department (and that make the department work!), in particular **Nazife**, who was always there for me.

I also want to thank all the members of the **Mitri-group** for being always so nice with me and for the insightful scientific discussion that we had during these years.

In the department I found amazing people to drink a tea in the afternoon and beers (or prosecco…) in the evening: thanks to **Ania**, **Paddy**, **Dimi**, **Marian**, **Roberto** and **Björn** for being such good friends to me. To my Italian-speaking-department-friends I have to say a huge "*grazie*"! Grazie to **Sam**, **Anto**, **Philippe** and in particular to **Rita**: you are one of the best persons I know, without whom I would not be where I am today (end of the PhD!) and you became one of my best friends in the world.

I was always supported also from my friends that are on the other side of the Swiss border (north and south!), thank you. A special thanks goes to **Billa** (Mona) and also to **Sofa**, my best friend since 28 years (since I have memories).

I want to thank all of my family for the endless support and inspiration. My parents for blindly believing in me and for their infinite love. Grazie **mamma**. Grazie **papà**. My brother for being the first triggering in me the passion for science and for listening always with interest to my little discoveries, grazie **Michele**.

I also want to thank the **Clément family** for being my family in Switzerland.

Last but not least, **Mathieu**, thanks for believing in me and supporting me more than anyone, for taking my hand and helping me to climb also to the top of this mountain.

Table of Contents

Résumé pour le grand public

Les mécanismes sous-jacents à l'évolution et à la coexistence des bactéries étroitement apparentées de l'intestin de l'abeille domestique, *Apis mellifera*

Silvia Brochet, Département de Microbiologie Fondamentale

On vit dans un monde dominé par des êtres minuscules : les *bactéries*. Ces microorganismes peuvent vivre dans n'importe quel environnement: du sol gelé en permanence des régions arctiques jusqu'aux lacs acides situés dans les cratères de volcans actifs. Mais cela n'est pas nécessaire d'aller si loin: les bactéries vivent aussi dans nos intestins, où elles constituent le microbiote intestinal. Ici, les bactéries sont organisées en communautés de plusieurs espèces parmi lesquelles peuvent coexister aussi bien des bactéries extrêmement similaires que des complétement différentes. Le principe de l'exclusion compétitive stipule que plus deux espèces sont similaires, moins elles ont de chances de coexister dans le même environnement. Ceci est dû au fait qu'elles ont des exigences similaires et se disputent l'espace et les nutriments jusqu'à qu'une des deux exclue l'autre. Mais comment les espèces étroitement liées du microbiote intestinal peuvent coexister? Et est-ce-que ces espèces, si similaires, ont toutes la même fonction?

Pour répondre à ces questions, dans cette thèse j'ai utilisé comme modèle des bactéries du microbiote intestinal des abeilles mellifères qu'on appelle *Lactobacillus* Firm5. Différentes variétés de Firm5 existent: ces bactéries se divisent en quatre espèces qui se divisent en différentes souches. En réalisant des expériences durant lesquelles j'ai colonisé des abeilles sans microbiote avec des souches de différentes espèces, j'ai démontré que celles-ci peuvent coexister grâce au partage des différents composants du pollen, la principale source alimentaire des abeilles, évitant ainsi l'exclusion compétitive. Au contraire, j'ai observé qu'il y a moins de chance que les souches de la même espèce coexistent. Pourtant, certaines souches de la même espèce peuvent coexister, ce qui indique que la coexistence de différentes souches pourrait être spécifique à la souche et non à l'espèce à laquelle la souche est associée. Enfin, j'ai observé que différentes espèces et souches de Firm5 peuvent avoir des fonctions différentes, ce qui pourrait avoir un impact sur les autres membres de la communauté et sur l'hôte.

En résumé, dans cette thèse j'ai souligné l'importance d'étudier les bactéries étroitement liées dans les communautés bactériennes naturelles, car en dépit d'être très similaires, elles peuvent toujours se comporter très différemment. En étudiant de cette façon les communautés bactériennes naturelles, on obtient des connaissances qui peuvent nous aider à mieux comprendre leurs fonctions dans le contexte de l'hôte et éventuellement des autres environnements naturels qu'elles occupent.

Summary

Bacterial communities colonize nearly all environments on our planet. One of the most remarkable feature of many of these communities is that they are highly diverse, and that closely related species and strains of the same species often co-exist. How such diversity has emerged and is maintained in natural microbial communities, and to what extent interactions at the species-level differ from those at the strain-level has largely remained elusive. In this thesis, I addressed these questions focusing on the honey bee gut symbiont *Lactobacillus* Firm5. This bacterial clade is one of the most abundant and widely distributed gut symbiont of social bees and has diverged into four species which coexist in honey bees and harbor a large extent of strain-level diversity. The experimental tractability of the bee system allowed me to investigate interactions of *Lactobacillus* Firm5 *in vitro* and *in vivo* at the species- and strain-level and to identify causes and consequences of bacterial coexistence.

In the first chapter (**Chapter I**) we investigated whether the coexistence of the four *Lactobacillus* Firm5 species in the honey bee gut is dependent on the diet of the host. We selected one strain for each species and passaged them *in vivo* and *in vitro* under two different nutrient conditions, i.e. in the presence of simple sugars only or in the presence of pollen. We found that the four species were able to coexist only in the presence of pollen. Metatranscriptomics and metabolomics analysis showed that the four species utilize distinct carbohydrates present in pollen suggesting that resource partitioning of the bee diet facilitates the coexistence of the four related Firm5 species.

In **Chapter II** we turned the focus towards the question of whether there are differences in interaction and coexistence between strains that belong to the same (conspecific) or a different (allospecific) species. We hypothesized that conspecific strains of *Lactobacillus* Firm5 would engage in stronger negative interactions than allospecific strains and be less likely to coexist when compared to allospecific strains. To test this, we colonized microbiota-depleted bees with all possible pairwise combinations of twelve strains (three strains of each species). The vast majority of strains mutually inhibited each other independent of the species affiliation. However, conspecific strains had less symmetrical abundances than allospecific strains in these two-member communities. Moreover, serial passaging of a subset of these communities through gnotobiotic bees revealed that allospecific pairs, were more likely to stably coexist than conspecific pairs. However, this was not the case for all pairs highlighting that sometimes strain-specific features and not species identity determine coexistence. These results reproduce the co-occurrence patterns observed in the microbiota of honey bees in nature, and suggest that the divergence of Firm5 into the four different species was driven by adaptation to different ecological niches provided by the pollen diet in the honey bee gut.

Finally, in **Chapter III** we addressed the question whether the presence of four different *Lactobacillus* Firm5 species and divergent strains of these species has any functional consequences. We used the same strains as studied in Chapter II to colonize microbiota-depleted (MD) bees, and assessed the production of different short chain fatty acids (SCFAs) in the gut and the hemolymph of the host. While we did not detect any accumulation of SCFAs in colonized (CL) vs MD bees in the hemolymph, five out of the six detected SCFAs had elevated levels in CL vs MD bees in the gut. While some SCFAs (acetate and succinate) were accumulated by all strains, others (lactate, butyrate and formate) were produced in a species- or strain-specific way. These results indicated that different species and strains of *Lactobacillus* Firm5 produce different metabolites in the gut. As SCFAs are key metabolites for the interaction with other bacteria and with the host, our results suggest that the presence of different species and strains of *Lactobacillus* Firm5 has functional consequences for the bee gut microbiota and the host.

In summary, our results show that resource partitioning of dietary nutrients plays a key role in facilitating the coexistence of closely related bacterial species in the animal gut. We find that sometimes strain identity rather than species identity determines the outcome of bacterial interactions in the gut. Moreover, closely related species and different strains within these species can generate different fermentation products. This is likely to be relevant for the interactions with bacterial communities and the impact on the host. Therefore, our findings highlight the importance of studying natural bacterial communities at the level of individual species and strains.

Résumé

Les communautés bactériennes colonisent presque tous les environnements de notre planète. L'une des caractéristiques les plus remarquables de la plupart de ces communautés est qu'elles sont très diverses et que des espèces et des souches de la même espèce étroitement apparentées coexistent souvent. Comment une telle diversité a-t-elle émergé et s'est maintenue dans les communautés microbiennes naturelles, et dans quelle mesure les interactions au niveau de l'espèce diffèrent-elle de celles au niveau de la souche sont des questions restées largement insaisissables. Dans cette thèse, j'ai abordé ces questions en me concentrant sur le symbiote intestinal de l'abeille mellifère *Lactobacillus* Firm5. Firm5 est l'un des symbiotes intestinaux les plus abondants et les plus largement distribués des abeilles sociales. Il a divergé en quatre espèces qui coexistent chez les abeilles mellifères et abritent une grande diversité au niveau des souches. La traçabilité expérimentale de l'abeille comme organisme model m'a permis d'étudier les interactions de *Lactobacillus* Firm5 *in vitro* et *in vivo* aussi bien au niveau de l'espèce que de la souche et également d'identifier les causes et les conséquences de la coexistence bactérienne.

Dans le premier chapitre (**Chapitre I**), nous avons examiné si la coexistence des quatre espèces de *Lactobacillus* Firm5 dans l'intestin de l'abeille dépendait du régime alimentaire de l'hôte. Nous avons sélectionné une souche pour chaque espèce et les avons passées *in vivo* et *in vitro* dans deux conditions nutritionnelles différentes, c'est-à-dire en présence uniquement de sucres simples ou en présence de pollen. Nous avons constaté que les quatre espèces ne pouvaient coexister qu'en présence de pollen. L'analyse métatranscriptomique et métabolomique a montré que les quatre espèces utilisent des glucides distincts présents dans le pollen, ce qui suggère que la répartition des ressources du régime alimentaire des abeilles facilite la coexistence des quatre espèces Firm5 apparentées.

Dans le **Chapitre II**, nous nous sommes concentrés sur la question de savoir s'il existe des différences d'interaction et de coexistence entre des souches appartenant à la même espèce (conspécifique) ou à une espèce différente (allospécifique). Nous avons émis l'hypothèse que les souches conspécifiques de *Lactobacillus* Firm5 s'engageraient dans des interactions négatives plus fortes que les souches allospécifiques et seraient moins susceptibles de coexister par rapport aux souches allospécifiques. Pour tester cela, nous avons colonisé des abeilles appauvries en microbiote avec toutes les combinaisons possibles par paires de douze souches (trois souches de chaque espèce). La grande majorité des souches s'inhibaient mutuellement indépendamment de l'affiliation à l'espèce. Cependant, les souches conspécifiques avaient des abondances moins uniformes que les souches allospécifiques dans ces communautés à deux membres. De plus, le passage en série d'un sous-ensemble de ces communautés à travers des abeilles

gnotobiotiques a révélé que les paires allospécifiques étaient plus susceptibles de coexister de manière stable que les paires conspécifiques. Cependant, ce n'était pas le cas pour toutes les paires, soulignant que parfois des caractéristiques spécifiques à la souche et non l'identité de l'espèce déterminent la coexistence. Ces résultats reproduisent les modèles de cooccurrence observés dans le microbiote des abeilles mellifères dans la nature et suggèrent que la divergence de Firm5 dans les quatre espèces différentes est due à l'adaptation à différentes niches écologiques fournies par le régime pollinique dans l'intestin des abeilles mellifères.

Enfin, dans le **Chapitre III**, nous avons abordé la question de savoir si la présence de quatre espèces différentes de *Lactobacillus* Firm5 et de souches divergentes de ces espèces a des conséquences fonctionnelles. Nous avons utilisé les mêmes souches que celles étudiées au chapitre II pour coloniser des abeilles appauvries en microbiote (MD) et évalué la production de différents acides gras à chaîne courte (AGCC) dans l'intestin et l'hémolymphe de l'hôte. Bien que nous n'ayons détecté aucune accumulation d'AGCC chez les abeilles colonisées (CL) vs MD dans l'hémolymphe, cinq des six AGCC détectés présentaient des niveaux élevés chez les abeilles CL vs MD dans l'intestin. Alors que certains AGCC (acétate et succinate) ont été accumulés par toutes les souches, d'autres (lactate, butyrate et formiate) ont été produits d'une manière spécifique à l'espèce ou à la souche. Ces résultats indiquent que différentes espèces et souches de *Lactobacillus* Firm5 produisent différents métabolites dans l'intestin. Comme les AGCC sont des métabolites clés pour l'interaction avec d'autres bactéries et avec l'hôte, nos résultats suggèrent que la présence de différentes espèces et souches de *Lactobacillus* Firm5 a des conséquences fonctionnelles pour le microbiote intestinal de l'abeille et l'hôte.

En résumé, nos résultats montrent que la répartition des ressources des nutriments alimentaires joue un rôle clé en facilitant la coexistence d'espèces bactériennes étroitement apparentées dans l'intestin des animaux. Nous constatons que parfois l'identité de la souche plutôt que l'identité de l'espèce détermine le résultat des interactions bactériennes dans l'intestin. De plus, des espèces étroitement apparentées et différentes souches au sein de ces espèces peuvent générer différents produits de fermentation. Ceci est susceptible d'être pertinent pour les interactions avec les communautés bactériennes et l'impact sur l'hôte. Par conséquent, nos résultats soulignent l'importance d'étudier les communautés bactériennes naturelles au niveau des espèces et des souches individuelles.

Preface

Why are there so many kinds of bacteria?

"Our planet, the Earth, is, as far as we know, unique in the universe. It contains life. Even in its most barren stretches, there are animals. Around the equator, where those two essentials for life, sunshine and moisture, are most abundant, great forests grow. And here plants and animals proliferate in such numbers that we still have not even named all the different species. Here, animals and plants, insects and birds, mammals and man live together in intimate and complex communities, each dependent on one another. Two thirds of the surface of this unique planet are covered by water, and it was here indeed that life began. From the oceans, it has spread even to the summits of the highest mountains as animals and plants have responded to the changing face of the Earth."

Sir David Attenborough (Opening narration of 'The Living Planet', 1984)

These beautiful words that Sir David Attenborough uses to describe the world of animals and plants could be used to describe the world of bacteria as well. Bacteria are unicellular microorganisms that colonize nearly all environments on our planet. Not only can they resist to extreme conditions such as high temperatures or pressure, but they also colonize all kinds of animals and plants, often providing beneficial services to their eukaryotic hosts. Bacterial life, as for the life that Sir Attenborough refers to, is extremely diverse. A large number of different phyla, families, species, and strains can simultaneously thrive in the same environment with emerging properties that are key for ecosystem functions.

But why are there so many kinds of bacteria? In his article "Homage to Santa Rosalia *or* why are there so many kinds of animals?", the ecologist George Evelyn Hutchinson asked himself a very similar question (Hutchinson 1959). Hutchinson was in Sicily, visiting the sanctuary of Santa Rosalia (Palermo), looking for a pond to observe some aquatic insects of the *Corixidae* family. He eventually found two different species of these insects and noticed that their breeding periods were offset: while one species was concluding the breeding season (as there were only females present) the other species was

starting it (as there was an equal mix of males and females). These observations inspired him to ask some questions:

> "It was not until I asked myself why the larger species should breed first, and then the more general question as to why there should be two and not 20 or 200 species of the genus in the pond, that ideas suitable to present to you began to emerge. These ideas finally prompted the very general question as to why there are such an enormous number of animal species."
>
> George Evelyn Hutchinson ("Homage to Santa Rosalia or why are there so many kinds of animals?" The American Naturalist, 1959)

To address this general question Hutchinson introduced the revolutionary concept of *ecological niche* as an "n-dimensional hypervolume", which includes the resources (e.g. space and nutrients) that a species requires to persist within an environment (Hutchinson 1957). His species-specific and quantitative definition led to the development of several niche-related theories such as the *niche partitioning* concept, i.e. the idea that coexisting species have different niche-requirements (Chase and Leibold 2009).

Some say that this article has provoked most of the ecological research on biodiversity in the last half a century. Certainly this paper still inspires generations of ecologists in search of the causes of diversity (including myself!). Indeed, the questions that Hutchinson asked himself 60 years ago are not yet fully answered at the present day.

In my PhD work, I used the honey bee gut microbiota to explore some of these questions. With this model I wanted to investigate why there are so many kinds of bacteria within natural bacterial communities. In particular, I wanted look for answers to questions such as: Which are the factors that facilitate the coexistence of several different kinds of bacteria within the same environment? Which interactions dominate within a community and do they depend on how related the different kinds of bacteria are? Is there a correspondence between how many kinds of bacteria there are and the amount of functions of a bacterial community?

Introduction

Diversity in natural bacterial communities

The structure of diversity

Diversity is one of the intrinsic properties of bacterial communities. The word 'diversity' literally means "a different range of things or people" (Cambridge English Dictionary). When microbiologists use the term diversity, they are referring to the number and types of microbial variants that can be found within a community. There are three main levels at which diversity can be measured or three main 'types' of diversity: *alpha* diversity (the diversity among species within one community), *beta* diversity (the diversity within the range of several communities or ecosystems) and *gamma* diversity (the diversity in the range of all communities) (Whittaker 1972). This thesis mostly focuses on alpha diversity, which measures diversity within a community and how individual species divide resources. In contrast, beta and gamma diversity measure diversity across communities (Sepkoski 1988).

The alpha diversity of a community is described by two main types of indexes: i) *species richness* (which measures the species content of the community, i.e. the number of species) and ii) *equitability* (which measures the evenness of species distributions within the community, i.e. the relative abundances of different species) (Thukral 2017). In particular, to assess the richness of a community, the phylogenetic diversity between individuals is measured and used to classify them in *taxonomic hierarchies* that start from the highest rank (i.e. *phylum*), passing by intermediate ranks (i.e. *family,* and *genus*) to the lowest ranks (i.e. *species* and *strains*). The definition of species, while still controversial to some extent (Caro-Quintero and Konstantinidis 2012; Doolittle 2012), is 'group of individuals who form a coherent genomics cluster' (Van Rossum et al. 2020).

At the same time, despite the genetic similarity, within species additional variance is found at the level of individual strains, for which there is not an universal definition yet. Strains are typically delimited by tens of single nucleotide variants (SNVs), even if in theory genomes with as few as one SNV difference could be referred to as different strains (Van Rossum et al. 2020). The lower the rank, the higher the number of features that are shared between the individuals of that rank (Philippot et al. 2010). Richness can be assessed by counting the number of variants per taxonomic rank.

The amount of diversity-richness that can be measured can greatly vary between different bacterial communities. In fact, depending on the rank that is used to describe a bacterial community, the amount of diversity that is detected can differ. For example, while soil-associated bacterial communities are very diverse at the deep-phylogenetic level in that they consist of more than 20 different bacterial phyla (Dunbar et al. 1999), the communities residing in the gut of animals (mammals, birds, amphibians, reptiles, fishes and invertebrates including insects) are typically dominated by only two different phyla (i.e. Firmicutes and Bacteroidetes) (Bäckhed et al. 2005; Eckburg et al. 2005). Instead, communities in the animal gut contain a lot of diversity at the shallow phylogenetic level, i.e. at the level of different genera and species, resulting a 'fanlike' structure (Bäckhed et al. 2005; Eckburg et al. 2005; Lozupone et al. 2012). These two patterns of diversity in microbial communities can be pictured like trees that can be found in nature with different types of branching (Dethlefsen *et al.*, 2007), as illustrated in **Figure 1**. The diversity in communities found in soil is represented by a tree whose branches arise at all distances from the root (**Figure 1**), while in gut-associated communities diversity is represented by a palm tree with few branches arising close to the root and with many branches arising close to the branch tips (**Figure 1**).

Figure 1: Trees with different types of branching representing different patterns of microbial diversity. On the left: tree representing microbial diversity in soil-associated communities: new branches arise at all distances from the root indicating high amounts of diversity at the level of distantly related lineages (phyla). On the right: tree representing microbial diversity in gut-associated communities: few branches arise close to the root and many arise close to the branch tips indicating low amounts of deep-diversity and high amounts of diversity at the shallow level. Adapted from Dethlefsen et al., 2007 and Ellegaard et al., 2019.

Bacterial communities frequently undergo temporal changes in taxonomic composition and hence in diversity. From when the first few founders or pioneering individuals colonize a new niche to the full establishment of the community, the amount of diversity present can dramatically shift. One example is the assembly of the human gut microbiota. It is known that the infant gut is at first colonized by few founder species, to which other species will add up over time and who will experience changes in their distribution within the community (La Rosa et al. 2014). In fact, the community composition changes gradually over time as the phylogenetic diversity increases while at the same time major taxonomic groups experience drastic shifts in abundance corresponding to changes in diet or health (Koenig et al. 2011). In particular, it was shown that there are age-specific community types succeeding in the infant gut until around 5 years of age, when the gut microbiota acquires an adult-like configuration but still lacks the richness typical of the adult gut microbiota (Roswall et al. 2021). Moreover, the diversity of gut-associated communities can also fluctuate during adult life, as it was

shown for example for the Hazda people from Tanzania, whose microbiome changes seasonally with some taxa becoming undetectable due to radical dietary shifts linked to the seasonal availability of different food sources (Schnorr et al. 2017). Another example of bacterial communities whose diversity fluctuates depending on season are marine bacterial communities, such as the ones found in the Western English channel, which experience strong seasonal patterns of species diversity and richness (Caporaso et al. 2012), as well as soil bacterial communities, especially within soil microaggregates (Upton, Bach, and Hofmockel 2019).

How to assess diversity in bacterial communities

Before DNA sequencing techniques became available, microbiologists were assessing diversity in bacterial communities by distinguishing bacteria based on their morphology (e.g. cells shape and size) and their physiology (e.g. relationship to oxygen, light and temperature) (Atlas 1984). From the very beginning, when for the first time in history Robert Hooke looked into a microscope and observed microorganisms around 300 years ago, microbiologists developed several methods to define different groups of bacteria based on this kind of characteristics, some of which are still used today. However, the majority of these methods require the cultivability of bacteria, which is not always possible, and they are often not enough accurate. A more accurate alternative to compare and classify bacteria is to sequence bacterial DNA to look at differences in the genetic material of the isolates. The DNA sequencing methods that are currently used to measure diversity in bacterial communities are mainly two: *16S rRNA gene sequencing* and *shotgun metagenomics sequencing* (**Figure 2**).

The 16S rRNA gene was used for the first time to classify bacteria in 1977 by Carl Woese, who defined it a "*molecular clock*". Since this gene is universally conserved among bacteria in that it experiences slow rates of evolution, it allows to infer phylogenetic relationships between bacteria and ultimately to classify them

taxonomically. With this method, DNA is extracted from the sample and one or more hypervariable regions of the 16S rRNA gene are amplified by PCR using nearly universal primers and sequenced (**Figure 2.A**). One of the first methods that was developed to analyze 16S rRNA data was based on clustering similar sequences in *Operational Taxonomic Units* (OTUs) using a 97% sequence-identity threshold, corresponding to the genus/species-level of diversity. Still, this method carries the risk of grouping multiple similar variants into a single OTU. A more recently-developed approach allows to group sequences into more fine-grained units of diversity referred to as ASVs or ESVs (*Amplicon Sequence Variants or Exact Sequence variants*) (Callahan, McMurdie, and Holmes 2017). This approach only clusters identical sequences into units (the ASVs/ESVs). This is possible thanks to a sequencing run-specific error model that is built during the analysis and which results in the correction and/or exclusion of low-abundant sequences suffering from sequencing errors (Callahan, McMurdie, and Holmes 2017). However, even with the ASV method, it needs to be noted that 16S rRNA gene based analyses have limitations for categorizing diversity in natural microbial communities (Ellegaard and Engel 2016). The same divergence in the *16S rRNA* gene sequence can correspond to different levels of divergence in the rest of the genome across different bacterial lineages. Moreover, even bacteria identical in the *16S rRNA* gene can harbor large differences in the accessory genome, which is often crucial for adaptation to specific environments (Biller et al. 2015; Ellegaard and Engel 2016, 2019; Touchon et al. 2009). While it is difficult to relate ASVs to a certain taxonomic level, this type of approach should be appropriate to resolve communities down to the genus- but not to the species- and strain-level (**Figure 2.A**).

Less conserved genes than the 16S rRNA have been targeted in amplicon sequencing to assess diversity in communities beyond the species-level (Ogier et al. 2019; Poirier et al. 2018; Wright, Erik S. 2016). These approaches are limited in the taxonomic range of the primers used to amplify the gene of interest if compared to the universal

primers targeting the 16S rRNA gene, but they provide more resolution. In fact, using primers targeting selected marker genes allows to discriminate between strains of the same species. For example, Wright et al. used an 80 nucleotide long sequence region of the *rpoB* gene to be able to successfully discriminate between 18 different strains of *Streptomyces* (Wright, Erik S. 2016).

In contrast to amplicon sequencing approaches, in shotgun metagenomics the entire gene content in natural microbial populations is sequenced (Ellegaard and Engel 2019; Engel, Martinson, and Moran 2012; Qin et al. 2012) (**Figure 2.B**). With this method, whole genomic DNA is extracted from a sample and directly subjected to sequencing without prior pre-amplification of a specific gene of interest. The entire genetic content is subsequently sequenced and *de novo* assembled into so-called MAGs (Metagenome Assembled Genomes) or mapped to a reference database for analysis (**Figure 2.B**). The entire gene content can be considered to place genomes into phylogenetic context or to assess diversity based on the accessory gene content or SNVs in the core genome. These approaches allow to discriminate bacteria to a deeper level of diversity, beyond the genus level, i.e. until the '*species'* and the '*strain'* level (**Figure 2.B**). Metagenomics also allows to assess the functional properties of a community, as functions can be predicted for the entire gene content of the sample (**Figure 2.B**). The drawback of shotgun metagenomics is that this method is relatively expensive, needs more input material, and may not allow to detect very rare species as the required sequencing depth cannot be achieved. Moreover, the analysis tools that are available for metagenomics data are not yet as standardized and ready-to-use as the ones for 16S rRNA data. However, as the sequencing costs have gone down, sequence library preparations protocols improved and new powerful analytical tools have been developed, shotgun metagenomics becomes more and more popular in recent years. One of the first studies using shotgun metagenomics sequencing was published in 2004, when Venter *et al.* analyzed seawater samples and identified 148 previously unknown bacterial genera,

1800 different bacterial species and over 1.2 million previously unknown genes (Venter et al. 2004). Metagenomics revealed that the diversity measured using the 16S rRNA method represents just the tip of the iceberg of the total diversity that is actually present in natural bacterial communities. Consequently, it is possible that at one point shotgun metagenomics approaches will to a large part replace amplicon sequencing methods.

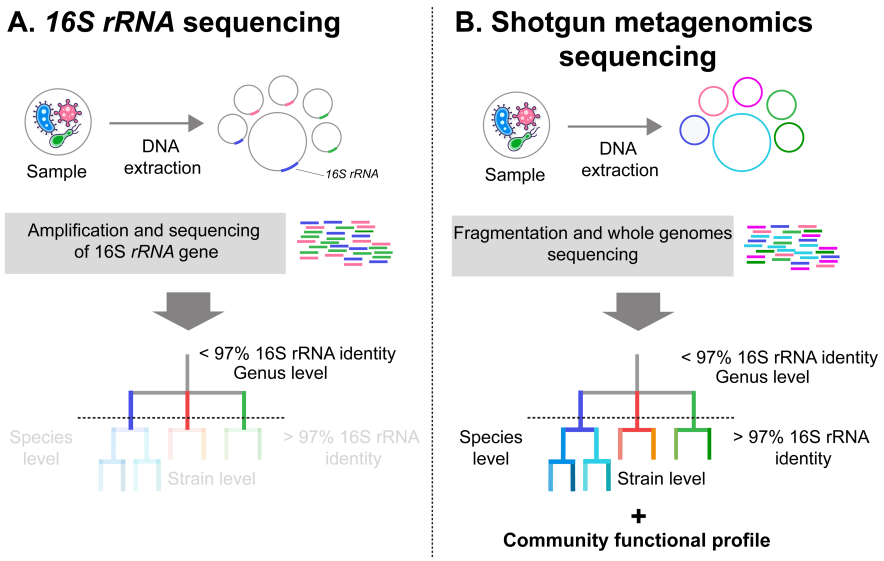


Figure 2: **Methods to study diversity within natural bacterial communities. A.** 16S rRNA sequencing. Following DNA extraction of the sample material, selected hypervariable regions included in the 16S rRNA gene are amplified using conserved primers and subjected to high throughput sequencing. Following the sequences clustering mapping bacteria can be assigned to different taxonomic ranks based on the $>$ 97 % similarity cut-off. **B.** Metagenomics sequencing. DNA is extracted from a sample and subsequently fragmented and all fragments are sequenced. Sequences are assembled to re-construct entire genomes or mapped against a reference-genome database. This methods allows to gain insights into the whole community diversity (and function) beyond the genus level.

Maintenance of diversity

While diversity within microbial communities can vary, natural bacterial communities rarely consist of pure cultures of single strains. They typically harbor a few to up to hundreds of different genera, species or strains. Different factors are likely to contribute to the maintenance of diversity in communities. First of all bacteria are metabolically extremely diverse, i.e. they can use distinct compounds available in the environment to obtain the carbon and nitrogen necessary for energy production and for

biomass building. Moreover, bacteria themselves can create new niches. They can modify physical, chemical or biological properties of the environment thereby facilitating the growth of other species (Madi et al. 2020). For instance a new bacterial variant may be able to integrate in a given community by growing on the metabolic waste products of another bacterium, thus establishing a cross-feeding interaction. For example, Blasche et al. showed that for the production of kefir, a fermented milk-drink, cross-feeding interactions between the community-fermenters are crucial, in that early milk-colonizers open the niche for the followers making available otherwise inaccessible metabolites (Blasche et al. 2021). This concept is referred to as 'Diversity Begets Diversity' (DBD), i.e. existing diversity promotes the accumulation of further diversity for instance *via* niche construction (Madi et al. 2020). In particular Madi et al., show that DBD is stronger in low diversity communities and weaker in more diverse communities, hypothetically because of niche saturation (Madi et al. 2020).

However, many bacteria have overlapping metabolic capabilities, and, since the diversity of nutrients that can be used is not infinite, there is a limit to the amount of cross-feeding interactions that can exist. In fact it has been recently shown that bacterial species benefit more from interacting with metabolically dissimilar partners, both considering their phylogenetic distance and the dissimilarity of their metabolic networks (Giri et al. 2021). Still, in nature we observe that often closely related bacteria with very similar metabolic capabilities are able to coexist. This is in contradiction with the *competition-relatedness hypothesis* (CRH), which predicts that the more closely two organisms are related the more likely they will compete strongly and eventually exclude each other due to overlapping niches. The overlap in the niche requirements will drive to the dominance of the variant with a fitness advantage over the other, e.g. a faster duplication time. This was recently tested experimentally by Ratzke et al., who established that more closely related bacteria interact more strongly and that stronger negative interactions resulted in the exclusion of species from the community and hence

a decrease in diversity (Ratzke, Barrere, and Gore 2020). This demonstrates that there is an important link between interbacterial interactions and diversity in microbial communities (Ratzke, Barrere, and Gore 2020). It thus appears puzzling why closely related bacteria would coexist and be stably maintained in natural communities.

The CRH concept was first suggested in Darwin's most notorious work "The origin of species" (1859) and it was then formulated by Gause in 1935 after having gathered experimental evidence such as in a study where he mixed different species of protozoa (Darwin 1858; Gause and Witt 1935). Gause observed that when the two species *Paramecium aurelia* and *Paramecium caudatum* were grown together the latter was always declining to the point of extinction (Gause and Witt 1935). On the other hand, when one of the two previously mentioned species were coupled with *Paramecium busaria*, the two protozoan species could coexist (Gause and Witt 1935). Gause could demonstrate that coexistence was promoted by the fact that while *P. aurelia* and *P. caudatum* consume more effectively bacterial components suspended in the upper layer of the liquid, *P. bursaria* prefers to feed on yeast cells settling on the bottom of the liquid (Gause and Witt 1935). Gause thus found one of the possible solutions for the diversity-CRH paradox: diversity maintenance thanks to ecological differentiation in space or nutrient source (Gause and Witt 1935). This is also what Darwin guessed for the 14 closely related finch species that he saw coexisting on the Galápagos islands and that harbor remarkable diversity in their beak form and consequently in their food-targets (Darwin 1858). This kind of phenomenon is known as *niche partitioning*, i.e. division of limited resources by species within an ecological niche. Evidence of niche partitioning within bacterial communities was gathered by Wilson et al. (1994), who worked on epiphytic bacterial populations and observed that the level of coexistence between different strains was inversely correlated with their ecological similarity (Wilson and Lindow 1994). By growing bacterial isolates individually on different substrates, they also obtained each strain approximate carbon sources utilization profile and identified a

number of overlapping carbon sources for which the more similar strains were likely competing (Wilson and Lindow 1994). In addition, Baran et al. (2015) investigated the role of bacterial substrate specialization in maintaining soil-communities diversity (Baran et al. 2015). They identified the pool of available metabolites in the shared niche and observed that individual isolates used only a small fraction of these, indicating resource partitioning among isolates (Baran et al. 2015).

The maintenance of diversity through niche partitioning among bacteria is likely an effect of natural selection acting on the community. At the same time, in host-associated communities, selection acting on the host can also promote diversity maintenance in order to augment redundancy and increase the stability of the community in the face of environmental disturbances (Foster et al. 2017; Ley, Peterson, and Gordon 2006).

Another solution is presented by the '*kill the winner'* hypothesis which proposes that classical predator-prey dynamics facilitate competing strains to coexist. The predator would decrease the population of a dominant strain, thereby facilitating other strains to take over the available niche space, only to be targeted by another predator when having reached high enough numbers. This would lead to oscillations of different strains in the community but never to the dominance of a single strain. Phages and protists are typically predators of bacteria in natural microbial communities and hence have been suggested to promote coexistence by hindering the growth of single-dominant strains (Koskella and Brockhurst 2014). For example, Brockhurst et al. showed that in presence of phages specifically infecting the dominating competitor in a community including *Pseudomonas fluorescens* and *Pseudomonas aeruginosa* enhanced the evenness of bacterial coexistence (Brockhurst et al. 2006)

At the same time, *neutral processes* can also play a role in maintaining diversity within natural bacterial community. The neutral theory predicts that all the individuals of a community are competitively equivalent and that differences in their abundance and

frequency are due to stochastic events (Hubbell 2001). An example of a neutral process is the *species-area relationship*, which predicts that there is a positive correlation between habitat size and diversity (Arrhenius 1921). Bell et al. experimentally validated this theory by measuring bacterial genetic diversity in water-filled tree-holes of different size and by concluding that diversity in this system increased with increasing three-hole size (Bell et al. 2005). Another example of neutral process are priority effects, for which the first species that colonizes an environment gains an advantage over the late arriving ones. Cheong et al. recently showed that priority effects (but also interbacterial competition) play a crucial role in shaping fungal-bacterial biofilm structure and in affecting their pathogenicity (Cheong et al. 2021). Finally, populations bottlenecks due to random sampling/dilution of bacterial populations, are another example of a neutral process that could contribute to diversity maintenance within bacterial communities (Wein and Dagan 2019).

Different types of bacterial interactions and how to measure them

We have seen from the previous chapters that interbacterial interactions are key for understanding the composition and dynamics of microbial communities. One way to study interactions between bacteria is by comparing the growth of a bacterium alone and when in presence of another bacterium in a pairwise manner. Depending on how the abundance of the first bacterium is affected by the presence of the second in *co-culture* if compared to *mono-culture*, the interaction is defined (Mitri and Foster, 2013) (**Figure 3**). If neither species is affected in their growth by the other's presence, it means that the two species are essentially non-interacting (*neutrality*) (**Figure 3**). Two species are interacting positively (*cooperation*) when their growth is enhanced in co-culture if compared to mono-culture (**Figure 3**). In this case the sum of the bacterial loads of the

two species when grown together is higher than the sum of the bacterial loads of the two species grown individually (**Figure 3**). On the other hand, two species are interacting negatively if one species is positively affected and the other is negatively affected or if they are both negatively affected (*parasitism* and *mutual inhibition* respectively) (**Figure 3**). In both of these cases the sum of the bacterial abundances when the species are together is lower if compared to when the two species are grown alone (**Figure 3**). If one of the two species is positively or negatively affected by the association while the other is not affected the interaction is defined as *commensalism* and *amensalism*, respectively (**Figure 3**). However, defining this type of interaction is ambiguous as in principle all interactions must have an effect on both partners, even if they could simply be too small to measure without the right method and the appropriate resolution (**Figure 3**).

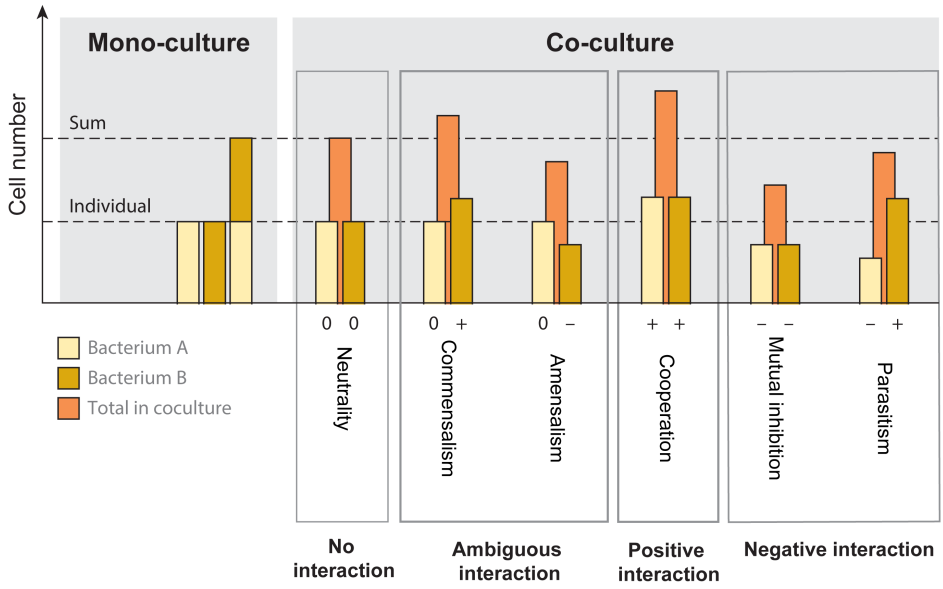


Figure 3. Assessing bacterial interactions by comparing growth in mono-culture vs co-culture. Each species (Bacterium A and B) is mono-cultured and co-cultured within the same condition. By comparing the growth (cell number) in mono-culture and co-culture for both species one can infer interactions. If there is no difference between mono and co-cultures, there are no interactions ongoing (neutrality). If one species is positively or negatively affected from the association and the other is not, the interaction is defined respectively commensalism and amensalism. The two species are cooperating if they are both positively affected by the association. If one species is positively affected and another is negatively affected or if they are both negatively affected by the association, they are interacting negatively (parasitism and mutual inhibition respectively). Adapted from Mitri and Foster (2013).

Studying pairwise interactions can be very valuable as it can provide a bird's-eye view of the stimulatory/inhibitory interactions in a community (Momeni, Xie, and Shou 2017). In addition, this approach can be complemented with other culture-based assays, such as growth in another species spent-media to check for instance for cross-feeding interactions, but also with screenings of the bacterial genomes to better understand the mechanisms underlying the interactions (Mitri and Richard Foster 2013). Pairwise interactions can also be combined and used to predict the composition of multispecies communities. The idea is that we can predict that in an community consisting of strains A, B, and C, if A and B (pairwise) interact positively with each other but negatively with C, C will likely go extinct. For example, Friedman et al. could predict survival in three-species competitions using pairwise competition outcomes with an accuracy of ~90% (Friedman, Higgins, and Gore 2017). However, in some cases higher order interactions (i.e. interactions between more than two species) are needed to predict the composition of multispecies communities. In this case C would not go extinct in the ABC community even if in mono-association with A or B it would be negatively affected because when the three are together the interaction dynamics would change. For example, Mickalide et al. observed that the outcome of a species invasion of a community was not predictable using pairwise interactions, suggesting that higher order interactions might be important (Mickalide and Kuehn 2019).

In addition to this, there are some important factors that need to be considered when studying bacterial interactions, such as the interaction-*distance*, the interaction-*environment* and the interaction-*level*. The interaction-distance is the range within which two species can interact and have an effect on each other growth. It was recently shown that bacteria interact only with their immediate neighbors, making the spatial arrangement of different species central for determining the type of interactions (Dal Co et al. 2020; Nadell, Drescher, and Foster 2016).

The environment also has an important influence on the interactions of bacteria. We know that an environment that is toxic for bacteria can promote positive interactions while an environment with increased nutrient availability can magnify negative interactions between genotypes (Piccardi, Vessman, and Mitri 2019; Ratzke, Barrere, and Gore 2020). It could be that a toxic environment requires more species that interact positively and facilitate each other, for instance to achieve detoxification, while within a more permissive environment the fact of having more species would just increase competition and mutual exclusion (Piccardi, Vessman, and Mitri 2019).

In addition, since bacterial communities include high extents of coexisting variants of the same species (i.e. strains), there is the need of carefully consider at which level of diversity it is appropriate to study interactions. Recently Goyal et al. tracked strain-level interactions within microbial communities associated with pitcher plants and observed that even extremely closely related strains exhibit vastly different interactions dynamics and that interactions in microbial communities might be strain-specific and not species-specific (Goyal et al. 2021).

Finally, it is important to keep in mind that in order to gain insights about the relevance of interactions between bacteria to determine the community-composition, it is necessary to study them over time. In fact, it could be that two species are interacting negatively in that they reach lower biomass when they are found together if compared to when they are alone, but this does not tell us much about these species ability to coexist over time. It could be that one strain ends up dominating and excluding the other, but it could as well be that the two strains are able to coexist even while competing (i.e. interacting negatively). The only way to investigate this is to measure interactions between species over-time, doing time-course studies.

From genetic to functional diversity

In the previous sections we described genetic diversity as an intrinsic property of bacterial communities, which can fluctuate depending on several factors. Nevertheless, it is still debated if the genetic diversity that we observe in such great amounts in natural bacterial communities corresponds to comparable amounts of *functional diversity* (Griffiths, Ritz, and Wheatley 1997). With metagenomics we can study the functional diversity of bacterial communities by measuring the number of genes that have a functional annotation, i.e. are predicted to carry out a certain function. One can speculate that the more different genes are present, the more the community will be functionally diverse. However, it is not yet clear if functional diversity at the genetic level necessarily leads to functional diversity at the phenotypic level, which could be for instance resistance to pathogen invasions or community metabolic-output in terms of fermentation products.

One way to investigate the correspondence between genetic and functional diversity in bacterial communities is by manipulating the genetic diversity of the community and monitoring changes in function. For example, Wagg et al. created a declining gradient of bacterial and fungal diversity among soil-plant ecosystems and observed that all monitored ecosystem functions (e.g. plant biomass production and litter decomposition) changed significantly depending on the soil biodiversity-level (Wagg et al. 2021).

Understanding the relationship between the functioning of an ecosystem and its diversity has broad implications that go far beyond the study of microbial communities. In particular, the negative impact of humans on almost all ecosystems on our planet and the resulting loss of biodiversity has received a lot of attention in past decades (Loreau et al. 2001). The general assumption that a more diverse community is a better community started to catch on, but it is not that simple and the link between the

diversity and the emergent functional properties of a community has not been established yet (Shade 2017).

Diversity within gut-associated communities

One of the most studied natural bacterial communities is the one of the microorganisms that live in the digestive tracts of animals: the *gut microbiota*. While gut microbiota research has received a lot of attention in recent years, this area of research is in fact not as new as we may think. The very first microbes that were ever observed were those that Antonie van Leeuwenhoek isolated from his oral and fecal microbiota (he called them "*animalcules*") around 1670-80s. However, it is true that with the access to the genetic information of microbes, in particular when sequencing technologies became broadly available, research on the gut microbiota took off. One of many important findings based on sequencing early on was the fact that the bacterial communities in our digestive systems are surprisingly diverse with few distantly related lineages but many closely related species and strains. Diversity within species is generated by continuous processes of variation generation, selection and drift (Van Rossum et al. 2020). As to why gut-associated communities are characterized by this kind of diversity structure, it could be because bacteria from only a few phyla contain the basic metabolic, physiological or structural features allowing them to adapt to and thrive in the gut environment (Ley, Peterson, and Gordon 2006). Indeed, the host is a harsh environment to colonize and adapt to, especially because of its immune system (Hooper and MacPherson 2010). This adaptation may have happened many times independently in different bacterial lineages within these phyla. Alternatively, it is possible that only a few founder lineages within these phyla were able to colonize the gut and subsequently diversified *in situ* into a vast array of ecologically different subdivisions (Moeller et al. 2016). There is evidence that this second hypothesis is the most likely: first, most gut-

associated bacteria live exclusively in this environment, second, it was shown that bacterial communities co-diversified with their hosts (Ley et al. 2008). It is thus likely that gut-bacteria have evolved *in situ* from a restricted number of founder species.

As to how the peculiar structure of diversity of gut-associated communities is maintained, there are few studies that approached this question experimentally. Although it has been suggested, there is no study that evaluated the impact of resource partitioning in terms of nutrients or space in facilitating the coexistence of closely related bacteria (Ley, Peterson, and Gordon 2006). Rawls et al. showed that the host can select its microbial constituents by transferring the zebrafish gut-community to germ-free mice and observing that the adult mouse cecum was able to support the microbial consortium after having heavily altered and shaped its composition (Rawls et al. 2006). The two main reasons for why the host could select for maintenance of diversity are: i) to promote functional redundancy and reduce the loss of function associated with the loss of variants in case of dysbiosis and ii) to get all the available gut-niches filled to prevent pathogen invasion (Foster and Bell 2012; Haag et al. 2012). Moreover, there are few studies that highlighted the role of phages in maintaining diversity within gut-associated bacterial communities either through kill-the-winner dynamics (Ventura et al. 2011) or by promoting the evolution of different phage-resistance mechanisms (Hussain et al. 2021). The 'diversity begets diversity' idea has also been explored in gut associated communities by Madi et al., who showed that in the animal distal gut (where the diversity is relatively low) there is a strong positive effect of diversity on diversification (Madi et al. 2020). Finally, the role of neutral processes such as priority effects in shaping and maintaining diversity in microbiomes has been established in several models (Debray et al. 2021). For example, Lee et al. showed that early arriving *Bacteroides* strains in the mouse gut saturate the available niches forcing the following strains to occupy the less protected ones (S. M. Lee et al. 2013).

Still, more experimental studies are needed to gain insights about diversity and its maintenance within gut-associated communities. For such studies, experimentally tractable communities of reduced complexity, usable for both *in vivo* and *ex vivo* studies (*in vitro* models), allowing to test the effect of different factors (e.g. nutrient availability) on diversity are necessary (Douglas 2019). The human gut microbiota, although it is an extremely relevant community, has several study-limits. In addition to the important ethical constraints that come with it, the bacterial community associated with the human gut is extremely complex, consisting of hundreds of different bacterial species and thousands bacterial strains of which only a minority can be cultured *ex vivo* (Ericsson and Franklin 2015). One of the most used vertebrate-models to study gut-associated communities are mice but, although they are extremely useful and allowed to make groundbreaking discoveries in the field, they are a relatively expensive model that also presents ethical constraints (Loan et al. 2015). On the other hand, lower vertebrate and invertebrate species have been proposed as cost-effective models of animal microbiomes of lower complexity that allow to perform extended experimental screenings with high numbers of replicates (Douglas 2019). One of these models is the microbiota associated with the gut of honey bees (Douglas 2019).

The honey bee gut microbiota

The honey bee gut microbiota as a model

Honey bees (*Apis mellifera*) have recently emerged as versatile models for gut microbiota research. The bacterial community associated with the bee gut is relatively simple as it is dominated by few core bacterial lineages that specifically colonize this environment and are present in the guts of most adult worker bees regardless of their geographic origin, their age, or the season (Kešnerová et al. 2020; Kwong et al. 2017).

Importantly, all these lineages can be cultured (and some even genetically modified) in the laboratory (Zheng et al. 2018). Moreover, microbiota-depleted bees can be reared in the laboratory without the use of antibiotics and colonized with single isolates or with defined synthetic communities (Kešnerová et al. 2017). This allows to perform experiments with large sample sizes that are inexpensive and without the ethical constraints of other animal models or humans (Zheng et al. 2018). With this model, coexistence, interactions, and function of gut bacteria can be tested both *in vitro* in culture tubes and *in vivo* within the honey bee gut.

The honey bee gut microbiota is not only a suitable model to study bacteria-bacteria interactions, but also to study bacteria-host interactions. In fact, several functions have been associated with the gut microbiota that are likely to impact bee health and disease, including nutrient digestion, production of fermentation products, immune system stimulation, behavioral modulation, or pathogen resistance (Engel, Martinson, and Moran 2012; F. J. Lee et al. 2015; Raymann, Shaffer, and Moran 2016; Steele et al. 2021; Zheng et al. 2016). The honey bee is a very important pollinator with huge ecologic and economic value and it was recently object of global concern due to high seasonal mortality rates (Johnson 2013). Therefore, studying the honey bee gut microbiota is not only relevant for fundamental science, but also because it can help us to come up with new strategies to improve bee health on the long run.

Diversity within the honey bee gut microbiota

The first studies investigating the honey bee gut community, using terminal restriction profiling of 16S rRNA gene fragments and metagenomics, identified a limited number bacterial groups or clusters, belonging to few distantly related bacterial lineages, that were not assigned to a specific taxonomic rank (Babendreier et al. 2007; Cox-Foster et al. 2007). Based on follow-up studies, it was then established that the community associated with the bee gut is composed of eight core bacterial lineages, named

phylotypes, which share > 97 % of sequence identity in the 16S rRNA gene and which belong to the Firmicutes, Actinobacteria and Proteobacteria phyla (Kwong and Moran 2016; Martinson et al. 2011; Powell et al. 2014) (**Figure 4.A**).

The assembly and maintenance of the bee gut microbiota across a bees' life is influenced by the holometabolism of honey bees, which have a larval and adult life stage. During the larval life stage, the gut is colonized by relatively few bacteria. These are shed or killed during metamorphosis, and hence young adult honey bees emerge from the brood cells without microbiota in their gut. It is supposed that the characteristic gut microbiota of the adults is picked up from the hive environment after pupal-eclosion or *via* social interactions with older nestmates (Powell et al. 2014). Bacteria colonize the gut during the 2 days post-eclosure but the establishment of more consistent core communities seems to be achieved within 4-6 days post-eclosure (Engel, Bartlett, and Moran 2015; Martinson, Moy, and Moran 2012; Powell et al. 2014). This kind of feature can be exploited to get microbiota depleted bees for the laboratory, i.e. by pulling bee pupae from frames before they eclose and by letting them ultimate their development in a sterile environment (Kešnerová et al. 2017; Zheng et al. 2018)

The abundance of the different community members and the overall diversity structure of the bee gut microbiota varies among the different organ sections (Martinson, Moy, and Moran 2012). The ileum and pylorus regions are dominated by the core Gram-negative species (*Snodgrassella alvi*, *Gilliamella apicola* and *Frischella perrara*) while the rectum region is dominated by the Gram-positive species (*Lactobacillus* Firm4 and Firm5 and *Bifidobacterium*) (Engel, Bartlett, and Moran 2015; Martinson et al. 2011; Martinson, Moy, and Moran 2012; Powell et al. 2014) (**Figure 4.A**). Moreover, the bee gut community structure also experiences changes related to the bee life cycle (Kešnerová et al. 2020). In fact, Kešnerová et al. showed that total bacterial loads in nurse-bees (young worker bees that stay inside the hive to take care of the larvae) and winter bees (bees with an extended lifespan that ensure colony survival during the winter season)

are higher than in foragers (older worker bees that forage pollen) (Kešnerová et al. 2020). Moreover, winter bees displayed the lowest community alpha diversity (Kešnerová et al. 2020). A possible explanation for these changes in community structure across these different 'bee-types' could be the dietary differences between them: while foragers mainly feed on nectar and honey, nurses and winter bees consume also pollen (Kešnerová et al. 2020).

Although the bee gut community is relatively simple at the phylotype-level, recent genomic analyses revealed that there is extensive diversity within phylotypes (Ellegaard et al. 2015; Ellegaard and Engel 2016, 2019; Engel, Martinson, and Moran 2012) (**Figure 4.B and C**). In fact, most phylotypes have diverged into 2-4 distinct bacterial lineages or sequence discrete populations that share $< 85\%$ pairwise average nucleotide identities (gANI) and that we call *species* (Ellegaard and Engel 2019) **(Figure 4.B)**. This is also true for other gut-communities associated with different bee species, such as *Apis ceranae* or different *Bombus* species, reminiscent of the fanlike architecture typical of gut-associated communities described above (**Figure 1**, palm tree) (Ellegaard et al. 2019, 2020). Moreover, species contain further diversity at the level of individual strains (**Figure 4.C**). While species are present in honey bee colonies worldwide and consistently co-occur in individual bees, the total diversity of strains present in a colony was found to be substantially higher than the diversity present within a given bee, suggesting that strains of the same species segregate into different host individuals (**Figure 4.C**) (Ellegaard et al. 2020; Ellegaard and Engel 2019). In fact, bees differ from one another based on their strain profiles (**Figure 4.D**) (Ellegaard and Engel 2019).

Why the phylotypes of the bee gut microbiota have diversified into different species and strains and how these can coexist has remained elusive. As strains of the same species can harbor markedly different gene content, the gut microbiota of individual bees may be functionally different across bees, despite being conserved at the species

level. This may have consequences for the impact of the gut microbiota on the host and its implication in health and disease.

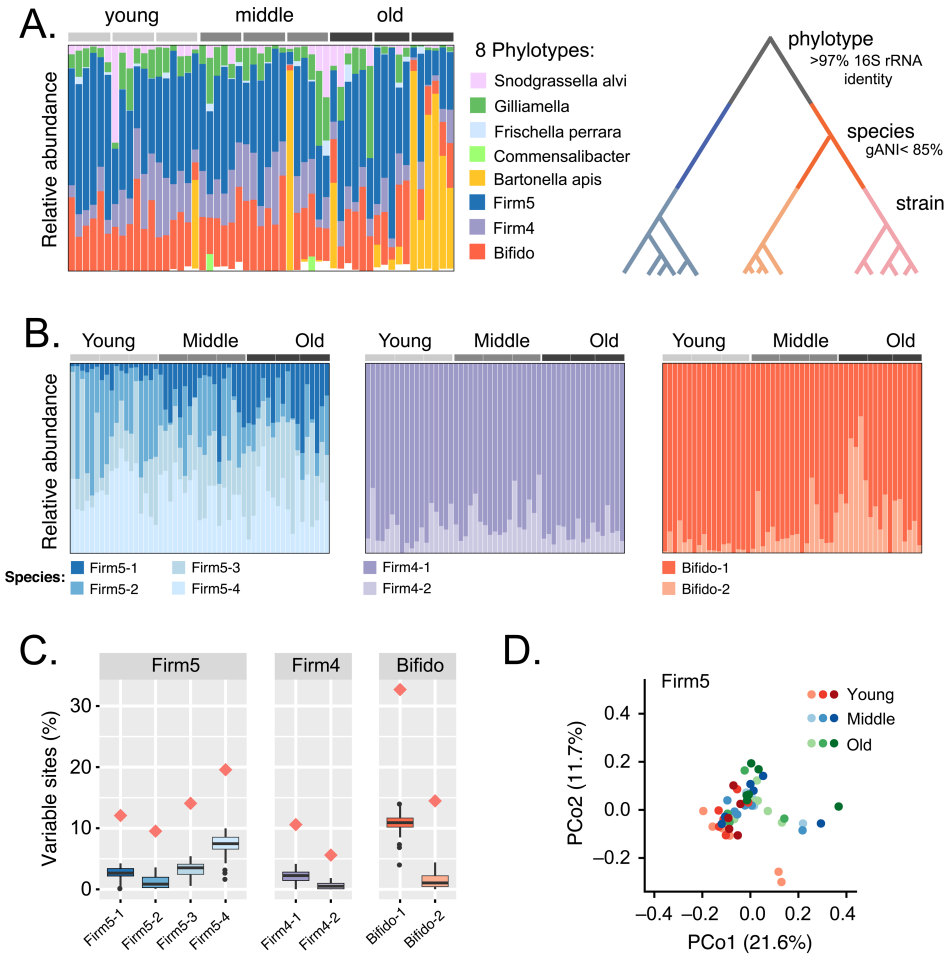


Figure 4: **Diversity of the honey bee gut microbiota. A.** Community composition across samples at the phylotype level. Bee age group and colony origin are indicated on top of the graph by grey bars. Phyla-classification information for each phylotype with correspondent gut-region occupied is displayed on the right. **B.** Community composition across samples at the species level for three of the core gut-microbiota members: *Lactobacillus* Firm5 (4 species), *Lactobacillus* Firm4 (2 species) and Bifido (2 species). Bee age group and colony origin are indicated on top of the graph by grey bars **C.** Boxes: percentage of polymorphic (variable) sites per species/phylotype within samples. Red diamonds: total fraction of polymorphic sites across the study. **D.** Principal coordinate analysis of Jaccard distances calculated between pairs of bees based on the shared variome for the phylotype *Lactobacillus* Firm5. Adapted from Ellegaard et al., 2019.

Diversity within the phylotype *Lactobacillus* Firm5

A striking example of how diversity beyond the phylotype level is structured in the honey bee gut microbiota is *Lactobacillus* Firm-5 (Ellegaard and Engel 2019) (**Figure**

5.A). *Lactobacillus* Firm5 consists of four species which co-occur within individual bees: *Lactobacillus apis* (Firm5-1), *Lactobacillus helsingborgensis* (Firm5-2), *Lactobacillus melliventris* (Firm5-3) and *Lactobacillus kullabergensis* (Firm5-4) (Ellegaard and Engel 2019) (**Figure 4.B,** and **5.A**). The four species harbor a very high amount of species-specific (or non-conserved) gene families, suggesting that they evolved by adapting to different ecological niches, which would allow them to coexist (Ellegaard et al. 2019) (**Figure 5.B**). Comparative genomics showed that this phylotype is characterized by an exceptionally large number of genes belonging to the COG functional category "G" and that these are mainly genes coding for phosphotransferase transporters (PTS) for the import of sugars (Ellegaard et al. 2015) (**Figure 5.C**). Interestingly, functions for the breakdown, transport and metabolism of carbohydrates are among the genes that vary the most across species, i.e. that are less conserved, suggesting that adaptation to different nutrient-derived carbohydrates might facilitate coexistence (Ellegaard et al. 2019). *Lactobacillus* Firm5 is a facultative anaerobe, which lacks a respiratory chain and converts sugars into organic acids via fermentation. Importantly, in a recent metabolomics study this phylotype was shown to specifically employ as metabolic substrates pollen-derived compounds such as flavonoids (e.g. Rutin) and pollen-coat phenolamides (Kešnerová et al. 2017) (**Figure 5.D**).

Each Firm5 species harbors further diversity at the strain level and the amount of variants detected differs between species (**Figure 4.C** and **5.A**). Moreover, in all four species, the amount of strain-level diversity is higher across bees than within bees (**Figure 4.C**). While adaptation to different ecological niches likely promoted the evolution of the four species allowing their coexistence, conspecific strains may be functionally too similar to coexist. Moreover, it is not clear if to the genetic diversity found within the Firm5 phylotype there is a correspondent functional diversity, for instance in the species or strains metabolic output.

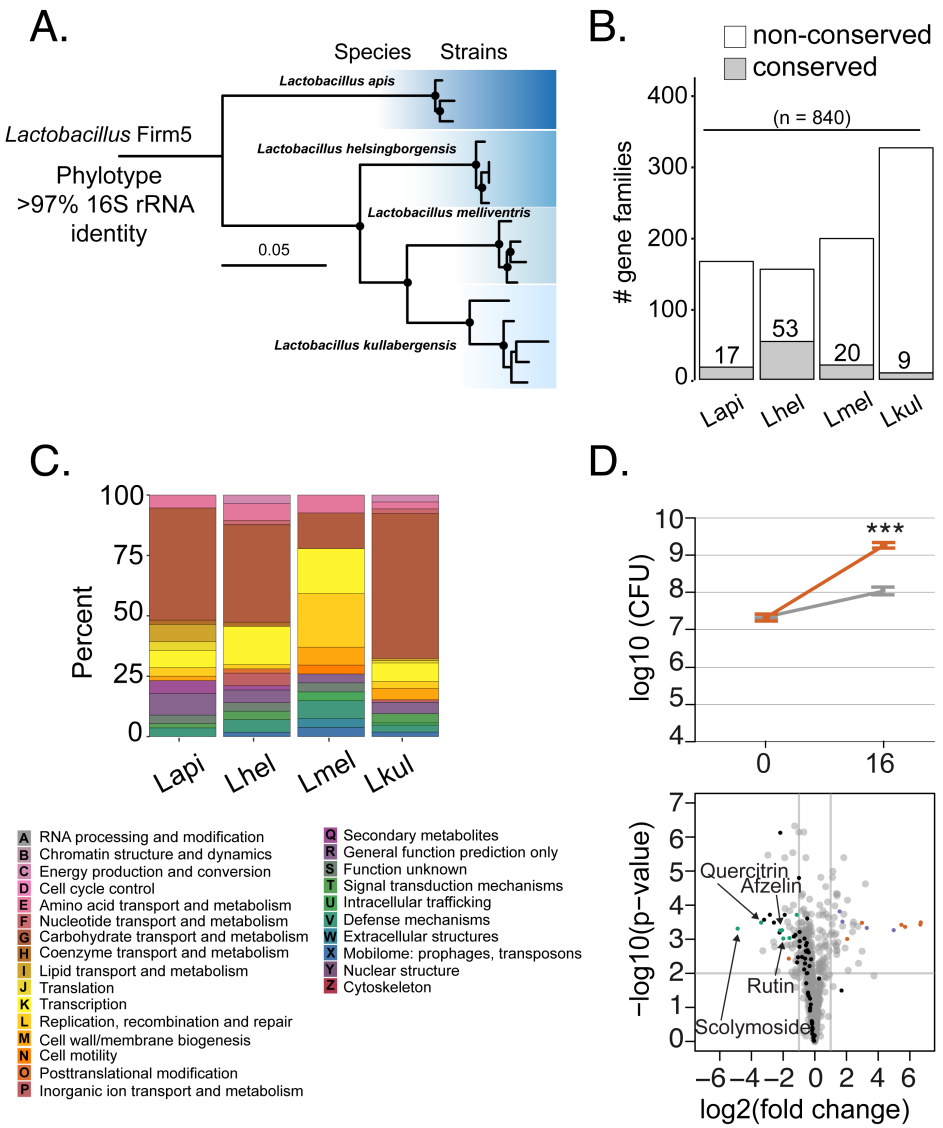


Figure 5: **The phylotype *Lactobacillus* Firm5. A.** Intra-phylotype diversity of *Lactobacillus* Firm5. Maximum-likelihood core genome phylogeny. Different shades of blue represent different species. The scalebar corresponds to 0.05 substitutions per site. Adapted from (Ellegaard and Engel 2019). **B.** Firm5 gene families. Non-conserved = fraction of gene families present in every genome of a species; conserved = fraction of gene families present in at least one genome of a given species. Numbers above the graph indicate total number of lineage-specific gene families for each host group. Adapted from (Ellegaard et al. 2019). **C.** Categories of Clusters of Orthologous Groups (COGs) for species specific gene families. Each color represents a different COG (see legend). Adapted from (Ellegaard et al. 2019). **D.** Top panel: line graph showing the growth of *Lactobacillus* Firm5 in control medium (grey lines) and in pollen-conditioned medium (orange lines) at time points 0h and 16h. Values are the mean of five replicates, with error bars indicating standard deviation. ***P < 0.001 (Welch's t test). Bottom panel: Volcano plots of significance (Welch's t test Benjamini and Hochberg adjusted [BH adj.] P value) versus $\log_2$(fold change) show metabolic changes in pollen-conditioned medium at time point 16 h relative to 0 h. Ions identified as pollen derived are highlighted in black. Ions annotated as glycosylated flavonoids, flavonoid aglycones (non-glycosylated flavonoids), or putative flavonoid breakdown products are shown in color (green: glycosylated flavonoids, red: flavonoid aglycones, violet: aromatic compound degradation intermediates, black: pollen metabolites) when they displayed $\log_2$(fold changes) $>= |1|$. Other annotated ions are plotted in grey. Adapted from (Kešnerová et al., 2017).

Objectives of the thesis

The overall objective of my PhD thesis was to study diversity within natural bacterial communities. In particular, I was interested in understanding how closely related bacteria are maintained within communities and how such species/strain-level diversity can impact the functional diversity of a bacterial community.

To reach these objectives, I focused on the honey bee gut microbiota, specifically on the phylotype *Lactobacillus* Firm5, which, as mentioned above, diverged into four species that consistently coexist within the honey bee gut and which harbor further diversity at the strain-level.

In the first results chapter (**Manuscript I**), I tested the hypothesis that the different *Lactobacillus* Firm5 species can coexist thanks to the partitioning of nutrients derived from the pollen-diet of the host. To do so, I selected one strain for each of the four species and I performed experiments both *in vivo* and *in vitro* under different dietary conditions. Moreover, to better understand the underlying mechanisms of coexistence I applied metatranscriptomics and metabolomics analyses.

In the second results chapter (**Manuscript II**), I changed my focus from species-level to strain-level diversity. In particular, I wanted to test if interactions and coexistence between Firm5 strains are different depending on whether they are conspecific (i.e. from the same species) or not. To this end, I selected three strains for each Firm5 species and tested interactions *in vivo* in a pairwise manner, characterizing them qualitatively and quantitatively and highlighting the differences between conspecific and allospecific strains. To test for coexistence differences, I serially passaged pairs of conspecific and allospecific strains *in vivo* across gnotobiotic bees and *in vitro* in culture tubes.

In the third results chapter (**Manuscript III**), I wanted to test if the genetic diversity within the Firm5 phylotype impacts the functional output of the community. In particular, I investigated if there were differences between Firm5 species or strains in the accumulation of main fermentation products, i.e. short chain fatty acids (SCFAs). I therefore mono-colonized microbiota depleted bees with the same strains that I selected for Manuscript II and I used metabolomics GC-MS analysis to evaluate the production of SCFAs in the bees guts and hemolymph.

Manuscript I: Niche partitioning facilitates coexistence of closely related honey bee gut bacteria

Summary

In this Manuscript we investigated the effect of resource partitioning in facilitating the coexistence of the four species of the honey bee gut symbiont *Lactobacillus* Firm5 (*Lactobacillus apis*, *Lactobacillus helsingborgensis*, *Lactobacillus melliventris* and *Lactobacillus kullabergensis*).

To this end, we selected one strain representative for each Firm5 species, we mixed them and we serially passaged them both *in vivo* through gnotobiotic bees and *in vitro* in liquid cultures under two different nutrient-treatments: in the presence of pollen (bee diet) or in the presence of simple sugars. Our hypothesis was that the presence of pollen, a rich and heterogeneous nutrient source consisting of a variety of substrates (e.g. different sugars especially in the form of polysaccharides), facilitates the coexistence between the four species favoring resource partitioning and holding back competitive exclusion. On the other hand, we hypothesize that the presence of simple sugars as the only carbon source available in the shared niche would not allow coexistence but rather promote stronger competition resulting in competitive exclusion and ultimately in the dominance of only one species (the one that has a fitness advantage over the others, e.g. a faster duplication time). Although we found that the four species were competing in both nutrient treatments, they were able to stably coexist, both *in vivo* and *in vitro*, only in presence of pollen, supporting our

Gut communes

Closely related bacteria in the gut of honeybees can coexist in communities by sharing diet-derived nutrients.

Digest · Aug 31, 2021

Our paper highlighted in eLife digest, i.e. a platform where plain language summaries of eLife papers are made available.

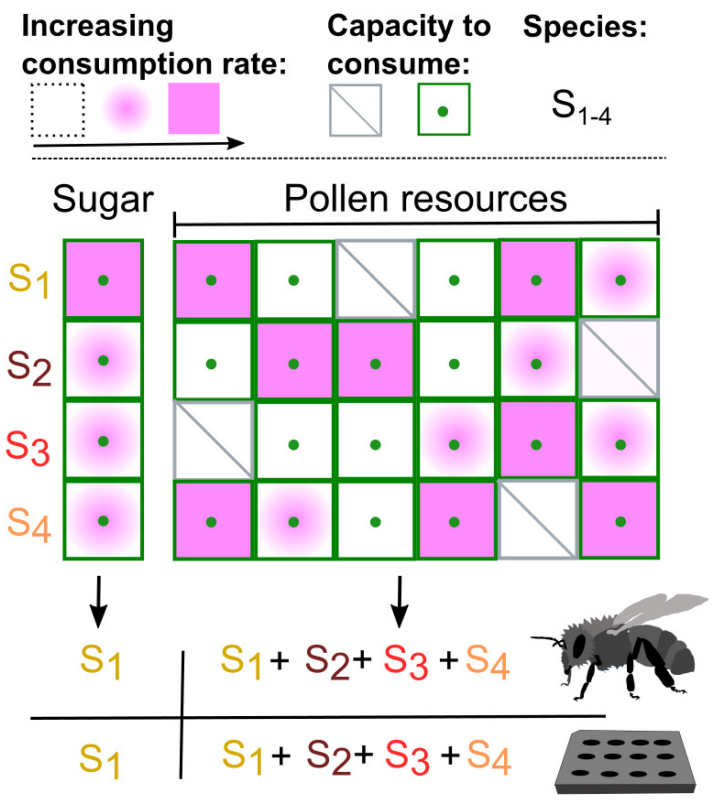
initial hypothesis. In presence of simple sugars only one dominant species was detectable at the end of the passaging experiments and interestingly it was not the same species to dominate *in vivo* and *in vitro*.

We then investigated the underlying molecular mechanism of coexistence *in vivo* and *in vitro* using both metatranscriptomics (to understand if different species were upregulating different genes for the utilization of pollen-substrates) and metabolomics (to understand if different species were consuming different pollen-substrates). Our results strongly indicate resource partitioning as the basis of the coexistence between the four Firm5 species, independent of the environment (*in vitro* or *in vivo*). In fact, using metatranscriptomics we found that all four species upregulated different genes for the utilization of different carbon sources (carbohydrate transport and metabolism functions) in presence of pollen both *in vivo* and *in vitro*. We also compared the different species transcriptomes *in vitro* when cultured individually (mono-cultures) and together (co-cultures). Interestingly, we identified no transcriptional changes when comparing mono-cultures and co-cultures of the four *Lactobacillus* species, indicating that they are likely not cross-feeding. In addition, using metabolomics on the pollen-supplemented medium in which we grew the four species separately and comparing before and after growth, we found that the four species consumed different pollen-metabolites, in particular distinct glycosylated secondary plant metabolites.

In summary, our results show that the four Firm5 species coexistence is facilitated by the partitioning of the pollen diet of the bee. Our results emphasize the advantages of using the honey bee model to study bacterial coexistence and interactions as they can be tested *in vivo* and recapitulated *in vitro*.

The Supplementary Figures of this paper can be found in the Annex (I) of this thesis. The Supplementary Files of this paper can be found online (https://elifesciences.org/articles/68583/figures#content, see Additional Files).

Summary Figure by Julia Schwartzman - highlight for our eLife paper (DOI: 10.7554/eLife.72380)



How a diet of pollen allows closely related species of bacteria to coexist in the gut of honey bees. Brochet et al. created an artificial microbial community made up of four species that are commonly found in the gut of honey bees (S1, S2, S3, S4) and studied the growth of this model community *in vivo* (bee symbol) and *in vitro*. When grown with just one resource (sugar) available, one species (S1 in this instance) consumed the resource at a faster rate than the other species (see key at top left), even though all four species were capable of consuming sugar (as indicated by green dots). However, pollen offers multiple nutrients (as represented by the six columns in the figure), which the four species of bacteria consume in different ways. For example, S3 is unable to consume the nutrient represented by column 1 (indicated by a diagonal line), but can consume the nutrients represented by the other five columns: moreover, it consumes some nutrients at a higher rate than other species. If the consumption profiles of the four species complement each other (as is the case for the four species studied), they can coexist when grown in the gut of honey bees fed a diet of pollen or when cultured on pollen in the laboratory.

RESEARCH ARTICLE

Niche partitioning facilitates coexistence of closely related honey bee gut bacteria

Silvia Brochet[1], Andrew Quinn[1], Ruben AT Mars[2], Nicolas Neuschwander[1], Uwe Sauer[2], Philipp Engel[1]*

[1]Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland; [2]Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland

Abstract Ecological processes underlying bacterial coexistence in the gut are not well understood. Here, we disentangled the effect of the host and the diet on the coexistence of four closely related *Lactobacillus* species colonizing the honey bee gut. We serially passaged the four species through gnotobiotic bees and in liquid cultures in the presence of either pollen (bee diet) or simple sugars. Although the four species engaged in negative interactions, they were able to stably coexist, both in vivo and in vitro. However, coexistence was only possible in the presence of pollen, and not in simple sugars, independent of the environment. Using metatranscriptomics and metabolomics, we found that the four species utilize different pollen-derived carbohydrate substrates indicating resource partitioning as the basis of coexistence. Our results show that despite longstanding host association, gut bacterial interactions can be recapitulated in vitro providing insights about bacterial coexistence when combined with in vivo experiments.

Introduction

Gut microbial communities are usually dominated by few bacterial phyla and families, but contain a high extent of species- and strain-level diversity (***Ley et al., 2008***; ***Dethlefsen et al., 2007***). According to the competition-relatedness hypothesis, the more closely two organisms are related the more likely it is that they will compete and exclude each other due to overlapping niches (***Elton, 1946***). Therefore, it has remained unclear how closely related microbes can be maintained in the gut, or in any other natural microbial ecosystem.

The high concentration of nutrients and the structured environment of the gut may allow functionally redundant species or strains to coexist (***Ley et al., 2006***). The host may even select for such redundancy, as it can increase the stability and resilience of the microbiota against environmental disturbance (***Ley et al., 2006***; ***Foster et al., 2017***). Phage predation can also contribute to the maintenance of diversity by imposing kill-the-winner dynamics and hindering the outgrowth of a single dominant strain (***Koskella and Brockhurst, 2014***). Another possibility is that closely related species, and even strains of the same species, have functionally diverged from each other and occupy distinct ecological niches (***Chesson, 2000***; ***Bittleston et al., 2019***). The genomic flexibility of bacteria facilitates adaptation to different nutrients, provided in the diet or by the host (***Berasategui et al., 2017***; ***Martens et al., 2008***), or result from interactions with other bacteria (***Madi et al., 2020***), such as cross-feeding (***Goldford et al., 2018***) or cooperative glycan breakdown (***Rakoff-Nahoum et al., 2016***).

Few experimental studies have investigated the coexistence of bacteria in host-associated microbial communities. The high diversity in these ecosystems and the resistance of many host-associated bacteria to experimental manipulations introduce considerable challenges for such approaches (***Ortiz et al., 2021***; ***Venturelli et al., 2018***). Moreover, community dynamics observed in vivo can be difficult to reproduce under laboratory conditions, as the host presents a highly specialized

***For correspondence:** philipp.engel@unil.ch

Competing interests: The authors declare that no competing interests exist.

Funding: See page 19

Preprinted: 12 March 2021
Received: 19 March 2021
Accepted: 14 July 2021
Published: 19 July 2021

Reviewing editor: Karina B Xavier, Instituto Gulbenkian de Ciência, Portugal

eLife digest Microbes colonize nearly every environment on Earth, from the ocean and soil to the inner and outer surfaces of animals, such as the gut or skin. They form communities that are usually made up of a diverse range of bacteria, often containing closely related species – a key factor for a successful community.

But closely related bacteria can battle for the same resources, so it is unclear how they manage to live alongside each other without competing against one another. While diet is thought to play a key role in enabling closely related bacterial species to co-exist in the gut of an animal, experimental evidence is lacking, due to the difficulty in replicating these systems in the laboratory.

One strategy for investigating microbial communities is using honeybees. A major dietary source for honeybees is pollen, which can also be applied in the laboratory to grow diverse types of bacteria found in the honeybee gut. In addition, scientists can generate bees that lack microbial communities in the gut, allowing them to add specific types of bacteria to study their impact.

Brochet et al. used this approach with Western honeybees to assess whether diet enables closely related bacteria to live alongside one another in the gut. First, they colonized bees that lacked gut microbes with four closely related bacteria of the genus *Lactobacillus*, alone or together, and fed the bees either sugar water or sugar water and pollen. After five days, the gut bacteria were analysed. This revealed that bees fed on sugar water only had one dominant *Lactobacillus* species present in their gut, while bees fed with additional pollen harboured all four *Lactobacillus* species. Further analysis of these four bacterial species revealed that each of them activates distinct genes when grown on pollen, allowing the different species to consume specific nutrients from broken down pollen.

These findings show that closely related bacteria can coexist in the gut by sharing the different nutrients provided in the diet of the host. Consequently, differences in dietary intake in honeybees and other animals may affect the diversity of gut bacteria, and potentially the health of an animal.

nutritional and spatial environment. Thus, there is a need for in vitro models that can reproduce ecological interactions observed in vivo, from simple co-culturing setups (***Li et al., 2019***) to sophisticated 'organoids-on-a-chip' systems (***Jalili-Firoozinezhad et al., 2019***; ***Nikolaev et al., 2020***).

The gut microbiota of the Western honey bee (*Apis mellifera*) is composed of a few deep-branching phylogenetic lineages (phylotypes) belonging to the Firmicutes, Actinobacteria, and Proteobacteria phyla (***Martinson et al., 2011***; ***Kwong and Moran, 2016***). Most of these lineages are composed of several closely related sequence-discrete populations, hereafter referred to as species, each of which contains further diversity at the strain-level (***Ellegaard and Engel, 2016***; ***Ellegaard and Engel, 2019***; ***Engel et al., 2012***; ***Ellegaard et al., 2015***). Microbiota-depleted bees can be generated and experimentally colonized with synthetic communities of different strains. Moreover, most community members can be cultured in pollen, which is the major dietary source of honey bees (***Kešnerová et al., 2017***). This experimental tractability offers an excellent opportunity to probe the coexistence of bacteria in the gut of their native host and in controlled laboratory cultures using similar nutritional conditions.

One of the most abundant and diverse phylotype of the honey bee gut microbiota is *Lactobacillus* Firm5 (***Ellegaard and Engel, 2019***). This phylotype consists of facultative anaerobes that ferment sugars into organic acids and utilize various pollen-derived glycosylated plant compounds, such as flavonoids (***Kešnerová et al., 2017***). *Lactobacillus* Firm5 is specific to social bees but has diverged into many different species of which four are specifically associated with the Western honey bee, *Apis mellifera: Lactobacillus apis* (Lapi), *Lactobacillus helsingborgensis* (Lhel), *Lactobacillus mellifer-ventris* (Lmel), and *Lactobacillus kullabergensis* (Lkul). The four species are consistently present in the gut of individual honey bees suggesting that they can share the available niches and stably coexist despite their phylogenetic relatedness. Genomic analysis has revealed that these species share <85% pairwise average nucleotide identities (gANI) and exhibit high levels of genomic variation in terms of carbohydrate metabolism (***Ellegaard and Engel, 2019***; ***Ellegaard et al., 2015***). However, whether the coexistence is facilitated by adaptation to different nutritional niches, and to what

extent the host environment, the diet, or the interactions with other community members matter is currently unknown.

Here, we tested under which conditions the four *Lactobacillus* Firm5 species can coexist and investigated the underlying molecular mechanism. We serially passaged the four species in vivo through gnotobiotic bees and in vitro in liquid cultures, and applied RNA sequencing and metabolomics analysis. Our results show that the coexistence of the four species is mediated by the partitioning of nutrients derived from the pollen diet of bees and is largely independent from the presence of the host or other community members.

Results

The coexistence of four related *Lactobacillus* species in the honey bee gut depends on the host diet

All experiments in this study were conducted with four bacterial isolates representing the four *Lactobacillus* Firm5 species (Lapi, Lhel, Lmel, and Lkul) associated with the Western honey bee. We first tested if the four species can establish in the gut of gnotobiotic bees under two different dietary conditions. To this end, we colonized microbiota-depleted bees with each of the four species, alone or together, and fed bees either sterilized sugar water (SW) or sterilized sugar water and pollen (SW+PG). Five days post-colonization, we assessed the bacterial loads in the gut by counting CFUs (***Figure 1A***, ***Supplementary file 3***). When mono-colonized, the four species established in the gut of microbiota-depleted bees independent of the dietary treatment (***Figure 1A***). In the SW treatment, the colonization levels were generally lower than in the SW+PG treatment (***Figure 1A***, ANOVA q-value < 0.01) confirming previous results that pollen increases the total bacterial load in the gut (***Kešnerová et al., 2020***). There was no statistically significant difference between the total bacterial loads of the mono-colonizations and the co-colonizations in either dietary treatment, with the exception of the mono-colonization with Lkul, which showed higher loads than the co-colonizations in SW (***Figure 1A***, ANOVA q-value < 0.01). Consequently, the sum of the bacterial loads of the mono-colonizations exceeded the total bacterial load of the co-colonizations in both dietary treatments, suggesting that the species engage in negative interactions when colonizing the honey bee gut together.

To test if the four species can stably coexist in the bee gut, we serially passaged the community seven times in microbiota-depleted bees under both dietary conditions (SW and SW+PG). After each passage (i.e. after 5 days of colonization), we used amplicon sequencing of a

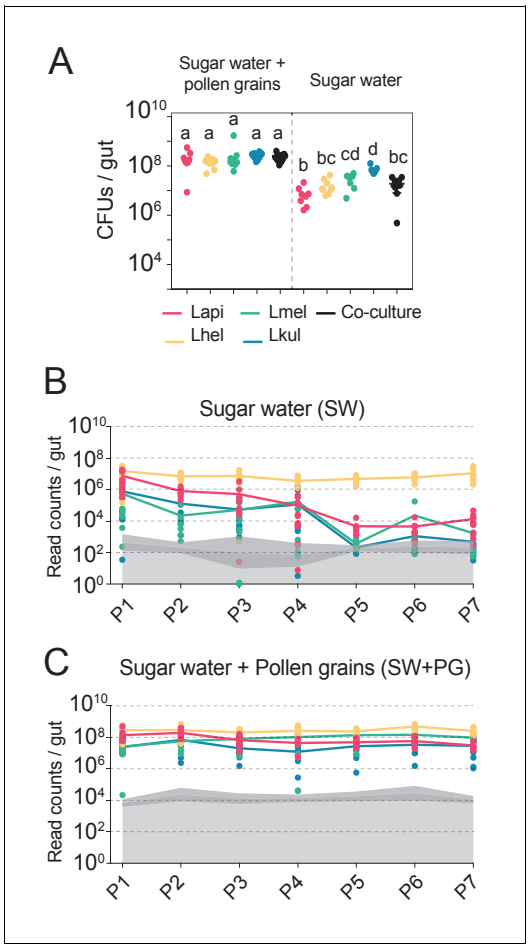


Figure 1. The presence of dietary pollen facilitates the stable coexistence of the four *Lactobacillus* species in the honey bee gut. (**A**) Bacterial abundance (CFUs) in the gut of gnotobiotic bees (n = 7–10) colonized with the four species separately or together under two different dietary conditions. Bees were sampled five days after colonization. Statistical differences (ANOVA with Tuckey post-hoc test and BH correction) are depicted by different letters. (**B, C**) Changes in the absolute abundance of each member of the four-species community across the seven serial passages (P1–P7) through the gut of gnotobiotic bees. The absolute abundance of each species was determined by multiplying the total number of CFUs with the relative abundance of each species in the community. Grey areas represent the limit of detection which can vary depending on the sequencing depth of each replicate (see Materials and methods). Therefore, the average limit of detection and the 95% confidence intervals are shown.

discriminatory housekeeping gene fragment (see Materials and methods) in combination with CFU counting to determine the absolute abundance of each species in the community. We observed clear differences between the two dietary conditions in the ability of the four species to coexist across the passages (***Figure 1B–C***, ***Supplementary file 4***). In the SW treatment, all species were initially detectable in most samples (P1, ***Figure 1B***). However, three species (Lapi, Lmel and Lkul) steadily decreased in abundance in the subsequent passages resulting in a rapid dominance of Lhel (***Figure 1B***). Lmel and Lkul reached the detection limit and Lapi decreased to around 10^4 bacteria/gut by passage five (P5, ***Figure 1B***). Only Lhel was stably maintained across all seven passages and was present at around 1000x higher abundance than Lapi at the end of the experiment (~10^7 bacteria/gut, ***Figure 1B***). In the contrary, in the SW+PG treatment, all four species were detectable in all passages at around 10^6 to 10^8 bacteria/gut, and displayed a highly stable abundance profile over time (***Figure 1C***).

In summary, these findings show that the four species can stably coexist in vivo when bees are fed pollen, but not when they are only fed sugar water. This is consistent with the idea that pollen facilitates niche partitioning in the honey bee gut by offering competing species different ecological niches facilitating their coexistence.

In vitro co-culture experiments recapitulate the nutrient-dependent coexistence of the four *Lactobacillus* species

We next tested if the four species can also coexist in vitro, outside of the host environment, under different nutrient conditions. To this end, we cultured the species alone or together in minimal medium supplemented with either glucose (G), pollen extract (PE), or entire pollen grains (PG). All four species were able to grow when cultured alone under the three nutrient conditions (***Figure 2—figure supplement 1***, ***Supplementary file 3***). Growth yields of Lhel, Lkul, and the co-culture were slightly lower in PE and PG than in G, and Lmel showed lower growth yields than some of the other species in PE and G (***Figure 2—figure supplement 1***, ANOVA q-value < 0.01). As in vivo, the total bacterial loads of the co-cultures were not consistently different from those of the mono-cultures (***Figure 2—figure supplement 1***), suggesting that the four species have overlapping metabolic niches and engage in negative interactions with each other.

We then serially passaged the co-cultures 21 times under the three different nutrient conditions by transferring an aliquot after 24 hr of growth into fresh culture medium (1:20). The absolute abundance of each strain was determined after every other passage by combining amplicon sequencing with qPCR (see Materials and methods). As for the in vivo experiment, we observed clear differences in the growth dynamics of the four species, both over time and between the glucose and the pollen culture conditions (***Figure 2***, ***Supplementary file 4***). In the presence of glucose, three of the four species (Lhel, Lmel, and Lkul) steadily decreased in abundance over time (***Figure 2A***), with two of them reaching the limit of detection (<10^5 bacteria/ml) after about 11 passages (P11). In contrast, Lapi was stably maintained at high abundance (10^9 bacteria/ml) until the last passage (***Figure 2A***) and hence dominated the co-culture for most of the transfer experiment. In the presence of PE or PG, the four species revealed very different growth behaviors (***Figure 2B and C***). None of the species decreased over time, and after 21 transfers all species still yielded between 10^6 and 10^9 bacteria/ml.

To look at changes in community composition over time, we measured the community stability (temporal mean divided by temporal standard deviation of the species abundances) in sliding windows of five passages. Little to no change in community stability was observed for the two pollen conditions throughout the experiment, whereas in glucose the community reached a stable state after ~11 transfers (***Figure 2D***). To compare the growth yields of each species across the three nutrient conditions, we only considered the passages after which community stability was reached (P13-21). With the exception of Lapi all species reached higher yields in the presence of pollen as compared to glucose (***Figure 2E***, ANOVA q-value < 0.01). Notably, Lmel was the only species that showed improved growth in PG as compared to PE (***Figure 2A–C***).

In summary, these findings show that the nutrient-dependent coexistence of the four *Lactobacillus* species observed in vivo can be recapitulated in vitro in a simple co-culture experiment, suggesting that the partitioning of pollen-derived nutrients is sufficient for enabling coexistence. Similar results were obtained for a second in vitro experiment which included the same nutrient conditions, but was only conducted for ten transfers (***Figure 2—figure supplement 2***).

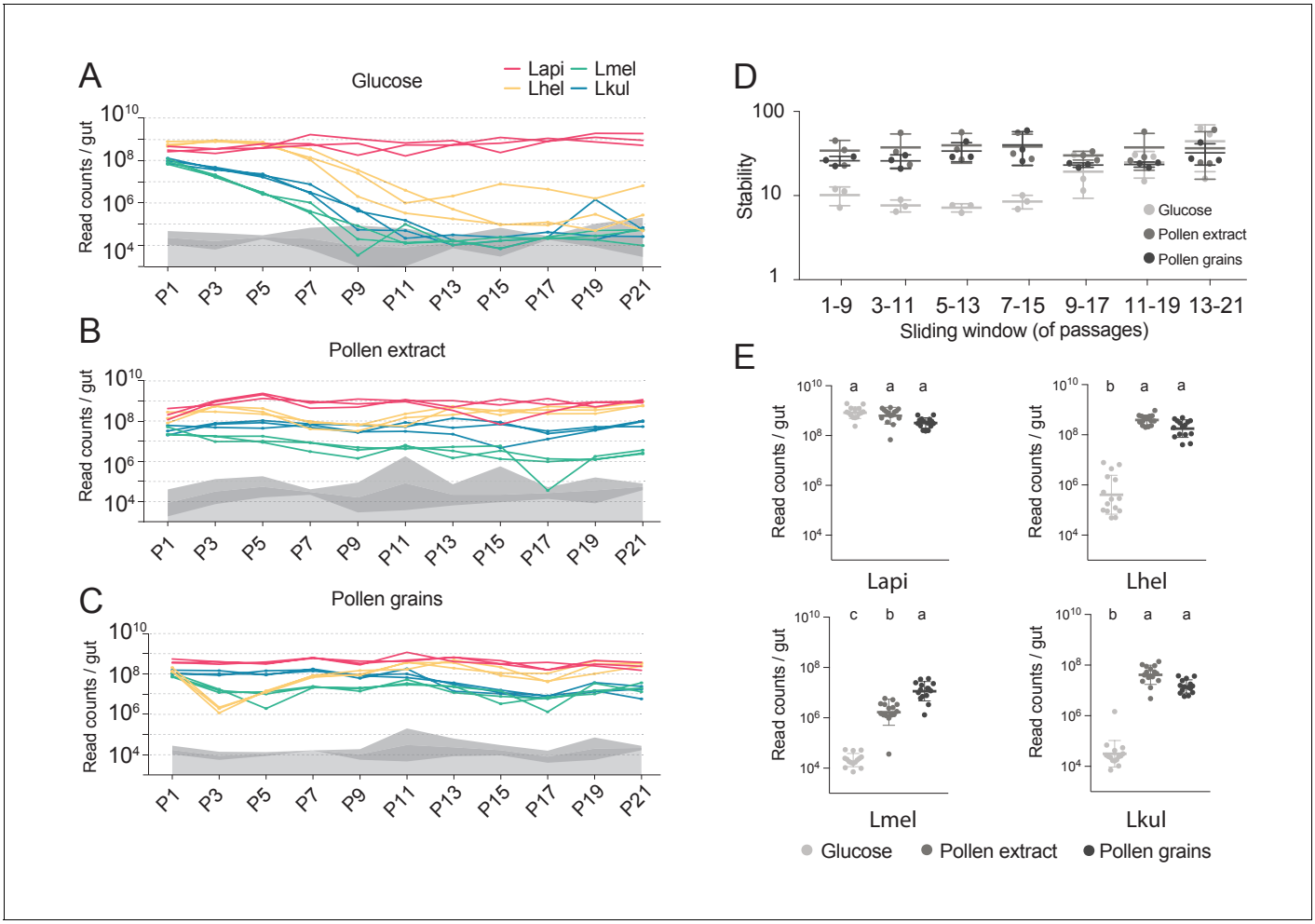


Figure 2. The stable coexistence of the four *Lactobacillus* species can be recapitulated in vitro in the presence of pollen. (**A–C**) Changes in total abundance of the four species when serial passaged in co-culture for 21 times in minimal medium supplemented with (**A**) 2% (w/v) glucose, (**B**) 10% (v/v) pollen extract, and (**C**) 10% (v/v) pollen grains. The absolute abundance of each species was determined by multiplying the total number of CFUs with the proportion of each strain in a given sample as based on amplicon sequencing (see Materials and methods). Gray areas indicate the limit of detection as explained in the Materials and methods. (**D**) Community stability of each replicate calculated based on the species abundances for a sliding window of five passages with a step size of 1. (**E**) Absolute abundance of each species across the three treatments considering the replicates of passages 13–21, which is when the community reached stability. Statistical differences (ANOVA with Tuckey post-hoc test and BH correction) are depicted by different letters.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Colony-forming units (CFUs) per ml of culture after 24 hr of growth of the four species in mono-cultures (n=3) or in co-culture (n=3) in the presence of 2% (w/v) glucose (**G**), 10% pollen extract (PE), or 10% pollen grains (PG).

Figure supplement 2. Second in vitro transfer experiment.

The four *Lactobacillus* species upregulate divergent carbohydrate transport and metabolism functions in the presence of pollen during gut colonization

Given the impact of pollen on the coexistence of the four *Lactobacillus* species, we tested if genes involved in nutrient acquisition and metabolism were differentially expressed between the dietary treatments. To this end, we carried out RNA sequencing of the four-species community in honey bees that were fed either sugar water (SW) or sugar water and pollen grains (SW+PG) (***Figure 3A***). Multidimensional scaling (MDS) of the normalized read counts mapped to each species revealed that most samples clustered by treatment (SW+PG versus SW) (***Figure 3—figure supplement 1***), indicating that all four species exhibited dietary-specific transcriptional responses.

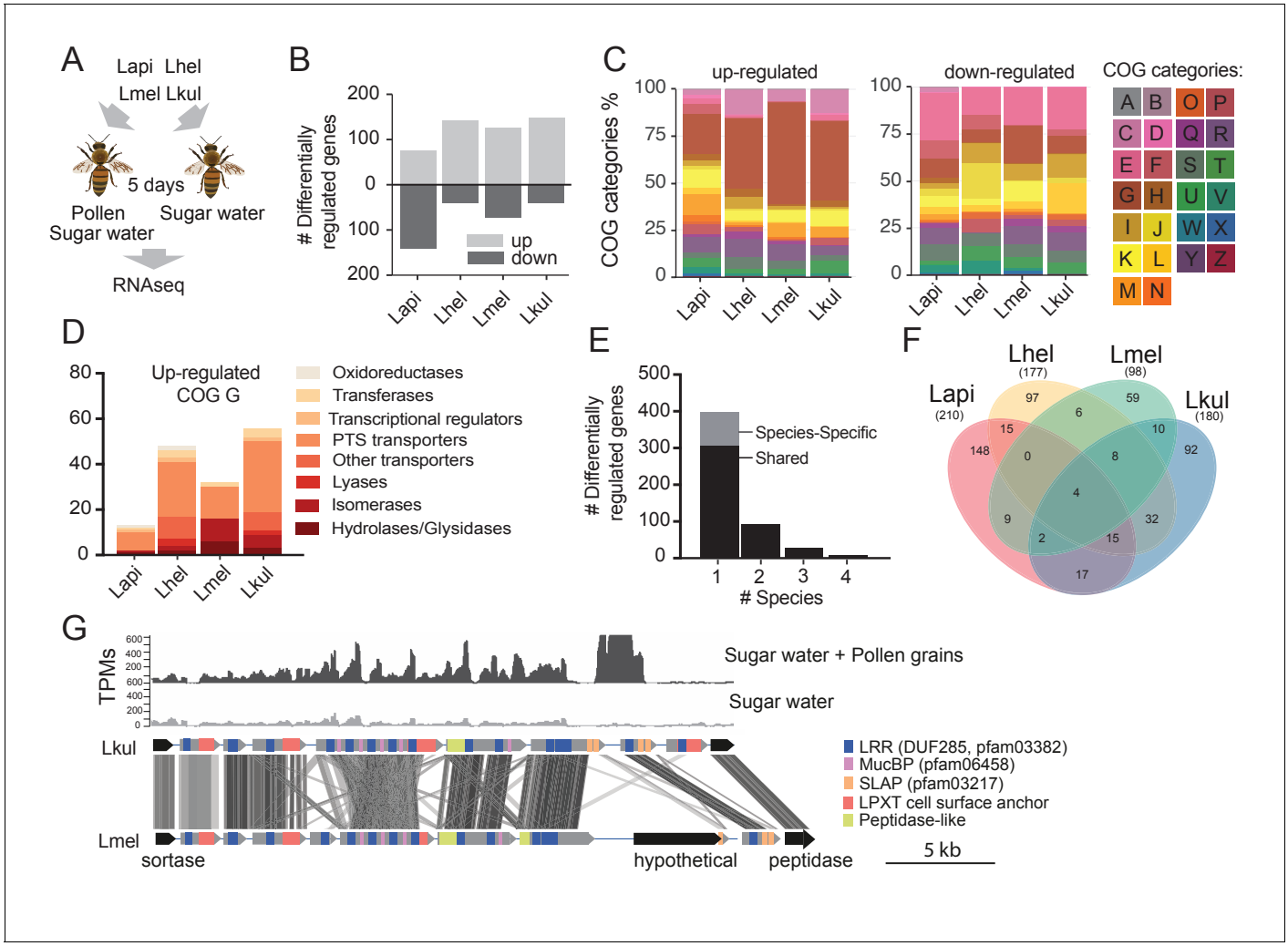


Figure 3. Transcriptome analysis of the four *Lactobacillus* species during co-colonization of gnotobiotic bees. (**A**) Schematic outline of the RNA-Seq experiment. (**B**) Number of differentially regulated genes (log2FC $\geq$ |2| and p-value $\leq$ 0.01) in each species during co-colonization of gnotobiotic bees fed either pollen and sugar water (PG+SW) or sugar water only (SW). Up- and down-regulated genes are shown in different gray tones. (**C**) COG categories of genes up- or down-regulated by the four species in SW+PG if compared to SW. For COG definitions, see ***Supplementary file 2*** (**D**) Functional sub-categories of COG 'G' genes upregulated in SW+PG if compared to SW. (**E**) Barplot displaying numbers of gene families differentially regulated in one species, two species, three species, or four species. Gene families differentially regulated in only one species are split into those that have homologs in the other species or that are species-specific.(**F**) Venn diagram showing overlap of gene families (based on gene homology) differentially regulated in the four species. (**G**) Transcripts per million (TPM) for two representative samples of the SW+PG and the SW treatments over a genomic region of Lkul encoding *Lactobacillus*-specific surface proteins. The genomic region of Lkul is compared to a similar region identified in Lmel which is also differentially regulated across the two treatments (expression profile not shown). Similarity between genes is shown by vertical lines. Gray tones indicate level of similarity. Surface protein-encoding genes are show in grey with the different domains and motifs shown according to the color legend.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. MDS plots of in vivo RNA-seq samples.

We found a total 687 genes (181 to 217 genes per species) to be differentially expressed ($log_2FC \geq |2|$ and p-value ≤ 0.01) between the two dietary treatments (***Figure 3B***). 'Carbohydrate transport and metabolism' (Cluster of orthologous group category G, COG G) was by far the most abundant functional category among the genes upregulated in the SW+PG treatment relative to the SW treatment (***Figure 3C***, 17.1–37.6% of all upregulated genes). In three of the four species (Lmel, Lhel, and Lkul), this category was significantly enriched among the upregulated genes (Fisher's exact test, p<0.01, ***Supplementary file 6***). The largest fraction of the upregulated COG G genes encoded PTS

transporters (***Figure 3D***, ***Supplementary file 5***), followed by other sugar transporters (e.g. ABC transporters), and enzymes involved in sugar cleavage and conversion (***Figure 3D***). Among the downregulated genes, COG G genes were not abundant (5.1–7.8%) (***Figure 3C***). Instead, the category 'Amino acid metabolism and transport' (COG E) was enriched in Lapi (Fisher's exact test, $p < 0.01$, ***Supplementary file 8***), and genes encoding ABC-type amino acid transporters were present among the downregulated genes in all species (***Supplementary file 5***).

We next clustered all genes by homology into gene families. While most of the differentially expressed genes (89%) belonged to gene families with homologs in multiple species, differential expression was typically observed for just one of the species (***Figure 3E–F***). This suggests that the presence of pollen triggers distinct transcriptional changes in the four species during gut colonization. Indeed, gene annotation analysis allowed us to identify several species-specific metabolic functions among the differentially regulated genes (***Figure 4***, ***Supplementary file 5***). For example, Lhel specifically upregulated three PTS gene clusters for the uptake and metabolism of sugar alcohols and one gene cluster for ribose utilization. In contrast, Lmel upregulated several gene clusters involved in the cleavage of xylose, mannose, rhamnose, and arabinose from polysaccharides or other glycosylated compounds. Lmel also upregulated a gene cluster for the synthesis and the transport of bacteriocins in the presence of pollen. Lkul upregulated a starch utilization gene cluster, which in

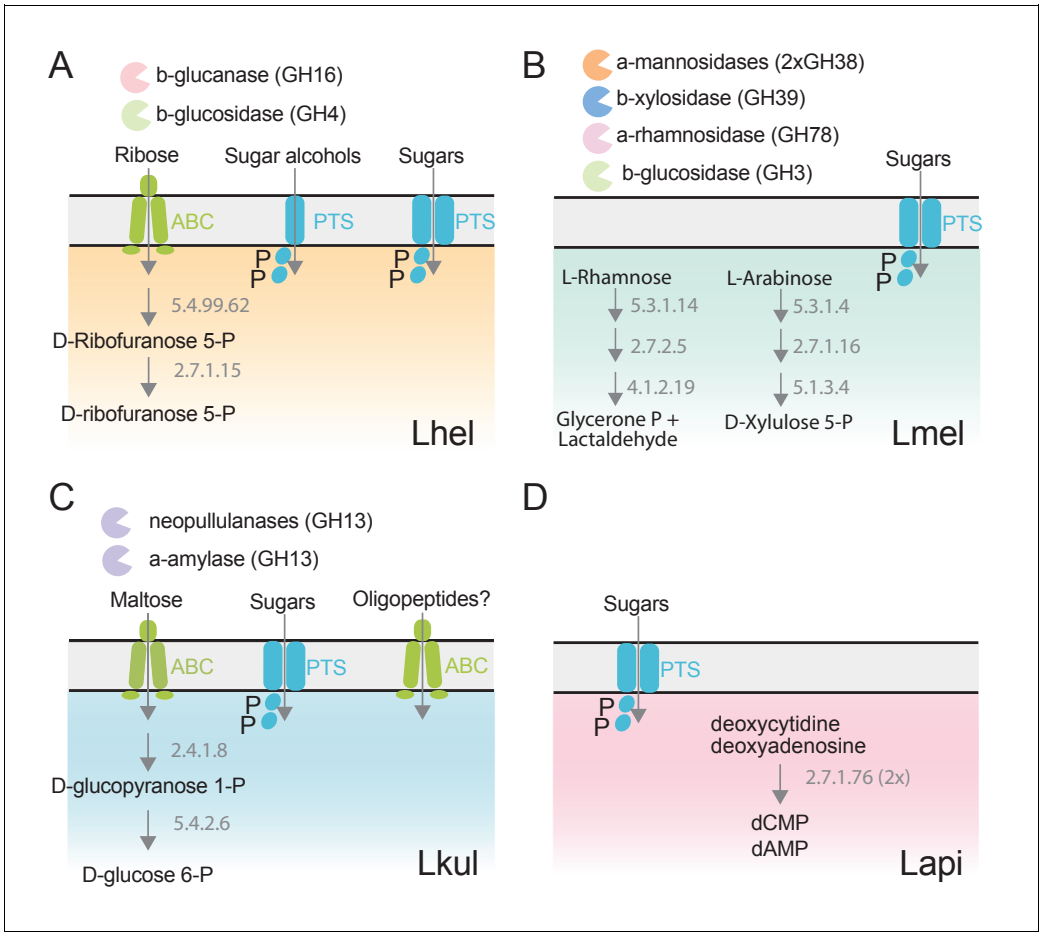


Figure 4. The four *Lactobacillus* species upregulate different carbohydrate metabolism functions during gut colonization of gnotobiotic bees. (**A**) Lhel, (**B**) Lmel, (**C**) Lkul, (**D**) Lapi. Only enzymes and transporters that are upregulated in a species-specific manner in the pollen treatment versus the sugar water treatment are shown. The figure is not exhaustive, but highlights the main differences that could be identified based on gene annotations among all differentially regulated genes (***Supplementary file 6***). Glycosidases belonging to different CAZyme families are represented by different colors. ABC: ABC transporters, PTS: phosphotransferase system transporters. Numbers indicate EC numbers of upregulated enzymatic steps.

part was also differentially regulated in Lmel. In addition, this species upregulated an oligopeptide transporter gene cluster that was present in some of the other strains but not differentially regulated. The fourth species, Lapi, also differentially expressed genes belonging to COG 'G' (mainly PTS transporters), but fewer ones, and with similar functional annotations as found in the other three species. However, Lapi was the only species that upregulated two conserved deoxycytidine kinase genes encoding enzymes involved in nucleoside salvage pathways.

Besides these species-specific transcriptional changes, a number of interesting functions were differentially regulated in more than one species. For example, we found evidence for citrate fermentation in Lhel and Lkul. Both species upregulated genes encoding a citrate lyase for the conversion of citrate into oxaloacetate and acetate in the presence of pollen (***Supplementary file 5***). Lhel, Lmel, and Lkul upregulated genes for the uptake and metabolism of glycerol. Moreover, all four species upregulated gene clusters encoding surface proteins with leucine-rich repeat (LRR) regions, LPXT cell-wall anchoring motifs, and SLAP (S-layer associated protein) domains (***Figure 3G***).

Altogether, these results suggest that the four species utilize different carbohydrate-related resources from pollen, which supports the niche partitioning hypothesis as the basis for coexistence.

Transcriptional responses to pollen are similar in vivo and in vitro

In vivo gene expression differences between the two dietary conditions could be influenced by the host or by bacteria-bacteria interactions. Therefore, we carried out an additional transcriptomics analysis to disentangle the contribution of each of these factors to transcriptional changes in the four *Lactobacillus* species. We grew the four species in vitro in either co-culture or mono-culture, and with either pollen extract (PE) or glucose as growth substrate (G) (***Figure 5A***). As for the in vivo RNA-Seq analysis, MDS plots of the normalized read counts indicated that the four species exhibit treatment-specific transcriptional responses (***Figure 5—figure supplement 1***).

For each species, whether grown alone or in co-culture, we found between 159 and 393 genes to be differentially regulated between the PE and the G treatment (***Figure 5B***, $\log_2FC \geq |2|$ and p-value$\leq$0.01). As in vivo, Carbohydrate transport and metabolism (COG 'G') was the predominant functional category among the upregulated genes in the presence of pollen (***Figure 5C***) and enriched in all eight comparisons (four species, each alone or in co-culture, Fisher's exact test p-value < 0.01, ***Supplementary file 8***). Moreover, 25.3–36.9% of the genes upregulated in vivo were also upregulated in vitro in the presence of pollen. In particular, the species-specific carbohydrate metabolism functions described above (***Figure 4***) showed a similar transcriptional response to pollen in vivo and in vitro (***Figure 5D***). In contrast, most of the putative adhesin genes upregulated in vivo were not upregulated in vitro during growth in pollen or had relatively low transcripts per million (TPM). This suggests that these genes are either expressed in response to the host environment, or the presence of entire pollen grains or sugar water, both of which were only included in the in vivo but not in the in vitro experiment (***Supplementary file 9***). It is also noteworthy that fewer genes were downregulated than upregulated in pollen relative to glucose, and that the COG category 'G' was not enriched among the downregulated genes, which is concordant with our in vivo transcriptome analysis. (***Supplementary file 8***). Based on these results, we conclude that each species upregulates specific operons for the transport and utilization of different carbohydrates (e.g. sugar alcohols and glycans) in response to the presence of pollen, independent of the host environment.

The presence of other community members has little impact on the transcriptional profile of the four species

We found that a large fraction of the genes upregulated in PE relative to G in the mono-cultures were also upregulated in the co-cultures (58.2–87.8%, ***Figure 5E***). In particular, the gene clusters identified to be regulated in a species-specific manner (see above) showed highly concordant gene expression profiles in vitro independent of the presence/absence of the other *Lactobacillus* species. This was confirmed by the direct comparison of mono-culture and co-culture conditions. In comparison to the nutritional treatments, fewer genes (9–149 genes) were differentially expressed between co-culture and mono-culture treatments ($\log_2FC \geq |2|$ and p-value$\leq$0.01), (***Figure 5F***).

We could not find any consistent pattern across the four species in terms of COG category enrichment (***Supplementary file 8***). Moreover, only a few genes were differentially expressed in more than one species (6.25–30%), or across both nutrient conditions (1.86–5.33%). Citrate

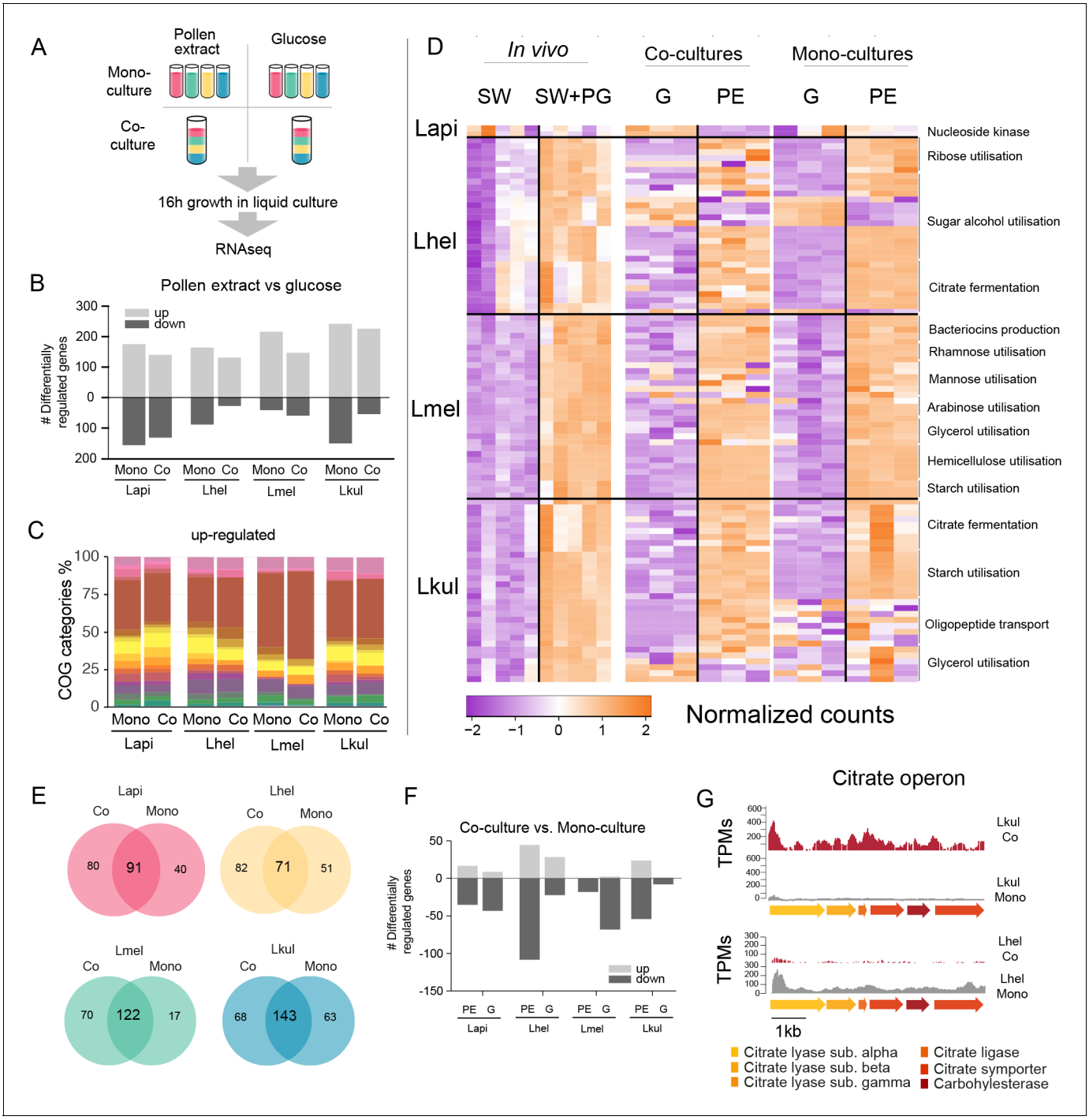


Figure 5. Transcriptome analysis of the four *Lactobacillus* species grown in vitro in pollen extract or in glucose. (**A**) Scheme of the 2x2 experimental design. Species were grown alone or together, in either glucose (**G**) or pollen extract (PE). (**B**) Number of differentially regulated genes in each of the four species in the presence of PE if compared to G. Mono, mono-culture, Co, co-culture. Up- and down-regulated genes are shown in different gray tones. (**C**) COG categories of genes up-regulated by the four species in the presence of PE if compared to G. The colors are the same as in ***Figure 3C***. For COG definitions, see ***Supplementary file 2*** Heatmap displaying normalized counts of selected genes differentially regulated across the in vivo and in vitro RNA-Seq experiments. We selected metabolic genes and gene clusters that were identified in the in vivo experiment to be differentially regulated across the two treatments and which could be assigned a putative function based on annotation. Counts were normalized for each gene and dataset separately, that is in vivo, co-cultures, and mono-cultures. (**E**) Venn diagrams displaying the overlap of the genes differentially regulated between the PE and G treatment when the four species were grown in co-culture (Co) and mono-culture (Mono). (**F**) Number of differentially regulated genes in each of the four species in co-culture relative to mono-culture. Up- and down-regulated genes are shown in different gray tones. (**G**)

Figure 5 continued on next page

Figure 5 continued

Transcripts per million (TPM) over a genomic region of Lkul and Lmel encoding genes for citrate fermentation (i.e. citrate operon) for a representative sample of the co-culture and the mono-culture treatment when grown in PE.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. MDS plots of in vitro RNA-seq samples.

fermentation genes were upregulated in Lkul in co-culture relative to mono-culture when grown in pollen, whereas in Lhel the opposite was observed (***Figure 5G***). Also of note, the oligopeptide transporter system which was upregulated in vivo in Lkul in the presence of pollen, was also upregulated in vitro in the presence of pollen, but only when other species were present. These two specific examples show that a few metabolic functions are differentially regulated in response to other bacteria, but not always in the same direction across species, or only in a specific nutrient condition. We thus conclude that the main factor driving changes in gene expression in the four strains is the presence of pollen, rather than the presence of other *Lactobacillus* species.

Metabolomics analysis reveals differences in flavonoid and sugar metabolism across the four *Lactobacillus* species

Our transcriptome analyses suggest that differences in sugar metabolism may enable the four species to coexist in the presence of pollen in vitro and in vivo. To assess species-specific metabolic changes when grown in pollen, we profiled the metabolome of the pollen extract medium before (t = 0 hr) and after bacterial growth (t = 16 hr) using Q-TOF-based untargeted metabolomics (***Fuhrer et al., 2011***). We annotated a total of 657 ions of which 406 could be reliably categorized as pollen-derived ions, as opposed to ions originating from the base medium (see Materials and methods, ***Supplementary file 10***, ***Figure 6—figure supplement 2***). The metabolomics data clearly separated the four species indicating distinctive metabolic changes and thus corroborating the transcriptome results (***Figure 6—figure supplement 1***). A total of 76 pollen-derived ions showed a significant decrease in abundance over-time (log2FC ≤ -1 and p-value$\leq$0.01, Student's t-Test, BH correction) (***Figure 6A***, ***Supplementary file 10***). Of those, 24 ions decreased in abundance in all four species, another 24 ions decreased in abundance in only a subset of the species, and the remaining 28 ions decreased in abundance in only a single species (***Figure 6A***). Ions annotated as glycosylated flavonoids were among the top ions responsible for the separation of the four species in the PCA (***Figure 6—figure supplement 1***). Lmel depleted six different ions annotated as flavonoids (isoorientin 2''-O-rhamnoside, quercetin-3-O-glucoside, vitexin, rutin, luteolin-7-O-(6''-malonylglucoside), and quercetin-3-O-beta-D-glucosyl-(1->2)-beta-D-glucoside), while Lapi depleted three ions annotated as flavonoids (isoorientin 2''-O-rhamnoside, quercetin-3-O-glucoside, vitexin) (***Figure 6A***, ***Figure 6—figure supplement 3***). In contrast, Lkul only depleted the flavonoid ion annotated as isoorientin 2''-O-rhamnoside, and no flavonoid ion changes were identified for Lhel (***Figure 6A***, ***Figure 6—figure supplement 3***).

To corroborate the species-specific utilization of flavonoids, we incubated each of the four species in base culture medium supplemented with rutin. We observed the formation of a yellow insoluble precipitate only in the wells incubated with Lmel (***Figure 6B***). Metabolomics analysis confirmed that rutin was depleted in these wells and that the yellow precipitate corresponded to an accumulation of quercetin, the water-insoluble, deglycosylated aglycone of rutin (***Figure 6C***). These findings are consistent with our transcriptome results which show that Lmel is the only species that upregulated a rhamnosidase gene known to cleave rhamnose residue from rutin (***Beekwilder et al., 2009***; ***Figure 4***).

Other ions with species-specific abundance changes included a plant-derived glycosylated compound belonging to the iridioids class (i.e. antirrhinoside, depleted in the presence of Lapi), a component of the outer pollen wall (i.e. 9,10,18-trihydroxystearate, accumulated in the presence of Lmel) and cyclic nucleotides (depleted in the presence of Lhel, Lmel, and Lkul) (***Figure 6A*** and ***Figure 6—figure supplement 3***). Lhel was the only species depleting an ion corresponding to sugar alcohols (mannitol, D-sorbitol, or L-glucitol) (***Figure 6A*** and ***Figure 6—figure supplement 3***) consistent with the specific upregulation of sugar alcohol PTS transporters in this species (***Figure 4***).

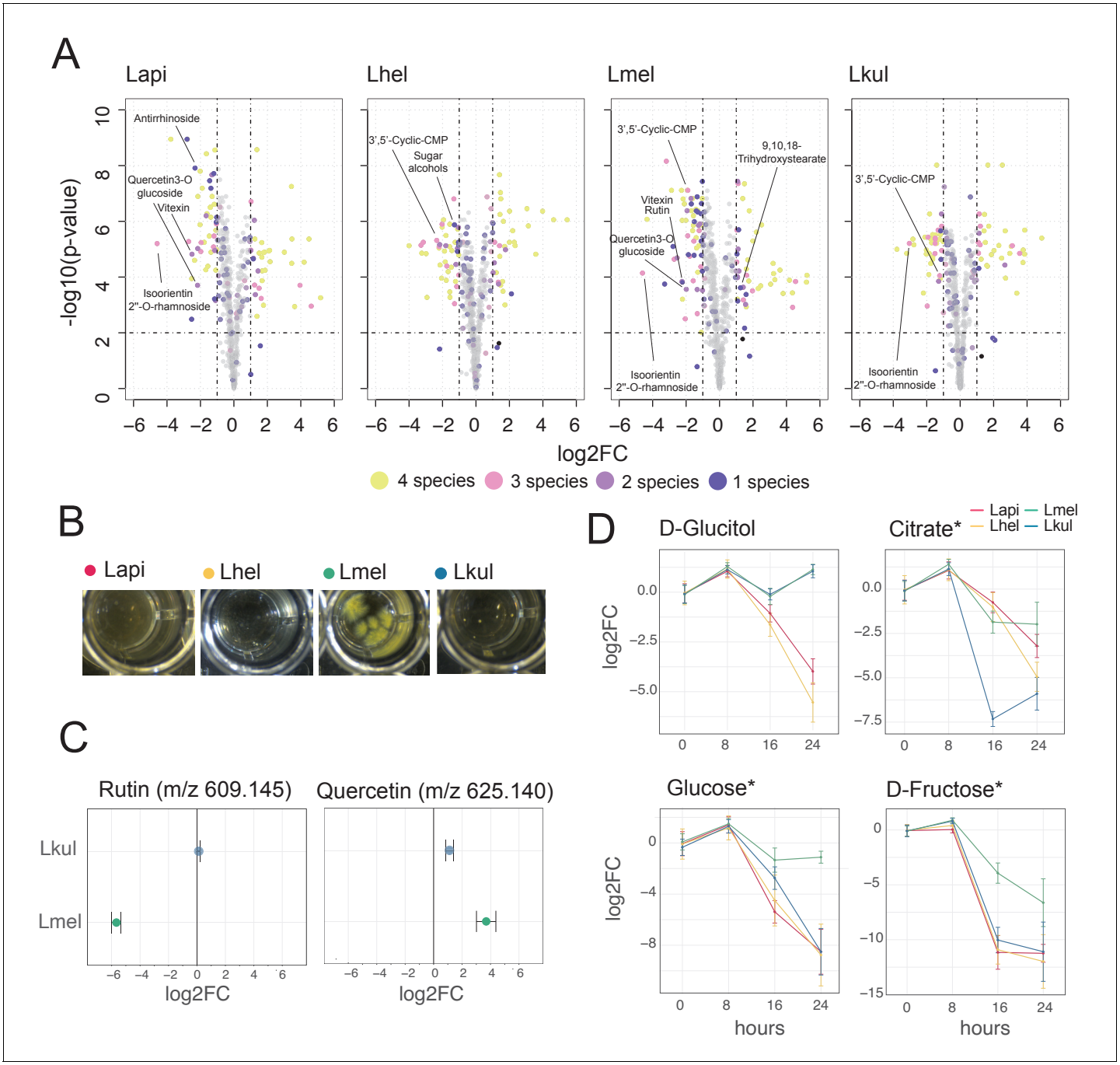


Figure 6. Metabolomics analysis shows differences in the utilization of pollen-derived glycosides across the four *Lactobacillus* species. (**A**) Volcano plots displaying ions with significant fold changes (FC) for each of the four species after 16 hr of growth in pollen extract versus glucose. Each dot corresponds to an ion in the untargeted metabolomics dataset. Different colors represent ions that significantly change over time in one, two, three, or four species. Dashed black lines represent the significance thresholds: p-value < 0.01 and log2FC < −1 or > 1. (**B**) Culture wells of the four species grown in cfMRS + 0.05% rutin after 16 hr of incubation. The yellow precipitate is only visible for Lmel. (**C**) Rutin and quercetin detection in spent medium of Lmel and Lkul grown in cfMRS + 0.05% rutin after 16 hr of incubation (n=5). (**D**) Changes in key metabolites during growth measured by GC-MS (n=5). Log2FC relative to time point 0 is plotted. Time is reported in hours. * Indicates metabolites whose identity was confirmed using analytical standards. For m/z values see ***Supplementary file 10***.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. PCA in vitro metabolomics.

Figure supplement 2. Definition of pollen-derived ions.

Figure supplement 3. Untargeted metabolomics: key metabolites discussed in the main text.

Figure 6 continued on next page

Figure 6 continued

Figure supplement 4. GC-MS detection of key metabolites over time.
Figure supplement 5. Logistic regression growth curve of the four species.

Based on the untargeted metabolomics analysis, we conclude that the four species target different metabolites, in particular secondary plant metabolites present in pollen. In order to assess differences in the utilization of simple sugars and acids in more detail, we analyzed the supernatants of cultures of the four strains after 0, 8, 16, and 24 hr of growth using GC-MS. We used a semi-targeted approach, where we identified a subset of metabolites by preparing analytical standards and the others by using a reference library (see Materials and methods). We identified 113 metabolites of which 46 showed a significant change in abundance in at least one strain between timepoint 0 hr and 24 hr ($\log_2$FC $\geq$ |2| and p-value $\leq$ 0.01, Student's t-Test, BH correction) (***Supplementary file 10***). All four species showed mixed substrate utilization, that is they utilized several substrates simultaneously. Moreover, most substrates were utilized by all four species, but often at different rates. Many metabolites that we identified with the GC-MS had annotations comparable to the ones found in the Q-TOF-based experiment. For example, we detected a compound annotated as sugar alcohol, that is glucitol, that was consumed most efficiently by Lhel as observed in the previous analysis (***Figure 6C***). Moreover, all four species consumed the carboxylic acids citrate and malate (***Figure 6C*** and ***Figure 6—figure supplement 4***), which corresponded with the results of the Q-TOF-based experiment. Interestingly, Lkul and Lhel consumed citrate at the fastest rate and they were also the two species that upregulated gene clusters for citrate fermentation in the presence of pollen in vivo and in vitro (***Figure 4***).

Lmel consumed several simple monosaccharides (such as glucose, fructose, allose, and mannose) at a slower rate than the other species, although having a similar growth profile (***Figure 6C***, ***Figure 6—figure supplements 4–5***). This could indicate that Lmel has specialized in the metabolism of pollen-derived glycosylated compounds (such as rutin, ***Figure 6B–C***) at the expense of fast consumption of generic substrates, which accords with the upregulation of several gene clusters for the cleavage of such sugars from polysaccharides or other glycosides (e.g. flavonoids) in presence of pollen (***Figure 4***).

In summary, our metabolomics results show that the four species specialize in the utilization of different pollen-derived compounds, and that the observed metabolite changes are to some extent consistent with the transcriptional changes observed in the presence of pollen relative to the presence of simple sugars.

Discussion

Ecological processes governing the coexistence of microbes have been probed in the laboratory using microbial communities of different complexity (***Goldford et al., 2018***; ***Ortiz et al., 2021***; ***Wright and Vetsigian, 2016***; ***Friedman et al., 2017***; ***Piccardi et al., 2019***; ***Deines et al., 2020***; ***Logan, 2017***). However, few studies have examined the impact of the host on the coexistence of bacterial symbionts of animals (***Ortiz et al., 2021***; ***Deines et al., 2020***). In particular, little is known about the extent to which closely related species and strains can be stably maintained (***Bittleston et al., 2019***). We capitalized on the experimental tractability of honey bees and their gut microbiota and used a bottom-up approach to study the coexistence of four closely related, naturally co-occurring *Lactobacillus* species. We disentangled the effect of the diet and the host on the interactions between the four species by serially passaging them through gnotobiotic bees or in culture tubes, under two nutrient conditions (pollen versus simple sugars). Our results show that the dynamics in the four-species community is governed by negative interactions, because the growth of each member was lower in co-culture than in mono-culture, independent of the environment (host or culture tube) and the nutrient condition (pollen or simple sugars). This is consistent with previous observations that negative interactions predominate in nutrient-rich environments (***Piccardi et al., 2019***; ***Foster and Bell, 2012***; ***Berry and Widder, 2014***; ***Coyte et al., 2015***; ***Ghoul and Mitri, 2016***). Moreover, the four *Lactobacillus* species harbor relatively small genomes (1.5–2 Mb) with a

conserved and streamlined core metabolism and similar auxotrophies, suggesting overlapping nutritional requirements (**Ellegaard et al., 2019**; **Ellegaard et al., 2015**; **Kwong and Moran, 2016**).

The coexistence of bacterial symbionts can be facilitated by the host, for example by providing a spatially structured environment that results in the physical separation of competing strains (**Gude et al., 2020**; **Kim et al., 2008**; **Mitri et al., 2016**; **Hallatschek et al., 2007**), or by secreting metabolites that support niche specialization (**Schluter and Foster, 2012**; **McLoughlin et al., 2016**). However, in the case of the four *Lactobacillus* species, such host-related features seem not to be sufficient to support coexistence, because the four-species community was rapidly dominated by a single species, when passaged through gnotobiotic bees that were fed a simple sugar diet. In contrast, when providing a more diverse nutrition in the form of pollen, we found that the four species were stably maintained both in vivo and in vitro. We thus conclude that the coexistence of the four *Lactobacillus* species in the honey bee gut primarily depends on the pollen diet of the host and not the host environment itself.

The challenges in replicating the native environment such that it is possible to study relevant interactions of host-associated microbes in vitro are formidable. These were highlighted in a recent study on the microbial community associated with the freshwater polyp hydra that could not recapitulate the coexistence of the dominant microbiota members in vitro (**Deines et al., 2020**). Here, we aimed to approximate the nutritional conditions in the honey bee gut by culturing the bacteria in pollen infused media, that is the natural diet of bees. In both the in vivo and in vitro transfer experiment, we assessed the effect of pollen on the dynamics of the community by comparing it to a simple sugar treatment. Although not identical, the nutritional conditions in vitro were sufficiently similar as to recapitulate the overall community dynamics observed in vivo: pollen nutrients supported the stable coexistence of the four species, while the simple sugars led to the dominance of a single species. As the bee and members of the bee gut microbiota pre-digest pollen and sugars upstream of the rectum, it is difficult to exactly replicate the metabolic environment of the rectum. For example, sucrose is largely absorbed via the midgut epithelium and cleaved into glucose and fructose by host enzymes, while fermentative bacteria such as *Gilliamella apicola* degrade and modify a diverse range of carbohydrates in the ileum (**Kešnerová et al., 2017**; **Crailsheim, 1988**). These metabolic alterations may explain some of the differences observed between the in vivo and in vitro experiments, such as the dominance of different species in the simple sugar conditions (sucrose and glucose, respectively). We therefore suspect that different species would dominate in vitro or in vivo with an alternative simple sugar composition.

Our findings are consistent with the consumer-resource model, which predicts that the number of species that can coexist depends on the number of available resources (**Tilman, 1986**). Correspondingly, in the presence of a single substrate, such as in the case of glucose in vitro, competition for the same nutrient results in the competitive exclusion of all but one species. However, depending on the nutrient availability, the dietary transit time, the crosstalk with the host, or the spatial structure of the gut, the ecological processes governing bacterial coexistence may differ across host-associated microbiomes. For example, the *Lactobacillus* species of the honey bee gut microbiota primarily colonize the luminal space of the rectum, where partially digested pollen accumulates. In contrast, some of the Proteobacteria of the bee gut microbiota adhere to the epithelial surface of the ileum (**Zheng et al., 2018**). We expect that in the latter case interactions with the host play a more important role for microbial coexistence than in the case of the Lactobacilli in the rectum.

Although ecological interactions in bacterial communities have been investigated across a wide range of experimental systems, few studies have tackled the molecular mechanisms underlying coexistence. In some cases, cross-feeding of metabolic by-products facilitates the maintenance of diversity in bacterial communities, such as after passaging leaf and soil samples in single carbon sources (**Goldford et al., 2018**). However, cross-feeding does not seem to play an important role in maintaining coexistence of the four *Lactobacillus* species in this study. Unlike the above example, feeding a single carbon source led to the extinction of all but one species. Our metabolomics analysis also did not reveal any major metabolites that could potentially be cross-fed, that is were produced by one species and utilized by another. Finally, we identified no transcriptional changes that would suggest cross-feeding activities when comparing mono-cultures and co-cultures of the four *Lactobacillus* species.

Instead, our combined transcriptomics and metabolomics analyses suggest that coexistence is facilitated by specialization toward distinct pollen-derived nutrients. We found that all four species

upregulated carbohydrate transport and metabolism functions dedicated to the utilization of different carbon sources in the presence of pollen when colonizing the bee gut, and these changes were reproducible in vitro. Our metabolomics analysis identified a number of pollen-derived glycosides that were utilized in a species-specific manner. In particular, Lmel specialized in the utilization of flavonoids at the expense of simple sugars, which may explain why this species rapidly went extinct in presence of only simple sugars during the transfer experiments. While the importance of pollen-derived flavonoids in niche partitioning needs to be validated, the species-specific deglycosylation of these secondary plant compounds illustrates that the four species have different hydrolytic capabilities that may also be involved in the cleavage of other carbohydrates. The metabolic specialization on plant glycans may be a common phenomenon in animal gut communities, as similar transcriptional changes have been described in other gut symbionts when the host diet was supplemented with specific plant glycans (**Sonnenburg et al., 2005**; **Zheng et al., 2019**).

In contrast to the species specific metabolism of glycoside, we observed few differences in the utilization of simple saccharides among the four species in our time-resolved GC-MS analysis. While this may seem surprising, theoretical work has established that resource preference for at least one substrate is sufficient to explain coexistence (**Meszéna et al., 2006**). Moreover, it is plausible that the four species utilize the same sugars, but extract them from different pollen-derived glycans, such as starch, hemicellulose, flavonoids, or other glycosylated secondary plant metabolites.

While this work focused on niche partitioning based on degradation of complex carbohydrates, it is noteworthy that all four *Lactobacillus* species engaged to a variable extent in co-fermentation of the carboxylic acids citrate and malate present in pollen. The two species, Lkul and Lhel, that upregulated citrate fermentation pathways in the presence of pollen also consumed citrate at the fastest rate. Citrate co-fermentation has been linked to competitive advantages in lactic acid bacteria, though whether the varying levels of co-fermentation contribute to colonization stability in this system remains an outstanding question (**Laëtitia et al., 2014**; **Magni et al., 1999**; **Jimeno et al., 1995**).

Previous work suggested that the large diversity of carbohydrate transport and metabolism functions in the accessory gene pool of *Lactobacillus* Firm5 is an adaptation to the pollen-based diet of the host and a consequence of the nutrient competition with closely related species (**Ellegaard and Engel, 2019**; **Ellegaard et al., 2019**). Our findings support this hypothesis and provide the first experimental evidence for a link between the coexistence of the four *Lactobacillus* species, the large diversity of carbohydrate metabolism functions in their genomes, and the pollen diet of the host. Moreover, these results suggest that dietary differences between host species or natural variation in pollen diversity influence the diversity of *Lactobacillus* Firm5 and could, for example explain why the Asian honey bee, *Apis cerana,* harbors only one species of this phylotype in its gut (**Ellegaard et al., 2020**).

However, we have only tested a single strain of each of the four species. Therefore, given the extensive genomic diversity within these species (**Ellegaard and Engel, 2019**), more work is needed to determine if the identified patterns of coexistence reflect stable ecological niches occupied by the four species or are rather the result of the specific strains selected for our experiments. In a recent study on pitcher plant microbiomes, it was shown that even strains that differ by only a few base pairs can have different ecological trajectories in communities and coexist over extended period of time (**Bittleston et al., 2019**). Expanding our approach to strains within species presents an exciting next step to understand at which level discrete ecological niches are defined in the bee gut and how diversity can be maintained in such ecosystems.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background (*Lactobacillus apis*)	Lapi	https://doi.org/10.1371/journal.pbio.2003467	Genome ID: 2684622912	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background (*Lactobacillus helsingborgensis*)	Lhel	https://doi.org/10.1371/journal.pbio.2003467	Genome ID: 2684622914	
Strain, strain background (*Lactobacillus melliventris*)	Lmel	https://doi.org/10.1371/journal.pbio.2003467	Genome ID: 2684622913	
Strain, strain background (*Lactobacillus kullabergensis*)	Lkul	https://doi.org/10.1371/journal.pbio.2003467	Genome ID: 2684622911	
Commercial assay or kit	QIAquick Gel Extraction Kit	Qiagen	#Cat 28706X4	
Commercial assay or kit	Nucleospin RNA clean-up kit	Macherey-Nagel	#Cat 740903	
Commercial assay or kit	Zymo-Seq RiboFree Total RNA Library kit	Zymo Research	#Cat R3000	
Software, algorithm	R Studio software	R Studio (https://www.rstudio.com)	RRID:SCR_000432	
Software, algorithm	Integrative Genomics Viewer	Integrative Genomics Viewer (https://software.broadinstitute.org/software/igv/)	RRID:SCR_011793	

Culturing of bacterial strains

We used the following four bacterial strains of Lhel, Lmel, Lapi, and Lkul for our experiments: ESL0183, ESL0184, ESL0185, and ESL0186 (***Kešnerová et al., 2017***). All strains were precultured on solid De Man – Rogosa – Sharpe agar (MRSA) (supplemented with 2% w/v fructose and 0.2% w/v L-cysteine-HCl) from glycerol stocks stored at −80˚C. MRSA plates were incubated for three days in anaerobic conditions at 34˚C to obtain single colonies. Single colonies were inoculated into a liquid carbohydrate-free MRS medium (cfMRS; ***O' Donnell et al., 2011***) supplemented with 4% glucose (w/v), 4% fructose (w/v), and 1% L-cysteine-HCl (w/v) and incubated at 34˚C in anaerobic conditions without shaking.

In vivo transfer experiments

Bacterial colonization stocks were prepared from overnight cultures by washing the bacteria in 1xPBS, diluting them to an OD_{600} = 1, and storing them in 25% glycerol at −80˚C until further use. For colonization stocks containing all four species, cultures adjusted to an OD_{600} = one were mixed at equal proportions. Microbiota-depleted bees were obtained from colonies of *Apis mellifera carnica* located at the University of Lausanne following the procedure described in ***Kešnerová et al., 2017***. Colonization stocks were diluted ten times in a 1:1 mixture of 1xPBS and sugar water (50% sucrose solution, w/v) and 5 µL were fed to each bee using a pipette. Five days post-colonization, 10 rectums were dissected and homogenized in 1xPBS. An aliquot of each homogenized gut was used for CFU plating to enumerate the total bacterial load and for amplicon sequencing to obtain the relative abundance of each community member (see below). To serial passage the community through microbiota-depleted bees, the ten homogenized gut samples from the same treatment were pooled together and stored in 25% glycerol at −80˚C until a new batch of microbiota-depleted bees was available. At the day of colonization, a frozen aliquot of the pooled gut homogenate was thawed, diluted ten times in a 1:1 mixture of 1xPBS and sugar water (50% sucrose solution, w/v), and fed to newly emerged microbiota-depleted bee as described above. This was repeated for a total of six serial passages. Throughout the experiments all bees were kept on either a sugar water or a sugar water/pollen diet according to the two dietary treatment. Food was provided ad libitum.

In vitro transfer experiment

Each of the four strains was cultured in liquid medium overnight for about 16 hr as described above. The cultures were re-inoculated at an OD_{600} = 0.3 in fresh medium and let grow for another 4 hr at

34˚C with shaking (700 rpm). Bacterial cells were then washed with 1xPBS, mixed in equal proportions, and inoculated at an OD_{600} = 0.05 in triplicates in 96-deep well plates (SIGMA) containing cfMRS medium supplemented with either 2% glucose (w/v), 10% pollen extract (v/v), or 10% pollen grains (v/v) in a final volume of 500 µL per well. Detailed information about pollen extract preparation can be found in the Supporting methods section of ***Kešnerová et al., 2017***. Pollen grain solutions were prepared by adding 1.250 ml of ddH_2O to 80 mg of pollen grains crushed with the bottom of a 15 mL falcon tube. The plates were incubated for 24 hr at 34˚C under anaerobic conditions without shaking (300 rpm). After 24 hr of incubation, an aliquot of each sample was subjected to CFU plating to enumerate the total bacterial load. Then, 1% of each culture (i.e. 5 µL) was transferred to a plate with fresh medium supplemented with the appropriate carbon sources and incubated again. These transfers were repeated 10, respectively, 20 times for the two independent experiments. After each transfer, cultures were washed once with 1xPBS and stored at −20˚C for amplicon sequencing analysis. CFUs were counted after 24 hr and at the final transfer.

Amplicon sequencing

The relative abundance of the four strains across all transfer experiments was obtained using amplicon sequencing of a 199 bp long fragment of a housekeeping gene encoding a DNA formamidopyrimidine-glycosylase which allows to discriminate the four strains from each other (***Ellegaard et al., 2019***).

For the in vitro transfer experiments, the PCR fragment was amplified from crude cell lysates. They were generated by mixing 5 µL of culture with 50 µL of lysis solution, containing 45 µL of lysis buffer (10 mM Tris- HCl, 1 mM EDTA, 0.1% Triton, pH 8), 2.5 µL of lysozyme (20 mg/ml, Fluka), and 2.5 µL of Proteinase K solution (10 mg/ml, Roth). The samples were incubated for 10 min at 37˚C, for 20 min at 55 ˚C, and for 10 min at 95 ˚C, followed by a short spin before preparing the PCR (1 min, 1500 rpm). For the in vivo transfer experiment, DNA was isolated from the homogenized gut samples using the hot phenol protocol used in ***Kešnerová et al., 2017***.

To amplify the gene fragment and to add the Illumina barcodes and adapters, the two-step PCR strategy published in ***Ellegaard et al., 2019*** was used. For the first PCR, 5 µL of DNA or 5 µL of cell lysate were mixed with 12.5 µL of GoTaq Colorless Master Mix (Promega), 1 µL of forward and reverse primer (5 µM, see ***Supplementary file 1***) and 5.5 µL of Nuclease-free Water (Promega). The PCR I was performed as follows: initial denaturation (95˚C – 3 min), 30 times denaturation-annealing-extension (95˚C – 30 s, 64˚C – 30 s, 72˚C – 30 s), final extension (72 ˚C – 5 min). To purify the amplicons, 15 µL of PCR product were mixed with 5 µL of a 5X Exo-SAP solution (15% Shrimp Alkaline Phosphatase – 1000 U/ ml – NEB, 10% Exonuclease I – 20,000 U/ ml – NEB, 45% glycerol 80% and 30% dH2O). and incubated for 30 min at 37˚C and for 15 min at 80˚C. For the second PCR reaction, 5 µL of purified PCR products were mixed with the same reagents as before. The PCR program was the same as above with the exception that the annealing temperature was set to 60˚C and the denaturation-annealing-extension steps were repeated for only eight times. The barcoded primers are listed in ***Supplementary file 1***. The amplicons of the second PCR were purified using the Exo-SAP program as described above.

To prepare the sequencing of the amplicons, DNA concentrations were measured using Quant-iT PicoGreen for dsDNA (Invitrogen). Each sample was adjusted to a DNA concentration of 0.5 ng/µL and 5 µL of each sample were pooled together. The pooled sample was loaded on a 0.9% agarose gel and gel-purified using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions. The purified DNA was prepared for sequencing following the Illumina MiniSeq System Guide for 'denaturate and dilute libraries' and then loaded on a Illumina MiniSeq Mid Output Reagent Cartridge using the correspondent MiniSeq flow cell. Illumina reads were demultiplexed by retrieving the unique barcodes of the different samples and quality-filtered using Trimmomatic (Trimmomatic-0.35) (LEADING:28 TRAILING: 29 SLIDING WINDOW:4:15 MINLEN:90). Each forward and reverse read pair was assembled using PEAR (-m 290 n 284 j 4 -q 26 v 10 -b 33) (***Zhang et al., 2014***), and the assembled reads were assigned to the different strains based on base pair positions with discriminatory SNP. See details in Supplementary material.

To obtain absolute abundance data for each strain, we combined the relative abundance data from the amplicon sequencing with CFU counts obtained from plating homogenized bee guts in the case of the in vivo experiments (see above) or by carrying out qPCR with *Lactobacillus*-specific primers as described in ***Kešnerová et al., 2017*** in the case of the in vitro co-culture experiments

(*Supplementary file 1*, *Supplementary file 4*). For the in vitro transfer, the stability of the four-species community over time was calculated using the codyn R package (*Hallett et al., 2020*).

RNA extraction and sequencing

For the in vivo RNA sequencing, microbiota-depleted bees were colonized with the four species community as described above and fed with either sugar water and pollen grains or with sugar water only. After 5 days of colonization, the rectums of five bees per treatment (all kept in the same cage) were dissected and snap-frozen in liquid nitrogen in separate tubes containing glass beads (0.1 mm dia. Zirconia/Silica beads; Carl Roth). For RNA extraction, the tissue samples were suspended in 1 mL of TE buffer and homogenized using a bead beater (45 m/s, 6 s). Then, 200 µL of an ice-cold STOP solution (95% v/v ethanol, 5% v/v Aqua phenol [Roth]) was added to 1 ml of homogenate and snap-frozen again in liquid nitrogen. Tubes were then thawed on ice and a previously developed hot phenol RNA extraction protocol was followed (*Sharma et al., 2010*). For the in vitro RNA sequencing, bacterial strains were cultured in triplicates in cfMRS supplemented with either 1% w/v glucose or 1% w/v pollen extract. After 16 hr of growth, 200 µL of STOP solution was added to 1 mL of culture followed by the same steps as described above.

After the precipitation step, samples were treated with DNaseI (NEB) to degrade DNA. RNA samples were purified using Nucleospin RNA clean-up kit (Macherey-Nagel) following the manufacturer's instructions. RNA was eluted in RNase free-water and stored at −80˚C until further use. RNA concentration and quality were assessed using Nanodrop (ThermoFisher Scientific), Qubit (ThermoFisher Scientific, RNA – High Sensitivity reagents and settings) and Bioanalyzer (Agilent). High-quality RNA samples were selected to prepare RNA libraries. For the in vivo RNA sequencing, libraries were prepared using the Zymo-Seq RiboFree Total RNA Library kit (Zymo Research). The libraries were sequenced by the GTF facility of the University of Lausanne using HiSeq 4000 SR150 sequencing (150 bp reads) (Illumina). For the in vitro RNA sequencing, libraries were prepared following the protocol developed by *Avraham et al., 2016*. Libraries were then prepared for sequencing following the Illumina MiniSeq System guide for denaturate and dilute libraries. Libraries were sequenced using the Illumina MiniSeq technology using High Output Reagent Cartridges (150 bp reads) and MiniSeq flow cells.

RNA sequencing analysis

For the in vitro samples, raw reads were demultiplexed using a script provided by Dr. Jelle Slager (Personal communication) For the in vivo samples, the reads were already demultiplexed by the sequencing facility. For both experiments, the reads were trimmed with Trimmomatic (Trimmomatic-0.35) (LEADING:30 TRAILING: 3 SLIDING WINDOW:4:15 MINLEN:20). The quality of the reads was checked using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For the in vivo samples, trimmed reads were sorted with sortmerna-4.2.0 to select only the non-rRNA reads for the downstream analysis. Reads were mapped onto the genomes of the selected strains (*Ellegaard and Engel, 2018*) (Lapi, Lhel, Lmel, and Lkul) using Bowtie (bowtie2-2.3.2). Gene annotations for the four genomes were retrieved from IMG/mer (*Chen et al., 2021*). Mapped reads were quality filtered for the alignment length (CIGAR > 100 bp) and for the allowed mismatches in the sequence (NM = 0–1). Quality filtered reads were then quantified using HTseq (Version 0.7.2). Differential gene expression between samples cultured in pollen extract and samples cultured in glucose, and between mono-cultures and co-cultures, was calculated using the R package EdgeR (*Robinson et al., 2010*). Counts per million were calculated and only genes with at least one count per million were used for the analysis. EdgeR fits negative binomial models to the data. The counts were normalized for RNA composition by adjusting the $\log_2$FC according to the library size, and the quantile-adjusted conditional maximum likelihood (qCML) method was used to estimate the common dispersion and the tag-wise dispersion. Finally, the differential gene expression was determined using the exact test with a false discovery rate (FDR) <5%. COG annotations were obtained from IMG/mer, and the enrichment analysis for COG categories tested using the Fisher's exact test. Transcripts per million (TPM) were visualized using the Integrated Genome Browser software (*Freese et al., 2016*).

Untargeted metabolomics

Metabolites were extracted from liquid cultures supplemented with 10% (w/v) pollen extract at the inoculation time and after 16 hr of incubation at 34˚C. For each liquid culture sample, 300 µL was collected and centrifuged (20,000 g, 4˚C, 30 min), then 200 µL supernatant was transferred to a new tube and stored at −80˚C. After collection of all samples, they were prepared for metabolomics analysis. The samples were thawed on ice and centrifuged again (20,000 g, 4˚C, 5 min), then diluted 10 times with ddH_2O. For metabolomics analysis, 25 µL of each diluted sample was sent in a 96-well plate on dry ice to the laboratory of Prof. Uwe Sauer for analysis (ETH Zürich, Switzerland). Three replicates of a pollen-extract dilution series (10 serial 2x dilutions) as well as undiluted pollen-extracts and water used for performing the dilution series were included in the metabolomics analysis. Because of the insolubility of flavonoid aglycones in a water matrix, metabolites from liquid cultures supplemented with rutin were extracted using a methanol-extraction protocol at the time of inoculation and after 16 hr of growth by adding 200 µL of methanol pre-cooled to −20˚C to 100 µL of culture. Tubes were vortexed thoroughly and incubated for 5 min (4˚C, shaking 14,000 rpm). Samples where then incubated at −20 ˚C for 1 hr and centrifuged (20'000 g, 5 min). A total of 200 µL of the supernatant was transferred to a new tube and diluted 10 times in 70% methanol and 25 µL of each diluted sample was sent to Zürich in Eppendorf tubes sealed with parafilm on dry ice. For untargeted metabolomics analysis, each sample was injected twice (technical replicate) into an Agilent 6550 time-of-flight mass spectrometer (ESI-iFunnel Q-TOF, Agilent Technologies) as detailed in ***Kešnerová et al., 2017***. In brief, m/z features (ions) were annotated by matching their accurate mass-to-sum formulas of compounds in the KEGG database accounting for deprotonation ($-H^+$). Alternative annotation can be found in ***Supplementary file 10***. When available, metabolites categories were assigned to ions based on KEGG ontology.

Metabolomics data analysis was carried out using R version 3.6.3. Variation of raw ion intensities obtained from untargeted metabolomics analysis for the two technical replicates was determined by assessing the correlation between ion intensities of the respective technical replicates. Then, mean ion intensities of technical replicates were calculated. Time point comparisons (T = 0 hr vs T = 16 hr) were performed using t-tests with Benjamini-Hochberg (BH) correction for multiple testing. log_2FC values between the two time-points were calculated with respect to the mean intensity in the T0 time point. To identify pollen-derived ions, and distinguish them from background originating from culture medium and experimental noise, the ion intensities of the pollen dilution series were plotted for each ion and the R (2) of the obtained linear fit was extracted. In addition, we calculated the log_2FC difference between undiluted pollen and water. The R (2) values were then plotted against the log_2FC values, and stringent thresholds ($R_2 > 0.75$ and $log_2FC > 2$) were chosen to discriminate ions that are likely pollen-derived (***Figure 6—figure supplement 3***). All ions were included for downstream analysis (e.g. PCA) and then they were discriminated between pollen-derived and non-pollen-derived.

Semi-targeted metabolomics via GC-MS

Soluble metabolites were extracted from liquid cultures supplemented with 10% pollen extract (w/v) at the inoculation time and after 8, 16, and 24 hr of incubation. For each liquid culture sample, 300 µL was collected and centrifuged (15,000 g, 4˚C, 15 min). Then, 200 µL was transferred to a new tube, snap-frozen in liquid nitrogen, and stored at −80˚C. Once that all the samples were collected, soluble metabolites were extracted. To extract soluble metabolites, tubes were thawed on ice, and 75 µL of sample was combined with 5 µL of 20 mM internal standard (norleucine and norvaline, (Sigma-Aldrich) and U-$^{13}C_6$ glucose [Cambridge Isotope laboratories]). A volume of 825 µL of cold methanol:water:chloroform (5:2:2) solution was added to the sample and vortexed for 30 s. The tubes were incubated at −20˚C for 90 min and vortexed 2x for 30 s during the incubation. Tubes were centrifuged for 5 min at 10,000 g at 4˚C. The supernatant was removed and extraction was repeated using 400 µL of ice cold chloroform:methanol (1:1), tubes were vortexed and left on ice for 30 min. Tubes were centrifuged 5 min at 8000 rpm at 4˚C and the liquid phase was transferred to the previous extracted aqueous phase. A total of 200 µL of water was added and tubes were centrifuged 5 min at 8000 rpm. The aqueous phase was transferred to a 2 mL microcentrifuge tube. The aqueous extract was dried using a vacuum concentrator at ambient temperature overnight (Univapo 150 ECH vacuum concentrator centrifuge). Once dried, the samples were dissolved in 50 µL of 20

mg/ml methoxyamine hydrochloride in pyridine for 1.5 hr at 33˚C followed by derivatization with N-Methyl-N-(trymethylsolyl)trifluoroacetamide (MSTFA, Sigma Aldrich) for 2 hr at 35˚C.

Aliquots (1 µL) were injected on an Agilent 8890/5977B GC-MSD. The samples were injected in split mode (20:1) with an inlet temperature of 250˚C. The VF-5ms (30 m x 250 µm x 0.25 µm) column was held initially at 125˚C for 2 min, ramped at 5˚C / min to 250˚C, ramped at 15˚C to 300˚C, and held for 5 min. The MS was run in full scan mode (50–500 m/z) at a speed of 5 Hz. Peaks from the total ion chromatogram (TIC) were identified by matching retention times and spectra to an in-house library that was built by comparing selected T=0h and T=24h samples against the NIST library, as well as our library of analytical standards. Compounds are noted as either confirmed with our own standards, or the best match and associated matching factor against the NIST library are reported (***Supplementary file 10***). Peaks were picked and integrated using the Agilent MassHunter Quantitative Analysis software. Peak areas were normalized to the internal standards. The data were processed using R version 3.6.3 and mean intensities and $\log_2$FC between time-points were calculated as described above for the untargeted metabolomics analysis.

Analysis-code and data availability

The complete custom code for all the analyses is available on GitHub: (https://github.com/silviabrochet/Brochet_2021_eLife, copy archived at swh:1:rev:237a27f757296372f0333d298dfb7c765686fe03; ***Brochet, 2021***). The amplicon sequencing data and the RNA sequencing data are available under the NCBI Bioproject PRJNA700984 and the GEO record GSE166724. All differential expression analysis results of this study are included in ***Supplementary file 10***.

Acknowledgements

We thank Kirsten Ellegaard for the critical reading of the manuscript and discussions on the experimental design. We also thank Sara Mitri for providing feedback on the manuscript. We are grateful to Simon Maréchal for his help with the transfer experiments. This project was funded by the ERC-StG 'MicroBeeOme' (grant 714804), the Swiss National Science Foundation (grant 31003A_160345 and 31003A_179487), the Human Frontier Science Program Young Investigator (grant RGY0077/2016) and the NCCR Microbiomes.

Additional information

Funding

Funder	Grant reference number	Author
H2020 European Research Council	714804	Silvia Brochet Philipp Engel
Swiss National Science Foundation	31003A_179487	Andrew Quinn Nicolas Neuschwander Philipp Engel
Swiss National Science Foundation	51NF40_180575	Philipp Engel
Swiss National Science Foundation	31003A_160345	Philipp Engel
Human Frontier Science Program	RGY0077/2016	Philipp Engel

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Silvia Brochet, Conceptualization, Visualization, Methodology, Writing - original draft, Writing - review and editing; Andrew Quinn, Conceptualization, Data curation, Methodology, Writing - review and editing; Ruben AT Mars, Resources, Investigation, Methodology, Writing - review and editing; Nicolas Neuschwander, Investigation; Uwe Sauer, Resources, Writing - review and editing; Philipp

Engel, Conceptualization, Software, Supervision, Funding acquisition, Visualization, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Silvia Brochet https://orcid.org/0000-0002-6443-185X
Andrew Quinn https://orcid.org/0000-0003-1401-1053
Uwe Sauer https://orcid.org/0000-0002-5923-0770
Philipp Engel https://orcid.org/0000-0002-4678-6200

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.68583.sa1
Author response https://doi.org/10.7554/eLife.68583.sa2

Additional files

Supplementary files

- Supplementary file 1. List of primers used in this study: the sample-specific barcodes used in the primers for the second PCR of the amplicon sequencing are highlighted in gray.
- Supplementary file 2. COG functional categories.
- Supplementary file 3. Bacterial abundance data (CFUs).
- Supplementary file 4. Amplicon sequencing processed data.
- Supplementary file 5. RNA sequencing processed data, statistical analysis results (enrichment tests), transcript per million data.
- Supplementary file 6. RNA sequencing processed data, statistical analysis results (enrichment tests), transcript per million data.
- Supplementary file 7. RNA sequencing processed data, statistical analysis results (enrichment tests), transcript per million data.
- Supplementary file 8. RNA sequencing processed data, statistical analysis results (enrichment tests), transcript per million data.
- Supplementary file 9. RNA sequencing processed data, statistical analysis results (enrichment tests), transcript per million data.
- Supplementary file 10. Metabolomics analysis data.
- Supplementary file 11. Differential expression analysis results.
- Transparent reporting form

Data availability

The amplicon sequencing data and the RNA sequencing data are available under the NCBI Bioproject PRJNA700984 and the GEO record GSE166724 respectively. All data generated or analysed during this study are included in the manuscript and supporting files. Bacterial abundance data (CFUs) are included into Supplementary file 3, amplicon sequencing processed data are included into Supplementary file 4, RNA sequencing processed data, statistical analysis results (enrichment tests) and transcript per million data are included into Supplementary file 5–9, metabolomics analysis data are included into Supplementary file 10. All differential expression analysis results of this study are included in Supplementary file 11.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Brochet S, Quinn Q, Mars RAT, Neuschwander N, Sauer U, Engel P	2021	Lactobacillus Firm5 amplicon sequencing	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA700984	NCBI BioProject, PRJNA700984
Brochet S, Quinn	2021	Lactobacillus Firm5 in vivo and	https://www.ncbi.nlm.	NCBI Gene Expression

Q, Mars RAT, Neuschwander N, Sauer U, Engel P	in vitro RNA sequencing	nih.gov/geo/query/acc.cgi?acc=GSE166724	Omnibus, GSE166724

References

Avraham R, Haseley N, Fan A, Bloom-Ackermann Z, Livny J, Hung DT. 2016. A highly multiplexed and sensitive RNA-seq protocol for simultaneous analysis of host and pathogen transcriptomes. *Nature protocols* **11**:1477–1491. DOI: https://doi.org/10.1038/nprot.2016.090, PMID: 27442864

Beekwilder J, Marcozzi D, Vecchi S, de Vos R, Janssen P, Francke C, van Hylckama Vlieg J, Hall RD. 2009. Characterization of Rhamnosidases from Lactobacillus plantarum and Lactobacillus acidophilus. *Applied and environmental microbiology* **75**:3447–3454. DOI: https://doi.org/10.1128/AEM.02675-08, PMID: 19346347

Berasategui A, Salem H, Paetz C, Santoro M, Gershenzon J, Kaltenpoth M, Schmidt A. 2017. Gut microbiota of the pine weevil degrades conifer diterpenes and increases insect fitness. *Molecular ecology* **26**:4099–4110. DOI: https://doi.org/10.1111/mec.14186, PMID: 28543918

Berry D, Widder S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology* **5**:1–14. DOI: https://doi.org/10.3389/fmicb.2014.00219, PMID: 24904535

Bittleston LS, Gralka M, Leventhal GE, Itzhak Mizrahi OXC. 2019. Context-dependent dynamics lead to the assembly of functionally distinct pitcher-plant microbiomes. *Nature Communications* **11** 1440:–. DOI: https://doi.org/10.1038/s41467-020-15169-0

Brochet S. 2021. Brochet_2021_eLife. *Software Heritage*. swh:1:rev:237a27f757296372f0333d298dfb7c765686fe03. https://archive.softwareheritage.org/swh:1:dir:16c27686b491f4891b450ee51815d6fe11fdcdec;origin=https://github.com/silviabrochet/Brochet_2021_eLife;visit=swh:1:snp:77902562ef58804af850328001cd27a512f2f5a0;anchor=swh:1:rev:237a27f757296372f0333d298dfb7c765686fe03

Chen IA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter S, Varghese N, Seshadri R, Roux S, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2021. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic acids research* **49**:D751–D763. DOI: https://doi.org/10.1093/nar/gkaa939, PMID: 33119741

Chesson P. 2000. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* **31**:343–366. DOI: https://doi.org/10.1146/annurev.ecolsys.31.1.343

Coyte KZ, Schluter J, Foster KR. 2015. The ecology of the microbiome: Networks, competition, and stability. *Science* **350**:663–666. DOI: https://doi.org/10.1126/science.aad2602, PMID: 26542567

Crailsheim K. 1988. Intestinal transport of sugars in the honeybee (Apis mellifera L.). *Journal of Insect Physiology* **34**:839–845. DOI: https://doi.org/10.1016/0022-1910(88)90117-5

Deines P, Hammerschmidt K, Bosch TCG. 2020. Microbial species coexistence depends on the host environment. *mBio* **11**:20. DOI: https://doi.org/10.1128/mBio.00807-20

Dethlefsen L, McFall-Ngai M, Relman DA. 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449**:811–818. DOI: https://doi.org/10.1038/nature06245, PMID: 17943117

Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vásquez A. 2015. Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* **16**:1–22. DOI: https://doi.org/10.1186/s12864-015-1476-6, PMID: 25880915

Ellegaard KM, Brochet S, Bonilla-Rosso G, Emery O, Glover N, Hadadi N, Jaron KS, van der Meer JR, Robinson-Rechavi M, Sentchilo V, Tagini F, Engel P, SAGE class 2016-17. 2019. Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. *Molecular ecology* **28**:2224–2237. DOI: https://doi.org/10.1111/mec.15075, PMID: 30864192

Ellegaard KM, Suenami S, Miyazaki R, Engel P. 2020. Vast Differences in Strain-Level Diversity in the Gut Microbiota of Two Closely Related Honey Bee Species. *Current biology : CB* **30**:2520–2531. DOI: https://doi.org/10.1016/j.cub.2020.04.070, PMID: 32531278

Ellegaard KM, Engel P. 2016. Beyond 16S rRNA community profiling: intra-species diversity in the gut Microbiota. *Frontiers in Microbiology* **7**:1–16. DOI: https://doi.org/10.3389/fmicb.2016.01475, PMID: 27708630

Ellegaard KM, Engel P. 2018. New reference genome sequences for 17 bacterial strains of the honey bee gut Microbiota. *Microbiology Resource Announcements* **7**:e00834. DOI: https://doi.org/10.1128/MRA.00834-18, PMID: 30533872

Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota. *Nature communications* **10**:1–13. DOI: https://doi.org/10.1038/s41467-019-08303-0, PMID: 30683856

Elton C. 1946. Competition and the structure of ecological communities. *The Journal of Animal Ecology* **15**:54–68. DOI: https://doi.org/10.2307/1625

Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. *PNAS* **109**:11002–11007. DOI: https://doi.org/10.1073/pnas.1202970109, PMID: 22711827

Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. 2017. The evolution of the host microbiome as an ecosystem on a leash. *Nature* **548**:43–51. DOI: https://doi.org/10.1038/nature23292, PMID: 28770836

Foster KR, Bell T. 2012. Competition, not cooperation, dominates interactions among culturable microbial species. *Current biology : CB* **22**:1845–1850. DOI: https://doi.org/10.1016/j.cub.2012.08.005, PMID: 22959348

Freese NH, Norris DC, Loraine AE. 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**:2089–2095. DOI: https://doi.org/10.1093/bioinformatics/btw069, PMID: 27153568

Friedman J, Higgins LM, Gore J. 2017. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution* **1**:109. DOI: https://doi.org/10.1038/s41559-017-0109, PMID: 28812687

Fuhrer T, Heer D, Begemann B, Zamboni N. 2011. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Analytical chemistry* **83**:7074–7080. DOI: https://doi.org/10.1021/ac201267k, PMID: 21830798

Ghoul M, Mitri S. 2016. The Ecology and Evolution of Microbial Competition. *Trends in microbiology* **24**:833–845. DOI: https://doi.org/10.1016/j.tim.2016.06.011, PMID: 27546832

Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, Segrè D, Mehta P, Sanchez A. 2018. Emergent simplicity in microbial community assembly. *Science* **361**:469–474. DOI: https://doi.org/10.1126/science.aat1168, PMID: 30072533

Gude S, Pinçe E, Taute KM, Seinen AB, Shimizu TS, Tans SJ. 2020. Bacterial coexistence driven by motility and spatial competition. *Nature* **578**:588–592. DOI: https://doi.org/10.1038/s41586-020-2033-2, PMID: 32076271

Hallatschek O, Hersen P, Ramanathan S, Nelson DR. 2007. Genetic drift at expanding frontiers promotes gene segregation. *PNAS* **104**:19926–19930. DOI: https://doi.org/10.1073/pnas.0710150104, PMID: 18056799

Hallett L, Avolio M, Carroll I, Jones S, MacDonald A, Flynn D, Slaughter P, Ripplinger J, Collins S, Gries C JM. 2020. *Codyn: Community Dynamics Metrics*. https://github.com/NCEAS/codyn

Jalili-Firoozinezhad S, Gazzaniga FS, Calamari EL, Camacho DM, Fadel CW, Bein A, Swenor B, Nestor B, Cronce MJ, Tovaglieri A, Levy O, Gregory KE, Breault DT, Cabral JMS, Kasper DL, Novak R, Ingber DE. 2019. A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. *Nature Biomedical Engineering* **3**:520–531. DOI: https://doi.org/10.1038/s41551-019-0397-0, PMID: 31086325

Jimeno J, Lazaro MJ, Sollberger H. 1995. Antagonistic interactions between propionic acid Bacteria and non-starter lactic acid Bacteria. *Le Lait* **75**:401–413. DOI: https://doi.org/10.1051/lait:19954-530

Kešnerová L, Mars RAT, Ellegaard KM, Troilo M, Sauer U, Engel P. 2017. Disentangling metabolic functions of Bacteria in the honey bee gut. *PLOS Biology* **15**:e2003467. DOI: https://doi.org/10.1371/journal.pbio.2003467, PMID: 29232373

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**:801–814. DOI: https://doi.org/10.1038/s41396-019-0568-8, PMID: 31836840

Kim HJ, Boedicker JQ, Choi JW, Ismagilov RF. 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *PNAS* **105**:18188–18193. DOI: https://doi.org/10.1073/pnas.0807935105, PMID: 19011107

Koskella B, Brockhurst MA. 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS microbiology reviews* **38**:916–931. DOI: https://doi.org/10.1111/1574-6976.12072, PMID: 24617569

Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. *Nature Reviews Microbiology* **14**:374–384. DOI: https://doi.org/10.1038/nrmicro.2016.43, PMID: 27140688

Laëtitia G, Pascal D, Yann D. 2014. The citrate metabolism in Homo- and heterofermentative LAB: a selective means of becoming dominant over other microorganisms in complex ecosystems. *Food and Nutrition Sciences* **05**:953–969. DOI: https://doi.org/10.4236/fns.2014.510106

Ley RE, Peterson DA, Gordon JI. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**:837–848. DOI: https://doi.org/10.1016/j.cell.2006.02.017, PMID: 16497592

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. *Science* **320**:1647–1651. DOI: https://doi.org/10.1126/science.1155725, PMID: 18497261

Li L, Abou-Samra E, Ning Z, Zhang X, Mayne J, Wang J, Cheng K, Walker K, Stintzi A, Figeys D. 2019. An in vitro model maintaining taxon-specific functional activities of the gut microbiome. *Nature Communications* **10**:4146. DOI: https://doi.org/10.1038/s41467-019-12087-8, PMID: 31515476

Logan M. 2017. Co-occurring soil Bacteria exhibit a robust competitive hierarchy and lack of non-transitive interactions. *bioRxiv*. DOI: https://doi.org/10.1101/175737

Madi N, Vos M, Murall CL, Legendre P, Shapiro BJ. 2020. Does diversity beget diversity in Microbiomes? *bioRxiv*. DOI: https://doi.org/10.1101/612739

Magni C, de Mendoza D, Konings WN, Lolkema JS. 1999. Mechanism of citrate metabolism in Lactococcus lactis: resistance against lactate toxicity at low pH. *Journal of bacteriology* **181**:1451–1457. DOI: https://doi.org/10.1128/JB.181.5.1451-1457.1999, PMID: 10049375

Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell host & microbe* **4**:447–457. DOI: https://doi.org/10.1016/j.chom.2008.09.007, PMID: 18996345

Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA. 2011. A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular ecology* **20**:619–628. DOI: https://doi.org/10.1111/j.1365-294X.2010.04959.x, PMID: 21175905

McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, Foster KR. 2016. Host Selection of Microbiota via Differential Adhesion. *Cell host & microbe* **19**:550–559. DOI: https://doi.org/10.1016/j.chom.2016.02.021, PMID: 27053168

Meszéna G, Gyllenberg M, Pásztor L, Metz JA. 2006. Competitive exclusion and limiting similarity: a unified theory. *Theoretical population biology* **69**:68–87. DOI: https://doi.org/10.1016/j.tpb.2005.07.001, PMID: 16243372

Mitri S, Clarke E, Foster KR. 2016. Resource limitation drives spatial organization in microbial groups. *The ISME journal* **10**:1471–1482. DOI: https://doi.org/10.1038/ismej.2015.208, PMID: 26613343

Nikolaev M, Mitrofanova O, Broguiere N, Geraldo S, Dutta D, Tabata Y, Elci B, Brandenberg N, Kolotuev I, Gjorevski N, Clevers H, Lutolf MP. 2020. Homeostatic mini-intestines through scaffold-guided organoid morphogenesis. *Nature* **585**:574–578. DOI: https://doi.org/10.1038/s41586-020-2724-8, PMID: 32939089

Ortiz A, Vega NM, Ratzke C, Gore J. 2021. Interspecies bacterial competition determines community assembly in the *C. elegans* intestine. *The ISME Journal* **15**:2131–2145. DOI: https://doi.org/10.1038/s41396-021-00910-4, PMID: 33589765

O' Donnell MM, Forde BM, Neville B, Ross PR, O' Toole PW. 2011. Carbohydrate catabolic flexibility in the mammalian intestinal commensal Lactobacillus ruminis revealed by fermentation studies aligned to genome annotations. *Microbial Cell Factories* **10**:1–11. DOI: https://doi.org/10.1186/1475-2859-10-S1-S12

Piccardi P, Vessman B, Mitri S. 2019. Toxicity drives facilitation between 4 bacterial species. *PNAS* **116**:15979–15984. DOI: https://doi.org/10.1073/pnas.1906172116, PMID: 31270235

Rakoff-Nahoum S, Foster KR, Comstock LE. 2016. The evolution of cooperation within the gut microbiota. *Nature* **533**:255–259. DOI: https://doi.org/10.1038/nature17626, PMID: 27111508

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. DOI: https://doi.org/10.1093/bioinformatics/btp616, PMID: 19910308

Schluter J, Foster KR. 2012. The evolution of mutualism in gut Microbiota via host epithelial selection. *PLOS Biology* **10**::e1001424. DOI: https://doi.org/10.1371/journal.pbio.1001424

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J. 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**:250–255. DOI: https://doi.org/10.1038/nature08756, PMID: 20164839

Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD, Gordon JI. 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**:1955–1959. DOI: https://doi.org/10.1126/science.1109051, PMID: 15790854

Tilman D. 1986. A Consumer-Resource approach to community structure. *American Zoologist* **26**:5–22. DOI: https://doi.org/10.1093/icb/26.1.5

Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, Hromada S, Northen T, Arkin AP. 2018. Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular Systems Biology* **14**:e8157. DOI: https://doi.org/10.15252/msb.20178157, PMID: 29930200

Wright ES, Vetsigian KH. 2016. Inhibitory interactions promote frequent bistability among competing bacteria. *Nature communications* **7**:11274. DOI: https://doi.org/10.1038/ncomms11274, PMID: 27097658

Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**:614–620. DOI: https://doi.org/10.1093/bioinformatics/btt593, PMID: 24142950

Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA. 2018. Honey bees as models for gut microbiota research. *Lab animal* **47**:317–325. DOI: https://doi.org/10.1038/s41684-018-0173-x, PMID: 30353179

Zheng H, Perreau J, Powell JE, Han B, Zhang Z, Kwong WK, Tringe SG, Moran NA. 2019. Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *PNAS* **116**:25909–25916. DOI: https://doi.org/10.1073/pnas.1916224116, PMID: 31776248

Manuscript II: Strain-strain interactions within the honey bee gut

Summary

In this Manuscript we investigated if strains from the same species (conspecific) and from different species (allospecific) of bee-gut phylotype *Lactobacillus* Firm5 differ in how they interact and coexist. Our hypothesis was that interactions and coexistence between closely related bacteria depends on their phylogenetic distance, with more closely related strains (conspecific) interacting more negatively and being less likely to coexist if compared to less closely related strains (allospecific).

To test this hypothesis we selected 12 different Firm5 strains, 3 for each of the four Firm5 species, and we colonized microbiota depleted bees with all possible conspecific and allospecific pairwise combinations (for a total of 66 different pairs) as well as with the 12 strains individually. 10 days post-colonization we analyzed each gut-community using an amplicon sequencing protocol that we designed to target and distinguish between the selected Firm5 strains.

We inferred interactions between strains by measuring the effect of each strain on another based on the comparison of the colonization levels of each strain in mono-colonization and in co-colonization. Negative interactions governed most pairs of strains that were interacting significantly during gut colonization (n = 59). The majority of Firm5 strains were mutually inhibiting each other (53/59) and in a subset of pairs (6/59) one strain was facilitated while the other was inhibited (parasitism). Neither the type nor the strength of the interactions that we measured correlated with the phylogenetic distance between strains or with their differences in the accessory gene content. Still, we

found that pairs of conspecific strains tended towards a more skewed community composition, with one strain dominating over the other, while allospecific pairs tended towards a more even community composition.

These results show that negative interactions dominate among Firm5 strains, independently of their phylogenetic relatedness. Still, depending on the species-affiliation, the competition outcome in strains proportions can vary, which we reasoned could potentially affect strains coexistence over time. To investigate if these differences in strains proportions within pairs corresponded to differences in the ability to coexist between conspecific and allospecific strains we randomly selected four conspecific strains pairs and four allospecific strains pairs and we passaged them for three times across microbiota-depleted bees. We also did a similar experiment *in vitro* in culture tubes supplemented with water extracts of pollen grains.

While allospecific strains were always able to coexist in the bee gut, three out of four combinations of conspecific strains could not coexist. This indicated that allospecific strains are more likely to coexist if compared to conspecific strains. Still, the fact that there was one combination of conspecific strains that could coexist suggests that coexistence within Firm5 species might be to some extent strain-specific. Further experiments including additional combinations of conspecific strains are needed to confirm these findings. In addition, we observed similar results *in vitro* but with a tendency of having a lower coexistence maintenance also within allospecific pairs of strains as well. The *in vitro* data is however still highly preliminary and the results need to be further validated.

Introduction

Natural microbial communities are considered to be highly complex. They usually consist of several different bacterial species that occupy the same physical space and that interact with each other. Moreover, there is additional variation found within species in terms of strain-level and genomic diversity (Dethlefsen, McFall-Ngai, and Relman 2007; Ellegaard and Engel 2016; Louca et al. 2017; Van Rossum et al. 2020; Venter et al. 2004).

Bacterial species are defined (with some controversies) as 'coherent genomic clusters' that are functionally distinct to one another and thus expected to coexist e.g. by avoiding mutual exclusion thanks to resource partitioning (Chun et al. 2018; Van Rossum et al. 2020). The definition of strain is also highly debated, as in principle one single nucleotide variant (SNV) is sufficient to define a strain but i) it is not clear if this is a biologically relevant definition and ii) even with metagenomics it is complicated to distinguish particularly closely related strains (especially in absence of a comprehensive database of reference sequences) (Van Rossum et al. 2020). Ratzke et al. recently established that more closely related bacteria interact more strongly and that stronger negative interactions result in the exclusion of the 'weaker' variants from the community (Ratzke, Barrere, and Gore 2020). Hence, conspecific strains are in principle not expected to coexist within bacterial communities, as they are too similar to each other. Nevertheless, it could be that conspecific strains harbor (little) but functionally distinct strain-specific (accessory) genes that could allow them to differentiate and coexist. In this case, the coexistence of bacteria would be determined at the level of the strain rather than at the level of species (Dana E. Hunt, Lawrence A. David, Dirk Gevers and Eric J. Alm 2008). Interestingly, a recent study tracking microbial communities associated with pitcher plants showed that interactions between closely related strains (differing as little as by 100 base pairs) can exhibit vastly different dynamics, suggesting that intra-specific interactions govern the community dynamics over time (Goyal et al. 2021). In addition,

it appears that macroecological properties of the human gut microbiome such as its stability is determined at the strain-level (Wolff, Shoemaker, and Garud 2021). Still, it is not yet clear which kind of interactions are common among conspecific strains, if they are different if compared to the interactions between the allospecific ones and if these differences influence strains coexistence.

An experimentally tractable bacterial community suitable to investigate about this this is the gut microbiota of the honey bee (*Apis mellifera*) (Zheng et al. 2018). This community is dominated by eight deep-branching phylogenetic lineages (phylotypes) that diverged into several closely related species, which contain further diversity at the strain level (Ellegaard et al. 2015; Ellegaard and Engel 2016, 2019; Engel, Martinson, and Moran 2012; Kwong and Moran 2016). This model allows to study inter- and intra-specific interactions both *in vitro*, as all the community-members can be cultured in the laboratory, as well as *in vivo*, as microbiota-depleted bees can be generated and experimentally colonized with synthetic communities of different strains (Kešnerová et al. 2017; Zheng et al. 2018).

In this study we used one of the most abundant phylotypes of the honey bee gut microbiota, i.e. *Lactobacillus* Firm5, to study differences in interactions and coexistence between conspecific and allospecific strains. Firm5 has diverged into four different species whose coexistence is facilitated by the partitioning of the pollen diet of the bee (Manuscript I). Strain-diversity is found within species and the amount of conspecific strains that can be detected within bees varies depending on the species (Ellegaard and Engel 2019). It's not known which interactions dominate at the strain level within the Firm5 phylotype and which is their impact on coexistence: it could be that conspecific strains compete more strongly as they are more closely related and exclude one another. On the other hand, it could also be that interactions are strain-specific and depend on which strain is present and not to which species it belongs to.

In this study we selected three strains for each of the four Firm5 species and we mixed them in all possible pairwise combinations (12 conspecific combinations and 54 allospecific combinations). We fed each combination (co-colonizations) as well as the 12 strains individually (mono-colonizations) to microbiota-depleted bees. To measure interactions, we calculated the effect of each strain on the other by comparing colonization levels in mono-colonizations and in co-colonizations in conspecific and allospecific pairs. Moreover, to investigate coexistence, we selected eight pairs of strains, four including conspecific strains and four including allospecific strains, and we passaged them both *in vivo* in microbiota-depleted bees and *in vitro* in culture tubes. Our results show that negative interactions dominate strain-level diversity within the bee gut and that their strength does not correlate with relatedness. However, conspecific pairs resulted in less even communities, with one strain generally dominating over the other. Our passaging experiments indicated that these pairs were less likely to co-exist than pairs of strains from different species.

Results

Systematic pairwise colonization of gnotobiotic bees with 12 strains of *Lactobacillus* Firm5 belonging to four different species

In order to explore strain-level interactions within the bee gut, we selected three strains from each of the four *Lactobacillus* Firm5 species (in total 12 strains, see Materials and methods, **Annex (II) - Supplementary Table 1**). We colonized microbiota-depleted bees with each of the 12 strains separately as well as in all 66 possible pairwise combinations (**Annex (II) - Supplementary Table 2**) (**Figure 1.A**). Ten days post-colonization, we dissected the distal hindgut of each bee (rectum) and determined the total bacterial abundance of each community member combining CFUs plating with amplicon sequencing of a house-keeping genes that allows to discriminate the selected

strains (**Figure 1.B**, see detailed methods in Manuscript I). We analyzed a minimum of three and a maximum of five individual bees per treatment resulting in a total of 332 bees that were analyzed for this experiment.

We obtained between 1,000 and 55,000 (in average = 8,000) reads per sample (**Figure 1 - figure supplement 1.A, Annex (II)**). Our limit of detection (LOD) for the two strains in a given sample varied from 10^4 to 10^6 copies per bee gut depending on the sequencing depth and the total bacterial abundance in each sample (**Figure 1 - figure supplement 1.B, Annex (II)**). The total number of CFUs per gut ranged from $5.0x10^8$ to $5.0x10^9$, indicating that all strains were able to colonize microbiota-depleted bees (**Figure 1 - figure supplement 1.C, Annex (II)**). Amplicon sequencing revealed that the two strains of the inoculum dominated the communities in most of the analyzed samples (94.7% of all samples had > 90% of all reads assigned to the two focal strains) suggesting that the colonization worked as expected (**Figure 1 - figure supplement 2, Annex (II)**). Only a few bees had marked cross-contaminations with other strains (i.e. > 10% of the reads were assigned to a strain not present in the inoculum) (**Figure 1 - figure supplement 2, Annex (II)**, red zone). These samples were excluded from further analysis.

While we aimed at colonizing all bees with equal proportions of the two strains (see Materials and methods), one of the strains was often more abundant than the other (**Figure 1 - figure supplement 3.A, Annex (II)**). However, there was no correlation between the proportions of the strains in the inoculum and in the gut ten days post inoculation, indicating that interactions with the host, the diet, or the other community members influence community composition (**Figure 1 - figure supplement 3.B, Annex (II)**).

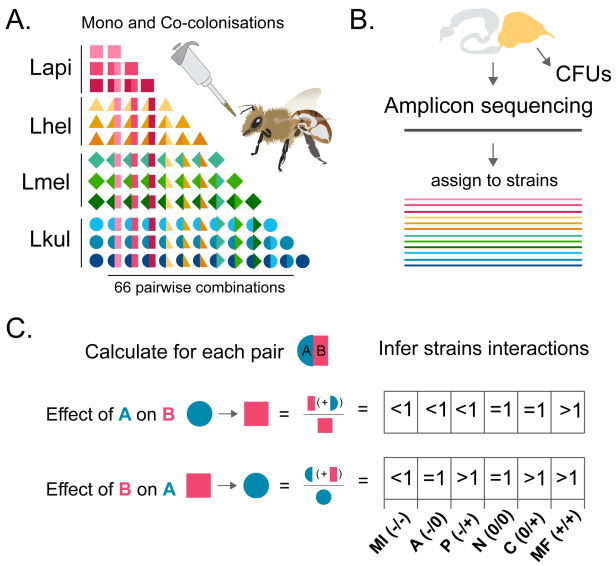


Figure 1: Experimental design and analysis strategy of the pairwise interaction experiment. A. Twelve strains (different colors) from the four Firm5 species (different shapes, Lapi, Lhel, Lmel and Lkul) were selected to analyze species- vs strain-level interactions in the bee gut. Microbiota-depleted bees were colonized with each of the 12 strains separately as well as in all 66 possible pairwise combinations (n=5). **B.** 10 days post-colonization the rectum of the gnotobiotic bees was subjected to CFUs plating and amplicon sequencing to determine the absolute abundance of each of the two community members in each strain pair. **C.** In order to determine the interactions governing each strain pair, the effect of one strain on the other was calculated by computing for each strain the ratio between the bacterial loads in the co-colonization versus the mono-colonization condition. The reciprocal effects of the two strains were used to determine the type of the interaction governing in each pair (mutual inhibition – MI, amensalism – A, parasitism – P, neutralism – N, commensalism – C, mutual facilitation – MF).

Negative interactions dominate among strains of *Lactobacillus* Firm5

In all samples, both strains were detected after 10 days of colonization suggesting that none of the strains was able to completely outcompete the other strain. To analyze the type of interaction governing each pair of strains we used a similar approach as Kehe et al. (Kehe et al. 2021). We determined each one-way interaction by calculating the effect of a given strain on another measuring the ratio between the yield of the co-colonization and the yield of the mono-colonization (see Materials and methods, **Formula 1**). If the yield of a strain in co-colonization was lower in presence of the second strain, the effect (< 1) was classified as negative (-), if it was higher the effect (> 1) was classified as positive (+), while if there was no effect ($= 1$) it was classified as neutral

(0). These one-way interactions were then used to determine the bidirectional interactions for each pair of strains both qualitatively and quantitatively. Six different bidirectional interactions can be determined based on the reciprocal one-way interactions (-/-, -/0, +/-, 0/0, +/0, or +/+; **Figure 1.C**). Importantly, bidirectional interactions can be measured both qualitatively and quantitatively. For example, we can infer that two strains are interacting through mutual inhibition (-/-) and also determine the extent to which they inhibit each other: one strain may have a much stronger inhibitory effect on the other strain than vice versa.

To be able to capture qualitative and quantitative aspects of interactions, we considered all bidirectional interactions as polar coordinates (or vectors) of angle (θ) and of radius (r), which respectively measure the type and the strength of the interaction (see Materials and methods, **Formula 2**) (**Figure 2.A**). To determine if a given pair exhibits a robust type of interaction we calculated the vectors dispersion (ρ, 0 = high dispersion, 1 = low dispersion) across the analyzed replicates. We then used circular statistics (see Materials and methods) to test which vectors had significantly high ρ to explain an interaction. For 59 out of 66 pairs, the two strains had a statistically significant interaction type (Pycke's test, $p < 0.05$), i.e. their correspondent vectors were concentrated around one preferred direction and were not dispersed around the circle (**Figure 2.B**) (Pycke 2010). Of these significant interactions, 53 were categorized as mutual inhibition (average θ = -65.6 ° +/- 10.9) and 6 as parasitism (average θ = -37.2 ° +/- 7.9) (**Figure 2.B**). The value of θ does not only indicate which type of bidirectional interaction exists but also how balanced this interaction is among the two strains (see **Figure 2.A**). The closer θ is to -90 °, 0 ° or +90 °, the more the two strains are respectively equally inhibiting or facilitating each other (**Figure 2.A**). The values of θ that we measured across the different interacting pairs variated substantially, from -24.2 ° to -87.8 ° (average = -60.8 °), indicating that in some pairs the interactions were more balanced than in other pairs (**Figure 2.A**).

The r values of the interaction vectors varied substantially across pairs ranging from 0.1-10.9 (average = 2.8) indicating that some pairs had stronger effects on each other than others.

All this taken together, we can conclude that the large majority of the interactions between Firm5 strains are negative, but that there is variation in the reciprocal inhibitory effect of the two strains on each other and the strength of their interactions.

Negative interactions between strains from different species are less biased

Our results suggest that in terms of the type of interaction, it does not matter whether two strains are from the same or a different species: negative interactions predominate among *Lactobacillus* Firm5. However, our results suggest that there may be quantitative differences between strains that could depend on the species-affiliation. For example, conspecific strains may interact more strongly with each other or have less even effects on each other.

We correlated both the r (interaction strength) and the θ (interaction type) values with both the genetic distance and the differences in the accessory gene content between strains, but could not find any significant correlation (**Figure 2 - figure supplement 1, Annex (II)**). However, strain pairs from different species (allospecific strains) had lower θ values (mean = -66.5 ˚ +/- 14.1) than strain pairs from the same species (conspecific strains) (mean = -71.4 ˚ +/- 12.6) (**Figure 2.C**), suggesting less even reciprocal effects between conspecific strains. These findings were consistent with differences observed in the relative abundance of strains in pairs consisting of conspecific versus allospecific strains. Allospecific pairs were normally distributed around equal proportions of the two strains in the community (Shapiro-Wilk normality test, $p < 0.07611$) (**Figure 2.D**). In contrast, conspecific pairs followed a non-normal distribution

with a larger variation the proportions of the two strains in the community (Shapiro-Wilk normality test, $p < 0.000279$) (**Figure 2.D**).

These results show that phylogenetic relatedness between strains does not influence the strength with which the strains interact nor the type of bidirectional interactions between strains. However, it appears that strains from the same species interact in a more biased way, i.e. strain A has a stronger influence on strain B, than *vice versa*, explaining the less even community compositions of conspecific pairs.

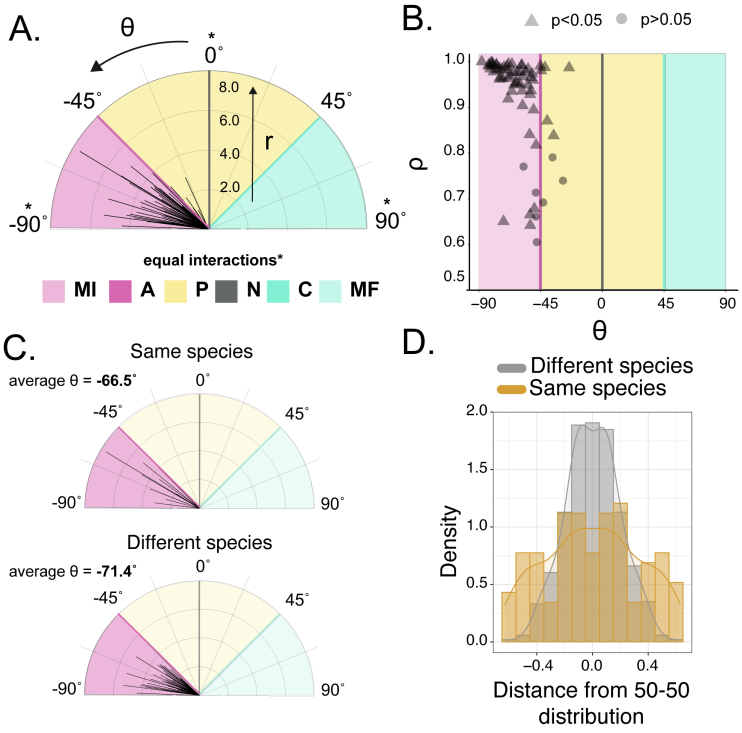


Figure 2: Polar coordinates describing Firm5 strain-strain interactions. A. Polar plot displaying means of replicates for each strain-pair (n=3-5) as vectors of angle θ (type of interaction) and length r (strength of interaction). The different colors represent the different interaction types (mutual inhibition – MI, amensalism – A, parasitism – P, neutralism – N, commensalism – C, mutual facilitation – MF). Note that A, N, and C only exist if the angle θ is exactly, -45 ˚ , 0 ˚ , and 45 ˚ , respectively. Asterisks show the θ values where interactions between strains are balanced. **B.** Statistical analysis of vectors dispersion between the replicates of each strain pair. Each point represents one pair of strains. ρ is a measure of the vectors dispersion between the different replicates of a strain pair. It can range from 0 (high dispersion) to 1 (low dispersion). The significance of ρ was tested for each vector for each replicate using circular statistics (see Materials and methods). **C.** Polar plots displaying means of vector replicates for each strain pair divided into pairs of strains from the same species and different species. **D.** Histogram displaying the distribution of the distance from even proportions between the two strains in each of the 66 pairs. Pairs including strains from the same species and pairs including strains from different species are displayed in yellow and grey respectively.

Coexistence of strains from different species, competitive exclusion of strains from the same species *in vivo*

Although none of the tested strains was able to outcompete any of the other strains in our pairwise colonization experiments, the disproportional abundances of some of the conspecific pairs suggest that these strains may not be able to coexist over time. To explore this further, we serially passaged four randomly selected pairs of each conspecific and allospecific strains through gnotobiotic bees for a total of three times (**Figure 3.A**). After each passage (i.e. after 7 days of colonization), we used amplicon sequencing in combination with CFUs counting (see Manuscript I) to determine the absolute abundance of each strain in each pair (**Figure 3.B**).

In three out of the four conspecific strain pairs (pairs 31, 57 and 65), one of the two strains steadily decreased in abundance, with two of them reaching the limit of detection after one and two passages, respectively (**Figure 3.B**). In contrast, in the fourth pair (pair 1), while the two strains exhibited clear differences in the bacterial loads in the gut, none of them decreased in abundance over the three passages suggesting stable coexistence (**Figure 3.B**). A similar pattern was observed for all four pairs of allospecific strains. Both strains were detected in all passages across all four pairs, with none of them reaching the limit of detection (**Figure 3.B**). None of the strains showed a trend towards decreasing total abundance from passage to passage (**Figure 3.B**). However, two pairs (pairs 11 and 13) exhibited slightly more variation in bacterial abundance over the three passages (especially at P1) than the other two pairs (pair 7 and 43) (**Figure 3.C**). In summary, we show that only one out of four combinations of conspecific strains, but all four combinations of allospecific strains can stably coexist over three passages through microbiota-depleted bees.

In order to investigate to what extent the host environment influences the interactions and coexistence patterns of conspecific and allospecific strains, we serially passaged the same four conspecific and allospecific pairs *in vitro* in a minimal medium

supplemented with water extracts of pollen grains for four times every 24 hours (**Figure 3 - figure supplement 1, Annex (II)**). The absolute abundance of each strain in the pair was determined as described above for the *in vivo* experiment. The *in vitro* results, although preliminary, show a similar trend as the *in vivo* results. In fact, the three conspecific pairs that could not coexist *in vivo*, also show a similar trend *in vitro*, with one of the two strains of the pair decreasing in its abundance and the other remaining constant (**Figure 3 - figure supplement 1, Annex (II)**). Interestingly, the same strain combination that was able to coexist *in vivo* also seemed to be able to coexist *in vitro* (**Figure 3 - figure supplement 1, Annex (II)**). In contrast, the results between the *in vitro* and *in vivo* experiment were less congruent for the allospecific strain pairs. In fact, only two out of four pairs (7 and 13) showed signs of coexistence over time (**Figure 3 - figure supplement 1, Annex (II)**).

In summary, we can conclude that the *in vitro* results for the conspecific, but not the allospecific strain pairs correspond to the *in vivo* ones.

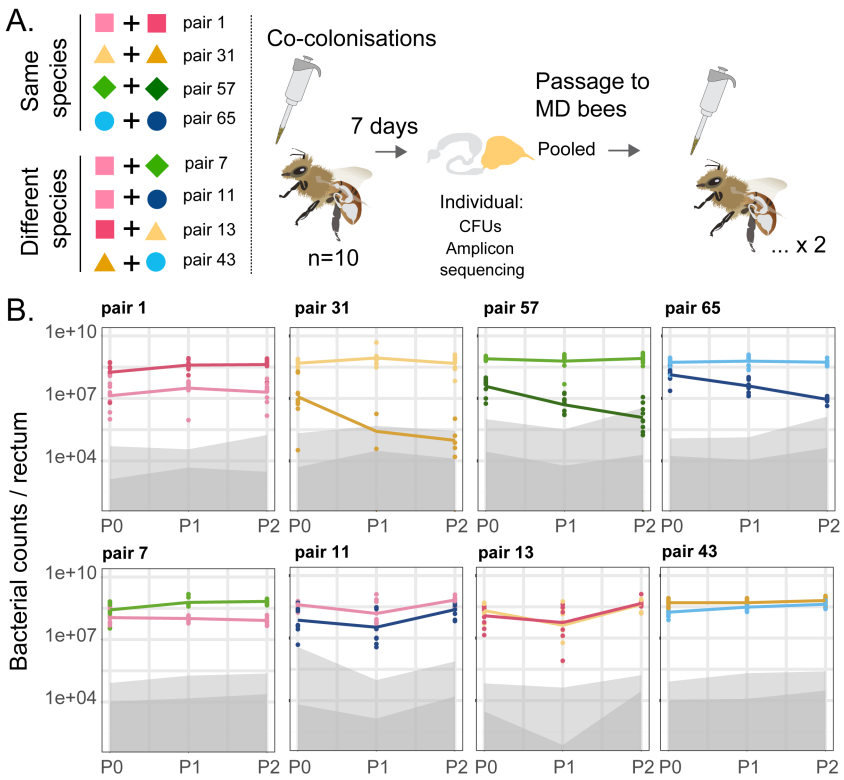


Figure 3: Differences in the coexistence of strain pairs in gnotobiotic bees. A. Experimental setup of the *in vivo* passaging experiment including four pairs of each two strains of the same (pairs 1, 31, 57, and 65) and two strains of different species (pairs 7, 11, 13, and 43). Microbiota-depleted bees were colonized with the selected strain pairs and seven days post-colonization the rectum was dissected. An aliquot was used to plate CFUs and perform amplicon sequencing. The remaining gut content of each replicate was pooled to colonize a new batch of microbiota depleted bees. In total three passages through gnotobiotic bees (P0-P2) were performed. **B.** Bacterial counts per rectum (n=10) for the two strains in each of the eight tested strain pairs across the three passages (P0-P2). Counts were determined by multiplying the total number of CFUs with the relative abundance of each strain in the community. Grey areas represent the limit of detection: the 95% confidence intervals of the limit of detection are shown in light grey.

Discussion

According to the bacterial species concept proposed by Cohan (2002) bacteria diversify into species because they occupy distinct ecological niches. In contrast, strains within species are supposed to have overlapping niches (Van Rossum et al. 2020). This implies that conspecific (from the same species) and allospecific (from different species) strains interact differently with each other, with different consequences on their ability

to coexist. However, both conspecific and allospecific strains often co-occur in natural microbial communities (Van Rossum et al. 2020), putting this concept into question.

Here we studied pairwise interactions and coexistence between conspecific and allospecific strains of the honey bee gut symbiont *Lactobacillus* Firm5 during gut colonization. The four selected species form a monophyletic clade within the lineage of *Lactobacillus* Firm5 and mostly occur in the gut of honey bees suggesting that these bacteria have diversified into different strains and species in the bee gut environment. This makes them particularly interesting candidates to investigate the type of interactions that governs them.

A first observation that we made was that the proportions of the strains in the rectum at the end of the experiment did not reflect the strains ratio in the inoculum. This finding is in line with the niche-theory, which predicts that the assembly of microbiomes is governed by deterministic processes, such as interactions with the partner strain, the diet or the host (Jeraldo et al. 2012; Sun et al. 2019).

In our case, all strains were mutually inhibiting each other except of six pairs in which we observed interactions of the parasitism type. This was independent of whether the two strains came from the same or a different species and confirms our previous results where we had tested a four strain community of *Lactobacillus* Firm5 and also found negative interactions (see Manuscript I). This is likely because, since a single Firm5 strain is sufficient to reach the carrying capacity (i.e. the average maximum population size, (McArthur 2006)) of this phylotype in the rectum, when other strains are present they are compelled to mutually inhibit one another to share the niche. Currently, we do not know if the observed negative interactions result from bacterial warfare, utilization of the same nutrients, or competition for space. We showed (Manuscript I) that the presence of pollen increases the carrying capacity of this phylotype in the gut. This indicates that the carrying capacity is limited either by the availability of 'broad spectrum' pollen-derived compounds (such as co-factors, nitrogen and carbon sources) in

the rectum or by the niche space that might increase (especially the amount of surface available) in presence of pollen. Moreover, as we found some statistically significant differences in the bacterial load between strain pairs in mono- and co-colonization, some strains may have less overlapping niches than others. Finally, we observed variation in the relative strains proportions across different strain pairs, indicating that the strength of the negative effects quantitatively differs between pairs. The measurement of interactions-balance (how skewed are strains proportions within pairs) and strength (how strongly are strains negatively affecting one another in co- vs to mono-colonization) allowed us to assess such quantitative differences between strain pairs and test whether they could be linked to features of the combined strains, such as their relatedness.

We didn't find any correlation between the variation in interactions-balance and strength neither with the phylogenetic distances between strains nor with their differences in the accessory gene content. This could be because the differences in interactions-balance and strength are due to few specific genes that don't have an effect on the phylogenetic distance between strains. In line with this, we showed previously (Manuscript I) that shared genes, and not necessarily strain-specific genes, are differentially regulated in conditions where divergent strains of Firm5 coexist. Moreover it could be that the accessory gene content includes genes that don't contribute to niche differentiation thus blurring any signal of niche differentiation.

The evidence that interactions within the Firm5 phylotype are mostly negative and don't correlate with phylogenetic relatedness is coherent with the fact that these bacteria are extremely closely related. They are all fermenters that share similar metabolic strategies and that lack a TCA cycle, suggesting that they are likely not engaging in cross-feeding interactions. Moreover, this is in line with the conclusions of several papers that highlighted that competitive interactions dominate in the bacterial world (Foster and Bell 2012; Kehe et al. 2021). In addition to the competition for similar resources, another reason that could explain why negative interactions dominate is that,

on the contrary, investing in positive interactions could be evolutionary disadvantageous as it creates interdependencies between bacteria that are likely to become unreliable and unstable for example facing environmental disturbances (Foster and Bell 2012).

But can measuring interactions help us to predict bacterial coexistence? In some cases, yes. For example, Piccardi et al. observed that when the shared environment is toxic, the interactions between the members of a four species community are positive, allowing them to coexist (Piccardi, Vessman, and Mitri 2019). This is likely due to the fact that a subset of species participates to the detoxification of the environment, making it less toxic for their own survival and at the same time accidentally allowing the other species to survive (Piccardi, Vessman, and Mitri 2019). On the other hand, when the shared environment is not toxic but rich in nutrients, negative interactions dominate among the four species and only the detoxifier ones can persist (Piccardi, Vessman, and Mitri 2019).

Contrarily, in our case the dominance of negative interactions is not translated in the absence of coexistence among strains. In fact, by serially passaging selected conspecific and allospecific strain pairs through gnotobiotic bees and monitoring the abundance of both strains at each passage we observed that the majority (5/8) of these pairs (although they were competing) could coexist over time. This is also what we showed in our previous study (Manuscript I) as the four Firm5 species could coexist in presence of pollen while competing with one another. It could be that certain Firm5 strains compete because of the saturated carrying capacity of the system in mono-colonization (as mentioned above) and at the same time manage to coexist by specializing on a subset of pollen-derived substrates. In addition, the fact that strains don't necessarily exclude one another could be advantageous in that neighboring strains can potentially leak metabolic by-products promoting cross-feeding interactions or can contribute to the resistance to the invasion of alien species by filling of all the available niches.

In particular, by studying coexistence over time, we found that strains were able to coexist within all four tested allospecific stain pairs. In contrast, in three out of the four tested conspecific strain pairs, one of the strains steadily decreased in abundance over the three passages. This suggests that conspecific strains are less likely to coexist. These results are in line with our previous study where we showed that four strains, each from a different *Lactobacillus* Firm5 species were stably coexisting in the bee gut in the presence of pollen (the same conditions as used here). In any case, an important conclusion that we can draw from our experiments is that not all combinations of strains from the same species will necessarily result in competitive exclusion. As a matter of fact, in one of the four tested conspecific strain pairs, both strains (ESL0185 and ESL0263) were maintained at similar bacterial loads across the three passages. Interestingly, these two strains were coexisting over time with non-balanced proportions, i.e. ESL0263 was always much more abundant than ESL0185, indicating that interactions-balance is also not always a good predictor of coexistence over time. ESL0185 and ESL0263 belong to *L. apis*, the species with the lowest number of COG 'G' genes (carbohydrate transport and metabolism category) and phosphotransferase transporters (PTS), that are contrarily very abundant in the other species and are likely important for the import and the utilization of pollen-derived carbohydrates (see Manuscript I, Ellegaard et al., 2020). It could be that these two strains, as they appear to be slightly less specialized to use pollen carbohydrates if compared to the other species, manage to coexist by partitioning other types of substrates.

How do these results compare to the distribution of different Firm5 species and strains in naturally colonized bees? It was shown by a recent metagenomics study from our lab that the four Firm5 species consistently coexist in individual bees. In contrast, strains of the same species showed the tendency to segregate between individual bees (Ellegaard and Engel 2019). Our results seem to recapitulate this as allospecific strains

coexisted more often than conspecific strains. Conspecific strains can also (more rarely) coexist perhaps thanks to niche segregation.

Our study presents some limitations. First of all, due to the laborious nature of gnotobiotic bee experiments and the availability of cultured isolates, we could only include a relatively small number of strains per species (three). This has limited the number of conspecific strain combinations we could test. More pairs of strains would need to be included in these experiments to test the statistical significance of the difference between allospecific and conspecific strain pairs. Moreover, our *in vitro* study needs to be expanded with more replicates and passages to better disentangle strains dynamics *ex vivo* and understand to which extent interbacterial interactions or host interference are important to predict community dynamics.

Finally, in this study we measured interactions and coexistence in a pairwise manner. We can therefore not exclude that there could also be higher order interactions arising when more than two species are mixed, even if there is evidence that pairwise interactions have been proven meaningful to describe bacterial dynamics within populations (Foster and Bell 2012). In addition, in our study we are only considering a subset of *Lactobacillus* Firm5 strains in absence of the rest of the bee-microbiota members. Interactions between Firm5 strains could change in presence of the other microbiota members as they would potentially fill niches that in our experiment remain available or they could engage in cross-feeding interactions based on byproducts of other phylotypes.

Despite these limitations altogether these results show that the outcome of bacterial interactions cannot simply be predicted from the phylogenetic relatedness of bacteria and that it is important to consider strain-level variation when monitoring bacterial community dynamics. This was also highlighted in a recent study by Goyal et al., who showed that even extremely closely related bacterial variants (differing by only 100 base pairs!) can exhibit vastly different coexistence patterns (Goyal et al. 2021).

Materials and methods

Culturing of bacterial strains

For this study we used 12 *Lactobacillus* Firm5 strains, 3 for each of the four species (*Lactobacillus apis*, *Lactobacillus helsingborgensis*, *Lactobacillus melliventris* and *Lactobacillus kullabergensis*, see **Supplementary Table 1, Annex (II)**). All strains were precultured on solid media and subsequently inoculated in liquid media for overnight growth as described in Manuscript I – Materials and methods.

in vivo colonization and passaging experiments

Bacterial colonization stocks were prepared as described in Manuscript I – Materials and methods. Colonization stocks detail (ID and content) is reported in **Supplementary Table 2** (**Annex (II)**). All colonization stocks were diluted ten times in a 1:1 mixture of 1xPBS and sugar water (50% sucrose solution, w/v) and colonization stocks for mono-colonizations were diluted 50 times more to achieve more comparable bacterial levels as for the co-colonizations. 5 µL of diluted stocks were fed to each bee using a pipette. Bees were kept on a sugar water + pollen diet and food was provided ad libitum. Ten days post-colonization, five rectums were dissected and homogenized in 1xPBS. An aliquot of each homogenized rectum was used for CFU plating to enumerate the total bacterial load and for amplicon sequencing to obtain the relative abundance of each community member.

For the eight Firm5 communities that were passaged across bees to investigate coexistence over time, a total of ten rectums were dissected seven days post-colonization, pooled together, stored and then used to feed another batch of microbiota-depleted bees

as described in Manuscript I – Materials and methods. At the same time, after every passage, an aliquot of each homogenized rectum was used to plate CFUs and to perform amplicon sequencing analysis. This was repeated for a total of three serial passages.

in vitro passaging experiment

Each strain was cultured in liquid medium overnight and subsequently re-inoculated as described in Manuscript I – Materials and methods. Strains were then mixed in equal volumes and inoculated at an OD_{600} = 0.05 in triplicates in 2mL Eppendorf tubes containing 1.350 mL of carbohydrate-free cfMRS medium (Mortera et al. 2013) supplemented with 10 % water extracts of pollen grains. Detailed information about pollen extracts preparation can be found in the Supporting methods section of Kešnerová et al., 2017. Then, 1 % of each culture (i.e. 15 µL) was passaged to a tube with fresh medium and incubated again. These passages were repeated 3 times and after each transfer cultures were pelleted by centrifuging (5 min, 5000 g), supernatants were discarded and pellets were stored at -20 ° C for amplicon sequencing analysis. At every passage an aliquot of each culture was used for CFU plating to enumerate the total bacterial load.

Amplicon sequencing

The relative abundance of each strain across all experiments was obtained using amplicon sequencing of a 199-bp long fragment which allows to discriminate the 12 strains from each other, as described in Manuscript I – Materials and methods. In brief the amplicon, after a lysis step, is amplified either from gut homogenates or from cells pellets resuspended in 1xPBS (200 µL) with a two-step PCR protocol interspersed with purification steps. During the first PCR the amplicon is amplified with specific primers (see Materials and methods of Manuscript I) and during the second PCR sample-specific

barcodes are annealed to it (**Supplementary Table 3, Annex (II)**). For barcode-combinations to strains-combinations correspondence, see **Supplementary Table 4** (https://drive.switch.ch/index.php/s/T5Yru1AnXjEMCfV).

Finally, DNA concentrations are measured and each sample is adjusted to a concentration of 0.5 ng/µL prior of being pooled together. After that, the pooled samples are loaded and purified from an agarose gel, the DNA is prepared for sequencing following the Illumina Miniseq guidelines and then loaded on a Illumina MiniSeq Mid Output Reagent Cartridge using the correspondent MiniSeq flow cell. The sequencing output is analyzed (as described in Manuscript I – Materials and methods) using the scripts available on github and the relative abundance of each strain in each samples is obtained. To obtain absolute abundance data for each strain, we combined the relative abundance data from the amplicon sequencing with CFU counts obtained during the *in vivo* and *in vitro* experiments.

in vivo interactions analysis

To infer interactions for each combination-pair, for example of strain A and strain B, the $\log_2$ of the ratio between the abundance of A in co-colonization with B and the abundance of A in mono-colonization is calculated (**Formula 1**).

$$log2\left(\frac{A\ in\ co-colonization\ with\ B}{mean(A\ in\ mono-colonization)}\right.$$

Formula 1: Calculation effect of one strain on the other comparing mono- and co-colonization abundances.

The resulting number can be either positive (strain B is facilitating strain A), negative (strain B is inhibiting strain A) or it can be "1" (strain B has no effect on strain A). In this study we never obtained a round "1" as a result when using Formula 1, thus neutrality and interactions such as amensalism (one strain inhibited and the other not

affected) or commensalism (one strain facilitated and the other not affected) were not considered. By using this formula in each direction for each strain-pair a measure of the effect of each strain on another is obtained (Effect of A on B; Effect of B on A). By combining each strain's effect on another the interaction type is inferred. If both strains inhibit each other the interaction is defined as mutual inhibition, while if both strains are facilitating each other the interaction is defined as mutual facilitation. If one strain is facilitated and the other is inhibited by the association the interaction is defined as parasitism. Moreover, by combining each strain's effect on another cartesian coordinates to plot and visualize each pair interaction on a cartesian plane are obtained (Effect of A on B; Effect of B on C, see **Supplementary Figure 1 - Annex (II)**). Based on the quadrant in which the point corresponding to the pair-interaction falls the interaction can be assessed: mutual facilitation pairs fall within the (+,+) quadrant, mutual inhibition pairs fall within the (-,-) quadrant and parasitism pairs fall within the (-,+) quadrant. In order to obtain a quantitative measure of interactions we converted cartesian coordinates (x,y) in polar coordinates (r, θ) (**Formula 2**).

$$(r, \theta) = (\sqrt{x\,2 + y2}, tan - 1\,(y\,x))$$

Formula 2: Conversion of cartesian coordinates into polar coordinates.

Every coordinate is thus converted from a point of coordinates x and y into a vector whose length (r) describes the distance of the vector from the center and whose angle (θ) describes the position of the vector on the plane. In particular r can be used as a measure of the strength of the interaction and θ can be used as a measure of which is the type of interaction as it indicates in which plane-quadrant the pair is located as well as a measure of how symmetrically strains are interacting. To determine within which pairs strains were interacting significantly, a ρ measure of the dispersion of the replicate vectors for each pair was calculated. Pycke's test was used to determine which pairs

replicate vectors were concordant and thus indicating that the measured interaction was significant (Pycke 2010).

Code and data availability

The complete custom code for all the analyses is available on GitHub: (https://github.com/silviabrochet/strain-level-interactions). We are currently uploading the amplicon sequencing data under an NCBI Bioproject.

Manuscript III: SCFAs production by different *Lactobacillus* Firm5 species and strains

Summary

In this Manuscript we investigated if to the genetic-diversity that is found in natural bacterial communities there is a correspondent functional-diversity. In other words, we asked if in microbiomes each genotype has a distinct function (making the community functionally synergistic) or if all genotypes have the same function (making the community functionally redundant).

To this end, we selected 12 different strains of the bee-gut phylotype *Lactobacillus* Firm5, 3 for each of the four Firm5 species, and we used them to mono-colonize microbiota-depleted bees. 5 days post-colonization we measured the production of short chain fatty acids (SCFAs) both in the bee gut as well as in the hemolymph. SCFAs are the main products resulting from the microbial fermentation of substrates that are not digestible by the host and they have been shown to be an important energy source for the host. Here we measured the production of the SCFAs acetate, butyrate, formate, lactate, propionate and succinate in bees mono-colonized with individual Firm5 strains if compared to non-colonized bees (microbiota-depleted control, MD).

While in bees hemolymph we could not detect any difference in SCFAs accumulation between colonized (CL) and MD bees, in the bee gut five out of six SCFAs that accumulated in at least one mono-colonization treatment if compared to the MD control. Some SCFAs, such as acetate and succinate, were produced by all Firm5 strains. Others, such as lactate and butyrate, were produced in a rather species-specific way, with two species accumulating lactate and two species accumulating butyrate.

Interestingly, the accumulation of lactate was species-specific only *in vivo* and not *in vitro*, suggesting that all four species have the ability of producing lactate. The fact that the two species that accumulate butyrate instead of lactate *in vivo* harbor specific lactate permeases genes and genes involved in the butyrate metabolism pathway suggests that they are producing lactate and subsequently converting it to butyrate. Finally, there was one SCFA, i.e. formate, that was accumulated in a strain-specific way.

These results suggest that, depending on the type of SCFA, its production can be common across all strains or it can be rather produced in a species/strain-specific way.

Introduction

Microbes play key roles for ecosystem functioning and health. For example, microbial communities associated with ocean and soil are essential to fulfill important functions such as nutrient recycling and carbon storage (Bardgett and Van Der Putten 2014; Moran 2015). Plant-associated microbial communities are crucial for promoting plants health by preventing pathogen-colonization through competition or through the production of toxic compounds (Berendsen, Pieterse, and Bakker 2012). Finally, host-associated microbial communities protect against pathogens and contribute to host nutrition by fermenting non-digestible dietary components into short chain fatty acids SCFAs) that can be absorbed by the host (Flint et al. 2012).

At the same time, microbial communities are genetically diverse, i.e. they typically consist of many different bacterial phyla, families, genera, species and strains. Moreover, the degree of their complexity can vary across different environments (Locey and Lennon 2016). For example it is estimated that while a single gram of soil can harbor up to 50'000 bacterial species, the human gut generally includes a maximum of 400 bacterial species (Lloyd-Price, Abu-Ali, and Huttenhower 2016; Roesch et al. 2010).

How the genetic diversity of microbiomes is linked to their functional output is a standing question in the field (Louca et al. 2018). On one hand, the function of a microbial community depends on the concerted action of functions carried by individual members of the community (Armstrong et al. 2018). On the other hand, many members of a community have identical functional capacities and occupy similar ecological niches indicating that microbial communities contain a lot of functionally redundancy (Moya and Ferrer 2016). For host microbial communities, it is possible that competition for nutrients in the gut drives bacterial diversification resulting in bacteria occupying divergent niches. However, the host may select for functional redundancy in the microbiota to increase the robustness against e.g. pathogen attacks, or the resilience against environmental fluctuations or to produce a more diverse set of host-beneficial bacterial products.

Here, we looked into this question by focusing on the honey bee gut microbiota. Honey bees harbor a specialized and experimentally tractable bacterial community in their gut. This community consists of 8 deep-branching bacterial lineages (or phylotypes) which can be cultured in the lab and that can be used to colonize microbiota-depleted bees (Kwong and Moran 2016; Martinson et al. 2011; Zheng et al. 2018). Here we focused on one of the most abundant phylotypes of the bee gut microbiota, i.e. *Lactobacillus* Firm5. *Lactobacillus* Firm5 is a facultative anaerobe that converts sugars into organic acids (such as SCFAs) via fermentation. In honey bees, this phylotype comprises four divergent species and each of these contain a substantial amount of strain-level diversity (Ellegaard and Engel 2019). In our previous work, we showed that the four Firm5 species consistently coexist in the bee gut due to niche differentiation and resource partitioning of carbohydrates present in the pollen diet (Manuscript I). In contrast, strains of the same species tend to segregate into different host animals and are less likely to coexist due to competitive exclusion (Ellegaard and Engel 2019). Whether differences in species- or strain-level composition of this phylotype result in any differences in the metabolic

output of the microbiota has not been addressed, but could have important implications for the host.

To test if the genetic diversity impacts the functional output of the Firm5 phylotype in the bee gut, we selected three divergent strains for each species and measured their effect on SCFAs production in the rectum and the hemolymph of gnotobiotic bees. We decided to look both at the gut and the hemolymph to assess the direct metabolic output of the community in the gut and its effect on the metabolism of the host.

We found that colonization with bacteria of the phylotype Firm5 affected the SCFA profiles in the gut but not in the hemolymph. Interestingly, while some SCFAs (i.e. acetate and succinate) were produced in similar quantities by all species indicating redundant metabolic output, other SCFAs (i.e. butyrate, lactate and formate) were produced in a species-specific or strain-specific manner indicating that the species- and strain-level diversity both may impact the metabolic output of the gut microbiota in bees.

Results

Colonization with *Lactobacillus* Firm5 strains impacts SCFAs production in the gut but not in the hemolymph

To study the production of SCFAs by *Lactobacillus* Firm5 in the gut and in the hemolymph we performed two independent replicates of the same gnotobiotic bee experiment. In both experiments, we mono-colonized microbiota-depleted (MD) bees with 12 different strains of *Lactobacillus* Firm5, three for each of the four species (see Manuscript II) (**Figure 1.A**). Five days post-colonization, we dissected the rectum and determined the bacterial loads based on CFUs plating (**Figure 1.A** and **Figure 1 - figure supplement 1, Annex (III)**). In both experiments all strains successfully colonized reaching bacterial abundances higher than $5x10^7$ cells per rectum. As expected,

no culturable bacteria were detected in the MD bees control (**Figure 1 - figure supplement 1, Annex (III)**). Homogenized rectum tissue and hemolymph from experiment 1 and experiment 2, respectively, were subjected to GC-MS analysis to measure the production of six SCFAs (acetate, butyrate, formate, lactate, propanoate and succinate). These SCFAs were selected based on a previous study about SCFAs in the honey bee gut (Zheng et al. 2017) (**Figure 1.A**).

There was a clear difference in the SCFAs profiles of the rectum and the hemolymph samples (**Figure 1.B** and **C**). The SCFAs profiles of colonized (CL) bee rectums clustered separately (first PCA axis) from the MD ones, while the all replicates tended to cluster together within treatments (**Figure 1.B**). Moreover, in the rectum the SCFAs profiles of the bees colonized with the different Firm5 strains clustered separately according to species-affiliation on the second PCA axis (**Figure 1.B**). In the rectum 5 out of 6 SCFAs (all but propanoate) were significantly accumulated in the CL bees if compared to the MD control (**Figure 2**). On the other hand, the SCFAs profiles that we detected within the hemolymph of CL bees did not cluster separately from the ones of MD bees and neither by species (**Figure 1.C**). Moreover in the hemolymph none of the 6 measured SCFAs was accumulated significantly in the CL bees if compared to the MD control (**Figure 2 - figure supplement 1, Annex (III)**).

These results suggest that the colonization with *Lactobacillus* Firm5 effect SCFAs production in the bee rectum rather than in the bee hemolymph.

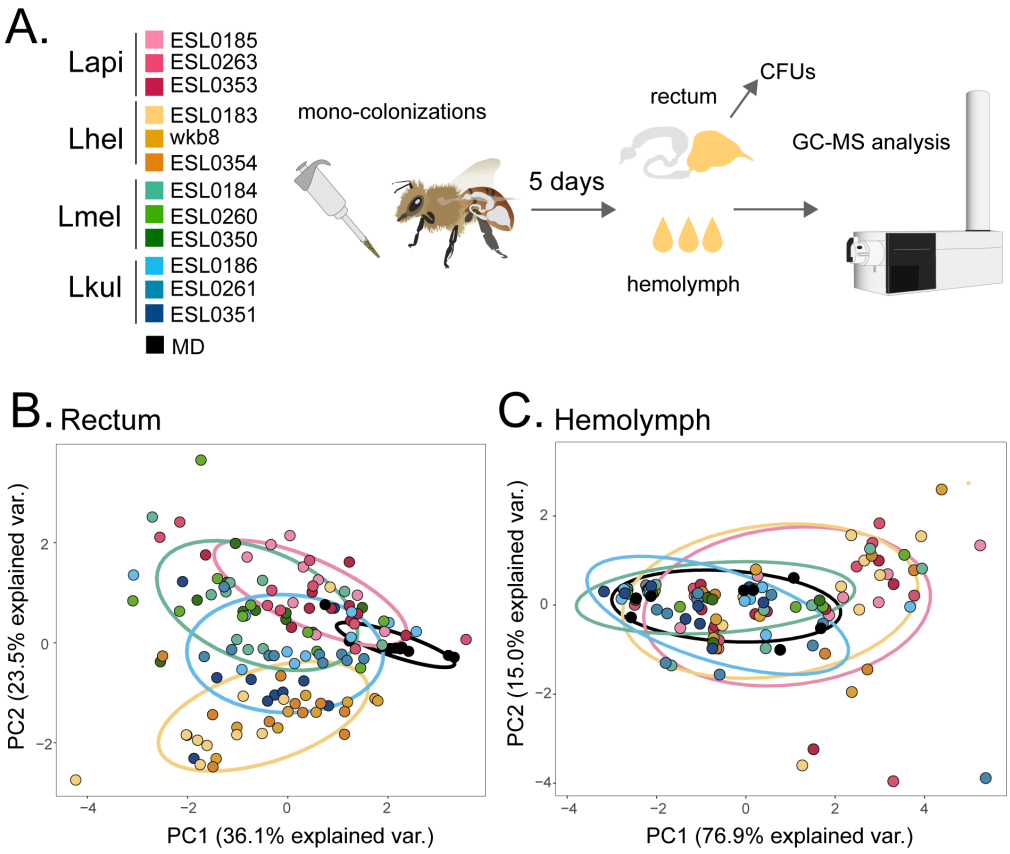


Figure 1: Production of SCFAs by *Lactobacillus* Firm5 strains. A. Scheme of the experimental design. Two independent replicates of the same gnotobiotic bee experiment were carried out. Microbiota depleted bees were mono-colonized with 12 strains of *Lactobacillus* Firm5 belonging to the four *Lactobacillus* Firm5 species (Lapi, Lhel, Lmel, Lkul). Five days post colonization, the rectum of each bee was dissected and an aliquot was used to plate CFUs. Aliquots of bee rectums or bee hemolymph (respectively from experiment 1 and experiment 2, n = 10) were subjected to GC-MS analysis to detect the presence of SCFAs. **B.** Principal component analysis (PCA) of SCFAs detected in the rectum samples and **C.** PCA of SCFAs detected in hemolymph samples. The first two PCA dimensions are displayed (PC1 and PC2). SCFAs responses were normalized by the internal standard and the gut mg or the hemolymph µL and subsequently converted to z-scores (see Materials and methods). All replicates are displayed (each point represents a bee, n = 10). The ellipses delimit the different species (different colors) and the MD control (black).

Different patterns of specificity for SCFAs production in the bee gut

We then focused on SCFAs production within the rectum in order to investigate differences within the Firm5 phylotype.

Acetate and succinate accumulated in the rectum in all colonizations relative to MD bees (**Figure 2**). In contrast, lactate only accumulated in bees colonized with the strains of the two species Lapi and Lmel while butyrate exclusively accumulated in bees

colonized with the strains of the other two Firm5 species, Lhel and Lkul (two out of three strains for Lkul) (**Figure 2**). Concordantly, lactate and butyrate were the two SCFAs that were driving most of the separation of the different Firm5 species in the rectum across the second PCA axis (**Figure 1 - figure supplement 2, Annex (III)**). Finally, formate was produced in a strain-specific way. Two strains of Lapi, all strains of Lhel, one strain of Lmel and two strains of Lkul produced this SCFA (**Figure 2**).

To link the differences in SCFA profiles to differences in genomic features of different Firm5 strains, we looked at the KEGG gene annotation of the publicly available genomes of the 12 strains. All Firm5 strains harbored the genes to produce acetate and lactate starting from pyruvate (**Figure 2 - figure supplement 2, Annex (III)**). Notably, the strains of Lhel and Lkul, which did not increase lactate levels in the gut of CL bees, harbored a gene encoding a lactate permease that can import extracellular lactate from the environment (**Figure 2 - figure supplement 2, Annex (III)**). Most of the genes of the butyrate metabolism were absent from Firm5 genomes, except for the gene encoding for an enoyl-CoA hydratase (converting hydroxybutyryl-CoA to crotonyl-CoA) in all strains of Lhel and in one strain of Lkul (ELS0351, **Figure 2 - figure supplement 2, Annex (III)**). Interestingly, these were precisely the strains which showed increased butyrate levels. Although propanoate levels were not increased in CL vs MD bees, all strains harbored the genes to produce propanoate starting from the oxidation of odd-chain fatty acids and a subset of amino acids (**Figure 2 - figure supplement 2, Annex (III)**). Moreover, all Firm5 strains had almost all the genes to produce the TCA cycle intermediate succinate (**Figure 2 - figure supplement 2, Annex (III)**), yet there were differences in the amount of succinate produced in the presence of different strains. Finally, Firm5 strains lack genes to convert pyruvate to formate despite the fact that some strains were able to accumulate this SCFA (**Figure 2 - figure supplement 2, Annex (III)**).

Differences in species-specific lactate accumulation *in vivo* vs *in vitro*

All Firm5 species harbor the genes necessary to produce lactate as a result of sugars fermentation (**Figure 2 - figure supplement 2, Annex (III)**) and we showed previously (Manuscript I) that all four Firm5 species can produce lactate *in vitro* (ESL0185, ESL0183, ESL0184 and ESL0186). Why do we see lactate production by all species *ex vivo* but only by two out of four species *in vivo*? It could be that lactate is accumulated by all four species *in vivo* as well and later re-absorbed to be reconverted to another product. Interestingly, the two species that don't accumulate lactate *in vivo* harbor specific lactate permeases that could promote the reabsorption of lactate (**Figure 2 - figure supplement 2, Annex (III)**). Alternatively, it could be that solely the lactate produced by Lhel and Lkul is taken up by the host. This could be possible if, for instance, Lhel and Lkul produce another isomeric form of lactate than Lapi and Lmel. Since GC-MS analysis cannot discriminate between the two isomeric forms of lactate (D-/L-lactate), to test this hypothesis we used a kit to measure the accumulation of D/L-lactate both *in vitro* and *in vivo* in presence of a lactate-accumulating strain (Lapi ESL0185) or non-accumulating strain (Lhel ESL0183) (**Figure 2 - figure supplement 3, Annex (III)**). While the results of this enzymatic assays were in accordance with our GC-MS analysis for the species-specific production of D-lactate *in vivo*, there were no differences in the isomeric form of lactate as all strains produced D-lactate only (**Figure 2 - figure supplement 3, Annex (III)**).

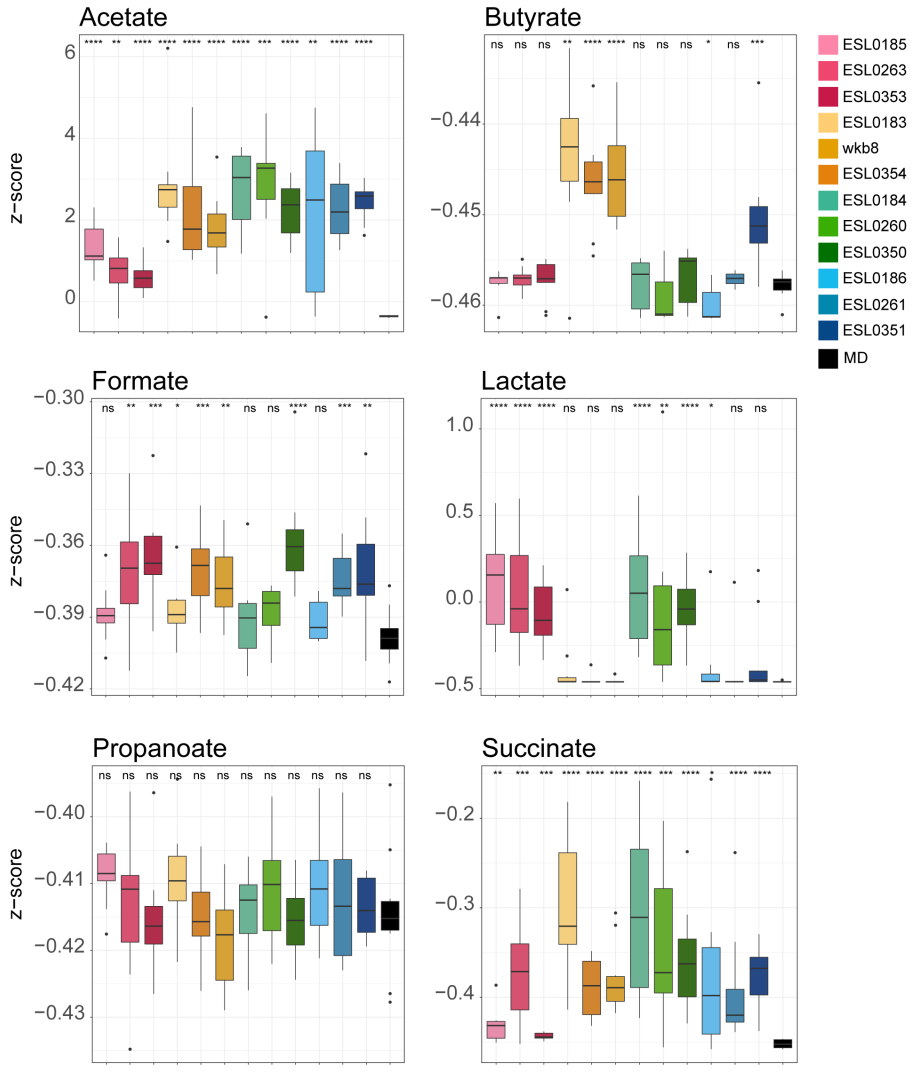


Figure 1: SCFAs production in the bee rectum upon colonization with different *Lactobacillus* Firm5 strains. Boxplots displaying the z-score for each measured SCFA for each treatment (see Materials and methods). Significant differences between the Firm5 treatments and the microbiota-depleted (MD) control was measured using Wilcoxon signed rank tests (ns: p > 0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001; ****: p <= 0.0001).

Discussion

In this study we investigated if different species and strains of the bee gut phylotype *Lactobacillus* Firm5 are functionally different by measuring SCFAs production in the bee gut and hemolymph.

We found SCFAs production (significant accumulation in CL vs MD bees) in bees rectums but not in hemolymph. This is in contrast with a previous study by Zheng et

al. where the authors showed that, although it's lower than in bees gut, there is an impact of the gut microbiota on the bees hemolymph metabolic profile (Zheng et al. 2017). In particular Zheng et al. could detect the SCFA butyrate only in the hemolymph of CL bees (Zheng et al. 2017). So why we didn't detect any SCFAs production in CL bees hemolymph? It could be because, while Zheng et al. used the entire community and we only used individual Firm5 strains, the effect could not be strong enough. In fact, it could be that the individual Firm5 strains produce very low amounts of SCFAs that are absorbed and used up so fast in the bee hemolymph that we cannot detect them. It could also be that the strains that we tested are not responsible for SCFAs accumulation in bees hemolymph or that our hemolymph sampling method does not allow to detect SCFAs accumulation as we could only sample limited amounts of material per bee.

As for SCFAs production within the bee gut, previous studies showed that, in presence of the entire community, acetate and succinate are the most abundant SCFAs produced by bee-gut bacteria (Callegari et al. 2021; Zheng et al. 2017). In addition, succinate was shown to be a cross-fed metabolite between two bee-gut symbionts (Kešnerová et al. 2017). Interestingly, acetate and succinate are the two SCFAs that were accumulated by all the Firm5 strains that we tested. While all strains harbor the complete gene-pathway to produce acetate, they miss few genes involved in the production of the TCA cycle intermediate succinate (in particular for the conversion of pyruvate into oxaloacetate). This could be because of insufficient/not accurate gene annotation or because they use alternative genes to produce this SCFA.

On the other hand, we showed that other SCFAs such as lactate, butyrate and formate are accumulated in a species- or strain-specific way in the bee gut. This is the first evidence that to the intra-phylotype genetic-diversity of *Lactobacillus* Firm5 there is a correspondent functional diversity. Previous studies showed that there are differences between hive-bees in the Firm5 composition both at the species and as well as at the strain level. In fact, Ellegaard and Engel (2019) showed that the relative abundances of

Lactobacillus Firm5 species in the gut can change depending on bees age and that different bees can display different strain-profiles, as strains (especially from the same species) are likely segregating across bees due to competition (Ellegaard and Engel 2019). This makes very relevant the fact that Firm5 species and strains are functionally different as it could mean that bees have functionally distinct microbiotas depending on the age group and even depending on the individual bee. Further studies are needed to investigate the impact of the microbiota functional differences on host physiology.

In particular, lactate was accumulated in the bee gut only by two out of four Firm5 species, i.e. Lapi and Lmel. The two other species, Lhel and Lkul, complementarily accumulated another SCFA, i.e. butyrate. One explanation for this could be that Lhel and Lkul produce lactate and then they use their species-specific lactate permease to re-absorb lactate from the extracellular space and reconvert it to pyruvate and subsequently to butyrate (Bourriaud et al. 2005; Detman et al. 2019). Interestingly, all strains from these species (except one) harbor a specific gene encoding for an enzyme involved in the butyrate metabolism pathway. As for why this difference in lactate accumulation is only found *in vivo* and not *in vitro* is not yet clear. In this study we demonstrated that this is not due to the differential absorption by the host of different lactate isomers produced by Lhel and Lkul. Instead, one possibility could be that we are sampling at different time-points of growth between the *in vivo* and *in vitro* experiments. Alternatively, it could be that *in vivo* the host environment triggers the expression of the lactate transporters. In support to this idea, in a previous study we showed that lactate permeases are among the genes upregulated by two Lhel and Lkul strains *in vivo* in presence of pollen if compared to in absence of pollen (See Manuscript I). Another possibility is that the difference in pH between the *in vivo* and the *in vitro* conditions is responsible of the differences observed in lactate accumulation. In fact Wang et al. recently showed that lower pH levels led to lactate accumulation and butyrate production reduction (Wang et al. 2020).

Finally, we found that formate was produced in a strain-specific way within the Firm5 phylotype. This finding, together with our findings of Manuscript II, i.e. that interactions and coexistence dynamics can be strain-specific, strongly suggests that there could be a subset of Firm5 strain that, independently of their species affiliation, have unique characteristics.

This study presents some limitations. In fact here we are not considering the entire bee-gut microbiota community but just a subset of strains from one of the eight phylotypes. It could be that the SCFAs profiles in the gut (and in the hemolymph) are different in presence of the entire conventional bee community (as showed in Zheng et al., 2019 and Callegari et al., 2021). This could be due to the fact that different phylotypes could be capable of producing additional SCFAs or that there could be cross-feeding interactions between them. Moreover, here we are measuring SCFAs production in the gut at a given time-point post-colonization while there could be differences in SCFAs accumulation before or after this time-point.

In this study we also investigated about the correspondence between the differences in SCFAs production and the differences in gene content within the Firm5 phylotype. However, only for acetate and lactate accumulation profiles in the gut we found consistent annotations in the correspondent metabolic pathways. As for the other SCFAs, we found either little or no correspondence between their accumulation profiles and their gene annotations for the correspondent metabolic pathways. This could be due to the fact that their gene-annotations might be missing, wrong or they not accurate enough. For example it could be that some enzymes, even if they can potentially catalyze more than one reaction, have a unique annotation and they are included in a single pathway.

Future studies should focus on uncovering the impact that differential SCFA production profiles in the gut could have on interbacterial interactions and on the bee physiology. Our results indicate that that Firm5 species and strains are not functionally

equal and thus their presence could affect in different ways the other members of the bee-gut community and the host. Lactate, for example, could be an important SCFA in this regard. In fact, it was shown that high levels of lactate in the gut are detrimental as this SCFA has the potential to lower the pH leading to acidosis, colitis and changes in the microbiota leading to the proliferation of pathogens that benefit from the perturbed gut environment (Gillis et al. 2018; Vernia et al. 1988). Lactate does not accumulate in the gut of healthy individuals thanks to the lactate-utilizing bacteria that can use lactate for growth (in particular *Firmicutes*) converting it into butyrate and propanoate (Duncan, Louis, and Flint 2004). It could thus be that the lactate 'homeostasis' in the bee-gut is maintained thanks to the equilibrium between the four coexisting Firm5 species: two lactate accumulators and two lactate converters.

Materials and methods

Culturing of bacterial strains

For this study we used the same 12 *Lactobacillus* Firm5 strains that we used in Manuscript II, 3 for each of the four species (*Lactobacillus apis*, *Lactobacillus helsingborgensis*, *Lactobacillus melliventris* and *Lactobacillus kullabergensis*, see Manuscript II, **Supplementary Table 1, Annex (II)**). All strains were cultured as described in Manuscript I – Materials and methods.

in vivo colonization

Bacterial colonization stocks for mono-colonizations using each of the 12 strains individually were prepared as described in Manuscript I – Materials and methods. Bees were colonized as described in Manuscript I and II (see Materials and methods) and were

kept on a sugar water + pollen diet. For the rectum experiment, five days post-colonization, ten rectums were dissected per treatment and homogenized in 1xPBS. For each rectum-homogenate sample, 300 µL were collected and centrifuged (15,000 g, 4°C, 15 min). Then, 200 µL were transferred to a new tube, snap-frozen in liquid nitrogen, and stored at −80°C prior GC-MS analysis. For the hemolymph experiment, five days post-colonization hemolymph was extracted from 10 bees per treatment as described in Borsuk et al. (Borsuk et al. 2017), snap-frozen in liquid nitrogen and stored at −80°C prior GC-MS analysis. In both experiments an aliquot of each homogenized rectum was used for CFU plating to enumerate the total bacterial load.

GC-MS analysis

Once that all the samples were collected, SCFAs metabolites were extracted. To extract soluble metabolites, tubes were thawed on ice, and 100 µL of sample was combined with 20 µL of 1 mM internal standard (^{12}C isovalerate - Cambridge Isotope Laboratories, Inc). The mixtures of samples and standard were acidified with 5 µL of 11% HCl and 500 µL of Diethyl Ether were added immediately after. Samples were incubated for 10 min in a thermoblock C (Eppendorf) (1 ˚ C, 2000 rpm shaking). Tubes were centrifuged (13'000 g, 4 ˚ C, 5 min) and 80 µL of the upper phase were transferred to glass GC-MS vials with 0.3 mL glass insert (Agilent).

Derivatization was performed by adding 20 uL MTBSTFA + 1% t-BDMS (Sigma-Aldrich), vortexing briefly and then incubating for 1h at 30°C. Samples were run on a single quadrupole GC-MS system (Agilent 8890-5977B) equipped with an autosampler. Samples were injected (1µL) in split mode (15:1) at an inlet temperature of 150 °C with a helium gas column flow rate of 1 ml/min on an Agilent VF-5ms column (30 m, 0.25 mm, 0.25 µm). The oven was held initially at 35 °C for 2 min and ramped at 5 °C / min to a final temperature of 250 °C for 1 min. The MS was operated in selected ion monitoring (SIM) mode.

Peaks from the total ion chromatogram (TIC) were identified by matching retention times and spectra to an in-house library. Peaks were picked and integrated using the Agilent MassHunter Quantitative Analysis software. Peak areas were normalized to the internal standards and to the rectum mg or to the hemolymph µL. Z-scores were calculated by multiplying the normalized data by the average of all the normalized data and dividing by the standard deviation across all the normalized data.

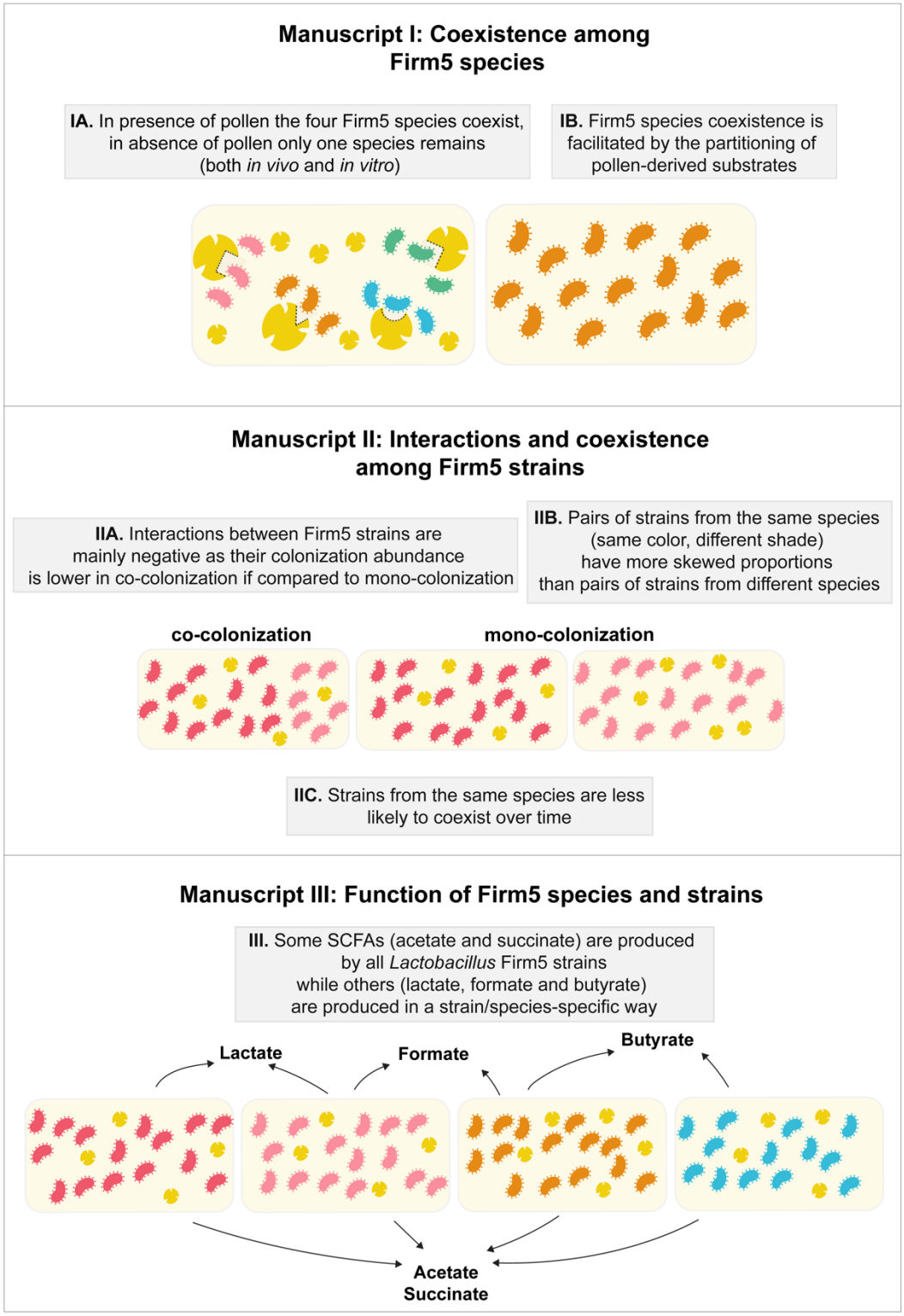
D-/L-lactate assays

D- and L-lactate isomers were detected within bee gut homogenates and bacterial liquid cultures using the D-/L-lactic Acid (D-/L-lactate) (Rapid) Assay Kit (Megazyme). The samples were either bees rectums that were colonized with individual strains for 5 days (as described above) and homogenized in 1XPBS or liquid cultures of individual strains after overnight growth in cfMRS supplemented with pollen extracts (see Materials and methods Manuscripts I and II). Briefly, 10 µL of sample, H_2O (negative control) or D-/L-lactic acid standard solution (positive control) were added to a suspension including the kit's buffer, NAD^+ and D-Glutamate-pyruvate transaminase suspension and the absorbances were read at 340 nm. After the addition of either L-lactate dehydrogenase to detect L-lactate or D-lactate dehydrogenase to detect D-lactate, the reaction of lactate oxidation to pyruvate with the conversion of NAD^+ in $NADH+H^+$ was produced. The amount of NADH formed in the above coupled reaction is stoichiometric with the amount of D or L-lactic acid. The difference of absorbance given by the NADH accumulation is thus measured (absorbance at 340 nm) and the g/L of the different lactate isomers is calculated as described by the kit manufacturer.

Code and data availability

The complete custom code for all the analyses is available on GitHub: (https://github.com/silviabrochet/SCFAs_production).

General Discussion



Schematic view of the main results of this thesis divided by Manuscript.

This thesis contributes to the question: Why are there so many kinds of bacteria?

As mentioned in the introduction, around 60 years ago the ecologist George Evelyn Hutchinson posed a question that became central in ecology and that still remains largely unanswered: *why are there so many kinds of animals*? (Hutchinson 1959). During my PhD I aimed at contributing to answer to this question and, as a microbiologist, I transposed it to the microbial world asking: *why are there so many kinds of bacteria*? More precisely, I asked how bacteria coexist and interact within communities and if this depends on their relatedness. Moreover, I asked if genetically distinct bacteria are also functionally distinct within communities.

To address these questions I used as a model the bacterium *Lactobacillus* Firm5, one of the symbionts of the bacterial community associated with gut of the honey bee (*Apis mellifera*). The honey bee gut microbiota is a highly relevant study model not only because it is associated with one of the most widespread pollinators worldwide (and it can impact its physiology, see Zheng et al., 2017, and protect it from pathogens, see Steele et al. 2021) but also because it is highly experimentally tractable (Johnson 2013; Zheng et al. 2018). In fact, this community is an ideal model to study bacterial coexistence and interactions both *in vivo* and *in vitro* (Kešnerová et al. 2017). In this thesis I selected the bee-gut phylotype *Lactobacillus* Firm5 because it is one of the most abundant members of the community and one of the most divergent ones, as it diversified into species and strains that can be cultured in the laboratory and used to colonize microbiota-depleted bees (Ellegaard et al. 2015; Ellegaard and Engel 2019; Kešnerová et al. 2017). This phylotype is thus particularly fit to study patterns of coexistence, interactions and function among closely related bacteria within communities.

The choice of *Lactobacillus* Firm5 was also motivated by the will of experimentally validate what was suggested in a recent metagenomics study from our lab (Ellegaard and Engel 2019). This study showed that *Lactobacillus* Firm5 species are consistently detected within individual bee guts while exhibiting high levels of genomic variation in terms of carbohydrate metabolism, suggesting that their coexistence might be facilitated by the adaptation to different niches (Ellegaard et al. 2015; Ellegaard and Engel 2019). In contrast, strains within species were proposed to be less likely coexisting as they were tending to segregate across individual bees (Ellegaard and Engel 2019). If this is the case, i.e. that different Firm5 species always coexist while strains within species do not, individual bees could have either functionally similar bacterial communities if functions are common to all species (or species-specific) either functionally distinct profiles if each strain is functionally different (Ellegaard and Engel 2019). Given these premises, we experimentally investigated coexistence, interactions and functions among the different *Lactobacillus* Firm5 species and strains.

Negative interactions dominate among *Lactobacillus* Firm5 strains

When we used four strains from different Firm5 species (Manuscript I) and pairs of conspecific or allospecific strains (Manuscript II) we observed that negative interactions were dominant, as the abundance of Firm5 strains in mono-colonization was always higher than in co-colonization. This could be because Firm5 strains are compelled to compete in order to partition a niche (the bee rectum) whose carrying capacity can be attained in mono-colonization. Our results thus strongly suggest that competition is dominant and the default interaction-type in the bacterial world (as it was also pointed out by Foster and Bell (2012) and Kehe et al. (2021)). Still, the fact that the interactions that we measured at one fixed time point are negative, cannot be directly linked to the ability of strains to coexist. To gain insights about coexistence (and the mechanism

allowing it) it is needed to serially passage communities and observe dynamics over time. The number of time-points (or passages) to consider is likely different depending on the community type. One way to determine the appropriate number of passages to perform is to measure when the community reaches stability by grouping the passages into sliding windows (as we did in Manuscript I).

Resource partitioning facilitates coexistence at the species level

Even if the four Firm5 species were negatively interacting, in Manuscript I we showed that they were at the same time able to coexist thanks to the partitioning of different pollen-derived substrates by passaging them through gnotobiotic bees in the presence or in the absence of pollen. Interestingly, we showed that the four species have sort of a streamlined way of using different pollen-derived substrates that is set in motion in response to the presence of pollen and not to the presence of other species. This together with the facts that i) Firm5 species harbor a large repertoire of genes for the transport and utilization of pollen-derived carbohydrates and ii) Firm5 species have never been isolated from other environments than the bee gut, ultimately suggest that these species have co-evolved within the gut adapting strongly to and specializing on the pollen diet of the bee.

It is puzzling to think that on the other hand in *Apis cerana*, a closely related bee species to *A. mellifera* that feeds on pollen as well, there is only one species of *Lactobacillus* Firm5 (specific to this bee type) (Ellegaard et al. 2020). Why are these two bee species so different in the diversity of this phylotype? Why in *A. cerana* we don't see the parallel evolution of different Firm5 species (with different pollen-preferences) as in *A. mellifera*? It could be that, as suggested by Ellegaard et al., the *A. mellifera* gut-community is more diverse if compared to *A. cerana* because *A. mellifera* has a more generalist pollen-diet including different floral pollen-types or because it occupies a larger

geographical area (species-area relationship) (Ellegaard et al. 2020). It could also be that in *A. cerana* the divergence within the Firm5 phylotype is still ongoing and that new species (and strains) will appear.

Conspecific strains are less likely to coexist than allospecific strains

Concordantly with our four-strain experiments, our pairwise-interaction experiments showed a tendency of strains from different species to maintain relatively equilibrated proportions and our pairwise-passaging experiments showed that they were able to coexist over time. On the contrary, strains from the same species had more biased proportions within pairs, with one strain often dominating over the other, and three out of four of the pairs tested not able to coexist over time.

Although these results show that allospecific strains are more likely to coexist than conspecific strains, one of the pairs of conspecific strains that we tested was able to consistently coexist over time (both *in vivo* and *in vitro*). This finding needs to be carefully considered as it suggests that by classifying bacterial strains into species we are not classifying them based on the ecological niche that they occupy. This could be due to the fact that these strains genomes are dynamic and in fact could still be involved in a process of diversification for example by horizontal gene transfer, as it was suggested by Ellegaard et al. (2019). Moreover, this result is in accordance with what was recently pointed out in a study by Goyal et al., where the authors conclude that even strains differing by few SNPs differences can result in very different community-dynamics over-time (Goyal et al. 2021). Altogether these results highlight the importance of studying strain-level diversity in bacterial communities.

Different Firm5 species and strains within species produce different SCFAs

Our results show that Firm5 species and strains not only differ in substrate-utilization and coexistence dynamics, but are also to some extent functionally distinct as they can produce different types of short chain fatty acids (SCFAs). This, similarly as before with coexistence, suggests that classifying strains into species can give us only a limited definition of their functional capabilities. These differences in strains functional profiles could be very relevant in that two bees harboring different Firm5 strain-profiles would consequently be functionally different, which could have repercussions on the other community members as well as on the physiology of individual hosts.

Regarding the species-specific patterns of SCFAs production that we observed, it is tempting to think that as among the four coexisting Firm5 species two are producing lactate and two are producing butyrate (likely following lactate reconversion), this could create a balance in the production-conversion rates of lactate. This might consequently help to maintain adequate pH levels in the gut and thereby facilitate the community stability or having a beneficial effect on the host. However, these conclusions are still speculative and would need to be validated for example by investigating if a community dominated by lactate accumulators and lacking lactate converters is more susceptible to pathogen-invasion. At the same time, it would be relevant to test if non-regulated lactate accumulation in the gut can have an effect on the physiology of the bee host.

Take home messages

In summary, our results highlight the importance of studying natural bacterial communities both at the species as well as at the strain level. In fact, we showed that while coexistence among *Lactobacillus* Firm5 species is the rule and it relies on resource partitioning, coexistence within species is not common but can happen and it could be

specific to a set of species or strains. Moreover, we showed that not only coexistence but also function can change depending on the species and on the strain that is present. These are aspects that can influence both the other community-members as well as the host (or eventually the environment in which the community is found). In addition, it looks like communities of Firm5 strains pairs assemble following deterministic rules. In fact, the assemble of Firm5 communities at the strain-level do not depend on their initial abundance. The assembly could thus be the result of species/strain-specific interbacterial interactions or, in alternative, it could be due to a host-promoted filtering-effect.

Secondly, our results hint that coexistence and competition are not incompatible bacterial 'behaviors': on the contrary, bacteria can coexist thanks to resource partitioning due to diet-specialization while competing at the same time (e.g. for space and for other common limiting resources). It could be that this community, in which there are very likely longstanding evolutionary relations between the different members, evolution selected for compatible genotypes that can be maintained either at the level of individual bees or in the colony or that have evolved to achieve a balance between competition and coexistence.

Future directions

How generalizable are our findings to the other bee-gut phylotypes?

In this study we focused on one of the eight phylotypes of the honey bee gut microbiota. In order to understand how generalizable are the results that we obtained with Firm5, future research should expand to study species- and strain-level diversity in the other bee-gut phylotypes. To this end, the amplicon sequencing approach that we used to study different Firm5 strains could be applied to other phylotypes as well. In fact, diversity within phylotypes is also common across the other bee-gut members:

Gilliamella diverged into three different species and Bifidobacterium and *Lactobacillus* Firm4 diverged into two different species (Ellegaard and Engel 2019).

It could be that we would observe different types of bacterial interactions and coexistence dynamics in phylotypes that colonize different compartments of the bee gut. This could be because, depending on the gut-region, the niche can change for example in terms of available nutrients and space. Moreover, we know that while Firm5 is mainly found in the luminal space of the rectum, other phylotypes such as Gilliamella are found adhering to the epithelial surface of the ileum (Zheng et al. 2018), which could suggest that interactions and coexistence within these latter phylotypes are more influenced by the host (as they are found in direct contact with it) than in Firm5. On the other hand, it could be that we would observe similar patterns of interactions and coexistence in phylotypes that colonize the same niche as Firm5, i.e. Bifidobacterium. Interestingly, also Bifidobacterium, as Firm5, was found to harbor genes encoding key functions for the uptake and the degradation of polysaccharides such as cellulose and hemicellulose (Ellegaard and Engel 2019).

In addition, by including other bee-gut phylotypes we could investigate if and how interactions and coexistence among Firm5 species and strains are influenced by their presence. It could be that the presence of strains from other phylotypes would restrict the amount of available niches and thus increase competition and mutual exclusion. It could also be that cross-feeding interactions might arise across different phylotypes, either in a species or strain specific way.

What is the effect on or of the host?

Future studies could investigate more on if and how species- and strain-level diversity impacts host physiology. It would be important to get insights on whether the amount of species and strain-level diversity has an effect (positive or negative) on the honey bee, as this insect is a very important pollinator whose activity was recently

threatened by high seasonal mortality rates (Johnson 2013). Moreover, it would be relevant to understand if certain combinations of strains are more beneficial or detrimental to the host physiology. To this end, there is the need of finding new and meaningful ways to measure bee physiological parameters that could be impacted in response to changes to the microbiota. A promising method are tracking systems that monitor how bees behave and interact with nestmates (Geffre et al. 2020). Studying how and if diversity has an impact on honey bees could be an important contribute to understand how to improve bees health.

In addition, future research should also focus on investigating the possible influence of the host on the community assembly as well as on bacterial interactions and coexistence. In this study we did this to some extent as we tested bacterial coexistence both *in vivo* and *in vitro* at the species level. In this way we could confirm that coexistence at the species level in Firm5 is facilitated by the partitioning of pollen and not by the host. It would be necessary to screen also interactions and coexistence at the strain level *ex vivo* to understand if the host plays a role at this level of community-diversity.

Which strain-number combinations is more insightful?

In this thesis we studied coexistence between species and strains combining *Lactobacillus* Firm5 isolates by four (one combination) in Manuscript I and by two (pairwise) in Manuscript II. Using more combinations of strains by four would be useful to understand if four allospecific strains will always coexist while for example a combination of two allospecific and two conspecific strains would result in the presence of only two strains by the end of the experiment (one per species). Moreover, it could be that for example the conspecific pair that we see coexisting in our experiment could not coexist if there were also strains representatives of other species as they could decrease

the amount of available niches. Still, this kind of experiment is extremely laborious, as the possible combinations by four that should be tested considering the 12 Firm5 strains from the four species are almost five hundred. Because by four combinations are so numerous, it is more convenient experimentally to test interactions and coexistence between strains in a pairwise manner and then eventually measure higher order interactions by inferring models using pairwise interactions data (Friedman, Higgins, and Gore 2017).

Still, even considering pairwise combinations of strains as we did in our case, the experiments sample size, due to experimental limitations, remains limited. A solution could be to test all 12 Firm5 strains (or even more strains) at the same time as a unique multi-species/strain community in several replicates and observe if, after passaging, there is only one strain left per species or multiple depending on the species. Moreover, one of the factors limiting the sample size of our experiments is the number of replicate bees that can be handled in a single run. In addition, bee experiments can be performed only during the summer season, when it is possible to harvest microbiota depleted bees. One way around this would be thus to try to screen for bacterial interactions and coexistence *in vitro* first.

To conclude, in this thesis we highlight the importance of studying closely related bacteria within microbial communities. No matter how closely related bacterial species and strains are, they can still be highly different in how they assemble and thrive within a community. Future research should continue to focus on studying this level of diversity to better understand community dynamics within the bee gut community and beyond.

References

Armstrong, Zachary et al. 2018. "Metagenomics Reveals Functional Synergy and Novel Polysaccharide Utilization Loci in the Castor Canadensis Fecal Microbiome." *ISME Journal* 12(11): 2757–69.

Arrhenius, Olof. 1921. "Species and Area." *British Ecological Society* 9(1): 95–99.

Atlas, Ronald M. 1984. "Diversity of Microbial Communities." *Advances in Microbial Ecology*: 1–47.

Babendreier, Dirk et al. 2007. "Bacterial Community Structures in Honeybee Intestines and Their Response to Two Insecticidal Proteins." *FEMS Microbiology Ecology* 59(3): 600–610.

Bäckhed, Fredrik et al. 2005. "Host-Bacterial Mutualism in the Human Intestine." *Science* 307(5717): 1915–20.

Baran, Richard et al. 2015. "Exometabolite Niche Partitioning among Sympatric Soil Bacteria." *Nature Communications* 6.

Bardgett, Richard D., and Wim H. Van Der Putten. 2014. "Belowground Biodiversity and Ecosystem Functioning." *Nature* 515(7528): 505–11.

Bell, Thomas et al. 2005. "Larger Islands House More Bacterial Taxa." *Science* 308(5730): 1884.

Berendsen, Roeland L., Corné M.J. Pieterse, and Peter A.H.M. Bakker. 2012. "The Rhizosphere Microbiome and Plant Health." *Trends in Plant Science* 17(8): 478–86.

Biller, Steven J., Paul M. Berube, Debbie Lindell, and Sallie W. Chisholm. 2015. "Prochlorococcus: The Structure and Function of Collective Diversity." *Nature Reviews Microbiology* 13(1): 13–27.

Blasche, Sonja et al. 2021. "Emergence of Stable Coexistence in a Complex Microbial Community through Metabolic Cooperation and Spatio-Temporal Niche Partitioning." *Nature Microbiology* 6: 196–208.

Borsuk, Grzegorz et al. 2017. "A New Method for Quick and Easy Hemolymph Collection from Apidae Adults." *PLoS ONE* 12(1): 1–9.

Bourriaud, C. et al. 2005. "Lactate Is Mainly Fermented to Butyrate by Human Intestinal

Microfloras but Inter-Individual Variation Is Evident." *Journal of Applied Microbiology* 99(1): 201–12.

Brockhurst, Michael A., Andrew Fenton, Barrie Roulston, and Paul B. Rainey. 2006. "The Impact of Phages on Interspecific Competition in Experimental Populations of Bacteria." *BMC Ecology* 6: 1–7.

Callahan, Benjamin J., Paul J. McMurdie, and Susan P. Holmes. 2017. "Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis." *ISME Journal* 11(12): 2639–43.

Callegari, Matteo et al. 2021. "Compartmentalization of Bacterial and Fungal Microbiomes in the Gut of Adult Honeybees." *npj Biofilms and Microbiomes* 7(1).

Caporaso, J. Gregory et al. 2012. "The Western English Channel Contains a Persistent Microbial Seed Bank." *ISME Journal* 6(6): 1089–93.

Caro-Quintero, Alejandro, and Konstantinos T. Konstantinidis. 2012. "Bacterial Species May Exist, Metagenomics Reveal." *Environmental Microbiology* 14(2): 347–55.

Chase, J.M, and M.A Leibold. 2009. "Ecological Niches." *University of Chicago Press*.

Cheong, J. Z.Alex et al. 2021. "Priority Effects Dictate Community Structure and Alter Virulence of Fungal-Bacterial Biofilms." *ISME Journal* 15(7): 2012–27.

Chun, Jongsik et al. 2018. "Proposed Minimal Standards for the Use of Genome Data for the Taxonomy of Prokaryotes." *International Journal of Systematic and Evolutionary Microbiology* 68(1): 461–66.

Cox-Foster, Diana L et al. 2007. "Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder." *Science* 318(October): 283–88.

Dal Co, Alma et al. 2020. "Short-Range Interactions Govern the Dynamics and Functions of Microbial Communities." *Nature Ecology and Evolution* 4(3): 366–75.

Dana E. Hunt, Lawrence A. David, Dirk Gevers, Sarah P. Preheim, and Martin F. Polz Eric J. Alm. 2008. "Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton." *Science* 320(May): 1081–85.

Darwin, Charles. 1858. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: Routledge.

Debray, Reena et al. 2021. "Priority Effects in Microbiome Assembly." *Nature Reviews*

Microbiology 0123456789.

Dethlefsen, Les, Margaret McFall-Ngai, and David A. Relman. 2007. "An Ecological and Evolutionary Perspective on Humang-Microbe Mutualism and Disease." *Nature* 449(7164): 811–18.

Detman, Anna et al. 2019. "Cell Factories Converting Lactate and Acetate to Butyrate: Clostridium Butyricum and Microbial Communities from Dark Fermentation Bioreactors." *Microbial Cell Factories* 18(1): 1–12.

Doolittle, W Ford. 2012. "Population Genomics : How Bacterial Species Form and Why They Don ' t Exist." *Current Biology* 22(11): R451–53.

Douglas, Angela E. 2019. "Simple Animal Models for Microbiome Research." *Nature Reviews Microbiology* 17(12): 764–75.

Dunbar, John et al. 1999. "Levels of Bacterial Community Diversity in Four Arid Soils Compared by Cultivation and 16S RRNA Gene Cloning." *Applied and Environmental Microbiology* 65(4): 1662–69.

Duncan, Sylvia H., Petra Louis, and Harry J. Flint. 2004. "Lactate-Utilizing Bacteria, Isolated from Human Feces, That Produce Butyrate as a Major Fermentation Product." *Applied and Environmental Microbiology* 70(10): 5810–17.

Eckburg, Paul B. et al. 2005. "Diversity of the Human Intestinal Microbial Flora." *Science* 308: 1635–38.

Ellegaard, K. M. et al. 2015. "Extensive Intra-Phylotype Diversity in Lactobacilli and Bifidobacteria from the Honeybee Gut." *BMC Genomics* 16(1): 1–22.

———. 2019. "Genomic Changes Underlying Host Specialization in the Bee Gut Symbiont Lactobacillus Firm5." *Molecular Ecology* 28(9): 2224–37.

Ellegaard, K. M., and P. Engel. 2016. "Beyond 16S RRNA Community Profiling: Intra-Species Diversity in the Gut Microbiota." *Frontiers in Microbiology* 7(SEP): 1–16.

Ellegaard, K. M., and Philipp Engel. 2019. "Genomic Diversity Landscape of the Honey Bee Gut Microbiota." *Nature Communications* 10.1(2019): 1–13.

Ellegaard, K. M., Shota Suenami, Ryo Miyazaki, and Philipp Engel. 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.

Engel, Philipp, Kelsey D. Bartlett, and Nancy A. Moran. 2015. "The Bacterium Frischella Perrara Causes Scab Formation in the Gut of Its Honeybee Host." *mBio* 6(3): 1–8.

Engel, Philipp, Vincent G. Martinson, and Nancy A. Moran. 2012. "Functional Diversity within the Simple Gut Microbiota of the Honey Bee." *Proceedings of the National Academy of Sciences of the United States of America* 109(27): 11002–7.

Ericsson, Aaron C., and Craig L. Franklin. 2015. "Manipulating the Gut Microbiota: Methods and Challenges." *ILAR Journal* 56(2): 205–17.

Flint, Harry J., Karen P. Scott, Petra Louis, and Sylvia H. Duncan. 2012. "The Role of the Gut Microbiota in Nutrition and Health." *Nature Reviews Gastroenterology and Hepatology* 9(10): 577–89.

Foster, Kevin R., and Thomas Bell. 2012. "Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species." *Current Biology* 22(19): 1845–50.

Foster, Kevin R., Jonas Schluter, Katharine Z. Coyte, and Seth Rakoff-Nahoum. 2017. "The Evolution of the Host Microbiome as an Ecosystem on a Leash." *Nature* 548(7665): 43–51.

Friedman, Jonathan, Logan M. Higgins, and Jeff Gore. 2017. "Community Structure Follows Simple Assembly Rules in Microbial Microcosms." *Nature Ecology and Evolution* 1(5): 1–7.

Gause, G.F., and A.A Witt. 1935. "Behavior of Mixed Populations and the Problem of Natural Selection." *The American Society of Naturalists* 69(725): 596–609.

Geffre, Amy C. et al. 2020. "Honey Bee Virus Causes Context-Dependent Changes in Host Social Behavior." *Proceedings of the National Academy of Sciences of the United States of America* 117(19): 10406–13.

Gillis, Caroline C. et al. 2018. "Dysbiosis-Associated Change in Host Metabolism Generates Lactate to Support Salmonella Growth." *Cell Host and Microbe* 23(4): 570.

Giri, Samir et al. 2021. "Metabolic Dissimilarity Determines the Establishment of Cross-Feeding Interactions in Bacteria." *Current Biology*.

Goyal, Akshit et al. 2021. "Interactions between Strains Govern the Eco-Evolutionary Dynamics of Microbial Communities." *bioRxiv*: 2021.01.04.425224.

Griffiths, B.S, K Ritz, and R.E Wheatley. 1997. Microbial Communities *Relationship between Functional Diversity and Genetic Diversity*.

Haag, Lea Maxie et al. 2012. "Intestinal Microbiota Shifts towards Elevated Commensal Escherichia Coli Loads Abrogate Colonization Resistance against Campylobacter Jejuni in Mice." *PLoS ONE* 7(5): 1–13.

Hooper, Lora V., and Andrew J. MacPherson. 2010. "Immune Adaptations That Maintain Homeostasis with the Intestinal Microbiota." *Nature Reviews Immunology* 10(3): 159–69.

Hubbell, S. P. 2001. "The Unified Neutral Theory of Biodiversity and Biogeography." *Princeton University Press*.

Hussain, Fatima Aysha et al. 2021. "Rapid Evolutionary Turnover of Mobile Genetic Elements Drives Bacterial Resistance to Phages." *Science* 374(6566): 488–92.

Hutchinson, G E. 1957. "Concluding Remarks." *Cold Spring Harbor Symp* 22: 415–427.

———. 1959. "Homage to Santa Rosalia or Why Are There So Many Kinds of Animals?" *The American Naturalist* 93(870): 145–59. D-494.

Jeraldo, Patricio et al. 2012. "Quantification of the Relative Roles of Niche and Neutral Processes in Structuring Gastrointestinal Microbiomes." *Proceedings of the National Academy of Sciences of the United States of America* 109(25): 9692–98.

Johnson, Renée. 2013. "Honey Bee Colony Collapse Disorder." *Honey Bees and Colony Collapse Disorder: Select Analyses*: 69–92.

Kehe, Jared et al. 2021. "Positive Interactions Are Common among Culturable Bacteria." *Sci. Adv.* 7159(November): 1–11.

Kešnerová, Lucie et al. 2017. "Disentangling Metabolic Functions of Bacteria in the Honey Bee Gut." *PLoS Biology* 15(12): 1–51.

———. 2020. "Gut Microbiota Structure Differs between Honeybees in Winter and Summer." *ISME Journal* 14(3): 801–14.

Koenig, Jeremy E. et al. 2011. "Succession of Microbial Consortia in the Developing Infant Gut Microbiome." *Proceedings of the National Academy of Sciences of the United States of America* 108(SUPPL. 1): 4578–85.

Koskella, Britt, and Michael A. Brockhurst. 2014. "Bacteria-Phage Coevolution as a Driver of Ecological and Evolutionary Processes in Microbial Communities." *FEMS Microbiology Reviews* 38(5): 916–31.

Kwong, Waldan K. et al. 2017. "Dynamic Microbiome Evolution in Social Bees." *Science*

Advances 3(3): 1–17.

Kwong, Waldan K., Amanda L Mancenido, and A Moran. 2014. "Members of the Firm-5 Clade from Honey Bee Guts." *Genome Announcements* 2(6): 5–6.

Kwong, Waldan K., and Nancy A. Moran. 2016. "Gut Microbial Communities of Social Bees." *Nature Reviews Microbiology* 14(6): 374–84.

Lee, Fredrick J. et al. 2015. "Saccharide Breakdown and Fermentation by the Honey Bee Gut Microbiome." *Environmental Microbiology* 17(3): 796–815.

Lee, S. Melanie et al. 2013. "Bacterial Colonization Factors Control Specificity and Stability of the Gut Microbiota." *Nature* 501(7467): 426–29. http://dx.doi.org/10.1038/nature12447.

Ley, Ruth E. et al. 2008. "Evolution of Mammals and Their Gut Microbes." *Science* 320(5883): 1647–51.

Ley, Ruth E., Daniel A. Peterson, and Jeffrey I. Gordon. 2006. "Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine." *Cell* 124(4): 837–48.

Lloyd-Price, Jason, Galeb Abu-Ali, and Curtis Huttenhower. 2016. "The Healthy Human Microbiome." *Genome Medicine* 8(1): 1–11.

Loan, Thi et al. 2015. "How Informative Is the Mouse for Human Gut Microbiota Research ?" *Disease models & mechanisms* 8: 1–16.

Locey, Kenneth J., and Jay T. Lennon. 2016. "Scaling Laws Predict Global Microbial Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 113(21): 5970–75.

Loreau, M. et al. 2001. "Ecology: Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges." *Science* 294(5543): 804–8.

Louca, Stilianos et al. 2017. "High Taxonomic Variability despite Stable Functional Structure across Microbial Communities." *Nature Ecology & Evolution* 1(1): 1–12.

———. 2018. "Function and Functional Redundancy in Microbial Systems." *Nature Ecology and Evolution* 2(6): 936–43.

Lozupone, Catherine A. et al. 2012. "Diversity, Stability and Resilience of the Human Gut Microbiota." *Nature* 489(7415): 220–30.

Madi, Naïma et al. 2020. "Does Diversity Beget Diversity in Microbiomes?" *eLife* 9: 1–26.

Martinson, Vincent G. et al. 2011. "A Simple and Distinctive Microbiota Associated with

Honey Bees and Bumble Bees." *Molecular Ecology* 20(3): 619–28.

Martinson, Vincent G., Jamie Moy, and Nancy A. Moran. 2012. "Establishment of Characteristic Gut Bacteria during Development of the Honeybee Worker." *Applied and Environmental Microbiology* 78(8): 2830–40.

McArthur, J Vaun. 2006. *Microbial Ecology: An Evolutionary Approach*. Elsevier.

Mickalide, Harry, and Seppe Kuehn. 2019. "Higher-Order Interaction between Species Inhibits Bacterial Invasion of a Phototroph-Predator Microbial Community." *Cell Systems* 9(6): 521-533.e10.

Mitri, Sara, and Kevin Richard Foster. 2013. "The Genotypic View of Social Interactions in Microbial Communities." *Annual Review of Genetics* 47: 247–73.

Moeller, Andrew H et al. 2016. "Cospeciation of Gut Microbiota with Hominids." *Science* 353(6297): 380–82.

Momeni, Babak, Li Xie, and Wenying Shou. 2017. "Lotka-Volterra Pairwise Modeling Fails to Capture Diverse Pairwise Microbial Interactions." *eLife* 6: 1–34.

Moran, Mary Ann. 2015. "The Global Ocean Microbiome." *Science* 350(6266).

Mortera, Pablo et al. 2013. "Ca2+ -Citrate Uptake and Metabolism in Lactobacillus Casei Atcc 334." *Applied and Environmental Microbiology* 79(15): 4603–12.

Moya, Andrés, and Manuel Ferrer. 2016. "Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance." *Trends in Microbiology* 24(5): 402–13.

Nadell, Carey D., Knut Drescher, and Kevin R. Foster. 2016. "Spatial Structure, Cooperation and Competition in Biofilms." *Nature Reviews Microbiology* 14(9): 589–600.

Ogier, Jean Claude et al. 2019. "RpoB, a Promising Marker for Analyzing the Diversity of Bacterial Communities by Amplicon Sequencing." *bioRxiv*: 1–16.

Olofsson, Tobias C. et al. 2014. "Lactobacillus Apinorum Sp. Nov., Lactobacillus Mellifer Sp. Nov., Lactobacillus Mellis Sp. Nov., Lactobacillus Melliventris Sp. Nov., Lactobacillus Kimbladii Sp. Nov., Lactobacillus Helsingborgensis Sp. Nov. and Lactobacillus Kullabergensis Sp. Nov., Isol." *International Journal of Systematic and Evolutionary Microbiology* 64(2014): 3109–19.

Philippot, Laurent et al. 2010. "The Ecological Coherence of High Bacterial Taxonomic Ranks." *Nature Reviews Microbiology* 8(7): 523–29.

Piccardi, Philippe, Björn Vessman, and Sara Mitri. 2019. "Toxicity Drives Facilitation between 4 Bacterial Species." *Proceedings of the National Academy of Sciences of the United States of America* 116(32): 15979–84.

Poirier, Simon et al. 2018. "Deciphering Intra-Species Bacterial Diversity of Meat and Seafood Spoilage Microbiota Using GyrB Amplicon Sequencing: A Comparative Analysis with 16S RDNA V3-V4 Amplicon Sequencing." *PLoS ONE* 13(9): 1–26.

Powell, J. Elijah, Vincent G. Martinson, Katherine Urban-Mead, and Nancy A. Moran. 2014. "Routes of Acquisition of the Gut Microbiota of the Honey Bee Apis Mellifera." *Applied and Environmental Microbiology* 80(23): 7378–87.

Pycke, Jean Renaud. 2010. "Some Tests for Uniformity of Circular Distributions Powerful against Multimodal Alternatives." *Canadian Journal of Statistics* 38(1): 80–96.

Qin, Junjie et al. 2012. "A Metagenome-Wide Association Study of Gut Microbiota in Type 2 Diabetes." *Nature* 490(7418): 55–60.

Ratzke, Christoph, Julien Barrere, and Jeff Gore. 2020. "Strength of Species Interactions Determines Biodiversity and Stability in Microbial Communities." *Nature Ecology and Evolution* 4(3): 376–83.

Rawls, John F., Michael A. Mahowald, Ruth E. Ley, and Jeffrey I. Gordon. 2006. "Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-Free Recipients Reveal Host Habitat Selection." *Cell* 127(2): 423–33.

Raymann, Kasie, Zack Shaffer, and Nancy A Moran. 2016. "Antibiotic Exposure Perturbs the Gut Microbiota and Elevates Mortality in Honeybees Antibiotic Exposure Perturbs the Gut Microbiota and Elevates Mortality in Honeybees." : 1–22.

Roesch, Luiz F W et al. 2010. "2007-Pyrosequencing Enumerates and Contrasts Soil Microbial.Pdf." 1(4): 283–90.

La Rosa, Patricio S. et al. 2014. "Patterned Progression of Bacterial Populations in the Premature Infant Gut." *Proceedings of the National Academy of Sciences of the United States of America* 111(48): 17336.

Van Rossum, Thea, Pamela Ferretti, Oleksandr M. Maistrenko, and Peer Bork. 2020. "Diversity within Species: Interpreting Strains in Microbiomes." *Nature Reviews Microbiology* 18(9): 491–506.

Roswall, Josefine et al. 2021. "Developmental Trajectory of the Healthy Human Gut Microbiota during the First 5 Years of Life." *Cell Host and Microbe* 29(5): 765-776.e3.

Schnorr, Stephanie et al. 2017. "Seasonal Cycling in the Gut Microbiome of the Hadza Hunter-Gatherers of Tanzania." *Science* 357(6353): 802–806.

Sepkoski, J. John. 1988. "Alpha, Beta, or Gamma: Where Does All the Diversity Go?" *Paleobiology* 14(3): 221–34.

Shade, Ashley. 2017. "Diversity Is the Question, Not the Answer." *ISME Journal* 11(1): 1–6.

Steele, Margaret I. et al. 2021. "The Gut Microbiota Protects Bees from Invasion by a Bacterial Pathogen." *Microbiology Spectrum* 9(2).

Sun, Yang et al. 2019. "The Gut Microbiota Heterogeneity and Assembly Changes Associated with the IBD." *Scientific Reports* 9(1): 1–8.

Thukral, Ashwani K. 2017. "A Review on Measurement of Alpha Diversity in Biology." *Agricultural Research Journal* 54(1): 1.

Touchon, Marie et al. 2009. "Organised Genome Dynamics in the Escherichia Coli Species Results in Highly Diverse Adaptive Paths." *PLoS Genetics* 5(1).

Upton, Racheal N., Elizabeth M. Bach, and Kirsten S. Hofmockel. 2019. "Spatio-Temporal Microbial Community Dynamics within Soil Aggregates." *Soil Biology and Biochemistry* 132(October 2018): 58–68.

Venter, J. Craig et al. 2004. "Environmental Genome Shotgun Sequencing of the Sargasso Sea." *Science* 304(5667): 66–74.

Ventura, Marco et al. 2011. "The Impact of Bacteriophages on Probiotic Bacteria and Gut Microbiota Diversity." *Genes and Nutrition* 6(3): 205–7.

Vernia, Piero et al. 1988. "Fecal Lactate and Ulcerative Colitis." *Gastroenterology* 95(6): 1564–68.

Wagg, Cameron et al. 2021. "Diversity and Asynchrony in Soil Microbial Communities Stabilizes Ecosystem Functioning." *eLife* 10: 1–19.

Wang, Shui Ping et al. 2020. "Pivotal Roles for PH, Lactate, and Lactate-Utilizing Bacteria in the Stability of a Human Colonic Microbial Ecosystem." *mSystems* 5(5): 1–18.

Wein, Tanita, and Tal Dagan. 2019. "The Effect of Population Bottleneck Size and Selective Regime on Genetic Diversity and Evolvability in Bacteria." *Genome biology and evolution*

11(11): 3283–90.

Whittaker, R H. 1972. “Evolution and Measurement of Species Diversity.” *International Association for Plant Taxonomy (IAPT)* 21(2/3): 213–51.

Wilson, Mark, and Steven E Lindow. 1994. “Coexistence among Epiphytic Bacterial Populations Mediated through Nutritional Resource Partitioning.” 60(12): 4468–77.

Wolff, Richard, William Shoemaker, and Nandita Garud. 2021. “Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome.” *bioRxiv*.

Wright, Erik S., and Kalin H. Vetsigian. 2016. “Inhibitory Interactions Promote Frequent Bistability among Competing Bacteria.” *Nat Commun* 7: 11274.

Zheng, Hao et al. 2016. “Metabolism of Toxic Sugars by Strains of the Bee Gut Symbiont Gilliamella Apicola.” *mBio* 7(6): 1–9.

———. 2017. “Honeybee Gut Microbiota Promotes Host Weight Gain via Bacterial Metabolism and Hormonal Signaling.” *Proceedings of the National Academy of Sciences of the United States of America* 114(18): 4775–80.

———. 2018. “Honey Bees as Models for Gut Microbiota Research.” *Lab Animal* 47(11): 317–25.

Curriculum vitae

CONTACT

13/12/1993 | Italian

Rue du Bugnon 40
1005, Lausanne

+41 78 716 22 17

silviabrochet@gmail.com

SKILLS

Biology

Wet lab:
General laboratory
Molecular Biology
Microbiology
Amplicon sequencing
RNA sequencing
Metabolomics (GC-MS)

Dry lab:
Sequencing-data analysis
Statistical analyses

Personal

Reliable - Persevering
Productive - Team-spirit
Curious - Proactive

LANGUAGES

Italian	–	Mother Tongue
English	–	C1
French	–	C1
Spanish	–	A2

HOBBIES

Silvia BROCHET

Microbiologist

I completed a Bachelor's degree in biology followed by a Master's degree in microbiology. I am now completing a PhD in this field. At the same time, I carried out several activities in the science communication domain. One can describe me as a enthusiastic and curious scientist.

PROFESSIONAL EXPERIENCE

University of Lausanne

Doctoral Researcher — March 2017 - Now

Research on diversity of bacterial communities associated to the honey bee gut
- Design and execute experiments such as *in vivo* colonizations of gnotobiotic bees and *in vitro* cultivation followed by genetic and metabolic screenings.
- Process and analyse data and perform statistical analyses.
- Lead technical training and mentoring for Master students.
- Present outcomes (posters and talks) at national and international conferences:

Swiss society of Microbiology (SSM):
2018 - Lausanne (Switzerland) - Oral presentation
2019 - Zurich (Switerland) - Oral presentation
2021 - Virtual - Oral presentation
International Society for Microbial Ecology (ISME):
2018 - Leipzig (Germany) - Poster presentation
Society for Molecular Biology and Evolution (SMBE):
2019 - Manchester (UK) Poster presentation
Microbial Ecology and Evolution (MEE):
2020 - Virtual - Poster presentation

Scientific Mediator — July 2019 - Now

- Organise and present laboratory activities to bring the broad public closer to science at the University public laboratory: Eprouvette.

Laboratory assistant — March 2017 - Now

- Organize, teach and supervise students during the Microbiology practical work of second year Bachelor students.

Volounteer experience - Scientific mediator — June 2018 - Now

- Organise and manage activities for kids and adults to discover science during the University open doors event ('Mystheres de l'UNIL').

EDUCATION

PhD in Life Sciences — March 2017 - Now
University of Lausanne (Switzerland)

Master in Molecular Life Sciences — September 2015 - January 2017
University of Lausanne (Switzerland)

Bachelor in Cell Molecular Biology — September 2012 - July 2015
University of Turin (Italy)

PUBLICATIONS

Brochet, S., Quinn, A., Mars, R. A., Neuschwander, N., Sauer, U., & Engel, P. (2021). Niche partitioning facilitates coexistence of closely related gut bacteria. eLife, 2021;10:e71105. https://doi.org/10.7554/eLife.68583

Vacheron, J., Pechy-Tarr, M., **Brochet, S.**, Heiman, C. M., Stojiljkovic, M., Maurhofer, M., & Keel, C. (2019). T6SS contributes to gut microbiome invasion and killing of an herbivorous pest insect by plant-beneficial Pseudomonas protegens. The ISME journal, 13(5), 1318-1329. https://doi.org/10.1038/s41396-019-0353-8

Ellegaard, K. M., **Brochet, S.**, Bonilla-Rosso, G., Emery, O., Glover, N., Hadadi, N., ... & 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

Annex

Annex (I)

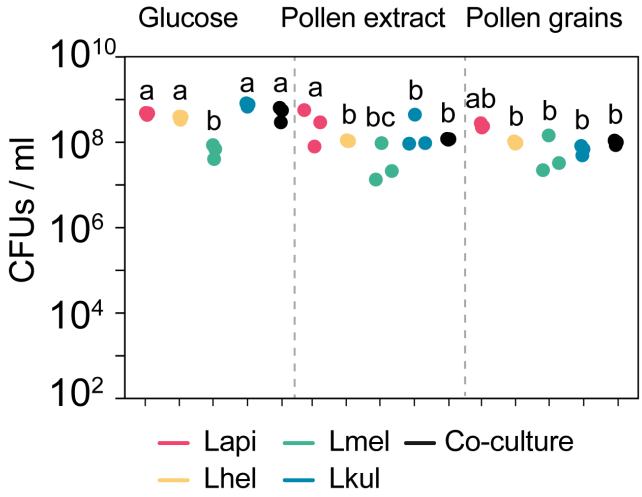


Figure 2—figure supplement 1: Colony-forming units (CFUs) per ml of culture after 24 hr of growth of the four species in mono-cultures (n=3) or in co-culture (n=3) in the presence of 2% (w/v) glucose (G), 10% pollen extract (PE), or 10% pollen grains (PG). Statistical differences: ANOVA with Tuckey post-hoc test (BH correction), represented by different letters.

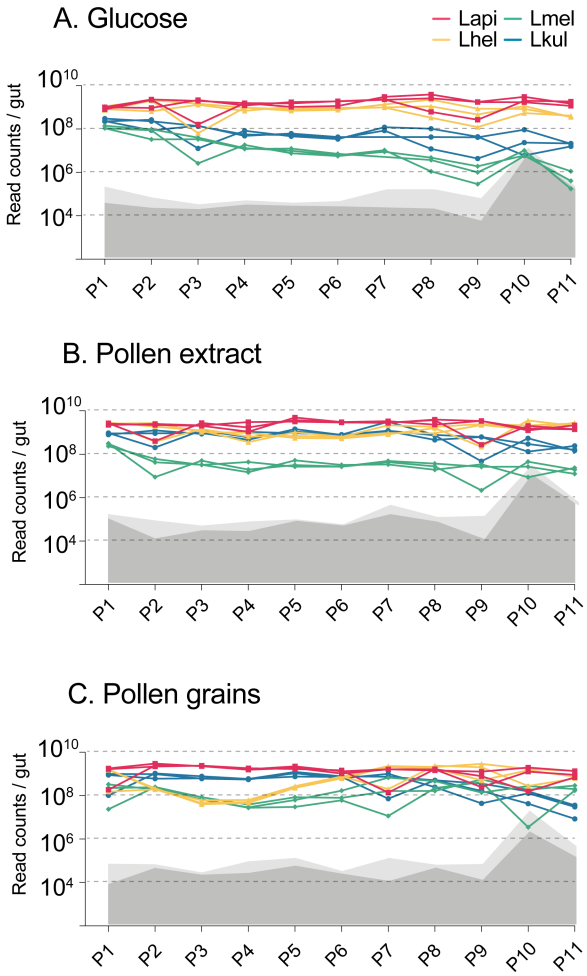


Figure 2—figure supplement 2: Second *in vitro* transfer experiment. Changes in total bacterial abundance of the four species across the 11 serially passaged co-cultures in minimal medium supplemented with either 2% (w/v) glucose (A), 10% pollen extract (B), or 10% pollen grains (C). The absolute abundance of each species was determined by multiplying the total number of CFUs with the proportion of each strain in a given sample as based on amplicon sequencing. Gray areas (light gray = 95% CI) represent the limit of detection as explained in Figure 1 (see Materials and method).

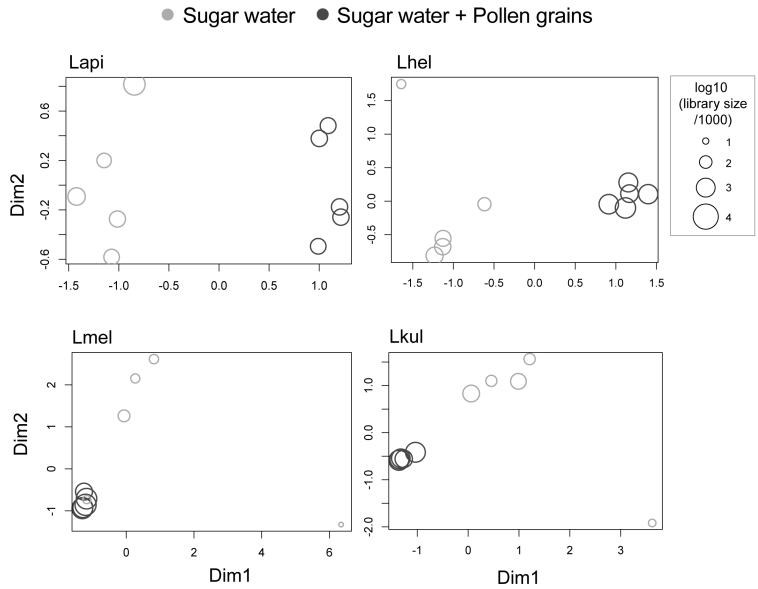


Figure 3—figure supplement 1: MDS plots of *in vivo* RNA-seq samples. Counts per million (cpm) were calculated for each sample (n=5) and visualized using multidimensional scaling (MDS) plots. X- and y-axis axes show first and second MDS dimension, respectively. Shapes size correspond to the samples libraries size. A few samples of the SW treatment did not cluster with the other replicates (two samples for Lmel, and one sample for each Lhel and Lkul), in part because relatively few reads mapped to the reference genomes of these strains.

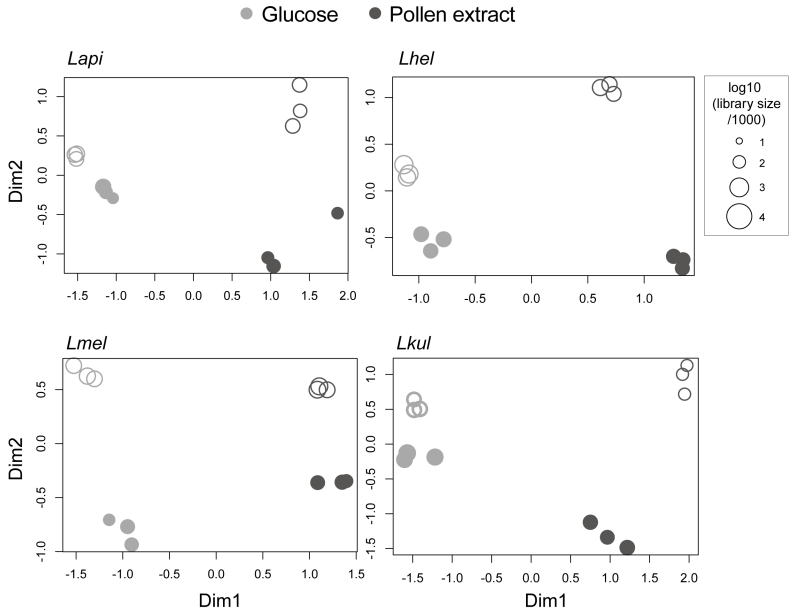


Figure 5—figure supplement 1: MDS plots of *in vitro* RNA-seq samples. Counts per million (cpm) were calculated for each sample (n=3) and visualized using multidimensional scaling (MDS) plots. X- and y-axis axes show first and second MDS dimension, respectively. Filled shapes represent mono-culture samples and empty shapes represent co-culture samples. Shapes size corresponds to the libraries size of that sample.

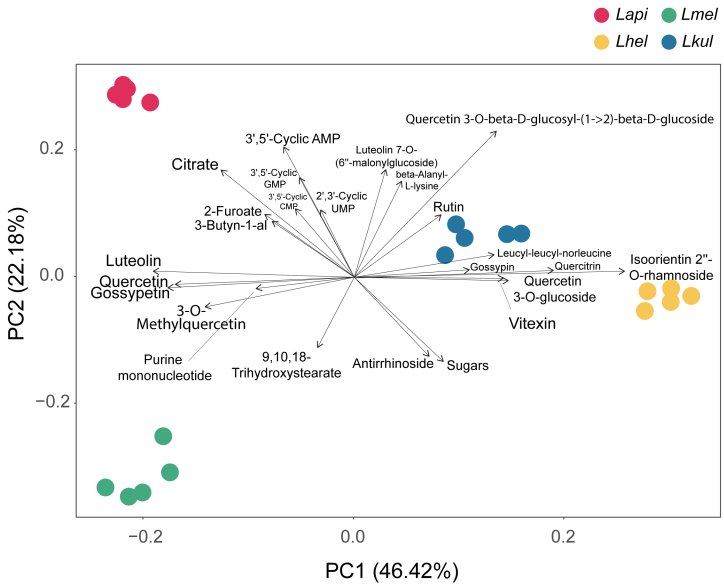


Figure 6—figure supplement 1: PCA *in vitro* metabolomics. Principal component analysis (PCA) of the metabolome profile of each species based on the log2FC values calculated between the two time-points for each ion. The larger the distance between species on the PCA axes, the more they differ in their metabolome profiles. The arrows, that is the environmental vectors, point in the direction of the maximum correlation with the environmental variable, that is the ions. The ions on the tip of the longest arrows are the ones that explain the most of the distribution of the data within the PCA. Only the top 24 ions explaining the data distribution are displayed.

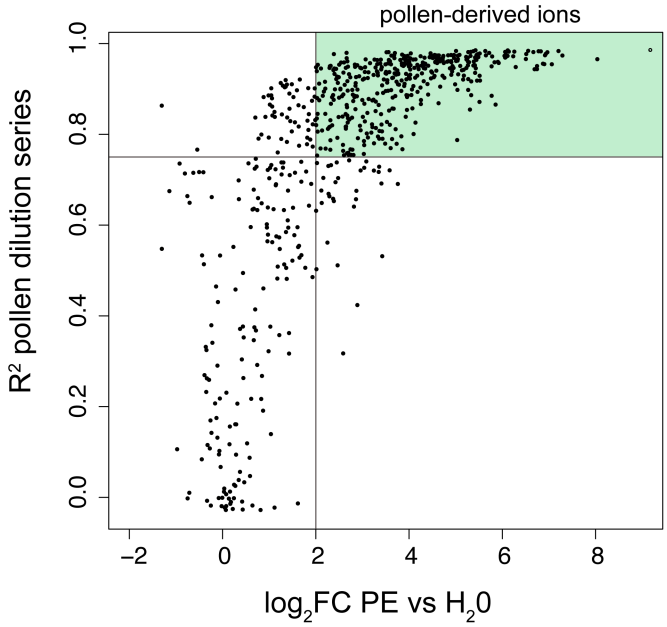


Figure 6—figure supplement 2: Definition of pollen-derived ions. Volcano plot displaying R2 values obtained from the pollen dilution series regression lines and the log2FC calculated between undiluted pollen extract and water. The lines represent the thresholds that we set to define an ion as pollen-derived: log2FC > two and R2 > 0.75. Within the light green area are included the ions that we consider pollen-derived (n = 406).

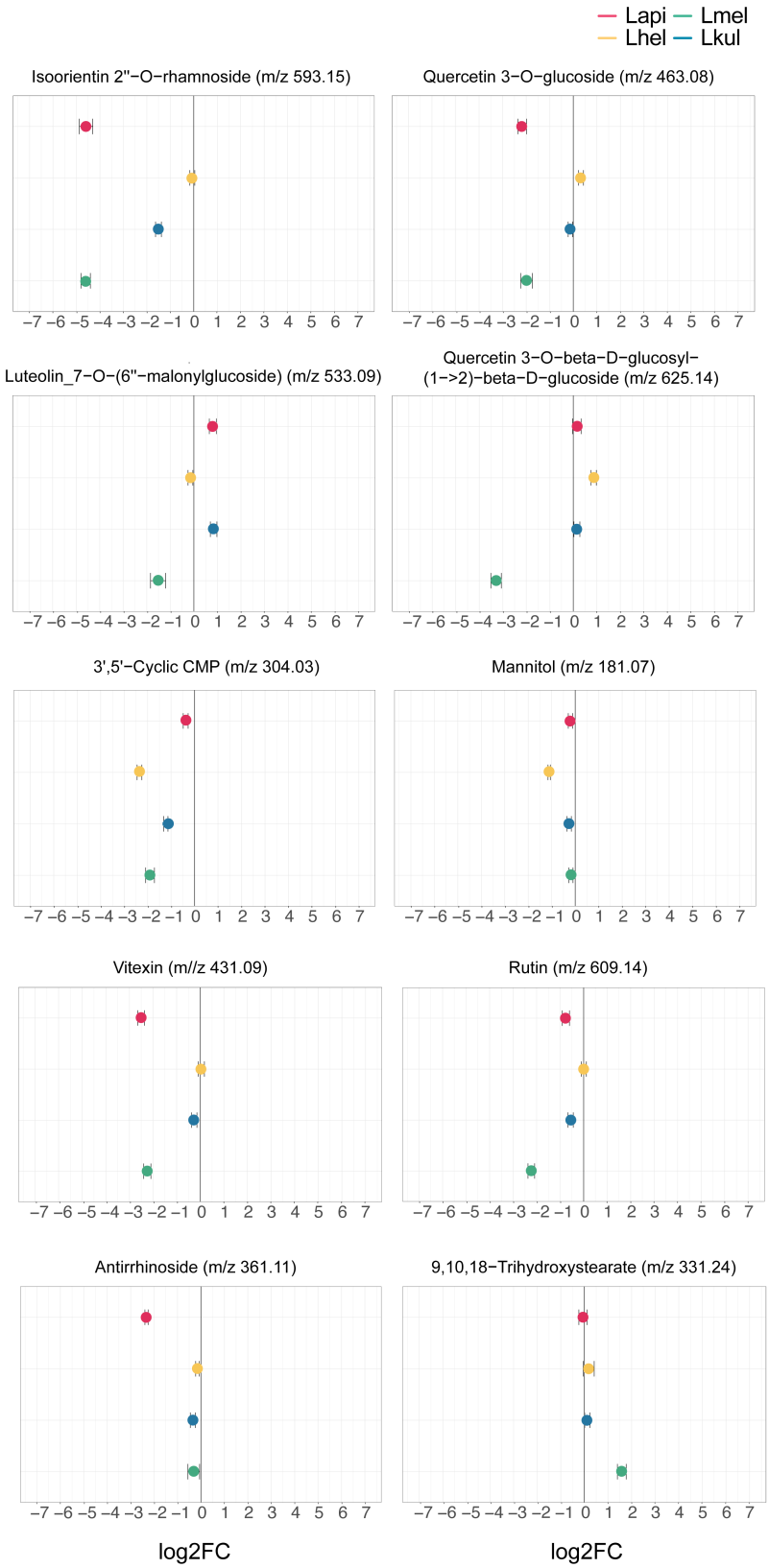


Figure 6—figure supplement 3: Untargeted metabolomics: key metabolites discussed in the main text. In vitro metabolomics of spent medium of the four species grown in cfMRS + PE for 16 hr. The log2FC was obtained comparing the ion intensities at the end and at the beginning of the experiment.

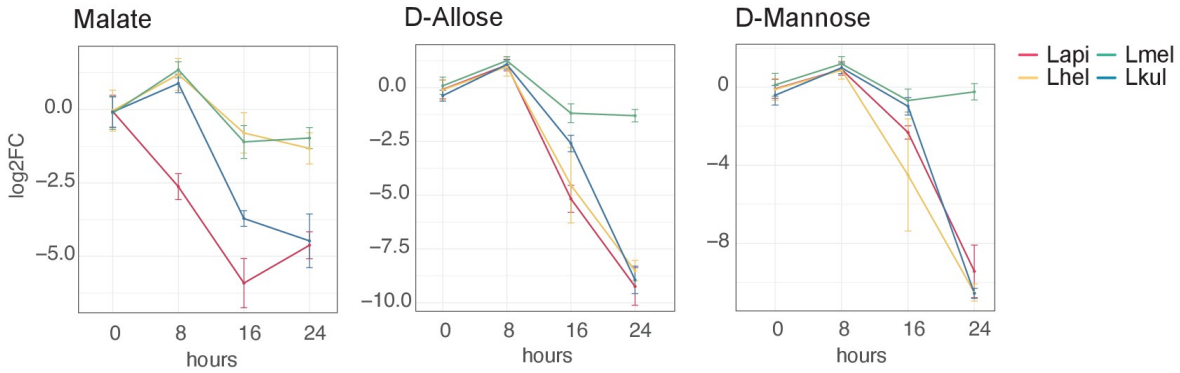


Figure 6—figure supplement 4: GC-MS detection of key metabolites over time. Log2FC relative to T0 is plotted. Time is reported in hours. For m/z values see Supplementary file 10.

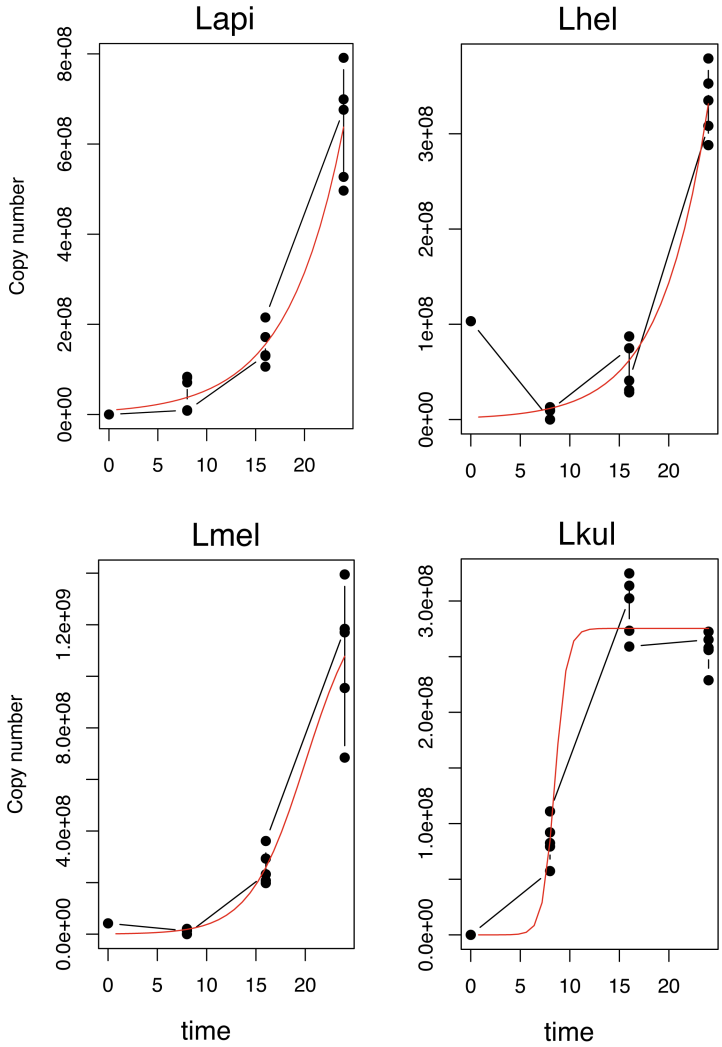


Figure 6—figure supplement 5: Logistic regression growth curve of the four species. Growth-curve data were obtained for the four species at the four time-points included in the second metabolomics experiment (growth in presence of pollen extract) by qPCR (copy number) and fitted to a standard form of the logistic equation. Each point represents one replicate (n=5).

Annex (II)

Supplementary Table 1: List of bacterial strains used in this study.

Species	Strain	Reference
Lactobacillus apis	ESL0263	(Ellegaard et al. 2019)
	ESL0185	(Kešnerová et al. 2017)
	ESL0353	(Olofsson et al. 2014)
Lactobacillus helsingborgensis	ESL0183	(Kešnerová et al. 2017)
	wkb8	(Kwong *et al*, 2014)
	ESL0354	(Olofsson et al. 2014)
Lactobacillus melliventris	ESL0184	(Kešnerová et al. 2017)
	ESL0260	(Ellegaard et al. 2019)
	ESL0350	(Olofsson et al. 2014)
Lactobacillus kullabergensis	ESL0186	(Kešnerová et al. 2017)
	ESL0261	(Ellegaard et al. 2019)
	ESL0351	(Olofsson et al. 2014)

Supplementary Table 2: List of all possible combinations between the 12 selected *Lactobacillus* Firm5 strains.

Combination ID	Strain 1	Strain 2
1	ESL0185	ESL0263
2	ESL0185	ESL0353
3	ESL0185	ESL0183
4	ESL0185	ESL0354
5	ESL0185	wkb8
6	ESL0185	ESL0184
7	ESL0185	ESL0260
8	ESL0185	ESL0350
9	ESL0185	ESL0186
10	ESL0185	ESL0261
11	ESL0185	ESL0351
12	ESL0263	ESL0353
13	ESL0263	ESL0183
14	ESL0263	ESL0354
15	ESL0263	wkb8
16	ESL0263	ESL0184
17	ESL0263	ESL0260
18	ESL0263	ESL0350
19	ESL0263	ESL0186
20	ESL0263	ESL0261

21	ESL0263	ESL0351
22	ESL0353	ESL0183
23	ESL0353	ESL0354
24	ESL0353	wkb8
25	ESL0353	ESL0184
26	ESL0353	ESL0260
27	ESL0353	ESL0350
28	ESL0353	ESL0186
29	ESL0353	ESL0261
30	ESL0353	ESL0351
31	ESL0183	ESL0354
32	ESL0183	wkb8
33	ESL0183	ESL0184
34	ESL0183	ESL0260
35	ESL0183	ESL0350
36	ESL0183	ESL0186
37	ESL0183	ESL0261
38	ESL0183	ESL0351
39	ESL0354	wkb8
40	ESL0354	ESL0184
41	ESL0354	ESL0260
42	ESL0354	ESL0350
43	ESL0354	ESL0186
44	ESL0354	ESL0261
45	ESL0354	ESL0351
46	wkb8	ESL0184
47	wkb8	ESL0260
48	wkb8	ESL0350
49	wkb8	ESL0186
50	wkb8	ESL0261
51	wkb8	ESL0351
52	ESL0184	ESL0260
53	ESL0184	ESL0350
54	ESL0184	ESL0186
55	ESL0184	ESL0261
56	ESL0184	ESL0351
57	ESL0260	ESL0350
58	ESL0260	ESL0186
59	ESL0260	ESL0261
60	ESL0260	ESL0351
61	ESL0350	ESL0186
62	ESL0350	ESL0261

63	ESL0350	ESL0351
64	ESL0186	ESL0261
65	ESL0186	ESL0351
66	ESL0261	ESL0351

Supplementary Table 3: List of barcoded primers used in this study.

ID	Sequence
B1F=	CAAGCAGAAGACGGCATACGAGATTAAGAGGCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT
B2F=	CAAGCAGAAGACGGCATACGAGATGCGAATTCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGTACGTAGACGGCCAGT
B3F=	CAAGCAGAAGACGGCATACGAGATACTGAGCTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGTACGTAGACGGCCAGT
B4F=	CAAGCAGAAGACGGCATACGAGATTTAGGCACGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTACGTAGACGGCCAGT
B5F=	CAAGCAGAAGACGGCATACGAGATCTCCGATTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT
B6F=	CAAGCAGAAGACGGCATACGAGATTTCACAGGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGTACGTAGACGGCCAGT
B7F=	CAAGCAGAAGACGGCATACGAGATCAGAGGTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGTACGTAGACGGCCAGT
B8F=	CAAGCAGAAGACGGCATACGAGATGGCATCATGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTACGTAGACGGCCAGT
B9F=	CAAGCAGAAGACGGCATACGAGATTATGACCGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT
B10F=	CAAGCAGAAGACGGCATACGAGATATACGCTGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGTACGTAGACGGCCAGT
B11F=	CAAGCAGAAGACGGCATACGAGATACGTTCTCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGTACGTAGACGGCCAGT
B12F=	CAAGCAGAAGACGGCATACGAGATAATTGGCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTACGTAGACGGCCAGT
B13F=	CAAGCAGAAGACGGCATACGAGATCATGGCATGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT
B14F=	CAAGCAGAAGACGGCATACGAGATACTCGGTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGTACGTAGACGGCCAGT
B15F=	CAAGCAGAAGACGGCATACGAGATGAGTTCCAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGTACGTAGACGGCCAGT
B16F=	CAAGCAGAAGACGGCATACGAGATGATGCGAAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTACGTAGACGGCCAGT
B17F=	CAAGCAGAAGACGGCATACGAGATCCTTACGTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT
B18F=	CAAGCAGAAGACGGCATACGAGATTCACGTTCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGTACGTAGACGGCCAGT
B19F=	CAAGCAGAAGACGGCATACGAGATCGTGATCAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGTACGTAGACGGCCAGT
B20F=	CAAGCAGAAGACGGCATACGAGATTGCATTCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTACGTAGACGGCCAGT
B21F=	CAAGCAGAAGACGGCATACGAGATTCTGGACAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT
B22F=	CAAGCAGAAGACGGCATACGAGATTGAGCAAGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGTACGTAGACGGCCAGT
B23F=	CAAGCAGAAGACGGCATACGAGATTGACCTGAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCGTACGTAGACGGCCAGT
B24F=	CAAGCAGAAGACGGCATACGAGATCCCTCGGCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACGTACGTAGACGGCCAGT

B1R=	AATGATACGGCGACCACCGAGATCTACACGCCTCTTATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTGACTGCCTATGACG
B2R=	AATGATACGGCGACCACCGAGATCTACACCAGCGTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTGACTGCCTATGACG
B3R=	AATGATACGGCGACCACCGAGATCTACACGAATTCGCTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGACTGCCTATGACG
B4R=	AATGATACGGCGACCACCGAGATCTACACAGCTCAGTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGACTGACTGCCTATGACG
B5R=	AATGATACGGCGACCACCGAGATCTACACGTGCCTAATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTGACTGCCTATGACG
B6R=	AATGATACGGCGACCACCGAGATCTACACACGTAAGGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTGACTGCCTATGACG

B7R=	AATGATACGGCGACCACCGAGATCTACACCCTGTGAATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGACTGCCTATGACG
B8R=	AATGATACGGCGACCACCGAGATCTACACTACCGAGTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGACTGACTGCCTATGACG
B9R=	AATGATACGGCGACCACCGAGATCTACACGGAATGCATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTGACTGCCTATGACG
B10R=	AATGATACGGCGACCACCGAGATCTACACAATCGGAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTGACTGCCTATGACG
B11R=	AATGATACGGCGACCACCGAGATCTACACGAGAACGTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGACTGCCTATGACG
B12R=	AATGATACGGCGACCACCGAGATCTACACGGCCAATTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGACTGACTGCCTATGACG
B13R=	AATGATACGGCGACCACCGAGATCTACACTTCGCATCTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTGACTGCCTATGACG
B14R=	AATGATACGGCGACCACCGAGATCTACACATGCCATGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTGACTGCCTATGACG
B15R=	AATGATACGGCGACCACCGAGATCTACACTGATCACGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGACTGCCTATGACG
B16R=	AATGATACGGCGACCACCGAGATCTACACTACCTCTGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGACTGACTGCCTATGACG
B17R=	AATGATACGGCGACCACCGAGATCTACACTGGAACTCTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTGACTGCCTATGACG
B18R=	AATGATACGGCGACCACCGAGATCTACACGAACGTGATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTGACTGCCTATGACG
B19R=	AATGATACGGCGACCACCGAGATCTACACCGGTCATATCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGACTGCCTATGACG
B20R=	AATGATACGGCGACCACCGAGATCTACACATGATGCCTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGACTGACTGCCTATGACG
B21R=	AATGATACGGCGACCACCGAGATCTACACCCCTACAGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACTGACTGCCTATGACG
B22R=	AATGATACGGCGACCACCGAGATCTACACACATTATTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTGACTGCCTATGACG
B23R=	AATGATACGGCGACCACCGAGATCTACACACCCATGTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACTGACTGCCTATGACG
B24R=	AATGATACGGCGACCACCGAGATCTACACAAGTGTTGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGACTGACTGCCTATGACG

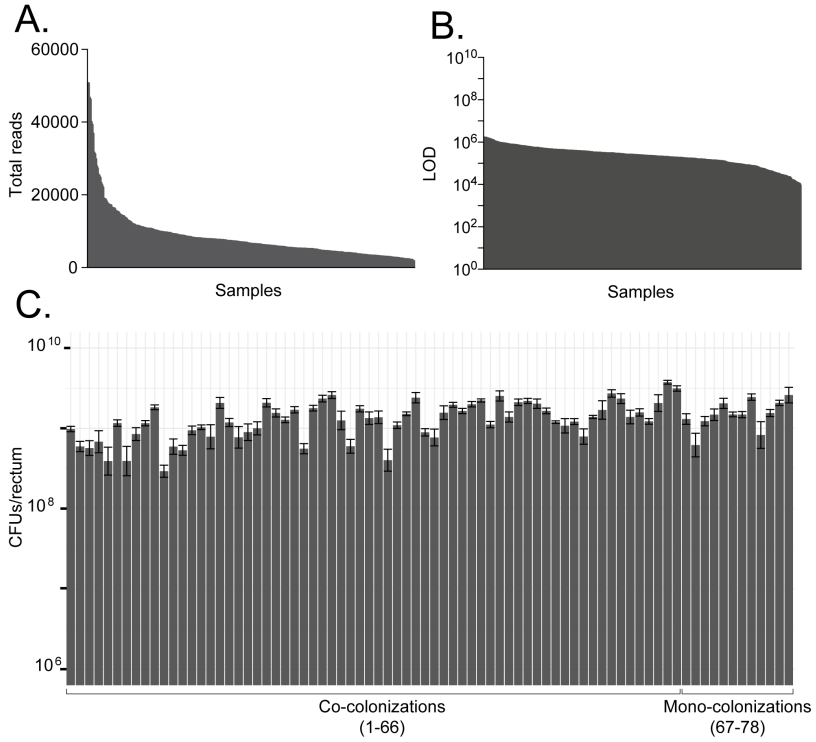


Figure 1 - figure supplement 1: Quality-control strain-strain interactions experiment. A. Total number of reads obtained after sequencing all samples (all replicates included). **B.** Limit of detection (LOD) calculated as the ratio of the CFUs and total number of reads for each sample (all replicates included) **C.** CFUs/rectum calculated for each sample (means of $n_{min} = 3$, standard deviation from the mean is displayed).

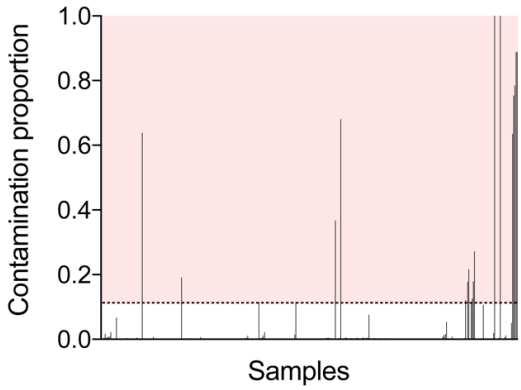


Figure 1 - figure supplement 2: Contamination verification strain-strain interactions experiment. The plot displays the proportion of contaminant-strains (that should not be detected in a given pair) within each sample, i.e. within each pair. The light red area (contamination proportion > 10%) includes all the samples that were excluded for the downstream analysis.

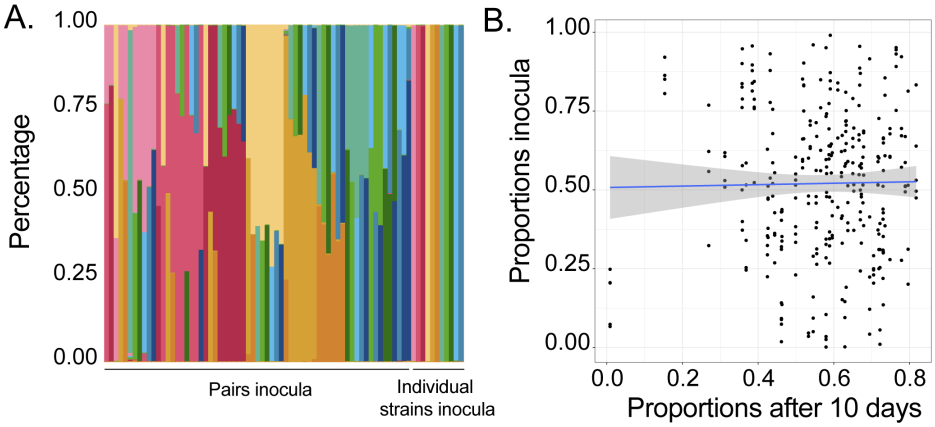


Figure 1 - figure supplement 3: Inocula-samples verification strain-strain interactions experiment. A. Barplot displaying the relative abundances found in the experiment inocula-samples, either in the paired strains condition or in the individual strains condition. Different colors represent different strains. **B.** Correlation between strains proportions within the inocula and strains proportions at the end of the experiment. $R^2 = 0.00022$.

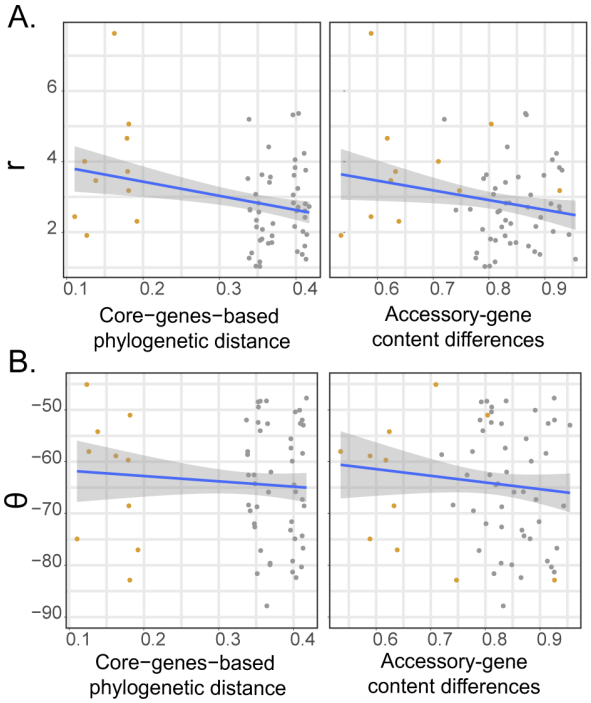


Figure 2 - figure supplement 1: A. Correlation between each vector's r (interaction-strength) and the 12 strains core-genes-based phylogenetic distance (R^2 = 0.076) or accessory gene content differences (R^2 = 0.043). Means of replicates (n=3-5) are displayed. **B.** Correlation between each vector's θ (interaction-type) and the 12 strains core-genes-based phylogenetic distance (R^2 = 0.0062) or accessory gene content differences (R^2 = 0.012). Means of replicates (n=3-5) are displayed. Dark yellow = conspecific strain-pairs; grey = allospecific strain-pairs.

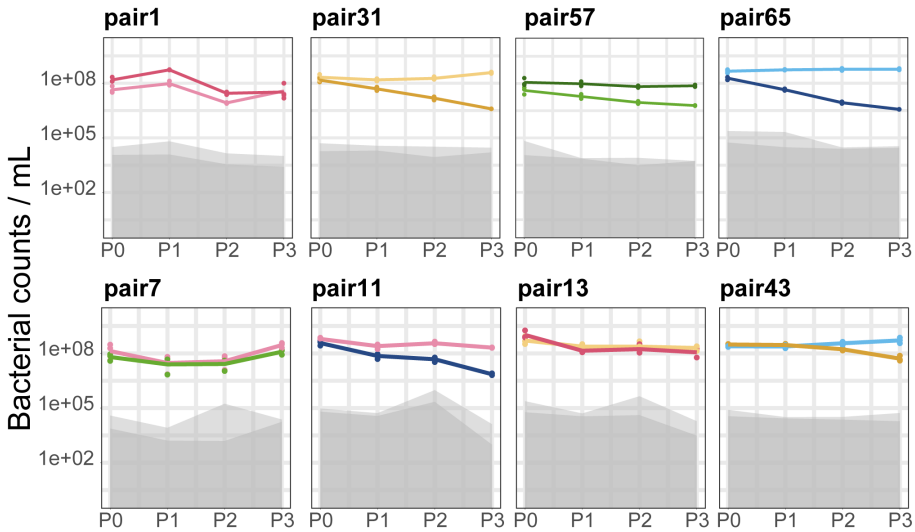
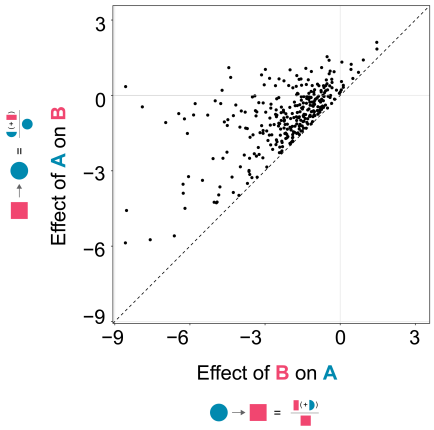


Figure 3 - figure supplement 1: *In vitro* passaging of same-species or different-species Firm5 pairs. Bacterial counts per mL (n=3) across the four *in vitro* passages (P0-P4) of the eight different Firm5 strains pairs (four including strains from the same species and four including strains from different species) were determined by multiplying the total number of CFUs with the relative abundance of each strain in the community. Grey areas represent the limit of detection: the 95% confidence intervals of the limit of detection are shown in light grey.



Supplementary Figure 1: Cartesian plot of strains effects on each other. The two cartesian axes (Effect of A on B and Effect of B on A) represent the results of the calculated effect of one strain on another, which is obtained by computing the ratio between the colonization abundance of the strain in presence of the other strain and the colonization abundance of the strain alone.

Annex (III)

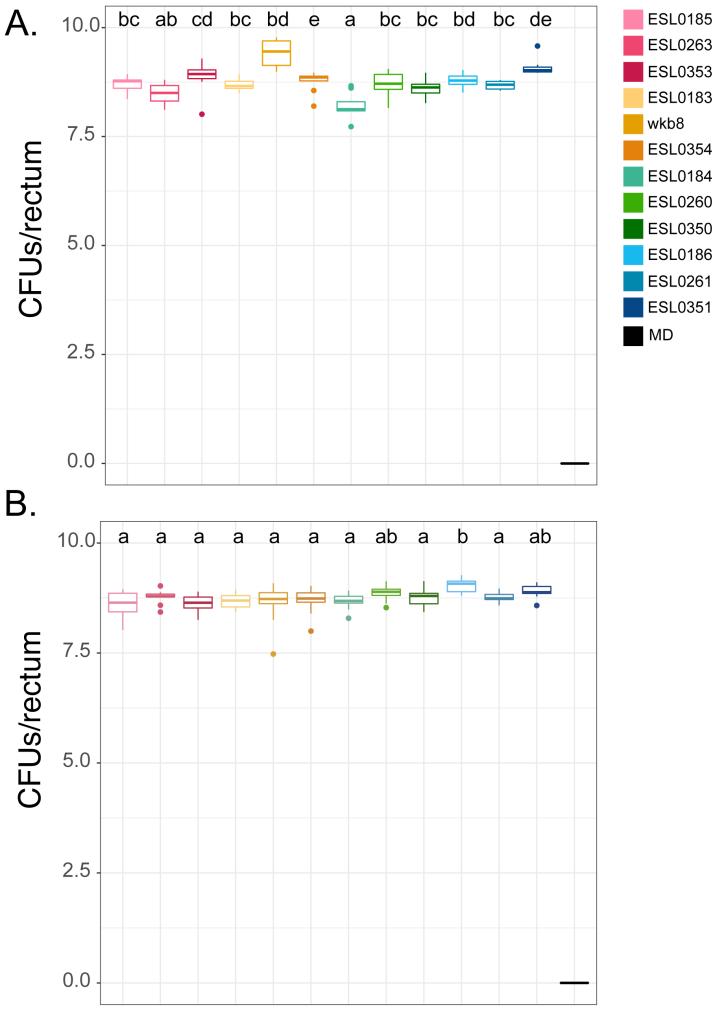


Figure 1 - figure supplement 1: CFUs measured within bees rectums. A. CFUs measured for the experiment 1, where we measured SCFAs production in the rectum. **B.** CFUs measured for the experiment 2, where we measured SCFAs production in the hemolymph. Statistical differences (ANOVA with Tuckey post-hoc test and BH correction) are depicted by different letters.

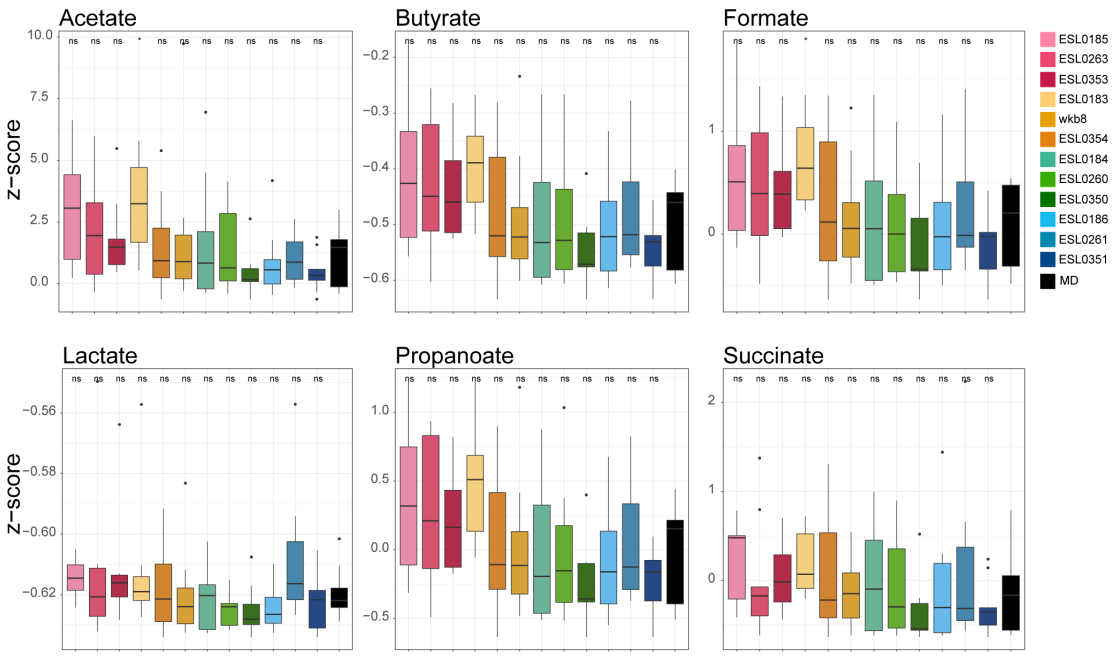


Figure 2 - figure supplement 1: SCFAs production in the bee hemolymph upon colonization with different *Lactobacillus* Firm5 strains. Boxplots displaying the z-score for each measured SCFA for each treatment (see Materials and methods). Significant difference between the Firm5 treatments and the microbiota-depleted (MD) control was measured using Wilcoxon signed rank tests (ns: p > 0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001; ****: p <= 0.0001).

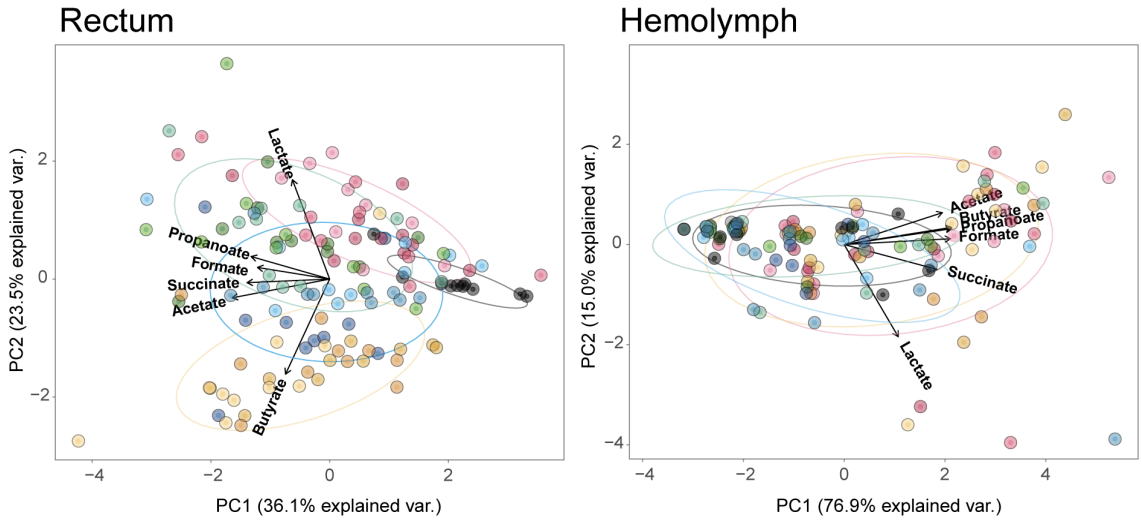


Figure 1 - figure supplement 2: PCA plot highlighting SCFAs that explain data dispersion. Principal component analysis (PCA) of SCFAs detected in the rectum and hemolymph samples. SCFAs responses were normalized by the internal standard and the gut mg or the hemolymph µL and converted to z-scores (see Materials and methods). All replicates are displayed (each point represent a bee, n = 10). The ellipses delimit the different species (different colors) and the MD control (black). The arrows, that represent the environmental vectors, point in the direction of the maximum correlation with the environmental variable, that are the different SCFAs. The SCFAs on the tip of the longest arrows are the ones that explain the most of the distribution of the data within the PCA.

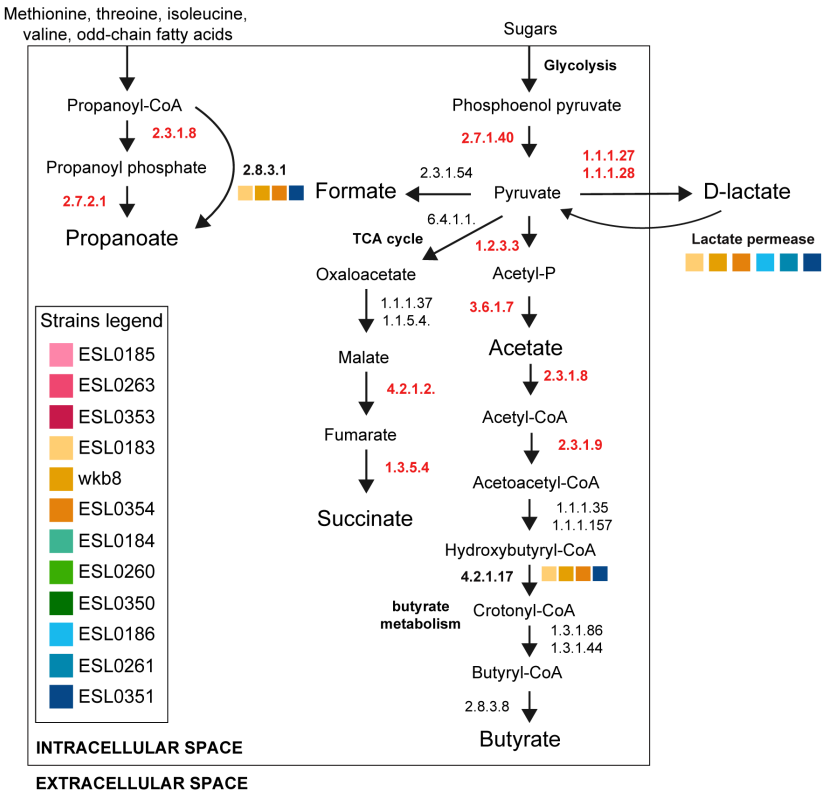


Figure 2 - figure supplement 2: Schematic view of enzymatic pathways leading to the production of the SCFAs analyzed in this study. EC numbers in red indicate enzymes that are present in all the selected Firm5 strains while EC numbers in black indicate enzymes that are not detected in the selected Firm5 strains or, in bold, detected in only a subset of Firm5 strains (see squares colors).

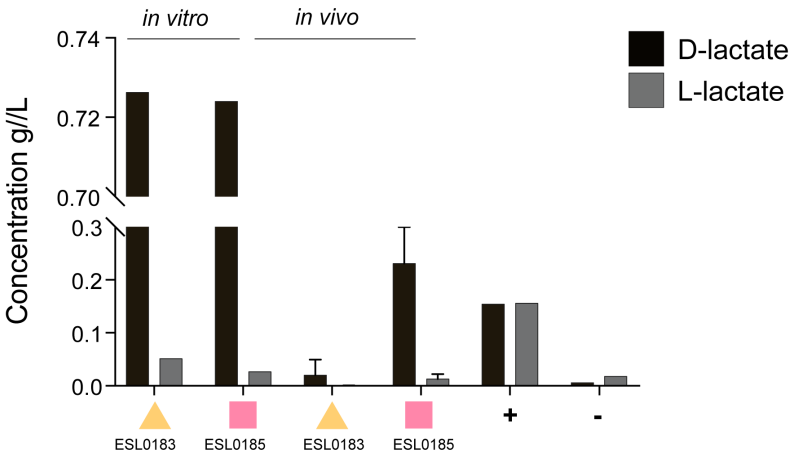


Figure 2 - figure supplement 3: D-/L-lactate measurements. The presence of D/L-lactate was measured both *in vitro* and *in vivo* for strains ESL0183 (Lhel) and ESL0185 (Lapi). The positive control consists of a equal mixture of D-/L-lactate.