Niche partitioning facilitates coexistence of closely related gut bacteria.

Silvia Brochet[a], Andrew Quinn[a], Ruben A.T. Mars[b], Nicolas Neuschwander[a], Uwe Sauer[b], Philipp Engel[a]

[a]Department of Fundamental Microbiology, University of Lausanne, 1015 Lausanne, Switzerland

[b]Institute of Molecular Systems Biology, ETH Zürich, Zürich, Switzerland

Correspondence:

Prof. Philipp Engel

Department of Fundamental Microbiology

University of Lausanne, CH-1015 Lausanne, Switzerland

Tel.: +41 (0)21 692 56 12

e-mail: philipp.engel@unil.ch

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, but 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 symbiotic associations, 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 phyla and families, but contain a high extent of species- and strain-level diversity[1,2]. According to the competitive-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[3]. Therefore, it has remained largely speculative how closely related microbes are 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 co-exist[4]. Such redundancy may be even selected by the host, as it can increase the stability and resilience of the microbiota against environmental disturbance[4,5]. 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[6]. It is also possible that closely related species, and even strains of the same species, have functionally diverged from each other so that they occupy distinct ecological niches[7,8]. Genomic flexibility allows bacteria to adapt their metabolism to utilize different nutrients, provided in the diet or by the host[9,10], or resulting from interactions with other bacteria[11], such as cross-feeding[12] or cooperative glycan breakdown[13].

Few experimental studies of bacterial coexistence in host-associated microbiota exist due to the challenges associated with the high diversity in these ecosystems and the resistance of many bacteria to manipulations[14,15]. Moreover, the community dynamics observed *in vivo* are difficult to reproduce under laboratory conditions, as the host provides a highly specialized (nutritional and spatial) environment that affects the interactions between the community members. There is a need for *in vitro* models that

can reproduce ecological interactions observed *in vivo*, from simple co-culturing models[16] to sophisticated 'organoids-on-a-chip'[17,18] systems.

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[19,20]. 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[21–24]. Microbiota-depleted bees can be generated and experimentally colonized with synthetic communities of different strains. Moreover, most community members can be cultured in the laboratory in pollen, which is the major dietary source of honey bees. This allows to experimentally investigate the coexistence of different gut symbionts in the native host as well as in controlled laboratory cultures using nutritional conditions similar to those found in the gut.

One of the most prevalent phylotypes (>97% 16S rRNA identity) colonizing the honey bee gut microbiota is *Lactobacillus* Firm5[22], which is composed of multiple species of facultative anaerobes that convert sugars into organic acids *via* fermentation. The *Lactobacillus* Firm5 phylotype has so far only been detected in social bees and contains at least four honey bee-specific species: *Lactobacillus apis* (Lapi), *Lactobacillus helsingborgensis* (Lhel), *Lactobacillus melliventris* (Lmel) and *Lactobacillus kullabergensis* (Lkul). Despite having more than 97% sequence identity in the 16S rRNA gene, the inter-species genomic average nucleotide identities (gANI) are below 85%. Comparative genomics revealed high levels of genomic plasticity in carbohydrate metabolism functions across the four species[22,24], and a recent metabolomics study showed that this phylotype utilizes a large amount of pollen-derived glycosylated plant compounds, such as flavonoids, in the bee gut[25]. Metagenomic analyses has shown that the four *Lactobacillus* Firm5 species consistently co-occur in the honey bee gut. However, it is

currently unknown whether the co-existence is facilitated by adaptation to different nutritional niches, and to what extent the host environment or interactions with other community members play important roles.

In this study, we tested under which conditions the four *Lactobacillus* Firm5 species can coexist and investigated underlying molecular mechanisms. To do so, 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 of this study were conducted with four bacterial isolates representing the four species of *Lactobacillus* Firm5 present in the honey bee gut (Lapi, Lhel, Lmel, and Lkul). We first tested if the four species could establish a community 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 Dataset 1**). 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 confirming previous results that pollen increases the total bacterial load in the gut[26] (**Figure 1A**, ANOVA q-value < 0.01). There was no statistically significant difference between the bacterial loads of the mono-colonizations and the co-colonizations in either dietary treatment, with the exception of Lkul, which showed higher loads than the co-colonizations in SW (**Figure 1A**, ANOVA q-value < 0.01). Accordingly, the sum of the bacterial loads of the mono-colonizations exceeds 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 (**Figure 1A**).

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 discriminatory housekeeping gene fragment (see methods) in combination with CFU counting to determine the absolute abundance of each species in the community. We observed striking differences between the two dietary conditions in the ability of the four species to coexist across the passages (**Figure 1B-C, Supplementary Dataset 2**). 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**). In contrast, 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**). On the other hand, for 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. These findings are in line with the idea that coexistence is facilitated by niche partitioning in the gut, but dependent on the nutritional environment.

In vitro **co-culture experiments recapitulate the diet-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), metabolites water-extracted from pollen (PE), or entire pollen grains (PG). All four species were able

to grow when cultured alone under the three nutrient conditions (**Supplementary Figure 1, Supplementary Dataset 1**). Growth yields of Lhel, Lkul, and the co-culture were slightly lower in PE and PG than in G, and Lmel showed lower CFU counts than some of the other species in PE and G (**Supplementary Figure 1,** ANOVA q-value < 0.01). As *in vivo*, the bacterial loads of the co-cultures were not much different than those of the mono-cultures, (**Supplementary Figure 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 h 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 methods). As for the *in vivo* experiment, we observed striking differences in the growth dynamics of the four species, both over time and between the glucose and pollen culture conditions (**Figure 2, Supplementary Dataset 2**). In 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 2C**). None of the species decreased over time, and after 21 transfers all species still yielded between 10^6-10^9 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 changes in community stability were observed for the two pollen conditions throughout the

experiment, whereas in glucose the community reached a stable state after about 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). However, *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 dietary-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 (**Supplementary Figure 2**).

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

Given the impact of pollen on the species co-existence , we tested if genes involved in nutrient acquisition and metabolism were differentially expressed between the different dietary treatments. To this end, we carried out RNA sequencing of the four-species community for 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) (**Supplementary Figure 3**), indicating that all four species exhibited dietary-specific transcriptional responses.

We found a total 687 genes (101 to 217 genes per species) to be differentially expressed ($\log_2 FC \geq |2|$ 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); and 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 Dataset 4)**. The largest fraction of the upregulated COG G genes encoded PTS transporters (**Figure 3D, Supplementary Dataset 3)**, 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 among the downregulated genes in Lapi (Fisher's exact test, $p < 0.01$, **Supplementary Dataset 6**), and genes encoding ABC-type amino acid transporters were present among the downregulated genes in all species (**Supplementary Dataset 3)**.

Clustering genes by homology revealed that 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 in the diet triggers distinct transcriptional changes in the gut. Gene annotation analysis allowed us to identify species-specific metabolic functions among the differentially regulated genes (**Figure 4, Supplementary Dataset 3**). For example, Lhel specifically upregulated three PTS gene clusters for the uptake and metabolism of sugar alcohols and one gene cluster for the 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 present in pollen. Lmel also upregulated a gene cluster for the synthesis and the transport of bacteriocins in presence of pollen. Lkul upregulated a starch utilization gene cluster, which in part was also differentially regulated in Lmel. In addition, this strain also upregulated an oligopeptide transporter gene cluster that was present in some of the other strains but not differentially regulated. As for the fourth species, Lapi also exhibited differentially expressed genes belong to COG G (mainly PTS transporters), although the number of differentially expressed genes was lower. However, it 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 presence of pollen (**Supplementary Dataset 3)**. 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 (see Lkul and Lmel example in **Figure 3G**).

Altogether, these results suggest that the four species utilize different carbohydrate-related resources from pollen, which facilitates niche specialization and hence coexistence.

Transcriptional responses to pollen are concordant *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 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 (**Supplementary Figure 4**).

For each species, whether grown alone or in co-culture, we found between 159-393 genes to be differentially regulated between the PE and the G treatment (**Figure 5B**, $\log_2 FC \geq |2|$ 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 vitro* 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 Dataset 6**). Moreover, 25.3%-36.9% of all genes upregulated *in vivo* in the presence of pollen were also upregulated *in vitro* in the presence of pollen. Specifically, most of the species-specific responses described above (**Figure 4**) were also differentially regulated *in vitro* in the same direction (**Figure 5D**). Interestingly, 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) suggesting that these genes are specifically expressed *in vivo* in presence of pollen (**Supplementary Dataset 7**). 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, as we found in the *in vivo* transcriptome analysis. (**Supplementary Dataset 6**). 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 growth in the presence of pollen, independent of the host environment.

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

We found that a large fraction of the genes upregulated in the mono-cultures were also upregulated in the co-cultures in presence of PE relative to G (58.2-87.8%, **Figure 5E**). In particular, the major gene clusters identified to be regulated in a species-specific manner (see above) showed highly concordant gene expression profiles *in vitro* independent of whether the other *Lactobacillus* species were present. This was confirmed by the direct comparison of mono-culture and co-culture conditions. Compared to the comparison of the nutritional treatments, fewer genes (9-149 genes) were differentially expressed between co-culture and mono-culture treatments ($\log_2$FC $\geq |2|$ and p-value$\leq$0.01), (**Figure 5F**).

We could not find any consistent pattern across the four species relative to COG categories enrichment (**Supplementary Dataset 6**). Moreover, we noted that few genes were differentially expressed in more than one species (6.25-30%), or across both nutrient conditions (1.86-5.33%). Citrate 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 considered to be important *in vivo* are differentially regulated in response to other bacteria, but not always in the same direction across species, or only in the presence of the relevant nutrient condition. We

thus conclude that the main factor driving changes in gene expression in the four species is the presence of pollen, while the presence of other species affects slightly gene expression, without consistent patterns across 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 = 0h) and after bacterial growth (t = 16h) using a Q-TOF-based untargeted metabolomics[27]. 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 methods, **Supplementary Dataset 8**, **Supplementary Figure 6**). The metabolomics data clearly separate the four species corroborating the transcriptomics results suggesting species-specific expression responses for different pollen-derived substrates (**Supplementary Figure 5**). A total of 76 pollen-derived ions showed a significant decrease in abundance over-time (log2FC ≤ -1 and p-value≤0.01, Student's t-Test, BH correction) across the four species (**Figure 6A, Supplementary Dataset 8**). Of those, 24 decreased in abundance in only a subset of the species and another 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 (**Supplementary Figure 5**). 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 different ions annotated as

flavonoids (isoorientin 2''-O-rhamnoside, quercetin-3-O-glucoside, vitexin) (**Figure 6A, Supplementary Figure 7**). 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, Supplementary Figure 7**).

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 RNA-seq results which revealed that Lmel is the only species that upregulated a rhamnosidase gene known to cleave rhamnose residue from rutin[28] (**Figure 4**).

Other ions with species-specific abundance changes included another 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 Supplementary Figure 7**). Lhel was the only species depleting an ion corresponding to sugar alcohols (mannitol, D-sorbitol, or L-glucitol) (**Figure 6A and Supplementary Figure 7**) consistent with the specific upregulation of sugar alcohol PTS transporters in this species (**Figure 4**).

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 h 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 methods). We identified 113 metabolites of which 46 showed a significant change in abundance in at least one strain between timepoint 0h and 24h ($\log_2 FC \geq |2|$ and p-value ≤ 0.01, Student's t-Test, BH correction) (**Supplementary Dataset 8**). All four species showed mixed substrate utilization, i.e. 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, i.e. 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 **Supplementary Figure 8**), which were also found to be consumed by three and all four species, respectively, in 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 if compared to the other species, although having a similar growth profile as the other species (**Figure 6C**, **Supplementary Figure 8** and **9**). 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. This would also be in agreement with the fact that this species upregulated 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[12,14,29–33]. However, few studies have examined the impact of the host on the coexistence of bacterial symbionts of animals[14,32]. Moreover, little is known to what extent closely related species and strains can be maintained under laboratory conditions[8]. We capitalized on the experimental amenability of honey bees and their gut microbiota and used a bottom-up approach to study the coexistence of four naturally co-occurring, closely related *Lactobacillus* species. By selecting one strain per species, we disentangled the effects of the diet and the host on the bacterial interactions by serially passaging the four species-community through gnotobiotic bees, and in culture tubes, under two nutrient conditions (pollen versus simple sugars).

Our results show that the dynamics in the four-species community are governed by negative interactions, as 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[31,34–37], such as the animal gut or culture media. In addition, the studied *Lactobacillus* species are likely to have similar nutritional requirements[38–40], as they harbor relatively small genomes (1.5-2Mb) with a conserved and streamlined core metabolism and overlapping auxotrophies.

The host can play a key role for the coexistence of bacterial symbionts, e.g. by providing a spatially structured environment that facilitates physical separation of competing strains[41–44], or by secreting metabolites that support niche specialization[45,46]. In the case of the four investigated *Lactobacillus* species, such host-related features seem

not to be sufficient to support coexistence, as 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 depends primarily on the pollen diet of the host and not on the host environment itself.

A recent study on the microbiome of the freshwater polyp hydra could not recapitulate the coexistence of the dominant community members *in vitro*, which highlights the difficulty to replicate host conditions and to study relevant interactions of host-associated microbes in culture tubes[32]. In the current study, we aimed to mimic the nutritional conditions in the honey bee gut by growing the bacteria on the natural diet of bees, i.e. pollen, thus disregarding any effect of the host or other community member on the nutritional value of pollen in the gut. This may explain why there were differences in the relative abundance of the four community members *in vitro* and *in vivo*. Yet, qualitatively we were able to recapitulate the *in vivo* observations *in vitro*: pollen nutrients facilitated the four species to stably coexist, while simple sugars led to competitive exclusion.

It is expected that the processes governing bacterial coexistence in host associated microbiomes differ between hosts and symbionts and depend on nutrient availability, dietary transit time, host-microbiota cross-talk, and the spatial structure of the gut. For example, the primary site of colonization of *Lactobacillus* species is the luminal space of the rectum, where partially digested pollen accumulates. In contrast, other members of the bee gut microbiota adhere to the epithelial folds of the ileum, where host interactions

may likely play a more important role for microbial coexistence than in case of the *Lactobacillus* species in the rectum.

While interactions governing microbial communities have been investigated across a range of experimental systems, few studies have tackled underlying molecular and genetic mechanisms of coexistence. Cross-feeding of metabolic by-products was shown to be responsible for the maintenance of diversity in bacterial communities that emerged after passaging leaf and soil samples in single carbon sources[12]. We have little evidence from our study that cross-feeding would play a major role for the coexistence of the four *Lactobacilli*. First, unlike in the study mentioned before, feeding a single carbon source led to the extinction of all species except for one. Second, in the presence of pollen, our metabolomics analysis identified very few metabolites that could potentially be cross-fed, i.e. were produced by one species and utilized by another. Third, we detected no transcriptional changes between mono-cultures and co-cultures that would suggest cross-feeding activities.

Rather our combined transcriptomics and metabolomics data suggest that the coexistence is facilitated by specialization towards distinct pollen-derived nutrients. We found that all four species upregulated carbohydrate transport and metabolism functions dedicated to the utilization of different carbon sources when grown in pollen. Similar transcriptional changes were observed for another honey bee gut symbiont and for human gut bacteria when the host diet was supplemented with specific plant-derived glycans[47,48]. The role of the carbohydrate metabolism-related gene content for the coexistence of the four species is corroborated by the fact that concordant transcriptional changes were observed *in vivo* and *in vitro*. Finally, using metabolomics we identified a number of pollen-derived nutrients, e.g. flavonoids or other glycosides, that were utilized in a species-specific manner. Lmel specialized on the utilization of flavonoids at the

expense of simple sugars, which may explain why this species was going rapidly extinct in presence of only simple sugars during the transfer experiments.

It remains to be determined if pollen-derived flavonoids offer a major nutrient source in the honey bee gut. However, 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. Notably, with the exception of a few compounds, such as glucitol, citrate, malate, or lactose, we saw relatively little differences in the utilization of simple carbon metabolites among the four species in our time-resolved GC/MS analysis. This seems to be congruent with the observation that the four species engage in negative interactions with each other. Yet, coexistence in pollen is still possible. Theoretical work has established that resource preference for at least one substrate is sufficient to explain coexistence[49]. Moreover, it is possible that the four species utilize similar sugars, but that they extract them from different plant-derived glycans present in pollen, such as starch, hemicellulose, flavonoids, or other glycosylated secondary plant metabolites. Whether the host benefits from harboring all four species by for example facilitating the breakdown of different recalcitrant dietary compounds or by providing different metabolic end-products needs to be determined.

It has previously been speculated that the large diversity of carbohydrate transport and metabolism functions in the accessory gene pool of *Lactobacillus* is an adaptation to the pollen-based diet of the host and a consequence of the nutrient competition with closely related species[22,38]. Our findings support this hypothesis and provide first experimental evidence for the link between the coexistence of diverse *Lactobacilli* species in the honey bee gut, the large diversity of carbohydrate metabolism related functions in their genomes, and the pollen diet of the host.

Under natural conditions and in the presence of the full microbiota, the four different species of *Lactobacillus* have been found to co-occur in every individual bee, while strains of the same species tend to segregate into different individuals. Since we have only tested a single strain of each species, and given the extensive genomic diversity within each species[22], more work is needed to determine if the identified patterns of coexistence reflect stable ecological niches occupied by the four species. It is possible that even strains of the same species could coexist in such experiments. In a recent study on pitcher plant microbiomes it was shown that strains that differ by only a few base pairs can have different ecological trajectories in communities and coexist over extended period of time[8]. Such experiments would help to further 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

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[50]) supplemented with 4 % w/v glucose, 4 % w/v fructose and 1 w/v % L-cysteine-HCl 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} = 1 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, ten 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 pool 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**

Each of the four strains was cultured in liquid medium overnight for about 16 h as described above. The cultures were re-inoculated at an OD_{600} = 0.3 in fresh medium and let grow for another 4 h 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 450 µL of cfMRS medium supplemented with either 2 % glucose w/v, 10 % pollen extract, or 10 % pollen grains 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 h at 34˚C under anaerobic conditions without shaking (300 rpm). After 24 h 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 respectively 10 and 20 times for the two independent experiments carried out. After each transfer, cultures were

washed once with 1xPBS and stored at -20˚C for amplicon sequencing analysis. CFUs were counted after 24 h 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 the housekeeping gene DNA formamidopyrimidine-glycosylase which allows to discriminate the four strains from each other[51].

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 Table 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 sec, 64˚C – 30 sec, 72˚C – 30 sec), 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 – 1000U/ ml – NEB, 10 % Exonuclease I – 20000 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 8 times. The barcoded primers are listed in Supplementary Table 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)[52], 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 Table 1**, **Supplementary Dataset 2**).

For the *in vitro* transfer the stability of the four-species community over time was calculated using the codyn R package[53].

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 seconds). 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 hot phenol RNA extraction protocol developed by Jörg Vogel lab was followed[54]. 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 h 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. [55]. 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[56] (Lapi, Lhel, Lmel and Lkul) using Bowtie (bowtie2-2.3.2). Gene annotations for the four genomes were retrieved from IMG/mer[57]. 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[58]. Counts per million were calculated and only genes with at least 1

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[59].

Untargeted metabolomics

Metabolites were extracted from liquid cultures supplemented with 10% w/v pollen extract at the inoculation time and after 16 h 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 ddH2O. 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 times 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 h 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 h and centrifuged (20'000 g, 5 min). 200 µL of the supernatant was transferred to a new tube and diluted 10x 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[25]. 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 dataset 8.** 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 assessed by assessing the correlation between ion intensities of the respective technical replicate. Then, mean ion intensities of technical replicates were calculated. Time point comparisons (T = 0 h vs T = 16 h) were performed using t-tests with Benjamini-Hochberg (BH) correction for multiple testing. $\log_2$FC 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 for each ion were plotted and the R^2 of the obtained linear fit was extracted. In addition, we calculated the $\log_2$FC difference between undiluted pollen and water. The R^2 values were then plotted against the $\log_2$FC values, and stringent thresholds ($R_2 > 0.75$ and $\log_2$FC > 2) were chosen to discriminate ions that are likely pollen-derived (**Supplementary Figure 7**). All ions were included for downstream analysis (e.g. PCA) and then they were discriminated between pollen-derived and non-pollen-derived.

GC-MS-based metabolomics

Soluble metabolites were extracted from liquid cultures supplemented with 10% pollen extract (w/v) at the inoculation time and after 8, 16 and 24 h 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 uL 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 sec. The tubes were incubated at -20°C for 90 min and vortexed 2x for 30 sec during the incubation. Tubes were centrifuged for 5 min at 10'000g 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 8'000 rpm at 4°C and the liquid phase was transferred to the previous extracted aqueous phase. 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 h at 33°C followed by derivatization with N-Methyl-N-(trymethylsolyl)trifluoroacetamide (MSTFA, Sigma Aldrich) for 2 h at 35°C.

Aliquots (1 µL) were injected on an Agilent 8890/5977B GC/MDS instrument GC-MS instrument. The samples were injected in split mode (20:1) with an inlet temperature of 250 °C. The column was held initially at 125 °C for 2 mins, 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). Peaks from the total ion chromatogram (TIC) were identified by matching retention times and spectra to an in-house library of analytical standards, as well as to the NIST v17 Mass Spectral Library (NIST). 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 and code

The complete custom code for all the analyses is available on GitHub: (https://github.com/silviabrochet). The amplicon sequencing data and the RNA sequencing data are available under the NCBI Bioproject PRJNA700984 and the GEO record GSE166724.

Acknowledgments

We thank Kirsten Ellegaard for the critical reading of the manuscript and discussions on the experimental design. We also thank Sara Mitri for reading the manuscript and giving us critical suggestions to improve the manuscript. We also thank Simon Maréchal for his support with the transfer experiments the lab. This project was funded by the ERC-StG "MicroBeeOme" (grant 714804), the Swiss National Science Foundation (grant 31003A_160345 and 31003A_179487) and the Human Frontier Science Program Young Investigator (grant RGY0077/2016).

Competing interests

The authors declare that they have no competing interests.

References

1. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science (80-.).* **320**, 1647–1651 (2008).
2. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on humang-microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
3. Elton, C. Competition and the Structure of Ecological Communities. *J. Anim. Ecol.* **15**, 54–68 (2016).
4. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848 (2006).
5. Foster, K. R., Schluter, J., Coyte, K. Z. & Rakoff-Nahoum, S. The evolution of the host microbiome as an ecosystem on a leash. *Nature* **548**, 43–51 (2017).
6. Koskella, B. & Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–931 (2014).
7. Chesson, P. Mechanisms of Maintainance of Species Diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–66 (2000).
8. Leonora S. Bittleston, Matti Gralka, Gabriel E. Leventhal, Itzhak Mizrahi, O. X. C. Context-dependent dynamics lead to the assembly of functionally distinct pitcher-plant microbiomes. *Nat Commun* **3**, 1–31 (2019).
9. Berasategui, A. *et al.* Gut microbiota of the pine weevil degrades conifer diterpenes and increases insect fitness. *Mol. Ecol.* **26**, 4099–4110 (2017).
10. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
11. Madi, N., Vos, M., Murall, C. L., Legendre, P. & Shapiro, B. J. Does diversity beget diversity in microbiomes? *bioRxiv* 1–26 (2020). doi:10.1101/612739
12. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science (80-.).* **361**, 469–474 (2018).
13. Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of cooperation within the gut microbiota. *Nature* **533**, 255–259 (2016).
14. Lopez, A. O., Vega, N. M. & Gore, J. Interspecies bacterial competition determines

community assembly in the C. elegans intestine. *ISME J.* (2021). doi:10.1101/535633

15. Venturelli, O. S. *et al.* Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.* **14**, 1–19 (2018).
16. Li, L. *et al.* An in vitro model maintaining taxon-specific functional activities of the gut microbiome. *Nat. Commun.* **10**, (2019).
17. Jalili-Firoozinezhad, S. *et al.* A complex human gut microbiome cultured in an anaerobic intestine-on-a-chip. *Nat. Biomed. Eng.* **3**, 520–531 (2019).
18. Nikolaev, M. *et al.* Homeostatic mini-intestines through scaffold-guided organoid morphogenesis. *Nature* **585**, 574–578 (2020).
19. Martinson, V. G. *et al.* A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol. Ecol.* **20**, 619–628 (2011).
20. Kwong, W. K. & Moran, N. A. Gut microbial communities of social bees. *Nat. Rev. Microbiol.* **14**, 374–384 (2016).
21. Ellegaard, K. M. & Engel, P. Beyond 16S rRNA community profiling: Intra-species diversity in the gut microbiota. *Front. Microbiol.* **7**, 1–16 (2016).
22. Ellegaard, K. M., Engel, P. Genomic diversity landscape of the honey bee gut microbiota. *Nat. Commun.* **10.1**, 1–13 (2019).
23. Engel, P., Martinson, V. G. & Moran, N. a. Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11002–11007 (2012).
24. Ellegaard, K. M. *et al.* Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* **16**, 1–22 (2015).
25. Kešnerová, L., Mars, R. A. T., Ellegaard, K. M., Sauer, U. & Engel, P. Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biol.* 1–28 (2017).
26. Kešnerová, L. *et al.* Gut microbiota structure differs between honeybees in winter and summer. *ISME J.* **14**, 801–814 (2020).
27. Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal. Chem.* **83**, 7074–7080 (2011).
28. Beekwilder, J. *et al.* Characterization of Rhamnosidases from Lactobacillus plantarum and Lactobacillus acidophilus. *Appl. Environ. Microbiol.* **75**, 3447–3454 (2009).

29. Wright, Erik S., and K. H. V. Inhibitory interactions promote frequent bistability among competing bacteria. *Nat Commun* **7**, 11274 (2016).
30. Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* **1**, 1–7 (2017).
31. Piccardi, P., Vessman, B. & Mitri, S. Toxicity drives facilitation between 4 bacterial species. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15979–15984 (2019).
32. Deines, P., Hammerschmidt, K. & Bosch, T. C. G. Microbial species coexistence depends on the host environment. *MBio* **11**, 1–13 (2020).
33. Logan M. Higgins, J. F. H. S. & J. G. Co-occurring soil bacteria exhibit a robust competitive hierarchy and lack of non-transitive interactions. *bioRxiv* 1–16 (2017).
34. Foster, K. R. & Bell, T. Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* **22**, 1845–1850 (2012).
35. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 1–14 (2014).
36. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks, competition, and stability. *Science (80-.).* **350**, 663–666 (2015).
37. Ghoul, M. & Mitri, S. The Ecology and Evolution of Microbial Competition. *Trends Microbiol.* **24**, 833–845 (2016).
38. Ellegaard, K., Brochet, S., Emery, O., Glover, N. & Hadadi, N. Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. *Mol. Ecol.* **28**, 2224–2237 (2019).
39. Ellegaard, K. M. *et al.* Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* **16**, 1–22 (2015).
40. Kwong, W. K. & Moran, N. A. Gut microbial communities of social bees. *Nat. Rev. Microbiol.* **14**, 374–384 (2016).
41. Gude, S. *et al.* Bacterial coexistence driven by motility and spatial competition. *Nature* **578**, 588–592 (2020).
42. Kim, H. J., Boedicker, J. Q., Choi, J. W. & Ismagilov, R. F. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *PNAS* **105**, 18188–18193 (2008).
43. Mitri, S., Clarke, E. & Foster, K. R. Resource limitation drives spatial organization in microbial groups. *ISME J.* **10**, 1471–1482 (2016).

44. Hallatschek, O., Hersen, P., Ramanathan, S. & Nelson, D. R. Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19926–19930 (2007).
45. Schluter, J. & Foster, K. R. The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection. *PLoS Biol.* **10**, (2012).
46. McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L. & Foster, K. R. Host Selection of Microbiota via Differential Adhesion. *Cell Host Microbe* **19**, 550–559 (2016).
47. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science (80-.).* **307**, 1955–1959 (2005).
48. Zheng, H. *et al.* Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25909–25916 (2019).
49. Meszéna, G., Gyllenberg, M., Pásztor, L. & Metz, J. A. J. Competitive exclusion and limiting similarity: A unified theory. *Theor. Popul. Biol.* **69**, 68–87 (2006).
50. Donnell, M. M. O., Forde, B. M., Neville, B., Ross, P. R. & Toole, P. W. O. Carbohydrate catabolic flexibility in the mammalian intestinal commensal Lactobacillus ruminis revealed by fermentation studies aligned to genome annotations. *Microb. Cell Fact.* **10**, 1–11 (2011).
51. Ellegaard, K. M. *et al.* Genomic changes underlying host specialization in the bee gut symbiont Lactobacillus Firm5. *Mol. Ecol.* **28**, 2224–2237 (2019).
52. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
53. Hallett L, Avolio M, Carroll I, Jones S, MacDonald A, Flynn D, Slaughter P, Ripplinger J, Collins S, Gries C, J. M. (2020). codyn: Community Dynamics Metrics. doi: 10.5063/F1N877Z6, R package version 2.0.5, https://github.com/NCEAS/codynHallett L, Avoli, https://github.com/NCEAS/codyn.
54. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250–255 (2010).
55. Avraham, R. *et al.* A highly multiplexed and sensitive RNA-seq protocol for simultaneous analysis of host and pathogen transcriptomes. *Nat. Protoc.* **11**, 1477–1491 (2016).
56. Ellegaard, K. M. & Engel, P. New Reference Genome Sequences for 17 Bacterial

Strains of. *Microbiol. Resour. Announc.* **7**, 4–6 (2018).

57. Chen, I. M. A. *et al.* The IMG/M data management and analysis system v.6.0: New tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).
58. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
59. Nowlan H. Freese, David C. Norris, and A. E. L. Integrated Genome Browser: Visual analytics platform for genomics. *Bioinformatics* (2016).

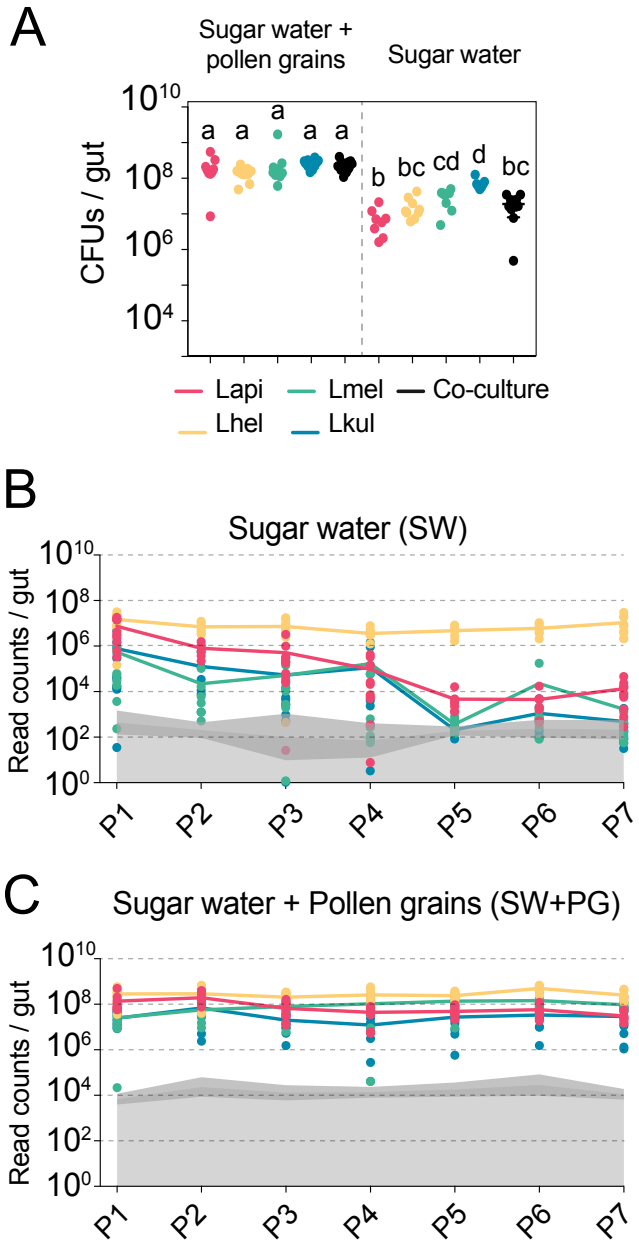

Figure 1. The presence of dietary pollen facilitates the stable coexistence of the four *Lactobacillus* species in the honey bee gut. (A) Bacterial abundance (CFUs) in the gut of gnotobiotic bees (n= 7-10) colonized with the four species separately or together under two different dietary conditions. 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 species across the seven serial passages (P1-P7) of the four-species community 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 methods). Therefore, the average limit of detection and the 95% confidence intervals are shown.

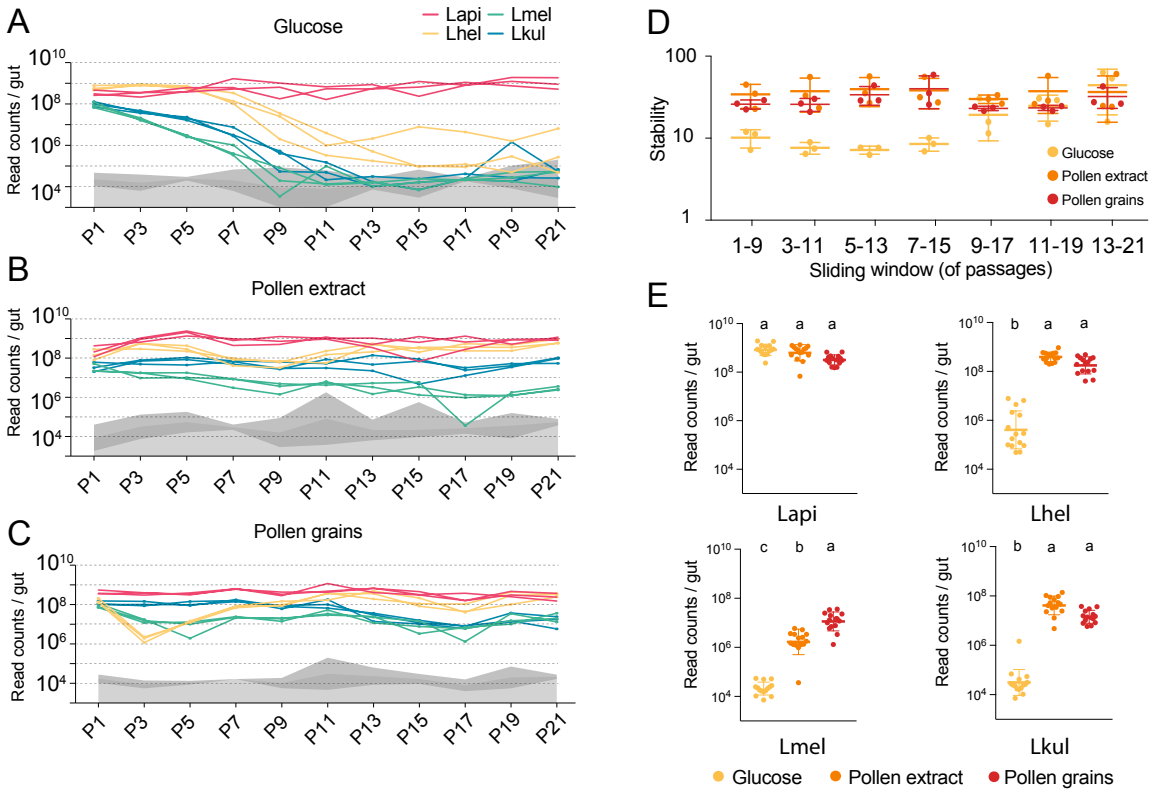

Figure 2. The stable coexistence of the four *Lactobacillus* species can be recapitulated *in vitro* in the presence of pollen. (A-C) Changes in total abundance of the four species when serial passaged in co-culture for 21 times in minimal medium supplemented with **(A)** 2% (w/v) glucose, **(B)** 10% pollen extract**,** and **(C)** 10% 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 methods). Grey areas indicate the limit of detection as explained in the 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.

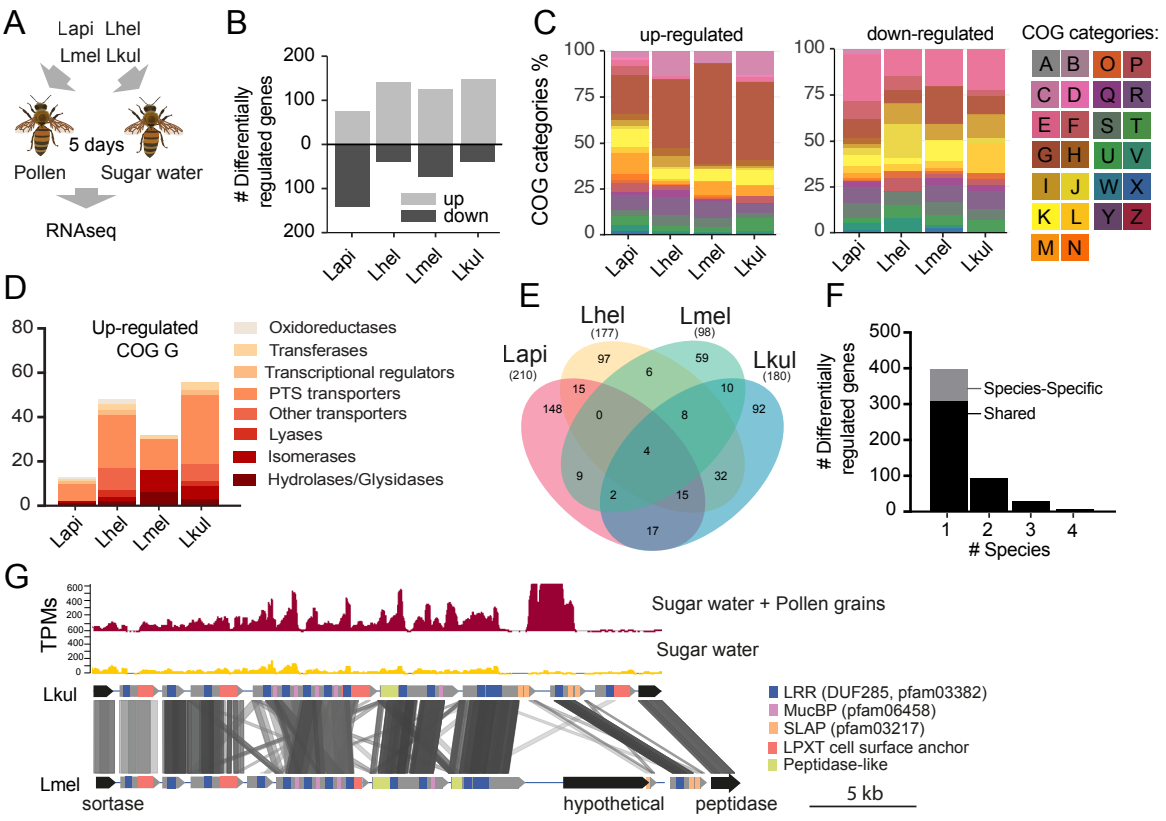

Figure 3: Transcriptome analysis of the four *Lactobacillus* species during co-colonization of gnotobiotic bees. **(A)** Experimental design scheme. **(B)** Number of differentially regulated genes ($\log_2 FC \geq |2|$ and p-value ≤ 0.01) in each species during co-colonization of gnotobiotic bees fed 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-regulated or down-regulated by the four species in SW+PG if compared to SW. **(D)** Functional sub-categories of COG 'G' genes upregulated in SW+PG if compared to SW. **(E)** Venn diagram showing overlap of gene families (based on gene homology) differentially regulated in the four species. **(F)** 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. **(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. Grey 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.

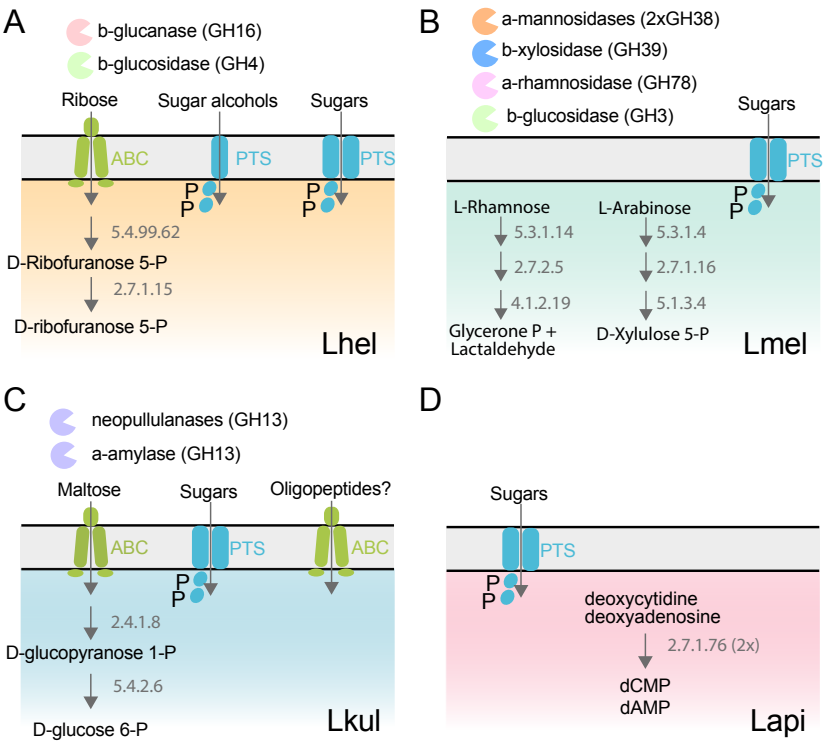

Figure 4. The four *Lactobacillus* species upregulate divergent carbohydrate metabolism functions during gut colonization of gnotobiotic bees. (A) Lhel, **(B)** Lmel, **(C)** Lkul, **(D)** Lapi. Only enzymes and transporters that are upregulated in a species-specific manner in the pollen diet treatment versus the sugar water diet 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 Dataset 4**). 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.

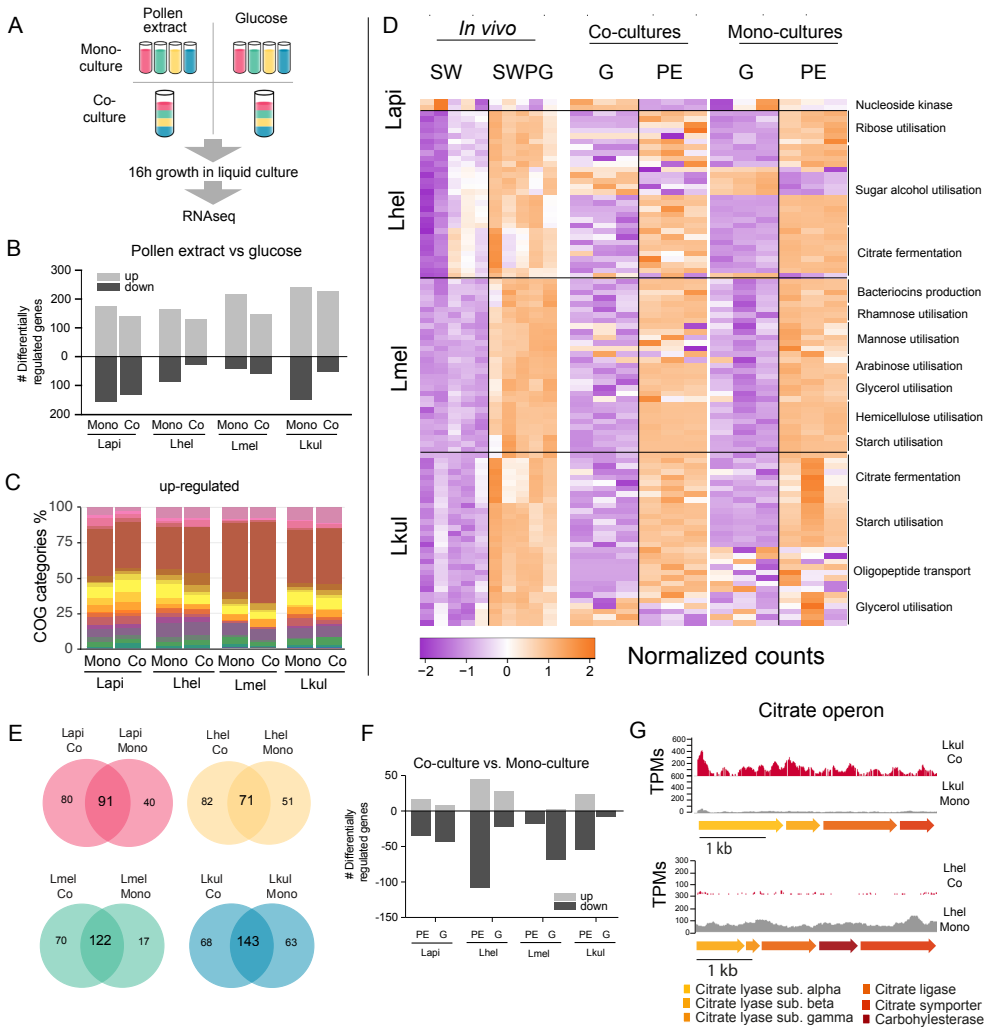

Figure 5: Transcriptome analysis of the four *Lactobacillus* species grown *in vitro* in pollen extract or in glucose. **(A)** Scheme of the 2x2 experimental design. Species were grown alone or together, in either glucose (G) or pollen extract (PE). **(B)** Number of differentially regulated genes by 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 presence of PE if compared to G. **(D)** 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 RNAseq 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, i.e. *in vivo*, co-cultures, and mono-cultures. **(E)** Venn diagrams displaying overlap between co-cultures and mono-cultures relative to the genes upregulated in PE relative to G. **(F)** Number of differentially regulated genes by the four species in co-culture relative to mono-culture**.** Up- and down-regulated genes are shown in different gray tones. **(G)** 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.

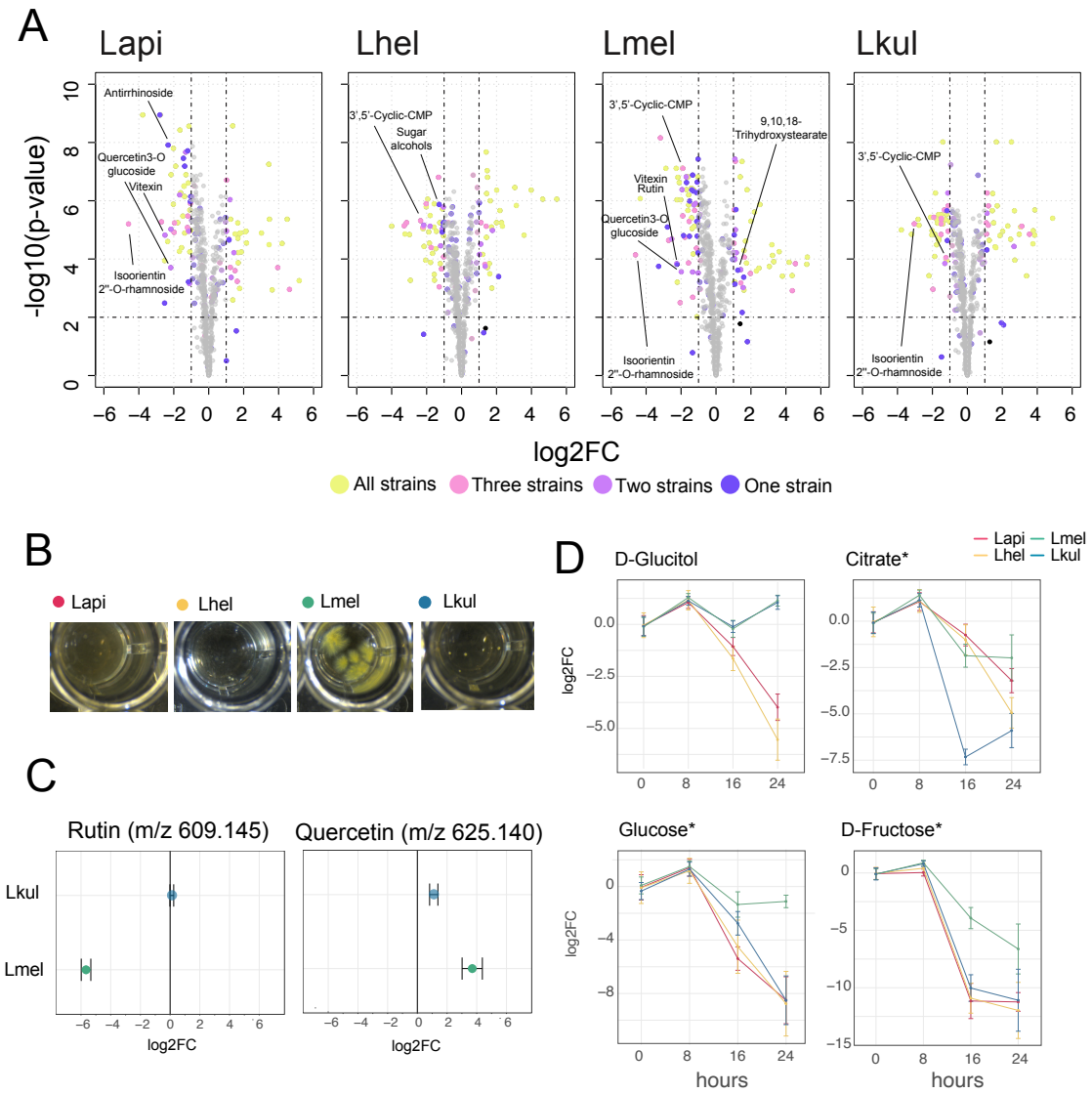

Figure 6. Metabolomics analysis confirms differences in the utilization of pollen-derived glycosides across the four *Lactobacillus* species. (A) Volcano plots displaying ions with significant fold changes (FC) for each of the four species after 16 h of growth in pollen extract versus glucose. Each dot corresponds to an ion in the untargeted metabolomics dataset. Different shades of blue represent ions that significantly change over time in one or more than one 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 h of incubation. 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 h of incubation. **(D)** GC-MS detection of key carbon metabolism metabolites over time. Log_2FC 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 Dataset 8.