1 **IHF regulates host colonization factors in the bee gut**

2 **symbiont *Frischella perrara***

3

4 K. Schmidt^{1,3}, S. Leopold-Messer², O. Emery¹, Y. El-Chazli¹, T. Steiner¹, J. Piel² & P. Engel¹

5

6 1 Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

7 2 Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, Zürich,
8 Switzerland

9 3 Current affiliation: Roche Diagnostics International Ltd, Rotkreuz, Switzerland

10

11 Correspondence:

12 Prof. Philipp Engel

13 Department of Fundamental Microbiology

14 University of Lausanne, CH-1015 Lausanne, Switzerland

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

16 e-mail: philipp.engel@unil.ch

17

Abstract

Bacteria colonize specific niches in the animal gut. However, the genetic basis of these associations is often unclear. The protobacterium *Frischella perrara* colonizes the hindgut of honey bees, where it causes a characteristic immune response leading to the formation of the scab phenotype. Genetic determinants required for the establishment of this specific association are unknown. Here, we isolated three point mutations in the genes encoding the DNA-binding protein integration host factor (IHF). All three mutants formed larger colonies than the wild type *in vitro* and did not produce an aryl polyene metabolite, which confers a yellow color to *F. perrara* colonies. Inoculation of microbiota-free bees with one of the mutants drastically decreased gut colonization of *F. perrara* and abolished scab development. Using RNAseq we find that IHF affects the expression of potential colonization factors, including a colibactin biosynthetic gene cluster, two Type 6 secretion systems, pili genes, and the aryl polyene biosynthesis pathway. Gene deletions of these components revealed different colonization defects. While some mutants were abolished in their capacity to colonize the gut, others colonized but did not trigger the scab phenotype. IHF is conserved across many bacteria and may thus regulate host colonization also in other animal symbionts.

Introduction

The digestive tract of many animals is colonized by specialized gut symbionts that occupy distinct physical niches and utilize diverse nutrients. Owing to the availability of genetic tools, gnotobiotic animal models and multi-omics approaches, we can now study the genetic features that allow gut symbionts to colonize various animal hosts including mammals, fishes, and insects (1-10).

The Western honey bee, *Apis mellifera*, is a particularly interesting model to characterize colonization factors of bacterial gut symbionts, due to its agricultural importance and the tractability of its gut microbiota. Honey bees harbor relatively simple yet highly specialized gut microbiota composed of 8-10 bacterial genera (11, 12). The wide distribution of this community across social bees suggests long evolutionary associations with the host (13). Moreover, different members of the bee gut microbiota colonize distinct physical niches along the gut. Lactobacilli and Bifidobacteria predominate in the posterior hindgut (rectum), while Gammaproteobacteria (*Frischella perrara* and *Gilliamella* species) and a betaproteobacterium (*Snodgrassella alvi*) preferentially colonize the anterior hindgut (ileum and adjacent pylorus, i.e., the transition zone between midgut and ileum) (11, 14-16). The partitioning of these bacteria into distinct gut compartments suggests the existence of specific bacterial mechanisms that facilitate colonization.

Most bacteria of the bee gut microbiota can be cultured and experiments with gnotobiotic bees have been established (17-21). Moreover, genomic analyses have provided important insights about the functional potential of bee gut symbionts and their adaptation to the gut environment (21-29). Yet, little is known about which genes are directly involved in establishing colonization in the bee gut and how these genes are regulated. The only symbiont that has been extensively studied in this respect is *S. alvi*. Using transposon sequencing and transcriptome analysis, Powell et al. determined genome-wide host colonization factors in *S. alvi* (7). Most genes with strong fitness effects were found to belong to three major categories: extracellular interactions, metabolism, and stress response. In particular, genes for attachment and biofilm formation were

highly beneficial for colonization, which is in agreement with the observation that this gut symbiont adheres to the host epithelium of the ileum and forms a multispecies biofilm with *Gilliamella*.

The honey bee gut symbiont *Frischella perrara* belongs to the recently described family Orbaceae within the Gammaproteobacteria (30). It is taxonomically close to the bee gut symbiont *Gilliamella* and also shows a similar metabolism as *Gilliamella* (31). However, compared to the other members of the bee gut microbiota, *F. perrara* shows a rather distinctive colonization phenotype. While *S. alvi* and *Gilliamella* have been reported to colonize the host epithelium of the entire ileum (32, 33), *F. perrara* preferentially colonizes the transition zone between the midgut and the ileum, i.e., the pylorus. Moreover, colonization with *F. perrara* leads to the appearance of a brown to black material on the luminal side of the epithelial surface between the cuticle layer of the host tissue and the adherent *F. perrara* cells (34). This so-called scab phenotype forms after 5-7 days post-colonization and has so far not been reported to be triggered by any other gut symbiont than *F. perrara*. Transcriptome analysis of the host showed that *F. perrara* elicits a specific immune response which includes the upregulation of the host melanization pathway likely responsible for the formation of the scab phenotype (17).

Genome sequencing of the type strain of *F. perrara* and comparison with other genomes of the Orbaceae family revealed the presence of several genomic islands that may be involved in the specific interaction of *F. perrara* with the host (35). These include a biosynthetic gene cluster for the production of the genotoxic metabolite colibactin (Cib), two distinct Type VI Secretion systems (T6SSs), and associated effector proteins, Type I secretion systems, and fimbrial low-molecular-weight protein (Flp) pili genes. However, currently no genetic tools are available for *F. perrara* precluding studies about the role of these genetic factors in gut colonization or the induction of the scab phenotype.

Here, we report the isolation of a spontaneous mutant of *F. perrara* that possesses a strong colonization defect *in vivo*. Resequencing of the mutant revealed a single nonsynonymous point mutation in the gene encoding the alpha subunit of the DNA-binding protein integration host

89 factor (IHF). Using a combination of gnotobiotic bee experiments, transcriptomics, and
 90 metabolite analyses, we characterized the genes regulated by IHF. We then established a gene
 91 deletion strategy for *F. perrara*, which allowed us to knockout some of the IHF-regulated genes
 92 and show that they impact gut colonization and scab development to different extent.

Results

Isolation of spontaneous IHF mutants affecting growth and colony morphology of *F. perrara*

Culturing *F. perrara* type strain PEB0191 on modified tryptone yeast glucose (mTYG) agar resulted in the formation of yellow colonies. However, we occasionally observed the appearance of larger white colonies among the yellow ones (**Figure 1A**). Restreaking white colonies on fresh mTYG agar usually resulted in yellow colonies again. However, three white colonies that we identified in independent experiments did not change their appearance anymore, suggesting that we had isolated stable 'white' variants of *F. perrara* PEB0191 (**Figure 1B**). Genome sequencing of the white variants revealed the presence of three different non-synonymous point mutations in the genes encoding the integration host factor (IHF). IHF is a widely distributed DNA-binding protein consisting of the IhfA/B heterocomplex (36-38). Strikingly, two point mutations were identical to each other, but occurred in the different subunits of IHF (*ihfA* and *ihfB*) resulting in a proline to lysine change at amino acid position 82 and 83, respectively (**Figure 1C**). The third point mutation resulted in a lysine to serine change at position 38 of IhfA (**Figure 1C**). Homology modelling showed that these amino acids are located in the region interacting with DNA, suggesting that the three mutations impact the DNA-binding properties of IHF (**Figure 1 D**, (39)). As two of the isolated mutants occurred when generating gene deletion of *F. perrara*, they harbored additional genetic modifications (see Methods). Therefore, we focused further characterization of IHF on the mutation Pro83Lys (hereafter *ihfA**) that occurred in the wild type (wt) background of *F. perrara*. While *IhfA** consistently formed larger colonies than the wt strain on mTYG agar (**Figure 1A and 1B**), there was no significant difference in growth in liquid culture (Permutation test, $p = 0.097$; **Supplementary Figure 1A**). However, light microscopy showed that cells of the mutant strain were on average slightly longer than cells of the wt (Kolmogorov-Smirnov test $p < 0.0001$, **Supplementary Figure 1B and 1C**).

***F. perrara* produces an aryl polyene secondary metabolite that is responsible for the yellow colony morphotype**

F. perrara PEB0191 encodes a genomic island that is homologous to aryl polyene (APE) biosynthetic gene clusters present in other Gammaproteobacteria (**Figure 2A**). APEs are polyunsaturated carboxylic acids conferring a yellow pigmentation to bacterial cells (40, 41). To assess if *F. perrara* wt, but not the *ihfA** mutant, produces an APE, we cultivated both strains, extracted the pellets, and analyzed the extracts by liquid chromatography coupled to heated electrospray ionization high-resolution mass spectrometry (HPLC-HES-HRMS). The data revealed a strongly UV-Vis-absorbent ion peak at m/z 323.1647 $[M+H]^+$, which had a suggested molecular formula of $C_{21}H_{23}O_3$ (**Figure 2B** and **2C**). In the *ihfA** mutant, this ion was only present at trace amounts (**Figure 2C**). To characterize the metabolite in greater detail a larger pellet of *F. perrara* PEB0191 wt cultures was extracted and purified by several HPLC runs. Mass spectrometry- (MS) and UV-Vis-guided fractionation yielded an enriched extract that was analyzed by nuclear magnetic resonance (NMR) spectroscopy. The characteristic ions detected in MS-MS fragmentation experiments (**Supplementary Figure 2**), the UV-Vis spectrum with an absorption maximum at 415 nm (**Supplementary Figure 3**) in conjunction with NMR data (**Supplementary Figure 4-10, Supplementary Table 1**) suggest an aryl polyene structure identical to that reported (42, 43) (**Figure 2D**). Unfortunately it was not possible to connect the NMR substructures because the central methines could not be assigned to chemical shifts (**Figure 2D, Supplementary Table 1**). Comparison of the organic extracts of *F. perrara* wt and *E. coli* CFT073 provided further evidence that both produce the same compound (**Supplementary Figure 11**). Combined, these results suggest that the APE pathway is responsible for the yellow color of the wt colonies of *F. perrara* and is suppressed in the *ihfA** mutant.

The *ihfA** mutant of *F. perrara* has a colonization defect and does not cause the scab phenotype

As APEs have been shown to increase protection from oxidative stress and contribute to biofilm formation (42, 43), we sought to test if the *ihfA** mutation impacts bee gut colonization. We mono-associated microbiota-free bees with either *F. perrara* wt or *ihfA**. Colonization with the wt strain resulted in a visible scab in 50% and 80% of all bees after five and ten days of colonization, respectively (n=18 and n=36 for both treatments for day 5 and day 10, respectively, **Figure 3A and 3B**). In contrast, none of the bees colonized with the *ihfA** mutant developed a visible scab phenotype. To determine whether this difference was due to a general colonization defect of *ihfA**, we quantified the colonization levels of *F. perrara* at day 5 and day 10 post colonization using colony forming units (CFUs). While there was a trend towards lower colonization levels (fewer CFUs and more bees without detectable colonization) for the *ihfA** mutant at day 5 post colonization, the difference was not statistically significant (**Figure 3C**, Wilcoxon rank-sum test p-value=0.076). However, at day 10 post-colonization, bees colonized with the wt strain showed significantly higher CFUs than the *ihfA** mutant (**Figure 3C**, Wilcoxon rank-sum test p-value<0.0001). In fact, in 50% of all bees (n = 36) the colonization levels of *ihfA** were below the detection limit of 500 CFUs (colony forming units) (**Figure 3C**).

As the quantification of *F. perrara* was based on CFUs obtained for whole-gut tissue, we carried out a second experiment, in which we specifically assessed the colonization levels in the pylorus and the ileum region of the honey bee gut, using both CFUs and quantitative PCR (**Supplementary Figure S12**). The results were comparable to those obtained for the whole gut: at day 10 post colonization, there was a significant difference in the colonization levels of the wt and *ihfA** in both the pylorus and the ileum (Wilcoxon rank-sum test p-value<0.05).

To obtain a better understanding of the colonization dynamics of *F. perrara* wt and the *ihfA** mutant, we conducted a third gnotobiotic bee experiment in which we inoculated microbiota-free bees with one of the two strains and followed the colonization levels over 12 timepoints from day

0 (i.e. 4 h post inoculation) until day 22 post inoculation (**Figure 3D**). We used the same optical density of the two strains for inoculation. However, this corresponded to fewer CFUs for the wt compared to the *ihfA** mutant. From the first time point at 4 h post inoculation until day 4 post inoculation, the bacterial levels were below the detection limit (i.e., below 500 CFUs) at both conditions. Between day 4 and day 8 post inoculation, the abundance of *F. perrara* wt increased rapidly to about 10⁶ CFUs per gut and then steadily further to 10⁷ CFUs per gut until the last timepoint. In contrast, the levels of the *ihfA** mutant remained low until day 10 post colonization and reached on average no more than 10⁵ CFUs per gut until the last timepoint at day 22 post colonization. In summary, these results show that *ihfA** has a colonization defect. It has a delayed colonization dynamics as compared to the wt, does not reach the same bacterial loads, and does not cause the scab phenotype.

Genes involved in symbiotic interactions are upregulated in *F. perrara* wt relative to the *ihfA mutant**

IHF may not have a direct effect on gut colonization, but rather regulate the gene expression of host colonization factors. To test this, we assessed the transcriptional differences by RNA sequencing (RNA-seq) between the wt and *ihfA** mutant when grown *in vitro*. We found that 358 out of 2,337 genes encoded in the genome of *F. perrara* PEB0191 were differentially expressed with a log2-fold change >|2| between the two strains (Fisher's exact test with p<0.05 and FDR<5%). Of those, 237 and 121 genes were up- and down-regulated in *F. perrara* wt versus *ihfA**, respectively (**Figure 4A and 4B, Supplementary Dataset 1**). Among the genes upregulated in the wt, 'Intracellular trafficking, secretion, and vesicular transport' (COG U), 'Extracellular structure' (COG W), 'Lipid transport and metabolism' (COG I), 'Mobilome: prophages and transposases' (COG X), and 'Secondary metabolites biosynthesis, transport and catabolism' (COG Q) were significantly enriched (Fisher's exact test, BH-adjusted p-value < 0.01; see **Supplementary Dataset 2**). Genes belonging to these three categories include different subunits and effectors of the two T6SSs of *F. perrara*, the Clb biosynthesis gene

199 cluster, various components of the Flp pili, and a RTX (repeats in toxin) toxin belonging to the
 200 Type I secretion system family (**Supplementary Dataset 1**). Also, the genes of the APE
 201 biosynthesis gene cluster were among the upregulated genes, which is in line with the
 202 production of the corresponding metabolites in the wt but not in the mutant strain (**Figure 2**).
 203 Interestingly, a relatively large proportion of the upregulated genes encoded hypothetical or
 204 poorly characterized proteins. In fact, genes without COG annotation were also enriched relative
 205 to the entire genome of *F. perrara* (Fisher's exact test, BH-adjusted p-value < 0.01; see
 206 **Supplementary Dataset 2**). Many of the upregulated genes were organized in genomic islands,
 207 with the largest one including the biosynthesis gene cluster of Clb and many hypothetical
 208 protein-encoding genes (**Figure 4C**). T6SS and Clb biosynthesis genes were among the genes
 209 with the highest fold changes relative to *ihfA** mutant (**Figure 4B**, 28 of 32 genes with log2-fold
 210 change >6). Moreover, 64% of the upregulated genes (152/237) belonged to the *F. perrara*-
 211 specific gene content as based on our previously published genome comparison of *F. perrara*
 212 PEB0191 with four other strains of the family Orbaceae (three of the genus *Gilliamella* and one
 213 of the genus *Orbus*, **Supplementary Dataset 1**) (35).

214 Among the 121 down-regulated genes, only COG category O ('Posttranslational modification,
 215 protein turnover, chaperones') was statistically enriched (Fisher's exact test, BH-adjusted p-
 216 value < 0.01, see **Supplementary Dataset 2**). Moreover, only a small fraction (12%) belonged
 217 to the "*F. perrara*-specific genes", and fewer genes were organized into genomic islands. A
 218 more detailed inspection of the annotation revealed that a large number of the down-regulated
 219 genes were involved in transport and metabolism (40 genes), transcriptional regulation (10
 220 genes), and protein folding (8 genes), highlighting clear differences in the functional roles of the
 221 up- and down-regulated genes. The two genes with the highest fold change (log2-fold change <-
 222 5) both encoded transcriptional regulators. One of them, *dksA* (Fpe_01158), is located upstream
 223 of the *mrsA/mrsB* antioxidant system (Fpe_01159 to Fpe_01162), which was also among the
 224 down-regulated genes. The other one is part of the two-component regulator system *basS/basR*
 225 (Fpe_02097 and Fpe_02098), which has been reported to act as an iron- and zinc-sensing

transcriptional repressor and activator in *E. coli* (44, 45). Taken together, these results show that many accessory genes known to be involved in symbiotic interactions (colibactin, Flp pili, T6SS) are upregulated in *F. perrara* wt as opposed to *ihfA*^{*}, providing a list of candidate genes responsible for the colonization defect of the *ihfA*^{*} mutant.

T6SS, pili, APE biosynthesis and Clb biosynthesis genes are expressed during bee gut colonization

To test if the genes upregulated *in vitro* in the wt relative to *ihfA*^{*} were expressed *in vivo*, we determined the transcriptome of *F. perrara* wt at day 5 and day 10 post colonization. A total of 260 (149 up and 111 down) and 298 (162 up and 136 down) genes were differentially expressed at day 5 and day 10 post colonization relative to growth *in vitro* (log2-fold change >|2|, quasi-likelihood F-test with p<0.05 and FDR<5%, **Supplementary Dataset 3**). There was a considerable overlap of the genes differentially regulated *in vivo* relative to *in vitro* between the two timepoints (115 and 80 shared up- and down-regulated genes, respectively). At both timepoints, the COG category 'Carbohydrate transport and metabolism' (COG G) was significantly enriched among the genes upregulated *in vivo* relative to the entire genome (**Supplementary Dataset 4**). In addition, at timepoint day 10, also the COG category (P) 'Inorganic ion transport and metabolism' was enriched (P adj < 0.01, Fisher's exact test, **Supplementary Dataset 4**). Genes belonging to these two categories encoded transporters for different sugars (Phosphotransferase systems), iron, and transferrin (**Supplementary Figure 13, Supplementary Dataset 3**). In addition, a catalase gene and several genes for the biosynthesis of the amino acid tryptophan were upregulated at both timepoints. However, only 14 and 19 genes of the ones upregulated *in vitro* in the wt relative to the *ihfA*^{*} mutant (see **Figure 5**), were also upregulated *in vivo* at day 5 and day 10 post-colonization, respectively (**Supplementary Dataset 3**). In fact, several of the T6SS, APE, and Clb biosynthesis genes remained unchanged or were slightly downregulated *in vivo* relative to *in vitro* (**Supplementary Dataset 3 and Supplementary Figure 13**). This was not unexpected, because the *in vitro*

RNAseq analysis our reference had shown that these genes are already expressed when grown on mTYG agar, which we used as a condition for the *in vivo* analysis. Indeed, when comparing count-normalized gene expression (as measured by TPM) across the different conditions, we found that most of the T6SS machinery, APE biosynthesis, pilus, and iron uptake genes were expressed at similar levels at both timepoints *in vivo* as during *in vitro* growth (**Figure 5, Supplementary Figure 14**). Only the Clb genomic island and some of the VgrG-like T6SS effector genes had clearly lower TPM values *in vivo* than *in vitro*, yet higher than in *ihfA** mutant (**Figure 5C and 5F**). These results suggest that most of the genes upregulated in the wt relative to *ihfA** *in vitro* were also expressed at high level at both timepoints *in vivo*.

Gene deletions of IhfA-regulated genes result in impaired gut colonization and/or abolish scab development

To test the direct impact of IhfA-regulated genes on host colonization and/or scab development we established a gene deletion strategy for *F. perrara* based on a two-step homologous recombination procedure (see methods). This allowed us to create six different non-polar in-frame gene deletion mutants of potential colonization factors regulated by IHF (Supplementary Table S2). We deleted an essential gene of the colibactin biosynthesis pathway ($\Delta clbB$), both *hcp* genes of the two T6SSs, either separately or as double mutant ($\Delta hcp1$, $\Delta hcp2$, and $\Delta hcp1/\Delta hcp2$), the gene encoding the major Flp pili subunit ($\Delta pilE$), and the entire APE biosynthesis gene cluster ($\Delta apeA-R$). Deletion mutants were confirmed by genome re-sequencing. The $\Delta apeA-R$ mutant was the only strain not forming yellow colonies anymore, which is consistent with the idea that the aryl polyene pathway is responsible for the yellow color. To compare the six gene deletion strains to *F. perrara* wt and the *ihfA** mutant in terms of their ability to colonize the honey bee gut and induce the scab phenotype, each strain was inoculated into microbiota-free bees and CFUs assessed in the pylorus/ileum region ten days post inoculation. As expected, *F. perrara* wt successfully colonized the pylorus/ileum region of

all analyzed gnotobiotic bees (median = $9.56 \times 10^6 \pm 5.06 \times 10^6$ CFUs per gut; n=18 bees, two independent experiments) and induced the scab phenotype in 16 of 18 bees (**Figure 6**). In contrast, *F. perrara ihfA** colonized poorly compared to the wt (median= $1.46 \times 10^4 \pm 3.19 \times 10^5$ CFUs per gut, n= 20 bees, Wilcoxon rank-sum test $P < 0.001$), and not a single bee developed the scab phenotype. Of the six tested gene deletion mutants, only the mutant of T6SS-1 ($\Delta hcp1$) reached wt colonization levels and also induced the scab phenotype in most of the bees (15 out of 20). The other five mutants all exhibited lower colonization levels than the wt strain. However, the severity of the colonization defect varied between the mutants, and while some of the mutants still caused the scab phenotype, others did not. For example, the Clb biosynthesis gene cluster mutant $\Delta clbB$ induced the scab phenotype in all but one bee, despite the fact that the colonization levels were about two-fold lower than for the wt (median= $4.24 \times 10^6 \pm 2.87 \times 10^6$ CFUs per gut). The gene deletion mutant of the T6SS-2 ($\Delta hcp2$) and the double mutant of both T6SSs ($\Delta hcp1/\Delta hcp2$) reached similar colonization levels as the $\Delta clbB$ mutant. Yet, none of the bees colonized with these two mutants developed the scab phenotype not even those that had very high bacterial counts (**Figure 6**). The deletion mutant of the APE biosynthesis pathway ($\Delta apeA-R$) showed again a different colonization phenotype: for some bees no colonization was detected, while others showed similar colonization levels as the wt of *F. perrara*. However, in contrast to the $\Delta hcp2$ and the $\Delta hcp1/\Delta hcp2$ mutants, the $\Delta apeA-R$ mutant still induced the scab phenotype, but only in the bees that had high bacterial loads. Finally, the $\Delta pilE$ mutant had the strongest impact on colonization. Only five out of 20 bees had detectable levels of *F. perrara* in the gut (limit of detection: 75 CFUs per gut) and none of the bees developed the scab phenotype. These results confirm that IHF regulates various host colonization factors that when deleted cause distinctive colonization defects.

Discussion

In this study, we identified genetic factors that allow the gut symbiont *F. perrara* to colonize its specific niche in the pylorus of the honey bee gut and to induce the scab phenotype. In summary, we found that a nonsynonymous mutation in the gene encoding the alpha subunit of the integration host factor (IhfA) leads to a strong colonization defect of *F. perrara* characterized by reduced bacterial loads in the gut and the absence of the scab phenotype. Using RNAseq, we showed that the *ihfA* mutation (*ihfA**) causes the downregulation of diverse genes known to be involved in symbiosis, including two T6SSs and associated effector proteins, the Clb biosynthetic gene cluster, Flp pili, and an RTX toxin locus. One of the IHF-regulated biosynthetic gene clusters encoded an APE, which we structurally characterized using specialized metabolite analysis, and which seems to be responsible for the yellow appearance of *F. perrara* colonies when grown on mTYG agar. Most of the genes regulated by IhfA were expressed during gut colonization, and gene deletion analysis revealed that the colibactin biosynthesis gene cluster, one of the T6SSs, the Flp pili, and the APE biosynthetic gene cluster impact gut colonization and in part also scab development.

These results advance our understanding of the genetic factors that facilitate symbionts to colonize niches in the animal gut. Specifically, we find that the DNA-binding protein IhfA plays an important role in gut colonization. IhfA is a histone-like, nucleoid-associated protein (NAP) that forms a heterodimer together with IhfB. IHF binds to and bends DNA in a sequence specific manner (37-39, 46), thereby facilitating the recombination of mobile DNA elements (37, 38) and influencing gene expression. We identified three different mutations in IHF, all located in the region of the protein interacting with the DNA. All three IHF mutants formed larger colonies than the wt strain on mTYG agar. Together, this suggests that the identified mutations change the DNA-binding properties of IHF resulting in broad transcriptional changes in *F. perrara* that provide a growth advantage in vitro relative to the wt strain.

Like other microbiota members inhabiting the pylorus/ileum region of the honey bee gut (*Gilliamella* and *S. alvi*), *F. perrara* colonizes the epithelial surface, where it adheres to the

cuticular lining and forms a thick biofilm-like layer (34). In case of the human pathogen *Haemophilus influenzae*, IHF has been shown to bind to extracellular DNA, which increases biofilm stability (47). Therefore, it is possible that IHF has a direct impact on gut colonization of *F. perrara* by stabilizing the biofilm formed on the host epithelium of the pylorus/ileum. However, based on our gene expression data and the results of the gene deletion experiments, it seems more likely that IHF influences gut colonization by regulating downstream genes involved in host interaction. Several of the genes regulated by IHF are known to play important roles in the adhesion to surfaces or the formation of biofilms. For example, pili are key factors for adhesion across many bacteria (48-51) and have already been shown in the case of the bee gut symbiont *S. alvi* to be beneficial for gut colonization and biofilm formation (7). Furthermore, several of the deletion mutants of IHF-regulated genes showed colonization defects in our gnotobiotic bee experiments providing further evidence that IHF impacts gut colonization through its effect on gene expression rather than through a direct role in biofilm formation.

The pili mutant ($\Delta pilE$) had the strongest colonization defect of the tested gene deletion strains, likely because pili mediate the adhesion to the host epithelium, and therefore allow *F. perrara* to persist and replicate in the pylorus/ileum region of the honey bee gut. The deletion mutant of the APE biosynthesis pathway ($\Delta apeA-R$) also showed a clear colonization defect. However, while some of the bees inoculated with this mutant had no detectable levels of *F. perrara* in the gut, others reached similar levels as bees colonized with the wt strain and also developed the scab phenotype. Aryl polyenes can protect bacteria from reactive oxygen species (42, 52) which are produced by insects as part of the host immune response in the midgut (53, 54). Our time course experiment revealed that in the first few days of colonization the bacterial numbers of *F. perrara* in the gut are much lower than in the initial inoculum (**Figure 2**) suggesting that the bacteria experience a significant population bottleneck at the beginning of the colonization process. It is possible that the $\Delta apeA-R$ mutant is impaired in its ability to resist reactive oxygen species (or other physicochemical stressors) when passing through the anterior sections of the honey bee gut and therefore reaches its niche in the pylorus in only a fraction of the inoculated

bees. However, APEs have also been implicated in other functions such as biofilm formation (52). Therefore, future studies will be needed to elucidate how the APE biosynthetic pathway of *F. perrara* impacts honey bee gut colonization. Given the widespread occurrence of these biosynthetic genes throughout different bacteria, this might help to understand their role in host colonization in a much wider context (43).

A somewhat unexpected result of our mutant analysis was that the deletion of one of the essential NRPS/PKS enzymes of the Clb biosynthetic gene cluster did not affect scab development and had only a weak impact on gut colonization (2-fold lower levels than wt). Studies on *E. coli* have shown that the Clb biosynthesis pathway induces DNA damage in eukaryotic cells (55) and contributes to tumorigenesis in the mammalian gut (56, 57). We have previously shown that *F. perrara* also induces DNA damage in eukaryotic cells and that this is dependent on a functional Clb biosynthesis pathway (35). Therefore, we had speculated that the genotoxic activity of colibactin may trigger the local melanization response and the development of the scab phenotype upon colonization with *F. perrara* (17, 34). However, the results presented here show that colibactin is not causing the scab phenotype. Therefore, other specific characteristics of *F. perrara* must explain why this bacterium but no other honeybee gut symbiont is causing the scab phenotype.

Our gnotobiotic bee experiments showed that the development of the scab phenotype depends to a large extent on the bacterial loads of *F. perrara* in the gut. Bees that had $<10^6$ bacteria in the gut did not develop visible scab at day 10 post colonization (**Figure 3 and Figure 7**), which suggests that a certain number of bacteria is needed to elicit this characteristic host response. On the contrary, not all bees with high levels of colonization developed a scab phenotype indicating that high loads are necessary but not sufficient to trigger the development of a visible scab phenotype. Specifically, the $\Delta hcp2$ mutant of T6SS-2 and the Δhcp double mutant of T6SS-1 and T6SS-2 both reached relatively high colonization levels, yet not a single bee developed the scab phenotype. T6SS are usually involved in interbacterial warfare by injecting toxins into neighboring bacterial cells (58-60). However, there is an increasing number of

385 studies showing that certain T6SS effectors can also target eukaryotic cells and modify diverse
 386 eukaryotic processes, including adhesion modification, stimulating internalization, cytoskeletal
 387 rearrangements and evasion of host innate immune responses (61). While *F. perrara* is not the
 388 only bee gut symbiont harboring T6SS, the effector protein repertoires differ tremendously
 389 between different species and even strains of the same species (62). Thus, it is possible that *F.*
 390 *perrara* carries effector proteins, which target the host rather than other bacteria eliciting the
 391 melanization response and scab phenotype in the bee gut.

392 We did not carry out a genome-wide screen for host colonization factors, and only generated
 393 deletion mutants of a few of the IHF-regulated genes. There are probably many other factors
 394 that also contribute to gut colonization. For example, our *in vivo* RNA-seq experiment indicated
 395 that metabolic genes involved in tryptophan biosynthesis, sugar transport, and iron uptake are
 396 upregulated during gut colonization. This is in line with the TnSeq screen performed in *S. alvi*,
 397 which revealed that genes of essential amino acid biosynthesis pathways and iron uptake are
 398 important colonization factors of the honey bee gut (63). It is however remarkable that *F. perrara*
 399 only upregulates the tryptophan biosynthesis pathway, but no other amino acid biosynthesis
 400 genes, which suggest a specific demand for the production of this specific amino acid during gut
 401 colonization.

402 While our study revealed that IHF is important for regulating host colonization factors in the
 403 honey bee gut, it remains to be elucidated whether IHF is a direct regulator of the identified
 404 genes. In *Vibrio fluvialis*, IHF binding sites were identified upstream of several T6SS gene
 405 clusters, indicating that direct transcriptional regulation of similar genes by IHF exists in other
 406 bacteria (64). Alternatively, IHF may act upstream of another regulator that controls the
 407 expression of the identified genes. In the plant pathogen *Dickeya zeae*, IHF was suggested to
 408 positively regulate different virulence factors through binding to the promoter region of a
 409 diguanylate cyclase gene, increasing the production of the secondary messenger c-di-GMP
 410 (65). In *F. perrara*, we found that the transcriptional regulators BasR and DksA were
 411 substantially downregulated in the wt versus the *ihfA** mutant. Hence, it is possible that at least

412 some of the differentially regulated genes may not be under the direct control of IHF but
 413 regulated via BasR or DskA. In the future, chromatin immunoprecipitation sequencing (ChIP-
 414 Seq) could be applied to experimentally identify the specific binding sites of IHF in *F. perrara*
 415 and to understand how IHF regulates the gene expression of host colonization factors across
 416 bacteria.

417 In conclusion, we identified important gut colonization factors of the bee gut symbiont *F. perrara*
 418 and show that they are regulated by IHF. The wide occurrence of these genes in host-
 419 associated bacteria suggests similar roles in other environments and calls for a more detailed
 420 functional characterization. Our approach to create clean gene deletion mutants expands the
 421 available genetic toolbox for bee symbionts (33) and will help to dissect the molecular
 422 mechanisms of the identified gut colonization factors in this animal model.

423

Material and methods

Bacterial cultivation

F. perrara strains were cultivated on modified tryptone yeast extract glucose (mTYG) medium (0.2% Bacto tryptone, 0.1% Bacto yeast extract, 2.2 mM D-glucose, 3.2 mM L-cysteine, 2.9 mM cellobiose, 5.8 mM vitamin K, 1.4 μ M FeSO₄, 72.1 μ M CaCl₂, 0.08 mM MgSO₄, 4.8 mM NaHCO₃, 1.36 mM NaCl, 1.8 μ M Hematine in 0.2 mM Histidine, 1.25% Agar adjusted to pH 7.2 with potassium phosphate buffer) or Columbia Blood Agar (CBA) containing 5% defibrinated sheep blood (Oxoid)) and incubated in anaerobic conditions at 35 °C (8% H₂, 20% N₂, 78% CO₂ in a Vinyl Anaerobic Chamber, Coy Lab). Fresh bacterial cultures were used for each experiment. To this end, pre-cultures streaked out from glycerol stocks were grown for 48 hours and re-inoculated onto fresh plates for 16-24 hours of growth. For liquid cultures, pre-cultures streaked out from glycerol stocks were grown on TYG or BHIA plates for 48 hours and then inoculated into fresh liquid TYG or BHI. Cultures were incubated at 34 °C in a ThermoMixer C (Eppendorf) at 800 rpm for 16-24 hours of growth. Strains of *F. perrara* used in this study are listed in **Supplementary Table S2**.

Rearing and experimental colonization of honey bees

Microbiota-depleted bees were generated as described by (18). For experimental mono-colonization, bees were starved for 1-3 hours by removal of the sugar water solution. Then, bees were cooled down to 4 °C in a refrigerator or on ice to transfer them (head side first) into 1.5 ml microfuge tubes with a hole at the bottom. Tubes with bees were kept at room temperature. For inoculation, each bee was fed 5 μ l of *F. perrara* resuspended in sugar water:PBS (1:1 v/v) through the hole at the bottom of the microcentrifuge tube. The bacterial inocula were adjusted to an OD₆₀₀ of 0.01 or 0.1 depending on the experiment. Colonized bees were kept at 30 °C with 70% humidity while having access to sugar water and sterilized bee pollen *ad libitum* until sampling.

Tissue dissection and bacterial quantification from colonized honey bees

Bees were anesthetized by exposure to CO₂. The whole gut or desired gut tissue, e.g., pylorus or ileum was dissected using a scalpel. Malpighian tubules were removed from the gut tissue and the presence of a scab was documented using a dissection stereomicroscope (Leica) as described in (14, 17). The tissues were placed into 2 ml screw-cap tubes containing glass beads (0.75-1 mm diameter, Roth) and 500-1000 µl PBS depending on the experiment. Homogenization of the sample was done by bead beating (FastPrep-24 5g MP Biomedicals) for 40 seconds at a speed of 7.5 m/s. Serial dilutions (1:10) were performed for each homogenate and plated onto mTYG or BHI agar. Single colonies were counted to determine the total number of bacteria per gut tissue by multiplying with the dilution factor.

RNA sequencing of bacterial *in vitro* cultures

Fresh *F. perrara* cultures were prepared on TYG plates, harvested, and directly transferred into a tube containing TRI reagent (Sigma-Aldrich, Merck) and silica beads (0.1 mm diameter, Roth). Samples were immediately snap frozen in liquid nitrogen and stored at -80 °C until RNA extraction. RNA was extracted using a modified TRI reagent protocol. Samples were homogenized by bead beating with a FastPrep instrument with CoolPrep adapter (FastPrep-24 5G, cooled with dry ice) for two cycles of 45 seconds at speed of 6 m/s, including a 30 seconds break between each cycle. Samples were kept at room temperature for 5 min, subsequently extracted using chloroform and RNA was precipitated using isopropanol overnight at -20 °C. Precipitated RNA was dissolved in RNase-free water (Gibco) and incubated with DNase (NEB) to degrade remaining DNA. RNA samples were cleaned up using NucleoSpin RNA clean up kit (Machery-Nagel) according to manufacturer's protocol, eluted in RNase free water and stored at -80 °C until further use. RNA quality was assessed using Nanodrop, Qubit RNA HS kit (ThermoFisher Scientific) and Bioanalyzer instrument (Agilent). High quality RNA samples were sent to Lausanne Genomic Technology Facility (GTF, University of Lausanne) for RNA sequencing. Samples were depleted of 16S rRNA using RiboZero reagent (Illumina) and Truseq

477 stranded-RNA Zero libraries were generated (Illumina) before sequencing on an Illumina
478 HiSeq2500 generating single-end 100 bp reads.

479

480 **RNA sequencing of bacteria during honey bee gut colonization**

481 Microbiota-depleted bees were colonized with freshly grown *F. perrara*. A fraction of the
482 inoculum was directly transferred into 2 ml tube containing TRI reagent (Sigma-Aldrich, Merck)
483 and silica beads (0.1 mm diameter, Roth). These samples were immediately snap-frozen in
484 liquid nitrogen and stored at -80 °C until RNA extraction, which was carried out in the same way
485 as described in the previous section. For the *in vivo* RNA samples, bees were colonized with *F.*
486 *perrara* as described above and sampled at day 5 and day 10 post inoculation. Guts were
487 removed from anesthetized bees, the pylorus and first part of the ileum was dissected, and the
488 presence of the scab was recorded. For each sample, 10 pylori and ileums from bees coming
489 from the same cage were pooled in a 2 ml screw cap tube containing 750 µl of TRIreagent
490 reagent (Sigma-Aldrich, Merck), glass beads (0.75-1 mm diameter, Roth), and silica beads (0.1
491 mm diameter, Roth). Immediately after collection, samples were snap frozen in liquid nitrogen
492 and stored at -80 °C until RNA extraction. RNA was extracted as described in the previous
493 section. At each time point, four biological replicate samples were used for RNA sequencing,
494 i.e., bees for one sample came from independent cages of gnotobiotic bees. The GTF at the
495 University of Lausanne generated the libraries for the sequencing of the *in vivo* RNA samples.
496 Poly-A depletion was performed in order to enrich for bacterial mRNA and Ribo-zero rRNA
497 depletion was performed in order to remove prokaryotic and eukaryotic rRNAs. Then TruSeq
498 stranded mRNA libraries (Illumina) were generated. The eight libraries were sequenced on an
499 Illumina HiSeq 2500 to obtain single-end 125 bp reads. Libraries of the *in vitro* RNA samples
500 were generated as described before. The four libraries corresponding to the *F. perrara* inocula
501 used for colonizing the bees of the four replicates were sequenced on an Illumina MiniSeq at the
502 Department of Fundamental Microbiology (125 bp single-end reads). Two *in vivo* samples

(sample identifier D10_3M and D5_1M) sequenced on the HiSeq 2500 at the GTF were included in the MiniSeq run which confirmed that the two runs gave comparable results.

Differential gene expression and gene enrichment analysis

Raw FASTQ files provided by the GTF containing all reads and corresponding tags indicating whether they were accepted or filtered out according to the CASAVA 1.82 pipeline (Illumina). Only reads tagged as accepted were kept for further analysis. FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to control the quality of the data, followed by Trimmomatic (Trimmomatic-0.35) to trim adapters. Filtered and trimmed reads were mapped onto the *F. perrara* genome using Bowtie (bowtie2-2.3.2). Mapped reads were quantified using HTseq (Version 0.7.2). Differential gene expression analysis was done with the Bioconductor package EdgeR (66) using R scripts. For the in vitro RNAseq analysis, negative binomial models were fitted to the data and quantile-adjusted conditional maximum likelihood (qCML) common and tagwise dispersion were estimated. The conditional distribution for the sum of counts in a group was used to calculate the significantly differently expressed genes using an Exact test with a false discovery rate <5%. Only genes detected in all samples with at least 1 count per million were used for the analysis. For the in vivo RNAseq analysis, we followed the recommendations specified in the EdgeR user guide for generalized linear models. In short, read counts were normalized by trimmed mean of M-values (67) producing scaling factors used by EdgeR to determine effective library sizes. Then, negative binomial generalized linear models were fitted for each condition and quasi-likelihood F-tests for each defined contrast (i.e., pairwise comparison between conditions) was used to assess the significance of differentially expressed genes. As for the in vitro RNAseq analysis, only genes with mapped reads with at least 1 count per million in all replicates and conditions were used for the analysis. Genes were considered as differentially expressed upon fulfilling the following criteria: (i) log2 fold change ≥ 2 (ii) a p-value of <0.05 and (iii) a false discovery rate <5%. Transcripts per million (TPM) were calculated as follows (68):

$$TPM \text{ of a gene} = \frac{\text{number of mapped reads} * \text{read length} * 10^6}{\text{total number of transcripts sampled} * \text{gene length in bp}}$$

530 Coverage plots in Figure 2A and Figure 4C were generated with bam2wig.pl contained in the
531 Bio-ToolBox-1.68 (<https://github.com/tjparnell/biotoolbox>) and visualized with the Integrated
532 Genome Browser (69). Scripts for analyzing the RNAseq data can be found under the following
533 Switchdrive link: <https://drive.switch.ch/index.php/s/kCNTp4g7n60ffMi>.

534

535 **DNA extraction of dissected gut tissue homogenates**

536 DNA/RNA extraction was performed using an adapted hot phenol protocol from (18) consisting
537 of sample thawing on ice, homogenization by bead beating at 7.5 m/s for 40 seconds (FastPrep-
538 24 5g MP Biomedicals), two phenol extractions, and an ethanol precipitation overnight at -80 °C.
539 DNA/RNA mixture was pelleted and eluted in RNase free water (Gibco). Samples were treated
540 with RNaseA and DNA was purified using a Gel and PCR purification kit (Machery-Nagel).

541

542 **Bacterial quantification by qPCR**

543 Bacterial absolute abundances were determined using quantitative PCR (qPCR) assays
544 targeting the 16S rRNA gene of *F. perrara*. Normalization was based on the number of host
545 actin gene copies as described in (18, 70). Primer sequences and other primer characteristics
546 are given in **Supplementary Table S3**. qPCR was conducted on a StepOne Plus instrument
547 (Applied Biosystems) with the following run method: a holding stage consisting of 2 min at 50 °C
548 followed by 2 min at 95 °C, 40 cycles of 15 sec at 95 °C, and 1 min at 65 °C. A melting curve
549 was generated after each run (15 sec at 95 °C, 20 sec at 60 °C and increments of 0.3 °C until
550 reaching 95 °C for 15 sec) and used to assess specificity of PCR products. qPCR reactions
551 were performed in 10 µL reactions in triplicates in 96-well plates, and each reaction consisted of
552 1 µL of DNA, 0.2 µM of forward and reverse primer and 1x SYBR green “Select” master mix
553 (Applied Biosystems). For each target, standard curves were generated for absolute
554 quantification using serial dilutions (from 10⁷ to 10 copies) of the target amplicon cloned into the

plasmid vector. Absolute abundance of bacteria was calculated by using the standard curve and were normalized by the median actin copy number per condition (to account for differences in gut size) and by the amount of 16S rRNA copies per genome. For the calculation we used the following formula:

$$\text{Absolute bacterial abundance} = 10^{\left(\frac{Ct_{\text{target}} - \text{Intercept}}{\text{Slope}}\right)} * \text{dilution factor} \div \text{number of 16S rRNA copies} \div (\text{median of actin copies per condition})$$

Here, Ct_{target} is the cycle threshold of the target bacterium and intercept and slope correspond to the values calculated for the standard curve of the target bacterium.

Single cell microscopy

F. perrara was freshly grown from stock on plate. Subsequently, liquid overnight cultures starting at an OD₆₀₀ 0.1 were inoculated in TYG. Liquid cultures were grown overnight 16-24 hours with shaking at 34 °C in a ThermoMixer C (Eppendorf) in anaerobic atmosphere. The OD₆₀₀ was measured and adjusted to 0.1 – 0.01 and cells were distributed onto small agar patches on a microscopy slide (Milan S.A., Menzel-Gläser). Cells were observed under 1000 times magnification using a ZEISS imager M1 microscope with phase contrast (PH3) condenser. Pictures were acquired randomly on the slides with VisiView software, 8-bit images were corrected with ImageJ and the contrast was set at min-value of 0 and max-value of 4000. A 10 µM scale bar was added with ImageJ (71). MicrobeJ plugin for ImageJ was used to measure cell size and area using the acquired pictures (72).

Isolation and identification of aryl polyene compounds

F. perrara was streaked from the -80 °C glycerol stock onto mTYG plates. Plates were placed in an anaerobic tank with a pack of AnaeroGel 3.5 L (AN0035A Thermo Scientific) at 37 °C and incubated. Under aerobic sterile conditions, colonies were picked and 5 mL of mTYG medium were inoculated in a Hungate tube. After purging the liquid for 1 h with gas (83% N₂, 10% CO₂, 7% H₂ v/v) bacteria were grown at 37 °C. After one day of growth, *F. perrara* wt and the *ihfA**

mutant were individually cultivated in Hungate tubes containing 30 mL mTYG for three days. Pellets were harvested, extracted with dichloromethane and methanol, then extracts were analyzed by HPLC-HESI-HRMS. The data revealed a strongly UV-Vis absorbent ion peak at m/z 323.1647 $[M+H]^+$. To enrich this compound, 400 mL of mTYG were inoculated with 1 mL of pre-culture in a Hungate tube and purged as described above. After 5-10 days all but a few mL of medium were harvested and 400 mL of fresh mTYG was added under sterile aerobic conditions, purged with anaerobic gas and incubated at 37 °C. A total of 3 L of mTYG were used to obtain 6.53 g of cells, which were stored at -80 °C. Light exposure was minimized during all following steps of the aryl polyene enrichment procedure. A mixture of 340 mL dichloromethane and 170 mL of methanol was added to *F. perrara* cell pellets. After 2 h stirring at room temperature (RT), the suspension was filtered and 250 mL KOH (0.5 M) were added. After an additional 1 h of stirring at RT the pH was set to 6 using H₂SO₄ (1 M). The organic layer was washed twice with 500 mL water, once with 250 mL brine, dried over Na₂SO₄ and concentrated. The extract was separated by reverse-phase high performance liquid chromatography (RP-HPLC, Phenomenex Luna 5 µm C18, φ 21.2 x 250 mm, 15.0 mL/min, λ=420 nm) with water +0.1% formic acid (solvent A) and MeCN +0.1% formic acid (solvent B). Solvent compositions of 5% B for 4 min, a gradient to 95% B for 18 min, 95% B for 9 min and 5% B for 4 min were used. Fractions around 24 min exhibited strong UV absorption at 420 nm and were further purified by RP-HPLC (Phenomenex Luna 5µ Phenyl-Hexyl, φ 10 x 250 mm, 2.0 mL/min, λ=420 nm) with 80% B for 5 min and a gradient of 80% B to 100% B for 30 min. Peaks around 17 min were combined, concentrated and analyzed by NMR spectroscopy. HPLC-HESI-HRMS was performed on a Thermo Scientific Q Exactive mass spectrometer coupled to a Dionex Ultimate 3000 UPLC system. NMR spectra were recorded on a Bruker Avance III spectrometer equipped with a cold probe at 500 MHz and 600 MHz for ¹H NMR and 125 MHz and 150 MHz for ¹³C NMR at 298 K. Chemical shifts were referenced to the solvent peaks of DMSO-*d*₆ at δ_H 2.50 ppm and δ_C 39.51 ppm.

608

609 **Structural characterization of the aryl polyene**

610 ¹H NMR in conjunction with HSQC data suggested approximately 15 methines and aromatic
611 protons, one methoxy group and one methyl group (**Supplementary Figure 3-6**). From the
612 COSY spectrum, four methines of the conjugated double bond system, and two of the aromatic
613 protons could be connected (**Supplementary Figure 7, 8**). HMBC correlations from the methyl
614 group to an aromatic carbon at δ_c 156 ppm placed it in *ortho*-position of a hydroxy group and
615 next to a singlet aromatic proton. The singlet of the methoxy-group was connected to a carbonyl
616 group by HMBC correlations (**Supplementary Figure 9**).

617

618 **Construction of gene deletion mutants**

619 Targeted in-frame gene deletions of *F. perrara* were constructed using a two-step homologous
620 recombination procedure (see **Supplementary Figure 14**). In a first step, suicide plasmids were
621 constructed harboring two adjacent 700-800 bp-long homology regions matching the up- and
622 downstream region of the gene or gene cluster to be deleted. To this end, the two homology
623 regions were amplified using 2 Phanta Max Master Mix (Vazyme) and cloned into pKS2 using
624 Golden Gate assembly. pKS2 is a derivative of pSEVA312 (73) which contains a sfGFP
625 expression cassette flanked by two BsaI restriction sites in the multiple cloning site. All clonings
626 were done in *E. coli* DH5a/ λ pir. To validate successful cloning of the two homology regions,
627 PCR and Sanger sequencing was performed. The plasmids derived from pKS2 containing the
628 homology regions were shuttled into the diaminopimelic acid (DAP)-auxotroph *E. coli* JKE201
629 and delivered to *F. perrara* wt via conjugation using bi-parental mating. To this end, *E. coli*
630 JKE201 carrying the plasmid of interest was cultivated for 16-20 hours at 37 °C in LB with
631 500 μ M DAP (Sigma, LB-DAP) and 30 μ g/ml Chloramphenicol. Sub-cultures were made by
632 diluting the culture at a ratio of 1:20 into fresh medium and incubated until exponential growth
633 was reached. A fresh culture of *F. perrara* was prepared by growing it for 36 h on mTYG agar.
634 *F. perrara* and *E. coli* were harvested in PBS and the OD₆₀₀ adjusted to 10. Equal quantities of

both bacteria were mixed and spotted onto a cellulose acetate membrane filter (0.2 μ m, 25 mm, Huberlab, Sartorius) placed on BHI agar with 500 μ M DAP, followed by incubation at 35 °C for 6 hours under microaerophilic condition. Filters with bacterial mixtures were removed from the plates with a forceps, placed into a tube with PBS. Bacteria were removed from the filter by pipetting, vortexing and agitation for 10 min at 1500 rpm (Thermomixer C, Eppendorf). The bacterial suspension was centrifuged at 8000x g for 10 minutes, the supernatant discarded to remove dead bacteria and the bacterial pellet was washed once with 1000 μ l PBS before resuspending in a reduced volume, i.e. 100 μ l PBS. The bacterial mixture was plated onto CBA and TYG with selection antibiotic and incubated at 35 °C for 5 days under anaerobic condition. Colonies were picked after 4-5 days of incubation and expanded to fresh plates. DNA was isolated and five different PCRs performed to check for successful integration of the plasmid into the chromosomal region of *F. perrara* targeted for deletion. Positive clones (loop-in strains) were stocked in liquid mTYG containing 20% (v/v) glycerol at -80 °C until further usage. To select for bacteria that have lost the integrated plasmid, *F. perrara* loop-in strains were grown on mTYG agar and subsequently transformed with plasmid pYE1. pYE1 is a derivative of pBZ485 (74) containing the restriction enzyme *I-SceI* under the control of the IPTG-inducible *lac* promoter. The activity of the restriction enzyme *I-SceI* will negatively select bacteria containing pKS2-derivatives as two *I-SceI* sites are flanking the multiple cloning sites on pKS2. To generate electrocompetent *F. perrara* cells, the bacteria were cultivated from glycerol stocks on mTYG agar and incubated at 35 °C for 2 days under anaerobic condition. Bacteria from plate were inoculated in BHI broth with a starting OD₆₀₀ of 0.1 and grown anaerobically at 40 °C for 12-16 hours. Subsequently, growth was stopped by placing cultures onto ice for 15 minutes. Cells were made electrocompetent by washing twice with ice-cold MOPS solution supplemented with 20% glycerol, with decreasing volumes per wash. Electrocompetent cells were mixed with >500 ng plasmid DNA and incubated for 15 minutes on ice before transferring into an electroporation cuvette. Electroporation was performed applying a voltage of 2.5 kV with resistance of 200 Ω and capacitance of 25 μ F. Bacteria were immediately resuspended in BHI and incubated anaerobically for 6 hours at 35 °C to allow phenotypic expression. Bacteria were plated on CBA,

mTYG and/or BHIA plates with 25 µg/ml Kanamycin and 100 µM IPTG to select for pYE1 and induce the restriction enzyme *I-SceI*. Plates were incubated in anaerobic atmosphere at 35 °C. After 3 to 5 days of incubation, bacterial colonies were replica plated onto fresh plates containing 15 µg/ml chloramphenicol (selection marker of pKS2-derivatives) or 25 µg/ml kanamycin (selection marker of pYE1) to identify clones that cannot grow on chloramphenicol anymore as a consequence of the loss of the integrated pKS2-derivative. DNA was extracted and a PCR screen performed with primers amplifying the gene targeted for deletion as well as the overspanning region of the deletion. All clones possessing the gene deletion were subsequently expanded on mTYG agar and stocked in glycerol as described above

672

673 **Genome re-sequencing**

All gene deletion strains and the three white variants of *F. perrara* PEB091 were subjected to Illumina genome re-sequencing at Novogene (2x150bp). Mutations in the re-sequenced genomes relative to the genome of the wt strain of *F. perrara* were analyzed with *breseq* (75). The detailed results of this analysis can be found: <https://drive.switch.ch/index.php/s/kCNTp4g7n60ffMi>.

679

680 **Statistical analysis**

Statistical analysis was performed using GraphPad Prism v6.01 or R. Results of all statistical tests can be found in **Supplementary Dataset 5**. Normality was analyzed using D'Agostino-Pearson and/or Shapiro-Wilk test. For comparison with two conditions, t-test was used for parametric data and Wilcoxon sum-rank test or Kolgomorov-Smirnov test for non-parametric data. For comparison of the growth curves of *F. perrara* wt and *ihfA*^{*}, we used a permutation test 'compareGrowthCruves' included in the statmod package of R (76).

687

688 **Data availability**

689 RNA sequence datasets are available under NCBI Gene Expression Omnibus ID GSE189728.
 690 Code, scripts and numeric data files of experimental data have been deposited on Switchdrive:
 691 <https://drive.switch.ch/index.php/s/kCNTp4g7n60ffMi>.

692

693 **Acknowledgments**

694 We thank Stephan Gruber for input regarding the characterization of the ihfA* mutant strain and
 695 general feedback for the manuscript. We are grateful to Nicolas Neuschwander and Lucie
 696 Kesner for helping with some of the bee experiments. This project was funded by the ERC-StG
 697 ‘MicroBeeOme’ (grant agreement No. 714804), the Swiss National Science Foundation (grant
 698 agreement No. 31003A_160345 and 31003A_179487), the NCCR Microbiomes (all awarded to
 699 PE). J.P. acknowledges funding by the European Research Council (ERC) under the European
 700 Union’s Horizon 2020 Research and Innovation Program (grant agreement No. 742739).

701

Figures

Figure 1. Isolation of a spontaneous *ihfA mutant of *F. perrara* displaying an altered colony morphotype.** (A) Colonies of *F. perrara* PEB0191 (wt) after 48h of growth on mTYG agar. Arrowhead points at a larger white colony in between many yellow colonies. The area in the white square is magnified. (B) Colony morphology of *F. perrara* wt and the isolated white *ihfA** mutant after growth on mTYG for 48h. (C) Protein sequence comparison of IhfA and IhfB of *F. perrara* wt and *E. coli*. The mutated lysine and proline residues in IhfA and IhfB are outlined with a green and orange rectangle and the amino acid change due to the point mutations indicated. Secondary structures are depicted above as ribbons (α-helix) and arrows (β-sheet) and are numbered according to their appearance in the protein. (D) Three-dimensional structure of *E. coli* IhfA/B heterocomplex with DNA (source protein databank NDB: PDT040). DNA is depicted in blue and IhfB in dark grey. IhfA is colored according to secondary structure: α-helix orange, β-sheet light blue and the rest in light grey. α-helices and β-sheets are numbered. The mutated Pro83 and Lys38 residues of *F. perrara* *ihfA* and the Pro82 residue of *ihfB* are marked with an orange and green circle, respectively.

Figure 3. *F. perrara* *ihfA mutant displays a colonization defect.** (A) Light microscopy pictures of pylorus region of bees colonized with *F. perrara* PEB0191wt or *ihfA** 10 days post colonization. (B) Quantification of scab phenotype of bees 5 and 10 days post colonization with n=18 and n=36 per treatment, respectively. (C) Quantification of colonization levels are measured by colony forming units (CFU) 5 (n=18) and 10 (n=36) days post colonization. Wilcoxon rank-sum test was used to assess significant differences. (D) Time-course experiment of bees colonized with *F. perrara* wt or *ihfA**. Colonization levels were measured by CFUs every second day until Day 10 and following at Day 14 and Day 21. n=12 per time point per treatment. Wilcoxon rank-sum test was used to assess significant differences per time point. Error bars represent median and interquartile range. Data from three independent experiments. * p<0.05, **p<0.01, ***p<0.001.

Figure 4. Differential gene expression between *F. perrara* wt and *ihfA mutant during *in vitro* growth.** (A) Chromosomal localization of all genes significantly differentially expressed (2-log fold change = |2|, Fisher's exact test p value <0.05, FDR of <0.05) between *F. perrara* wt and the *ihfA** mutant. Starting from outside, the first circle shows the scale of the genome representation of *F. perrara* in gray and white steps of 100 kb. The second and third circles (gray) depict the genes on the plus and minus strands of *F. perrara*. The fourth (beige) and fifth (light blue) circle depicts genes upregulated and downregulated in wt compared to *ihfA**. Genomic islands are highlighted by coloration. (B) Bar plot of the genes differentially expressed between *F. perrara* wt and *ihfA** with a log2-fold change >2 (Fisher's exact test p value <0.05, FDR of <0.05). (C) Comparison of the transcriptional profile of the genomic location encoding the colibactin biosynthetic gene cluster between *F. perrara* wt and the *ihfA** mutant. Transcripts per million were visualized using the Integrative Genome Browser (69). The colibactin operon is schematically depicted below (green arrows).

Figure 5. Gene expression of Ihf-regulated genes of *F. perrara* ten days post-inoculation of gnotobiotic honey bees. Transcripts per million were calculated for all replicates of the *in vitro* and the *in vivo* RNAseq experiments. For the *in vitro* experiment (Exp. 1), all three replicates of the wt and the *ihfA** mutant are shown. For the *in vivo* experiment (Exp. 2), the four replicates of the day 10 time-point and the *in vitro* reference condition are shown. Data for day 5 time-point in comparison to day 10 time-point is shown in **Supplementary Figure14**.

Figure 6. Colonization levels of T6SSs, Clb, Flp pili, and ape gene deletion mutants of *F. perrara*. Colonization levels were assessed 10 days after inoculation by counting colony-forming units (CFU) in dilutions of homogenized bee guts plated on BHI agar. Only the pylorus and ileum section of the gut were analyzed. Limit of detection (LOD) corresponds to the lowest colonization level detectable in our assay, i.e. points below the LOD correspond to bees for which no CFUs were detected. Statistically significant differences of the colonization levels of each mutant relative to the wt of *F. perrara* were determined using the Wilcoxon rank-sum test with BH correction. Bees were inoculated with an OD₆₀₀ of 0.01. Data come from two independent experiments. **Supplementary Figure 15** shows the data points by experiments. * p<0.05, **p<0.01, ***p<0.001. Filled circle colors indicate whether a scab was detected during dissection (green = scab; yellow = no scab).

Supplementary Figures

Supplementary Figure 1. In vitro characterization of *F. perrara ihfA.** (A) Growth curve of *F. perrara* PEB0191 wt and *ihfA** in liquid TYG. The optical density (OD600) was measured every hour for 14 hours (Permutation test, p = 0.097). (B) Quantification of single cell length of *F. perrara* wt PEB0191 and *ihfA** from the single cell microscopy experiment shown in C. Significance was tested using Kolgomorov-Smirnov test, stars depict significance: *** p<0.0001. (C) Single cell light microscopy of *F. perrara* wt and *ihfA**. The scale bar (5μm) is depicted in the right lower corner.

Supplementary Figure 2. Ms-Ms fragmentation spectrum of $m/z = 323.1647$ in extracts of *F. perrara* wt. Many of the prominent ions were reported from aryl polyenes by (42): $m/z = 121.00, 131.00, 145.00, 147.08, 171.08, 183.00, 197.08, 209.08, 223.08$. The spectrum was recorded at 30 kV collision energy.

Supplementary Figure 3. UV-spectrum of the extracts indicates isomerization of the aryl polyene. UV spectrum of the extracts indicates isomerization of the aryl polyene. **A** Total ion chromatogram of extracts of *F. perrara* wt. **B** Extracted ion chromatogram of $m/z = 323.1640$ Da (± 5 ppm). Storage led to the formation of an additional peak at 19.28 min, possibly by isomerization. **C** The UV-spectrum at 18.94 min has the maximum around 415 nm, similar to that reported by (Cimermanic et al., 2014) **D** The UV-spectrum at 19.25 min shows an additional maximum at 315 nm, characteristic for a *trans-cis* conversion of a double bond.

818

819 **Supplementary Figure 4.** ^1H NMR spectrum of enriched aryl polyene in $\text{DMSO}-d_6$ at 298 K.

820

821

822 **Supplementary Figure 5.** ^1H NMR spectrum of enriched aryl polyene in $\text{DMSO}-d_6$ at 298 K.

823 Enhanced view of the aromatic region and integrals of assigned peaks.

824

825

826 **Supplementary Figure 6.** HSQC spectrum of enriched aryl polyene in DMSO- δ_6 at 298 K.

827

828

829 **Supplementary Figure 7.** HSQC spectrum of enriched aryl polyene in DMSO- δ_6 at 298 K.

830 Enhanced view of the aromatic region.

831

832

833 **Supplementary Figure 8.** COSY spectrum of enriched aryl polyene in DMSO- δ_6 at 298 K.

834

835

836 **Supplementary Figure 9.** COSY spectrum of enriched aryl polyene in DMSO- δ_6 at 298 K.

837 Enhanced view of the aromatic region.

838

Supplementary Figure 10. HMBC spectrum of enriched aryl polyene in DMSO- d_6 at 298 K.

Supplementary Figure 11. HPLC-HRESIMS analysis of *F. perrara* and *E. coli* CFT073. At 18.85 min both extracts show strong absorbance at $\lambda = 420$ nm. The MS-MS fragmentation pattern of the ion m/z 323.1642 at that retention time shows a similar pattern in extracts from both strains.

Supplementary Figure 12. Quantification of *F. perrara* wt and *ihfA in pylorus and ileum.**

Mono-colonization of bees with *F. perrara* wt and *ihfA**. Colonization levels are measured by colony forming units (CFU) and quantitative PCR (qPCR) 10 days post colonization in the specified gut regions: pylorus and ileum. Wilcoxon rank-sum test was used to assess the statistical significance of differences. n=6.

Supplementary Figure 13. RNAseq comparison of *F. perrara* during *in vivo* colonization compared to growth *in vitro*. Significantly differentially expressed genes with a $|\log_2\text{-fold change}| > 2$ of *F. perrara* PEB0191 wt 5(**A**) and 10 (**B**) days post colonization in comparison to growth *in vitro*. Genes are colored according to their category. The experiment was performed in quadruplicates, all genes displayed were significantly differentially expressed with $p < 0.05$ and a FDR of < 0.05 (Exact test).

867

868

869 **Supplementary Figure 14.** Same graph as shown in Figure 7 but datapoints are labelled by
870 experiment.

871

Supplementary Figure 15. Gene expression of Ihf-regulated genes of *F. perrara* five days post-inoculation of gnotobiotic honey bees. Same as Figure 5, but only time point Day 5 and Day 10 are depicted in comparison to *F. perrara ihfA in vitro.**

Supplementary Figure 16. Scheme of the gene deletion strategy based on a two-step homologous recombination procedure. **A**, A non-replicative plasmid pKS2 integrates via homologous recombination of one of the two cloned 'homology regions' (HRs) upstream or downstream of the gene that is targeted for deletion. **B**, In a second step, a replicative plasmid harboring the restriction enzyme I-SceI is transformed. I-SceI targets corresponding recognition sites located on pKS2 resulting in the selection of either revertant (wt genotypes) or mutant genotypes that underwent a second homologous recombination event in the region targeted for deletion. PCR screening and replica plating on different selective allows to identify to correct clones.

Supplementary Tables

Supplementary Table 1. Chemical shifts of enriched aryl polyene in DMSO- δ_6 at 298 K. Key COSY and HMBC correlation of the enriched polyene.

No	δ_C [ppm]	δ_H [ppm], mult.	coupling constant	integral
1	166.7			
2	119.3	5.97, d	$J=15.1$ Hz	0.5
3	144.6	7.33, dd	$J=11.6$ Hz $J=15.1$ Hz	1.2
4	129.5	6.47, dd overlap	$J=11.6$ Hz $J=14.4$ Hz	1.0
5	141.5	6.82, dd overlap	$J=11.4$ Hz $J=14.5$ Hz	0.9
6				
7				
8				
9				
10				
11				
12				
13				
14				
15	128.7	7.22, d	$J=1.7$ Hz	1.0
16	124.4			
17	155.9			
18	114.6	6.73, d	$J=8.3$ Hz	1.2
19	125.2	7.13, dd	$J=1.9$ Hz $J=8.3$ Hz	1.1
20	15.6	2.11, s		2.8
21	50.9	3.66, s overlap		3.5

Supplementary Table 2. Strains used in this study.

Species	Strain	Genotype*	Mutated gene	Reference	Parent strain
<i>F. perrara</i>	PEB0191/ ESL0157	wild type (wt)	None	(31)	Primary isolate
<i>F. perrara</i>	ESL0158	P83L (CCA→CTA) in <i>ihfA</i>	Fpe_00769	This study	PEB0191
<i>F. perrara</i>	ESL0910	P82L (CCA→CTA) in <i>ihfB</i> ; Δ <i>apeO</i>	Fpe_00769; Fpe_01928	This study	PEB0191
<i>F. perrara</i>	ESL0922	L38S (TTA→TCA) in <i>ihfA</i> ; Δ <i>pilE</i> *	Fpe_00769; Fpe_00791*; Fpe_01328	This study	PEB0191
<i>F. perrara</i>	ESL0654	Δ <i>hcp1</i> *	Fpe_01742	This study	PEB0191
<i>F. perrara</i>	ESL0655	Δ <i>hcp2</i>	Fpe_01947	This study	PEB0191
<i>F. perrara</i>	ESL0656	Δ <i>hcp1</i> , Δ <i>hcp2</i>	Fpe_01742, Fpe_01947	This study	PEB0191
<i>F. perrara</i>	ESL0888	Δ <i>clbB</i> *	Fpe_00576	This study	PEB0191
<i>F. perrara</i>	ESL0921	Δ <i>pilE</i> *	Fpe_01328	This study	PEB0191
<i>F. perrara</i>	ESL0957	Δ <i>apeA-R</i> *	Fpe_01915- Fpe_01932	This study	PEB0191

*re-sequencing of the genomes revealed that some of the strains (indicated with an asterisk) carried an additional point mutation. However, these mutations were located in random genes and included synonymous as well as non-synonymous base substitutions unlikely to have an effect on the observed phenotypes. The genomic analysis of the re-sequenced strains can be found on the Switchdrive: <https://drive.switch.ch/index.php/s/kCNTp4g7n60ffMi>.

904 **Supplementary Table 3. Primers used for qPCR quantification**

Primer name	Target	Sequence 5' to 3'	Tm (°C) ^a	Amplicon size (bp)	# Gene loci per genome	Standard curve			Reference
						Primer Efficiency, R ²	Slope, Intercept	LOD Cq ^b (#copies)	
prPEN0013	Actin AB023025 A. mellifera	TGCCAACACTGT CCTTTCTG	58.4	156	-	1.896 = 89.6 %, 1.0	-3.6, 37.57	33.7 (10)	(78)
prPEN0014		AGAATTGACCCA CCAATCCA	56.4						
prLK-Frisch-042-F	F. perrara 16S rRNA gene	GGAAGTTATGTG TGGGATAAGC	60.1	185	4	1.946 = 94.6 %, 0.995	-3.45, 38.04	31.7 (100)	(18)
prLK-Frisch-043-R		CTATTCTCAGGTT GAGCCCG	60.5						

905 a - Melting temperatures were calculated with the online tool described in (Kibbe 2007)

906 b - LOD refers to the limit of detection of primer sets, expressed here as the lowest number of
 907 plasmid copies reliably detected by qPCR when standard curves were performed

908

References

1. Wu M, *et al.* (2015) Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut *Bacteroides*. *Science* 350(6256):aac5992.
2. Lee SM, *et al.* (2013) Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* 501(7467):426-429.
3. Townsend GE, *et al.* (2019) Dietary sugar silences a colonization factor in a mammalian gut symbiont. *Proceedings of the National Academy of Sciences* 116(1):233-238.
4. Townsend GE, *et al.* (2020) A Master Regulator of Bacteroides thetaiotaomicron Gut Colonization Controls Carbohydrate Utilization and an Alternative Protein Synthesis Factor. *mBio* 11(1):e03221-03219.
5. David LA, *et al.* (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505(7484):559-563.
6. Nakajima A, *et al.* (2018) IgA regulates the composition and metabolic function of gut microbiota by promoting symbiosis between bacteria. *Journal of Experimental Medicine* 215(8):2019-2034.
7. Powell JE, Leonard SP, Kwong WK, Engel P, & Moran NA (2016) Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. *Proceedings of the National Academy of Sciences of the United States of America*.
8. Ohbayashi T, *et al.* (2015) Insect's intestinal organ for symbiont sorting. *Proceedings of the National Academy of Sciences* 112(37):E5179-E5188.
9. Salem H, *et al.* (2017) Drastic Genome Reduction in an Herbivore's Pectinolytic Symbiont. *Cell* 171(7):1520-1531.e1513.
10. Kikuchi Y, Ohbayashi T, Jang S, & Mergaert P (2020) Burkholderia insecticola triggers midgut closure in the bean bug Riptortus pedestris to prevent secondary bacterial infections of midgut crypts. *The ISME journal*.
11. Kwong WK & Moran NA (2016) Gut microbial communities of social bees. *Nature reviews. Microbiology* 14(6):374-384.
12. Bonilla-Rosso G & Engel P (2018) Functional roles and metabolic niches in the honey bee gut microbiota. *Curr Opin Microbiol* 43:69-76.
13. Kwong WK, *et al.* (2017) Dynamic microbiome evolution in social bees. *Sci Adv* 3(3):e1600513.
14. Engel P, Bartlett KD, & Moran NA (2015) The Bacterium Frischella perrara Causes Scab Formation in the Gut of its Honeybee Host. *mBio* 6(3).
15. Martinson VG, *et al.* (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular ecology* 20(3):619-628.
16. Powell JE, Martinson VG, Urban-Mead K, & Moran NA (2014) Routes of Acquisition of the Gut Microbiota of the Honey Bee *Apis mellifera*. *Appl Environ Microbiol* 80(23):7378-7387.
17. Emery O, Schmidt K, & Engel P (2017) Immune system stimulation by the gut symbiont Frischella perrara in the honey bee (*Apis mellifera*). *Molecular ecology* 26(9):2576-2590.
18. Kesnerova L, *et al.* (2017) Disentangling metabolic functions of bacteria in the honey bee gut. *bioRxiv*.
19. Raymann K, Bobay LM, & Moran NA (2018) Antibiotics reduce genetic diversity of core species in the honeybee gut microbiome. *Mol Ecol* 27(8):2057-2066.
20. Zheng H, Powell JE, Steele MI, Dietrich C, & Moran NA (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):4775-4780.

21. Brochet S, *et al.* (2021) Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *Elife* 10.
22. Ellegaard KM, *et al.* (2019) Genomic changes underlying host specialization in the bee gut symbiont *Lactobacillus Firm5*. *Molecular ecology* 28(9):2224-2237.
23. Ellegaard KM & Engel P (2019) Genomic diversity landscape of the honey bee gut microbiota. *Nature communications* 10(1):446.
24. Engel P, Martinson VG, & Moran NA (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):11002-11007.
25. Kwong WK, Engel P, Koch H, & Moran NA (2014) Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proceedings of the National Academy of Sciences of the United States of America* 111(31):11509-11514.
26. Kwong WK, *et al.* (2017) Dynamic microbiome evolution in social bees. *Science Advances* 3(3).
27. Ludvigsen J, Porcellato D, Amdam GV, & Rudi K (2018) Addressing the diversity of the honeybee gut symbiont *Gilliamella*: description of *Gilliamella apis* sp. nov., isolated from the gut of honeybees (*Apis mellifera*). *Int J Syst Evol Microbiol* 68(5):1762-1770.
28. Steele MI, Kwong WK, Whiteley M, & Moran NA (2017) Diversification of Type VI Secretion System Toxins Reveals Ancient Antagonism among Bee Gut Microbes. *mBio* 8(6).
29. Zheng H, *et al.* (2019) Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proceedings of the National Academy of Sciences* 116(51):25909-25916.
30. Volkmann M, *et al.* (2010) *Orbus hercynius* gen. nov., sp. nov., isolated from faeces of wild boar, is most closely related to members of the orders 'Enterobacteriales' and Pasteurellales. *Int J Syst Evol Microbiol* 60(Pt 11):2601-2605.
31. Engel P, Kwong WK, & Moran NA (2013) *Frischella perrara* gen. nov., sp. nov., a gammaproteobacterium isolated from the gut of the honeybee, *Apis mellifera*. *International Journal of Systematic and Evolutionary Microbiology* 63(10):3646-3651.
32. Martinson VG, Moy J, & Moran NA (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl Environ Microbiol* 78(8):2830-2840.
33. Leonard SP, *et al.* (2018) Genetic Engineering of Bee Gut Microbiome Bacteria with a Toolkit for Modular Assembly of Broad-Host-Range Plasmids. *ACS Synth Biol* 7(5):1279-1290.
34. Engel P, Bartlett KD, & Moran NA (2015) The Bacterium *Frischella perrara* Causes Scab Formation in the Gut of its Honeybee Host. *MBio* 6(3):e00193-00115.
35. Engel P, Vizcaino MI, & Crawford JM (2015) Gut symbionts from distinct hosts exhibit genotoxic activity via divergent colibactin biosynthesis pathways. *Appl Environ Microbiol* 81(4):1502-1512.
36. Friedman DI (1988) Integration host factor: A protein for all reasons. *Cell* 55(4):545-554.
37. Freundlich M, Ramani N, Mathew E, Sirko A, & Tsui P (1992) The role of integration host factor in gene expression in *Escherichia coli*. *Molecular microbiology* 6(18):2557-2563.
38. Goosen N & van de Putte P (1995) The regulation of transcription initiation by integration host factor. *Molecular microbiology* 16(1):1-7.
39. Rice PA, Yang S-w, Mizuuchi K, & Nash HA (1996) Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn. *Cell* 87(7):1295-1306.
40. Goel AK, Rajagopal L, Nagesh N, & Sonti RV (2002) Genetic Locus Encoding Functions Involved in Biosynthesis and Outer Membrane Localization of Xanthomonadin in *Xanthomonas oryzae* pv. *oryzae*. *J Bacteriol* 184(13):3539-3548.
41. Poplawsky AR, Urban SC, & Chun W (2000) Biological role of xanthomonadin pigments in *Xanthomonas campestris* pv. *campestris*. *Appl Environ Microbiol* 66(12):5123-5127.

42. Schoner TA, *et al.* (2016) Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products, Are Functionally Related to Antioxidative Carotenoids. *Chembiochem* 17(3):247-253.
43. Cimermancic P, *et al.* (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158(2):412-421.
44. Lee LJ, Barrett JA, & Poole RK (2005) Genome-wide transcriptional response of chemostat-cultured *Escherichia coli* to zinc. *J Bacteriol* 187(3):1124-1134.
45. Hantke K (2001) Iron and metal regulation in bacteria. *Current opinion in microbiology* 4(2):172-177.
46. Zulianello L, van Ulsen P, van de Putte P, & Goosen N (1995) Participation of the Flank Regions of the Integration Host Factor Protein in the Specificity and Stability of DNA Binding. *Journal of Biological Chemistry* 270(30):17902-17907.
47. Jurcisek JA, Brockman KL, Novotny LA, Goodman SD, & Bakaletz LO (2017) Nontypeable *Haemophilus influenzae* releases DNA and DNABII proteins via a T4SS-like complex and ComE of the type IV pilus machinery. *Proc Natl Acad Sci U S A* 114(32):E6632-E6641.
48. Utada AS, *et al.* (2014) *Vibrio cholerae* use pili and flagella synergistically to effect motility switching and conditional surface attachment. *Nat Commun* 5:4913.
49. Tomich M, Planet PJ, & Figurski DH (2007) The tad locus: postcards from the widespread colonization island. *Nature reviews. Microbiology* 5(5):363-375.
50. Shime-Hattori A, *et al.* (2006) Two type IV pili of *Vibrio parahaemolyticus* play different roles in biofilm formation. *FEMS Microbiol Lett* 264(1):89-97.
51. Craig L, Pique ME, & Tainer JA (2004) Type IV pilus structure and bacterial pathogenicity. *Nature reviews. Microbiology* 2(5):363-378.
52. Johnston I, *et al.* (2021) Identification of essential genes for *Escherichia coli* aryl polyene biosynthesis and function in biofilm formation. *NPJ Biofilms Microbiomes* 7(1):56.
53. Kuraishi T, Hori A, & Kurata S (2013) Host-microbe interactions in the gut of *Drosophila melanogaster*. *Front Physiol* 4:375.
54. Ha EM, Oh CT, Bae YS, & Lee WJ (2005) A direct role for dual oxidase in *Drosophila* gut immunity. *Science* 310(5749):847-850.
55. Nougayrede JP, *et al.* (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* 313(5788):848-851.
56. Cuevas-Ramos G, *et al.* (2010) *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A* 107(25):11537-11542.
57. Arthur JC, *et al.* (2012) Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 338(6103):120-123.
58. Mougous JD, *et al.* (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* 312(5779):1526-1530.
59. Coyne MJ & Comstock LE (2019) Type VI Secretion Systems and the Gut Microbiota. *Microbiol Spectr* 7(2).
60. Coulthurst S (2019) The Type VI secretion system: a versatile bacterial weapon. *Microbiology (Reading)* 165(5):503-515.
61. Monjaras Fera J & Valvano MA (2020) An Overview of Anti-Eukaryotic T6SS Effectors. *Front Cell Infect Microbiol* 10:584751.
62. Steele MI & Moran NA (2021) Evolution of Interbacterial Antagonism in Bee Gut Microbiota Reflects Host and Symbiont Diversification. *mSystems* 6(3).
63. Powell E, Ratnayeke N, & Moran NA (2016) Strain diversity and host specificity in a specialized gut symbiont of honeybees and bumblebees. *Mol Ecol* 25(18):4461-4471.
64. Pan J, *et al.* (2018) Integration Host Factor Modulates the Expression and Function of T6SS2 in *Vibrio fluvialis*. *Frontiers in microbiology* 9(962).
65. Chen X, Yu C, Li S, Li X, & Liu Q (2018) Integration Host Factor Is Essential for Biofilm Formation, Extracellular Enzyme, Zeamine Production, and Virulence in *Dickeya zeae*. *Molecular Plant-Microbe Interactions*® 32(3):325-335.

66. Nikolayeva O & Robinson MD (2014) edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. *Methods Mol Biol* 1150:45-79.
67. Robinson MD & Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11(3):R25.
68. Wagner GP, Kin K, & Lynch VJ (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci Theor Den Biowissenschaften* 131.
69. Freese NH, Norris DC, & Loraine AE (2016) Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32(14):2089-2095.
70. Kesnerova L, *et al.* (2019) Gut microbiota structure differs between honeybees in winter and summer. *The ISME journal*.
71. Schneider CA, Rasband WS, & Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671-675.
72. Ducret A, Quardokus EM, & Brun YV (2016) MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nat Microbiol* 1(7):16077.
73. Martinez-Garcia E, *et al.* (2020) SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic acids research* 48(6):3395.
74. Harms A, *et al.* (2017) A bacterial toxin-antitoxin module is the origin of inter-bacterial and inter-kingdom effectors of Bartonella. *PLoS Genet* 13(10):e1007077.
75. Deatherage DE & Barrick JE (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol* 1151:165-188.
76. Elso CM, *et al.* (2004) Leishmaniasis host response loci (Imr1-3) modify disease severity through a Th1/Th2-independent pathway. *Genes Immun* 5(2):93-100.
77. Guy L, Kultima JR, & Andersson SG (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26(18):2334-2335.
78. Zufelato MS, Lourenço AP, Simões ZL, Jorge JA, & Bitondi MM (2004) Phenoloxidase activity in Apis mellifera honey bee pupae, and ecdysteroid-dependent expression of the prophenoloxidase mRNA. *Insect biochemistry and molecular biology* 34(12):1257-1268.