environmental microbiology

Environmental Microbiology (2019) **21**(11), 4253–4269 doi:10.1111/1462-2920.14786

The microbiome of the *Melitaea cinxia* butterfly shows marked variation but is only little explained by the traits of the butterfly or its host plant

Guillaume Minard,[1,2,3*†] **Gleb Tikhonov**,[1†] **Otso Ovaskainen**[1,4] **and Marjo Saastamoinen**[1,5]

[1]*Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland.*

[2]*Université de Lyon, Lyon, France.*

[3]*Ecologie Microbienne, UMR CNRS 5557, UMR INRA 1418, VetAgro Sup, Université Lyon 1, Villeurbanne, France.*

[4]*Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, N-7491, Trondheim, Norway.*

[5]*Helsinki Institute of Life Sciences, University of Helsinki, Helsinki, Finland.*

Summary

Understanding of the ecological factors that shape intraspecific variation of insect microbiota in natural populations is relatively poor. In Lepidopteran caterpillars, microbiota is assumed to be mainly composed of transient bacterial symbionts acquired from the host plant. We sampled Glanville fritillary (*Melitaea cinxia*) caterpillars from natural populations to describe their gut microbiome and to identify potential ecological factors that determine its structure. Our results demonstrate high variability of microbiota composition even among caterpillars that shared the same host plant individual and most likely the same genetic background. We observed that the caterpillars harboured microbial classes that varied among individuals and alternated between two distinct communities (one composed of mainly Enterobacteriaceae and another with more variable microbiota community). Even though the general structure of the microbiota was not attributed to the measured ecological factors, we found that phylogenetically similar microbiota showed corresponding responses to the sex and the parasitoid infection of the caterpillar and to those of the host plant's microbial and chemical composition. Our results indicate high among-individual variability in the microbiota of the *M. cinxia* caterpillar and contradict previous findings that the host plant is the major driver of the microbiota communities of insect herbivores.

Received 11 July, 2018; revised 14 August, 2019; accepted 20 August, 2019. *For correspondence. E-mail guillaume.minard@univ-lyon1.fr; Tel. +33 472431143. †These authors contributed equally to this work.

Introduction

All animals interact with microorganisms (McFall-Ngai *et al*., 2013), with interactions between hosts and their microbes ranging from mutualistic to competitive (Douglas, 2010). Insects harbour highly diversified host–symbiont interactions with various examples of fitness benefits (Douglas, 2011), such as the control of the host's reproduction (Werren *et al*., 2008; Engelstädter and Hurst, 2009), the enhancement of nutrition via effects on the digestion process (Warnecke *et al*., 2007), the degrading of toxic metabolites (Kikuchi *et al*., 2012; Ceja-Navarro *et al*., 2015), and the production of nutrients essential for the host (Akman Gunduz and Douglas, 2009; Salem *et al*., 2014). Endosymbionts can also protect their hosts against abiotic stressors and pathogens (Montllor *et al*., 2002; Dunbar *et al*., 2007; King *et al*., 2016). The literature may, however, be biased towards mutualistic and parasitic/pathogenic interactions, as commensal or neutral interactions may be understudied or underreported (reviewed by Hammer *et al*., 2019). In general, the microbiota is a multilayer system in which prevalent members compose the core microbiota and a more flexible pool of microbial members compose the non-core community (Shapira, 2016).

Host–microbiota interactions are often complex, involve multiple taxa and multiple transmission processes, and consequently laboratory-based studies may fail to realistically portray natural systems. Indeed, several studies have highlighted pronounced differences in the microbiota of laboratory-reared versus field-captured individuals (Rani *et al*., 2009; Staubach *et al*., 2013; Tinker and Ottesen, 2016). Characterizing and determining the impact of microbiota in natural populations remain challenging, due to the multiple confounding factors that can affect the microbiota composition. Consequently, we still know little of the ecological factors that shape among-

individual variation of microbial communities in natural populations. Another challenge is related to the data analyses: microbiota data typically include large numbers of taxonomical units, most of which are rare, complicating the use of conventional statistical frameworks.

The gut microbiota of insects is often highly heterogeneous both among species and among individuals within single species, with relatively high variation reported even across different gut sections (Douglas, 2015). The consumed diet has been suggested to be the major determinant of the microbiota composition, as it can shape the microbial communities both directly (e.g. acquisition of food-associated microorganisms or growth of microorganisms that utilize the consumed food) and indirectly (e.g. through impacts on immunity, anatomy or digestive function; Douglas, 2015). However, several studies that have controlled for the transient effects of diet (e.g. in fruit flies and Asian tiger mosquitoes), still report strong inter-individual variation in the microbiota composition (Minard *et al*., 2015; Adair *et al*., 2018), suggesting the importance of diet-unrelated factors. Gut microbiota can, for example, be acquired via maternal or horizontal transmission (Engel and Moran, 2013), influenced by host genotype or environmental conditions unrelated to food (Yun *et al*., 2014), or be driven mainly by stochastic processes (Douglas, 2015; Zeng *et al*., 2015). In Lepidoptera, there is only little evidence on the transfer of symbiotic bacteria among individuals (Paniagua Voirol *et al*., 2018). Consistently, the Lepidopteran gut microbiome has been shown to be highly variable compared with other insect orders, with only few resident bacteria (Hammer *et al*., 2017). The importance of the gut microbiota on the performance of Lepidoptera has also been studied, even though the general knowledge on the bacterial associations across species is still very limited (see Paniagua Voirol *et al*., 2018 for a review).

To improve our understanding of the potential ecological determinants influencing associations between insect hosts and their gut symbionts, we exploit here the natural metapopulation of the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland islands, Finland. With *M. cinxia* caterpillars and their *Plantago lanceolata* host plants sampled across this system at a single timepoint, our overall aim is to associate the midgut microbiota of the caterpillars with ecological variables, and thus to identify potential drivers of variation that could impact these communities. In particular, we ask (i) what is the composition of *M. cinxia* microbiota and that of its host plant *P. lanceolata* (ii); is there a correspondence between the host plant microbiota and that of the caterpillar microbiota; (iii) is the host plant microbiota and the caterpillar microbiota influenced by the metabolite profile of the host plant; (iv) are the caterpillar microbial communities structured according to the sex and parasitoid infection status of the host; (v) after accounting for the above mentioned factors, is the variation in the microbiota communities structured by the caterpillars living in the same family on the same host plant individual or is it idiosyncratic among individuals independent of the family structure; and (vi) is the variation in the microbiota with respect to the questions i–v phylogenetically structured. Furthermore, to examine if and how microbial variation influences the fitness of the host, we ask (vii) whether the over-winter survival of caterpillar nests can be explained by their microbiota composition. To address these questions, we apply a joint species distribution model (Ovaskainen *et al*., 2017) to evaluate both species- and community-level responses to the abovementioned covariates, as well as residual co-occurrence patterns of the microbiota both at the levels of individual caterpillars and caterpillar families.

Results

Factors influencing caterpillar microbiota

Overall, the caterpillar microbiota was composed of variable microbiota among which the dominant taxa (>1% of the relative abundance across all samples) were *Uruburuella* (Proteobacteria, Betaproteobacteria), *Cloacibacterium* (Bacteroidetes, Flavobacteriia), *Moraxella* (Proteobacteria, Gammaproteobacteria), *Acinetobacter* (Proteobacteria, Gammaproteobacteria), *Dermacoccus* (Actinobacteria), *Hymenobacter* (Bacteroidetes, Cytophagia), *Corynebacterium* (Actinobacteria), *Paracoccus* (Proteobacteria, Alphaproteobacteria), *Wolbachia* (Proteobacteria, Alphaproteobacteria), *Methylobacterium* (Proteobacteria, Alphaproteobacteria), and unclassified Actinobacteria, Enterobacteriaceae (Proteobacteria, Gammaproteobacteria) and Corynebacteriaceae (Actinobacteria) (Figs 1A and 2). *Uruburuella* was most prevalent but still detected in only 58.8% of the samples, suggesting that there is either (i) a no core microbiota or (ii) that there is a core microbiota but it is not dominant across all individuals. To investigate the potential ecological factors explaining variation in the occurrences and abundances of microbial taxa, we used joint-species modelling framework.

Our model on the occurrence of the operational taxonomic units (OTUs) in the larvae had only little predictive power through its fixed effects (Prediction P1; Table 1). In line with this, model's fixed factors (caterpillar sex, parasitoid infection status and the host plant's bacterial and metabolic composition) did not show a community-consistent correlation with the occurrence patterns of the bacterial community ([5%, 95%] credibility interval for community-level mean value of species response overlapped with zero Table S1). Accounting for the residual species-to-species associations substantially increased the predictive power of the model (Table 1), meaning that

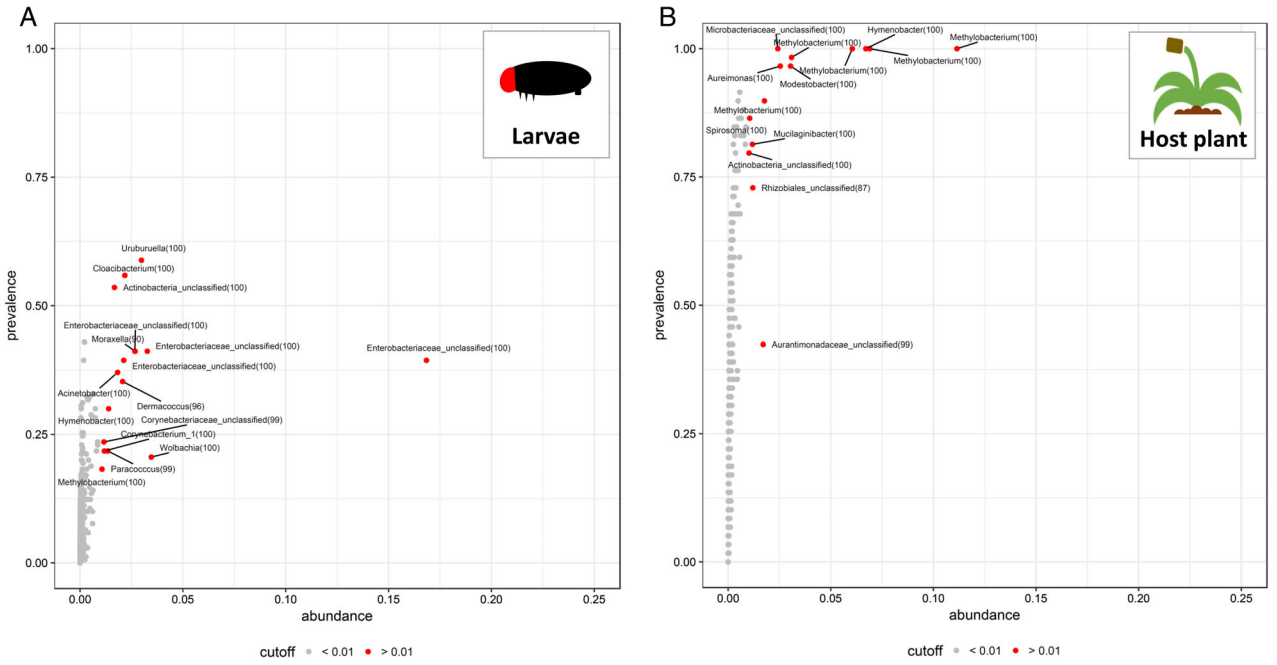


Fig. 1. Prevalence and average abundance of the bacteria within caterpillar midguts (A) and plant leaves (B). The OTUs (dots) are represented according to the proportion of individual samples in which they were detected (prevalence) and their average relative abundance across all the samples (abundance). The classification of the most abundant OTUs (cutoff >0.1) is provided. The bootstrap associated with each taxonomical classification is reported in brackets.

the bacteria show substantial residual co-occurrence patterns across the individuals. The same fixed factors and the co-occurrence between bacterial OTUs explained roughly equal amount of the variation ($R^2 = 0.22 \pm 0.17$ and 0.36 ± 0.22 respectively) in the model for OTU abundances (Table 1). This means that the ecological covariates and the bacterial co-occurrence patterns have approximately equivalent contribution to the variation of the bacterial communities associated with *M. cinxia*. As none of the fixed effects had a consistent correlation with the OTU abundance patterns, their impacts are taxon-specific. Comparisons based on variance partitioning among the explanatory factors showed consistent results to the above-presented comparisons based on predicted

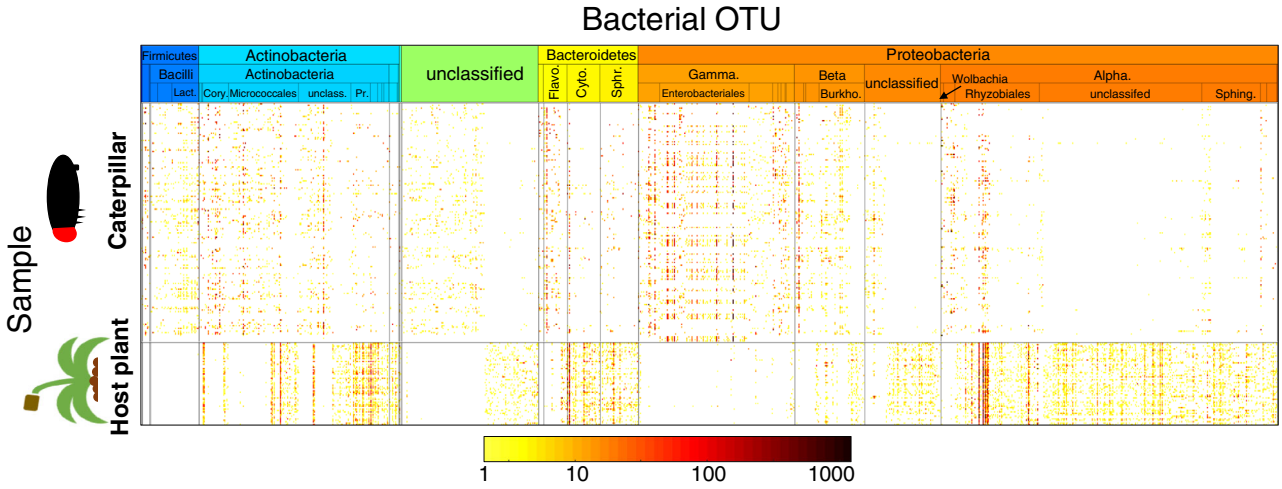


Fig. 2. Abundances of bacterial OTUs in caterpillar and plant samples. The OTUs (columns) have been ordered according to their taxonomical classification (for details, see Table S2). The colour scale shows OTU abundance (number of normalized sequences) for each caterpillar and plant sample on a logarithmic scale, and white colour indicates absence of OTU in given sample.

Table 1. Predictive powers of the larval and plant models.

Model	Prediction	R^2 (mean ± SD)
Caterpillar model		
Presence–absence	P1	0.017 ± 0.014
	P2	0.12 ± 0.10
	P3	0.14 ± 0.13
Abundance	P1	0.22 ± 0.17
	P2	0.36 ± 0.22
	P3	0.42 ± 0.23
Plant model		
Presence–absence	P1	0.024 ± 0.020
	P2	0.075 ± 0.077
	P3	0.083 ± 0.093
Abundance	P1	0.11 ± 0.13
	P2	0.22 ± 0.19
	P3	0.23 ± 0.20

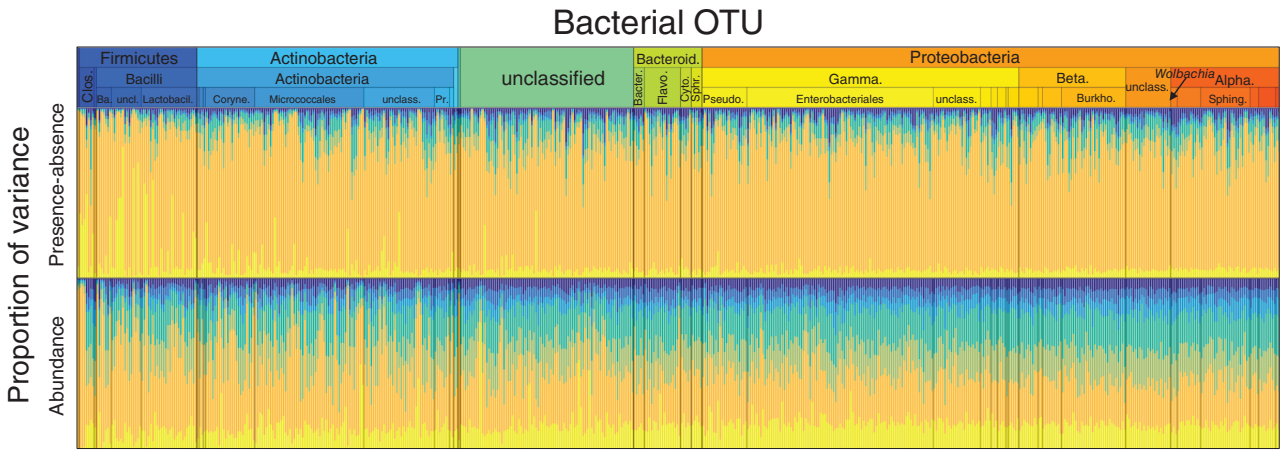
Predictive power is measured by Tjur R^2 for the occurrence models and by the standard R^2 for the abundance models. The values show the mean ± SD over the OTUs. As detailed in the *Statistical Methods*, Prediction P1 measures the predictive power solely due to the fixed effects part of the models, whereas P2 and P3 also account for species-to-species associations, with P2 being based on cross-validation across species and P3 on fitted model's self-explanatory predictive power.

power, as more variance in the occurrence (72%) than abundance (33%) of OTUs was attributed to the random effect of the individual caterpillar (Fig. 3).

Despite the overall community structure not being affected by the ecological factors assessed (Fig. 3), some taxa did show responses to the fixed effects (Fig. 4). Specifically, the occurrence probabilities of some of the OTUs decreased with the presence of the parasitoid infection (mostly Clostridia, Alphaproteobacteria and Betaproteobacteria), and were lower in males than in females (mostly Alphaproteobacteria: Rhodobacterales and Betaproteobacteria: Neisseriales). The presence of *Wolbachia*, on the other hand, was positively associated with the parasitoid infection of the caterpillars. Only a minority of the OTUs, classified as *Hymenobacter* and *Methylobacterium*, showed increased occurrence probability in the caterpillar with the increased abundance of the same OTU in the host plant. In the abundance model, very few individual OTUs' responded to the fixed effects with a high level of statistical support (Fig. 4).

Both our models had a very high phylogenetic signal (estimated posterior of phylogenic strength was 0.98 $\pm$ 0.002 and 0.86 $\pm$ 0.025), suggesting that related bacteria share similar niches and responded similarly to the fixed effects. This result is prominently visible in Fig. 4, which represents the responses of bacterial taxa ordered by taxonomy, and where the positive and negative effects (the red and blue colours respectively) are presented clearly as contiguous blocks rather than randomly distributed across the OTUs. The majority of the Betaproteobacteria and Burkolderiales (Alphaproteobacteria), for example, have lower occurrence probability in individuals infected by the parasitoid. Similarly, the majority of the Corinebacteriales (Actinobacteria) show correlated occurrence with the bacterial composition of the host plant. The occurrence of the microbial OTUs was phylogenetically structured not only with respect to the measured covariates, but also in their residual variation, as the OTUs split into two groups in a markedly pronounced manner (Fig. 5A). One of these two groups consisted of microbiota dominated by, with minor exceptions, the Enterobacteriaceae family. The other group, on the other hand, harboured microbiota that consisted of multiple taxa including *Uruburella*, *Cloacibacterium*, *Moraxella*, *Acinetobacter*, *Dermacoccus*, *Hymenobacter*, *Corynebacterium* and *Paracoccus*. Thus, some of the caterpillars were characterized by a high representation of Enterobacteriaceae in their microbiota, while the remaining individuals were characterized by a low representation of Enterobacteriaceae. Given its dominant role in variance partitioning, this pattern is the strongest signal related to OTU occurrences in our data (Fig. 5A), and its validity is supported by similar results of a complementary analysis based on Dirichlet mixture modelling (Fig. S1AC). None of the fixed factors assessed (sex, host plant and parasitoid), however, explained the occurrence of these two distinct microbial communities of phylogenetically related bacteria. In contrast to the strong patterns recorded in the occurrence model, only few statistically supported associations were found in the abundance model (Fig. 5B). The fact that caterpillars belonging to the same family nest and that were collected on the same host plant individual did not share similar microbial communities was somewhat unexpected. This was evident in both for the occurrence and abundance models, where host plant attributed only a minor proportion of the variance (Plant level in Fig. 3) and almost no statistically supported residual associations were found.

To summarize, the bacterial community of the caterpillars exhibited a complex structure, with a highly variable bacterial taxa that also showed marked among-individual variation. In terms of variation among caterpillar individuals, we found that neither the fixed effects (sex, host plant and parasitoid) assessed nor the family relationships (individuals collected from the same family nest on the same host plant) were capable to explain the very strong segregation of individuals into two groups with very distinct microbiota composition: about 40% of the individuals were characterized by a microbiota with a co-occurrence of phylogenetically related Enterobacteriaceae, whereas the rest of the individuals were characterized by a more complex microbiota, composed of *Uruburella*, *Cloacibacterium*, *Moraxella*, *Acinetobacter*, *Dermacoccus*, *Hymenobacter*, *Corynebacterium*, *Paracoccus*, *Wolbachia*, *Methylobacterium*, and some unclassified Actinobacteria and Corynebacteriaceae. Although, the fixed effects included in our model accounted for minor part of this variation, and the microbial OTU responses to the fixed effects were not synchronized across whole community, we found that phylogenetically similar OTUs responded to these effects in similar manner.

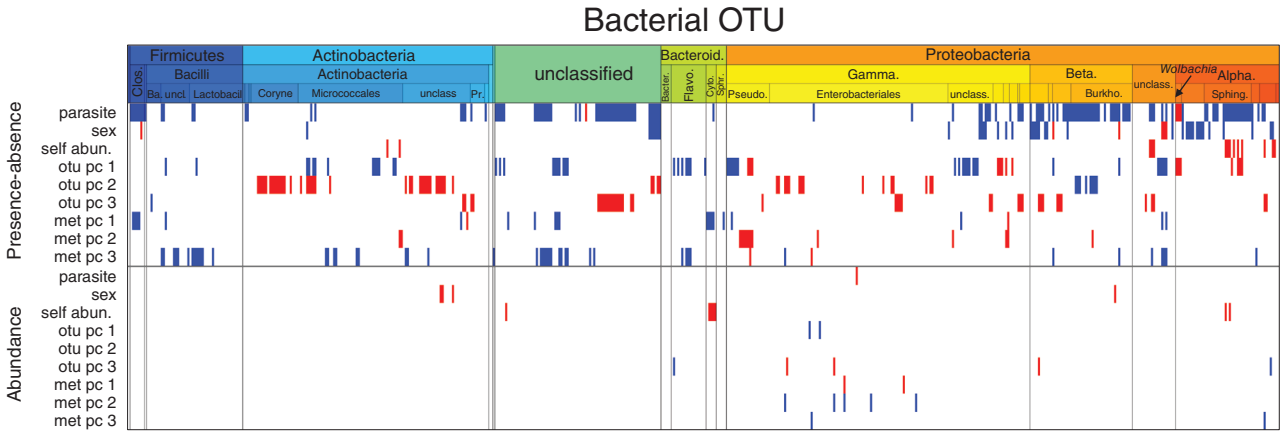


Variable	Type	Color	P-A (%)	Ab (%)
Presence of the parasitoid *Hyposoter horticola*	Fixed effect		3.5	5.6
The sex of the caterpillar	Fixed effect		2.4	5.9
Focal OTU abundance in the plant	Fixed effect		1.0	6.0
Plant OTU composition (PC1,PC2,PC3)	Fixed effect		7.7	18
Plant metabolic composition (PC1,PC2,PC3)	Fixed effect		6.7	17
Caterpillar level	Random effect		72	33
Plant level	Random effect		6.5	14

Fig. 3. Partitioning of variation in caterpillar microbiota to components explained by different types of fixed and random effects. The coloured bars show, for each OTU, the proportions of variance attributed to each group of explanatory variables. The average variance proportions over OTUs are shown in the legend, with P-A corresponding to the occurrence and Ab to the abundance model. The ordering of OTUs follows the ordering of Fig. 2 except for OTUs that were recorded only in plant samples and are omitted here (for details, see Table S2). See *Statistical Methods* for a full description of the included fixed and random effects.

Factors influencing host plant foliar microbiota

Contrary to the microbiota within the caterpillar gut, the plant microbiota was composed of highly prevalent bacterial taxa (detected in more than 90% of the collected samples; Fig. 1B). The bacteria in this core microbiota were assigned to *Methylobacterium* (Proteobacteria, Alphaproteobacteria), *Hymenobacter* (Bacteroidetes, Cytophagia), *Aureimonas*



Variable	Running name
Presence of the parasitoid *Hyposoterhorticola*	parasite
The sex of the caterpillar	sex
Focal OTU abundance in the plant	self abun.
Plant OTU composition (PC1,PC2,PC3)	otu pc 1-3
Plant metabolic composition (PC1,PC2,PC3)	met pc 1-3

Fig. 4. Influence of measured covariates on caterpillar microbiota. Regression coefficients that were estimated to be positive (respectively, negative) with 95% credibility level are shown in red (respectively, blue). The ordering of OTUs is identical to that of Fig. 3. The covariates included in the model are listed in the legend alongside with their running names used in axis labelling.

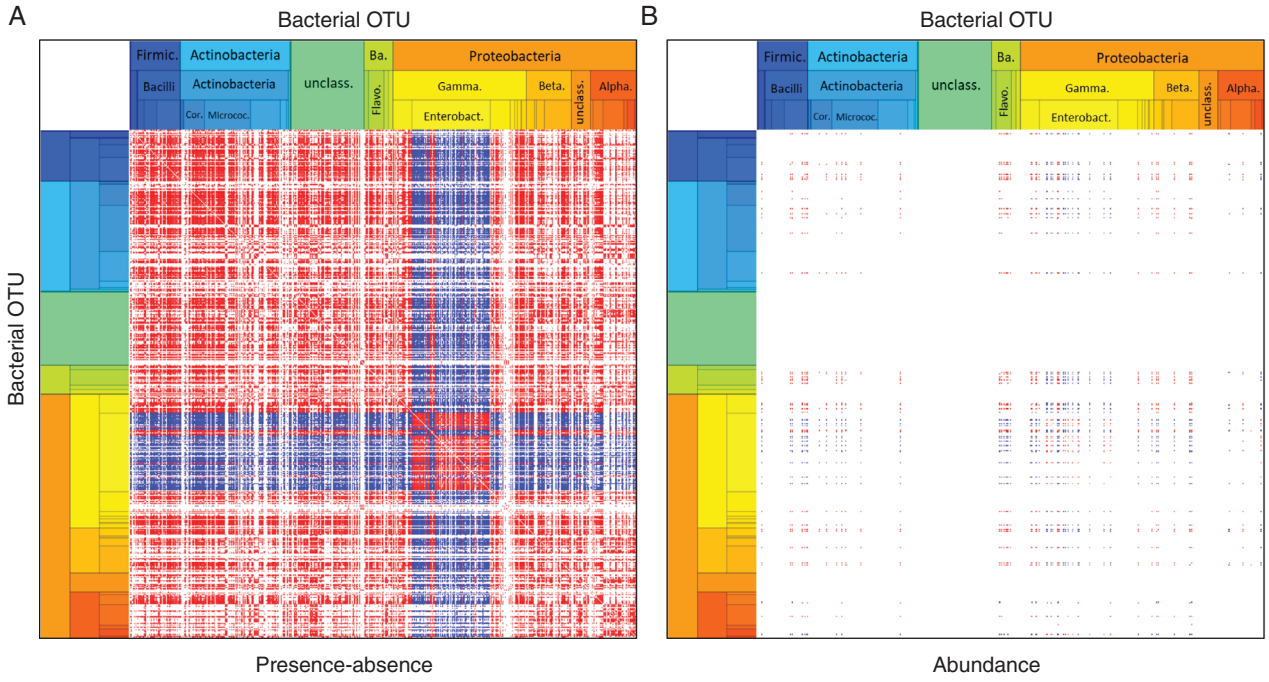


Fig. 5. Residual associations among caterpillar microbiota. The panels illustrate the caterpillar-level random effects for the occurrence (A) and abundance (B) parts of the caterpillar model. OTU pairs for which the residual correlation was estimated to be positive (respectively, negative) with 95% credibility level are shown in red (respectively, blue) colour. The ordering of OTUs is identical to that of Fig. 3.

(Proteobacteria, Alphaproteobacteria), *Modestobacter* (Actinobacteria) and an unclassified Microbacteriaceae (Actinobacteria). Whenever possible, we also assessed the plant metabolome by ^{1}H-NMR spectrometry (Fig. 6). The identified metabolites included amino acids (valine, threonine, alanine, arginine, glutamate and glutamine), sugars (xylose, α-glucose, β-glucose and sucrose), organic acids (fumaric acid, acetic acid and cis-aconic acid), ethanol and defensive metabolites (aucubin, catalpol and verbascosides). The latter included both terpenoids (aucubin, catalpol) and phenolic compounds (verbascosides), which constitute the main chemical defence of *Plantago lanceolata* against herbivores and pathogenic microorganisms. Most of the variation in the plant metabolites across samples was explained by unannotated metabolite signals. The most part of the variation was due to unannotated carbohydrates and amino acid residues (PC1 in Fig. S3), while the annotated metabolites including defensive metabolites showed only limited variations (PC1, PC2 and PC3 in Fig. S3).

Similar to the OTUs in caterpillars, we applied joint species modelling to assess the occurrence and the abundance patterns of the bacterial taxa retrieved from the host plant leaves. Of the explained variation, the metabolite composition of the plants was the key determinant in both the occurrence and the abundance models (Table 1, Table S1, Fig. S4). Both models had a strong high phylogenetic signal (estimated posterior of phylogenic strength was 0.97 ± 0.01 and 0.51 ± 0.08), suggesting that related bacteria responded in a similar manner to the variation in the plant metabolite composition (see Figs S5 and S6). In particular, lower residues of carbohydrates and amino acids (PC1 in Fig. S3) were negatively associated with Alphaproteobacteria and Actinobacteria (PC1 of the occurrence model in Fig. S7). Contrary to the microbial communities in the caterpillar, unexplained associations between bacteria OTUs in the host plant were much weaker. The Dirichlet-multinomial modelling results for plant-inhabiting communities indicated that this variation is best explained by a single-component distribution (Fig. S1B).

Influence of microbiota on overwinter survival

As the caterpillars of the Glanville fritillary overwinter gregariously, mainly in family groups, we were interested in testing whether the microbiota composition of the samples collected from the field would correlate with the survival of the siblings remaining in the wild. This could have demonstrated an important fitness benefit of the microbiota composition in wild populations. However, we did not find the over-wintering mortality of families to be related to the microbiota composition or the metabolite profile of the host plants in which the caterpillars were residing on (see Supporting Information).

Discussion

Symbionts that are highly competitive, strongly adapted to their host, and frequently colonize host populations form

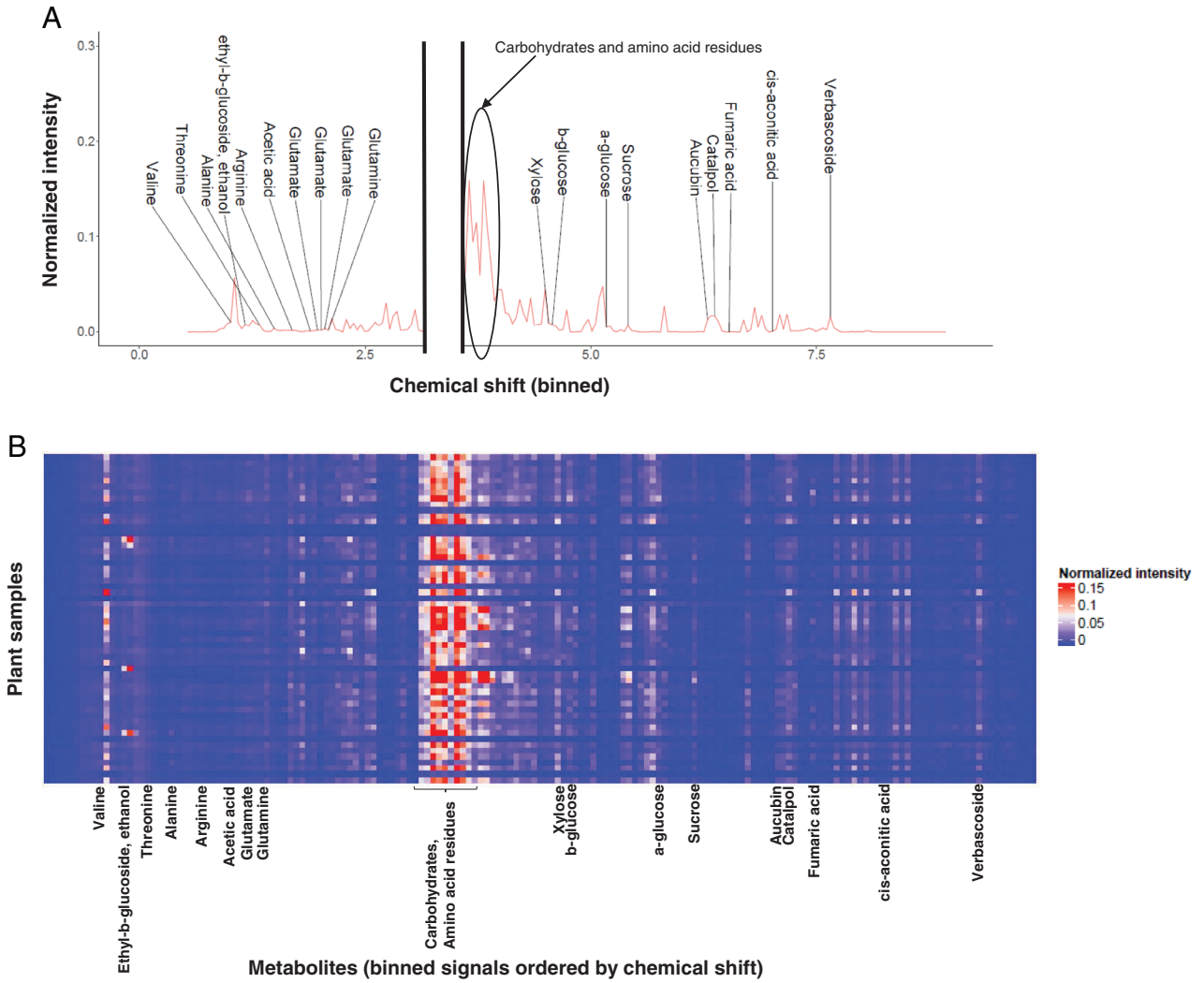


Fig. 6. Metabolomic analysis of *Plantago lanceolata* leaves.
A. An example spectrum (binned at 0.04 ppm) obtained for a leaf sample of *Pl. lanceolata* and showing the chemical shift of the different annotated metabolites. The gap within the graph represents the removal of the solvent pick.
B. A heatmap of the different binned signals showing their normalized intensity across all plant samples.

the core microbiota shared among individuals of the same species (Shapira, 2016). On the contrary, symbionts that are competitively inferior, less adapted to the intestinal conditions (e.g. pH and digestive enzymes), and/or are rarely acquired or transmitted among individuals, tend to form a pool of transient bacteria that consequently are subject to higher fluctuations among hosts (Shapira, 2016; Macke *et al*., 2017). We show that the natural midgut microbial community of *M. cinxia* caterpillar is highly variable, and that only a minor proportion of that variation is related to the measured caterpillar's traits or the properties of the host plant the caterpillar feeds on. Those minor taxa that responded to our assessed covariates were phylogenetically related. We further document a strong co-occurrence pattern of OTUs among caterpillar individual that was independent of the covariates included in our analyses. These co-occurrence patterns in the microbiota were also strongly phylogenetically structured, suggesting two mutually exclusive groups of bacterial communities. One of these co-occurring groups consisted of mainly OTUs that belonged to the Enterobacteriaceae family, whereas the other group consisted of the remaining taxa. Enterobacteriaceae contains several taxa specifically associated with animal digestive system with a broad range of host–microbe interactions ranging from pathogenic to mutualistic (Douglas, 1998; Weiss *et al*., 2006; Chandler *et al*., 2011; Parmentier *et al*., 2016). Enterobacteriaceae are one of the most widespread bacterial family also known to be associated with Lepidoptera (Paniagua Voirol *et al*., 2018), and in *Heliconius erato*, for example, they dominate the gut microbiota already in the early developmental stages (Hammer *et al*., 2014). Consistent with our results, the microbiota of *Drosophila melanogaster* has also been shown to be phylogenetically structured (Adair *et al*., 2018). In general,

Lepidopteran-associated microbiota are suggested to be highly variable (Staudacher *et al*., 2016): a study on caterpillars representing 124 Lepidopteran species showed high inter- and intra-specific variation in the gut microbiota, with a poorly abundant core microbiota (Hammer *et al*., 2017). The dominance of co-occurring taxa, such as Enterobacteriaceae in our study, may be driven by several factors, such as priority effects (dominance of a group of microbes that were the first to colonize the gut), the specific association of bacteria involved in mutualistic interactions, or by a niche overlap among the co-occurring bacteria that grow under similar conditions (Kennedy and Bruns, 2005; Peay *et al*., 2012; Sprockett *et al*., 2018). Due to the limitation in the biological material, we could not quantify the absolute abundance with e.g. qPCR. As our results are based on relative abundances, we cannot exclude the possibility that the individuals have otherwise a uniform microbiota but some individuals are additionally massively colonized by Enterobacteriaceae. This later scenario would hence suggest a potential core microbiota. Under some specific circumstances, which are not determined yet (e.g. decrease of the competition with the core, changes in the immune system of the larvae or random chance of acquisition), this core microbiota may be then supplemented by a community of Enterobacteriaceae leading to the dominance of the latter.

Sex and parasitoid infection are correlated with variation of marginal bacterial taxa

The occurrence of OTUs within Rhodobacterales (Proteobacteria, Alphaproteobacteria) and Neisseriales (Proteobacteria, Betaproteobacteria) orders was generally higher in female than male caterpillars. Due to the absence of sexual dimorphism and proper genetic markers, most studies conducted on immature developmental stages of insects fail to consider sex differences in the microbiome. However, sex-specific differences may greatly impact the microbiota from early caterpillar instar onwards. In the silkworm, where sexes can be identified in the caterpillars (Zhang *et al*., 2010), no strong difference was evident in the global β-diversity structure of the bacterial microbiota. However, marginal differences in the relative abundances of some bacterial taxa were reported, as females were shown to preferentially harbour *Delftia*, *Aurantimonas* and *Staphylococcus* while males were mostly colonized by *Enterococcus* (Sun *et al*., 2016). In adult *H*. *erato*, males and females share similar microbial communities (Hammer *et al*., 2014), whereas in *Spodoptera littoralis* the sexes harbour divergent bacterial communities, with higher Enterobacteriaceae proportion found in females (Chen *et al*., 2016). It is noteworthy that even when found, the consequences of sex-dimorphic microbiota in Lepidoptera are not well understood. Chen *et al*. (2016) showed enrichment of bacteria carrying genes involved in the energetic metabolism in females. Some of these bacterial taxa colonizing females were partly retrieved from the eggs. Those bacteria may be vertically transmitted from the mother to their eggs.

We found that the parasitoid infection was also associated with lower occurrence probability of some taxonomical groups, such as Clostridia, Rhizobiales, Neisseriales and Burkholderiales. This may result from parasitoid infection modifying host's immune (Tan *et al*., 2018) or metabolic (Potter and Woods, 2012; Mrinalini *et al*., 2015) homeostasis that can further influence the intestinal microbial community. Several studies have recently reported an impact of polydnaviruses injected in the caterpillars through the venoms of parasitoid wasps (Cusumano *et al*., 2018; Tan *et al*., 2018; Zhu *et al*., 2018). These symbiotic viruses induce changes in the caterpillar–plant interactions as well as in host immunity. Even though it has never been specifically studied, these viruses might also directly or indirectly impact the microbiota of the caterpillar. Alternatively, individuals not carrying specific symbionts may be more attractive or susceptible to the parasitoid infection. Such processes have been described, for example, in aphids where facultative symbionts interfere with the volatile signals released by the plant to attract parasitoid (Frago *et al*., 2017). *Wolbachia* sp., on the contrary, were more likely to occur in the gut of parasitized individuals. Previous screening of *M*. *cinxia* adults have not found the presence of *Wolbachia*, whereas the parasitoid, *H*. *horticola*, is naturally infected by a *Wolbachia* strain *w*Hho, with an infection rate of approximately 50% of the study population in the Åland islands (Duplouy *et al*., 2015). Therefore, our results suggest that *Wolbachia* may be horizontally transferred by the parasitoid. However, due to the high mortality of individuals to the parasitoid infection it may be extremely rare to find *Wolbachia*-infected adults. Furthermore, we do not know whether *Wolbachia* is able to persist in the individuals across the development or if they are viable only within the caterpillar gut. As recently reported, only 16.3% of the Lepidopteran caterpillars are infected by *Wolbachia* with different impacts of the endosymbiotic bacteria on the reproduction and the sex ratio of their host e.g. male killing, feminization and cytoplasmic incompatibility (Duplouy and Hornett, 2018).

Effects of host plant's microbiota and metabolite composition

The microbiome of plant phyllosphere is partially conserved across species with the presence of recurrent taxa such as *Methylobacterium*, *Pseudomonas* and *Sphingobium* (Delmotte *et al*., 2009). However, the plant microbiome is also generally considered highly variable and subject to spatial and temporal fluctuation in response to several

abiotic factors (Lindow, 1996; Turner *et al*., 2013). In addition, biotic factors, such as plant genotype, developmental stage or chemical composition are known to affect the microbiome (Delmotte *et al*., 2009; Berlec, 2012; Bodenhausen *et al*., 2014; Gargallo-Garriga *et al*., 2016; González-Arenzana *et al*., 2017). Consistently with previous results, we showed that the bacterial community of *P. lanceolata* is highly conserved and dominated by a set of core microorganisms, mainly OTUs classified as *Methylobacterium* that are present in the majority of the samples. These epiphytic Alphaproteobacteria are particularly adapted to the plant phyllosphere and recycle parts of the metabolites secreted by the stomata (methanol and amino acids), and contribute to plant quality, growth and defence (Sy *et al*., 2005; Madhaiyan *et al*., 2006; Kutschera, 2007; Madhaiyan *et al*., 2015).

When considering the whole community structure of the plant microbiota, most of the taxa were correlated with the metabolite profile of the host plant, so that the microbiota tended to decrease with decreasing carbohydrates and amino acid residues. This suggests that these plant metabolites either drive the bacterial communities that successfully colonize the leaves or that the leaf bacterial communities impact plant metabolism. Somewhat surprisingly, the defensive compounds (iridoid glycosides and verbascoside) showed little variation and were not correlated with the plant microbiota.

In general, the microbial communities of host plants and caterpillars were very different. The predominant bacteria in the plants, such as *Methylobacterium* sp., *Hymenobacter* sp., *Modestobacter* sp. and *Aureimonas* sp., were not dominant or even prevalent in the caterpillars. However, a high abundance of OTUs in the host plant did positively affect the same OTUs in the caterpillars in at least few taxonomic groups: in Methylobacteriaceae and some other Alphaproteobacteria, high abundance in the host plant increased their occurrence probabilities in the caterpillars, and in Cytophagaceae and some Methylobacteriaceae, high abundance in the host plant increased the OTU abundance in the caterpillars.

We suggest four potential reasons explaining the observed poor correspondence between caterpillars and host plant microbiota and/or metabolite composition. First, despite the high variability, the bacterial taxa associated with *M. cinxia* gut may be well adapted to their host and consequently weakly impacted by food intake, including the variation in secondary metabolites, such as iridoid glycosides and verbascoside concentrations. Second, the observed caterpillar gut microbiota variability might reflect high abundance of transient bacteria, which are rapidly acquired and eliminated with high turnover. Third, the microbiota of diapausing caterpillars may shift quickly in the beginning of the diapause, in the absence of plant microbial load or metabolites ingested. Fourth, several species of Lepidoptera harbour horizontally acquired bacterial genes that detoxify plants secondary metabolites. Such gene acquisitions might have relaxed any selective pressure in favour of the maintenance of bacterial symbionts within the gut leading to high variability of these communities (Hammer *et al*., 2017; Paniagua Voirol *et al*., 2018). Our observations are somewhat contrasting with other systems in which nutritionally acquired metabolites of the host plant have been observed to strongly shape the animal gut communities (Koropatkin *et al*., 2012; Etxeberria *et al*., 2013; Lu *et al*., 2014; Xu *et al*., 2016). Our results also contrast several other studies in insects that have highlighted the importance of host plant in shaping the gut microbiota community (Broderick *et al*., 2004; Xiang *et al*., 2006; Pinto-Tomás *et al*., 2011; Gayatri Priya *et al*., 2012; Mason and Raffa, 2014; Berman *et al*., 2018; Jones *et al*., 2019), including a study of actively feeding late instar stage of *M. cinxia* (Ruokolainen *et al*., 2016). The microbiota of actively feeding individuals are evidently affected by the plant material that they feed on, which can lead to rapid and reversible changes in the microbiota community depending on the organic matter, defensive metabolites. In actively feeding caterpillars, the microbiota found in faecal samples has been shown to resemble that of the host plant (Hammer *et al*., 2017). Our result of microbiota in the midgut not representing similar microbial community to that of the host plant suggests that the bacterial community of the host plant is actively transported through the digestive tract of the caterpillar while they are eating plant material, and that this community is excreted through the faeces and is not maintained within the gut of the caterpillar after they stopped eating. On the other hand, we cannot exclude the hypothesis that the excretion of the microbiota has happened during the moulting right before the individuals enter into diapause. Furthermore, a recent study on several *Lycaenid* butterfly species showed that starved carnivorous or herbivorous caterpillars did not present any differences in their intestinal communities in comparison to each other (Whitaker *et al*., 2016). In our study, we did not consider the soil below the host plant. The microbial communities of the soil have been previously highlighted as a potential source of microorganisms for the foliar caterpillar *Mamestra brassicae* (Hannula *et al*., 2019). However, if the soil was the microbiota source, we would expect caterpillars from the same host plant growing on the same soil to carry a more similar microbiota. This was not the case here, since most of the variation present occurred among individuals independent of their host plant.

The over-winter survival probability of caterpillars families is spatially structured but does not correlate with the microbiota or metabolite composition of the host plant

We did not find influence of microbiota composition on overwinter survival. However, this may be due to the

indirect assessment of this relationship: as the microbiota of the caterpillars from the same family nest on the same host plant did not resemble each other, the microbiota of the sampled individuals were not likely to be representative of the microbiota of the individuals for which the survival was scored in the field. Previous studies on Lepidoptera have documented contradictory results on the impact of gut microbiota on survival. Experimental perturbation of *Manduca sexta* microbiota by antibiotic treatments had no effect on survival and development (Hammer *et al*., 2017), whereas the removal of *Enterococcus mundtii* symbionts colonizing *Galleria mellonella* decreased individual survival during the adult stage (Johnston and Rolff, 2015). The observed over-winter survival of the *M*. *cinxia* families in the wild exhibited some spatial structure, suggesting that the mortality is strongly influenced by some spatially autocorrelated environmental factor such as summer drought (Saastamoinen *et al*., 2013; Tack *et al*., 2015, Kahilainen *et al*. 2018) or host plant density, not accounted for in our study.

Conclusions

The caterpillars of the Glanville fritillary butterfly present a highly variable gut microbiota even among caterpillars from the same family living on the same host plant individual. Variation in gut microbiota is predominantly related to Enterobacteriaceae, which show marked variation in their diversity among the individuals. Additionally, the occurrence probabilities of some OTUs were impacted by the presence of the parasitoid and by the sex of the caterpillar. The highly variable herbivore microbial communities differed markedly from those of the more conserved host plant microbiota communities. In particular, while the plant leaf metabolites influenced the plant microbiota, these effects did not penetrate to microbiota of the caterpillars feeding on those leaves. Future prospects on other developmental stages (pupae, adults and eggs) should be conducted to broaden our understanding of the variation and potential role of the Glanville fritillary microbiota.

Experimental procedures

The study system

The Glanville fritillary, *Melitaea cinxia*, butterfly occurs across the Eurasian continent, and in northern Europe has a univoltine life cycle (Ehrlich and Hanski, 2004). In Finland, the butterfly occurs only in the SW archipelago, the Åland islands, where it persists as a classic metapopulation within a network of ~4.000 discrete habitat patches consisting of meadows and pastures (Ojanen *et al*., 2013). The habitat patches have been annually surveyed since 1993 for the presence of caterpillar family nests (Hanski, 1994; van Nouhuys and Hanski, 2005; Ojanen *et al*., 2013). Females lay clutches of eggs on two caterpillars host plant species, *Veronica spicata* and *Plantago lanceolata* (Kuussaari *et al*., 2000). The gregarious caterpillars develop within the host plant, and in the fall, they build a thick and conspicuous winter nest, terminate feeding and moult into diapausing morphotype (Wahlberg, 2000). The diapause is broken in spring when the caterpillars continue their development until pupation. Approximately, 30% of the caterpillar families die during the winter (Tack *et al*., 2015). In addition, a conserved proportion of approximately 30% of the individuals get infected by a specialist parasitoid *Hyposoter horticola* (Ehrlich and Hanski, 2004; van Nouhuys and Ehrnsten, 2004). The parasitism occurs during the egg stage, after which the parasitoid develops within the host until it hatches from the seventh instar caterpillars early in the spring and kills the host. Several reasons make this system suitable for the present study: (i) the caterpillars and their host plant can be easily found from the field due to the gregarious life-history of the caterpillars and the conspicuous silk nest they spin for over-wintering; (ii) the over-wintering caterpillars are synchronized in their development prior diapause and have an empty gut at this developmental stage (Ojanen *et al*., 2013), which reduces confounding factors in the analyses; (iii) several individuals, from mainly full-sib families (Fountain *et al*., 2018), can be sampled from the same over-wintering nest on one host plant individual, which allows us to assess individual variation both within and among families; (iv) the local populations are well-described due to the long-term ecological monitoring; and (v) the host sex can be identified at the caterpillar stage using molecular markers (Rastas *et al*., 2013).

Sample collections

Caterpillar and plant samples were collected from natural populations of the *M*. *cinxia* in the region of Sund in the Åland islands within 3-day period in September 2015. This region was selected due to generally high occupancy of the butterfly in three connected networks ensuring sample availability (Supporting Information), and the possibility to control for some potentially confounding factors due to dominance of only one host plant species (*P*. *lanceolata*) and one specialist parasitoid species (*H*. *horticola*) (Nair *et al*., 2016; Hanski *et al*., 2017). The survey followed the general framework of the long-term survey of the *M*. *cinxia* butterfly (described in Ojanen *et al*., 2013): a total of 189 dry meadows i.e. potential habitats were surveyed for the presence of winter nests. Once located, the GPS coordinates were registered using the Earthcape biodiversity platform (http://www.earthcape.com). From each nest, three 5th instar caterpillars and one leaf from the host plant

on which the caterpillars resided were collected with disinfected forceps and stored individually in sterile 1.5 and 15 ml tubes respectively. A total of 191 caterpillars from 66 nests and 63 host plant samples were collected from the 15 patches that were occupied by the butterfly in 2015. In few cases, the entire host plant had already been consumed, and hence no plant sample was collected. The caterpillars were dissected in order to detect the presence of the potential parasitoid and to separate midgut from rest of the carcass (for more details about sample conservation and preparation see Supporting Information). Insect digestive tract is separated in three sections (foregut, midgut and hindgut) with often-observed heterogeneity in their physiology but also in the composition of the microbial communities (Engel and Moran, 2013). Results on Lepidoptera have, however, been somewhat contradictory, with differences in the microbiome across the different gut sections being evident in *Spodoptera littoralis* (Tang *et al*., 2012) but not in *Bombyx mori* (Chen *et al*., 2018). To avoid merging communities that potentially differ, we focused specifically on the microbiota localized within the midgut of the caterpillars. This section is the largest section, most important for food digestion, and its microbiota often shows interaction with host plant secondary metabolites (Terra and Ferreira, 2012; Pentzold *et al*., 2014). The over-winter survival of caterpillars nests in the field (i.e. from which the three individuals were sampled from) was assessed in the spring 2016, by checking the presence of active post-diapause caterpillars (Ojanen *et al*., 2013).

High throughput rrs amplicon sequencing

DNA was extracted from midgut samples with Qiagen DNeasy Blood and Tissue kit (Qiagen, Germany) using an optimized protocol for extraction of bacterial DNA from low matrix (Minard *et al*., 2015). For plant samples, a piece of 0.5 cm^2 was separated from the centre of the leaf, crushed in liquid nitrogen using a sterile pestle, and DNA was extracted following the protocol described for midgut samples. To avoid bias due to the possible confounding effect of extraction set, the samples were randomized before extraction. In addition, three independent extractions were carried out without any matrix and processed with the rest of the samples to identify potential bacterial DNA contamination that could affect results obtained from low biomass samples (Salter *et al*., 2014).

The 280 bp hypervariable V5-V6 region of the *rrs* gene was amplified in duplicates and sequenced with Miseq v.3. sequencing platform (Illumina, San Diego, CA). Details on the protocol are available in the Supporting Information. Analysis of sequences was performed using mothur v.1.37.6 following the Miseq Standard Operating Procedure described by the developers (http://www.mothur.org/wiki/MiSeq_SOP) (Schloss *et al*., 2009). A total of 16,710,206 sequences were obtained after alignment of forward and reverse reads. Aligned sequences were selected within a size range of 250–350 bp with less than eight homopolymers and any ambiguous position. All sequences that did not align to the *rrs* Silva v.123 database were filtered out. *De novo* chimera detection was performed using UCHIME implemented in mothur (Edgar *et al*., 2011). Clustering was performed using a maximum of 3% distance within each OTU according to the average neighbour method. After quality trimming and clustering, every contaminant sequence was trimmed out from the sample × OTU shared table as previously described (Minard *et al*., 2015). The samples were first rarefied at 3000 reads in order to control for sequencing depth biases. Same OTUs were considered as contaminant if they were present in the negative controls and if their proportion in a given sample was not at least 10 times higher than their proportion in the negative controls. After trimming and quality control, the samples were rarefied at 1500 reads per sample for further analysis. Twenty caterpillar samples and two plant samples, which did not contain the minimum amount of sequences, were discarded for the rest of the analysis. Miseq sequences have been deposited on the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under the accession project number PRJEB26629.

Metabolomic analysis of the leaf samples of host plant Plantago lanceolata

After subtraction of the extremities, the remaining parts of each leaf sample were crushed with a sterile pestle in liquid nitrogen and the frozen powder was freeze-dried for 48 h. The extraction was processed using previously described protocol, and ^{1}H-NMR spectra of the metabolites were recorded (Supporting Information; Kim *et al*., 2010). NMR spectra were processed with MNOVA software v.10.0.2 (Mestrelab research S.L., Spain). Model compounds of Aucubin (Sigma-Aldrich, Germany), Catalpol (Sigma-Aldrich) and Verbascoside (Extrasynthese, France) were used for signal assignments of *P. lanceolata* defensive metabolites. Other primary or secondary metabolite shifts and J-coupling constants obtained from plant material using similar solvents were used as reference (Kim *et al*., 2010; Lubbe *et al*., 2011; Yang *et al*., 2012; Agudelo-Romero *et al*., 2014; Gallo *et al*., 2014). For multivariate analysis, the signal was binned to 0.04 ppm and integrated. The trimethylsilylpropanoic acid (TSP) and methanol signals were removed and the relative intensity of the chemical signals was normalized according to the dry mass of the samples and the TSP intensity.

Sex determination

As caterpillar's sex cannot be determined based on morphology, we employed a panel of 24 SNP markers linked

to the Z chromosome to differentiate the sexes (Tables S3 and S4). The sensitivity and specificity of this method were estimated to be 0.81 and 0.89 respectively, based on a group of 150 adult individuals with known gender (75 males and 75 females). A total of 15 individuals could not be annotated based on the SNP panel.

Statistical analyses

We analysed the data with hierarchical modelling of species communities (HMSC; Ovaskainen *et al*., 2017), which approach belongs to the class of joint species distribution modelling (Warton *et al*., 2015). HMSC provide simultaneously species- and community-level inference on how species occurrences and/or abundances relate to environmental covariates, and how these relationships are structured with respect to species traits and phylogenetic relationships. HMSC additionally assesses the structure of co-occurrence patterns among the species that cannot be attributed to responses of the species to the measured covariates, either in spatially hierarchical or in spatially explicit context, depending on the nature of the study design (Ovaskainen *et al*., 2017).

We performed two separate analyses, called hereafter caterpillar and plant models, which differed in whether the OTU data were derived from caterpillar or plant material. In both models, the response variable was the vector of rarified sequence counts of the microbial OTUs. We employed a hurdle approach, in which we first used a probit model for OTU presence–absence, and then a log-normal model for OTU abundances conditional on presence. We restricted the analyses to OTUs that were present in at least five samples (562 and 610 OTUs for caterpillars and plants respectively). We further excluded samples for which plant OTUs or metabolites were missing. The analysed data set consisted of 142 caterpillars collected from 55 host plants (Fig. 2).

In the caterpillar model, our aim was to examine how the OTU composition depended on the properties of the focal caterpillar, and on the OTU and metabolite compositions of its host plant. We included as fixed effects (i) the sex of the individual (0 for female and 1 for male), (ii) the infection status of the individual (0 for non-infected and 1 for infected by the parasitoid wasp), (iii) the abundance of the focal OTU in the host plant where the individual was residing, (iv) the plant OTU community composition and (v) the plant metabolite composition. We measured plant OTU abundance as log-transformed sequence count and described plant OTU community composition and plant metabolite composition by the first three principal components that explained respectively 22% and 92% of their total variations (Figs. S3 and S7). To examine whether the responses of the species to the explanatory variables showed a phylogenetic signal, we included in the analysis a phylogenetic correlation matrix among the OTUs, obtained with FastTree method assuming the general time reversible evolution model (see Fig. S8) (Price *et al*., 2010). To examine residual co-occurrence patterns among the OTUs that cannot be attributed to the fixed effects, we further included in the model a spatial random effect that corresponds to individuals belonging to the same family on the same host plant individual (i.e. host plant level), and a non-structured random effect corresponding to the level of the individual caterpillars. In the plant model, we included as the sole fixed effect the plant metabolite composition, and single spatial random effect of the plant.

We fitted both the caterpillar and the plant models using the HMSC-Matlab implementation of Ovaskainen *et al*. (2017) with default prior distributions. To examine how much of the variation in OTU occurrences can be attributed to the fixed effects and to associations among the OTUs, we evaluated the predictive power of the model in three different ways. All of these accounted for the fixed effects, but differed on how the random effects were treated. Prediction P1 is aimed at measuring the predictive power based solely on fixed effects, and thus we integrated the random effects over their prior distributions rather than using sampling unit-specific fitted values. Prediction P2 is aimed at measuring the predictive power that can be gained by accounting for species-to-species associations. To generate P2, we split the species randomly to two groups, and made the predictions for each species group conditionally on the known occurrences of species belonging to the other group (see Supporting Information for details). Prediction P3 is aimed at measuring the full explanatory power of the model, and thus here the random effects were included based on their fitted values. Therefore, the performance of P1 measures the importance of fixed effects, and the difference between P2 and P1 (respectively, between P3 and P1) gives a minimum (respectively, maximum) estimate for the importance of species-to-species associations. This is because the difference between P3 and P2 may either be a true effect of species-to-species associations that is not captured by our approach of dividing the species into two groups, or then it may be due to overfitting of the random effects. We measured predictive powers by Tjur's R^2 (Tjur, 2009) for the probit models and standard R^2 for the log-normal models. Given that HMSC framework has not previously been used in microorganism studies and thus may be unfamiliar to microbial scientific community, we ran a series of complementary analyses with more traditional methods to support our HMSC-based results. Specifically, both for caterpillar and plant OTU communities we exploited the Dirichlet-mixture approach, proposed by Holmes *et al*. (2012) to test how many of distinct clusters does the data segregate to. The details are given in Supporting Information.

Finally, we analysed whether the overwintering survival of caterpillar nests (siblings of the caterpillars assessed above) was dependent on metabolite and OTU composition of the host plant they were residing on. We performed this analysis with a logistic regression model estimated with STAN (Carpenter *et al*., 2017), in which model we accounted for the spatial locations of the nests using a Gaussian process approach (see Supporting Information) (Rasmussen and Williams, 2006).

ACKNOWLEDGEMENTS

Funding for this project was provided by grants from the European Research Council (Independent Starting Grant No. 637412 'META-STRESS' to M.S.) and the Academy of Finland (Decision Nos. 273098 and 265641 to M.S. and 1273253, 250444 and 284601 to O.O.) and the Research Council of Norway (CoE Grant 223257), as well as by the Biocenter Finland related to the NMR core facility at the Institute on Biotechnology. We acknowledge Juha-Matti Pitkänen for help with DNA extraction, Sami Ojanen for the coordination of sampling, and field assistants for sample collection. We would also like to thank Aapo Kahilainen for his help with genotyping analysis.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest related to this work.

References

Adair, K.L., Wilson, M., Bost, A., and Douglas, A.E. (2018) Microbial community assembly in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J* **12**: 959–972.

Agudelo-Romero, P., Ali, K., Choi, Y.H., Sousa, L., Verpoorte, R., Tiburcio, A.F., and Fortes, A.M. (2014) Perturbation of polyamine catabolism affects grape ripening of *Vitis vinifera* cv. Trincadeira. *Plant Physiol Biochem* **74**: 141–155.

Akman Gunduz, E., and Douglas, A.E. (2009) Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc R Soc B Biol Sci* **276**: 987–991.

Berlec, A. (2012) Novel techniques and findings in the study of plant microbiota: search for plant probiotics. *Plant Sci* **193–194**: 96–102.

Berman, T.S., Laviad-Shitrit, S., Lalzar, M., Halpern, M., and Inbar, M. (2018) Cascading effects on bacterial communities: cattle grazing causes a shift in the microbiome of a herbivorous caterpillar. *ISME J* **12**: 1952–1963.

Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M., and Vorholt, J.A. (2014) A synthetic community approach reveals plant genotypes affecting the Phyllosphere microbiota. *PLoS Genet* **10**: e1004283.

Broderick, N.A., Raffa, K.F., Goodman, R.M., and Handelsman, J. (2004) Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. *Appl Environ Microbiol* **70**: 293–300.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., *et al*. (2017) Stan: a probabilistic programming language. *J Stat Softw* **76**: 1–32.

Ceja-Navarro, J.A., Vega, F.E., Karaoz, U., Hao, Z., Jenkins, S., Lim, H.C., *et al*. (2015) Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nat Commun* **6**: 7618.

Chandler, J.A., Lang, J.M., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011) Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. *PLoS Genet* **7**: e1002272.

Chen, B., Du, K., Sun, C., Vimalanathan, A., Liang, X., Li, Y., *et al.* (2018). Gut bacterial and fungal communities of the domesticated silkworm (*Bombyx mori*) and wild mulberry-feeding relatives. *ISME J* **12**: 2252–2262.

Chen, B., Teh, B.-S., Sun, C., Hu, S., Lu, X., Boland, W., and Shao, Y. (2016) Biodiversity and activity of the gut microbiota across the life history of the insect herbivore *Spodoptera littoralis*. *Sci Rep* **6**: 29505.

Cusumano, A., Zhu, F., Volkoff, A.-N., Verbaarschot, P., Bloem, J., Vogel, H., *et al*. (2018) Parasitic wasp-associated symbiont affects plant-mediated species interactions between herbivores. *Ecol Lett* **21**: 957–967.

Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., *et al*. (2009) Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc Natl Acad Sci U S A* **106**: 16428–16433.

Douglas, A.E. (2011) Lessons from studying insect symbioses. *Cell Host Microbe* **10**: 359–367.

Douglas, A.E. (2015) Multiorganismal insects: diversity and function of resident microorganisms. *Annu Rev Entomol* **60**: 17–34.

Douglas, A.E. (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annu Rev Entomol* **43**: 17–37.

Douglas, A.E. (2010) *The Symbiotic Habit*. Princeton: Princeton University Press.

Dunbar, H.E., Wilson, A.C.C., Ferguson, N.R., and Moran, N. A. (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* **5**: e96.

Duplouy, A., Couchoux, C., Hanski, I., and van Nouhuys, S. (2015) Wolbachia infection in a natural parasitoid wasp population. *PLoS One* **10**: e0134843.

Duplouy, A., and Hornett, E.A. (2018). Uncovering the hidden players in Lepidoptera biology: the heritable microbial endosymbionts. *PeerJ* **6**: e4629.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Ehrlich, P.R., and Hanski, I. (2004) *On the Wings of Checkerspots: A Model System for Population Biology*. New York: Oxford University Press.

Engel, P., and Moran, N.A. (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiol Rev* **37**: 699–735.

Engelstädter, J., and Hurst, G.D.D. (2009) The ecology and evolution of microbes that manipulate host reproduction. *Annu Rev Ecol Evol Syst* **40**: 127–149.

Etxeberria, U., Fernández-Quintela, A., Milagro, F.I., Aguirre, L., Martínez, J.A., and Portillo, M.P. (2013) Impact

of polyphenols and polyphenol-rich dietary sources on gut microbiota composition. *J Agric Food Chem* **61**: 9517–9533.

Fountain, T., Husby, A., Nonaka, E., DiLeo, M.F., Korhonen, J. H., Rastas, P., *et al*. (2018) Inferring dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary butterfly. *Evol Appl* **11**: 287–297.

Frago, E., Mala, M., Weldegergis, B.T., Yang, C., McLean, A., Godfray, H.C.J., *et al*. (2017) Symbionts protect aphids from parasitic wasps by attenuating herbivore-induced plant volatiles. *Nat Commun* **8**: 1860.

Gallo, V., Mastrorilli, P., Cafagna, I., Nitti, G.I., Latronico, M., Longobardi, F., *et al*. (2014) Effects of agronomical practices on chemical composition of table grapes evaluated by NMR spectroscopy. *J Food Compos Anal* **35**: 44–52.

Gargallo-Garriga, A., Sardans, J., Pérez-Trujillo, M., Guenther, A., Llusià, J., Rico, L., *et al*. (2016) Shifts in plant foliar and floral metabolomes in response to the suppression of the associated microbiota. *BMC Plant Biol* **16**: 78.

Gayatri Priya, N., Ojha, A., Kajla, M.K., Raj, A., and Rajagopal, R. (2012) Host plant induced variation in gut bacteria of *Helicoverpa armigera*. *PLoS One* **7**: e30768.

González-Arenzana, L., Portu, J., López, R., Garijo, P., Garde-Cerdán, T., and López-Alfaro, I. (2017) Phenylalanine and urea foliar application: effect on grape and must microbiota. *Int J Food Microbiol* **245**: 88–97.

Hammer, T.J., Janzen, D.H., Hallwachs, W., Jaffe, S.P., and Fierer, N. (2017) Caterpillars lack a resident gut microbiome. *Proc Natl Acad Sci U S A* **114**: 9641–9646.

Hammer, T.J., McMillan, W.O., and Fierer, N. (2014) Metamorphosis of a butterfly-associated bacterial community. *PLoS One* **9**: e86995.

Hammer, T.J., Sanders, J.G., and Fierer, N. (2019) Not all animals need a microbiome. *FEMS Microbiol Lett* **366**: 1–11.

Hannula, S.E., Zhu, F., Heinen, R., and Bezemer, T.M. (2019) Foliar-feeding insects acquire microbiomes from the soil rather than the host plant. *Nat Commun* **10**: 1254.

Hanski, I. (1994) A practical model of metapopulation dynamics. *J Anim Ecol* **63**: 151–162.

Hanski, I., Schulz, T., Wong, S.C., Ahola, V., Ruokolainen, A., and Ojanen, S.P. (2017) Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. *Nat Commun* **8**: 14504.

Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS One* **7**: e30126.

Johnston, P.R., and Rolff, J. (2015) Host and symbiont jointly control gut microbiota during complete metamorphosis. *PLoS Pathog* **11**: e1005246.

Jones, A.G., Mason, C.J., Felton, G.W., and Hoover, K. (2019) Host plant and population source drive diversity of microbial gut communities in two polyphagous insects. *Sci Rep* **9**: 2792.

Kahilainen, A., van Nouhuys, S., Schulz, T., & Saastamoinen, M. (2018). Metapopulation dynamics in a changing climate: Increasing spatial synchrony in weather conditions drives metapopulation synchrony of a butterfly inhabiting a fragmented landscape. *Glob Change Biol* **24**: 4316–4329.

Kennedy, P.G., and Bruns, T.D. (2005) Priority effects determine the outcome of ectomycorrhizal competition between two *Rhizopogon* species colonizing *Pinus muricata* seedlings. *New Phytol* **166**: 631–638.

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T. (2012) Symbiont-mediated insecticide resistance. *Proc Natl Acad Sci U S A* **109**: 8618–8622.

Kim, H.K., Choi, Y.H., and Verpoorte, R. (2010) NMR-based metabolomic analysis of plants. *Nat Protoc* **5**: 536–549.

King, K.C., Brockhurst, M.A., Vasieva, O., Paterson, S., Betts, A., Ford, S.A., *et al*. (2016) Rapid evolution of microbe-mediated protection against pathogens in a worm host. *ISME J* **10**: 1915–1924.

Koropatkin, N.M., Cameron, E.A., and Martens, E.C. (2012) How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **10**: 323–335.

Kutschera, U. (2007) Plant-associated methylobacteria as co-evolved phytosymbionts. *Plant Signal Behav* **2**: 74–78.

Kuussaari, M., Singer, M., and Hanski, I. (2000) Local specialization and landscape-level influence on host use in an herbivorous insect. *Ecology* **81**: 2177–2187.

Lindow, S.E. (1996) Role of immigration and other processes in determining epiphytic bacterial populations. In *Aerial Plant Surface Microbiology*, Morris, C.E., Nicot, P. C., and Nguyen-The, C. (eds). New York: Springer, pp. 155–168.

Lu, K., Abo, R.P., Schlieper, K.A., Graffam, M.E., Levine, S., Wishnok, J.S., *et al*. (2014) Arsenic exposure perturbs the gut microbiome and its metabolic profile in mice: an integrated metagenomics and metabolomics analysis. *Environ Health Perspect* **122**: 284–291.

Lubbe, A., Choi, Y.H., Vreeburg, P., and Verpoorte, R. (2011) Effect of fertilizers on galanthamine and metabolite profiles in narcissus bulbs by 1H NMR. *J Agric Food Chem* **59**: 3155–3161.

Macke, E., Tasiemski, A., Massol, F., Callens, M., and Decaestecker, E. (2017) Life history and eco-evolutionary dynamics in light of the gut microbiota. *Oikos* **126**: 508–531.

Madhaiyan, M., Alex, T.H.H., Ngoh, S.T., Prithiviraj, B., and Ji, L. (2015) Leaf-residing *Methylobacterium* species fix nitrogen and promote biomass and seed production in *Jatropha curcas*. *Biotechnol Biofuels* **8**: 222.

Madhaiyan, M., Reddy, B.V.S., Anandham, R., Senthilkumar, M., Poonguzhali, S., Sundaram, S.P., and Sa, T. (2006) Plant growth–promoting *Methylobacterium* induces defense responses in groundnut (*Arachis hypogaea* L.) compared with rot pathogens. *Curr Microbiol* **53**: 270–276.

Mason, C.J., and Raffa, K.F. (2014) Acquisition and structuring of midgut bacterial communities in gypsy moth (Lepidoptera: Erebidae) larvae. *Environ Entomol* **43**: 595–604.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., *et al*. (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* **110**: 3229–3236.

Minard, G., Tran, F.H., Van, V.T., Goubert, C., Bellet, C., Lambert, G., *et al*. (2015) French invasive Asian tiger mosquito populations harbor reduced bacterial microbiota and genetic diversity compared to Vietnamese autochthonous relatives. *Front Microbiol* **6**: 970.

Montllor, C.B., Maxmen, A., and Purcell, A.H. (2002) Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. *Ecol Entomol* **27**: 189–195.

Mrinalini, Siebert, A.L., Wright, J., Martinson, E., Wheeler, D., and Werren, J.H. (2015) Parasitoid venom induces metabolic cascades in fly hosts. *Metabolomics* **11**: 350–366.

Nair, A., Fountain, T., Ikonen, S., Ojanen, P.S., and van Nouhuys, S. (2016) Spatial and temporal genetic structure at the fourth trophic level in a fragmented landscape. *Proc R Soc B Biol Sci* **283**: 1–8.

van Nouhuys, S., and Ehrnsten, J. (2004) Wasp behavior leads to uniform parasitism of a host available only a few hours per year. *Behav Ecol* **15**: 661–665.

van Nouhuys, S., and Hanski, I. (2005) Metacommunities of butterflies, their host plants, and their parasitoids. In *Metacommunities Spatial Dynamics and Ecological Communities*, Vol. **99**. Chicago: University of Chicago Press.

Ojanen, S.P., Nieminen, M., Meyke, E., Pöyry, J., and Hanski, I. (2013) Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): survey methods, data management, and long-term population trends. *Ecol Evol* **3**: 3713–3737.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., *et al*. (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol Lett* **20**: 561–576.

Paniagua Voirol, L.R., Frago, E., Kaltenpoth, M., Hilker, M., and Fatouros, N.E. (2018) Bacterial symbionts in lepidoptera: their diversity, transmission, and impact on the host. *Front Microbiol* **9**: 556.

Parmentier, L., Meeus, I., Mosallanejad, H., de Graaf, D.C., and Smagghe, G. (2016) Plasticity in the gut microbial community and uptake of Enterobacteriaceae (Gammaproteobacteria) in *Bombus terrestris* bumblebees' nests when reared indoors and moved to an outdoor environment. *Apidologie* **47**: 237–250.

Peay, K.G., Belisle, M., and Fukami, T. (2012) Phylogenetic relatedness predicts priority effects in nectar yeast communities. *Proc Biol Sci* **279**: 749–758.

Pentzold, S., Zagrobelny, M., Rook, F., and Bak, S. (2014) How insects overcome two-component plant chemical defence: plant β-glucosidases as the main target for herbivore adaptation. *Biol Rev* **89**: 531–551.

Pinto-Tomás, A.A., Sittenfeld, A., Uribe-Lorío, L., Chavarría, F., Mora, M., Janzen, D.H., *et al*. (2011) Comparison of midgut bacterial diversity in tropical caterpillars (Lepidoptera: Saturniidae) fed on different diets. *Environ Entomol* **40**: 1111–1122.

Potter, K.A., and Woods, H.A. (2012) *Trichogramma* parasitoids alter the metabolic physiology of *Manduca* eggs. *Proc R Soc B Biol Sci* **279**: 3572–3576.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Rani, A., Sharma, A., Rajagopal, R., Adak, T., and Bhatnagar, R.K. (2009) Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected *Anopheles stephensi*-an Asian malarial vector. *BMC Microbiol* **9**: 96.

Rasmussen, C.E., and Williams, C.K.I. (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.

Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., and Auvinen, P. (2013) Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* **29**: 3128–3134.

Ruokolainen, L., Ikonen, S., Makkonen, H., and Hanski, I. (2016) Larval growth rate is associated with the composition of the gut microbiota in the Glanville fritillary butterfly. *Oecologia* **181**: 895–903.

Saastamoinen, M., Ikonen, S., Wong, S.C., Lehtonen, R., and Hanski, I. (2013) Plastic larval development in a butterfly has complex environmental and genetic causes and consequences for population dynamics. *J Anim Ecol* **82**: 529–539.

Salem, H., Bauer, E., Strauss, A.S., Vogel, H., Marz, M., and Kaltenpoth, M. (2014) Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. *Proc R Soc Lond B Biol Sci* **281**: 20141838.

Salter, S., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W. O., Moffatt, M.F., *et al*. (2014) Reagent contamination can critically impact sequence-based microbiome analyses. *BMC Biol* **12**: 87.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Shapira, M. (2016) Gut microbiotas and host evolution: scaling up symbiosis. *Trends Ecol Evol* **31**: 539–549.

Sprockett, D., Fukami, T., and Relman, D.A. (2018) Role of priority effects in the early-life assembly of the gut microbiota. *Nat Rev Gastroenterol Hepatol* **15**: 197–205.

Staubach, F., Baines, J.F., Künzel, S., Bik, E.M., and Petrov, D. A. (2013) Host species and environmental effects on bacterial communities associated with *Drosophila* in the laboratory and in the natural environment. *PLoS One* **8**: e70749.

Staudacher, H., Kaltenpoth, M., Breeuwer, J.A.J., Menken, S.B.J., Heckel, D.G., and Groot, A.T. (2016) Variability of bacterial communities in the moth *Heliothis virescens* indicates transient association with the host. *PLoS One* **11**: e0154514.

Sun, Z., Lu, Y., Zhang, H., Kumar, D., Liu, B., Gong, Y., *et al*. (2016) Effects of BmCPV infection on silkworm *Bombyx mori* intestinal bacteria. *PLoS One* **11**: e0146313.

Sy, A., Timmers, A.C.J., Knief, C., and Vorholt, J.A. (2005) Methylotrophic metabolism is advantageous for *Methylobacterium extorquens* during colonization of *Medicago truncatula* under competitive conditions. *Appl Environ Microbiol* **71**: 7245–7252.

Tack, A.J., Mononen, T., and Hanski, I. (2015) Increasing frequency of low summer precipitation synchronizes dynamics and compromises metapopulation stability in the Glanville fritillary butterfly. *Proc. Biol. Sci.* **282**: 20150173.

Tan, C.-W., Peiffer, M., Hoover, K., Rosa, C., Acevedo, F.E., and Felton, G.W. (2018) Symbiotic polydnavirus of a parasite manipulates caterpillar and plant immunity. *Proc Natl Acad Sci U S A* **115**: 5199–5204.

Tang, X., Freitak, D., Vogel, H., Ping, L., Shao, Y., Cordero, E.A., *et al.* Boland, W. (2012). Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS One* **7**: e36978.

Terra, W.R., and Ferreira, C. (2012) 11 - biochemistry and molecular biology of digestion. In *Insect Molecular Biology and Biochemistry*, Gilbert, L.I. (ed). San Diego: Academic Press, pp. 365–418.

Tinker, K.A., and Ottesen, E.A. (2016) The core gut microbiome of the American cockroach, *Periplaneta americana*, is stable and resilient to dietary shifts. *Appl Environ Microbiol* **82**: 6603–6610.

Tjur, T. (2009) Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. *Am Stat* **63**: 366–372.

Turner, T.R., James, E.K., and Poole, P.S. (2013) The plant microbiome. *Genome Biol* **14**: 209.

Wahlberg, N. (2000) Comparative descriptions of the immature stages and ecology of five Finnish melitaeine butterfly species (Lepidoptera: Nymphalidae). *Ann Zool Fenn* **11**: 167–174.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., *et al*. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560–565.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., and Hui, F.K.C. (2015) So many variables: joint modeling in community ecology. *Trends Ecol Evol* **30**: 766–779.

Weiss, B.L., Mouchotte, R., Rio, R.V.M., Wu, Y.-N., Wu, Z., Heddi, A., and Aksoy, S. (2006) Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. *Appl Environ Microbiol* **72**: 7013–7021.

Werren, J.H., Baldo, L., and Clark, M.E. (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6**: 741–751.

Whitaker, M.R.L., Salzman, S., Sanders, J., Kaltenpoth, M., and Pierce, N.E. (2016) Microbial communities of Lycaenid butterflies do not correlate with larval diet. *Front Microbiol* **7**: 1920.

Xiang, H., Wei, G.-F., Jia, S., Huang, J., Miao, X.-X., Zhou, Z., *et al*. (2006) Microbial communities in the larval midgut of laboratory and field populations of cotton bollworm (*Helicoverpa armigera*). *Can J Microbiol* **52**: 1085–1092.

Xu, L., Shi, Z., Wang, B., Lu, M., and Sun, J. (2016) Pine defensive monoterpene α-pinene influences the feeding behavior of *Dendroctonus valens* and its gut bacterial community structure. *Int J Mol Sci* **17**: 1734.

Yang, S.-O., Shin, Y.-S., Hyun, S.-H., Cho, S., Bang, K.-H., Lee, D., *et al*. (2012) NMR-based metabolic profiling and differentiation of ginseng roots according to cultivation ages. *J Pharm Biomed Anal* **58**: 19–26.

Yun, J.-H., Roh, S.W., Whon, T.W., Jung, M.-J., Kim, M.-S., Park, D.-S., *et al*. (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl Environ Microbiol* **80**: 5254–5264.

Zeng, Q., Sukumaran, J., Wu, S., and Rodrigo, A. (2015) Neutral models of microbiome evolution. *PLoS Comput Biol* **11**: e1004365.

Zhang, Y., Yu, X., Shen, W., Ma, Y., Zhou, L., Xu, N., and Yi, S. (2010) Mechanism of fluorescent cocoon sex identification for silkworms *Bombyx mori*. *Sci China Life Sci* **53**: 1330–1339.

Zhu, F., Cusumano, A., Bloem, J., Weldegergis, B.T., Villela, A., Fatouros, N.E., *et al*. (2018) Symbiotic polydnavirus and venom reveal parasitoid to its hyperparasitoids. *Proc Natl Acad Sci U S A* **115**: 5205–5210.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1: Supplementary Material and Methods.

Supplementary Figure S1. Results of Dirichlet-multinomial analysis. Panels A and B depict the Laplace goodness-of-fit measures for fitted mixture models with different number of mixture components (lower values corresponds to better fit). Panel C visualize the 2 mixture components of the best model for microbial OTUs in caterpillars.

Supplement Fig. S2. Results of permutation tests. Each panel depicts the Spearman rank correlation coefficient between the assigned mixture component of the best 2-component Dirichlet-multinomial model for caterpillar OTUs data and available predictors. Red line corresponds to the real value and the black curve depict the density of permutation-based values. Dashed blue lines depict the 2.5% and 97.5% quantiles of the permutation-based density.

Supplement Fig. S3. Principal Component Analysis (PCA) of the metabolites associated with the host plant. The PCA plot represents the ordination of the plant metabolites on the three first Principal Components (A) PC1 and PC2, (B) PC1 and PC3, (C) PC2 and PC3. The signal corresponding to the chemical shift of carbohydrates and amino acid residues are coloured in red while other signals are coloured in blue.

Supplement Fig. S4. Partitioning of the explained variance of bacterial OTUs among the fixed and random effects in plant models. The coloured bars show, for each OTU, the proportions of variance attributed to each of explanatory variables. The average variance proportions over the OTUs are shown in the legend box. The ordering of OTU is following ordering of Fig. 1 except for the OTUs that were recorded only in larvae samples (for details, see Supplementary Table S2). See *Statistical Methods* for a full description of the included fixed and random effects.

Supplement Fig. S5. The influence of metabolic covariates on plant microbiota. Regression coefficients that were estimated to be positive (respectively, negative) with 95% credibility level are shown by red (respectively, blue). The ordering of the OTUs is identical to that of Fig. S4.

Supplement Fig. S6. Residual associations among plant microbiota. The panels illustrate the random effects for the presence-absence (A) and abundance (B) parts of the plant model. OTU-pairs for which the residual correlation was estimated to be positive (respectively, negative) with 95% credibility level are shown by red (respectively, blue) colour. The ordering of the OTUs is identical to that of Fig. S4.

Supplement Fig. S7. Principal Component Analysis (PCA) of the bacterial microbiota associated with the host plant. The PCA plot represents the ordination of the bacterial Operational Taxonomic Units (OTUs) on the three first Principal Components (A) PC1 and PC2, (B) PC1 and PC3, (C) PC2 and PC3. The colour scale represents the OTU classification at the Phylum level.

Supplement Fig. S8. OTUs phylogenic relationship matrices. Phylogenetic relationship among OTUs are represented for larvae (A) and plant (B) microbial The relationships between the OTUs, used for analysis of bacterial community, were obtained with FastTree method assuming the GTR evolution model. Colour of each cell encodes the relationship between the OTUs, located at those row and column with the gradation of red indicating the level of relatedness. The order of the OTUs is selected according to the available approximate taxonomic classification and is further aligned according to phylogenic similarity, with the colours and relative ordering following the Fig. 2 in main text. A detailed similarly ordered lists of individual OTU with their full taxonomic classification are provided in Supplementary Table S2.

Supplementary Table S1. Predictive powers of the larval and plant models. Predictive power is measured by Tjur R^2 for the presence-absence models and by the standard R^2 for the abundance models. The values show the mean $\pm$ standard deviation over the OTUs. As detailed in the *Statistical Methods*, Prediction P1 measures the predictive power solely due to the fixed effects part of the models, whereas P2 and P3 also account for species-to-species associations, with P2 being based on cross-validation across species and P3 in on predicting the same data that were used to fit the model.

Supplementary Table S2. Taxonomic classification of the bacterial Operational Taxonomic Units.

Supplementary Table S3. Proportion of Heterozygous loci observed in males and females within the validation sample panel with known gender.

Supplementary Table S4. Genotyping for sex determination of *M. cinxia*.