

Host-structured adult butterfly microbiomes share a common core across continents

Arne Weinhold^{1*}, Andrea Pinos^{1,2#}, Yenny Correa-Carmona^{3#}, Kim L. Holzmann⁴, Pedro Alonso-Alonso⁴, Felipe Yon⁵, Ingolf Steffan-Dewenter⁴, Marcell K. Peters^{4,6}, Gunnar Brehm³, Alexander Keller¹

¹Center for Organismic Adaptation, Cellular and Organismic Networks, Ludwig-Maximilians-Universität München, Munich, Germany

²Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Carl Gustav Carus Dresden, Dresden, Germany

³Institut für Zoologie und Evolutionsbiologie mit Phyletischem Museum, Friedrich-Schiller University Jena, Jena, Germany

⁴Department of Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, Würzburg, Germany

⁵Departamento de Ciencias Biológicas y Fisiológicas, Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima, Peru

⁶Animal Ecology, BIOM, University of Bremen, Germany

*** Corresponding Author:**

arne.weinhold@bio.lmu.de

[#] authors contributed equally

Keywords: Lepidoptera, Butterflies, Pollination, Microbiome, Neotropics, Palearctic, Assembly, Host-filtering, Pieridae

Abstract

The assembly of host-specific microbiota is critical for health and functioning of many insect pollinators. While social pollinators maintain core microbiota through social transmission, the factors driving microbiota assembly in solitary pollinators remain poorly understood. Lepidoptera are an important group of pollinators, but microbiome research has largely focused on their larval stage, while nectar-feeding adults have been widely ignored. Field-based studies of the adult butterfly microbiota are rare and geographically and taxonomically restricted.

Here, we characterize the microbiota of adult butterflies from natural environments along elevational and seasonal gradients across two continents spanning the Neotropical and Palearctic realms. Microbiota diversity and composition were primarily explained by host taxonomic identity, whereas geographic location and temperature had little effect. We found high microbial titers and a common core microbiota across seven butterfly subfamilies from temperate and tropical regions. A comparison with sympatric wild bees revealed partial taxonomic overlap but marked distinctions in lactic acid bacteria between Lepidoptera and Hymenoptera.

Together, our results demonstrate that host taxonomic identity, rather than environmental drivers, is the dominant force structuring the microbiota of adult butterflies. This highlights that the host-filtering capacity of solitary species has been largely underestimated, challenging previous assumptions that the microbiota of this important pollinator group is primarily environmentally driven and of limited functional significance.

2 Introduction

Microbes play an important role in pollination, acting as a third interaction partner between flowers and pollinating insects, providing benefits for both sides^{1,2}. Social pollinators are well known for their ability to maintain microbes via conspecific interactions, resulting in conserved and host-specific core microbiota^{3,4}. In contrast, solitary pollinators account for most pollinating species worldwide and their microbiota is assumed to be mainly shaped by environmental factors, as they acquire their microbiota primarily from environmental sources via plant-insect interaction networks⁵. Microbial sharing between flowers and insects has long been hypothesized^{6,7} and experimentally confirmed in solitary bees and butterflies^{8,9}. Despite this, microbiota research has largely focused on hymenopteran pollinators¹⁰, while other insect orders (e.g. Hemiptera, Diptera or Lepidoptera) have received comparatively little attention^{11,12}. Comparing pollinator microbiota across insect orders co-occurring in the same habitat enables a better evaluation of host-specific enrichment.

Lepidoptera are the second-largest insect order in terms of species number and play a crucial ecological role as important non-bee pollinators^{13,14}. Despite a long research tradition, studies of the lepidopteran microbiota have focused almost exclusively on the phytophagous larval stage (caterpillars)^{15,16}, while the adult stage (imago) has been widely ignored^{17–20}. This bias was further amplified by laboratory assays using moth pest species, where adults had no access to flowers or nectar microbes, but were fed sterile sugar solutions^{21,22} or sampled immediately after emergence without prior feeding²³. These studies on larvae and captive adults reinforced the widespread perception that Lepidoptera lack specific microbial associations^{16,19,24}. However, adult butterflies differ markedly in gut morphology, lacking the simple tube-like guts of larvae but contain folds and villi-like structures that provide surfaces for microbial colonization²⁵. Unlike larvae, adult guts are less alkaline, with more acidic pH conditions in the hindgut²⁶ and digestive enzyme activity optimal at pH 6.4²⁷. When directly compared, adults of true butterflies (Papilionoidea) harbor microbiota compositions clearly distinct from those of larvae^{28–30}. That adults sampled from natural environments show contrasting microbiota compositions and

substantial microbial loads has only recently been recognized in lepidopteran microbiome research³¹.

Nevertheless, comprehensive studies of the adult butterfly microbiota from natural environments remain scarce and are largely limited to Heliconiinae (Nymphalidae) from Central and South America^{28,32–35}. These studies revealed high microbial titers in Neotropical butterflies and communities clearly distinct from those of food sources or trap baits^{32,35}. Microbiota compositions were best explained by host taxonomic placement, rather than diet, contrasting with initial expectations that feeding preferences—nectar versus fruit, or pollen versus non-pollen—would be primary drivers^{32,33}. Neotropical butterflies were consistently associated with a characteristic set of microbial taxa, dominated by insect-specific groups (*Orbus*, *Wolbachia*, *Spiroplasma*), lactic acid bacteria (*Enterococcus*, *Lactococcus*) and acetic acid bacteria (*Asaia*, *Commensalibacter*)^{28,32–35}. Whether this pattern is specific to Neotropical species or a general signature of the adult butterfly microbiota remains unclear, as systematic samplings across distant climate zones are still lacking. Defining a ‘common core’ microbiota across different host families or locations is a crucial first step to distinguish potentially host-enriched taxa from transient microbes^{3,36}. To date, global investigations have focused primarily on vertically transmitted endosymbionts such as *Wolbachia*^{37,38}, and the presence of other microbial taxa must be inferred from reports of individual bacterial isolates^{39,40} or from non-host reads mined from genomic datasets⁴¹.

Here, we address this knowledge gap by systematically investigating the microbiota of adult butterflies across distinct geographic regions and climate zones, comparing species from the Neotropical and Palearctic realms. Samples were collected along an elevational gradient in Peru (subfamilies: Dismorphiinae, Pierinae, Coliadinae, Danainae and Heliconiinae), spanning largely undisturbed habitats from the species-rich lowland Amazon basin to montane forests, and in Germany (subfamilies: Dismorphiinae, Pierinae, Coliadinae Satyrinae and Nymphalinae), through repeated seasonal sampling from suburban areas. This design allows us to use temperature as environmental predictor of microbiota variation along a gradient ranging from 6.7°C to 26.1°C, to disentangle environmental influences from host filtering effects. For broader host context, we compare

butterfly microbiota to that of sympatric wild bees from a previous study sampled along the same Peruvian gradient⁴².

Expanding upon previous work centered on Heliconiinae, our sampling covers a broader taxonomic range with approximately 70% of samples belonging to Pierinae and Coliadinae. These ‘whites’ and ‘sulfurs’ are evolutionary model groups to study diversification dynamics in response to environmental temperatures⁴³. Their widespread distribution across both study regions enabled us to assess whether environmental temperature, geographic location, or host taxonomy are the primary drivers of the adult butterfly microbiota. Within this context, we test the following hypotheses:

- 1) Environmental temperature shapes microbiota diversity in adult butterflies.
- 2) Microbial community composition is structured by host taxonomic identity.
- 3) Adult butterflies harbor a common core microbiota independent of geographic location.

We further compare common core taxa of adult butterflies with those reported from hymenopteran pollinators sampled from the same area in Peru⁴², highlighting distinctions among bee- and butterfly-associated groups. Together, these results offer a comprehensive characterization of the adult butterfly microbiota, including the first broad microbiome survey of European populations. This lays a foundation to further explore the ecological relevance and functional roles of adult-stage microbiota in butterflies.

3 Results

We compared the microbiota of adult butterflies across two distant countries (Peru and Germany) representing the Neotropics and the Palearctic realms (Figure 1A) within seven butterfly subfamilies: Dismorphiinae (n = 4), Pierinae (n = 73) and Coliadinae (n = 41) from the Pieridae family, and Danainae (n = 3), Satyrinae (n = 10), Heliconiinae (n = 27) and Nymphalinae (n = 9) belonging to the Nymphalidae family.

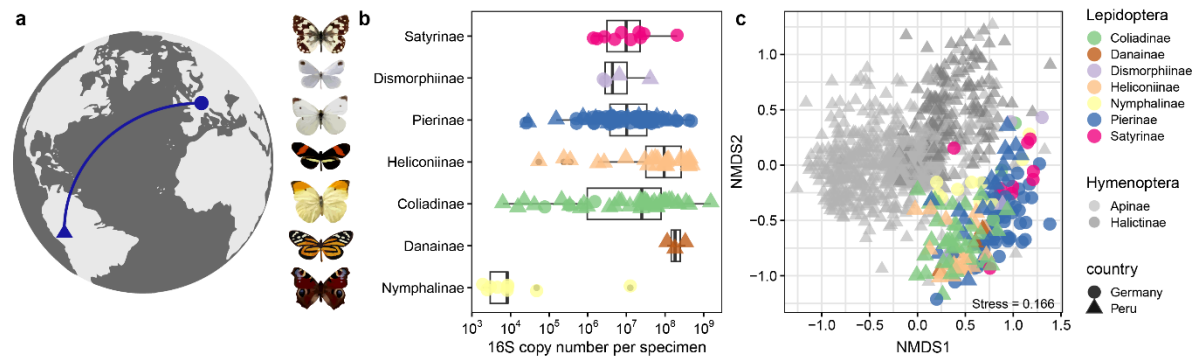


Figure 1 Adult butterflies harbor abundant microbiomes with host-structured community composition across continents

(a) Overview of adult butterfly sampling locations in Peru and Germany. Representative species are shown for each butterfly subfamily. (b) Absolute quantification of the bacterial load by 16S rRNA gene qPCR. (c) NMDS ordination based on Bray-Curtis dissimilarities comparing microbiome composition of Lepidoptera and Hymenoptera. Adult butterflies from Peru ($n = 90$) cluster together with butterflies from Germany ($n = 77$), while their microbiomes are distinct from those of sympatric wild bees from Peru (Apinae $n = 670$ and Halictinae $n = 204$).

3.1 Adult butterflies host abundant microbiota distinct from Hymenoptera

Quantitative analysis by 16S qPCR showed that the bacterial load of the different butterfly subfamilies spanned a broad range between 10^3 – 10^9 16S copy numbers per specimen (Figure 1B). Adult butterflies harbored a median bacterial load around 1.85×10^7 16S copies (interquartile range: 2.6×10^6 – 6.1×10^7) which was strongly predicted by host genus identity in a linear model (adj. $R^2 = 0.51$, $F_{23,143} = 8.6$, $p < 10^{-15}$) (Supplementary Figure S1A). A notable exception in abundance were *Aglais* (Nymphalinae), which were collected very early in season at low temperatures. Within individual subfamilies the bacterial load did not differ between countries (Supplementary Figure S1B). Using Type II ANOVA on a linear model showed a strong effect of host subfamily on bacterial load (partial $R^2 = 0.30$, $F_{6,159} = 11.32$, $p < 10^{-9}$), whereas country had no significant effect (partial $R^2 = 0.003$, $F_{1,159} = 0.49$, $p = 0.48$).

For a broader insect order comparison, we combined the butterfly microbiota with that of wild bees, collected from the same locations in Peru⁴². Butterflies from Peru grouped

more closely with those sampled from Germany, than with sympatric bees (Figure 1C). Overall microbial community composition differed between Hymenoptera and Lepidoptera, with host order explaining 8.2% of the variation in community composition (PERMANOVA, $R^2 = 0.082$, $F_{1,1037} = 92.1$, $P < 0.001$).

Within adult butterflies, microbiota diversity was better explained by host taxonomic placement than by sampling location (Supplementary Figure S1C). A linear model including host subfamily and location explained 31% of variation in Shannon diversity (adj. $R^2 = 0.31$, $F_{9,157} = 9.4$, $p < 10^{-11}$). We found a strong effect of host subfamily (Type II ANOVA partial $R^2 = 0.17$, $F_{6,157} = 5.5$, $p < 0.0001$), whereas sampling location had no significant effect (partial $R^2 = 0.04$, $F_{3,157} = 2.1$, $p = 0.11$). A similar result was obtained with a model including host tribe and country as predictor variables (Supplementary Figure S1C). As the sampling was not equal across countries, we additionally tested the effect of country on Shannon diversity within individual butterfly subfamilies (Supplementary Figure S1D). Still, country had no significant effect on microbial alpha diversity for Coliadinae ($p = 0.12$) or Pierinae ($p = 0.62$), while inferences are limited for Dismorphiinae ($p = 0.33$) due to the small sample size (Supplementary Figure S1D). Although samples had been collected from different continents, sampling location had minor influence on Shannon diversity compared to host taxonomic identity.

3.2 Influence of environmental gradients on adult butterfly microbiota diversity

Butterflies from Peru came from an elevational gradient, which showed strong linear correlation with environmental temperature (adj. $R^2 = 0.91$). A linear model with only the Peru samples including host subfamily and temperature as predictors explained 26% of the variation in Shannon diversity (adj. $R^2 = 0.26$, $F_{5,84} = 7.2$, $p < 10^{-4}$). While the main effect was accounted to host subfamily (partial $R^2 = 0.14$, $F_{4,84} = 3.5$, $p = 0.011$), temperature showed a lower effect size but was still significant (partial $R^2 = 0.06$, $F_{1,84} = 5.5$, $p = 0.021$). However, most of the butterflies were sampled below 1500 m, and only individual cold adapted Pierinae from higher elevations (*Catantix* at 2390 m and *Tatochila* at 3350 m) (Supplementary Figure S2A).

In contrast, butterflies from Germany were resampled from the same locations at different times of the year between March and September. Time of sampling (calendar week) showed a quadratic relationship with temperature (adj. $R^2 = 0.82$) but was a better predictor of Shannon diversity than temperature in the same linear model (Supplementary Figure S2B). When excluding *Aglaia* as low biomass outgroup the full model explained 33% of the variation in Shannon diversity ($F_{6,62} = 6.5$, $p < 10^{-4}$), which could be mainly accounted to calendar week (partial $R^2 = 0.19$, $F_{1,61} = 14.9$, $p = 0.0003$), while host subfamily (partial $R^2 = 0.04$, $F_{4,62} = 0.6$, $p = 0.63$) nor temperature (partial $R^2 = 0.02$, $F_{1,62} = 1.5$, $p = 0.23$) contributed significantly. The outcome was similar when tested with host genus instead of host subfamily. This indicates that diversity of the adult butterfly microbiota shows a seasonal increase over the course of the year, independent from temperature or host taxon turnover.

For the complete dataset from both countries, we used temperature as a universal environmental predictor. Although microbial load showed an apparent quadratic relationship with temperature (Supplementary Figure S2C), this effect was not significant after accounting for host subfamily (partial $R^2 = 0.02$, $F_{2,158} = 1.9$, $p = 0.16$), whereas host subfamily explained a significant proportion of the variation (partial $R^2 = 0.13$, $F_{6,158} = 3.9$, $p = 0.0011$). For all diversity estimates, we always accounted for host subfamily when evaluating the influence of temperature on microbial richness, Shannon diversity and inverse Simpson diversity. A full model including host subfamily and temperature explained 36% of the variation in Shannon diversity (adj. $R^2 = 0.36$, $F_{7,151} = 13.89$, $p < 10^{-13}$). The main effect could be attributed to host subfamily (partial $R^2 = 0.22$, $F_{6,151} = 7.2$, $p < 10^{-6}$), whereas temperature was not significant (partial $R^2 = 0.02$, $F_{1,151} = 3.2$, $p = 0.076$). Similar patterns were observed for microbial richness and inverse Simpson diversity (Supplementary Figure S2D–F), for which temperature was significantly associated with diversity, but explained substantially less variation than host subfamily. Notably, distribution of host subfamilies was strongly predicted by temperature ($R^2 = 0.48$, $F_{6,152} = 25$, $p < 10^{-15}$). This indicates that host taxonomic identity plays a larger role in shaping adult butterfly microbiota diversity than environmental temperature per se, though the occurrence of host taxa is temperature dependent.

3.3 Host taxonomic identity largely explains microbial community composition

Butterfly subfamilies showed significant differences in beta dispersion ($F_{6,160} = 7.94$, $p < 0.0001$), which was even more pronounced on host tribe level ($F_{11,155} = 15.55$, $p < 0.0001$) (Figure 2A). Differences among microbial community compositions were further tested by univariate PERMANOVA based on Bray-Curtis dissimilarities. As taxonomic ranks were all nested within each other, they were tested separately. Beforehand, host genera with less than two replicates were removed from the dataset. This showed increasing explanatory power (marginal R^2) with finer taxonomic resolution: Variance explained 3.6% by host family ($F_{1,159} = 5.8$), 18% by host subfamily ($F_{6,154} = 5.7$), 25% by host tribe ($F_{10,150} = 4.9$) (Figure 2B), 33% by host genus ($F_{17,143} = 4.1$) (Supplementary Figure S3A) and 40% by host species ($F_{28,132} = 3.1$, all $p < 0.0001$). Host genera differed significantly in beta dispersion ($F_{17,143} = 6.4$, $p < 0.0001$) (Supplementary Figure S3B), but there were no such differences observed between countries ($F_{1,159} = 1.2$, $p = 0.28$) or sampling locations ($F_{3,157} = 1.96$, $p = 0.12$). Univariate PERMANOVA on these geographic predictors showed overall lower explanatory power (marginal R^2): 6% by country ($F_{1,159} = 9.9$) and 8% by location ($F_{3,157} = 4.5$, all $p < 0.0001$).

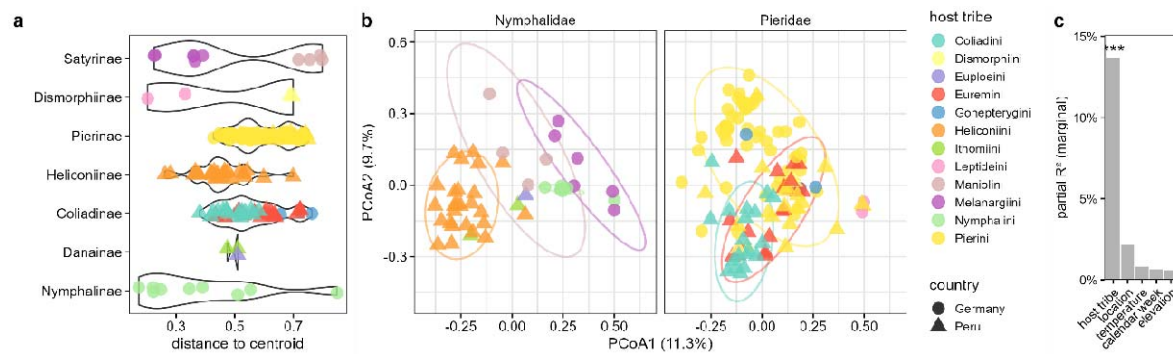


Figure 2 Host taxonomic identity largely explains dissimilarity of the adult butterfly microbiota composition

(a) Beta dispersion of the adult butterfly microbiota differs among host subfamilies (groups) and host tribes (colors). (b) Beta diversity of the adult butterfly microbiota using PCoA based on Bray-Curtis distances. Samples are colored by host tribe and panels faceted by host family for clarity. (c) Host tribe explained more variation (partial R^2) than environmental predictors and remained the strongest predictor of microbial community composition after excluding host groups with low replication ($n < 10$), retaining only Satyriinae, Pierinae, Heliconiinae, and Coliadinae.

When including host taxonomy, geographic and environmental predictors in the same multivariate model we used host tribe level to avoid collinearity among geographic predictors. As there was little autocorrelation between temperature, elevation and calendar week in the full dataset ($R^2 < 0.01$), we use them together in the same model. As further robustness measure we removed host groups with low replication ($n < 10$) or high multivariate dispersion (i.e. Dismorphiinae, Danainae and Nymphalinae). The full model with only Satyriinae, Pierinae, Heliconiinae, and Coliadinae and explained 26% of the variation in microbial community composition ($F_{12,134} = 3.9$, $p < 0.001$), while the majority was accounted to host tribe (marginal R^2 0.14, $F_{6,134} = 4.1$, $p < 0.001$) (Figure 2B,C), all other environmental factors did not contribute significantly to community composition (location: marginal R^2 0.02, $F_{3,134} = 1.3$, $p = 0.069$; elevation: marginal R^2 0.006, $F_{1,134} = 1.1$, $p = 0.35$; calendar week: marginal R^2 0.007, $F_{1,134} = 1.2$, $p = 0.25$; temperature: marginal R^2 0.008, $F_{1,134} = 1.5$, $p = 0.09$) (Figure 2C), indicating that the predominance of host taxonomy was not driven by sparsely replicated groups.

In addition, we performed variance partitioning to separate the unique contribution of host taxonomic identity and environmental or geographic predictors, as well as their shared fraction. For this we used again the reduced dataset containing only Satyrinae, Pierinae, Heliconiinae, and Coliadinae, and combined temperature, calendar week, elevation and location into a single environmental predictor. This confirmed the stronger contribution of host identity as a predictor of microbial community composition compared to environment influences. Host tribe showed a unique contribution (adj. R^2) of 10.3%, with a shared fraction of 5.3% and a unique contribution of the combined environmental predictors of only 4.8%. The effect in variance partitioning becomes more pronounced when using only temperature as environmental predictor (adj. R^2 of host tribe 12.8%, shared fraction 2.8%, unique temperature contribution 0.7%). Therefore, host taxonomic identity remained the strongest predictor in explaining the observed variation in microbial community composition than environmental and geographic factors combined.

3.4 General structure and composition of the adult butterfly microbiota

The microbial community composition of adult butterflies showed a high abundance and prevalence for lactic acid bacteria (LAB, order Lactobacillales rel. ab. 21.6%), enteric bacteria (order Enterobacterales rel. ab. 21.7%) and acetic acid bacteria (AAB, order Rhodospirillales rel. ab. 16.3%) (Figure 3A). Besides these, the adult butterfly microbiota contained host associated groups like Orbales, Rhizobiales & Rickettsiales (cumulative rel. ab. of 18.0%). The eight top prevalent bacterial orders made 85.6% rel. abundance of the bacterial community across all samples (Figure 3B).

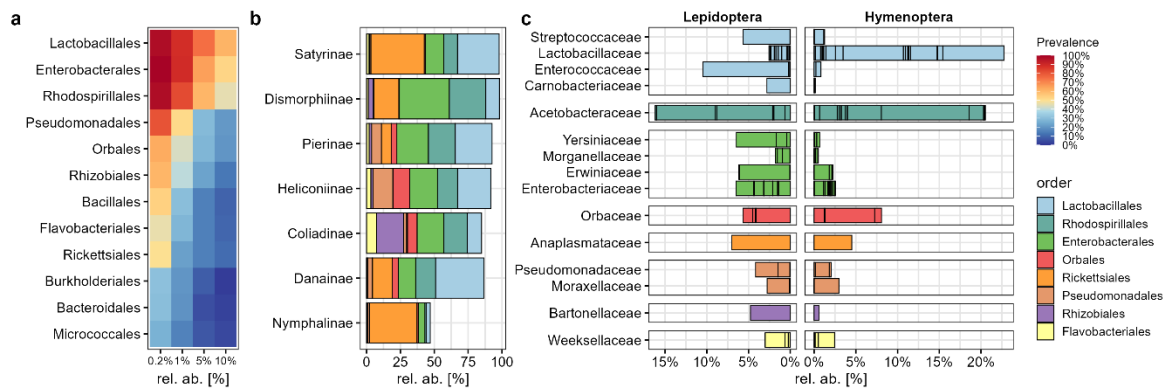


Figure 3 Compositional similarity of the adult butterfly microbiota across butterfly subfamilies and in comparison to bees

(a) Heatmap of most prevalent and abundant bacterial orders among all butterfly samples. (b) Microbial community composition grouped by butterfly subfamily showing the eight major bacterial orders which account for 85.6% relative abundances across all samples. (c) Relative abundances of major bacterial families from the microbiota of Lepidoptera compared to sympatric Hymenoptera from Peru (Apinae and Halictinae). Only top 16 families from butterflies are shown. Bars are colored by bacterial order and subdivided into bacterial genera.

Notably, several of the bacterial orders contained only a single bacterial family which was in turn often dominated by a single genus (Figure 3C). This was the case for insect host associated groups like Weeksellaceae (*Apibacter*), Bartonellaceae (*Bartonella*), Anaplasmatocaceae (*Wolbachia*) and Orbaceae (*Orbus*). Similar, Acetobacteraceae were dominated only by two genera (*Asaia* & *Commensalibacter*). Also, within the Lactobacillales most families were dominated by a single genus i.e. Enterococcaceae (*Enterococcus*), Streptococcaceae (*Lactococcus*) or Carnobacteriaceae (*Carnobacterium*). For enteric bacteria this pattern was less obvious, but some families contained also abundant genera i.e. Erwiniaceae (*Pantoea*) or Yersiniaceae (*Rahnella*), while others were more diverse.

By contrast, the family Lactobacillaceae showed only minor abundance in the Lepidoptera microbiota, with no dominant genus, but were the most abundant microbial family in the Hymenoptera dataset, comprising mainly *Apilactobacillus*, *Lactobacillus* and *Fructobacillus* (Figure 3C). Hymenoptera differed also in Acetobacteraceae, dominated by

Bombella alongside *Asaia* and *Commensalibacter*, and in Orbaceae with the main genus *Gilliamella*. Although adult butterflies shared several dominant microbial groups with wild bees, they showed varying proportions and substantial differences in the enriched bacterial genera (Figure 3C).

3.5 The common genus core microbiota of adult butterflies

To identify the most abundant and prevalent bacterial genera, we performed a threshold based core analysis on the entire dataset (minimum of 20% prevalence at 1% rel. ab.), which resulted in 12 common core genera enriched in adult butterflies (Supplementary Figure S4AB). These include the LAB *Enterococcus* (Enterococcaceae) & *Lactococcus* (Streptococcaceae) showing 68% and 37% prevalence, AAB *Asaia* & *Commensalibacter* (Acetobacteraceae) with 48% and 44% prevalence, enteric bacteria *Pantoea* (Erwiniaceae) & *Rahnella* (Yersiniaceae) with 41% and 28% prevalence, as well as *Bartonella* (Bartonellaceae) 33%, *Acinetobacter* (Moraxellaceae) 24%, *Orbus* (Orbaceae) 26%, *Wolbachia* (Anaplasmataceae) 20%, *Apibacter* (Weeksellaceae) 21% and *Entomomonas* (Pseudomonadaceae) with 20% prevalence, at 1% rel. abundance. (Supplementary Figure S4A–C). Notably, several of the remaining non-core genera belonged to the same bacterial groups e.g. LAB (*Carnobacterium*), enteric bacteria (*Cedecea*, *Raoultella*, *Serratia* and *Klebsiella*) or Orbacea (*Frischella*), but showed more host-specific occurrence in only individual butterfly genera (Supplementary Figure S4D) or certain butterfly subfamilies (Supplementary Figure S5).

To place these common core taxa into broader taxonomic context, they were shown together with the 80 most abundant genera in a phylogenetic tree. Common core genera stand out due to their high abundance and prevalence across all butterfly samples (Figure 4). For comparison, the same threshold was applied to the bee dataset, identifying 15 common core genera that showed high prevalence in the sympatric Hymenoptera (black dots in Figure 4, Supplementary Figure S6). Comparison of both core sets identified six genera that occurred only in the common core of Lepidoptera: *Orbus*, *Enterococcus*, *Rahnella*, *Entomomonas*, *Bartonella*, and *Lactococcus* (Supplementary Figure S7A). Six genera showed a taxonomic overlap between the common cores of Lepidoptera and

324 Hymenoptera: *Asaia*, *Pantoea*, *Commensalibacter*, *Wolbachia*, *Apibacter*, and
 325 *Acinetobacter* (Supplementary Figure S7B). The remaining nine genera occurred only in
 326 the common core of Hymenoptera: *Pseudomonas*, *Fructobacillus*, *Snodgrassella*,
 327 *Gilliamella*, *Lactobacillus*, *Bifidobacterium*, *Apilactobacillus*, *Bombella*, and *Pectinatus*
 328 (Supplementary Figure S7C).

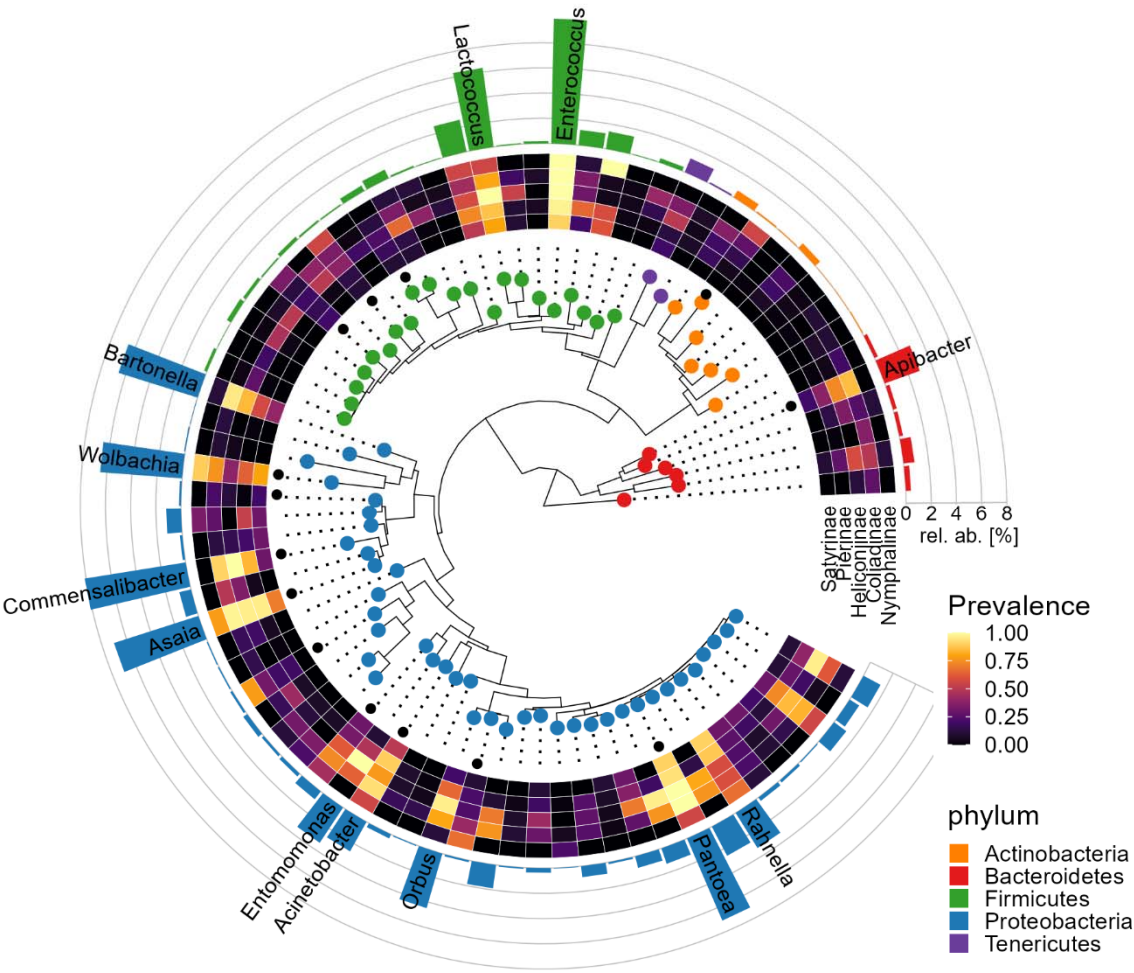


Figure 4 Taxonomic placement of the common core microbiota of adult butterflies

Phylogenetic grouping (SINA RAxML tree) and relative abundance of the 80 most abundant bacterial genera across all butterfly samples (represent about 98% of the total dataset). Only the common core taxa are shown with name, which show high prevalence in most butterfly subfamilies (only subfamilies with > 8 samples shown). Black dots indicate common core genera from a sympatric wild bee dataset for comparison, showing an overlap for *Apibacter*, *Wolbachia*, *Commensalibacter*, *Asaia*, *Acinetobacter*, and *Pantoea*. For a full genus-level annotation see Supplementary Figure S6.

3.6 Common core taxa distribution among countries and subfamilies

The distribution of the 12 common butterfly core genera was not equal between countries (Figure 5A). *Rahnella*, *Enterococcus*, and *Asaia* were both more abundant in samples from Germany (adj. $p < 0.05$), whereas *Wolbachia*, *Lactococcus*, and *Acinetobacter* were equally found in samples from both countries (adj. $p > 0.05$). On the other hand, *Orbus*, *Commensalibacter*, *Apibacter*, *Pantoea*, *Entomomonas*, and *Bartonella* were more abundant in samples from Peru (adj. $p < 0.001$) (Supplementary Figure S8). For some common core taxa, the same ASVs appear to be shared between samples from both countries (i.e. *Apibacter*, *Enterococcus*, *Lactococcus* *Orbus* *Pantoea* & *Rahnella*) (Supplementary Figure S9). Other common core taxa exhibited a mixture of country-specific and shared ASVs (i.e. *Asaia*, *Bartonella* & *Entomomonas*), with *Acinetobacter* & *Commensalibacter* showing the highest overall ASV diversity. Only *Wolbachia* showed clearly distinct ASVs in samples from both countries (Supplementary Figure S9). Among the 24 butterfly genera within this dataset *Wolbachia* was the most common endosymbiont and dominant in five different butterfly genera from five different subfamilies (Supplementary Figure S10). Others, like *Spiroplasma*, showed overall a minor abundance and more a loose association with individual samples from Peru but were absent in samples from Germany (Supplementary Figure S10).

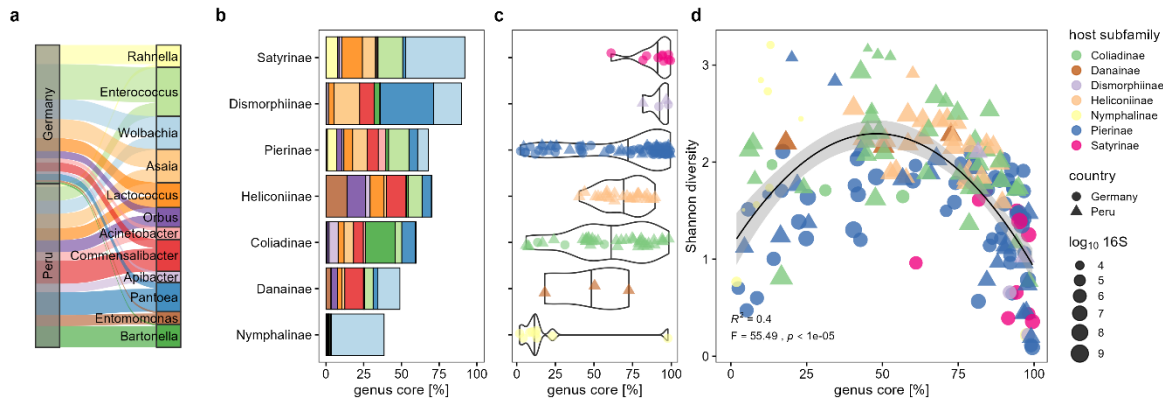


Figure 5 Abundance and distribution of the adult butterfly common genus core microbiota

(a) Country-specific distribution of the 12 common core genera. Samples from Germany (n = 76) had higher relative abundances of *Rahnella*, *Enterococcus*, and *Asaia*, while samples from Peru (n = 91) contained higher abundance of *Orbus*, *Commensalibacter*, *Apibacter*, *Pantoea*, *Entomomonas* & *Bartonella*. (b) Mean relative abundance of the 12 common core genera among butterfly subfamilies (average abundance of 67% across subfamilies). Color legend shared with panel A. (c) Distribution of the common genus core among individual butterfly samples within each subfamily (see Supplementary Figure S11 for details). (d) Genus core abundance and microbial load showed a quadratic relationship with Shannon diversity.

Subsetting the dataset to the 12 common core genera highlights their contribution to the total relative abundance across the different butterfly subfamilies (Figure 5B). The cumulative amount of common core genera ranged from 92% for Satyrinae, 90% for Dismorphiinae, 68% for Pierinae, 70% for Heliconiinae, 60% for Coliadinae, 49% for Danainae to 38% for Nymphalinae (Figure 5B). However, abundance distribution varied among individual samples within each host subfamily (Figure 5C). There was a high variation in core genus abundance among conspecific individuals, ranging from nearly 0% to 100% as seen for *Pieris* (Supplementary Figure S5A, Supplementary Figure S11A). The low biomass outgroup *Aglaia* (Nymphalinae) showed overall the lowest common core abundance, while *Polygonia* from the same subfamily showed even the highest, due to a high relative abundance of *Wolbachia* (Supplementary Figure S4C, Supplementary Figure S11A). The relative abundance of all common core genera varied significantly among

butterfly subfamilies (all genera KW adj. $p < 0.007$) (Supplementary Figure S11B). Some groups like *Bartonella* occurred mainly in Coliadinae, while *Entomomonas* seemed characteristic for Heliconiinae (Figure 5B) (Supplementary Figure S11B), which explains partly the country-specific distribution.

Interestingly, genus core abundance as well as microbial load both showed highly significant quadratic relationships with Shannon diversity (overall adj. $R^2 = 0.50$, $F_{4,162} = 42$, $p < 10^{-15}$). Both retained independent nonlinear associations when included in the same model, with genus core abundance explaining more variation (partial $R^2 = 0.39$, $F_{2,162} = 52$, $p < 10^{-15}$) than microbial load (partial $R^2 = 0.18$, $F_{2,162} = 18$, $p < 10^{-7}$) (Figure 5D). This result provides an unexpected link between total sample diversity, total microbial abundance, and abundance of common genus core taxa. Nonetheless, this pattern could provide a basis for the mechanistic understanding of the principles during community assembly in solitary pollinators.

4 Discussion

Our systematic assessment of the microbiota composition from adult butterfly populations across different continents revealed that adult Lepidoptera can assemble abundant gut microbiota distinct from those of Hymenoptera. The microbiota of adult butterflies showed conserved common core taxa shared across a variety of different butterfly subfamilies, with only partial overlap with those associated with sympatric bees. Our analysis showed that the variation of microbial communities was best explained by host taxonomic placement with only little environmental influence. Previous assumptions of uniform microbial community compositions within these mostly generalist nectar feeders do not hold true for adult butterflies⁴⁴.

4.1 Host taxonomic identity as best predictor of the adult butterfly microbiota

Our finding is in line with the observations by Ravenscraft, *et al.*³², who found a stronger influence of host species identity compared to host feeding preferences in Neotropical butterflies (with > 80% of samples from Nymphalidae including many fruit-feeding species). Similar, Hammer *et al.*³³ described phylogenetically structured microbiomes within Neotropical Heliconiinae, but found only a minor influence of pollen-feeding. Extending on these two previous studies, our dataset included Neotropical as well as Palearctic species and a broader taxonomic range with 70% of samples within the family Pieridae (subfamilies: Pierinae, Coliadinae & Dismorphiinae). This study represents the first cross-continental comparison of adult butterfly microbiota, and some taxonomic groups might still be underrepresented. Ongoing sampling efforts of local butterfly populations, as done within the framework of the LepEU consortium⁴⁵, will enable a targeted pan European comparison of selected host species. Still, we found a strong influence of host taxonomic placement, while our alternative environmental predictors based on biogeographic pattern, elevation or temperature showed only minor influence on the adult butterfly microbiota. Temperature is usually an important predictor of environmental microbial diversity⁴⁶, as well as insect abundance at the study site in Peru^{47,48}. Therefore, temperature appears to influence microbiota diversity of butterflies only indirectly, by affecting the presence of host taxa. Individual cold-adapted species

sampled from high elevations (*Catasticta* and *Tatochila*) or early in season (*Aglaia*) showed lower microbial richness and lower occurrence of common core taxa, with predominant colonization by generalist bacteria (*Pseudomonas* or *Bacillus*) (Supplementary Figure S4D). Microbial associations of insects appear to be less specific under colder conditions, fitting to similar observations made with mosquitoes from Greenland⁴⁹, or bees and butterflies from montane environments^{42,50}.

In contrast, environmental factors and geographic locations can be dominant drivers of microbiota composition in other model insect systems. A cross-continental survey of *Drosophila* populations showed that abiotic factors explain most of the variation in microbiome diversity⁵¹. *Drosophila* exhibit largely neutral microbiome assembly^{52,53} with limited influence of host phylogeny⁵⁴. However, their ecology differs fundamentally as fruit flies develop and proliferate on decaying fruits, which can be a reliable microbial pool when adults emerge. In contrast, butterflies undergo pronounced ontogenetic niche shift and nectar-feeding adults encounter a transient and variable distribution of nectar microbes⁵⁵. Nevertheless, even fruit-feeding and nectar-feeding species showed similar microbial richness and only minor differences in microbiota composition³². This contrasts findings in lepidopteran larvae, which contain transient, environmental strains which often resemble that of their local food plants¹⁶. Although adult butterflies in this study have been collected from different climate zones, they were colonized by a consistent set of common core taxa, typically found in insect gut environments. Remarkably, we found that several core taxa ASVs were identical in samples from both countries. This suggests that host-filtering is a dominant factor driving butterfly microbiota, resulting in the consistent enrichment of specific microbial groups from diverse environmental sources. At the same time, the widespread occurrence of these taxa indicates that they possess general traits that facilitate persistence and successful colonization of various insect gut environments.

4.2 Insect-associated bacteria as adult butterfly common core taxa

We identified twelve common core genera, all of which were also reported as prevalent groups in all previous studies of Neotropical butterflies^{28,32–35}. *Wolbachia* is probably the most common endosymbiont reported in Lepidoptera^{37,38} and showed host-

genus specific distribution due to its mainly vertical transmission mode. Other endosymbionts like *Spiroplasma* occurred in lower frequency in our dataset, but were also previously reported in *Heliconius*^{33–35}. Interestingly, *Heliconius* harbors either *Spiroplasma* or *Kinetoplastibacterium*, but never both within the same specimen (see Supplementary Figure S10). Another key taxon is *Orbus*, which was originally described as isolate from an adult butterfly *Sasakia charonda* (Apaturinae)³⁹. Later, the family of Orbaceae became mainly known as major gut symbiont of bees⁵⁶, but has recently been obtained as well from drosophilid and tephritid fruit flies⁵⁷. *Entomomonas* shows broad insect-associations and seems to be another characteristic core taxon of *Heliconius* butterflies^{33,58}.

4.3 Overlap and distinction of the lepidopteran and hymenopteran microbiota

The characteristic social core microbiota of corbiculate bees (Hymenoptera: Apinae) was, as expected, largely missing in solitary butterflies (i.e. *Bifidobacterium*, *Snodgrassella*, *Gilliamella* & *Lactobacillus*) (Supplementary Figure S7C). However, some common core genera observed in adult butterflies showed an overlap with the common core of sympatric hymenopteran pollinators: *Apibacter*, *Acinetobacter*, *Bartonella* and *Commensalibacter* are also commonly associated with bees. *Apibacter* and *Acinetobacter* are considered as environmentally acquired strains, but widely found in the microbiota of true social bees (*Apis*, *Bombus*)^{59,60}. In our analysis, they showed significantly higher abundance in Hymenoptera than in Lepidoptera (Supplementary Figure S7B). *Bartonella* (Alpha-1) and *Commensalibacter* (Alpha-2.1) occur in lower amounts in honey bees⁶¹, but can become abundant during overwintering⁶². Likewise, *Commensalibacter papalotli* has been isolated from overwintering monarch *Danaus plexippus* (Nymphalidae) from Mexico⁴⁰. Other *Commensalibacter* isolates have been obtained from different Nymphalidae and Pieridae butterflies in North America, Europe and Asia^{63–65}, indicating that this common core genus occurs across a broad range of butterfly species worldwide. Other AAB, such as *Asaia*, show generally very broad insect associations⁶⁶, whereas *Bombella* is characteristic of bees. Overall, AAB represent dominant members of the microbiota of Neotropical orchid bees (Euglossini) and stingless bees (Meliponini)^{3,42}. *Pectinatus* was the only genus in our dataset unique to bees and was entirely absent from butterflies (Supplementary Figure S7C). This genus has been mainly reported from

Meliponini^{3,67} and giant honey bees⁶⁸ and is considered a strict anaerobe. Its absence from adult butterflies suggests that their gut environment may favor more aerotolerant bacteria and may be less oxygen-depleted than the bee hindgut.

Another interesting distinction between the microbiota of Hymenoptera and Lepidoptera can be seen in lactic acid bacteria (LAB). Lactobacillaceae are a prominent bacterial group of social bees (*Apis*, *Bombus*, *Melipona*)^{67,69,70} and comprise a high diversity of genera, including *Lactobacillus* and *Apilactobacillus*. Ground nesting bees (Colletidae & Halictidae) can even be dominated by a single strain of *Apilactobacillus*^{71,72}. In contrast, these specialized fructophilic and acid-tolerant LAB are underrepresented and nearly absent from butterflies. Though it might be intriguing to hypothesize that Lactobacillaceae are acquired via pollen feeding, they were not particularly enriched in *Heliconius*, which represents the only pollen-feeding butterfly group in our dataset. Only *Tithorea* (Danainae) harbored considerable abundances of *Lactiplantibacillus*, *Secundilactobacillus*, and *Levilactobacillus*.

Instead, most adult butterflies show high abundance and prevalence for *Enterococcus* (Enterococcaceae) and *Lactococcus* (Streptococcaceae), which represent LAB with broader or more moderate pH tolerance, respectively. *Enterococcus* is a physiologically highly versatile bacterium and associated with a broad variety of different hosts giving them a character as generalist⁷³. They are also frequently reported as dominant member of the larval gut microbiota of certain moth families, providing potential antimicrobial and antiparasitic functions³¹. This makes *Enterococcus* a signature taxon with a prominent role in lepidopteran life cycle. *Enterococcus* is usually rare in Hymenoptera, while *Lactococcus* can occur in some bees. In stingless bees (*Melipona* spp.) a related, flower-associated genus *Floriccoccus* (Streptococcaceae) has been described recently^{67,69}.

4.4 Environmental acquisition and microbial community assembly in butterflies

Adult butterflies are exposed to a highly stochastic microbial pool of nectar microbes, driven by passive dispersal and random visitation of pollinators⁵⁵. Nevertheless, insects can acquire host-specific microbiota from ephemeral environmental sources. Ground-nesting

sweat bees (Halictidae) can selectively enrich a single *Apilactobacillus* strain from pollen provision, but will lose them under artificial rearing conditions⁷². Even obligate symbioses can evolve from environmental acquisition, which may make these associations more vulnerable to disturbances, while providing ecological flexibility that can foster host adaptations to changing environmental conditions^{74,75}. Unlike social insects, butterflies are exposed to environmental microbes immediately after emergence and lack persistent social interactions and direct transmission routes such as trophallaxis, resembling a lifestyle more comparable to that of solitary bees⁵. Nevertheless, aggregations at shared feeding sites may increase the probability of microbial exchange among conspecifics^{76,77}. Many butterfly species aggregate when performing mud puddling behavior to obtain additional salts and amino acids from moist soil, dung or even carcasses⁷⁸. This has been frequently observed in the tropics but also in temperate regions among various Pieridae and Nymphalidae, and is an intriguing but yet untested potential microbial source for adult butterflies³².

Nonetheless, the composition of the adult butterfly microbiota largely resembles that of insects feeding on sugar-rich substrates⁶⁶, and contains taxa associated with floral nectar, such as *Acinetobacter*, also common in other pollinators⁶⁰. We measured for most butterflies microbial loads around 10^6 to 10^8 16S copies which were in the range of previous reports from Heliconiini³³ or stingless bees³, while others have reported even higher amounts from butterflies³². Notably, similar microbial quantities have been reported during microbiome assembly in honey bees⁷⁹. The gut environment of nectar-feeding butterflies is probably highly selective, enriching microbes adapted to high sugar concentrations, while host-filtering others. Though, it cannot be excluded that other ecological or physiological factors may contribute to this (e.g. host immune parameter)⁴. The observed relationship between Shannon diversity and common genus core abundance (Figure 5D) seems, at first, surprising but could provide insights about the microbiota assembly process. Initial colonization may be driven by stochastic acquisition and priority effects⁴, resulting in communities dominated by non-core bacteria. Secondary microbial uptake may initially increase both microbial diversity and the representation of core taxa. With a transition toward a more established and core-dominated community, diversity may subsequently decline due to the increasing dominance of core taxa. Thus, core microbiota

dominance may provide an indicator of microbiome establishment that could complement microbial load measurements. Future studies with more extensive replication of individual butterfly species under comparable environmental settings will be required to disentangle the relative contributions of deterministic and stochastic processes during butterfly microbiome assembly. In addition, we identified a seasonal component, indicating that microbial diversity changes over the course of a year independent from temperature or host turnover. Considering ongoing climate-driven range shifts documented for butterflies globally⁸⁰, consequences for host-microbiome associations have yet to be determined.

The low bacterial biomass and almost lack of a common core microbiota in the early spring sampled *Aglaia* butterflies is not considered an artifact, but rather evidence that these butterflies have not yet acquired a host-specific microbiota and start ‘empty’ during their very first flights in March. This is consistent with the physiology of overwintering adults, as they usually purge their entire gut content before diapausing to avoid the presence of ice nucleators during hibernation⁸¹. Spore forming Bacilli are likely only inactive and transient gut members that can become dominant reads in metabarcoding studies when the amount of active bacteria is low⁸². But even *Aglaia* can assemble common core taxa as revealed in another study that sampled later in season (July – October)⁸³. Still, our data suggests that overwintering adults do not retain gut microbes from the previous year but rather acquire them newly when flying in spring.

4.5 Functional implications of the adult butterfly-associated microbiota

Even nectar-feeding adult butterflies need amino acid supplementation in their diet⁸⁴, which could be facilitated by symbiotic bacteria⁸⁵. Several enteric bacteria commonly found in insect guts have potential nitrogen fixing abilities, including the butterfly core members *Rahnella* & *Pantoea*⁸⁶. However, direct functional evaluations of the adult butterfly microbiota are rare. Antibiotic treatment shortened adult life span in the Indian Cabbage white *Pieris canidia* (Pieridae)⁸⁷, and a similar correlation between bacterial colonization and adult longevity was observed in *Speyeria mormonia* (Nymphalidae)⁶³. In *Pieris brassicae* (Pieridae), experimental depletion of the adult microbiota revealed an unexpected trans-generational phenotype on the offspring, influencing larval biomass gain

and immune responses during a host plant shift²⁹. Maternal and paternal effects on offspring performance are well documented across butterfly species⁸⁸, yet the potential role of adult microbial load in these phenomena has not been explicitly considered.

4.6 Conclusion and Outlook

Our findings provide insights into how host taxonomic identity and environmental factors shape the adult butterfly microbiota. It becomes obvious that even basic principles about the assembly and functional relevance of the adult butterfly microbiota remain poorly investigated aspects of Lepidopteran ecology. More extensive samplings across a broader taxonomic range among local populations could help to consolidate these findings and further disentangle environmental influences from host-specific effects. Our work provides a foundation to test how butterfly age or season influences core taxa assembly and diversity over time. Such work could help to evaluate the extent to which anthropogenic influences could potentially disturb adult butterfly microbiota in natural context.

5 Material and Methods

5.1 Sample collection and environmental temperature

Samples from Peru were collected between November 2022 and May 2023 within the Kosñipata valley from two main locations (provinces Paucartambo and Manu) as part of the ANDIV project (<https://www.andiv.biozentrum.uni-wuerzburg.de/>). Sampling sites include nine sublocations located along an elevational gradient reaching from 245 m up to 3350 m above sea level⁸⁹. Target groups were ‘whites’ and ‘sulfurs’ from the Pieridae family as well as *Heliconius* (Nymphalidae) as reference group. Butterflies were sampled by hand using a sweep net and directly dissected at local research stations. Entire abdomens were stored in DNA/RNA Shield buffer (Zymo Research) or in 100% Ethanol for preservation and transport at ambient temperatures. From each specimen a single leg was stored in 100% Ethanol for DNA barcoding and species identification at the Canadian Centre for DNA Barcoding (CCDB) in Guelph (Canada). The obtained barcodes of the cytochrome oxidase (COI) sequences are accessible on <https://portal.boldsystems.org/> (see Data availability

section). Wings and remaining body parts were desiccated and stored in glassine envelopes to aid with species identification. All butterflies were classified up to genus level, and if possible, to species level. For comparison we included a sympatric wild bee microbiota dataset from Peru⁴², which was processed under the same laboratory conditions. European butterfly species were collected near Nuremberg in middle Franconia in Germany. The two major sampling locations are approx. 26 km apart at 310 m and 350 m above sea level, respectively, and include four sublocations. Butterflies were collected between March and September in three consecutive years (2021, 2022 and 2023) by hand using a sweep net and identified to species level morphologically. Most samples were from two common species *Pieris napi* and *P. rapae* (Pierinae). Samples were stored at -20°C and abdomen dissected under laboratory conditions. Storage conditions had no influence on microbial community composition when accounting for host tribe and location within the same model (PERMANOVA Storage: partial $R^2 = 0.006$, $F_{1,146} = 1.17$, $p = 0.27$). A total of 199 butterfly samples were analyzed, 104 from Peru and 95 from Germany.

The average air temperature in the seven days before sample collection was taken as the sampling timepoint ‘temperature’ for each individual sample. The environmental temperatures at each sampling plot in Peru were directly measured with temperature loggers, showing a temperature gradient from 8.1°C to 26.1°C across the elevational gradient⁴⁸. For locations within Germany, measured temperatures for each sampling timepoint were taken from <https://meteostat.net>, showing a temperature gradient from 6.7°C to 24.8°C across season.

5.2 DNA Isolation, PCR conditions and library preparation

DNA isolation of entire abdomen was performed using ZymoBIOMICS™ 96 kits (Zymo Research). Bead beating of the ZR BashingBead™ lysis tubes was performed at 2500 rpm for 30 min on a multi-tube vortex mixer (Ohaus) and an additional Lysozyme digestion step with 50 U/μl Ready-Lyse Lysozym solution (Biozym) for 15 min at 25°C included. The ZymoBIOMICS Microbial Community Standard as well as ZymoBIOMICS Spike-in Control I (Zymo Research) was used as positive control. On each extraction plate a minimum of six negative extraction controls (NECs) were included. DNA isolation was

performed with an Integra Assist Plus pipetting robot placed under a vertical laminar air flow cabinet AirClean 600 (Syngene). DNA was eluted in a volume of 50 µl and stored at -20°C until library preparation.

The PCR for metabarcoding was performed according to the dual-indexing strategy by Kozich *et al.*⁹⁰ and the pipetting scheme from Sickel *et al.*⁹¹. Amplification target was the V4 region of the 16S SSU rRNA gene. The primers contained Illumina adaptor P5 (forward) or P7 (reverse), a unique 8 base barcode, a Pad sequence followed by the earth microbiome primer sequences 515F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT)⁹². PCR reactions were performed in triplicates, with 1 µl of DNA template in 10 µl final reaction volume using Thermo Phusion Plus PCR-Mastermix (Thermo Fisher Scientific) and 0.33 µM of each primer. The PCR protocol consisted of an initial activation at 98°C for 30 sec, followed by 30 cycles of 98°C for 10 sec, 55°C for 10 sec, 72°C for 30 sec and a final extension at 72°C for 5 min. PCR amplicons were assessed on precast E-Gel™ 96 Agarose Gels with SYBR Safe on a E-Gel Power Snap Plus Electrophoresis System (Thermo Fisher Scientific). Samples with no visible band on the agarose gel were subjected to a second PCR. Samples for which repeated amplification remained negative were later excluded during analysis (n = 17 samples). Exceptions were made for *Aglaia* samples that were kept on purpose as an outgroup within the dataset (see explanation below). Amplicons were normalized using SequalPrep™ Normalization Plate Kit (Thermo Fisher Scientific) and the plate pools were concentrated using AMPure XP Beads (Beckman Coulter). A final linearization PCR was performed with each plate pool library including primer based on the Illumina adapter sequences. Each 25 µl reaction contained 11 µl of plate pool library, 0.3 µM forward primer P5 (AATGATACGGCGACCACCGA), 0.3 µM reverse primer P7 (CAAGCAGAAGACGGCATACGA) and 12.5 µl PCR mix Thermo Phusion Plus PCR-Mastermix. The cycling conditions were as follows: 98°C for 3 min, followed by one cycle of 98°C for 1 min, 50°C for 1 min and 72°C for 5 min. Final products were cleaned by AMPure XP Beads and quantified using 1×dsDNA High Sensitivity Assay on a Qubit 4 Fluorometer (Thermo Fisher Scientific). Fragment size and library quality was checked on a 2100 Bioanalyzer using High Sensitivity DNA Chip (Agilent Technologies). The final

library was pooled equimolarly at 2 nM for paired-end sequencing and 5% PhiX was added to increase read quality. Sequencing was performed at the Genomics Service Unit of the LMU Biocenter on an Illumina MiSeq platform using Illumina Reagent Kit v2 (2 × 250bp).

5.3 Absolute quantification of 16S rRNA gene copy numbers by qPCR

Quantification of 16S rRNA gene copy numbers was performed using adapter-free versions of the 515F/806R primers. PCR reactions were performed in triplicates, with 1 µl DNA template per 10 µl final reaction on a CFX96 Touch Real-Time PCR Detection System (BioRad) using the Power Track™ SYBR Green Master Mix (Applied Biosystems). Cycling conditions followed a two-step protocol with initial activation at 95°C for 2 min, followed by 40 cycles of 95°C for 10 sec and a combined Annealing/Extension step at 60°C for 30 sec, including a final dissociation curve. Standard curves were generated from serial dilutions of the Femto Bacterial DNA Quantification Kit Standards (Zymo Research) ranging from 10 ng to 10 fg genomic DNA per reaction, corresponding to log₁₀ 16S rRNA gene copy numbers of 7.15 to 1.15. Standard curves showed coefficients of determination (R^2) > 0.99 and a mean amplification efficiency (E) of 80%. Absolute copy numbers per specimen were calculated considering a 2.5-fold dilution during DNA isolations and the final 50 µl elution volume. Non-template controls (NTCs) were used to determine the limit of detection (LOD), corresponding to a mean of 1.11 log₁₀ 16S copies per reaction (equivalent to 3.21 log₁₀ 16S copies per specimen).

5.4 Bioinformatics and metabarcoding sequence processing

Sequencing reads were processed using the custom metabarcoding pipeline available at https://github.com/chiras/metabarcoding_pipeline⁹³. Paired end forward and reverse reads were joined using VSEARCH v2.21.1⁹⁴ with a minimum overlap of 20 bp and a maximum of 10 mismatches. Reads were removed if shorter than 170 bp, contained ambiguous bases (N), or had a maximum expected error greater than 1 (maxEE = 1). Sequences were dereplicated and denoised to generate amplicon sequence variants (ASVs) using UNOISE⁹⁵, followed by singleton removal and chimera filtering⁹⁶. ASVs were mapped for iterative classification against the RDP (v18), Greengenes (v13.5) and SILVA

(v123) database using a global alignment identity threshold of 97%. Remaining reads without taxonomic allocation were hierarchically classified against the RDP (v18) database using SINTAX⁹⁷. Individual ASV sequences were manually validated using NCBI BLASTN v2.16.0 via the web interface and by the 16S-based identification available at <https://www.ezbiocloud.net/identify>. Data was further processed in R version 4.4.2⁹⁸ using a PHYLOSEQ v1.50.0 pipeline⁹⁹, including the packages MICROVIZ v0.12.5¹⁰⁰, GGPLOT2 v3.5.1¹⁰¹, DECONTAM v1.26.0¹⁰², VEGAN v2.6-8¹⁰³. Non-bacterial reads, including plant-derived chloroplast and mitochondrial sequences, and reads unassigned at phylum level were removed from the dataset, accounting for an average of 0.38% of total reads. Samples were excluded when non-bacterial reads exceed both 500 reads and 2% rel. abundance. ASVs with less than 60 reads (<0.01% of total reads) and genera below 600 reads (<0.1% of total reads) were filtered from the dataset, removing 1.1% of total reads. ASVs belonging to the microbial community standard used as positive controls were removed from the entire dataset. The resulting cleared positive controls were used together with negative extraction controls (NECs) to identify potential reagent contaminations using the prevalence based method implemented in the DECONTAM package¹⁰², resulting in an average removal of 0.13% of reads. Samples were evaluated for sequencing completeness rather than sequencing depth using the INEXT package v3.0.2¹⁰⁴, retaining only samples with high sample coverage (SC > 98%) for analysis

The final dataset contained 167 samples with 5,792,721 total reads and a median of 34,989 reads per sample, including 951 ASVs representing 146 bacterial genera. Final read depth and sample coverage per sample with all processing steps are provided in Supplemental Table S1. Differences in sequencing depth were accounted for by total sum scaling (TSS) to obtain relative read abundances (RRA) per sample. For most analyses ASVs were collated on genus level.

Aglaia samples showed overall the lowest sampling throughput (mean 5,245 reads) and highest proportions of filtered reads (average of 6.5% non-bacterial reads, 9.7% low abundant taxa and 3.1% removed by the DECONTAM package). Their microbiomes were mainly dominated by *Bacillus*. This was indicative for a low microbial biomass, which was confirmed by qPCR. Still, *Aglaia* samples were intentionally retained in the dataset for

comparative purposes as they represent an outgroup sampled at very low temperatures from an exceptionally early sampling time point. Due to their low microbial biomass and consequently low sequencing depth, they were excluded from diversity estimates to avoid any depth-related bias.

The combined Lepidoptera-Hymenoptera dataset was processed similarly. As bee samples showed substantially higher proportions of plant-associated reads, samples were excluded only when non-bacterial reads exceeded both 5,000 reads and 20% rel. abundance. The final dataset contained 872 bee samples from the subfamilies Apinae (Apini n = 69, Bombini n = 45, Euglossini n = 139, Meliponini n = 415) and Halictinae (n = 204), mainly from the tribe Augochlorini.

5.5 Core microbiota analysis and phylogenetic placement

Common core genera were defined based on abundance and prevalence thresholds using the `core_members` function implemented in the `MICROBIOME` package v1.28.0¹⁰⁵. Given the heterogeneous assemblage of butterflies in our dataset (35 species from 24 genera, 12 tribes and 7 subfamilies), we used less stringent and more general definitions compared to other studies on single social species. A medium abundance threshold was chosen as an intermediate core definition (minimum relative abundance of 1% with prevalence of 20%), which resulted in 12 common core genera across all butterfly samples (Supplementary Figure S4B). Similar results were obtained using a high abundance definition (5% rel. abundance and 10% prevalence) which resulted in a subset of 11 common core genera. The low abundance definition (0.2% rel. abundance with 40% prevalence) recovered 10 of the 12 common core members. For direct comparison, we used the same operational definition (1% rel. abundance with 20% prevalence) for the Hymenoptera dataset, identifying 15 common core genera among the 872 sympatric bee samples representing 34 genera, 6 tribes, and 2 subfamilies.

To build the phylogenetic tree of the butterfly microbiota, the dataset was simplified to the 80 most abundant bacterial genera meeting a minimum prevalence of 5% and minimum abundance of 2000 reads, accounting for 98% of total reads in the dataset. The most

abundant ASV of each genus was exported as representative sequence using BIOSTRINGS v 2.74.0¹⁰⁶ and APE v 5.8¹⁰⁷. ASV sequences were combined with a custom reference set and imported into SINA aligner v1.2.12 and aligned against neighboring query reference sequences from the SILVA database using minimum identity threshold of 0.95 (<https://www.arb-silva.de/aligner/>). A maximum-likelihood tree was constructed in RAXML using a GTR substitution model and Gamma likelihood. The resulting tree was re-imported into R, reference sequences were removed, and the tree was rooted to Chitinophagaceae as outgroup using GGTREEEXTRA v 1.16.0¹⁰⁸.

5.6 Statistical analysis

As predictors for microbial alpha and beta diversity, we tested nested host taxonomic ranks individually (i.e. host ‘family’, ‘subfamily’ or ‘tribe’). To avoid model overfitting most tests were done with these higher taxonomic ranks or on a reduced dataset where unique genera have been removed. Collinearity among all predictors was assessed in each model using the variance inflation factor ($VIF < 3$), and each model was checked for aliased coefficients. As butterfly ‘genus’ was structurally confounded with ‘country’ (2 levels) and ‘sublocation’ (13 level), we did not include them in the same model. To account for this, we primarily used only higher taxonomic ranks together with ‘location’ (4 levels) as the main geographic predictor, which includes two major sampling locations per country (Paucartambo $n = 67$, Manu $n = 24$, Nuremberg $n = 70$ and Hersbrucker Alb $n = 6$). We used ‘temperature’ as a unified environmental predictor as well as ‘elevation’ above sea level for the Peru dataset and sampling timepoint (‘calendar week’) for the Germany dataset.

For alpha diversity estimations, all retained samples met the minimum sample coverage threshold ($SC > 98\%$). We calculated observed richness, the Shannon index (H') and inverse Simpson index, corresponding to Hill numbers of orders $q = 0, 1$, and 2 , respectively¹⁰⁹. Predictors were tested with linear models (lm) and the Anova function with Type II sums of squares from the CAR package v3.1-3¹¹⁰. All tests were two-sided. Model residuals were assessed for normality by Shapiro-Wilk tests and Q-Q plots. Adjusted R^2 values were used to quantify overall model fit and multiple predictors were retained in the

models if they improved support according to Akaike Information Criterion (AIC). The contribution of each individual predictor was estimated using Type II sums of squares, and expressed as partial R^2 , verified by partial eta-squared values from the EFFECTSIZE package v1.0.1¹¹¹.

Beta diversity was visualized using Principal Coordinates Analysis (PCoA) based on Bray-Curtis distances from relative abundance data. In the combined Lepidoptera-Hymenoptera dataset, non-metric multidimensional scaling (NMDS) was based on Bray-Curtis dissimilarities of square-root-transformed relative abundance data. Differences in community composition were tested using a PERMANOVA based on a Bray-Curtis distance matrix via the `adonis2` function, and by permutational tests on group dispersion (distance to the centroid), both with 999 permutations as implemented in the VEGAN package¹⁰³. PERMANOVA was conducted using marginal (Type III-like) sums of squares to assess the contribution of each predictor (partial R^2) after accounting for all other terms in the model. To evaluate the unique and shared fractions of explained variation, we performed variance partitioning on a Hellinger transformed community matrix using the `varpart` function from the VEGAN package. This provides the unique fraction explained by host factors versus environmental factors (adjusted R^2) as well as the variation shared among predictors. As a robustness analysis we used a reduced dataset retaining only well replicated subfamilies ($n \geq 10$). Differences in abundance distribution of common core taxa among host groups or countries was assessed using individual Wilcoxon rank-sum Tests or Kruskal-Wallis-Tests with Benjamini-Hochberg correction accounting for multiple comparisons.

6 Data availability

The raw sequencing reads generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1426605 (<https://www.ncbi.nlm.nih.gov/bioproject/1426605>). The obtained barcodes of butterfly species are accessible via the Bold Systems portal (<https://portal.boldsystems.org/>) via the process IDs: ANDIV1591-23 to ANDIV1615-23, ANDIV1901-23 to ANDIV1928-23 and ANDIV11933-24 to ANDIV11983-24. Bee microbiome data from Pinos *et al*⁴² is available under BioProject: PRJNA1294982.

7 Code availability

The full analysis pipeline, including all code and data necessary to reproduce the 16S sample processing, figures and statistical analyses, is available on GitHub (<https://github.com/micromole/butterfly-core-microbiota>).

8 Acknowledgement

The authors would like to thank Uschi Schkölziger and Anika Schopf for technical help during DNA isolation, library preparation and qPCR as well as Andreas Brachmann for sequencing. We also thank Constantin Stefan for assistance with field sampling in Peru. We are grateful to the National Service of Natural Protected Areas (SERNANP) and the National Forestry and Wildlife Service (SERFOR) of Peru for permissions for field work. Samples have been collected under the SERFOR permit number AUT-IFS-2022-102 (N° D000178-2022-MIDAGRI-SERFOR-DGGSPFFS-DGSPFS) as well as by SERNAP permit N°011-2022-SERNANP-JEF. Export was done in accordance with permit N° 003756-SERFOR.

9 Funding

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project numbers 460916175; 546907793. Fieldwork was conducted as part of the ANDIV project (KE 1743/12-1) and sample analysis supported by an LMUexcellent Postdoc Support Fund and WE 5403/3-1 to AW.

10 Author contributions

Conceptualization, A.W.; Methodology, A.W., A.K.; Investigation, A.W., A.P., Y.C.-C., K.L.H., P.A.-A., F.Y.; Sample acquisition, A.W., A.P., Y.C.-C.; Data curation, A.W., K.L.H.; Data analysis, A.W.; Writing – Original Draft, A.W.; Writing –Review & Editing, A.W., A.P., Y.C.-C., K.L.H., P.A.-A., F.Y., I.S.-D., M.K.P., G.B., A.K.; Funding Acquisition, A.W., I.S.-D., M.K.P., G.B., A.K.; Resources, A.K., F.Y.

11 References

1. Cullen, N. P., Fethers, A. M. & Ashman, T. L. Integrating microbes into pollination. *Curr. Opin. Insect Sci.* **44**, 48–54 (2021).
2. Steffan, S. A. *et al.* Microbes, the ‘silent third partners’ of bee–angiosperm mutualisms. *Trends Ecol. Evol.* **39**, 65–77 (2024).
3. Kwong, W. K. *et al.* Dynamic microbiome evolution in social bees. *Sci. Adv.* **3**, 1–17 (2017).
4. Mazel, F., Prasad, A. & Engel, P. Host specificity of gut microbiota associated with social bees: patterns and processes. *Microbiol. Mol. Biol. Rev.* **89**, 715–735 (2025).
5. Voulgari-Kokota, A., McFrederick, Q. S., Steffan-Dewenter, I. & Keller, A. Drivers, Diversity, and Functions of the Solitary-Bee Microbiota. *Trends Microbiol.* **27**, 1034–1044 (2019).
6. McFrederick, Q. S. *et al.* Flowers and Wild Megachilid Bees Share Microbes. *Microb. Ecol.* **73**, 188–200 (2017).
7. Keller, A. *et al.* (More than) Hitchhikers through the network: the shared microbiome of bees and flowers. *Curr. Opin. Insect Sci.* **44**, 8–15 (2021).
8. Olson, M. M., Ravikanthachari, N., Blackwell, M. & Boggs, C. L. The dispersal of microbes among and within flowers by butterflies. *Ecol. Entomol.* **48**, 458–465 (2023).
9. Argueta-Guzmán, M., Spasojevic, M. J. & McFrederick, Q. S. Solitary Bees Acquire and Deposit Bacteria via Flowers: Testing the Environmental Transmission Hypothesis Using *Osmia lignaria*, *Phacelia tanacetifolia*, and *Apilactobacillus micheneri*. *Ecol. Evol.* **15**, e71138 (2025).
10. Rech de Laval, V., Dainat, B., Engel, P. & Robinson-Rechavi, M. The BeeBiome data portal provides easy access to bee microbiome information. *BMC Bioinformatics* **26**, 198 (2025).
11. Wei, N., Russell, A. L., Jarrett, A. R. & Ashman, T. L. Pollinators mediate floral microbial diversity and microbial network under agrochemical disturbance. *Mol. Ecol.* **30**, 2235–2247 (2021).
12. Li, J. *et al.* Divergence and convergence of gut microbiomes of wild insect pollinators. *MBio* **14**, e0127023 (2023).
13. Requier, F. *et al.* Bee and non-bee pollinator importance for local food security. *Trends Ecol. Evol.* **38**, 196–205 (2023).
14. Diniz, U. M. *et al.* The neglected pollinators: settling moths are keystone floral visitors essential to network connectivity and tropical forest recovery. *Proc. R. Soc. B Biol. Sci.* **292**, (2025).

- 863 15. Robinson, C. J., Schloss, P., Ramos, Y., Raffa, K. & Handelsman, J. Robustness of
864 the bacterial community in the cabbage white butterfly larval midgut. *Microb. Ecol.*
865 **59**, 199–211 (2010).
- 866 16. Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P. & Fierer, N. Caterpillars
867 lack a resident gut microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9641–9646
868 (2017).
- 869 17. Engel, P. & Moran, N. A. The gut microbiota of insects – diversity in structure and
870 function. *FEMS Microbiol. Rev.* **37**, 699–735 (2013).
- 871 18. Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M. & Fatouros, N. E. Bacterial
872 symbionts in lepidoptera: Their diversity, transmission, and impact on the host.
873 *Front. Microbiol.* **9**, 1–14 (2018).
- 874 19. Moran, N. A., Ochman, H. & Hammer, T. J. Evolutionary and Ecological
875 Consequences of Gut Microbial Communities. *Annu. Rev. Ecol. Evol. Syst.* **50**, 451–
876 475 (2019).
- 877 20. Kucuk, R. A. Gut bacteria in the holometabola: A review of obligate and facultative
878 symbionts. *J. Insect Sci.* **20**, 1–12 (2020).
- 879 21. Chen, B. *et al.* Biodiversity and Activity of the Gut Microbiota across the Life
880 History of the Insect Herbivore *Spodoptera littoralis*. *Sci. Rep.* **6**, 29505 (2016).
- 881 22. Staudacher, H. *et al.* Variability of bacterial communities in the moth *Heliothis*
882 *virescens* indicates transient association with the host. *PLoS One* **11**, 1–21 (2016).
- 883 23. González-Serrano, F. *et al.* The Gut Microbiota Composition of the Moth *Brithys*
884 *crini* Reflects Insect Metamorphosis. *Microb. Ecol.* **79**, 960–970 (2020).
- 885 24. Hammer, T. J., Sanders, J. G. & Fierer, N. Not all animals need a microbiome.
886 *FEMS Microbiol. Lett.* **366**, 1–11 (2019).
- 887 25. Poolprasert, P. *et al.* Light Microscopic Observations of the Mesentero-Proctodeal
888 Regions in Adult *Catopsilla Pomona* (Fabricius, 1758) (Lepidoptera: Pieridae).
889 *Suranaree J. Sci. Technol.* **22**, 99–103 (2015).
- 890 26. Waterhouse, D. F. The hydrogen ion concentration in the alimentary canal of larval
891 and adult lepidoptera. *Aust. J. Biol. Sci.* **2**, 428–437 (1949).
- 892 27. Gaikwad, S. M. & Bhawane, G. P. Partial characterization of midgut enzymes in
893 butterfly *Papilio polytes polytes* L. (Lepidoptera: papilionidae). *An Int. Q. J. Life Sci.*
894 **10**, 43–53 (2015).
- 895 28. Hammer, T. J., McMillan, W. O. & Fierer, N. Metamorphosis of a Butterfly-
896 Associated Bacterial Community. *PLoS One* **9**, e86995 (2014).
- 897 29. Paniagua Voirol, L. R., Weinhold, A., Johnston, P. R., Fatouros, N. E. & Hilker, M.
898 Legacy of a butterfly's parental microbiome in offspring performance. *Appl.*

- 899 *Environ. Microbiol.* **86**, 1–16 (2020).
- 900 30. Sanaei, E. *et al.* Microbiome analysis of monarch butterflies reveals effects of
901 development and diet. *FEMS Microbiol. Ecol.* **100**, 1–11 (2024).
- 902 31. Shao, Y., Mason, C. J. & Felton, G. W. Toward an Integrated Understanding of the
903 Lepidoptera Microbiome. *Annu. Rev. Entomol.* **69**, 326–342 (2024).
- 904 32. Ravenscraft, A., Berry, M., Hammer, T., Peay, K. & Boggs, C. Structure and
905 function of the bacterial and fungal gut microbiota of Neotropical butterflies. *Ecol.*
906 *Monogr.* **89**, 1–25 (2019).
- 907 33. Hammer, T. J., Dickerson, J. C., McMillan, W. O. & Fierer, N. Heliconius
908 Butterflies Host Characteristic and Phylogenetically Structured Adult-Stage
909 Microbiomes. *Appl. Environ. Microbiol.* **86**, 1–14 (2020).
- 910 34. Luna, N. *et al.* Geography shapes the microbial community in Heliconius butterflies.
911 *FEMS Microbiol. Ecol.* **99**, 1–8 (2023).
- 912 35. Khazan, E. S. *et al.* Abdominal microbiome composition and diversity of two
913 Heliconius species (Lepidoptera: Nymphalidae) in the Colombian Andes. *Trop.*
914 *Lepid. Res.* **33**, 70–76 (2023).
- 915 36. Risely, A. Applying the core microbiome to understand host–microbe systems. *J.*
916 *Anim. Ecol.* **89**, 1549–1558 (2020).
- 917 37. Ahmed, M. Z., Araujo-Jnr, E. V., Welch, J. J. & Kawahara, A. Y. Wolbachia in
918 butterflies and moths: Geographic structure in infection frequency. *Front. Zool.* **12**,
919 1–9 (2015).
- 920 38. Duploux, A., Pranter, R., Warren-Gash, H., Tropek, R. & Wahlberg, N. Towards
921 unravelling Wolbachia global exchange: a contribution from the Bicyclus and
922 Mylothris butterflies in the Afrotropics. *BMC Microbiol.* **20**, 1–9 (2020).
- 923 39. Kim, J. Y. *et al.* *Orbus sasakiae* sp. nov., a bacterium isolated from the gut of the
924 butterfly *Sasakia charonda*, and emended description of the genus *Orbus*. *Int. J. Syst.*
925 *Evol. Microbiol.* **63**, 1766–1770 (2013).
- 926 40. Servín-Garcidueñas, L. E., Sánchez-Quinto, A. & Martínez-Romero, E. Draft
927 genome sequence of *Commensalibacter papalotli* MX01, a symbiont identified from
928 the guts of overwintering monarch butterflies. *Genome Announc.* **2**, 0–1 (2014).
- 929 41. Jones, M. *et al.* Recovering ecological interactions by mining non-target data from
930 whole genome re-sequencing projects. *bioRxiv* 0–3 (2025)
931 doi:10.1101/2025.01.17.633498.
- 932 42. Pinos, A. *et al.* Host identity, more than elevation, shapes bee microbiomes along a
933 tropical elevation gradient. *Front. Microbiol.* **16**, 1–14 (2025).
- 934 43. Carvalho, A. P. S. *et al.* Comprehensive phylogeny of Pieridae butterflies reveals

- 935 strong correlation between diversification and temperature. *iScience* **27**, (2024).
- 936 44. Phalnikar, K., Kunte, K. & Agashe, D. Dietary and developmental shifts in butterfly-
937 associated bacterial communities. *R. Soc. Open Sci.* **5**, 171559 (2018).
- 938 45. Beldade, P. *et al.* LepEU²: A Consortium to Study the Population Genomics of
939 Butterflies and Moths Across Europe. 1–8 (2026) doi:10.1111/eva.70246.
- 940 46. Nottingham, A. T. *et al.* Microbes follow Humboldt: temperature drives plant and
941 soil microbial diversity patterns from the Amazon to the Andes. *Ecology* **99**, 2455–
942 2466 (2018).
- 943 47. Holzmann, K. L. *et al.* Temperature boosts and constrains dung beetle diversity
944 along an Andean-Amazonian elevation gradient. *Proc. R. Soc. B Biol. Sci.* **293**,
945 (2026).
- 946 48. Holzmann, K. L. *et al.* Cold waves in the Amazon rainforest and their ecological
947 impact. *Biol. Lett.* **21**, 1–6 (2025).
- 948 49. Rojas-Guerrero, D. *et al.* Cold and lonely: low microbial abundance and no core
949 microbiota in mosquito populations across Greenland. *bioRxiv* (2024)
950 doi:10.1101/2024.08.27.609970.
- 951 50. Taş, P., Mouly, A. & Lucek, K. A microbial view on secondary contact between two
952 Alpine butterflies. *BMC Ecol. Evol.* **26**, 22 (2026).
- 953 51. Gale, J. T. *et al.* Environment and diet shape the geography-specific *Drosophila*
954 *melanogaster* microbiota composition. *Appl. Environ. Microbiol.* **91**, e00883-25
955 (2025).
- 956 52. Adair, K. L., Wilson, M., Bost, A. & Douglas, A. E. Microbial community assembly
957 in wild populations of the fruit fly *Drosophila melanogaster*. *ISME J.* **12**, 959–972
958 (2018).
- 959 53. Henry, L. P. & Ayroles, J. F. *Drosophila melanogaster* microbiome is shaped by
960 strict filtering and neutrality along a latitudinal cline. *Mol. Ecol.* **31**, 5861–5871
961 (2022).
- 962 54. Wong, A. C. N., Chaston, J. M. & Douglas, A. E. The inconstant gut microbiota of
963 *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J.* **7**, 1922–1932
964 (2013).
- 965 55. de Vega, C. *et al.* The role of plant–pollinator interactions in structuring nectar
966 microbial communities. *J. Ecol.* **109**, 3379–3395 (2021).
- 967 56. Kwong, W. K. & Moran, N. A. Cultivation and characterization of the gut symbionts
968 of honey bees and bumble bees: Description of *Snodgrassella alvi* gen. nov., sp.
969 nov., a member of the family Neisseriaceae of the betaproteobacteria, and
970 *Gilliamella apicola* gen. nov., sp. nov., a memb. *Int. J. Syst. Evol. Microbiol.* **63**,
971 2008–2018 (2013).

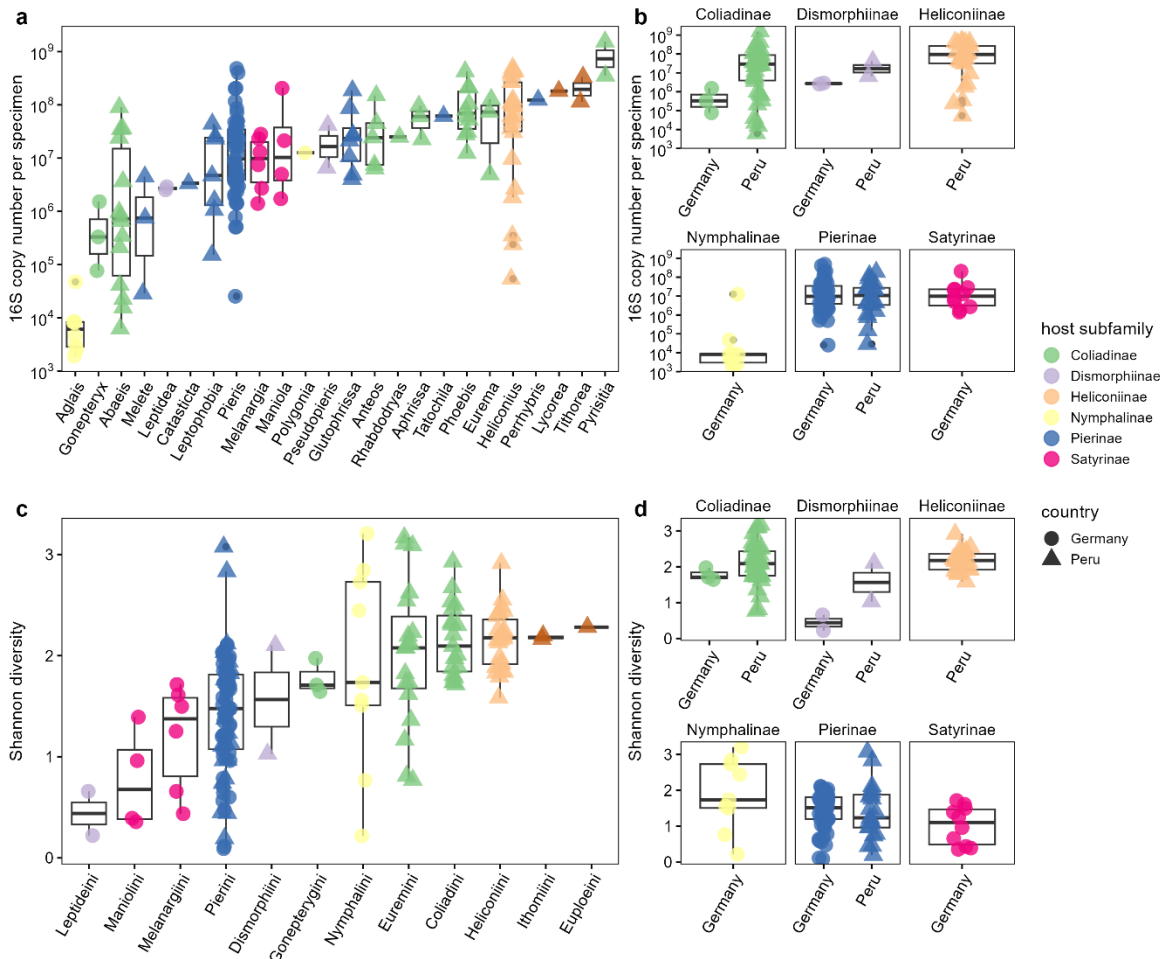
- 972 57. Phillips, L. E., Sotelo, K. L. & Moran, N. A. Characterization of gut symbionts from
973 wild-caught *Drosophila* and other Diptera: description of *Utexia brackfieldae* gen.
974 nov., sp. nov., *Orbus sturtevantii* sp. nov., *Orbus wheelerorum* sp. nov., and *Orbus*
975 *mooreae* sp. nov. *Int. J. Syst. Evol. Microbiol.* **74**, (2024).
- 976 58. Wang, J. *et al.* *Entomomonas moraniae* gen. Nov., sp. nov., a member of the family
977 *Pseudomonadaceae* isolated from asian honey bee gut, possesses a highly reduced
978 genome. *Int. J. Syst. Evol. Microbiol.* **70**, 165–171 (2020).
- 979 59. Praet, J. *et al.* *Apibacter mensalis* sp. Nov.: A rare member of the bumblebee gut
980 microbiota. *Int. J. Syst. Evol. Microbiol.* **66**, 1645–1651 (2016).
- 981 60. Alvarez-Perez, S. *et al.* *Acinetobacter pollinis* sp. nov., *Acinetobacter baretiae* sp.
982 nov. and *Acinetobacter rathckeae* sp. nov., isolated from floral nectar and honey
983 bees. *Int. J. Syst. Evol. Microbiol.* **71**, 004783 (2021).
- 984 61. Kešnerová, L., Moritz, R. & Engel, P. *Bartonella apis* sp. nov., a honey bee gut
985 symbiont of the class Alphaproteobacteria. *Int. J. Syst. Evol. Microbiol.* **66**, 414–421
986 (2016).
- 987 62. Kešnerová, L. *et al.* Gut microbiota structure differs between honeybees in winter
988 and summer. *ISME J.* **14**, 801–814 (2020).
- 989 63. Ravenscraft, A., Kish, N., Peay, K. & Boggs, C. No evidence that gut microbiota
990 impose a net cost on their butterfly host. *Mol. Ecol.* **28**, 2100–2117 (2019).
- 991 64. Botero, J. *et al.* A phylogenomic and comparative genomic analysis of
992 *Commensalibacter*, a versatile insect symbiont. *Anim. Microbiome* **5**, 25 (2023).
- 993 65. Yukphan, P., Charoenyingcharoen, P., Muangham, S., Mhuanong, W. & Yamada,
994 Y. *Commensalibacter nepenthis* sp. nov. and *Commensalibacter oyaizuii* sp. nov.,
995 from fluid in a *Nepenthes* pitcher cup and a butterfly (*Junonia lemonias*) in Thailand.
996 *Int. J. Syst. Evol. Microbiol.* **74**, (2024).
- 997 66. Crotti, E. *et al.* Acetic acid bacteria, newly emerging symbionts of insects. *Appl.*
998 *Environ. Microbiol.* **76**, 6963–6970 (2010).
- 999 67. Haag, K. L. *et al.* Conserved gut microbiomes with cross-species spillover between
1000 sympatric Neotropical stingless bees and honey bees. *Appl. Environ. Microbiol.* **92**,
1001 1–19 (2026).
- 1002 68. Prasad, A. *et al.* Evolution of gut microbiota across honeybee species revealed by
1003 comparative metagenomics. *Nat. Commun.* **16**, 1–15 (2025).
- 1004 69. Cerqueira, A. E. S. *et al.* Extinction of anciently associated gut bacterial symbionts
1005 in a clade of stingless bees. *ISME J.* **15**, 2813–2816 (2021).
- 1006 70. Weinhold, A., Grüner, E. & Keller, A. Bumble bee microbiota shows temporal
1007 succession and increase of lactic acid bacteria when exposed to outdoor
1008 environments. *Front. Cell. Infect. Microbiol.* **14**, 1–14 (2024).

- 1009 71. Hammer, T. J. *et al.* Bee breweries: The unusually fermentative, lactobacilli-
1010 dominated brood cell microbiomes of cellophane bees. *Front. Microbiol.* **14**, 1–16
1011 (2023).
- 1012 72. Mayr, A. V., Weinhold, A., Nolzen, A., Keller, A. & Ayasse, M. The neonicotinoid
1013 Acetamiprid alters the chemical profile of the primitive eusocial bee *Lasioglossum*
1014 *malachurum*. *Ecotoxicol. Environ. Saf.* **306**, 119311 (2025).
- 1015 73. Schwartzman, J. A. *et al.* Global diversity of enterococci and description of 18
1016 previously unknown species. *Proc. Natl. Acad. Sci. U. S. A.* **121**, 1–12 (2024).
- 1017 74. Hartmann, A. C., Baird, A. H., Knowlton, N. & Huang, D. The Paradox of
1018 Environmental Symbiont Acquisition in Obligate Mutualisms. *Curr. Biol.* **27**, 3711-
1019 3716.e3 (2017).
- 1020 75. Ravenscraft, A. & Coon, K. L. Transient Microbes in Insects: Fleeting but
1021 Functional. *Annu. Rev. Entomol.* **71**, 253–273 (2026).
- 1022 76. Onchuru, T. O., Javier Martinez, A., Ingham, C. S. & Kaltenpoth, M. Transmission
1023 of mutualistic bacteria in social and gregarious insects. *Curr. Opin. Insect Sci.* **28**,
1024 50–58 (2018).
- 1025 77. Weinhold, A. Bowel Movement: Integrating Host Mobility and Microbial
1026 Transmission Across Host Taxa. *Front. Microbiol.* **13**, 1–8 (2022).
- 1027 78. Lamie, E., Morton, E. R. & Parzer, H. F. Puddling in butterflies: current knowledge
1028 and new directions. *Ann. Entomol. Soc. Am.* **118**, 110–118 (2025).
- 1029 79. Powell, J. E., Martinson, V. G., Urban-Mead, K. & Moran, N. A. Routes of
1030 acquisition of the gut microbiota of the honey bee *Apis mellifera*. *Appl. Environ.*
1031 *Microbiol.* **80**, 7378–7387 (2014).
- 1032 80. Chowdhury, S. *et al.* Extensive climate-induced range shifts in butterflies across the
1033 globe. *Nat. Ecol. Evol.* 1–14 (2026) doi:10.1038/s41559-026-03117-y.
- 1034 81. Pullin, A. S. & Bale, J. S. Effects of low temperature on diapausing *Aglais urticae*
1035 and *Inachis io* (Lepidoptera: Nymphalidae): Cold hardiness and overwintering
1036 survival. *J. Insect Physiol.* **35**, 277–281 (1989).
- 1037 82. Mugo-Kamiri, L., Querejeta, M., Raymond, B. & Herniou, E. A. The effect of diet
1038 composition on the diversity of active gut bacteria and on the growth of *Spodoptera*
1039 *exigua* (Lepidoptera: Noctuidae). *J. Insect Sci.* **24**, 13 (2024).
- 1040 83. Manthey, C., Johnston, P. R., Nakagawa, S. & Rolff, J. Complete metamorphosis
1041 and microbiota turnover in insects. *Mol. Ecol.* 1–9 (2022) doi:10.1111/mec.16673.
- 1042 84. Mevi-Schütz, J. & Erhardt, A. Amino acids in nectar enhance butterfly fecundity: A
1043 long-awaited link. *Am. Nat.* **165**, 411–419 (2005).
- 1044 85. Hansen, A. K., Pers, D. & Russell, J. A. Symbiotic solutions to nitrogen limitation

- 1045 and amino acid imbalance in insect diets. *Adv. In Insect Phys.* **58**, 161–205 (2020).
- 1046 86. Behar, A., Yuval, B. & Jurkevitch, E. Enterobacteria-mediated nitrogen fixation in
1047 natural populations of the fruit fly *Ceratitis capitata*. *Mol. Ecol.* **14**, 2637–43 (2005).
- 1048 87. Wang, Y. *et al.* Diversity, composition and functional inference of gut microbiota in
1049 indian cabbage white pieris canidia (Lepidoptera: Pieridae). *Life* **10**, 1–11 (2020).
- 1050 88. Woestmann, L. & Saastamoinen, M. The importance of trans-generational effects in
1051 lepidoptera. *Curr. Zool.* **62**, 489–499 (2016).
- 1052 89. Holzmann, K. L. *et al.* Net primary productivity but not its remote sensing proxies
1053 predict mammal diversity in Andean Amazonian rainforests. *Ecology* **106**, e70059
1054 (2025).
- 1055 90. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D.
1056 Development of a dual-index sequencing strategy and curation pipeline for analyzing
1057 amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ.*
1058 *Microbiol.* **79**, 5112–5120 (2013).
- 1059 91. Sickel, W. *et al.* Increased efficiency in identifying mixed pollen samples by meta-
1060 barcoding with a dual-indexing approach. *BMC Ecol.* **15**, 1–9 (2015).
- 1061 92. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of
1062 sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4516–4522 (2011).
- 1063 93. Leonhardt, S. D., Peters, B. & Keller, A. Do amino and fatty acid profiles of pollen
1064 provisions correlate with bacterial microbiomes in the mason bee *Osmia bicornis*?
1065 *Philos. Trans. R. Soc. B* **377**, 20210171 (2022).
- 1066 94. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile
1067 open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
- 1068 95. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS
1069 amplicon sequencing. *bioRxiv* 081257 (2016).
- 1070 96. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for
1071 next-generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).
- 1072 97. Edgar, R. C. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS
1073 sequences. *bioRxiv* 074161 (2016) doi:10.1101/074161.
- 1074 98. R Development Core Team. R: A Language and Environment for Statistical
1075 Computing. Vienna: *R Foundation for Statistical Computing* at [http://www.r-](http://www.r-project.org)
1076 [project.org](http://www.r-project.org) (2024).
- 1077 99. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive
1078 Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013).
- 1079 100. Barnett, D., Arts, I. & Penders, J. microViz: an R package for microbiome data

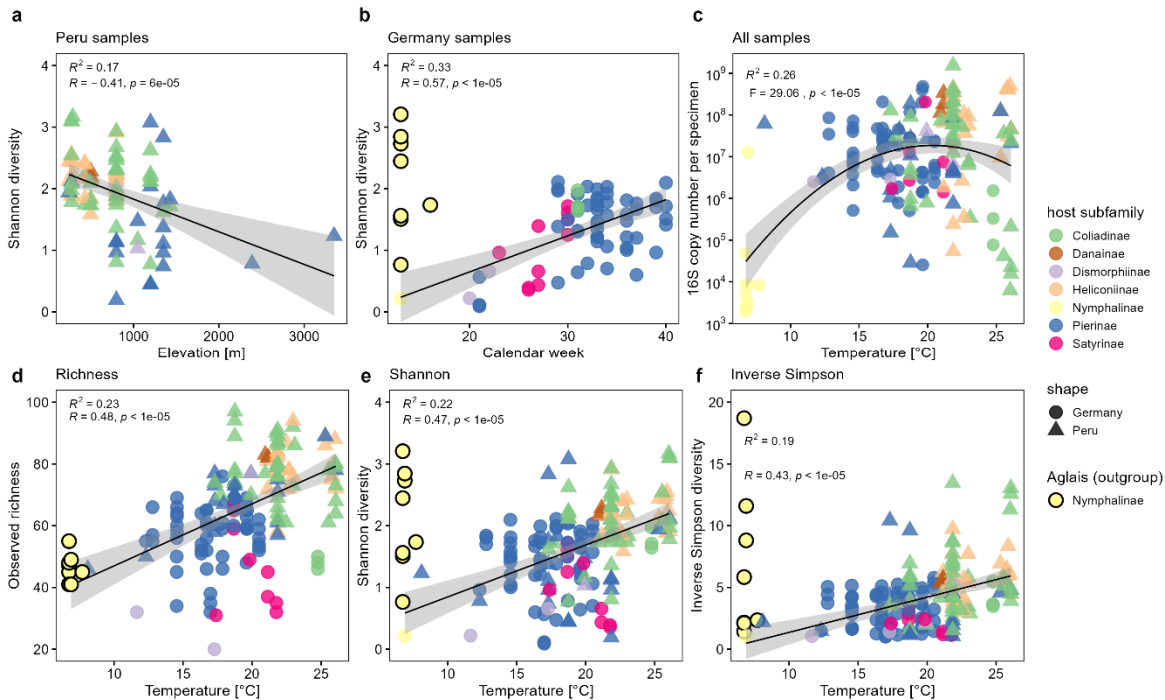
- visualization and statistics. *J. Open Source Softw.* **6**, 3201 (2021).
101. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).
doi:10.1007/978-3-319-24277-4_9.
102. Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A. & Callahan, B. J.
Simple statistical identification and removal of contaminant sequences in marker-
gene and metagenomics data. *Microbiome* **6**, 226 (2018).
103. Oksanen, J. *et al.* *vegan: Community Ecology Package*. R package version 2.6-8. at
<https://cran.r-project.org> (2024).
104. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and
extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456
(2016).
105. Lahti, L. & Shetty, S. microbiome R package. at
<https://www.bioconductor.org/packages/release/bioc/html/microbiome.html> (2019).
106. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: Efficient
manipulation of biological strings. *R package version 2.74.0* at
<https://doi.org/10.18129/B9.bioc.Biostrings> (2024).
107. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and
evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
108. Yu, G. *Data Integration, Manipulation and Visualization of Phylogenetic Trees*.
Data Integration, Manipulation and Visualization of Phylogenetic Trees (Chapman
and Hall/CRC, 2022). doi:10.1201/9781003279242.
109. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: A framework for
sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67
(2014).
110. Fox, J. & Weisberg, S. *An R Companion to Applied Regression: Appendices*. (Sage,
2019).
111. Ben-Shachar, M. S., Lüdtke, D. & Makowski, D. effectsize: Estimation of Effect
Size Indices and Standardized Parameter. *Open J.* **5**, 2815 (2020).

1110 12 Supplementary Material



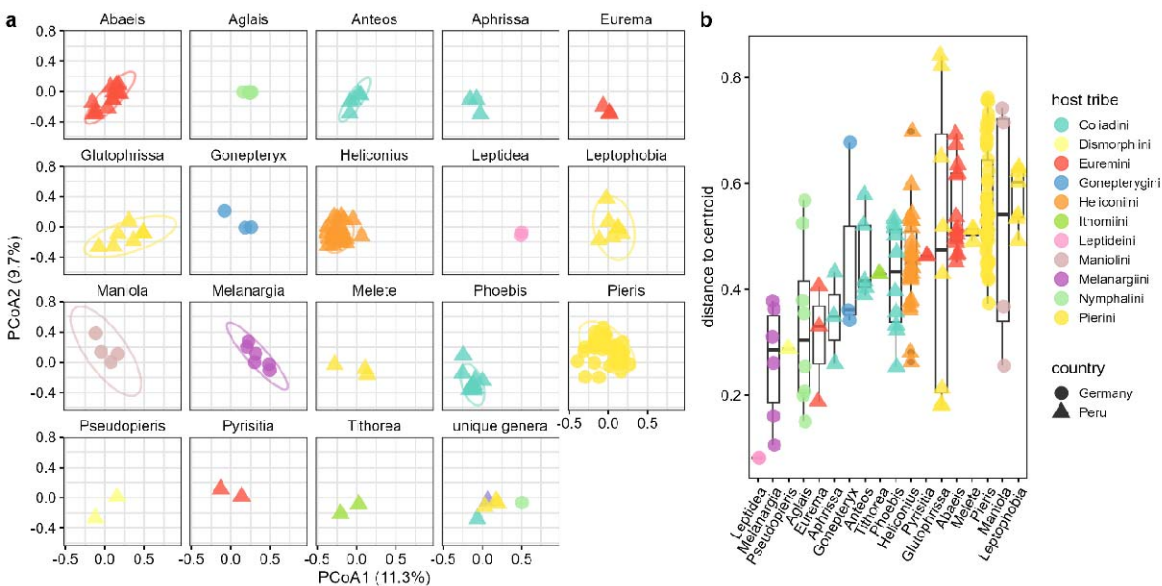
Supplementary Figure S1 Bacterial load and diversity across butterfly host taxa and countries.

(a) Total bacterial load estimated by 16S rRNA gene copy number qPCR per butterfly genus. (b) Bacterial load did not differ between countries in individual subfamilies. Wilcoxon rank-sum tests: Coliadinae ($p = 0.051$), Dismorphiinae ($p = 0.33$) and Pierinae ($p = 0.81$). (c) Shannon Index grouped by host tribe and colored by host subfamily. The full model explained 30% of variation in Shannon diversity, with the main effect accounted to host tribe (partial $R^2 = 0.24$, $F_{11,154} = 4.3$, $p < 10^{-4}$), there was no effect of country (partial $R^2 < 0.01$, $F_{1,154} = 0.03$, $p = 0.87$). (d) Shannon Index did not differ by country in individual Wilcoxon rank-sum tests: Coliadinae ($p = 0.12$), Dismorphiinae ($p = 0.33$) and Pierinae ($p = 0.62$).



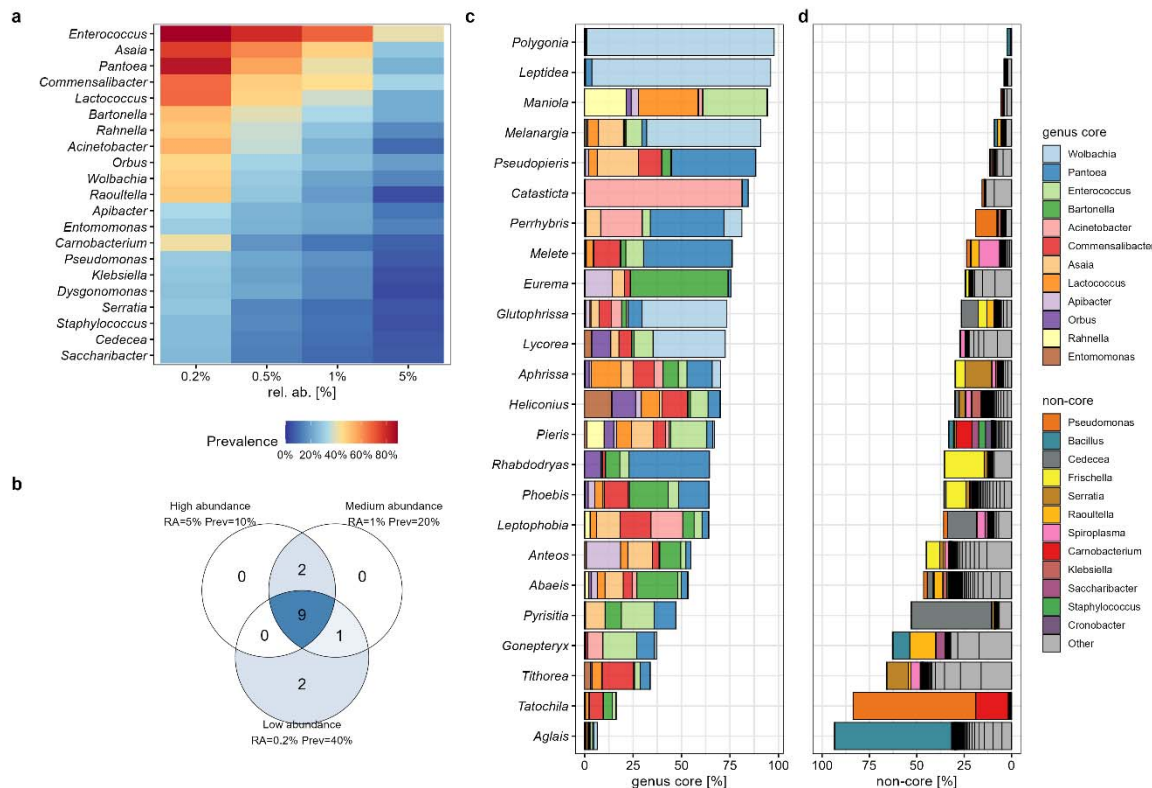
Supplementary Figure S2 Influence of environmental gradients on adult butterfly microbiota diversity

(a) Alpha diversity (only Peru samples) shows linear decrease with elevation (meters above sea level). (b) Alpha diversity (only Germany samples) shows linear increase by calendar week. *Aglaia* were removed from diversity analyses, due to their low microbial abundance, but illustrated as outgroup within the figure. (c) Temperature did not significantly affect bacterial load after accounting for host subfamily. (d) For observed microbial richness, both host subfamily and temperature were significant, with host subfamily having the stronger effect (partial $R^2 = 0.27$, $F_{6,151} = 9.42$, $p < 10^{-8}$), while temperature had a smaller but significant effect (partial $R^2 = 0.03$, $F_{1,151} = 4.63$, $p = 0.033$). (e) For Shannon diversity, the effect was attributable to host subfamily, whereas temperature was not significant. (f) For inverse Simpson diversity, host subfamily had the stronger effect (partial $R^2 = 0.14$, $F_{6,151} = 3.97$, $p = 0.001$), while temperature was still significant (partial $R^2 = 0.03$, $F_{1,151} = 4.32$, $p = 0.039$).



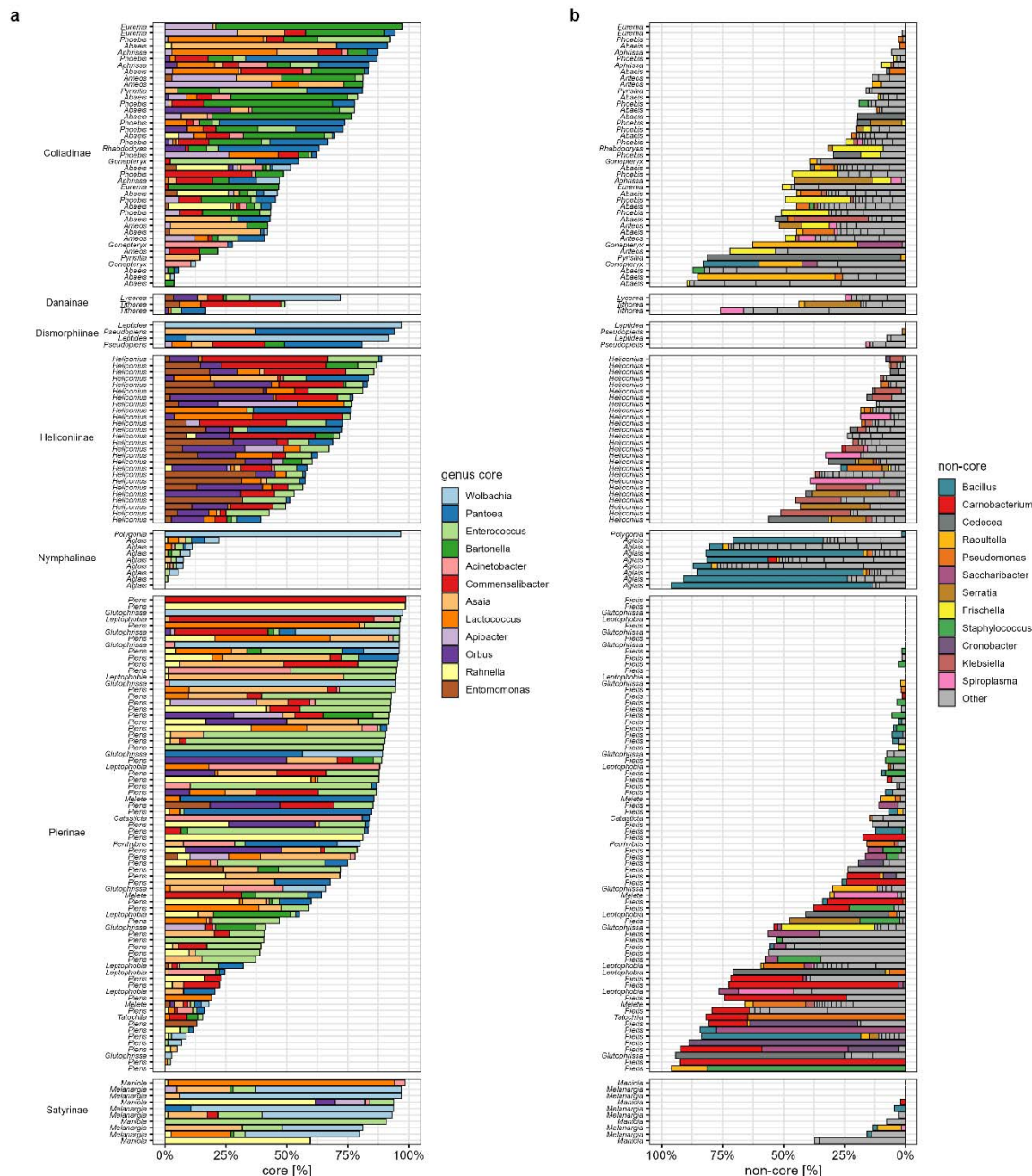
Supplementary Figure S3 Adult butterfly microbiome dissimilarity and dispersion across host genera.

(a) Beta diversity of the adult butterfly microbiota using PCoA based on Bray-Curtis distances. Samples are colored by host tribe and panels faceted by host genus (unique genera with less than two replicates were excluded). (b) Microbiota of adult butterfly genera differs in beta dispersion (colored by host tribe).



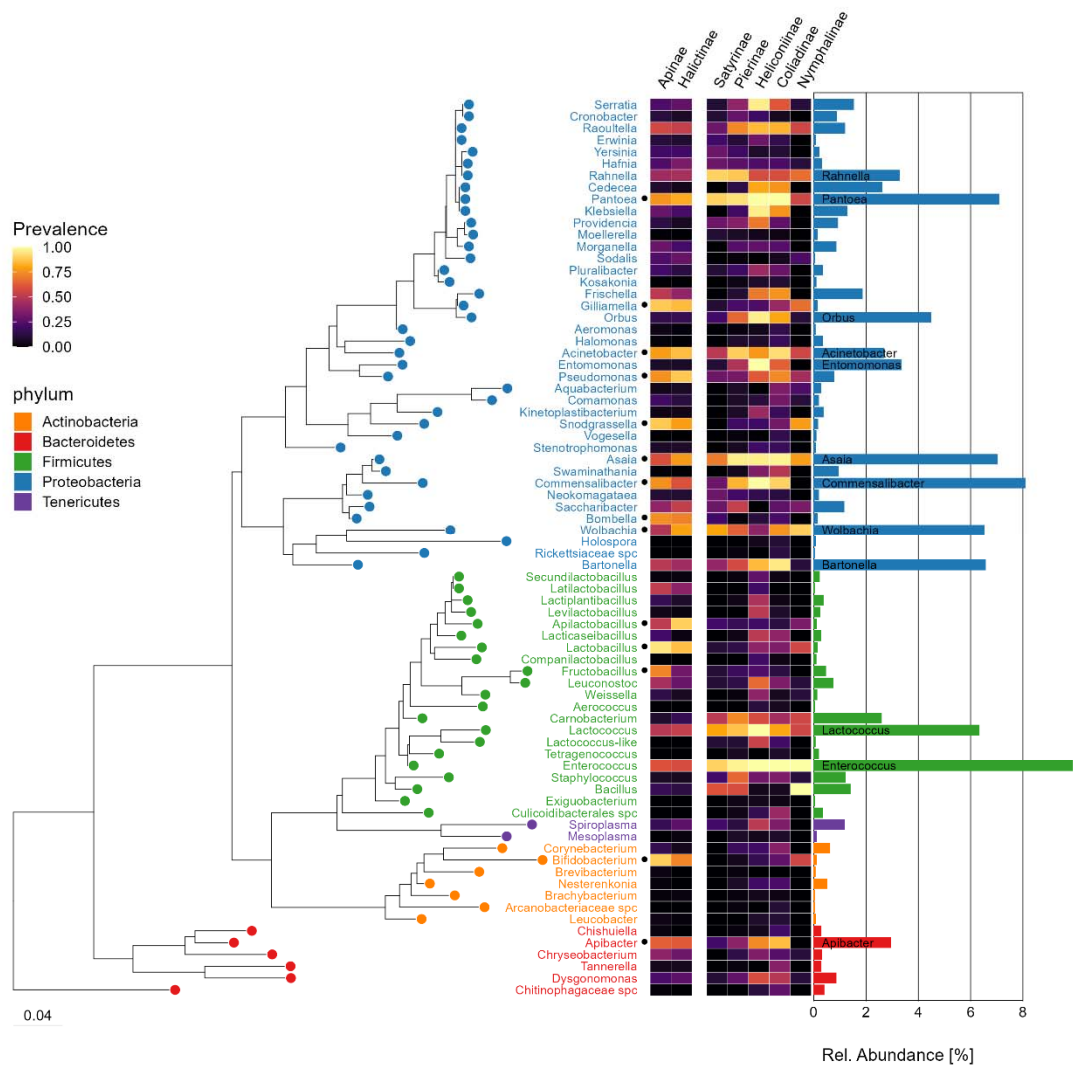
Supplementary Figure S4 Selection of the common genus core microbiota.

(a) Heatmap highlighting the most abundant and prevalent genera among all butterfly samples. (b) Venn diagram shows overlap of common core genera when applying different core definitions. High abundance definition (5% rel. abundance 10% prevalence) results in 11 core genera (missing *Acinetobacter*); Medium abundance definition (1% rel. abundance 20% prevalence) results in 12 core genera (as used in this study). Low abundance definition (0.2% rel. abundance 40% prevalence) results in 12 core genera replacing *Entomomonas* & *Apibacter* with *Carnobacterium* & *Raoultella*. (c) Average relative abundance of the 12 common genus core taxa among individual butterfly genera. (d) Remaining non-core genera belonged to the same bacterial groups e.g. LAB (*Carnobacterium*), enteric bacteria (*Cedecea*, *Serratia*, *Raoultella*), or Orbacea (*Frischella*).



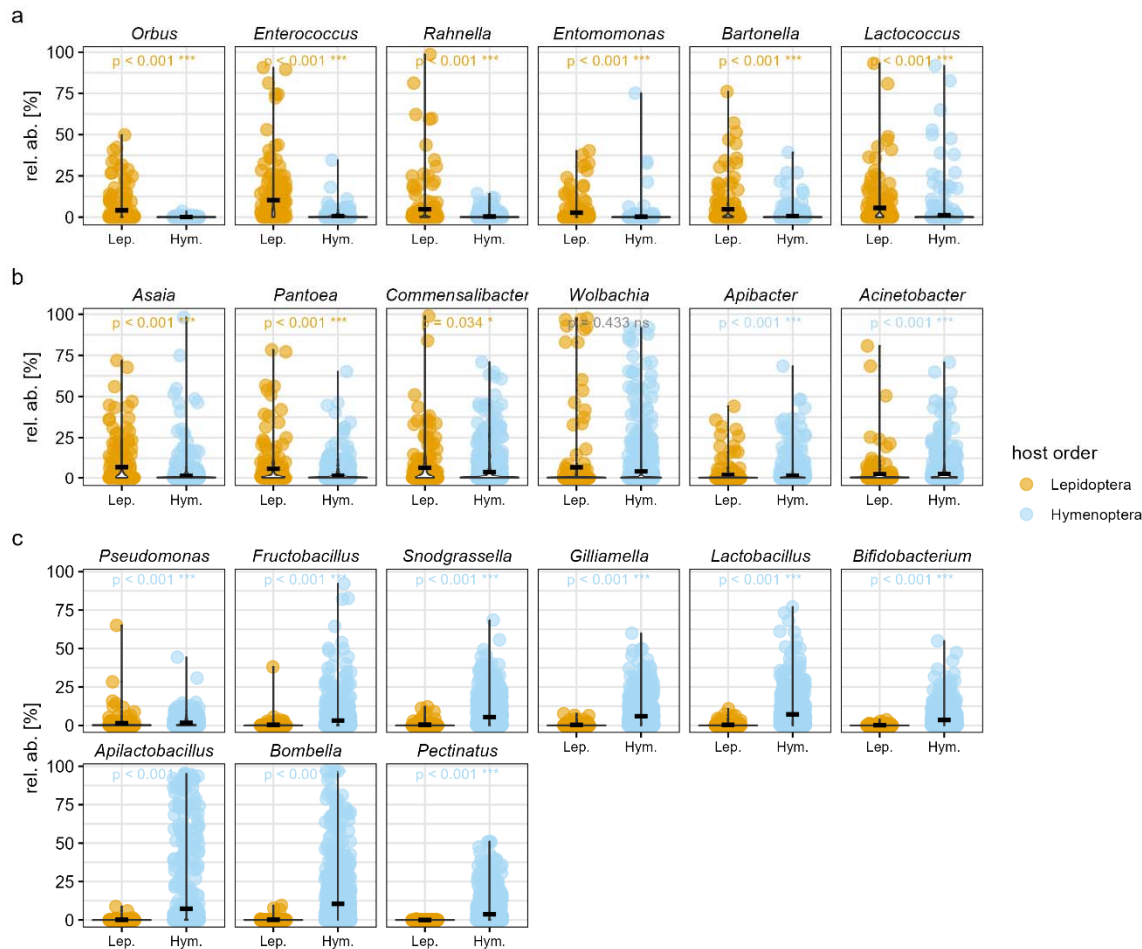
Supplementary Figure S5 Adult butterfly microbiome composition across individual samples

(a) Average relative abundance of the 12 common genus core taxa per individual sample grouped by host subfamily. (b) Remaining non-core genera per sample grouped by host subfamily. Relative abundances < 1% are not shown.



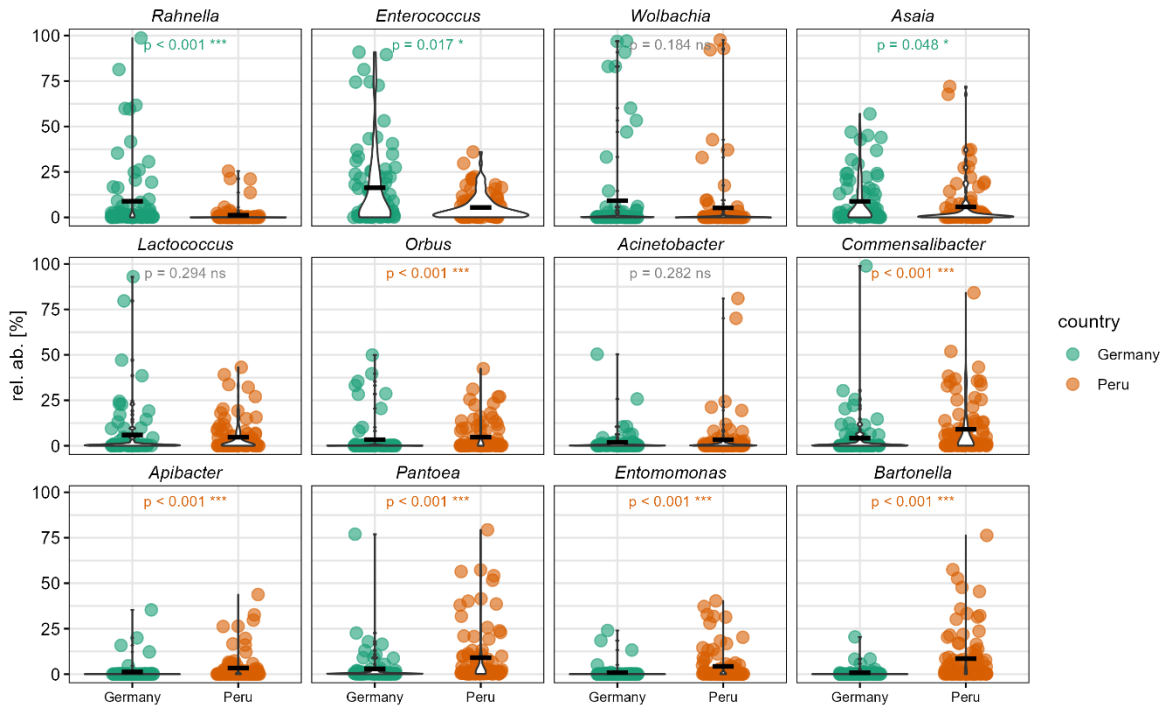
Supplementary Figure S6 Full genus-level annotation, prevalence and abundance of the adult butterfly microbiota.

Phylogenetic grouping (SINA RAXML tree) and relative abundance of the 80 most abundant bacterial genera from adult butterflies. Expansion of Figure 4 including all bacterial genus names. Prevalence in wild bee microbiomes from sympatric Apinae and Halictinae is shown for comparison. Black dots indicate common core genera identified from the bee dataset.



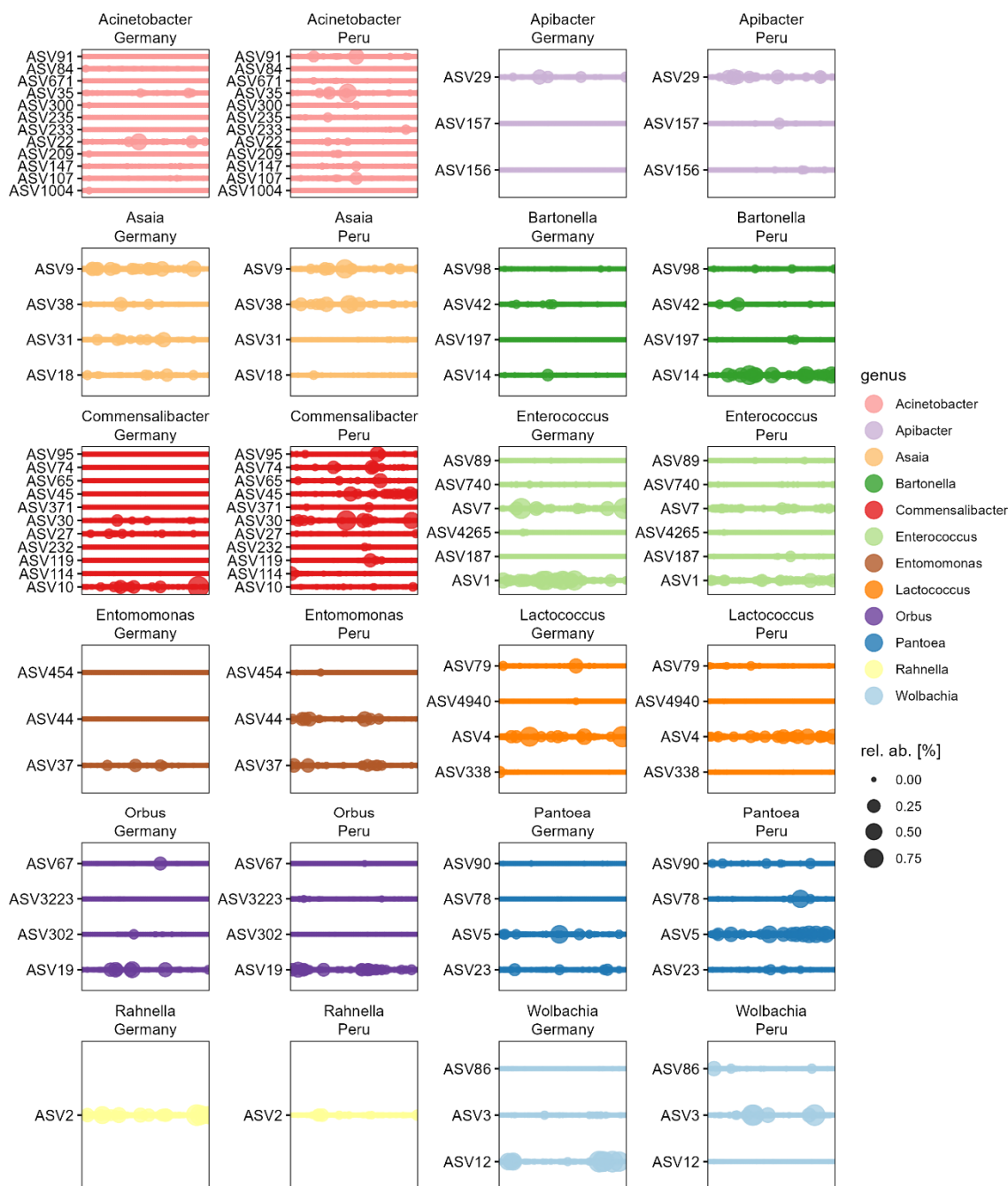
Supplementary Figure S7 Comparison of the common genus core microbiota of Lepidoptera and Hymenoptera

(a) Six genera are unique to the common core of Lepidoptera. (b) Six genera show taxonomic overlap between the common cores of Lepidoptera and Hymenoptera. (c) Nine genera are unique to the common core of Hymenoptera. Using the same operational common core definition for Lepidoptera and Hymenoptera should facilitate comparability between datasets but does not necessarily represent an optimized threshold for the Hymenoptera dataset, nor is it intended to specifically identify socially transmitted taxa. Wilcoxon rank-sum tests with Benjamini–Hochberg correction for multiple comparisons (Lepidoptera $n = 167$, Hymenoptera $n = 872$).

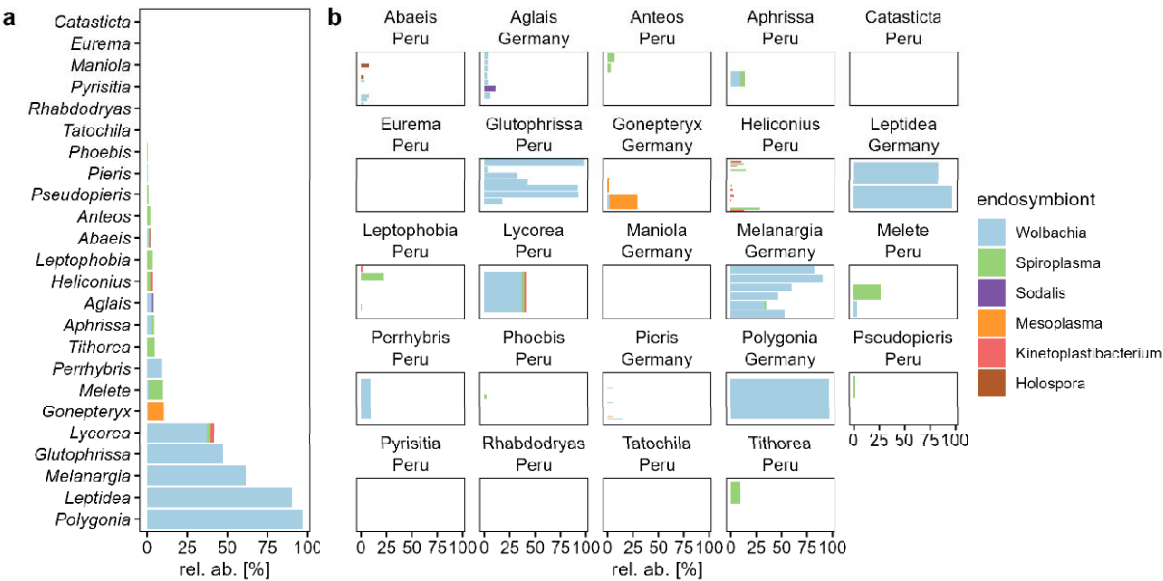


Supplementary Figure S8 Country-specific distribution of common genus core taxa

Relative abundance of individual core genera among countries and results of Wilcoxon rank-sum Tests following Benjamini–Hochberg correction accounting for multiple comparisons (Germany n = 76, Peru n = 91).

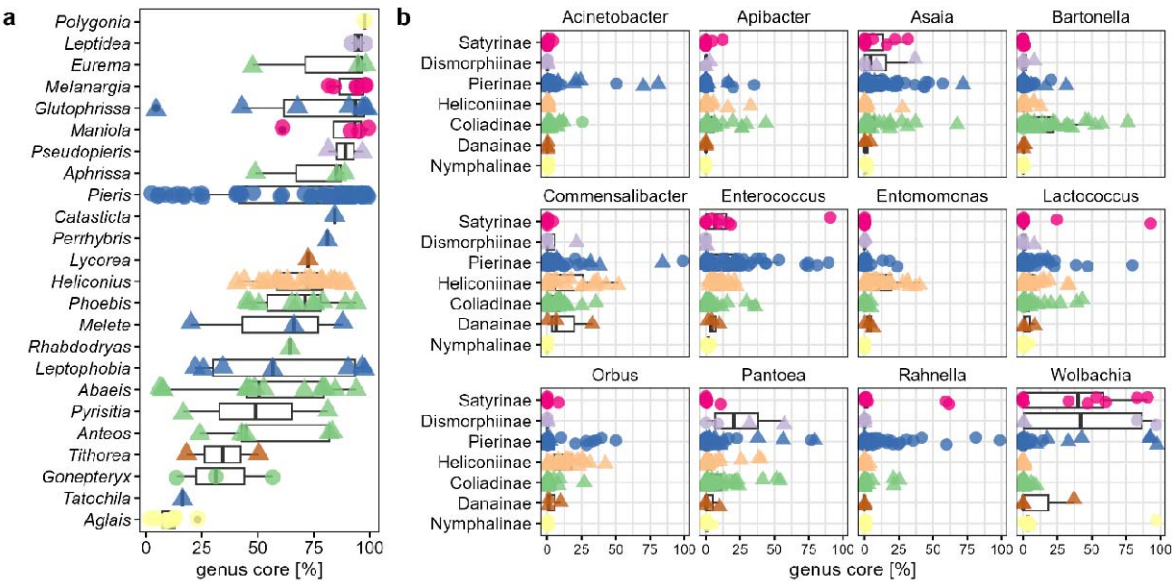


Supplementary Figure S9 Distribution of common genus core ASVs across countries
Relative abundance and distribution of individual ASVs of the 12 common genus core taxa. Only ASVs with $\geq 2\%$ relative abundance in at least one sample were shown.



Supplementary Figure S10 Endosymbionts in the adult butterfly microbiota

(a) Average relative abundance of endosymbionts grouped per butterfly genus. (b) Relative abundance of endosymbionts among individual samples grouped by genus and country. Relative abundances < 1% are not shown.



Supplementary Figure S11 Distribution of common genus core taxa among butterfly samples

(a) Total genus core abundance among individual butterfly genera. (b) Distribution of individual genus core taxa across butterfly subfamilies. Symbols depict different sampling countries (Circles for Germany, triangles for Peru).

File: Suppl_table_sample_filter_metadata.csv

Supplementary Table S1 Filtering steps and sample metadata

Removed reads per individual sample (total and percentage) including all metadata and individual BioProject accession numbers.