Heliconius butterflies host characteristic and phylogenetically structured adult-stage microbiomes

Tobin J. Hammer[1*], Jacob C. Dickerson[2,3], W. Owen McMillan[2], Noah Fierer[3,4]

1. Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA

2. Smithsonian Tropical Research Institute, Apartado 0843-03092, Panama City, Panama

3. Department of Ecology & Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309, USA

4. Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, CO 80309, USA

*Corresponding author: tobin.hammer@utexas.edu

Keywords: symbiosis, gut bacteria, Lepidoptera, microbiota, parasites

Abstract

Lepidoptera (butterflies and moths) are diverse and ecologically important, yet we know little about how they interact with microbes as adults. Due to metamorphosis, the form and function of their adult-stage microbiomes might be very different from microbiomes in the larval stage (caterpillars). We studied adult-stage microbiomes of *Heliconius* and closely related passion-vine butterflies (Heliconiini), which are an important model system in evolutionary biology. To characterize the structure and dynamics of heliconiine microbiomes, we used field collections of wild butterflies, 16S rRNA gene sequencing, quantitative PCR, and shotgun metagenomics. We found that *Heliconius* harbor simple and abundant bacterial communities that are moderately consistent among conspecific individuals and over time. Heliconiine microbiomes also exhibited a strong signal of host phylogeny, with a major distinction between *Heliconius* and other butterflies. These patterns were largely driven by differing relative abundances of bacterial phylotypes shared among host species and genera, as opposed to the presence or absence of host-specific phylotypes. We suggest that phylogenetic structure in heliconiine microbiomes arises from conserved host traits that differentially filter microbes from the environment. While the relative importance of different traits remains unclear, our data indicate that pollen-feeding (unique to *Heliconius*) is not a primary driver. Using shotgun metagenomics, we also discovered trypanosomatids and microsporidia to be prevalent in butterfly guts, raising the possibility of antagonistic interactions between eukaryotic parasites and co-localized gut bacteria. Our discovery of characteristic and phylogenetically structured microbiomes provides a foundation for tests of adult-stage microbiome function, a poorly understood aspect of lepidopteran biology.

Importance

Many insects host microbiomes with important ecological functions. However, the prevalence of this phenomenon is unclear, because in many insect taxa microbiomes have only been studied in part of the life cycle, if at all. A prominent example is the butterflies and moths, in which the composition and functional role of adult-stage microbiomes are largely unknown. We comprehensively characterized microbiomes in adult passion-vine butterflies. Butterfly-associated bacterial communities are generally abundant in guts, consistent within populations, and composed of taxa widely shared among hosts. More closely related butterflies harbor more similar microbiomes, with the most dramatic shift in microbiome composition occurring in tandem with a suite of ecological and life history traits unique to the genus *Heliconius*. Butterflies are also frequently infected with previously undescribed eukaryotic parasites, which may interact with bacteria in important ways. These findings advance our understanding of butterfly biology and of insect-microbe interactions generally.

Introduction

Insect microbiome research has historically focused on hosts with highly stable, specific, and functionally important microbiomes (1). For example, the obligate nutritional endosymbionts of sap- and blood-feeding insects, which exemplify this scenario, have long served as important model systems. More recently, however, wider microbial explorations have shown that this kind of association does not apply to all insect taxa. In some insect groups, microbiomes are more

47 variable, and the constituent microbes have low or no specificity to their hosts (2, 3). Moreover,
48 the functional importance of these microbiomes is often unclear—and possibly variable, low, or
49 nonexistent in some groups (4). Understanding the nature of microbiomes in these insects is not
50 only important to their biology, but also for a broader understanding of insect microbiome
51 evolution.

52 The Lepidoptera (butterflies and moths) are a group in which microbiome form and
53 function have been particularly difficult to resolve. Recent work indicates that larvae
54 (caterpillars) typically harbor microbiomes that are variable, low-abundance, and transient (5–9).
55 However, a crucial consideration for Lepidoptera and other insects that undergo complete
56 metamorphosis is that microbial associations may change between life stages (10). In
57 Lepidoptera, larvae and adults do often differ in the composition and absolute abundance of their
58 microbiomes (6, 11–15). Beyond this pattern, however, the ecology of adult-stage microbiomes
59 remains poorly understood.

60 There are three major unknowns about microbiomes of adult Lepidoptera that we sought
61 to address in this study. First, it is unclear how variable microbiomes are among host individuals
62 and taxa, and which ecological or evolutionary factors (e.g., host phylogeny) are associated with
63 microbiome variation. Understanding these patterns could give insights into whether
64 microbiomes are important for certain ecological traits, such as feeding, and into the likelihood
65 of host-microbe coevolution. Second, it is unclear whether butterfly-associated microbes have
66 any specificity to their host species or to higher taxonomic levels; tight specificity is a hallmark
67 of obligate insect-microbe symbioses (3). Previous studies of adult Lepidoptera microbiomes
68 have largely used 97% operational taxonomic units (OTUs) based on short fragments of the 16S
69 rRNA gene. OTUs can encompass substantial strain-level diversity (16), hindering analyses of

70 specificity. Third, the full diversity of eukaryotic microbes that may be associated with adult
71 Lepidoptera has not been explored, as previous studies have exclusively focused on bacteria or
72 fungi.

73 Our study focuses on the neotropical passion-vine butterflies (Nymphalidae: Heliconiini).
74 This tribe, which includes the genus *Heliconius*, is a foundational system in evolutionary biology
75 with over 150 years of scientific study (17). Given the scientific relevance of heliconiines (18),
76 and the myriad roles that microbes can play in animal biology (19), it is important to know the
77 structure and function of heliconiine-associated microbiomes. Heliconiine butterflies are also a
78 potentially useful system for the field of host-microbe interactions, given the wealth of genomic
79 data available (20) and the ability to rear and genetically manipulate members of this group (21).
80 A handful of amplicon sequencing-based microbiome studies have included adult heliconiines,
81 but these were limited to a small number of individuals and species (11, 22, 23). Consequently,
82 we lack even basic information about the microbes harbored by heliconiine butterflies and how
83 they may vary across individuals and the host phylogeny.

84 In addition to the practical advantages mentioned above, heliconiines are also an
85 interesting system to address how host ecological traits evolve in concert with microbiomes.
86 Within Heliconiini, the genus *Heliconius* exhibits unique feeding and life history traits relative to
87 other genera: larvae feed on young passion-vine shoots and leaves as opposed to old foliage,
88 adults of most species (besides *H. aoede*) feed on pollen as well as nectar, and they have
89 distinctive defensive chemistry and a greatly extended adult-stage lifespan of several months (18,
90 24). These traits could influence and be influenced by microbiomes in many ways. For example,
91 we hypothesized that pollen-feeding might be associated with a distinctive microbiome, either
92 because it is a source of microbes (25) or of nutrients that shift resident microbiome composition

93 (26), or, because the trait itself depends on metabolic contributions from novel microbes (11). A
94 useful starting point to uncovering these kinds of interactions is to examine whether *Heliconius*
95 have microbiomes that are internally consistent yet distinct from those of other heliconiine
96 genera. If so, then we can zero in on the specific ecological and evolutionary drivers of variation,
97 and the potential functional roles of microbiomes in *Heliconius* biology.

98 Here we addressed the following main questions: *i*) How variable are heliconiine
99 microbiomes among host individuals, species, genera, and across the heliconiine phylogeny? *ii*)
100 Are host traits, such as pollen-feeding, associated with microbiome variation? *iii*) How host-
101 specific are butterfly-associated microbes? At the same time, we sought to facilitate future
102 microbiome research on heliconiine butterflies by answering some additional questions: *iv*) How
103 are microbes distributed within the butterfly body? *v*) How abundant are these microbes? *vi*)
104 How reliable is amplicon sequencing of 16S rRNA genes for studying butterfly-associated
105 bacterial communities, and *vii*) are there eukaryotic microbes we might be missing with targeted
106 amplicon sequencing?

107 We collected 214 wild adult butterflies, representing 23 species and subspecies of
108 Heliconiini, and characterized their microbiomes with 16S rRNA gene sequencing. While many
109 adult Lepidoptera microbiome studies have used reared, captive specimens (6, 12, 13, 22, 27),
110 these may give a biased picture of microbial community structure in the wild (11). We also used
111 shotgun metagenomic sequencing on a subset of butterflies both to provide an untargeted
112 assessment of microbial diversity (including eukaryotes) and to provide a finer-resolution picture
113 of microbial specificity than is available from amplicon sequencing of short 16S rRNA gene
114 regions. Furthermore, we used quantitative PCR to estimate the absolute abundance of bacteria
115 across different heliconiine taxa and tissue types. Our work illustrates ecological and

116 evolutionary dynamics of microbiomes in adult heliconiine butterflies and advances our general
117 understanding of Lepidoptera-microbe interactions.
118
119 **Materials and Methods**
120
121 *Field collections*
122
123 The wild adult butterflies used for whole-body microbiome sequencing were collected
124 from seven locations in Panama and Ecuador in May-August 2014 (more detail is provided in the
125 supplemental file "Collection_localities.txt"). Butterflies were euthanized with ethyl acetate and
126 stored in DMSO after removal of wings, following (28). We also stored two DMSO-only blanks
127 to use as negative controls. In June 2016, we collected additional adult butterflies for gut and
128 head/thorax sequencing from Gamboa and Pipeline Road, Panama. For these specimens, we
129 dissected the gut (hindgut, midgut, and the distal ~1/2 of the foregut) using sterilized tools prior
130 to storage in DMSO. The whole head and thorax (including the proximal foregut) were stored
131 separately. Species or subspecies were identified based on morphology. Butterflies were
132 collected under permit # SC/A-7-11 from Panama's Autoridad Nacional del Ambiente and #
133 005-13 IC-FAU-DNB/MA from Ecuador's Ministerio del Ambiente.
134
135 *Sample processing, qPCR, PCR and sequencing*
136
137 We removed whole bodies and head/thorax samples from DMSO and, after
138 homogenization, used approximately 50 mg subsamples of homogenate for DNA extractions

139 with the MoBio PowerSoil kit following the manufacturer's instructions. We added entire guts
140 directly to DNA extraction tubes, in which they were homogenized during the first bead-beating
141 step of the protocol. Two DMSO blanks and 30 DNA extraction blanks were also processed in
142 tandem with the butterfly samples and sequenced. For dissected gut and head/thorax samples, we
143 estimated total bacterial abundance using quantitative PCR (qPCR) with 16S rRNA gene primers
144 (515F/806R) following the protocol described in ref. (5).

145 PCR amplifications (515F/806R primers, V4 region) and 2 X 150 bp Illumina MiSeq
146 sequencing of 16S rRNA genes followed standard Earth Microbiome Project protocols
147 (dx.doi.org/10.17504/protocols.io.nuudeww). For gut and head/thorax samples of 29 butterfly
148 individuals of nine species, we also attempted PCR amplification with primers that target the ITS
149 gene region of fungi (29). However, amplification success with these fungal-specific primers (as
150 assessed by gel electrophoresis) was very low, suggesting a lack of abundant fungal DNA that
151 was later corroborated with the shotgun metagenomic data (see below). DNA extracts from a
152 subset of 15 amplicon-sequenced gut samples were used for shotgun metagenomic sequencing
153 following the approach described previously (30) with an input DNA concentration of 0.75
154 ng/ul, KAPA HiFi HotStart ReadyMix and bead cleanup with Ampure XP beads at a 0.9x ratio.
155

156 *Amplicon data processing*

157

158 Amplicons from the 2014 whole-body samples and 2016 gut and head/thorax samples
159 were sequenced on separate runs, demultiplexed using idemp (https://github.com/yhwu/idemp),
160 and combined for further processing. Cutadapt (31) was used to remove primer sequences. We
161 then used the DADA2 pipeline (32) to quality-filter (max EE value = 1) and trim (150 bp

162 forward, 140 bp reverse) reads, infer exact sequence variants (ESVs), merge paired-end reads,
163 and remove chimeras. We classified ESVs using the RDP Naive Bayesian Classifier algorithm
164 (33) against the SILVA training set v. 132 (34).

165 Further data processing and analyses were conducted in R v. 3.6.0 (35). We used
166 decontam (36) for prevalence-based identification of putative contaminant ESVs based on 34
167 negative controls (DMSO and DNA extraction blanks and PCR no-template controls). The
168 median percentage of contaminant sequences across butterfly samples was 0.09%, but two
169 samples had >10% contaminants and were removed from further analysis. ESVs with < 100 total
170 sequences across all samples (out of a combined total of 5.6 million sequences) were removed,
171 as were ESVs classified as mitochondria or chloroplast, or bacteria lacking sub-domain
172 identification. These ESVs combined typically made up a low proportion of reads from the
173 libraries (median of 2.6% across all samples). We also used these data to correct qPCR-based
174 absolute abundances for non-bacterial amplification. Specifically, we multiplied the proportion
175 of bacteria in sequence libraries by the total number of 16S rRNA gene copies to obtain
176 estimates of bacterial absolute abundance.

177 As the 16S rRNA gene amplicon libraries were highly variable in read depth across
178 samples, we rarefied all samples to 5,000 sequences, filtering out 13 samples with lower
179 sequence depth. We relabeled *Pantoea*, *Erwinia*, *Kluyvera*, *Citrobacter*, *Klebsiella*, and
180 *Cronobacter* taxonomic assignments to *Enterobacter*. This step was taken as genera within
181 Enterobacteriaceae are often polyphyletic and are difficult to resolve from short 16S rRNA gene
182 regions (e.g., (37)), and we wanted to avoid spurious separation of ESVs among genera and
183 resulting idiosyncrasies in genus distributions across butterflies. Another case of ambiguous
184 taxonomy occurred with sequences classified as *Pseudomonas* by the SILVA reference database,

185 which we label below as "*Entomomonas/Pseudomonas*". *Entomomonas* is a recently described
186 genus of apparently insect-specialized Pseudomonadaceae and includes several sequences
187 retrieved from *Heliconius* genome assemblies (38).
188
189 *Amplicon and qPCR data analysis*
190
191 Beta diversity statistics and plots are based on Bray-Curtis dissimilarities and/or
192 UniFrac distances (weighted and unweighted). To obtain a bacterial phylogeny for the latter, we
193 used the fragment-insertion method (39) to place our ESV sequences into the Greengenes
194 reference tree (40). The butterfly phylogeny is from (41) (TreeBASE #Tr77496). Four of the
195 butterfly species in our sample set contained specimens from two distinct subspecies (e.g.,
196 *Heliconius sara sara* and *Heliconius sara magdalena*). To include these in the species-level host
197 phylogeny, we inserted subspecies tips halfway along the terminal branches to their sister
198 subspecies. Hereafter we refer to these subspecies as "species" for simplicity.
199 To test for host-phylogenetic signal in microbiomes, we used Mantel tests with 9999
200 permutations to calculate the correlation between microbial community dissimilarities/distances
201 and host phylogenetic distances (42). Intraspecific variation in microbiomes was handled by
202 averaging the pairwise dissimilarities/distances between all individuals of one species and all
203 individuals of another species. We used the phytools package (43) to visualize concordance
204 between topologies of the host phylogeny and a dendrogram of bacterial community
205 dissimilarities. Nodes were rotated with the "cophylo" function in phytools to maximize tip
206 matching between the two trees.

207 Differences in overall community composition between host genera were tested with
208 PERMANOVA as implemented in the vegan package (44). Using the "betadisper" function we
209 corroborated that significant test results were due to host genus-level differences in location and
210 not dispersion (45). We used a nonparametric statistical test (Wilcoxon rank-sum) to identify
211 bacterial genera that differed in relative abundance between host taxa or between sample types
212 (gut versus head/thorax) and applied a false discovery rate correction to the resulting p values.
213 We tested whether total bacterial abundance differed between *Heliconius* and other
214 butterfly genera using qPCR data from gut samples and head/thorax samples. Each sample type
215 was analyzed separately using log-transformed counts of bacterial 16S rRNA gene copies. We
216 verified that residuals were approximately normally distributed and used a linear mixed-effects
217 model as implemented in the nlme package (46), treating host genus (*Heliconius* vs. others) as a
218 fixed effect and host species as a random effect.
219
220 *Metagenome data processing and analysis*
221
222 For 15 gut samples, we obtained shotgun metagenomic data to complement the
223 bacterial 16S rRNA gene amplicon dataset. We quality-filtered these reads with sickle (47) and
224 trimmed adapters with cutadapt (31). We then used Bowtie 2 (48) to filter out reads matching a
225 given sample's corresponding host species' genome, obtained from Lepbase (20). The two
226 *Dryadula phaetusa* metagenomes were mapped to a genome of the sister species *Dryas iulia* as
227 no *Dryadula* genome was available. Since there was a high proportion of host-derived reads, we
228 focused here on describing microbial diversity using ribosomal RNA gene reads present in the
229 metagenomes. With the host-filtered reads, we used phyloFlash (49) to find and classify

230 eukaryotic and bacterial SSU rRNA reads. Bacterial community composition was compared
231 between the amplicon and shotgun metagenomic datasets using a Mantel test. We also used
232 phyloFlash to assemble 16S rRNA genes from the 150 bp shotgun reads. These longer sequences
233 allowed us to estimate the phylogeny of *Orbus*, the dominant bacterium in these 15 samples.
234 *Orbus* sequences were aligned with MUSCLE (50), curated with Gblocks (51), and used for
235 maximum likelihood reconstruction with the phylogeny.fr implementation (52) of PhyML (53).
236
237 *Data availability*
238
239 Amplicon data, metadata, and R code are available from figshare
240 (figshare.com/projects/Heliconius_butterfly_microbiomes/70520). Metagenomes are available
241 from MG-RAST (project no. MGP89563).
242
243 **Results**
244 Adult heliconiine butterflies host whole-body microbiomes that are typically low in
245 diversity and evenness. Across all individuals, the median ESV richness was 26, of which only
246 11 accounted for 95% or more of the sequences. The composition of these communities is also
247 reasonably consistent within host species. For example, within our most deeply sampled
248 population (*H. erato demophoon* in Gamboa, Panama; N = 23), a median of 84% of the 16S
249 rRNA gene reads obtained from a given individual's microbiome belonged to a core set of 10
250 bacterial genera (Fig. S1). Some of these genera are present in roughly similar relative
251 abundances across individuals and across our two sampling years (Fig. S1). Applying the
252 aforementioned calculation to all heliconiine species from which five or more individuals were

253 sequenced, we found that this intraspecific consistency extends beyond *H. erato demophoon*. For
254 most host species, over 80% of the constituent individuals' microbiome reads belonged to the 10
255 most abundant bacterial genera for that species (median = 88%, range = 77-94%).

256 Microbiomes from whole, homogenized butterfly bodies are mainly composed of gut-
257 associated taxa. Isolated guts are similar to conspecific whole-body microbiomes in their
258 bacterial community profiles (Fig. S1), with many bacterial genera occurring at similar relative
259 abundances in guts and head/thorax tissue (Table S1). Some bacteria did differ between whole-
260 body and gut samples, such as *Wolbachia*, which usually inhabit reproductive tissues (54) (Fig.
261 S1). Likewise, in both *Heliconius* and other heliconiine genera, the relative abundance of
262 *Acinetobacter* was ~10-fold higher in head/thorax samples as compared with guts (Table S1). In
263 contrast, *Orbus*, *Enterobacter*, *Asaia*, and some other dominant bacterial genera are enriched in
264 gut tissue (Table S1).

265 We used metagenomes (N = 15) to examine the accuracy of amplicon sequencing for
266 describing bacterial community composition in our broader sample set. Interindividual
267 microbiome variation was highly correlated between the amplicon-based dataset and the
268 metagenomic dataset (Mantel $r = 0.63$, $p < 0.001$). Furthermore, for six of the most abundant
269 bacterial genera, relative abundances in amplicon libraries were highly predictive of relative
270 abundances in metagenomes (Fig. S2). These results support the use of amplicon sequencing for
271 evaluating bacterial community composition in adult heliconiine butterflies.

272 Adult *Heliconius* butterflies generally harbor a high absolute abundance of gut
273 bacteria, though titers sometimes varied substantially even among conspecific individuals
274 collected from the same location (Fig. 1A). On average, *Heliconius* harbored significantly higher
275 (by ~5.5-fold) gut bacterial titers than species belonging to other passion-vine butterfly genera (p

276 = 0.039). Bacterial communities associated with head and thorax tissue did not significantly
277 differ in titer between *Heliconius* and other genera (p = 0.38; Fig. S3).
278 *Heliconius* also harbor compositionally distinct whole-body bacterial communities as
279 compared with other passion-vine butterfly genera (Fig. 1B). Microbiomes clustered by host
280 genera to varying degrees depending on the distance metric used. The effect was strongest with
281 taxonomic or phylogenetic metrics that incorporate information on the relative abundances of
282 exact sequence variants (ESVs) (Bray-Curtis: $R^2 = 0.13$, p = 0.001; weighted UniFrac: $R^2 = 0.18$,
283 p = 0.001) as opposed to purely presence/absence-based metrics (Jaccard: $R^2 = 0.11$, p = 0.001;
284 unweighted UniFrac: $R^2 = 0.10$, p = 0.001). This separation of *Heliconius* microbiomes from
285 related butterfly genera is not fully explicable by adult-stage diet, as individuals of the non-
286 pollen-feeding species *Heliconius aoede* mostly clustered with pollen-feeding *Heliconius* species
287 (Fig. 1B). An analysis of species-by-species Bray-Curtis dissimilarities confirmed that *H. aoede*
288 microbiomes are not uniquely distinct from those of pollen-feeding *Heliconius* species (Fig. S4).
289 As suggested by the host genus-level clustering, we found that variation in butterfly
290 microbiome composition was associated with host relatedness (phylogenetic distance). More
291 closely related butterfly lineages harbored more similar microbiomes (Bray-Curtis: Mantel r =
292 0.40, p = 0.01), and the topologies of the butterfly phylogeny and the dendrogram of microbiome
293 similarity were moderately congruent (Fig. 2). This pattern was still statistically significant when
294 microbiome variation was measured using weighted UniFrac distances (Mantel $r = 0.31$, p =
295 0.03), but not when using unweighted UniFrac (Mantel $r = 0.03$, p = 0.41) or Jaccard distances
296 (Mantel $r = 0.19$, p = 0.08), which do not incorporate relative abundance information.
297 We then determined which bacterial taxa contribute to the observed taxonomic (Fig.
298 1B) and phylogenetic (Fig. 2) structuring in overall community composition. The dominant

299 bacterial genera *Enterobacter*, *Orbus*, and *Entomomonas*/*Pseudomonas* were proportionally
300 more abundant in *Heliconius* versus the non-pollen-feeding butterfly genera, while *Asaia* and
301 *Apibacter* showed the opposite pattern (Fig. 3). None of these bacterial genera, however, were
302 exclusively restricted to one host feeding guild, genus, or species, and their relative abundances
303 were occasionally highly variable (Fig. 3). We also analyzed the data at the ESV-level, including
304 all individual ESVs that were reasonably abundant within at least one host species (≥ 5% within-
305 species average). We did not find evidence for prevalent host species- or genus-restricted ESVs
306 (Fig. 4). Most ESVs are present, albeit with varying relative abundances, across host genus and
307 species boundaries.

308 Amplicon-derived ESVs are limited in their ability to resolve bacterial strains as they
309 represent only a short region of the 16S rRNA gene (here, ~250 bp). To test for potential host
310 specificity at a finer level of resolution, we obtained near-full-length 16S rRNA gene sequences
311 from the bacterium *Orbus*, which is highly prevalent across gut and whole-body samples (Fig. 3,
312 Fig. S1) and almost exclusively composed of a single ESV (Fig. 4). These sequences were
313 reconstructed from the 12 metagenomes in which *Orbus* was sufficiently abundant. A phylogeny
314 based on these sequences and other Orbaceae shows that butterfly-associated *Orbus* form a
315 unique clade (Fig. 5). Host-phylogenetic or geographic structure was not evident within this
316 clade. In fact, many of the Panamanian heliconiine butterflies harbored *Orbus* that have identical
317 or near-identical 16S rRNA gene sequences to an *Orbus* strain isolated from an East Asian
318 butterfly, *Sasakia charonda* (55).

319 We also used the shotgun metagenomes to search for microbial eukaryotes in butterfly
320 gut samples and found that microsporidia and trypanosomatids were prevalent (27% and 47% of
321 the 15 individual samples analyzed, respectively) (Fig. S5). Fungi and other microeukaryotic

322 taxa were only sporadically detected, and at very low relative abundances. When
323 trypanosomatids were detected in a given sample, we also sometimes detected 16S rRNA gene
324 reads in the metagenomic data classified as *Kinetoplastibacterium*, an obligate bacterial
325 endosymbiont of certain trypanosomatids (56). *Kinetoplastibacterium* was not detected without
326 its corresponding trypanosomatid host (Fig. S5). This observation led us to reexamine the larger,
327 amplicon-based bacterial dataset. We found that 9% of *Heliconius* individuals and 4% of
328 individuals belonging to other heliconiine genera were infected with a trypanosomatid, as
329 inferred by the presence of *Kinetoplastibacterium*. These proportions were not significantly
330 different (Fisher's Exact Test, $p = 0.28$). They are also likely underestimates, given the
331 aforementioned prevalence of trypanosomatids among metagenomes (47%) and the potential for
332 false negatives—in five metagenomes, trypanosomatids were detected without
333 *Kinetoplastibacterium* (Fig. S5).
334
335 **Discussion**
336
337 We found that adult-stage microbiomes in heliconiine butterflies are generally similar in
338 composition among conspecific individuals and over time, and are also abundant within gut
339 tissue (especially in *Heliconius*). Our estimates of total bacterial abundances (via qPCR), in
340 parallel with those of ref. (23), support the hypothesis that bacteria actively colonize adult
341 butterflies as opposed to passing through the gut transiently. These features contrast with the
342 typical situation in lepidopteran larvae (5), reinforcing the idea that insect-microbe associations
343 can differ strongly between life stages (10).

344 Although stable relative to larvae, dominant butterfly-associated microbes do exhibit
345 higher levels of interindividual variability as compared with obligate nutritional endosymbionts
346 or gut microbes of some other insect groups. For example, in honey bees, >98% of sequences
347 belong to a honey bee-specific set of 5-9 bacterial species (57). We found that ~80-90% of
348 sequences in a given butterfly species' microbiome belonged to a set of 10 core bacterial genera.
349 This contrast suggests that microbiomes are more facultative for butterflies or that there is
350 greater functional redundancy among butterfly-associated microbes.

351 We also discovered that *Heliconius* butterfly microbiomes are distinct from those of other
352 heliconiine genera, in terms of overall community structure (Fig. 1B), relative abundances of
353 specific bacterial genera (Fig. 3), and total numbers of gut bacteria (Fig. 1A). This finding opens
354 the possibility that one or more of the ecological traits specific to *Heliconius* influences, and may
355 be influenced by, the microbiome. As these traits evolved in tandem (18) it is difficult to
356 disentangle their potential links to microbes, but our analysis of the species *Heliconius aoede*
357 suggests that pollen-feeding is not a primary driver. *H. aoede* does not pollen-feed (58, 59), yet
358 its microbiomes were not uniquely distinct from pollen-feeding *Heliconius* species (Fig. 1B, Fig.
359 S4). This result contrasts with previous work on butterflies finding a strong association between
360 adult-stage feeding ecology (nectivory versus frugivory) and adult gut microbiomes (23).

361 Larval feeding ecology, however, warrants further investigation as a potential driver of
362 variation in adult gut microbiomes. Microbiomes of butterfly larvae and adults are not fully
363 decoupled: in *H. erato*, some adult-stage microbes appear to be carried over from the larval stage
364 (11). All *Heliconius* (including *H. aoede*) feed on young passion-vine foliage while other
365 heliconiine genera feed on mature foliage (18). Related variation in diet-derived microbes, or in

366 host processes that determine which microbes persist through metamorphosis (10, 60), could
367 potentially underlie variation in adult-stage microbiomes.
368 In addition to the clear separation of *Heliconius* from other heliconiine genera (Fig.
369 1B), there was also a strong signal of the host phylogeny in heliconiine butterfly microbiomes
370 (Fig. 2). The strength of this signal was comparable to that observed in gut microbiomes of
371 mammals (61) and other host groups (42). This result is also in agreement with a recent survey of
372 neotropical butterflies, in which microbiomes were found to be phylogenetically structured
373 across a broad diversity of hosts (six different families) (23). Thus, even in a comparatively
374 young radiation such as the one studied here (41), host phylogenetic history is clearly an
375 important factor shaping the composition of adult butterfly microbiomes. The question then
376 becomes, how has this pattern (also known as phylosymbiosis (62)) arisen, and what does it
377 signify?
378 There are at least two, non-mutually exclusive processes that can lead to
379 phylosymbiosis: host-microbe co-diversification, and contemporary host filtering of
380 environmental microbes (42, 63). In the former process, there is a shared evolutionary history of
381 host and microbial lineages. This does not necessarily apply to the latter process, which occurs
382 when host traits that influence microbial colonization, such as diet preference, are conserved to
383 some degree across the host phylogeny. We suggest that co-diversification is unlikely in this
384 system given the broad distribution of bacterial phylotypes across host species and genera
385 (discussed below). Rather, host filtering of environmental microbes likely explains phylogenetic
386 structure in heliconiine microbiomes, as is the case in a wide variety of other animal groups (42).
387 A priority for future work on butterflies is to identify the specific traits underlying host filtering.
388 In Heliconiini, two candidates are larval host plant use (i.e., *Passiflora* species identity and age

389 of tissue consumed) and adult foraging behavior (especially whether pollen is collected, and
390 from which plant species), both of which exhibit phylogenetic signal (18, 64, 65).
391 Bacterial community-level variation among butterfly genera and across the phylogeny
392 was largely driven by shifts in the relative abundance of shared bacterial genera and ESVs, as
393 opposed to the presence or absence of host-specific bacterial taxa (Figs. 3, 4). Moreover, a finer-
394 resolution analysis of the nearly ubiquitous bacterium *Orbus* did not find evidence for specificity
395 within butterflies (Fig. 5). This pattern is notable in part because it weighs against the co-
396 diversification model described above, and in part because it provides a contrast to the high
397 degree of host specificity documented in a number of other insect-microbe symbioses (66–69).
398 While not highly host-specific, at least some of the dominant bacteria in butterflies appear to be
399 insect specialists as opposed to cosmopolitan environmental taxa. For example, phylogenetic
400 evidence supports a general insect association for *Entomomonas* (38), *Apibacter* (70), and *Orbus*
401 (Fig. 5) (71). The ecology of these groups is poorly understood; one important unknown is how
402 they are transmitted among butterflies, which may be via flowers (72) or other shared resources.
403 Shotgun metagenomes allowed us to corroborate the amplicon-based bacterial data (Fig.
404 S2) and led to the discovery of putative eukaryotic parasites (trypanosomatids and microsporidia)
405 in gut tissue of many adult heliconiine butterflies (Fig. S5). Beyond fungi (12, 23), which do not
406 appear to be abundant in heliconiines, microeukaryotes are almost unknown from adult-stage
407 Lepidoptera—with the exception of the neogregarine *Ophryocystis elektroscirrha* in monarch
408 butterflies (73). Given their potential interactions with gut bacteria and relevance to host fitness,
409 more targeted analyses of butterfly-associated microeukaryote diversity are clearly warranted.
410 While effects of endosymbionts such as *Wolbachia* and *Spiroplasma* have been
411 documented (74), a major open question is what functional role gut microbes play in the biology

412 of adult heliconiines and Lepidoptera generally (75). In *Heliconius*, our hypothesis that microbes
413 mediate pollen-feeding is not strongly supported. To directly influence pollen digestion—which
414 occurs extraorally using saliva exuded from the proboscis (76)—microbes would likely need to
415 inhabit the proboscis or salivary gland. Yet we saw no clear signal of elevated microbial
416 abundances or unique microbial taxa in these tissues in *Heliconius* as compared with non-pollen-
417 feeding butterflies (Fig. S3). Endogenous pollen-digestion mechanisms (59, 77, 78) may be
418 sufficient.

419 Other microbial roles in host nutrition are possible, although a recent experiment on the
420 butterfly *Speyeria mormonia* did not find evidence for nutrition-related functions (79). We
421 suggest that colonization resistance (i.e., protection from pathogens and parasites), which is a
422 feature common to many symbioses (63, 80), could be a primary ecological function of adult
423 butterfly gut microbiomes. We found that heliconiines are frequently infected by
424 trypanosomatids and microsporidia, as well as *Serratia* (Fig. 4), which are common opportunistic
425 bacterial pathogens in insects (81, 82). Adult butterflies may be similar to pyrrhocorid bugs,
426 tsetse flies and social bees, in which gut bacteria provide an important layer of defense against
427 related parasites and pathogens (83–87). Importantly, colonization resistance can readily evolve
428 even in symbioses lacking strong host-microbe specificity (88), such as that between heliconiines
429 and their adult-stage microbiomes.

430

431 **Conclusions**

432

433 This study provides an in-depth characterization of *Heliconius* and other passion-vine
434 butterfly microbiomes, adding a new dimension to a classic model system in evolutionary

435 biology. The characteristics of adult-stage microbiomes we report contrast with larval
436 Lepidoptera, emphasizing that holometabolous insects are able to interact with microbes in very
437 different ways across life stages (10). However, many Lepidoptera do not feed or even have a
438 digestive tract as adults (89), and we do not yet know whether patterns from heliconiines or other
439 butterflies are generalizable to these other groups. We further suggest that heliconiines may serve
440 as a useful system for exploring how animal ecology and life history relate to microbes.
441 *Heliconius* host phylogenetically structured microbiomes that are markedly distinct from those of
442 their close relatives, and this finding sets the stage for experiments to test the specific host traits
443 that may be involved. Finally, we uncovered novel microeukaryote diversity in butterflies, and
444 hypothesize that they are parasites that interact antagonistically with gut bacteria. Further
445 research on adult-stage microbiomes will help advance our understanding of both insect-
446 microbiome evolution and of the biology of Lepidoptera, a diverse group of considerable
447 ecological, societal, and scientific importance (89, 90).
448

449 **Acknowledgments**

450

451 This material is based on work supported by the Cooperative Institute for Research in
452 Environmental Sciences, the Ecology and Evolutionary Biology Department at the University of
453 Colorado Boulder, a National Science Foundation Graduate Research Fellowship (1144083), and
454 a postdoctoral fellowship from the USDA National Institute of Food and Agriculture (2018-
455 08156), to TJH. This work was further supported by grants from the Simons Foundation
456 (429440) and by funding from the Smithsonian Tropical Research Institute. We thank Jessica
457 Henley for assistance with sequencing, the Smithsonian Tropical Research Institute staff for

458 administrative support and the Panamanian and Ecuadorian Authorities for permission to collect
459 butterflies. We also acknowledge the reviewers for helpful comments that improved the
460 manuscript.
461
462 **References**
463
464 1. Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. John Wiley &
465 Sons.

466 2. Jones RT, Sanchez LG, Fierer N. 2013. A Cross-Taxon Analysis of Insect-Associated
467 Bacterial Diversity. PLoS ONE 8:e61218.

468 3. Engel P, Moran NA. 2013. The gut microbiota of insects – diversity in structure and
469 function. FEMS Microbiology Reviews 37:699–735.

470 4. Hammer TJ, Sanders JG, Fierer N. 2019. Not all animals need a microbiome. FEMS
471 Microbiology Letters 366:fnz117.

472 5. Hammer TJ, Janzen DH, Hallwachs W, Jaffe SP, Fierer N. 2017. Caterpillars lack a
473 resident gut microbiome. Proceedings of the National Academy of Sciences 114:9641–
474 9646.

475 6. Staudacher H, Kaltenpoth M, Breeuwer JAJ, Menken SBJ, Heckel DG, Groot AT. 2016.
476 Variability of bacterial communities in the moth *Heliothis virescens* indicates transient
477 association with the host. PLoS ONE 11:e0154514.

478 7. Priya NG, Ojha A, Kajla MK, Raj A, Rajagopal R. 2012. Host Plant Induced Variation in
479 Gut Bacteria of *Helicoverpa armigera*. PLoS ONE 7:e30768.

480 8. Whitaker M, Pierce N, Salzman S, Kaltenpoth M, Sanders J. 2016. Microbial communities
481 of lycaenid butterflies do not correlate with larval diet. Frontiers in Microbiology 7:1920.

482 9. Phalnikar K, Kunte K, Agashe D. 2019. Disrupting butterfly caterpillar microbiomes does
483 not impact their survival and development. Proc R Soc B 286:20192438.

484 10. Hammer TJ, Moran NA. 2019. Links between metamorphosis and symbiosis in
485 holometabolous insects. Phil Trans R Soc B 374:20190068.

486 11. Hammer TJ, McMillan WO, Fierer N. 2014. Metamorphosis of a Butterfly-Associated
487 Bacterial Community. PLoS ONE 9:e86995.

488 12. Harrison JG, Urruty DM, Forister ML. 2016. An exploration of the fungal assemblage in
489 each life history stage of the butterfly, *Lycaeides melissa* (Lycaenidae), as well as its host
490 plant *Astragalus canadensis* (Fabaceae). Fungal Ecology 22:10–16.

491 13. Paniagua Voirol LR, Weinhold A, Johnston PR, Fatouros NE, Hilker M. 2020. Legacy of a
492 Butterfly's Parental Microbiome in Offspring Performance. Appl Environ Microbiol
493 86:e00596-20.

494 14. Phalnikar K, Kunte K, Agashe D. 2018. Dietary and developmental shifts in butterfly-
495 associated bacterial communities. Royal Society Open Science 5:171559.

496 15. Wang X, Sun S, Yang X, Cheng J, Wei H, Li Z, Michaud JP, Liu X. 2020. Variability of
497 Gut Microbiota Across the Life Cycle of *Grapholita molesta* (Lepidoptera: Tortricidae).
498 Front Microbiol 11:1366.

499 16. Chase AB, Karaoz U, Brodie EL, Gomez-Lunar Z, Martiny AC, Martiny JBH. 2017.
500 Microdiversity of an Abundant Terrestrial Bacterium Encompasses Extensive Variation in
501 Ecologically Relevant Traits. mBio 8:e01809-17.

502 17. Bates HW. 1861. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera:
503 Heliconidae. Transactions of the Linnean Society of London 23:495–566.

504 18. Jiggins CD. 2016. The ecology and evolution of *Heliconius* butterflies. Oxford University
505 Press.

506 19. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE,
507 Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH,
508 Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby
509 EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world,
510 a new imperative for the life sciences. Proceedings of the National Academy of Sciences
511 110:3229–3236.

512 20. Challis RJ, Kumar S, Dasmahapatra KK, Jiggins CD, Blaxter M. 2016. Lepbase: the
513 Lepidopteran genome database. bioRxiv http://dx.doi.org/10.1101/056994.

514 21. Concha C, Wallbank RWR, Hanly JJ, Fenner J, Livraghi L, Rivera ES, Paulo DF, Arias C,
515 Vargas M, Sanjeev M, Morrison C, Tian D, Aguirre P, Ferrara S, Foley J, Pardo-Diaz C,
516 Salazar C, Linares M, Massardo D, Counterman BA, Scott MJ, Jiggins CD, Papa R, Martin

517 A, McMillan WO. 2019. Interplay between Developmental Flexibility and Determinism in
518 the Evolution of Mimetic Heliconius Wing Patterns. Current Biology 29:3996-4009.e4.

519 22. van Schooten B, Godoy-Vitorino F, McMillan WO, Papa R. 2018. Conserved microbiota
520 among young *Heliconius* butterfly species. PeerJ 6:e5502.

521 23. Ravenscraft A, Berry M, Hammer T, Peay K, Boggs C. 2019. Structure and function of the
522 bacterial and fungal gut microbiota of Neotropical butterflies. Ecological Monographs
523 89:e01346.

524 24. Sculfort O, Castro ECP, Kozak KM, Bak S, Elias M, Nay B, Llaurens V. 2020. Variation of
525 chemical compounds in wild Heliconiini reveals ecological factors involved in the
526 evolution of chemical defenses in mimetic butterflies. Ecol Evol 10:2677–2694.

527 25. Ambika Manirajan B, Ratering S, Rusch V, Schwiertz A, Geissler-Plaum R, Cardinale M,
528 Schnell S. 2016. Bacterial microbiota associated with flower pollen is influenced by
529 pollination type, and shows a high degree of diversity and species-specificity: The bacterial
530 microbiota of flower pollen. Environmental Microbiology 18:5161–5174.

531 26. Kešnerová L, Emery O, Troilo M, Liberti J, Erkosar B, Engel P. 2020. Gut microbiota
532 structure differs between honeybees in winter and summer. ISME J 14:801–814.

533 27. Kingsley VV. 1972. Persistence of intestinal bacteria in the developmental stages of the
534 monarch butterfly (*Danaus plexippus*). Journal of Invertebrate Pathology 20:51–58.

535 28. Hammer TJ, Dickerson JC, Fierer N. 2015. Evidence-based recommendations on storing
536 and handling specimens for analyses of insect microbiota. PeerJ 3:e1190.

537 29. Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes -
538 application to the identification of mycorrhizae and rusts. Mol Ecol 2:113–118.

539 30. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015.
540 Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. PLOS ONE
541 10:e0128036.

542 31. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
543 reads. EMBnet.journal 17:10–12.

544 32. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016.
545 DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods
546 13:581–583.

547 33. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid
548 assignment of rRNA sequences into the new bacterial taxonomy. Applied and
549 Environmental Microbiology 73:5261–5267.

550 34. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.
551 2012. The SILVA ribosomal RNA gene database project: improved data processing and
552 web-based tools. Nucleic Acids Research 41:D590–D596.

553 35. R Core Team. 2016. R: A language and environment for statistical computing. R
554 Foundation for Statistical Computing, Vienna, Austria.

555 36. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical
556 identification and removal of contaminant sequences in marker-gene and metagenomics
557 data. Microbiome 6:226.

558 37. Brady C, Cleenwerck I, Venter S, Coutinho T, De Vos P. 2013. Taxonomic evaluation of
559 the genus *Enterobacter* based on multilocus sequence analysis (MLSA). Systematic and
560 Applied Microbiology 36:309–319.

561 38. Wang J, Su Q, Zhang X, Li C, Luo S, Zhou X, Zheng H. 2020. *Entomomonas moraniae*
562 gen. nov., sp. nov., a member of the family Pseudomonadaceae isolated from Asian honey
563 bee gut, possesses a highly reduced genome. International Journal of Systematic and
564 Evolutionary Microbiology 70:165–171.

565 39. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado
566 DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic Placement of Exact
567 Amplicon Sequences Improves Associations with Clinical Information. mSystems
568 3:e00021-18.

569 40. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL,
570 Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for
571 ecological and evolutionary analyses of bacteria and archaea. The ISME Journal 6:610–
572 618.

573 41. Kozak KM, Wahlberg N, Neild AF, Dasmahapatra KK, Mallet J, Jiggins CD. 2015.
574 Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic *Heliconius*
575 Butterflies. Systematic Biology 64:505–524.

576 42. Mazel F, Davis KM, Loudon A, Kwong WK, Groussin M, Parfrey LW. 2018. Is Host
577 Filtering the Main Driver of Phylosymbiosis across the Tree of Life? mSystems 3:e00097-
578 18.

579 43. Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other
580 things). Methods in Ecology and Evolution 3:217–223.

581 44. Oksanen J, Kindt R, Legendre P. 2019. vegan: Community Ecology Package. R package
582 version 2.5-6.

583 45. Anderson MJ, Walsh DCI. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face
584 of heterogeneous dispersions: What null hypothesis are you testing? Ecological
585 Monographs 83:557–574.

586 46. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2019. Linear and Nonlinear Mixed
587 Effects Models.

588 47. Joshi, NA, Fass, JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for
589 FastQ files. https://github.com/najoshi/sickle.

590 48. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods
591 9:357–359.

592 49. Gruber-Vodicka HR, Seah BK, Pruesse E. 2019. phyloFlash — Rapid SSU rRNA profiling
593 and targeted assembly from metagenomes. bioRxiv https://doi.org/10.1101/521922.

594 50. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
595 throughput. Nucleic Acids Research 32:1792–1797.

596 51. Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use
597 in Phylogenetic Analysis. Molecular Biology and Evolution 17:540–552.

598 52. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon
599 S, Lefort V, Lescot M, Claverie J-M, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic
600 analysis for the non-specialist. Nucleic Acids Research 36:W465–W469.

601 53. Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood
602 phylgenies with PhyML. Methods in Molecular Biology 537:113–137.

603 54. Saridaki A, Bourtzis K. 2010. Wolbachia: more than just a bug in insects genitals. Current
604 Opinion in Microbiology 13:67–72.

605 55. Kim JY, Lee J, Shin N-R, Yun J-H, Whon TW, Kim M-S, Jung M-J, Roh SW, Hyun D-W,
606 Bae J-W. 2012. *Orbus sasakiae* sp. nov., a bacterium isolated from the gut of the butterfly
607 *Sasakia charonda*, and emended description of the genus *Orbus*. International Journal of
608 Systematic and Evolutionary Microbiology 63:1766–1770.

609 56. De Souza W, Motta MCM. 1999. Endosymbiosis in protozoa of the Trypanosomatidae
610 family. FEMS Microbiology Letters 173:1–8.

611 57. Zheng H, Steele MI, Leonard SP, Motta EVS, Moran NA. 2018. Honey bees as models for
612 gut microbiota research. Lab Animal 47:317–325.

613 58. Brown KS. 1981. The Biology of *Heliconius* and Related Genera. Annual Review of
614 Entomology 26:427–456.

615 59. Gilbert LE. 1972. Pollen Feeding and Reproductive Biology of *Heliconius* Butterflies.
616 Proceedings of the National Academy of Sciences 69:1403–1407.

617 60. Johnston PR, Rolff J. 2015. Host and Symbiont Jointly Control Gut Microbiota during
618 Complete Metamorphosis. PLOS Pathogens 11:e1005246.

619 61. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, Alm EJ. 2017.
620 Unraveling the processes shaping mammalian gut microbiomes over evolutionary time.
621 Nature Communications 8:14319.

622 62. Lim SJ, Bordenstein SR. 2020. An introduction to phylosymbiosis. Proc R Soc B
623 287:20192900.

624 63. Moran NA, Ochman H, Hammer TJ. 2019. Evolutionary and Ecological Consequences of
625 Gut Microbial Communities. Annu Rev Ecol Evol Syst 50:451-475.

626 64. Estrada C, Jiggins CD. 2002. Patterns of pollen feeding and habitat preference among
627 *Heliconius* species. Ecological Entomology 27:448–456.

628 65. Boggs CL, Smiley JT, Gilbert LE. 1981. Patterns of pollen exploitation by *Heliconius*
629 butterflies. Oecologia 48:284–289.

630 66. Ohkuma M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites.
631 Trends in Microbiology 16:345–352.

632 67. Kikuchi Y, Hosokawa T, Nikoh N, Meng X-Y, Kamagata Y, Fukatsu T. 2009. Host-
633 symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of
634 acanthosomatid stinkbugs. BMC Biol 7:2.

635 68. Kwong WK, Moran NA. 2015. Evolution of host specialization in gut microbes: The bee
636 gut as a model. Gut Microbes 6:214–220.

637 69. Kaltenpoth M, Roeser-Mueller K, Koehler S, Peterson A, Nechitaylo TY, Stubblefield JW,
638 Herzner G, Seger J, Strohm E. 2014. Partner choice and fidelity stabilize coevolution in a
639 Cretaceous-age defensive symbiosis. Proceedings of the National Academy of Sciences
640 111:6359–64.

641 70. Kwong WK, Moran NA. 2016. *Apibacter adventoris* gen. nov., sp. nov., a member of the
642 phylum Bacteroidetes isolated from honey bees. International Journal of Systematic and
643 Evolutionary Microbiology 66:1323–1329.

644 71. Martinson VG, Carpinteyro-Ponce J, Moran NA, Markow TA. 2017. A Distinctive and
645 Host-Restricted Gut Microbiota in Populations of a Cactophilic *Drosophila* Species.
646 Applied and Environmental Microbiology 83:e01551–17.

647 72. Graystock P, Goulson D, Hughes WOH. 2015. Parasites in bloom: flowers aid dispersal and
648 transmission of pollinator parasites within and between bee species. Proceedings of the
649 Royal Society B: Biological Sciences 282:20151371.

650 73. Altizer SM, Oberhauser KS. 1999. Effects of the Protozoan Parasite *Ophryocystis*
651 *elektroscirrha* on the Fitness of Monarch Butterflies (*Danaus plexippus*). Journal of
652 Invertebrate Pathology 74:76–88.

653 74. Duplouy A, Hornett EA. 2018. Uncovering the hidden players in Lepidoptera biology: the
654 heritable microbial endosymbionts. PeerJ 6:e4629.

655 75. Paniagua Voirol LR, Frago E, Kaltenpoth M, Hilker M, Fatouros NE. 2018. Bacterial
656 Symbionts in Lepidoptera: Their Diversity, Transmission, and Impact on the Host. Frontiers
657 in Microbiology 9:556.

658 76. Eberhard SH, Hikl AL, Boggs CL, Krenn HW. 2009. Saliva or Regurgitated Nectar? What
659 *Heliconius* Butterflies (Lepidoptera: Nymphalidae) Use for Pollen Feeding. Annals of the
660 Entomological Society of America 102:1105–1108.

661 77. Harpel D, Cullen DA, Ott SR, Jiggins CD, Walters JR. 2015. Pollen feeding proteomics:
662 Salivary proteins of the passion flower butterfly, *Heliconius melpomene*. Insect
663 Biochemistry and Molecular Biology 63:7–13.

664 78. Smith G, Macias-Muñoz A, Briscoe AD. 2016. Gene Duplication and Gene Expression
665 Changes Play a Role in the Evolution of Candidate Pollen Feeding Genes in *Heliconius*
666 Butterflies. Genome Biology and Evolution 8:2581–2596.

667 79. Ravenscraft A, Kish N, Peay K, Boggs C. 2019. No evidence that gut microbiota impose a
668 net cost on their butterfly host. Molecular Ecology 28:2100–2117.

669 80. Kamada N, Chen GY, Inohara N, Núñez G. 2013. Control of pathogens and pathobionts by
670 the gut microbiota. Nature Immunology 14:685–690.

671 81. Raymann K, Coon KL, Shaffer Z, Salisbury S, Moran NA. 2018. Pathogenicity of *Serratia*
672 *marcescens* Strains in Honey Bees. mBio 9:e01649-18.

673 82. Grimont F, Grimont PAD. 2006. The Genus *Serratia*, p. 219–244. *In* Dworkin, M, Falkow,
674 S, Rosenberg, E, Schleifer, K-H, Stackebrandt, E (eds.), The Prokaryotes. Springer New
675 York.

676 83. Onchuru TO, Martinez AJ, Kaltenpoth M. 2018. The cotton stainer's gut microbiota
677 suppresses infection of a cotransmitted trypanosomatid parasite. Molecular Ecology
678 27:3408–3419.

679 84. Raymann K, Shaffer Z, Moran NA. 2017. Antibiotic exposure perturbs the gut microbiota
680 and elevates mortality in honeybees. PLoS Biology 15:e2001861.

681 85. Koch H, Schmid-Hempel P. 2011. Socially transmitted gut microbiota protect bumble bees
682 against an intestinal parasite. Proceedings of the National Academy of Sciences 108:19288–
683 19292.

684 86. Weiss BL, Maltz MA, Vigneron A, Wu Y, Walter KS, O'Neill MB, Wang J, Aksoy S.
685 2019. Colonization of the tsetse fly midgut with commensal *Kosakonia cowanii* Zambiae
686 inhibits trypanosome infection establishment. PLOS Pathogens 15:e1007470.

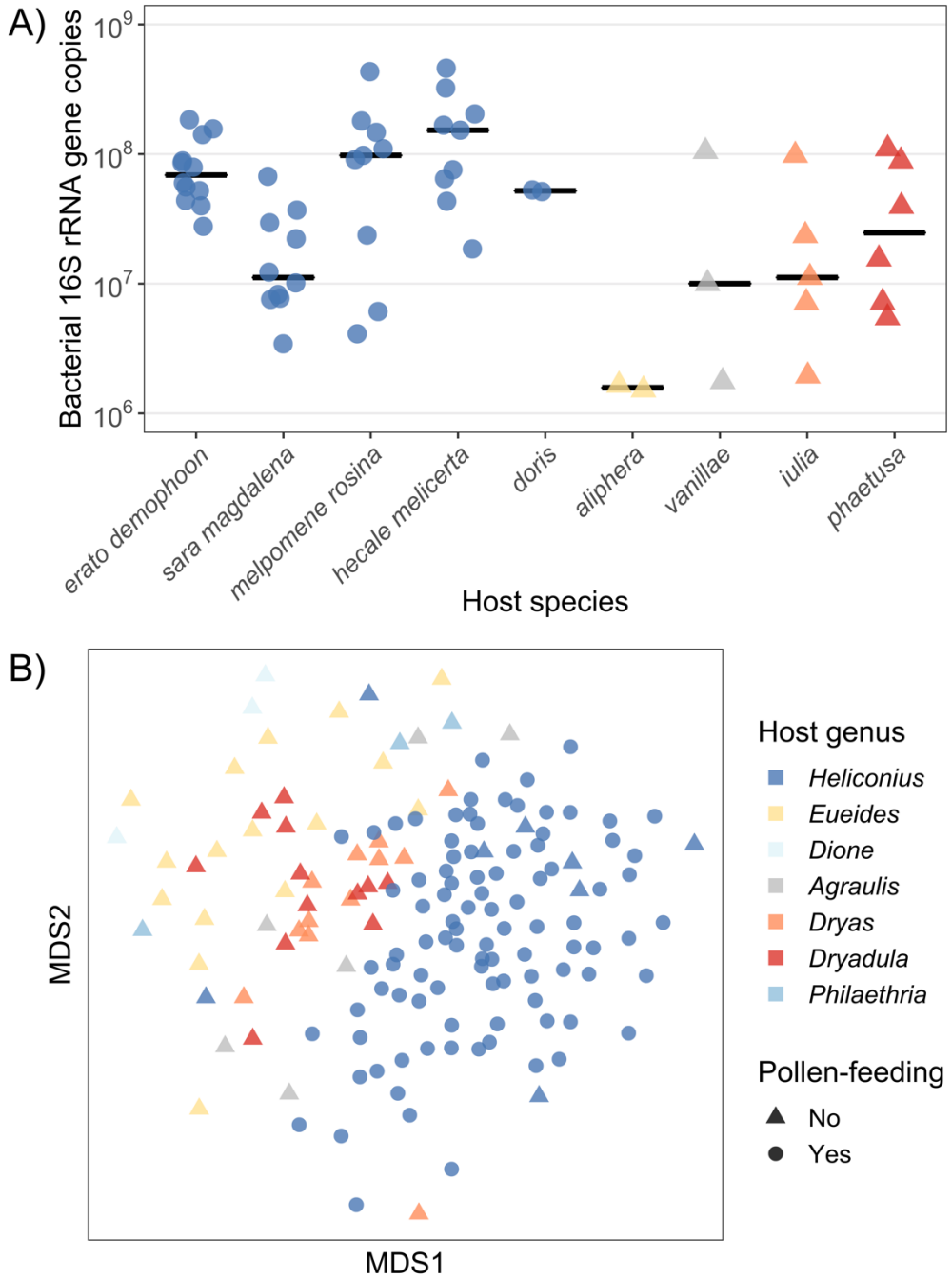
687 87. Wu Y, Zheng Y, Chen Y, Chen G, Zheng H, Hu F. 2020. *Apis cerana* gut microbiota
688 contribute to host health though stimulating host immune system and strengthening host
689 resistance to *Nosema ceranae*. R Soc Open Sci 7:192100.

690 88. McLaren MR, Callahan BJ. 2020. Pathogen resistance may be the principal evolutionary
691 advantage provided by the microbiome. Phil Trans R Soc B 375:20190592.

692 89. Scoble MJ. 1992. The Lepidoptera. Oxford University Press.

90. Boggs CL, Watt WB, Ehrlich PR. 2003. Butterflies: ecology and evolution taking flight. The University of Chicago Press.

696 **Figures**
697



698
699
700 Figure 1. A) Adult *Heliconius* butterflies have high, though often variable, titers of gut bacteria. Shown
701 are quantitative PCR-derived estimates of bacterial abundances in terms of the number of bacterial 16S
702 rRNA gene copies per individual gut. These individuals were collected from Gamboa, Panama in 2016 (N
703 = 42 *Heliconius*, 16 other Heliconiini). B) *Heliconius* host distinct adult-stage bacterial communities
704 compared with related genera. Shown is an ordination of microbiome variation (Bray-Curtis
705 dissimilarities) among all whole-body samples collected at various sites in Panama and Ecuador in 2014
706 (N = 104 *Heliconius*, 52 other Heliconiini). With the exception of *H. aoede* (dark blue triangles), pollen-
707 feeding is exclusive to, and ubiquitous within *Heliconius*.

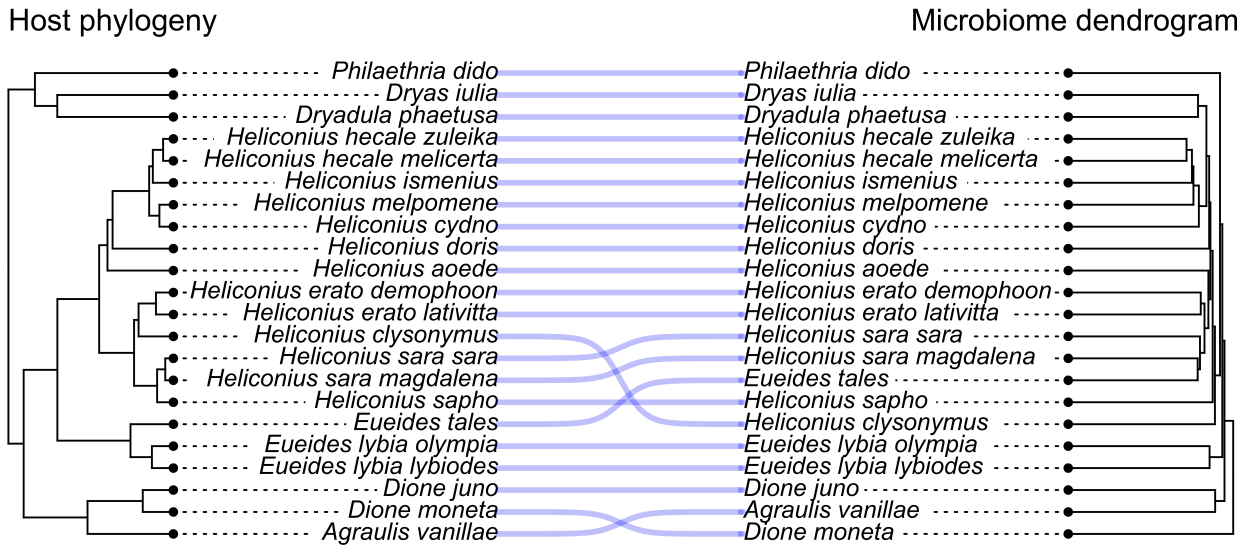


Figure 2. Variation in heliconiine butterfly microbiomes is correlated with host relatedness. Shown is the concordance between the host phylogeny (from (41)) and a dendrogram representation of microbiome variation among species (Bray-Curtis dissimilarities). Here, all nodes have been rotated to maximize tip matching. Note, however, that sets of parallel lines connecting tips between the phylogeny and the dendrogram do not always signify congruent branching structure.

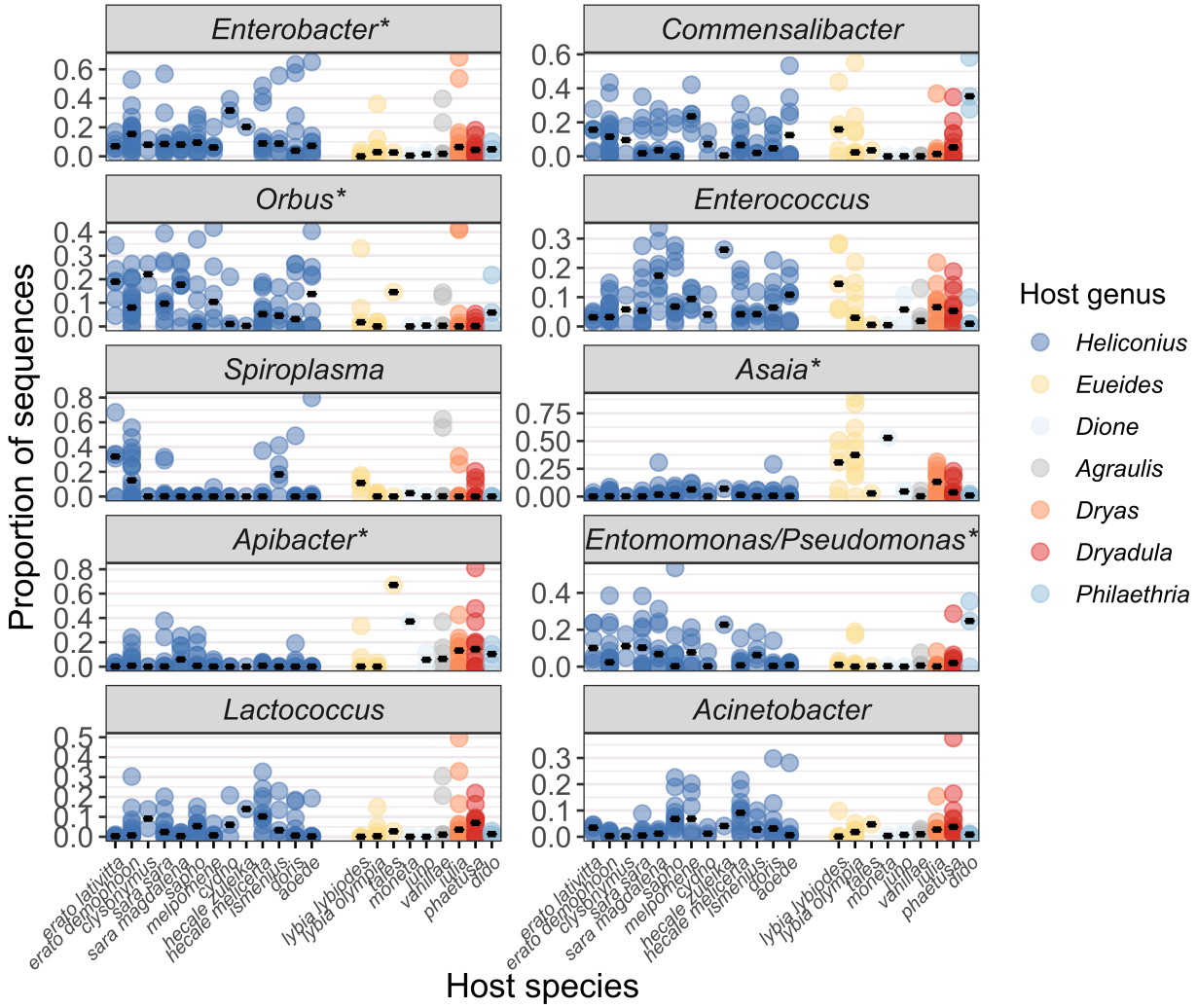


Figure 3. Dominant bacterial genera are largely shared among *Heliconius* and other butterflies, although some are differentially abundant. Shown are the relative abundances of the top 10 bacterial genera, ranked by mean abundance, in whole-body microbiomes. Dots indicate replicate individuals, and black bars indicate median proportions within a host species. Starred bacterial genera differed significantly in relative abundance between *Heliconius* and non-*Heliconius* butterflies ($p < 0.05$ after FDR correction). The arrangement of host species on the x axis corresponds to the phylogeny shown in Fig. 4. Note that *Enterobacter* here includes sequences originally assigned as *Klebsiella* and some other closely related Enterobacteriaceae genera (see Methods).

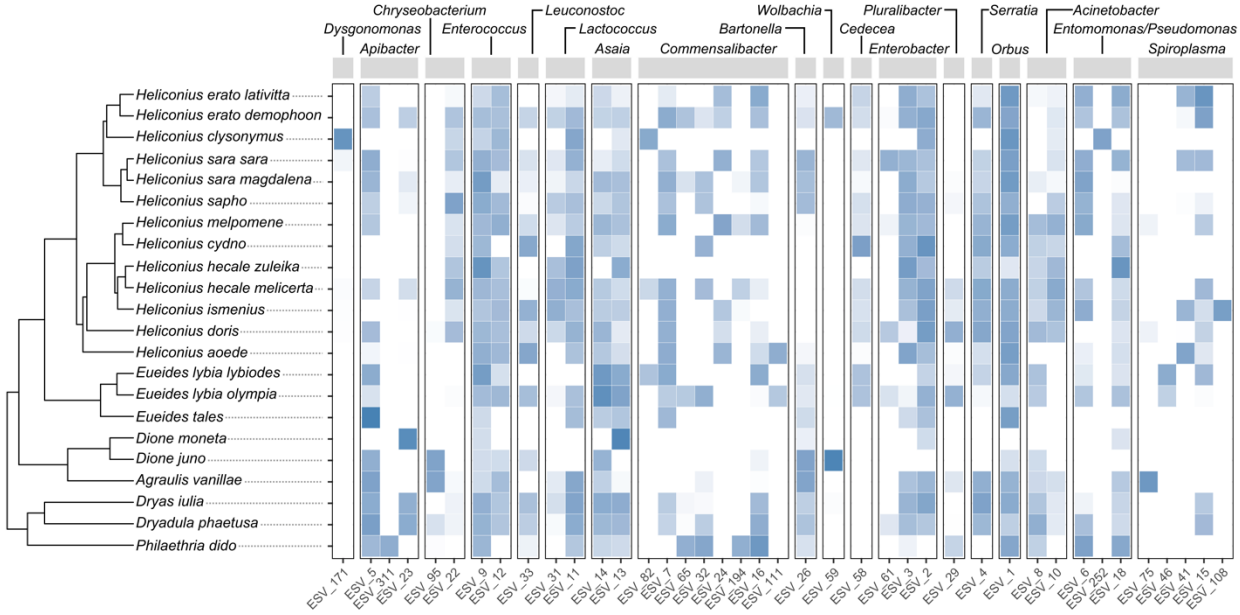


Figure 4. Most bacterial exact sequence variants (ESVs) are not host species- or host genus-specific. Shown are all ESVs in the dataset that had ≥ 5% mean relative abundance across conspecific individuals for one or more host species. For each ESV, the species-level mean relative abundance in whole-body samples, after log transformation, is indicated by the color of the cells (white = not detected in that species). Note that bacterial genera (labeled at the top) contained varying numbers of ESVs that met the aforementioned prevalence threshold.

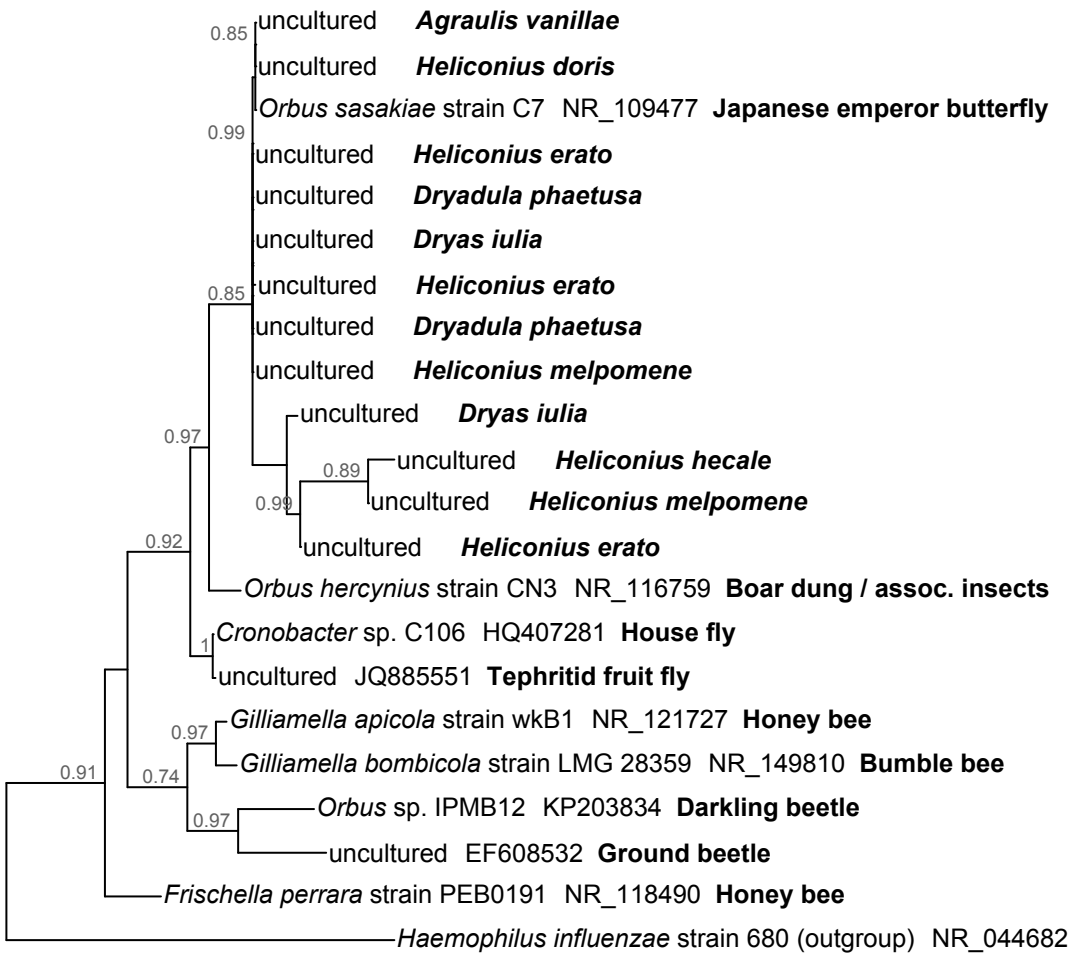
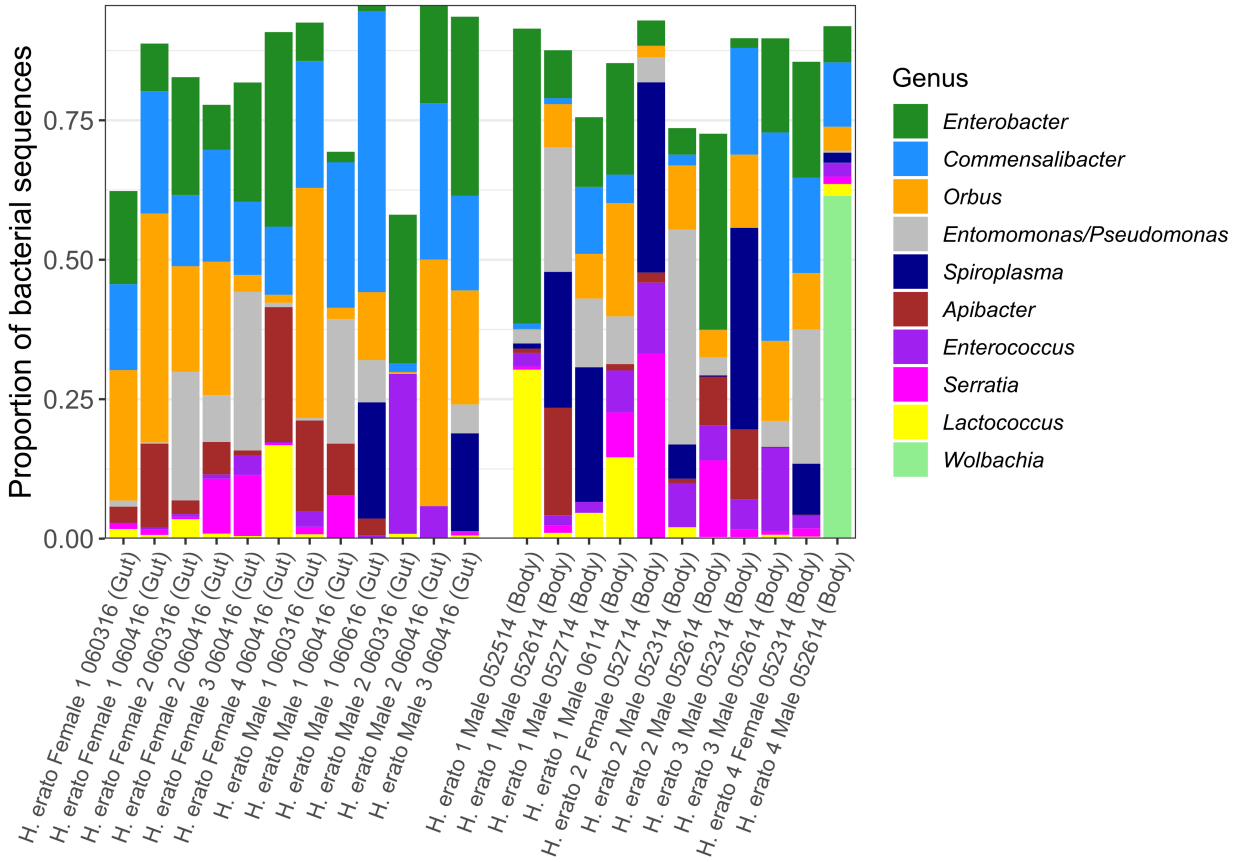


Figure 5. Maximum likelihood phylogenetic reconstruction of the bacterial family Orbaceae (Gammaproteobacteria: Pasteurellales). Some host taxonomic structure is apparent at the order level (i.e. Lepidoptera, beetles, bees, flies) but not within butterflies. 16S rRNA gene sequences from *Agraulis*, *Dryadula*, *Dryas*, and *Heliconius* were assembled from short metagenomic reads. Other 16S rRNA gene sequences are from GenBank. Branch support values are shown next to nodes. *Haemophilus influenzae* (Pasteurellaceae) was used as the outgroup.

741 **Supplementary Figures**

742



743

744

745 Figure S1. Whole-body bacterial communities are consistent among individuals and predominantly
746 represent gut-associated taxa. Shown are the relative abundances of the top 10 bacterial genera (ranked by
747 mean abundance) across *H. erato demophoon* collected from Gamboa, Panama. Remaining white space
748 represents sequences belonging to other genera; these made up a median 16% of sequence libraries across
749 individuals. Samples on the left are isolated guts from individuals collected in 2016, while samples on the
750 right are whole-body homogenates from individuals collected in 2014. Note that *Enterobacter* here
751 includes sequences originally assigned as *Klebsiella* and some other closely related Enterobacteriaceae
752 genera (see Methods).

753

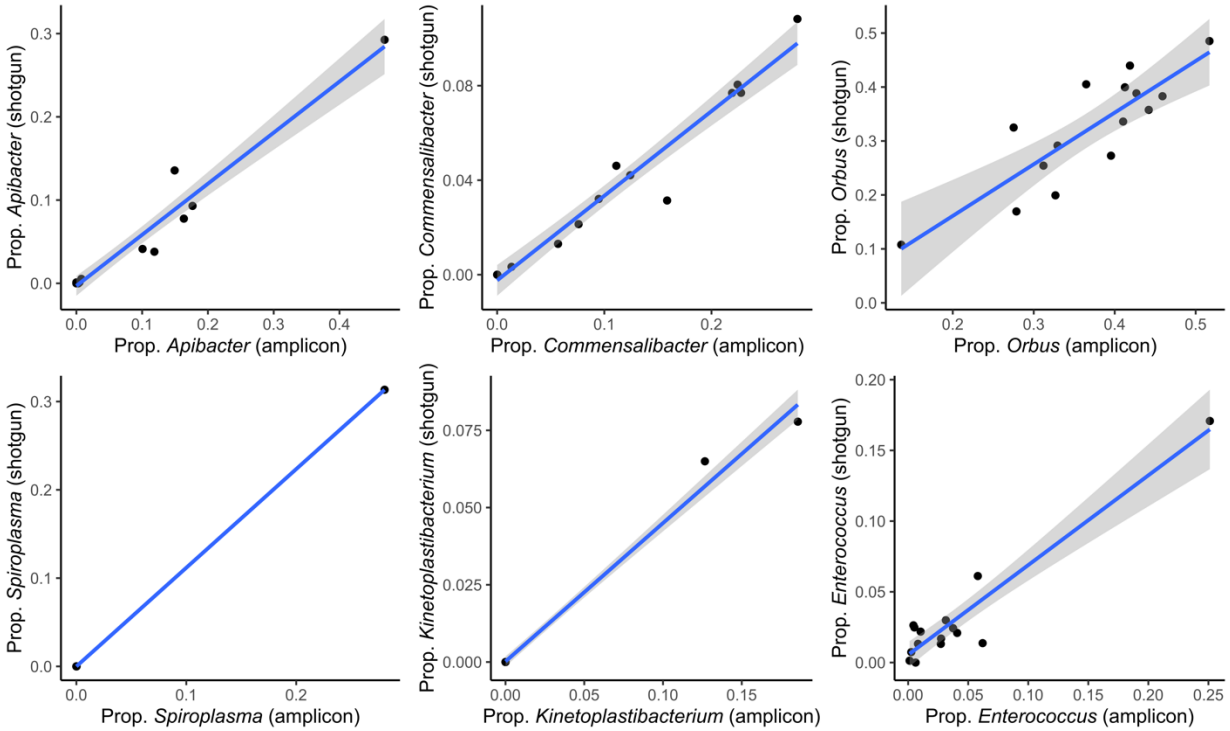


Figure S2. Bacterial genus-level relative abundances are highly correlated between the amplicon sequence libraries and metagenomes. Shown are relative abundances of six of the most abundant genera across the 15 butterfly gut samples for which we obtained both amplicon and shotgun metagenomic sequence data.

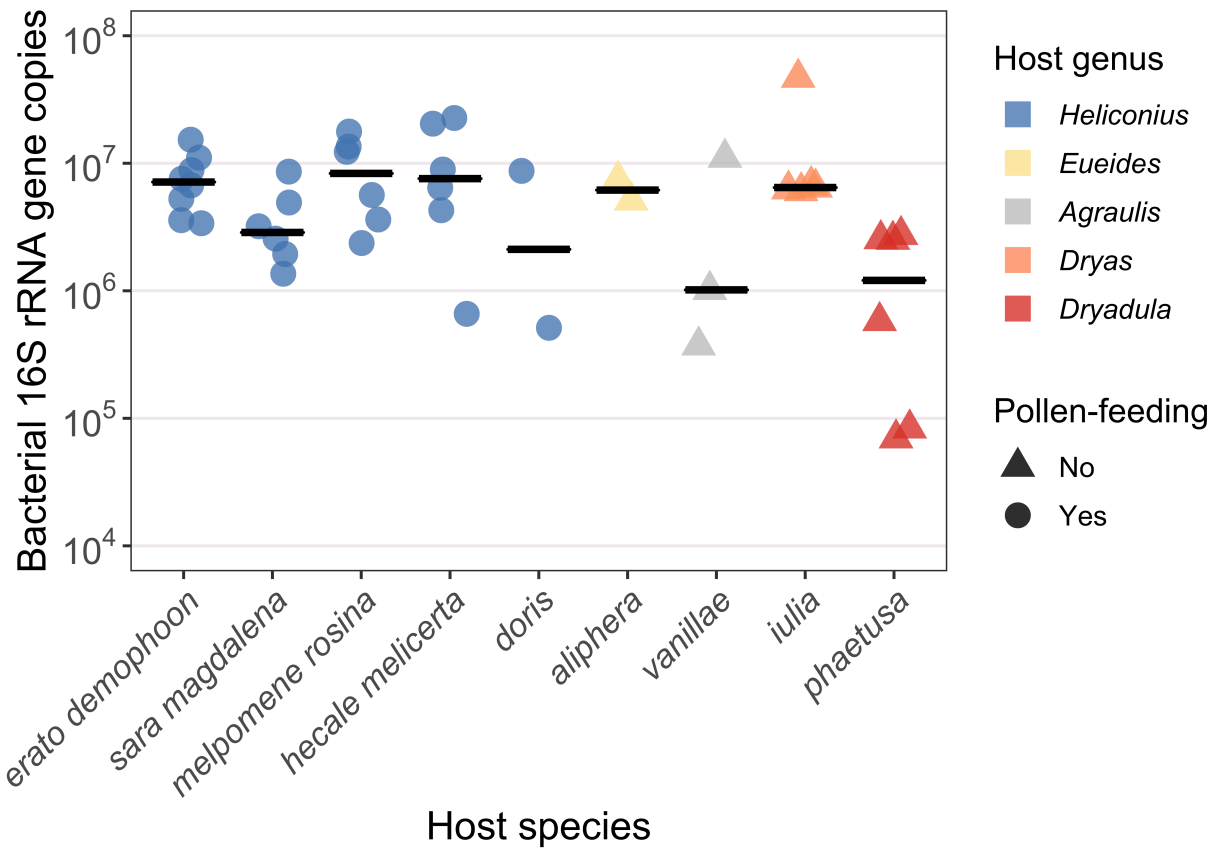


Figure S3. Absolute abundances of bacteria in the head and thorax of adult heliconiine butterflies, derived from qPCR. Shown are the number of bacterial 16S rRNA gene copies per subsample (approx. 50 mg) of homogenized, combined head and thorax tissue. These individuals were collected from Gamboa, Panama in 2016 (N = 28 *Heliconius*, 16 other Heliconiini).

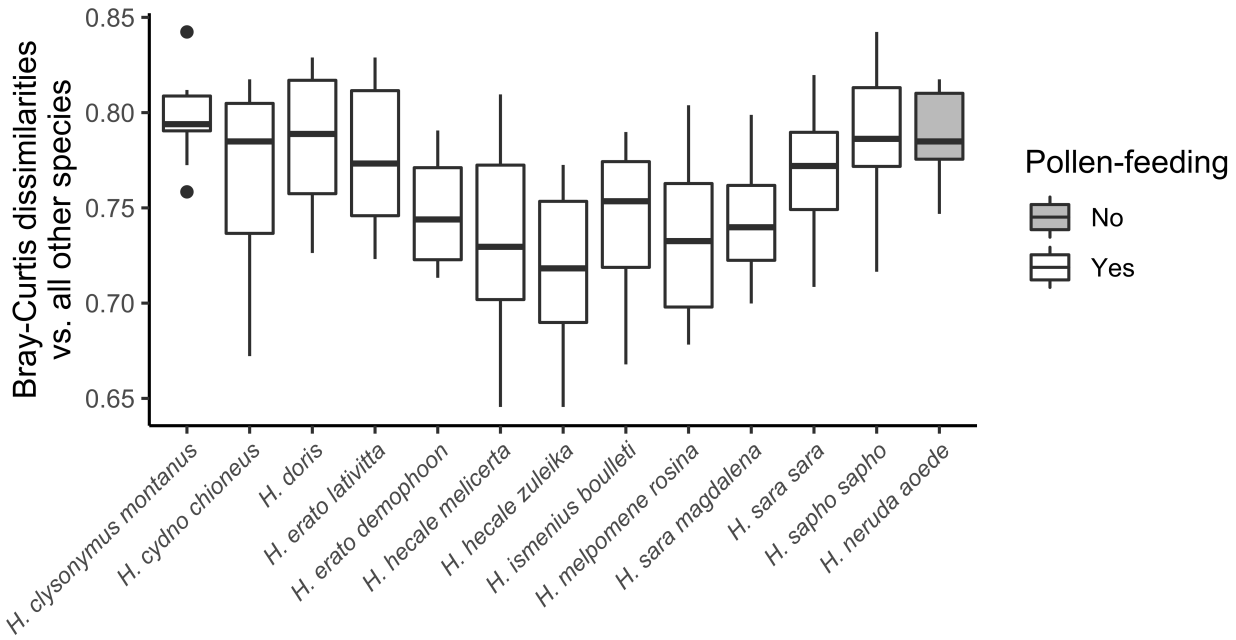


Figure S4. The non-pollen-feeding species *Heliconius aoede* is not uniquely distinct from pollen-feeding *Heliconius* species. Whole-body microbiomes from each focal species (x axis) are compared to all other species using Bray-Curtis dissimilarities. For a given comparison of two species' microbiomes, intraspecific replication was handled by averaging dissimilarities among all pairs of individuals.

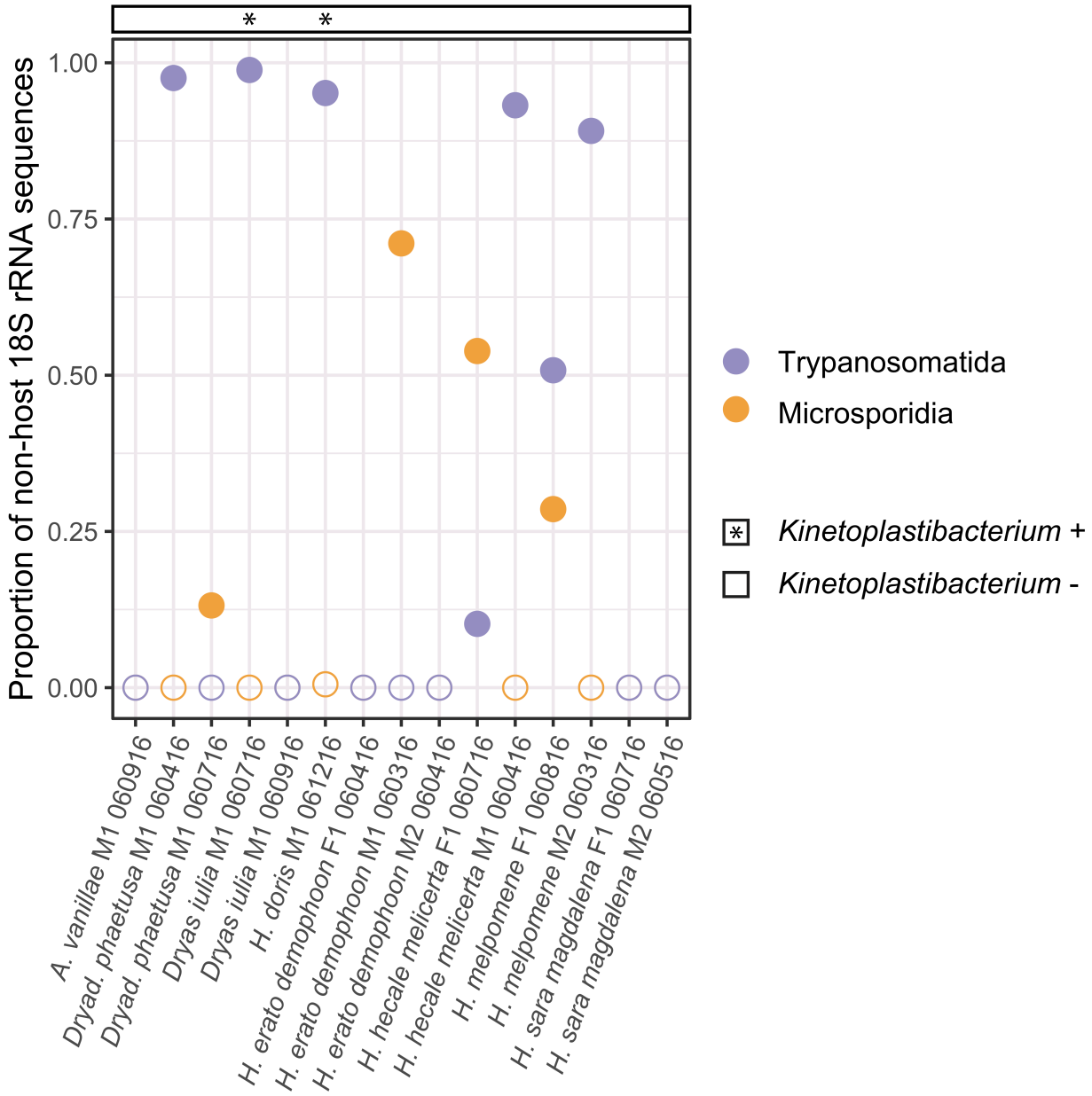


Figure S5. Microsporidia and Trypanosomatidae (related to *Crithidia*, *Leishmania*, and *Trypanosoma*) are prevalent in the 15 butterfly individuals with sequenced gut metagenomes. Points show the proportion of each microeukaryotic taxon out of all non-host 18S rRNA sequences identified by phyloFlash. Circles are filled if the proportion of microsporidia or trypanosomatids exceeded 0.01. Asterisks indicate samples in which *Kinetoplastibacterium*, a bacterial endosymbiont of trypanosomatids, was detected in metagenomes. Other microeukaryotes not shown here comprise a variety of very low-abundance taxa (i.e. ≤ 30 total reads per metagenome) classified mainly as coccidia, ascomycete fungi, acanthamoeba, and algae.

784 **Supplementary Table**

785

Host taxon	Phylum	Class	Order	Family	Genus	Mean prop. gut +/- SEM	Mean prop. head/thorax +/- SEM	p value
Heliconius	Bacteroidetes	Bacteroidia	Flavobacteriales	Weeksellaceae	*Apibacter*	0.036 +/- 0.009	0.002 +/- 0.001	< 0.001
					Chishuiella	0.007 +/- 0.003	0.095 +/- 0.027	0.001
					Chryseobacterium	0.011 +/- 0.004	0.096 +/- 0.024	0.001
	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	*Enterococcus*	0.031 +/- 0.008	0.021 +/- 0.006	0.117
				Streptococcaceae	*Lactococcus*	0.02 +/- 0.006	0.033 +/- 0.009	0.793
	Proteobacteria	Alphaproteobacteria	Acetobacterales	Acetobacteraceae	*Asaia*	0.024 +/- 0.007	0.004 +/- 0.002	0.001
					Commensalibacter	0.12 +/- 0.02	0.007 +/- 0.003	< 0.001
					Swaminathania	0.013 +/- 0.01	0.016 +/- 0.014	0.745
		Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Variovorax*	0.01 +/- 0.004	0.021 +/- 0.006	0.017
			Enterobacteriales	Enterobacteriaceae	*Enterobacter*	0.233 +/- 0.034	0.031 +/- 0.007	< 0.001
					Serratia	0.037 +/- 0.01	0.027 +/- 0.017	0.013
			Orbales	Orbaceae	*Orbus*	0.163 +/- 0.024	0.021 +/- 0.01	< 0.001
			Pseudomonadales	Moraxellaceae	*Acinetobacter*	0.023 +/- 0.007	0.293 +/- 0.041	< 0.001
				Pseudomonadaceae	*Entomomonas/Pseudomonas*	0.086 +/- 0.017	0.024 +/- 0.01	0.291
	Tenericutes	Mollicutes	Entomoplasmatales	Spiroplasmataceae	*Spiroplasma*	0.012 +/- 0.007	0.003 +/- 0.002	0.927
Other Heliconiini	Bacteroidetes	Bacteroidia	Flavobacteriales	Weeksellaceae	*Apibacter*	0.139 +/- 0.045	0.008 +/- 0.004	0.083
					Chishuiella	0.009 +/- 0.003	0.138 +/- 0.029	0.016
					Chryseobacterium	0 +/- 0	0.015 +/- 0.012	0.808
	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	*Enterococcus*	0.056 +/- 0.027	0.031 +/- 0.011	0.843
				Streptococcaceae	*Lactococcus*	0.036 +/- 0.026	0.012 +/- 0.011	0.098
	Proteobacteria	Alphaproteobacteria	Acetobacterales	Acetobacteraceae	*Asaia*	0.153 +/- 0.039	0.004 +/- 0.001	0.01
					Commensalibacter	0.031 +/- 0.015	0.002 +/- 0.002	0.106
					Swaminathania	0.007 +/- 0.004	0 +/- 0	0.226
		Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae	*Variovorax*	0.005 +/- 0.002	0.016 +/- 0.004	0.292
			Enterobacteriales	Enterobacteriaceae	*Enterobacter*	0.156 +/- 0.042	0.014 +/- 0.006	0.01
					Serratia	0.02 +/- 0.012	0.001 +/- 0	0.075
			Orbales	Orbaceae	*Orbus*	0.139 +/- 0.043	0.007 +/- 0.004	0.042
			Pseudomonadales	Moraxellaceae	*Acinetobacter*	0.046 +/- 0.018	0.45 +/- 0.054	< 0.001
				Pseudomonadaceae	*Entomomonas/Pseudomonas*	0.02 +/- 0.017	0.011 +/- 0.004	0.393
	Tenericutes	Mollicutes	Entomoplasmatales	Spiroplasmataceae	*Spiroplasma*	0.068 +/- 0.038	0.052 +/- 0.052	0.843

786
787
788 Table S1. Relative abundances and within-body distribution of the top 15 bacterial genera (ranked by mean
789 abundance) across butterflies collected from Gamboa, Panama in 2016. These individuals were dissected
790 to compare microbiomes between isolated gut tissue and the combined head and thorax. Abundances are
791 shown for *Heliconius* (top) and species belonging to other heliconiine genera (bottom) separately. P values
792 are from a nonparametric statistical test of proportions in guts versus head/thorax samples, after FDR
793 correction (see Methods).