SI Appendix for *Caterpillars lack a resident gut microbiome*

SI Figures

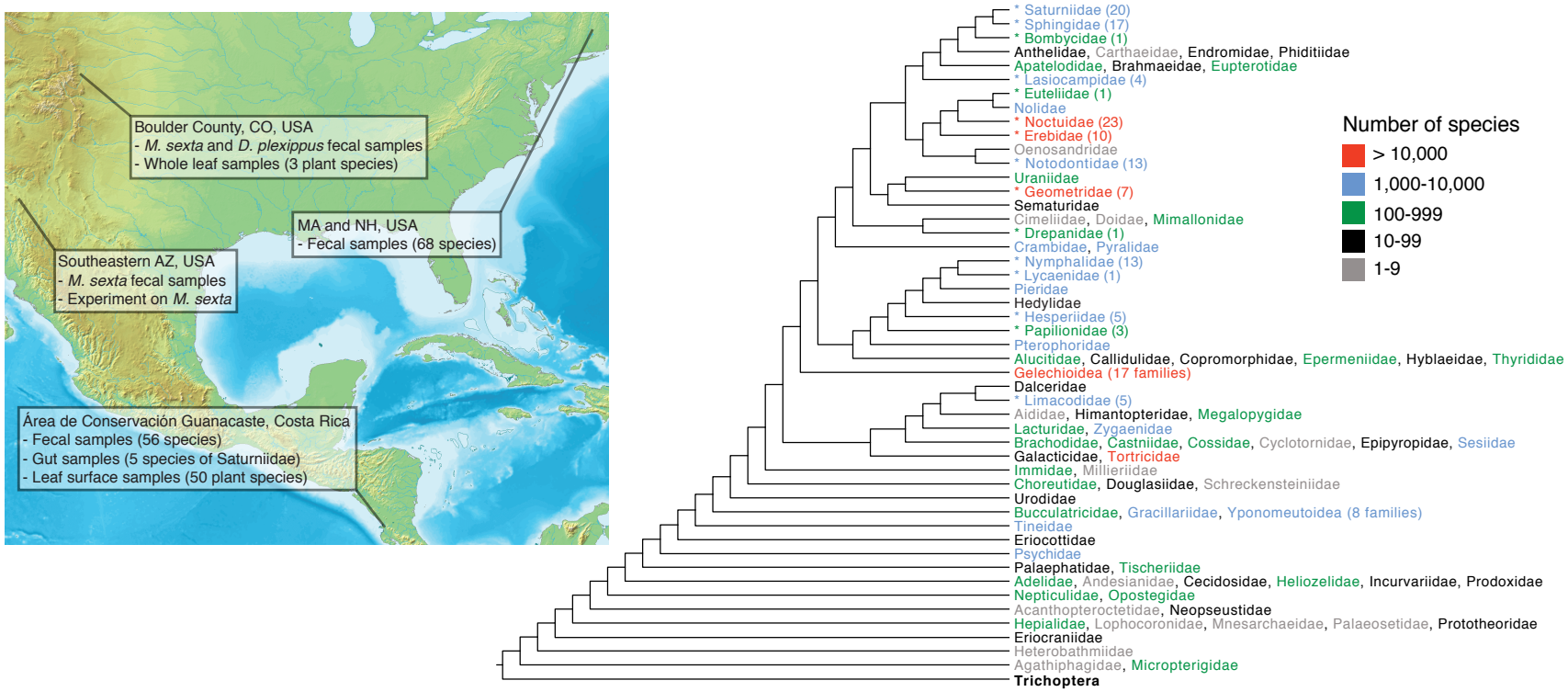

Figure S1. The caterpillar species included in this study and their phylogenetic placement (cladogram adapted from the "nt123_degen1" tree of Regier et al. (1)). Asterisks show sampled families of Lepidoptera, with the number of sampled species given in parentheses. To indicate the distribution of diversity (and taxonomic effort) across the tree, family names are colored by the number of described species given in (2). Inset: caterpillar sampling or experiment localities. Map from Wikimedia Commons.

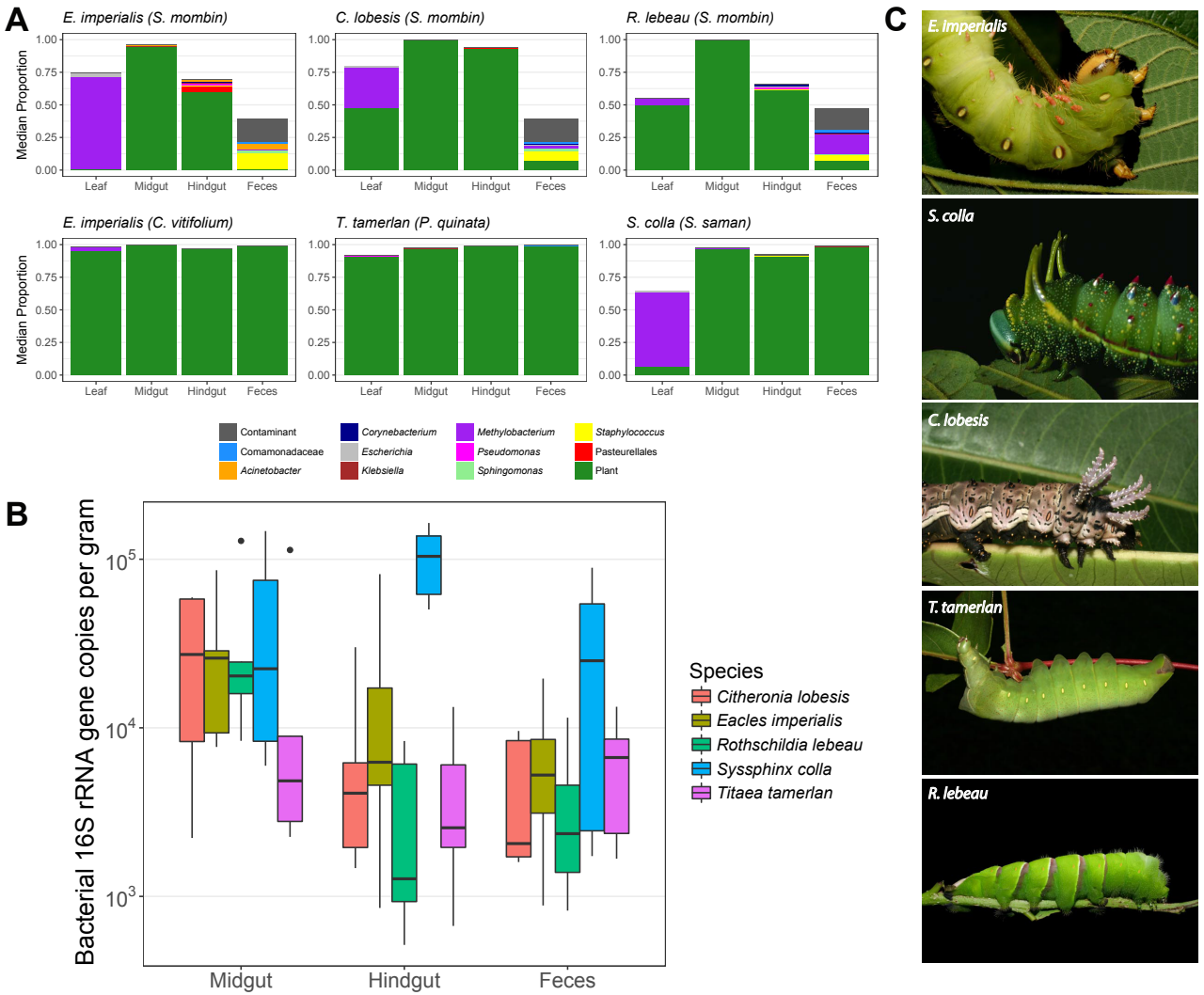

Figure S2. Bacterial community composition and density in five caterpillar species (Saturniidae). A) The composition of sequence libraries from the leaf surface, midgut, hindgut, and feces. The median across five replicate individuals is displayed. The food plant species is indicated in parentheses. Note that one species, *Eacles imperialis*, was reared separately on two plant species. Only plant chloroplast or mitochondrial sequences, reagent contaminants, and the top 10 bacterial genera (among the dissected individuals only) are shown; the remainder of the community (summing to 1) represents sequences from a variety of low-abundance taxa. B) The number of bacterial 16S rRNA gene copies per gram (fresh weight) in homogenized midgut or hindgut tissue and feces (N=5 individuals per species, except *E. imperialis* with 10 individuals (5 each on two plant species)). C) Photographs of each species taken in ACG, Costa Rica.

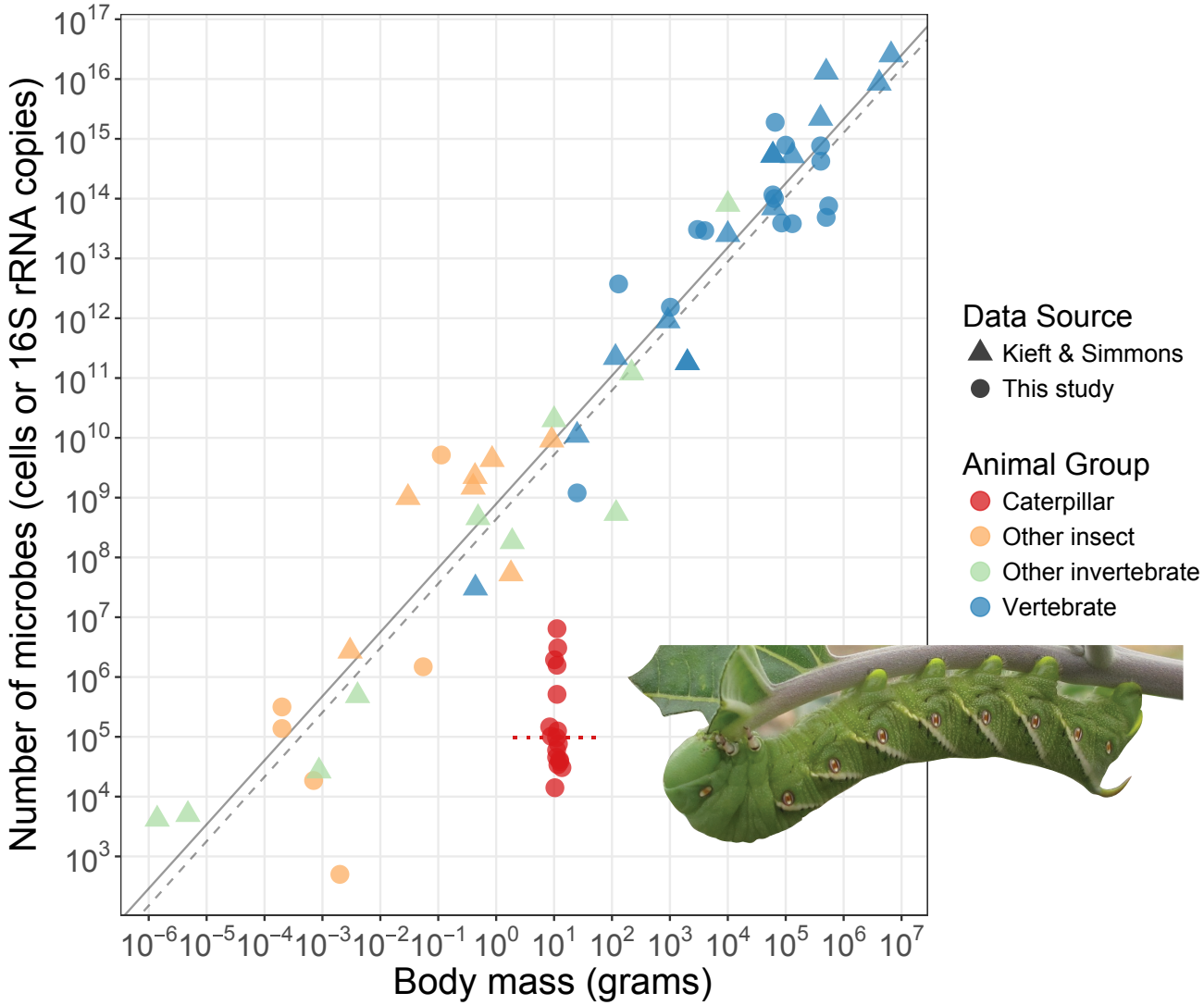

Figure S3. Allometric scaling of whole-individual microbial loads with body size. Triangles and the solid line show data replotted from (49), which were originally measured using microscopy or culturing. Circles show data generated in this study, using quantitative PCR. The dashed regression line is calculated from a model only including non-caterpillar species analyzed in this study, limited to those species with bacterial densities not less than 1/100 of the group median. The red horizontal dotted line indicates the median per-caterpillar bacterial load for *Manduca sexta* individuals collected in Colorado (N=15) or Arizona (N=2). The photograph is *M. sexta* feeding on *D. wrightii*.

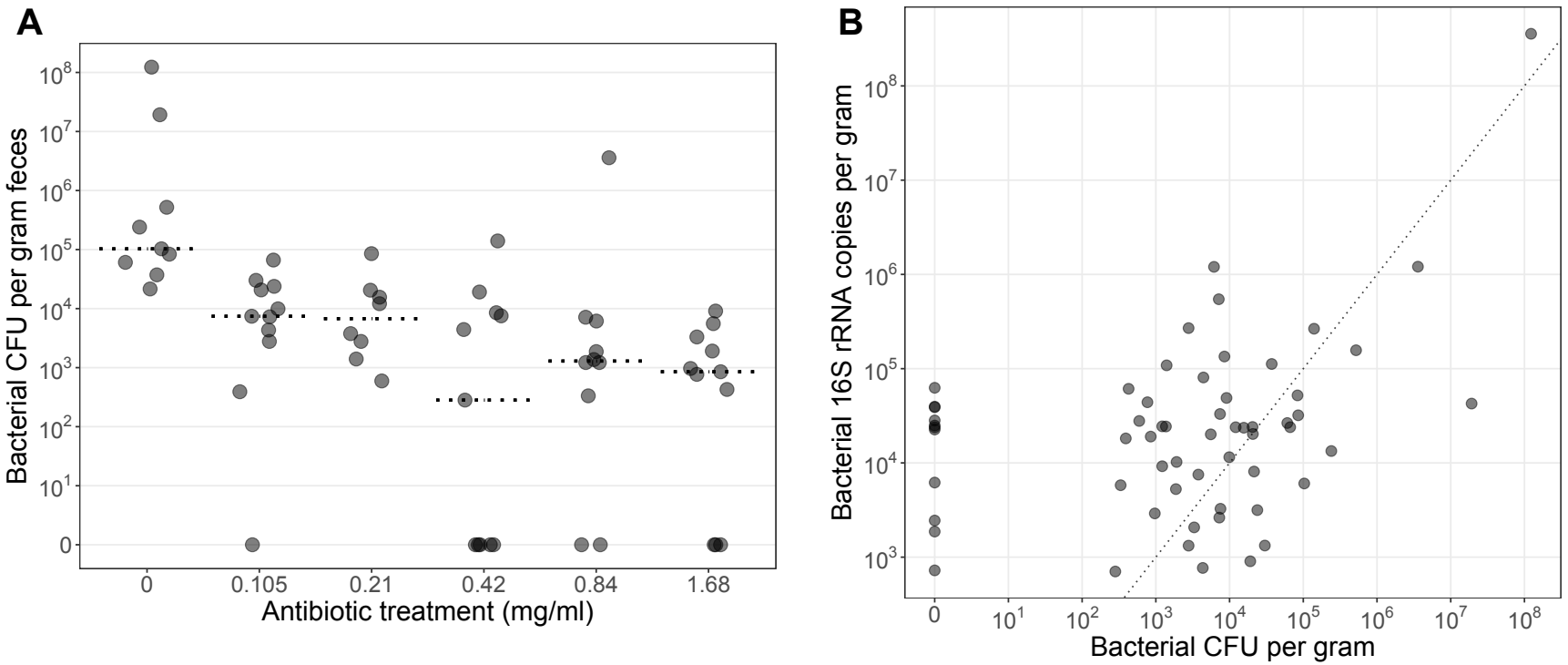

Figure S4. The relationships between antibiotic dose, the number of bacterial colony-forming units cultured on LB media, and the number of bacterial 16S rRNA gene copies measured by qPCR. Fecal samples that yielded no cultured colonies are plotted at 10^0 on $\log_{10}$ axes; for these, nonzero estimates of 16S rRNA gene copies are likely due, in large part, to amplification of DNA from dead or nonviable cells (see *SI Methods*). A) Effect of antibiotic treatment on the number of culturable bacteria in caterpillar feces. Points are individual caterpillars (N=60) and are horizontally jittered for clarity. Dashed lines are medians for each treatment. B) Correlation of bacterial density as measured by culturing versus by DNA quantification. The 1:1 line between the two variables is shown.

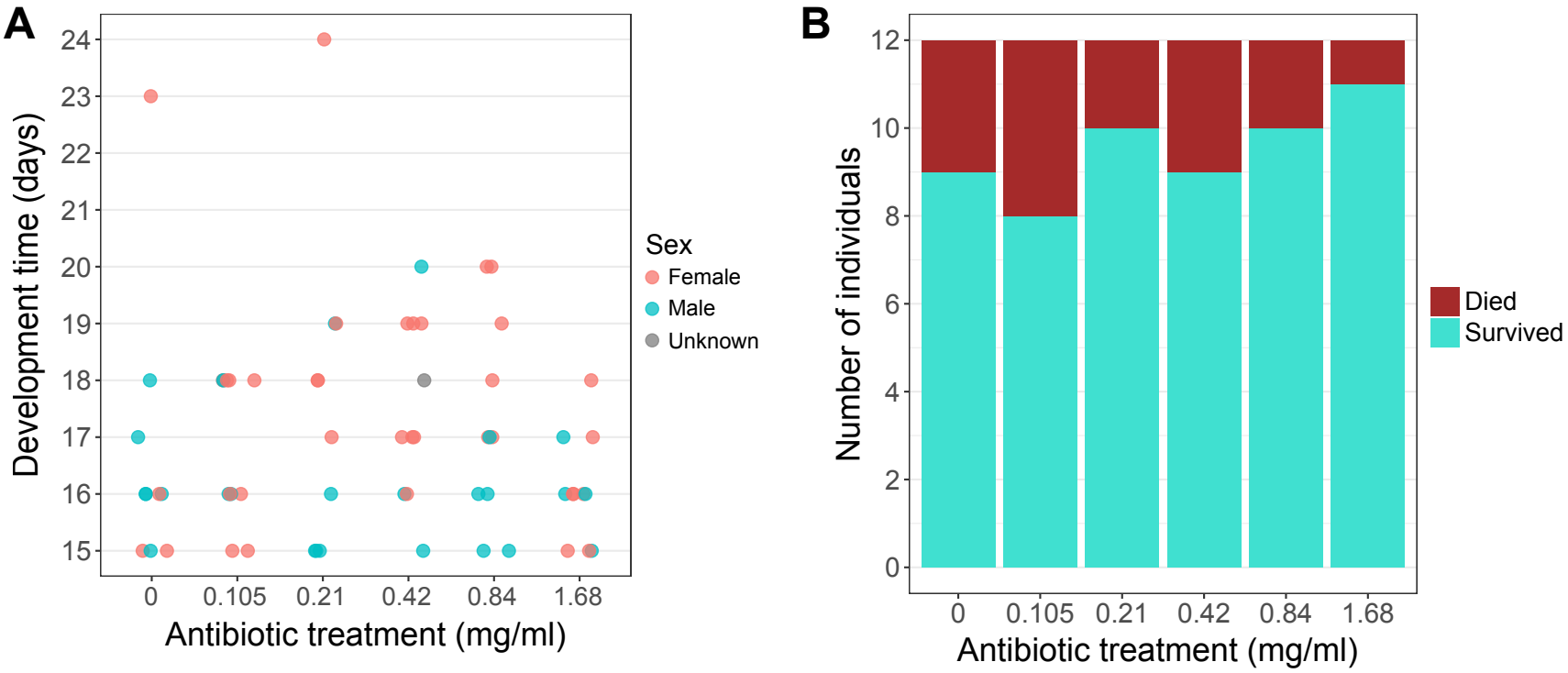

Figure S5. The relationship between antibiotic treatment and other components of *M. sexta* fitness. A) Number of days from larval hatching from eggs to the cessation of feeding, which marks the beginning of the prepupal stage. Shown are the 64 individuals that survived to this point. We were unable to identify the sex of two individuals that died as a prepupa or pupa. B) The proportion of individuals surviving from larval hatch to adult eclosion, for the control group and each antibiotic treatment.

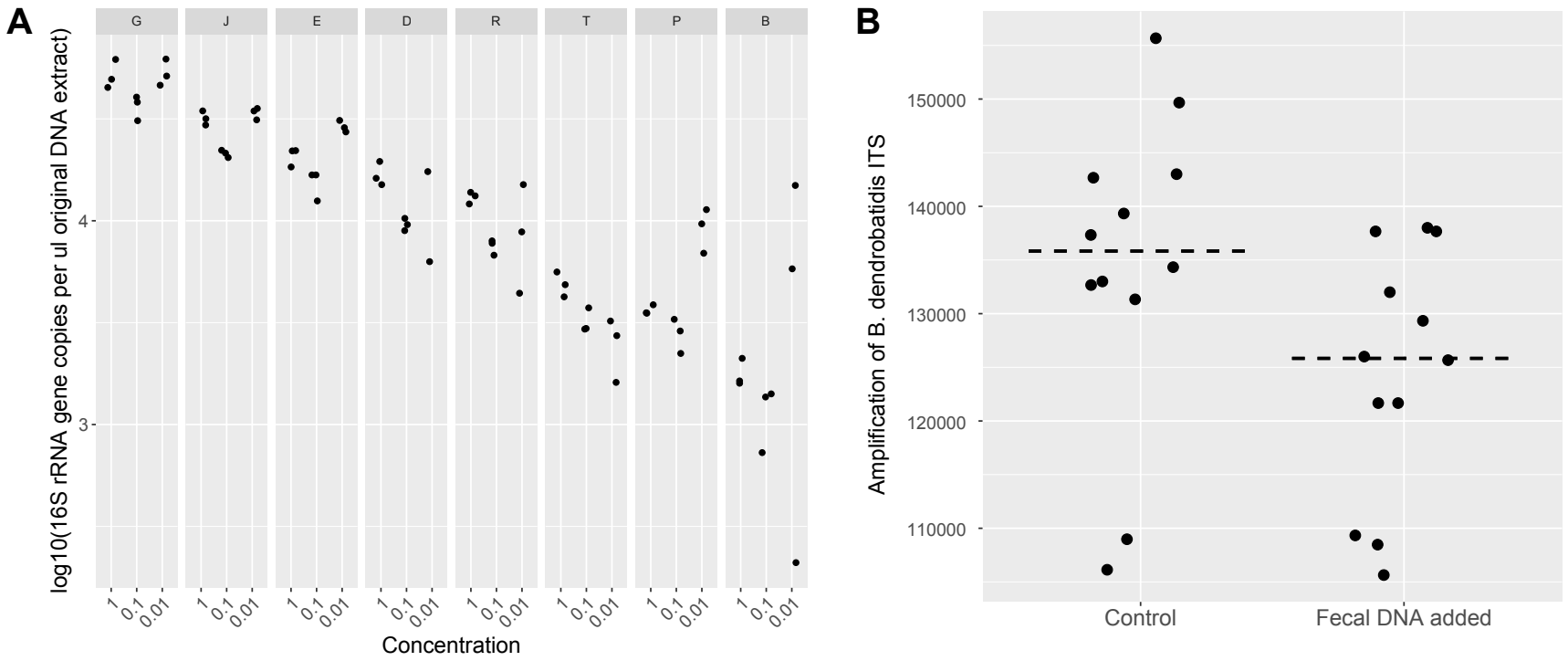

Figure S6. Two tests for PCR inhibitory substances in caterpillar feces. A) Fecal DNA from eight *M. sexta* individuals, arranged left-right by decreasing 16S rRNA gene copy number in original extracts. For each individual, $\log_{10}$(16S rRNA gene copies) is shown for the original sample, and for and extracts diluted 1:10 and 1:100 in pure water. Copy number estimates are standardized per ul of original DNA extract. Note that variability between technical replicates increases with low concentrations of template DNA. One sample, D-0.01, had less amplification than negative controls and is not shown. B) Amplification (arbitrary units) of rDNA ITS of *B. dendrobatidis*, a chytrid fungus of amphibians, showing 12 replicate controls (PCR-grade water only) versus 12 reactions to which 5 µl of caterpillar fecal DNA was substituted for water. Means of triplicate reactions are shown. The twelve caterpillar species with the lowest total 16S rRNA gene copy number were used for this test. Dashed lines show medians for each group.

SI Tables

Table S1. Microbial statistics and basic metadata for species included in the study.

Species	Type	Family	Group	Diet	Sample type	Replicate individuals per species	Bacterial rDNA per g	Fungal rDNA per g	Prop. Plant DNA	Prop. Core
Tarchon felderi	Caterpillar	Bombycidae	Insect	Herbivore	Fecal	1	4006.17	3313.70	0.83	NA
Drepana arcuata	Caterpillar	Drepanidae	Insect	Herbivore	Fecal	1	10146.23	1084.33	0.98	NA
Pachydota saduca	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	8497.66	283.10	0.14	NA
Ophisma tropicalis	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	9752.53	121.10	0.91	NA
Hypercompe icasia	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	6362.74	457.26	0.96	NA
Pyrrharctia isabella	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	33430.51	56.81	0.64	NA
Dasychira basiflava	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	117597.62	3510.38	0.99	NA
Calyptra canadensis	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	88325.96	2189.90	0.98	NA
Melese Espinoza01	Caterpillar	Erebidae	Insect	Herbivore	Fecal	2	36701.88	3884.78	0.53	NA
Dysschema viuda	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	72459066.45	340371.41	0.41	NA
Euchaetes egle	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	22054151.04	32617.59	0.63	NA
Diphthera festiva	Caterpillar	Erebidae	Insect	Herbivore	Fecal	1	59178.87	906.04	0.96	NA
Eutelia furcata	Caterpillar	Euteliidae	Insect	Herbivore	Fecal	1	810.62	18.20	0.33	NA
Nemoria bistriaria	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	14216.53	59.99	0.98	NA
Plagodis serinaria	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	8084.20	634.05	0.99	NA
Eutrapela clemataria	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	157782.77	372.07	0.26	NA
Tetracis cachexiata	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	4948.69	78.75	0.23	NA

Tetracis crocallata	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	2200.31	1.89	0.93	NA
Lytrosis unitaria	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	201447.66	927.73	0.07	NA
Selenia alciphearia	Caterpillar	Geometridae	Insect	Herbivore	Fecal	1	3542.71	11.76	0.62	NA
Udranomia kikkawai	Caterpillar	Hesperiidae	Insect	Herbivore	Fecal	1	6025.22	747.27	0.12	NA
Bungalotis quadratum	Caterpillar	Hesperiidae	Insect	Herbivore	Fecal	1	5459.42	167.10	0.39	NA
Epargyreus Burns06	Caterpillar	Hesperiidae	Insect	Herbivore	Fecal	1	94146.45	2848.23	0.26	NA
Polygonus leo	Caterpillar	Hesperiidae	Insect	Herbivore	Fecal	1	249671.14	19216.34	0.95	NA
Narcosius samson	Caterpillar	Hesperiidae	Insect	Herbivore	Fecal	1	377412.42	7361.54	0.31	NA
Phyllodesma americana	Caterpillar	Lasiocampidae	Insect	Herbivore	Fecal	1	615.86	5.98	0.96	NA
Malacosoma disstria	Caterpillar	Lasiocampidae	Insect	Herbivore	Fecal	1	11821.39	97.70	0.97	NA
Tolype velleda	Caterpillar	Lasiocampidae	Insect	Herbivore	Fecal	1	923.88	0.00	0.52	NA
Malacosoma americanum	Caterpillar	Lasiocampidae	Insect	Herbivore	Fecal	1	281308.04	40186.86	1.00	NA
Acharia stimulea	Caterpillar	Limacodidae	Insect	Herbivore	Fecal	1	15174.26	10296.82	0.99	NA
Tortricidia pallida	Caterpillar	Limacodidae	Insect	Herbivore	Fecal	1	1474.93	3049.52	0.97	NA
Euclea delphinii	Caterpillar	Limacodidae	Insect	Herbivore	Fecal	1	80879.30	77883.77	0.95	NA
Isa textula	Caterpillar	Limacodidae	Insect	Herbivore	Fecal	1	440.91	0.00	0.99	NA
Satyrium liparops	Caterpillar	Lycaenidae	Insect	Herbivore	Fecal	1	13999.94	549.02	1.00	NA
Pseudbarydia crespula	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	3923249.45	2906.99	0.04	NA
Spragueia lepus	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	2982.20	110.80	0.19	NA
Panthea acronyctoides	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	5590.63	2907.54	0.23	NA
Euscirrhopterus poeyi	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	17626.52	3130.87	0.79	NA
Lithophane baileyi	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	5862.53	22.55	0.97	NA
Harrisimemna trisignata	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	13192.59	5157.60	0.98	NA

Orthosia alurina	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	52211.88	0.00	0.99	NA
Copivaleria grotei	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	2409212.79	4333.12	0.97	NA
Catocala cara	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	71.13	0.58	0.50	NA
Spodoptera latifascia	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	2	9145.14	93.52	0.49	NA
Lithophane hemina	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	2872202.73	3346.44	0.69	NA
Orthosia rubescens	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	205996.07	9716.80	0.99	NA
Eudryas grata	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	1330.47	0.00	0.01	NA
Gonodonta unknownspecies	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	0.00	0.00	0.30	NA
Catocala epione	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	85409.54	0.00	0.99	NA
Acronicta americana	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	3360.63	88.22	0.94	NA
Panthea furcilla	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	800.16	0.00	0.13	NA
Lithophane grotei	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	229782.01	1337.89	0.90	NA
Feralia jocosa	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	140932.68	22278.88	0.94	NA
Cucullia convexipennis	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	6215.24	248.61	0.99	NA
Lithophane antennata	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	22275.94	150.51	0.97	NA
Xylena curvimacula	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	931.63	0.00	0.97	NA
Acronicta lepusculina	Caterpillar	Noctuidae	Insect	Herbivore	Fecal	1	8415.05	291.99	0.79	NA
Nystalea ebalea	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	886.24	96.05	0.52	NA
Rifargia felderi	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	641388.03	161600.81	0.96	NA
Sericochroa Janzen01	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	676325.13	49855.09	0.18	NA
Furcula borealis	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	18300.71	0.00	0.99	NA
Dasylophia thyatiroides	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	58873.76	3158.65	0.96	NA
Cecrita probaDHJ04	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	36027.44	1351.87	0.40	NA

Schizura concinna	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	544.01	0.00	0.91	NA
Notodonta torva	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	3385.37	0.00	0.71	NA
Pheosia rimosa	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	3901.59	0.00	0.67	NA
Nystalea squamosa	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	1227.42	30.47	0.28	NA
Crinodes besckei	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	137343.03	55371.50	0.99	NA
Symmerista sp	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	4212.32	28.30	0.93	NA
Hemiceras nigrescens	Caterpillar	Notodontidae	Insect	Herbivore	Fecal	1	1430.33	51.84	0.95	NA
Nymphalis antiopa	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	601.08	0.00	0.27	NA
Polygonia interrogationis	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	2	12245.63	1284.86	0.98	NA
Archaeoprepona demophoon	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	90460.15	6000.60	0.88	NA
Antirrhea lindigii	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	16759.35	121.96	0.19	NA
Danaus plexippus	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	529404.79	71541.19	0.95	NA
Catonephele numilia	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	6408679.22	5158461.47	0.03	NA
Caligo atreus	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	2	139516.96	691.36	0.29	NA
Memphis proserpina	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	112625.74	15737.27	0.68	NA
Archaeoprepona camilla	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	37516.03	1754.57	0.36	NA
Morpho amathonte	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	14123.86	268.49	0.22	NA
Morpho catalina	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	60979.21	1520.17	0.47	NA
Taygetis rufomarginata	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	16338612.30	5941.10	0.11	NA
Morpho granadensis	Caterpillar	Nymphalidae	Insect	Herbivore	Fecal	1	2415575.47	5258.48	0.04	NA
Papilio glaucus	Caterpillar	Papilionidae	Insect	Herbivore	Fecal	1	929.15	272.79	0.95	NA
Pterourus bryki	Caterpillar	Papilionidae	Insect	Herbivore	Fecal	2	1209928.18	11553.19	0.08	NA
Papilio troilus	Caterpillar	Papilionidae	Insect	Herbivore	Fecal	1	473.20	28.35	0.89	NA

Citheronia lobesis	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	5	2057.27	0.74	0.08	0.58
Eacles imperialis	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	16	5346.08	89.17	0.98	0.60
Rothschildia lebeau	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	5	1652.27	0.00	0.07	0.52
Schausiella santarosensis	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	5	36835.67	362.21	0.97	0.70
Syssphinx colla	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	5	25000.37	6217.02	0.98	0.40
Titaea tamerlan	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	5	6663.49	90.05	0.99	0.35
Antheraea polyphemus	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	7630.89	0.00	0.97	NA
Automeris postalbida	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	114148.96	3190.42	0.04	NA
Citheronia regalis	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	4283.77	3157.75	0.98	NA
Hemileuca lucina	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	57692.21	945.77	1.00	NA
Anisota virginiensis	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	3039.32	23.74	0.98	NA
Adeloneivaia jason	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	416249.53	14130.65	0.59	NA
Hylesia continua	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	320113.93	5680.18	0.18	NA
Callosamia promethea	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	1327.54	131.81	0.99	NA
Actias luna	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	37319.84	0.00	1.00	NA
Copaxa rufinans	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	87530.01	24309.40	0.47	NA
Sphingicampa bisecta	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	3538.19	491.41	0.31	NA
Callosamia angulifera	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	1769393.41	18566.11	0.76	NA
Oxytenis beprea	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	43077.75	1271.71	0.29	NA
Hyalophora cecropia	Caterpillar	Saturniidae	Insect	Herbivore	Fecal	1	5030260.25	9463.59	0.51	NA
Manduca sexta	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	18	9847.13	127.92	0.95	0.30
Aellopos titan	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	4	312675.68	6439.52	1.00	NA
Unzela japix	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	351.81	16.90	0.90	NA
Lapara bombycoides	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	676.87	71.68	0.79	NA

Nyceryx tacita	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	25549.44	1626.02	0.57	NA
Ceratomia undulosa	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	5090.77	445.87	0.81	NA
Aellopos fadus	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	2	5935.51	3528.85	0.98	NA
Eumorpha achemon	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	2078.38	0.00	0.09	NA
Sphinx poecila	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	0.00	0.00	0.00	NA
Ceratomia amyntor	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	79295.79	1174.75	0.94	NA
Hemaris diffinis	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	986309.61	659654.01	0.80	NA
Erinnyis crameri	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	806923.46	9066.56	0.99	NA
Perigonia ilus	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	6710.32	166.90	0.50	NA
Xylophanes loelia	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	54084.35	610.44	0.05	NA
Amphonyx duponchel	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	29325.85	7618.17	0.97	NA
Unzela pronoe	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	380.74	54.29	0.96	NA
Sphinx kalmiae	Caterpillar	Sphingidae	Insect	Herbivore	Fecal	1	15495.27	1981.95	0.99	NA
Branta bernicla nigricans	Other animal	Other	Bird	Herbivore	Fecal	3	518382.97	4983.42	0.48	0.95
Aphodius pedellus	Other animal	Other	Insect	Dung	Whole body (homogenate)	3	45689675894.29	20911404.35	0.00	0.94
Apis mellifera	Other animal	Other	Insect	Herbivore	Whole body (homogenate)	6	28807184.26	0.00	0.01	1.00
Tamalia coweni	Other animal	Other	Insect	Herbivore	Whole body (homogenate)	3	1578929729.82	90537.48	0.00	1.00
Tamalia inquilinus	Other animal	Other	Insect	Herbivore	Whole body (homogenate)	3	687978120.71	624186.28	0.00	1.00
Geotrupes stercorosus	Other animal	Other	Insect	Dung	Whole body (homogenate)	2	2851028.70	14836.51	0.00	NA
Pseudomyrmex spinicola	Other animal	Other	Insect	Herbivore	Whole body (homogenate)	1	12524618.61	99309.25	0.01	NA
Macrosiphum euphorbiae	Other animal	Other	Insect	Herbivore	Whole body (homogenate)	1	633845944.81	0.00	0.01	NA
Bison bison	Other animal	Other	Mammal	Herbivore	Fecal	3	137539946.03	0.00	0.00	0.95
Bos taurus	Other animal	Other	Mammal	Herbivore	Fecal	3	1907811339.05	17566372.08	0.00	0.99

Capra hircus	Other animal	Other	Mammal	Herbivore	Fecal	3	1922339809.30	5329055.19	0.00	0.97
Cercopithecus ascanius	Other animal	Other	Mammal	Omnivore	Fecal	3	10102019566.73	408997.09	0.00	0.94
Cercopithecus mitis	Other animal	Other	Mammal	Omnivore	Fecal	3	6620159251.40	0.00	0.00	0.95
Equus caballus	Other animal	Other	Mammal	Herbivore	Fecal	3	97169889.91	31087798.43	0.00	0.89
Mus musculus	Other animal	Other	Mammal	Herbivore	Fecal	3	47987023.24	44935.59	0.00	0.89
Myotis lucifugus	Other animal	Other	Mammal	Insectivore	Fecal	3	668425.17	32665.36	0.15	0.89
Neotoma lepida	Other animal	Other	Mammal	Herbivore	Fecal	3	28816733387.87	2505713.07	0.00	0.91
Odocoileus virginianus	Other animal	Other	Mammal	Herbivore	Fecal	3	460167255.40	315910782.40	0.00	0.90
Ovis dalli	Other animal	Other	Mammal	Herbivore	Fecal	3	1567164071.09	10278456.07	0.00	0.97
Rangifer tarandus	Other animal	Other	Mammal	Herbivore	Fecal	3	7838726610.14	68135.34	0.00	0.99
Sus scrofa	Other animal	Other	Mammal	Omnivore	Fecal	7	293518641.49	84175.96	0.01	0.95
Sylvilagus nuttallii	Other animal	Other	Mammal	Herbivore	Fecal	3	1497616430.42	130260411.03	0.00	0.98
Alces alces	Other animal	Other	Mammal	Herbivore	Fecal	2	1048643100.98	56134.09	0.00	NA
Homo sapiens	Other animal	Other	Mammal	Omnivore	Fecal	1	28723208492.57	735549.51	0.00	NA

Table S2**.** The taxonomy and distribution of the 10 most-common bacterial phylotypes in caterpillar feces.

Phylotype	Phylum	Family	Genus	Prop. samples detected in	Median rRNA copies per gram when present	Among top 10 plant phylotypes?	Top BLAST hits to named isolates	Highest seq. identity
OTU_19	Firmicutes	Staphylococcaceae	*Staphylococcus*	0.79	280	No	*Staphylococcus*	100%
OTU_23	Proteobacteria	Enterobacteriaceae	*Escherichia*	0.73	132	Yes	*Escherichia coli*	100%
OTU_11	Proteobacteria	Methylobacteriaceae	*Methylobacterium*	0.72	692	Yes	*Methylobacterium*	100%
OTU_2	Proteobacteria	Enterobacteriaceae	*Klebsiella*	0.65	202	Yes	*Klebsiella, Enterobacter, others*	100%
OTU_5	Firmicutes	Enterococcaceae	*Enterococcus*	0.65	109	Yes	*Enterococcus*	100%
OTU_79	Proteobacteria	Sphingomonadaceae	*Sphingomonas*	0.54	155	Yes	*Sphingomonas*	100%
OTU_8	Firmicutes	Lachnospiraceae	*Clostridium*	0.49	40	Yes	*Clostridium*	93%
OTU_16	Proteobacteria	Moraxellaceae	*Acinetobacter*	0.44	86	No	*Acinetobacter*	100%
OTU_160	Actinobacteria	Corynebacteriaceae	*Corynebacterium*	0.42	108	No	*Corynebacterium*	100%
OTU_36	Actinobacteria	Microbacteriaceae	unclassified	0.42	141	Yes	*Curtobacterium, Flavobacterium*	100%

The taxonomic classification from Greengenes is given in columns 2-4. For each phylotype, the median absolute abundance was calculated only including samples in which it was present. We also indicate whether each phylotype was one of the 10 most common among all leaf samples. The representative sequence was used in a BLAST search against the NCBI database, restricted to named bacterial isolates, and the sequence identity and genus-level classification of the top hit(s) are shown.

SI Methods

Sampling

Fecal samples were obtained from wild populations of caterpillars in four regions: Área de Conservación Guanacaste (Costa Rica), New Hampshire and Massachusetts (USA), Boulder County, Colorado (USA), and Portal, Arizona (USA). Caterpillars were collected in ACG under permit #ACG-PI-027-2015 and in Arizona under a Scientific Use Permit from the United States Forest Service. For more details about the ACG landscape and collection, rearing, and identification protocols, see (3-5). Most species were collected as caterpillars, but some ACG specimens were reared from eggs either found on foliage or laid by females caught at light traps (see file "Additional_ACG_SampleData"; all files are downloadable from the figshare repository linked in the main text). For some caterpillars we had information on whether they died of parasitoids or disease after sampling, and these samples were discarded in order to focus on apparently healthy individuals. Most caterpillars were sampled in the final or penultimate instar.

All samples were preserved within 30 minutes of defecation, as preliminary evidence suggested rapid (by 6-12 hours) bacterial and fungal growth in excreted fecal pellets, which would render old feces unsuitable as a proxy for gut microbial communities. In five caterpillar species, we did not find evidence for abundant bacterial populations in the midgut (including both ecto- and endoperitrophic spaces) or hindgut that were not captured in feces (Fig. S2A), supporting a previous finding that caterpillar feces approximates the whole-body microbial community (6). Further supporting the use of fresh feces to sample microbes in the caterpillar gut, we found that the inter-individual variation in sequence composition (including nonbacterial DNA) was reflected in fecal samples (Mantel tests: midgut $r = 0.33$, $p = 0.001$; hindgut $r = 0.39$, $p = 0.001$).

We preserved gut and fecal samples using either dry storage at -20°C or 95% ethanol (see file "SampleData"); both methods are suitable for storing insect microbiome samples and do not substantially alter community composition (7). Approximately 50 mg (fresh weight) of sample was used for DNA extraction. Prior to DNA extraction, ethanol-preserved samples were dried in a vacuum centrifuge; since this also evaporated water, their fresh weight equivalent was estimated using percent water content calculated from *M. sexta* guts or feces. To test whether microbial biomass estimates may have been biased by ethanol storage, we compared PCR amplification for paired ethanol-stored and frozen fecal pellets from eight *M. sexta* individuals. From a collection of pellets defecated by each individual during a 1-2 hour window, separate pellets were randomly chosen for each storage type (note that pre-storage inter-pellet microbial variation is possible even under these relatively controlled conditions). As assessed by a linear mixed-effects model treating individual as a random effect, there was no significant influence of storage method on 16S rRNA gene copy number ($\chi^2(1) = 1.09$, $p = 0.30$).

For caterpillars in Costa Rica and Colorado, we also sampled microbes from leaves of the same branch as that fed to the caterpillar prior to feces collection. With this strategy we aimed to maximize microbial similarity between the leaves that were sampled and those consumed by the caterpillar, although leaf microbiomes can also vary substantially within a branch (8). These leaves appeared clean and had not, to our knowledge, come into contact with any caterpillars prior to sampling. Leaves from Colorado

plants were frozen dry at -20°C and ground under liquid N_2 with a mortar and pestle prior to DNA extraction (thus including endophytes as well as surface-associated microbes). Leaves from Costa Rican plants were stored in 95% ethanol, and surface-associated microbes were concentrated in a vacuum centrifuge and resuspended in molecular grade water prior to DNA extraction. As this sampling method was not quantitative, we did not perform qPCR on plant samples from Costa Rica and used them only for analyses of microbial composition.

Non-lepidopteran animals were sampled using the same procedures outlined above, with five species preserved in ethanol, two in 15% glycerol at -80°C, and 17 preserved dry at -20°C or -80°C (see file "Other_animal_Metadata"). With the exception of two dung beetles feeding on herbivore dung, and the insectivorous bat *M. lucifugus*, these species are either predominantly or exclusively herbivorous, although the type of plant matter consumed (sap, leaves, seeds, fruit, pollen, etc.) varies. We extracted DNA from feces for vertebrates and from subsamples of homogenized whole bodies for insects (as some insects house the majority of symbionts in organs outside the gut). By including all tissue from these insects, we may have underestimated bacterial densities in the particular organs where microbes are housed (Fig. 1A).

DNA extraction, PCR and sequencing

Following previous studies of insect microbiomes (6,7,9), we used the MoBio Powersoil kit to extract DNA (100 µl eluate) from measured amounts of sample material. We then PCR-amplified a portion of the 16S rRNA gene with barcoded 515f/806r primers (10). PCR products were cleaned and normalized (up to 25 ng DNA/sample) using the SequalPrep Normalization kit (Thermo Fisher Scientific), and then sequenced on an Illumina MiSeq. Paired-end sequences of 16S rRNA amplicons were merged, quality-filtered, and clustered into operational taxonomic units ("phylotypes") at the 97% sequence similarity level using UPARSE (11), and classified using the RDP classifier and Greengenes (12,13) as previously described (14). The representative sequences of phylotypes unclassified at this stage, and mitochondrial rRNA phylotypes (which could be from plant, insect, fungal or other mitochondria) were aligned to the NCBI nonredundant nucleotide database (nt) using BLAST for taxonomic identification.

As bacterial DNA is ubiquitous in laboratory reagents used for DNA extraction and PCR, and especially problematic with low-biomass samples (such as caterpillar feces) (15), we removed contaminants from our samples using information from the 22 DNA extraction blanks and PCR no-template controls that yielded >100 bacterial sequences. Importantly, phylotypes detected in these blanks are not exclusively composed of reagent contaminants, because they receive some input from sample DNA during laboratory processing (16). As high-biomass samples are both least likely to experience reagent contamination (15), and themselves most likely to be the source of "real" sample phylotypes identified in blanks, they can be used to distinguish between laboratory contaminants and true sample sequences (16). We classified contaminants as phylotypes present at ≥1% abundance in one or more blank samples, excepting phylotypes present at ≥1% abundance in one or more of the best-amplifying samples (the top third in 16S rRNA gene copy number as measured by qPCR). These 25 phylotypes were removed from the dataset prior to analyses of bacterial abundance and composition (they are retained only in Fig. S2A). This approach does not include other types of contaminants introduced prior to DNA

extraction, such as those from human skin. Finally, we note that the high relatedness between microbes commonly present in laboratory reagents (listed in (15)) and those present in soil, water and leaves—all possible genuine microbial inputs to the caterpillar gut—precludes a taxonomy-based approach to removing contaminants.

Sequence Data Analysis

All analyses were conducted in R version 3.3.2 (17) and are available in the file "Hammer2017_Rcode_June2017.R". Analyses involving bacterial composition were limited to samples with at least 100 bacterial sequences. To calculate phylotype-level overlap between fecal and plant samples, "phylotypes detected on leaves" are defined as those present at any abundance in any plant sample in our dataset. New England and Arizona fecal samples which lack paired plant samples were excluded from this comparison. In measuring core microbiome size in caterpillars and other animals, we excluded species with fewer than three replicate individuals. Further, to be conservative, only caterpillars sampled from the same location, and feeding on the same species of plant were compared. As the number of replicates could affect this metric, and varied among species, we iterated these analyses over multiple combinations of only three replicates per species.

Quantitative PCR

We measured 16S rRNA gene copy number using quantitative PCR with the same primers and DNA extracts as above. Reaction conditions and other details are specified in (18). Each sample was run in triplicate (except 11 non-caterpillar species for which limited DNA was available, which were run singly) and the mean of these technical replicates was used for subsequent analyses. Standard curves were calculated using purified genomic DNA from *E. coli* DH10B, which has seven 16S rRNA operons per genome (19). The median copy number of 31 qPCR'd DNA extraction blanks was subtracted from sample copy numbers. Resulting counts of total 16S rRNA genes in samples were then multiplied by the proportion of bacterial sequences identified from the same DNA extract—excluding contaminants and DNA from plants and other eukaryotes—resulting in estimates of bacterial 16S copy numbers.

It is unlikely that the low amplification we found in caterpillar samples results from primer bias against abundant bacterial taxa. First, these primers successfully amplified bacteria in non-lepidopteran animals, even when in some cases (such as aphids (20)), the dominant symbiont has been strictly vertically transmitted between hosts for tens of millions of years. Even in this case, divergence from free-living relatives has not been so great that its 16S rRNA gene is un-amplifiable using 515f/806r primers. Second, the caterpillar gut-associated microbial taxa we found are similar to those reported as being relatively (i.e., in terms of the proportion of sequence libraries) abundant in metagenomic surveys (21,22) and amplicon-based studies using different 16S rRNA-targeting primer pairs (e.g., (23-27)).

To estimate the relationship between body size and whole-animal microbial loads (Fig. S3), we combined published data from (28) with body mass data we calculated directly or derived from other studies (see file "Body_mass_data"). To restrict the allometric scaling relationship for noncaterpillar animals to those species likely to harbor resident microbiomes, we removed species that had bacterial densities < $1/100^{th}$ of the group median. These species were the goose *Branta bernicla*, the bat *Myotis lucifugus*, and the dung beetle *Geotrupes stercorosus*. The body size of two *M. sexta* individuals from Arizona was not recorded and so we substituted the median from other *M. sexta*. Furthermore, as we only had direct gut mass measurements for *M. sexta* (30-40% of body mass), for species sampled using feces (including *M. sexta*) we calculated total microbial loads by multiplying 16S rRNA gene density in feces by body mass. This procedure is likely to have slightly overestimated the microbial load for these species. Despite the numerous methodological uncertainties, microbial counts from (28) and our qPCR-based data, and their allometric scaling relationship with body size (excepting *M. sexta*) were remarkably similar (compare solid and dashed line in Fig. S3).

PCR inhibition assays

To examine whether low 16S rRNA gene copy number estimates in caterpillar samples are an artifact of caterpillar-specific PCR inhibitors, we used two distinct approaches. First, we tested whether diluting extracted DNA improves PCR amplification by minimizing inhibitor effects (29). However, 1:10 and 1:100 dilutions of fecal DNA from eight *M. sexta* individuals did not have this effect (Fig. S6A). Second, we individually added the twelve lowest-amplifying caterpillar fecal samples—which might be especially likely to contain PCR inhibitors—to qPCR reactions with targeted primers and a template highly unlikely to be present in caterpillar feces (rDNA ITS region of *Batrachochytrium dendrobatidis* strain JEL270, a chytrid fungus pathogenic to amphibians). As compared to replicate reactions with pure molecular-grade water, adding caterpillar fecal DNA reduced amplification of *B. dendrobatidis* rDNA by 7.4% (Fig. S6B). This inhibition effect, which is also present in feces of humans (29) and likely many other species, is miniscule relative to the difference in bacterial loads between caterpillars and non-lepidopterans spanning multiple orders of magnitude (Fig. 1A). Therefore, the relatively low PCR amplification of 16S rRNA genes from caterpillar feces is most likely due to low microbial biomass rather than high PCR-inhibitory substances.

Additional information on the antibiotic experiment

M. sexta larval feces production was measured by collecting, drying (50°C for 24 hours), and weighing all fecal pellets in the final instar. Pupae were weighed six days after pupation and monitored daily for adult eclosion. We collected a fresh fecal pellet from each caterpillar midway through the final instar, from which one subsample was cultured on LB media, and another used for qPCR and sequencing using the aforementioned protocol. To culture bacteria, we plated a dilution series (in sterilized phosphate-buffered saline) of weighed (10-20 mg) subsamples of feces, incubated in aerobic conditions at 37°C. After 24 hours, visible colonies were counted and then, if present, collected *en masse* from the agar surface for sequencing using a sterile swab. This plate-scrape method

produces a list of the most abundant bacterial phylotypes potentially culturable using our approach. It should be noted that the presence of fecal bacteria in culture demonstrates that these taxa were viable, but not necessarily growing or metabolically active, while in the caterpillar gut.

Comparison of biomass estimates and evidence of extracellular DNA

Among *M. sexta* fecal samples collected during the antibiotic experiment, we found that qPCR-estimated bacterial abundances were correlated with the number of cultured bacterial colonies (see Results; Fig. S4B). Eleven individuals' fecal pellets did not produce any bacterial colonies whatsoever, but did contain measurable levels of DNA (Fig. S4B), and excluding these "zero-colony" samples yielded a stronger association between bacterial colony counts and 16S rRNA gene copy number (Pearson correlation, r = 0.51, p = 0.0002). This result could stem from the presence of bacteria that cannot grow aerobically or on LB. Alternatively, it may be due to PCR amplification of extracellular DNA or DNA from dead or otherwise nonviable cells (18). To evaluate these possibilities, we compared the phylotypes (identified by 16S rRNA gene sequencing) in zero-colony fecal samples to those from other samples that did yield colonies, in which bacterial biomass was swabbed directly from the agar surface and sequenced. Most of the 16S rRNA gene sequences in the zero-colony fecal samples (median 84%, interquartile range: 74-95%) belong to phylotypes cultured from other samples, suggesting that qPCR may have overestimated viable bacterial loads by amplifying DNA from lysed or nonviable cells. If the fraction of the gut microbiome originating from dead or nonviable cells is disproportionately high in caterpillars in general (e.g., due to their digestive physiology – see Discussion), then the difference in living, active microbial biomass between caterpillars and other animals (Fig. 1A) may have been underestimated.

SI References

1. Regier JC, et al. (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS One* 8(3):e58568.
2. van Nieukerken EJ, et al. (2011) Order Lepidoptera Linnaeus, 1758. *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*, ed Zhang Z-Q (Magnolia Press), pp 212–221.
3. Janzen DH, Hallwachs W (2011) Joining inventory by parataxonomists with DNA barcoding of a large complex tropical conserved wildland in northwestern Costa Rica. *PLoS One* 6(8):e18123.
4. Janzen DH, et al. (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour* 9:1–26.
5. Janzen DH, Hallwachs W (2016) DNA barcoding the Lepidoptera inventory of a large complex tropical conserved wildland, Área de Conservación Guanacaste, northwestern Costa Rica. *Genome* 59(9):641–660.

6. Hammer TJ, McMillan WO, Fierer N (2014) Metamorphosis of a butterfly-associated bacterial community. *PLoS One* 9(1):e86995.
7. Hammer TJ, Dickerson JC, Fierer N (2015) Evidence-based recommendations on storing and handling specimens for analyses of insect microbiota. *PeerJ* 3:e1190.
8. Espinosa-Garcia FJ, Langenheim JH (1990) The endophytic fungal community in leaves of a coastal redwood population diversity and spatial patterns. *New Phytol* 116(1):89–97.
9. Hammer TJ, et al. (2016) Treating cattle with antibiotics affects greenhouse gas emissions, and microbiota in dung and dung beetles. *Proc R Soc B* 283(1831):20160150.
10. Caporaso JG, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8):1621–1624.
11. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10(10):996–8.
12. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73(16):5261–5267.
13. McDonald D, et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6(3):610–618.
14. Ramirez KS, et al. (2014) Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc R Soc B Biol Sci* 281:20141988.
15. Salter SJ, et al. (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12(1):87.
16. Lazarevic V, Gaïa N, Girard M, Schrenzel J (2016) Decontamination of 16S rRNA gene amplicon sequence datasets based on bacterial load assessment by qPCR. *BMC Microbiol* 16(1):73.
17. R Core Team (2016) R: A language and environment for statistical computing. *R Found Stat Comput*. Available at: http://www.r-project.org.
18. Carini P, et al. (2016) Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol* 53(9):680840.
19. Durfee T, et al. (2008) The complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse. *J Bacteriol* 190(7):2597–2606.
20. Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol Biol Evol* 16:1586–1598.
21. Belda E, et al. (2011) Microbial diversity in the midguts of field and lab-reared populations of the European corn borer *Ostrinia nubilalis*. *PLoS One* 6(6):e21751.
22. Xia X, et al. (2017) Metagenomic sequencing of diamondback moth gut microbiome unveils key holobiont adaptations for herbivory. *Front Microbiol*. doi:10.3389/fmicb.2017.00663.

23. Broderick NA, Raffa KF, Goodman RM, Handelsman J (2004) Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. *Appl Environ Microbiol* 70(1):293–300.
24. Staudacher H, et al. (2016) Variability of bacterial communities in the moth *Heliothis virescens* indicates transient association with the host. *PLoS One* 11(5):1–21.
25. Mason CJ, Raffa KF (2014) Acquisition and structuring of midgut bacterial communities in gypsy moth (Lepidoptera: Erebidae) larvae. *Environ Entomol* 43(3):595–604.
26. Pinto-Tomás AA, et al. (2011) Comparison of midgut bacterial diversity in tropical caterpillars (Lepidoptera: Saturniidae) fed on different diets. *Environ Entomol* 40(5):1111–1122.
27. Brinkmann N, Martens R, Tebbe CC (2008) Origin and diversity of metabolically active gut bacteria from laboratory-bred larvae of *Manduca sexta* (Sphingidae, Lepidoptera, Insecta). *Appl Environ Microbiol* 74(23):7189–7196.
28. Kieft TL, Simmons KA (2015) Allometry of animal-microbe interactions and global census of animal-associated microbes. *Proc R Soc B* 282:20150702.
29. Nechvatal JM, et al. (2008) Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J Microbiol Methods* 72(2):124–132.