Structure, culture, and predicted function of the gut microbiome of the Mormon cricket *Anabrus simplex* (Orthoptera: Tettigoniidae)

Chad C. Smith[1], Robert B. Srygley[2], Frank Healy[3] Karthikeyan Swaminath[1], Ulrich G. Mueller[1]

[1]Department of Integrative Biology, University of Texas at Austin, 1 University Station C0990, Austin, TX 78724

[2]USDA-Agricultural Research Service, Northern Plains Area Research Laboratory, 1500 Central Avenue, Sidney MT 59270

[3]Department of Biology, Trinity University, 1 Trinity Pl., San Antonio, TX 78212

Corresponding author: Chad C. Smith

Email: chad.smith123@gmail.com

Header: Structure of the Mormon cricket gut microbiome

SUMMARY

1. The gut microbiome of insects plays an important role in their ecology and evolution, participating in nutrient acquisition, immunity, and behavior. Microbial community structure within the gut is heavily influenced by differences among gut regions in morphology and physiology, which determine the niches available for microbes to colonize.
2. We present a high-resolution analysis of the structure of the gut microbiome in the Mormon cricket *Anabrus simplex*, an insect known for its periodic outbreaks in the western United States and nutrition-dependent mating system. The Mormon cricket microbiome was dominated by eleven taxa from the Lactobacillaceae, Enterobacteriaceae, and Streptococcaeae. While most of these were represented in all gut regions, there were marked differences in their relative abundance, with lactic-acid bacteria (Lactobacillaceae) more common in the foregut and midgut and enteric (Enterobacteriaceae) bacteria more common in the hindgut.
3. Differences in community structure were driven by variation in the relative prevalence of three groups: a *Lactobacillus* in the foregut, *Pediococcus* lactic-acid bacteria in the midgut, and *Pantoea agglomerans*, an enteric bacterium, in the hindgut. These taxa have been shown to have beneficial effects on their hosts in insects and other animals by improving nutrition, increasing resistance to pathogens, and modulating social behavior.
4. Using PICRUSt to predict gene content from our 16S rRNA sequences, we found enzymes that participate in carbohydrate metabolism and pathogen defense in other orthopterans. These were predominately represented in the hindgut and midgut, the most important sites for nutrition and pathogen defense.

5. Phylogenetic analysis of 16S rRNA sequences from cultured isolates indicated low levels of divergence from sequences derived from plants and other insects, suggesting that these bacteria are likely to be exchanged between Mormon crickets and the environment.
6. Our study shows strong spatial variation in microbiome community structure, which influences predicted gene content and thus the potential of the microbiome to influence host function.

Keywords: symbiosis, alimentary tract, immunity, nutrition, PICRUSt, carbohydrate metabolism, lactic acid bacteria, Enterobacteriaceae, katydid,bacteria, metagenomic

INTRODUCTION

Insects are the most speciose and abundant taxa in the animal kingdom, playing a key ecological role in many of the world's ecosystems. Symbioses between insects and their microbial associates has undoubtedly contributed to their success, providing the capability to degrade recalcitrant food, to supplement nutrient-deficient diets, to protect them from their natural enemies, and to modulate the expression of social behavior (Engel & Moran 2013; Douglas 2015). Among the niches available to occupy the host, the gut houses the largest and most diverse microbiome in insects (Engel & Moran 2013; Douglas 2015) and other animals (Ley *et al*. 2008; Cho & Blaser 2012). Gut morphology and physiology vary markedly along the alimentary tract, resulting in an environmental gradient that influences, and is influenced by, the microbial communities that populate it (Dillon & Dillon 2004; Engel & Moran 2013).

The insect gut consists of three regions that are analogous to that in mammals, the foregut, the midgut and the hindgut, each of which contributes to a different aspect of gut function (Douglas 2013). The foregut serves as the entry point for food, where it is stored in the crop before passing through the proventriculus, a valve that can also be modified to mechanically filter food (Woodring & Lorenz 2007; Douglas 2013) and even microbes (Lanan *et al.* 2016). Digestion and absorption of nutrients begins at the midgut, which, in some species, contains specialized crypts that house microbes that aid in insect nutrition (Kikuchi, Meng & Fukatsu 2005; Bistolas *et al.* 2014). Host immune factors have been shown to play an important role in regulation of commensal microbes in the midgut (Ryu, Ha & Lee 2010; Buchon, Broderick & Lemaitre 2013), some of which protect the host from pathogens (Forsgren *et al.* 2010). Following the midgut is the hindgut, which is comprised of the ileum, colon, and rectum. Malphigian tubules permeate the anterior hindgut, excreting nitrogenous waste and other solutes

72 from the hemocoel that can provide nutrients for dense populations of microbes (Bignell 1984).
73 In some species, dense bristle-like structures in the ileum (Woodring & Lorenz 2007) and rectal
74 papillae (Hunt & Charnley 1981) provide attachment sites for bacteria, some of which fix
75 nitrogen (Tai *et al.* 2016), degrade recalcitrant plant polymers (Kaufman & Klug 1991; Engel &
76 Moran 2013), and prevent infection (Dillon & Charnley 2002).

77 The Mormon crickets *Anabrus simplex* (Orthoptera: Tettigoniidae) is an economically
78 important shield-backed katydid distributed throughout the Western United States. Mormon
79 crickets can form dense aggregations of millions of individuals spread over 10 kilometers long
80 and several kilometers wide, feeding on forbes, grasses, and agricultural crops as they march in
81 migratory bands across the landscape (MacVean 1987; Simpson *et al.* 2006). Mormon crickets
82 are also emerging as a model for the study of how social interactions and diet influence
83 immunity (Srygley *et al.* 2009; Srygley & Lorch 2011) and the microbiome (Smith *et al.* 2016).
84 Differences in population density are linked to reproductive behavior, as in high density
85 populations, protein-limited females compete for access to males to gain access to a
86 proteinaceous “nuptial gift” males produce for females during copulation (Gwynne 1984). While
87 consumption of male nuptial gifts by females does not influence the composition of the
88 microbiome, sexually inactive females experience a dramatic decline in *Pediococcus* lactic-acid
89 gut bacteria compared to sexually active females (Smith *et al.* 2016). Lactic-acid bacteria are
90 common associates of the alimentary tract and regarded for their beneficial effects on immune
91 function and nutrition in animals, including insects (Forsgren *et al.* 2010; Storelli *et al.* 2011;
92 Erkosar *et al.* 2015).

93 We characterize the structure of the gut microbiome of Mormon crickets and infer their
94 evolutionary relationships using a combination of culture-dependent and culture-independent

95 approaches. Our aims are to determine whether gut microbial communities vary along the
96 alimentary tract in the Mormon cricket and to infer their potential to influence host function
97 based on their known taxonomic associations with other insects and by employing bioinformatic
98 tools that predict metabolic capabilities from 16S rRNA sequences. We also establish methods
99 for isolating Mormon cricket gut microbiota in culture to permit future experimental
100 manipulations of the gut microbiome and build genomic resources to infer their evolution and
101 function.

102 ## MATERIALS AND METHODS

103 *Animal collection and tissue processing*

104 Mormon crickets were obtained from field (n=5) and laboratory-raised (n=8) collections. Wild
105 females were caught in EK Mountain (43°47'58"N, 106°50'31"W, 1752 m) near Kaycee,
106 Wyoming in the summer of 2014, immediately preserved in 100% ethanol, and stored at -80°C
107 until dissection. Laboratory-raised Mormon crickets were derived from eggs collected from
108 individuals caught on Paint Rock Road (44°27'52"N, 107°27'37"W, 2654 m) in the Bighorn
109 Mountains and fed a mixture of wheat bran, wheat germ, sunflower, mixed bird seeds, tropical
110 fish flakes, fresh Romaine lettuce (added daily), and water *ad libitum*.

111 Mormon crickets were dissected using flame-sterilized tools after rinsing in 1% bleach
112 for 3 min followed by two rinses in autoclaved distilled water to remove bacteria on the
113 exoskeleton. DNA from the foregut (crop and proventriculus), midgut (ventriculus), ileum, and
114 rectum (Fig. 1) of laboratory-raised crickets was extracted with MoBio Powersoil© as in Smith et
115 al. (2016). Foregut (crop and proventriculus), midgut (ventriculus), and hindgut tissue (ileum and
116 rectum combined) was extracted from field-collected animals using a bead-beating/phenol-
117 chloroform extraction protocol (see supplementary material). DNA extraction methods can

influence the representation of taxa in 16S rRNA metagenomic studies (Yuan *et al.* 2012), however our aim here is not to make inferences about differences between field and laboratory-raised animals but differences among tissue types. We include the source of the animal (field or laboratory) and as a covariate in our statistical analyses to account for variation due to source/DNA extraction method (see Statistics).

Sequencing and Bioinformatics

The variable V4 region of 16S rRNA gene was amplified with universal primers (Hyb515F: 5'-GTGYCAGCMGCCGCGGTA -3', Hyb806R: 5'-GGACTACHVGGGTWTCTAAT-3') and sequenced on the Illumina Miseq V3 platform. DADA2 1.1.5 (Callahan *et al.* 2016) was used to process the raw sequencing data and taxonomy was assigned with the Greengenes 13.8 database at 97% identity (see supplementary material). Sequence variants that comprised an average of less than 1% of the reads recovered within a given Mormon cricket were removed prior to analysis using phyloseq 1.16.2 (McMurdie & Holmes 2013).

Metagenomic predictions

We used PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) v1.1.0 (Langille *et al.* 2013) to estimate the functional gene content of our samples. PICRUSt generates metagenomic predictions from 16S rRNA data using annotations of sequenced genomes in the IMG database. Nearest sequenced taxon index (NSTI) values were small (mean $\pm$ sd: 0.03 $\pm$ 0.01, range: 0.006-0.040), indicating the taxa in our samples were closely related to the genomes in the IMG database (see supplementary material). Greengenes IDs used by PICRUSt to construct the phylogenetic tree were assigned to sequence variants using Qiime 1.9 (Caporaso *et al.* 2010), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used for functional classification.

Bacterial Abundance

The abundance of bacteria was estimated using qPCR following Powell et. al (2014) from laboratory-raised (n=8) and field-caught (n=8) Mormon crickets (see supplementary information). Universal 16S rRNA gene primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 355R (5'-CTGCTGCCTCCCGTAGGAGT-3') were used to amplify all copies of the 16S rRNA gene in tissue samples from laboratory (n=8) and field caught individuals (n=8) and copy number quantified with standard curves from the cloned target sequence (Powell *et al.* 2014)

Culturing and phylogenetic analysis

Five lab-reared female Mormon crickets were surface sterilized in 1% bleach for three min, rinsed twice in sterile water and dissected using flame-sterilized tools. Gut tissue was homogenized for 10 s with a bead beater using autoclaved 3.2mm stainless steel beads in sterile PBS. Homogenates were plated onto trypsin soy agar, brain heart infusion agar, nutrient agar, or Man–Rogosa–Sharpe agar (BD), cultured in anaerobic or Campy (low O_2) Gaspak pouches (Becton, Dickinson and Company, Franklin Lakes, NJ) at 37°C for 24-48 hours, and individual colonies passaged three times to obtain pure isolates. DNA was then extracted with chelex and the 16S gene amplified for Sanger sequencing using 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3') primers (see supplementary material).

Phenotypic assays

Fresh overnight cultures of all isolates were used for microscopic analysis. Lactobacillaceae isolates were cultured in Man–Rogosa–Sharpe medium and Enterobacteriaceae were cultured in nutrient broth or LB medium. Biochemical tests were done following Bridson (1998). Motility

was determined using SIM medium and microscopic examination of culture wet mounts. Man–Rogosa–Sharpe or nutrient broth containing 1 g/L potassium nitrate was used for nitrate reduction tests. Fermentation tests were done anaerobically in Man–Rogosa–Sharpe and nutrient broth media with the addition of indicated sugars to 1% w/v final concentration.

Statistics

Analyses were performed in R 3.3.1 (R Core Development Team 2013). Sequence tables were rarified at 1300 reads using phyloseq (McMurdie & Holmes 2013), resulting in the exclusion of hindgut samples from two field-caught females that had a low number of reads. Alpha diversity was compared among tissue types and between origin of subject (field vs. lab) with a linear mixed model (Bates *et al*. 2013), entering the individual ID as a random effect to account for within-subject correlations in diversity. Species richness, Chao1, and the Shannon-Weiner diversity index were calculated. Post-hoc comparisons among gut regions were performed using a Tukey test (Hothorn, Bretz & Westfall 2008).

Beta diversity among gut tissue types and between animal source (field vs. lab) was assessed with a distance-based redundancy analysis (db-RDA) in vegan 2.3 (Oksanen *et al.* 2015), specifying a principal components ordination of Bray-Curtis distances. Statistical significance of the terms in the db-RDA model was determined by 999 permutations of the distance matrix in vegan, restricting the permutations to within each individual to retain the nested structure of the data. The same procedure was also used to examine variation among tissue types in the abundance of KEGG pathways, except nonmetric multidimensional scaling was used for ordination.

We assessed the difference in taxon abundance among tissue types in univariate analyses by fitting the data to a negative binomial generalized linear mixed model (Bates *et al.* 2013),

specifying the individual ID as the random effect and the tissue type and animal source (field vs. lab) as fixed effects. A similar procedure was used to assess differences in 16S rRNA gene copy number between tissue types and animal source, except a normal distribution was specified. Likelihood ratio tests were used to determine the statistical significance of each factor (Venables & Ripley 2002). Goodness-of-fit was assessed by with a Chi-square test (Faraway 2006) and homoscedasticity was assessed by examination of residual plots. Nonparametric methods were used in univariate analyses of the metagenomic predictions because no distribution provided a reasonable fit to the data. P-values were adjusted for multiple tests using the false discovery rate (Benjamini & Hochberg 1995).

RESULTS

Spatial structure of the gut microbiome

We recovered 11 dominant sequence variants from field and lab-raised individuals (Fig. 2), with the remaining 749 sequence variants comprising <1% of the sequences from a given Mormon cricket. Field and laboratory-raised individuals shared 7 of the 11 sequence variants, including the most abundant *Pediococcus acidilactici* OTU that varied with mating status in a previous study (*P. acidilactici* 102222; Smith *et al*. 2016). The remaining five shared sequence variants were two Lactobacilliaceae (*Lactobacillus sp* and *P. acidilactici* 2), two Enterobacteriaceae (*Pantoea agglomerans* and a *Klebsiella sp*) and one Streptococcaceae (*Lactococcus garvieae*). Field-caught Mormon crickets had three taxa that were not shared with laboratory-raised individuals, while lab-raised individuals had two taxa that were not shared with field individuals (Fig. 2). Guts from two laboratory individuals were almost completely comprised of the enteric bacterium *Pantoea agglomerans* (99.3% and 80.8% of reads respectively), so we conducted our analysis with and without these individuals.

Species richness and diversity differed among gut regions and were higher in field compared to lab-raised animals (Table 1, Fig. 3). There was no significant interaction between collection source and tissue type (Table 1), indicating that differences in alpha diversity among tissue types were shared between lab and field-caught animals. We found that the midgut was the most diverse part of the gut with two of the three measures of alpha diversity (species richness and the Chao1 diversity estimator), while the hindgut and foregut had similar levels of richness and diversity. The third metric (Shannon-Weiner) also found the foregut to be the least diverse region, but differed in that the midgut and hindgut had similar levels of species diversity (Table 1, Fig 3).

The db-RDA analysis revealed that the structure of the gut microbiome also varied among gut regions and between field and laboratory animals (Table 2, Fig. 4a). The non-significant interaction in this analysis, however, indicates that the differences in community structure among tissue types were consistent between field and laboratory-raised individuals (Table 2). To determine which members of the gut microbiome varied among gut regions, we plotted the taxa scores from db-RDA analyses of field and laboratory Mormon crickets (Figure S1a). Three groups of bacteria appeared to separate along the gut axis: a *Lactobacillus sp.* lactic-acid bacterium associated with the foregut, *Pediococcus* lactic-acid bacteria were associated with the midgut, and *Pantoea agglomerans,* an enteric bacterium, was found in association with the hindgut. Inspection of the plots from laboratory animals, where the ileum and rectum of the hindgut were dissected separately, indicate that *P. agglomerans* is more abundant in the rectum, while the composition of the ileum, which is separated from the rectum by the colon, closely resembled that of the midgut (Figure S1b).

Univariate analyses of these three groups largely confirmed the pattern in the ordination (Table 3, Fig. 2, Fig. S2). The interaction between tissue type and source was not significant in any of the analyses and dropped to estimate the differences in abundance between tissue types. *Lactobacillus sp*. was three times more common in the foregut than in the midgut ($\beta=1.4 \pm 0.50$, $p=0.02$) and seven times more abundant in the foregut than in the hindgut ($\beta=2.0 \pm 0.51$, $p<0.001$). *Pediococcus* were similar in abundance in the midgut and hindgut but 4.7 times more common in these areas than the foregut ($\beta=1.1 \pm 0.36$, $p=0.006$). *P. agglomerans* was 209 times more abundant in the hindgut than in the foregut ($\beta=3.8 \pm 0.87$, $p<0.001$) and twelve times more abundant in the hindgut than in the midgut ($\beta=2.5 \pm 0.82$, $p=0.007$).

Abundance of bacteria from 16S rRNA qPCR

The number of copies of bacterial 16S rRNA genes was significantly different among tissue types, as indicated by the significant interaction between tissue type and the source of the Mormon crickets (Analysis of deviance: Source, $F_{1,14}=25.9$, $p<0.001$; tissue type, $F_{3,161}=7.8$, $p<0.001$; Interaction, $F_{3,161}=2.8$, $p=0.04$, Fig. 7). We decomposed the interaction to determine how the total number of 16S rRNA copies differed among tissue types within field and laboratory-raised animals. The major difference between the two sources was that in wild Mormon crickets, the midgut had the lowest abundance of all gut regions, while in laboratory-raised individuals, both the midgut and the ileum had the lowest abundance of bacterial 16S rRNA genes (Table S1, Fig. 5).

PICRUSt metagenomic predictions

PICRUSt analysis of 16S rRNA sequence variants recovered 5,891 KEGG orthologs associated with 328 metabolic pathways. The representation of the predicted KEGG pathways differed significantly among gut regions in both the full and reduced datasets, while the source of the

254 animals had a significant influence in the full dataset but not the reduced dataset (Table 2, Fig.
255 4b). Neither analysis, however, showed an interaction between tissue type and whether an animal
256 was wild or lab-reared, indicating that metagenomic predictions differed among tissue types in
257 similar ways (Table 2). Univariate analyses found significant differences among tissue types in
258 most KEGG pathways (Table S2), including those that could affect host-microbe interactions via
259 their role in nutrition, immunity, degradation of xenobiotics, and production of secondary
260 metabolites (Fig. 6). In these functional groups, the hindgut exhibited the most abundant
261 representation of each KEGG category, followed by the midgut and then the foregut (Fig. 6).

262 Nutrition

263 We searched our metagenomic predictions for specific bacterial genes known to
264 contribute to host nutrition in orthopterans. We queried our database for enzymes capable of
265 metabolizing the complex plant carbohydrates xylan, pectin, raffinose, and galactomannan,
266 which are metabolized by gut bacteria in the house cricket *Achetus domesticus* (Kaufman & Klug
267 1991), and cellulose, an important component of the plant cell wall. We found KEGG orthologs
268 involved in the metabolism of all these complex plant polymers, except galactomannan. The
269 pectin metabolic pathway was also incomplete. Only pectinesterase, the first of three enzymes
270 involved in pectin metabolism, was found among the eleven dominant taxa in Mormon crickets,
271 although the remaining two enzymes were represented in the minority members (i.e. <1% of 16S
272 rRNA sequences, see *Sequencing and Bionformatics*) of the gut microbiome.

273 The abundance of KEGG orthologs for carbohydrate metabolism in our samples were
274 most pronounced in the hindgut (Table S3, Fig. 7a) and dominated by the enteric bacteria,
275 particularly *Klebsiella sp*. and Enterobacteriaceae 1 (Fig 7b). Lactic-acid bacteria, however, were
276 also represented in predictions for raffinose metabolism and enzymes capable of participating in

277 the degradation of cellubiose to glucose via cellubiose glucohydrolase, but not in degrading
278 cellulose to cellubiose (Fig 7b).

279 Gut bacteria might also play a role in the production of the essential amino acid
280 phenylalanine via the shikimate pathway, which is found in microbes and plants but not in
281 animals (Herrmann & Weaver 1999). Phenylalanine is required for stabilization of the cuticle
282 following molting (Bernays & Woodhead 1984) and is converted to tyrosine, the precursor of
283 melanin, a key component of the insect immune response (González-Santoyo & Córdoba-
284 Aguilar 2012). All enzymes in the shikimate pathway were represented in our metagenomic
285 predictions (Fig. S3), although prephenate hydrogenase, which is required for phenylalanine
286 synthesis, was only represented in *Lactococcus garviae*. *L. garviae* abundance thus might
287 influence the availability of phenylalanine for Mormon crickets, unless they are able to acquire it
288 in sufficient quantities directly from their diet.

289 Immunity

290 In the locust *Schistocera gregaria* (Orthoptera), four phenols have been shown to
291 increase resistance to microbial pathogens (Dillon & Charnley 1988, 1995): hydroquinone, 3,4-
292 dihydroxybenzoic acid, p-hydroxybenzoic acid, and 4,5-dihydroxybenzoic acid. We found
293 enzymes associated with the production of all these compounds except for 4,5-dihydroxybenzoic
294 acid, which was not annotated in the KEGG database. Hydroquinone production was represented
295 by the enzyme arbutin 6-phosphate glucohydrolase, which metabolizes arbutin, a phenolic
296 glycoside present in leaf and fruit tissue of many plants (Xu *et al.* 2015).

297 Two enzymes were found capable of producing 3,4-dihydroxybenzoic acid. The first,
298 vanillate monooxygenase, demethylates vanillic acid, a compound derived from lignin (Bugg *et*
299 *al.* 2011). This is also the pathway proposed for 3,4-dihydroxybenzoic acid production in locusts

300 based on the abundance of vanillic acid in their feces (Dillon & Charnley 1988, 1995). The
301 second, p-hydroxybenzoate 3-monooxygenase, oxidizes p-hydroxybenzoic acid, one of the other
302 antimicrobial phenols in locusts (Dillon & Charnley 1995). The most likely source of p-
303 hydroxybenzoic acid in the diet of Mormon crickets is benzoic acid, which is a precursor to
304 salicyclic acid in plants (Raskin 1992). The enzyme responsible for catalyzing the conversion of
305 benzoic acid to p-hydroxybenzoic acid (benzoate 4-monooxygenase), however, was not found
306 among the 11 dominant taxa in our samples, although it was present in the minority members of
307 the Mormon cricket gut microbiome. Production of p-hydroxybenzoic acid in appreciable
308 concentrations is thus less likely than for hydroquinone or 3,4-dihydroxybenzoic acid.
309 Like carbohydrate metabolism, the hindgut (Fig. 7c) and enteric bacteria (Fig. 7d)
310 dominated the abundance of KEGG orthologs implicated in the production of antimicrobial
311 phenols in our samples, with the exception of hydroquinone, which was represented to varying
312 degrees among the lactic-acid bacteria. Notably, *P. agglomerans*, which has been reported to
313 participate in the production of 3,4-dihydroxybenzoic acid in locusts (Dillon & Charnley 1995),
314 was not among taxa responsible for the occurrence of vanillate monooxygenase in our samples
315 (Fig 7d).
316 Finally, we searched for three other known contributors to pathogen defense: bacterocins,
317 antibiotics, and lactate dehydrogenase, which provides protection from pathogens in the gut by
318 reducing pH (Servin 2004). We found lactate dehydrogenase to be equally represented among
319 gut regions (Fig 7c), and lactic-acid bacteria were the main contributors to our samples (Fig. 7d).
320 We found three bacteriocins in the KEGG database: nisin, mutacin, and *blp*-derived bacterocins.
321 None of these were found in our metagenomics predictions, perhaps not surprising considering
322 their association with *Streptococcus*, which was not among the top 11 taxa in our samples (Fig.

2). The bacteriocins we would expect to find based on taxonomy (e.g. pediocin for *Pediococcus*) were not annotated in the KEGG database.

Turning to the antibiotics, we found enzymes involved in the production of streptomycin, penicillin, and novobiocin, but not all enzymes required for their synthesis were present (data not shown). We did find β-lactamase, which confers resistance to β-lactam antibiotics (e.g. penicillins, cephalosporins, monobactams, and carbapenems; Drawz & Bonomo 2010), represented among all the Enterobacteriaceae and *Pediococcus* taxa in our samples, but not among *Lactobacillis sp.* other Lactobacillaceae (data not shown). This suggests that lactate production and antibiotic resistance could play a role in microbe-microbe interactions in the Mormon cricket gut microbiome.

Phylogenetic analysis of cultured isolates

Thirteen strains were cultured from the Mormon cricket gut based on 99% sequence similarity of their near full-length 16S rRNA genes (mean $\pm$ sd: 1406 $\pm$ 30bp). Six were lactic-acid bacteria (Lactobacillaceae) and seven were enteric bacteria (Enterobacteriaceae).

The lactic-acid bacteria fell into two clades in our phylogenetic analysis (Fig. 8). The first clade was comprised of *Pediococcus acidilactici* isolates derived from environmental sources, such as plants and various human foodstuffs, as well as strains from the human gut. Similarity to sequences from the BLAST search was high (>99.5%) and branch lengths were short, indicating that *Pediococcus* from the Mormon cricket gut are not highly derived from their relatives, as has been found for *Lactobacillus* species isolated from bees (Fig. 8; McFrederick *et al.* 2013).

Our search for *Pediococcus* sequences from insect guts in Genbank recovered sequences from the termites *Macrotermes bellicosus* and *M. subhyalinus*, which formed their own well-supported clade (Fig 8). Cultured *Pediococcus acidilactici* shared 100% sequence identity in the

346 V4 region with the *P. acidilactici 1* phylotype sequenced using the Illumina platform in this
347 study and with the *P. acidilactici* (102222) phylotype associated with variation in mating status
348 in Mormon crickets (Smith *et al.* 2016). Morphologically, *Pediococcus acidilactici* were
349 nonmotile and spherical (0.8 – 1.0 μm), often dividing to form pairs as described for other
350 *Pediococcus.* As other members of the genus, the *P. acidilactici* were gram-positive, non-motile,
351 facultatively anaerobic, grow at low pH, and produce lactate from lactose (Table S2).
352 The second clade of lactic-acid bacteria was comprised primarily of plant-associated
353 *Lactobacillus.* Unlike *P. acidilactici*, these *Lactobacillus* formed a distinct clade with good
354 branch support (Fig 8), indicating it is genetically distinct enough at the 16S rRNA locus to
355 distinguish itself from other clades in the phylogeny. Similar to *P. acidilacitici*, these
356 *Lactobacillus* had high sequence similarity (>99.5%) to other members of the clade and a short
357 branch length, indicating that while it is distinct enough to form its own clade, it is not highly
358 derived from its relatives at the 16S rRNA locus.
359 Our Genbank search for *Lactobacillus* isolated from insect guts found sequences from
360 ants, bees, and termites, and fruit flies, all of which fell into a different clade than *Lactobacillus*
361 isolated from Mormon crickets. *Lactobacillus* from these taxa thus appear to have a different
362 evolutionary history. *Lactobacillus* isolates shared 100% sequence identity in the V4 region with
363 the Lactobacillaceae 2 phylotype sequenced using the Illumina platform in this study.
364 Morphologically, these *Lactobacillus* appear as non-motile straight rods, approximately 1.3-2
365 μm in length and 0.8-1.0 μm wide and are gram-positive, non-motile, facultatively anaerobic,
366 grow at low pH, and produce lactate from lactose (Table S2).
367 The seven Enterobacteriaceae strains were most similar to *Enterobacter* strains in our
368 BLAST search, which recovered sequences from a variety of plant and animal sources (sequence

similarity=98.7-99.8%). Our survey of Genbank found *Enterobacter* from alimentary tracts of a diverse group of insects, including termites, cockroaches, flies, beetles, stink bugs, bees, ants, and moths. Like other studies (Brenner *et al.* 2005), however, the 16S rRNA gene did not have enough signal to resolve relationships among *Enterobacter* and its relatives (data not shown) so we present a simpler phylogeny with the Mormon cricket isolates and type strains from the family (Fig. 9).

We found that our Mormon cricket isolates were interspersed with *Enterobacter*, *Klebsiella*, and *Escherichia* type strains. A multilocus sequencing approach is thus needed to improve the inference (Brenner *et al.* 2005). All seven strains isolated from Mormon crickets had 100% identity at the V4 region with the *Klebsiella* phylotype sequenced on the Illumina platform, however the phylogenetic (Fig. 9) and phenotypic data (Table S2) suggest that *Klebsiella* is unlikely to be a correct taxonomic assignment. Unlike most *Klebsiella*, cultured strains were motile, which is more typical of *Enterobacter* and other Enterobacteriaceae (Brenner *et al*. 2005). Morphologically, all isolates were straight rods, approximately 0.8-1.0 μm in length and 0.6-0.8 μm wide. Strains were gram-negative and facultatively anaerobic (Table S2).

DISCUSSION

We found striking differences in the diversity and structure of the gut microbiome in the Mormon cricket *Anabrus simplex*. While most taxa were represented in the foregut, midgut and hindgut, there were dramatic differences in relative abundance within the Lactobacillaceae and between the Lactobacillaceae and Enterobacteriaceae, the main families recovered in our culture and culture-independent studies. Predictions of their metabolic capabilities using PICRUSt suggest the potential for these gut bacteria to participate in the metabolism of complex

carbohydrates and defense against microbial pathogens, particularly among the enteric bacteria in the midgut and hindgut, and to a lesser extent, the lactic-acid bacteria. Finally, our phylogenetic analysis of cultured isolates found that Mormon cricket gut bacteria are not highly derived from related bacteria associated with plants or the guts of other animals, suggesting that gut bacteria are either acquired from the environment in each generation or have not been restricted to Mormon crickets over appreciable periods of evolutionary time. Our findings have important implications for our understanding of the ecological and evolutionary processes that influence the assembly and function of gut microbial communities, as it suggests that host-microbe and microbe-microbe interactions shape the abundance and distribution of the gut microbiome.

Our finding that bacterial abundance is lower in the midgut is in agreement with reports from other orthopterans (Ulrich, Buthala & Klug 1981; Hunt & Charnley 1981) and insects (Köhler *et al.* 2012), and has been attributed to characteristics that make the midgut less hospitable to bacteria than other regions of the alimentary tract (Douglas 2015). The midgut in insects secretes a host of digestive enzymes, is immunologically active, and lined by the peritrophic membrane, which acts as a protective barrier that restricts microbes to the lumen and protects the epithelium (Douglas 2015). In the two orthopterans that have been studied in detail, bacteria are found in the midgut lumen but not in association with the epithelium (Hunt & Charnley 1981; Mead, Khachatourians & Jones 1988). As a consequence, midgut bacteria might need to be continually replenished from ingested food (Blum *et al.* 2013) because the peritrophic membrane is continually shed into the hindgut. In some insects, specialized midgut crypts provide niches that microbes colonize (Kikuchi *et al.* 2005; Bistolas *et al.* 2014), however we did not observe analogous structures in Mormon crickets (Fig. 1).

The midgut is particularly vulnerable to pathogens because the lack of an endocuticle leaves the epithelium exposed once the peritrophic membrane is penetrated (Lehane & Billingsley 1996). The Mormon cricket midgut was populated by lactic-acid bacteria, with *Pediococcus* specifically exhibiting greater abundance in the midgut (and hindgut) than in the foregut. Lactic-acid bacteria are known for their beneficial effects in insects, increasing resistance to parasites in bees (Forsgren *et al.* 2010) and promoting development in fruit flies by enhancing proteolytic activity (Erkosar *et al.* 2015) and upregulating host ecdysone and insulin-like peptides (Storelli *et al.* 2011). Lactic-acid bacteria are also known to suppress pathogenic bacteria by reducing pH through the production of lactate and by producing a number of antimicrobial compounds, such as hydrogen peroxide and bacteriocins (Cintas *et al.* 2001).

A previous study found that sexual interactions in Mormon crickets influences the abundance of three *Pediococcus* phylotypes (Smith *et al.* 2016), however spatial information on where in the gut *Pediococcus* is located has been unavailable until now. *Pediococcus* in the midgut could provide immunological or nutritional benefits to Mormon crickets, as has been shown for *P. acidilactici* in other animals (Castex *et al.* 2008, 2009). We found that the capacity for lactate production in our samples was dominated by *Pediococcus* and other lactic-acid bacteria, although the abundance of the enzyme mediating lactate production was not higher in the midgut relative to other regions based on our metagenomics predictions. The cultured isolates of *P. acidilactici* obtained from Mormon crickets in this study will enable future experimental and comparative genomic approaches to evaluate these hypotheses.

Lactic-acid bacteria were also common in the foregut, which was dominated by a *Lactobacillus* that averaged 73.9% of the sequences recovered from this region. Bignell (1984) noted that the foregut of insects tends to be the most acidic compartment, however studies that

438 measure the physiochemical environment and characterize microbiome composition of the
439 foregut are rare (but see Köhler *et al.* 2012). This is because the endocuticle, lack of
440 differentiated cells for absorption of nutrients, and frequent purging of consumed material into
441 the midgut provides little opportunity for foregut microbes to contribute to host nutrition. The
442 large differences in community structure between the foregut and the rest of the alimentary tract
443 in our study does illustrate the dramatic transition in microbial communities between what is
444 ingested and what can colonize the more distal regions of the gut. Our metagenomics predictions
445 also suggest that the foregut is not the site of extensive carbohydrate metabolism or pathogen
446 defense for most of the pathways we examined.

447 In contrast to the foregut and midgut, the hindgut was characterized by a dramatic
448 increase in enteric bacteria (Enterobacteriaceae). Ordination of the laboratory Mormon cricket
449 samples indicated that the rectum, not the ileum, was primarily responsible for the difference in
450 community structure in the hindgut. Enterobacteriaceae comprised 83.5% of the sequences from
451 the rectum compared to 57.5% from the ileum in laboratory-raised animals, which was more
452 similar to the midgut in community structure (Fig. 4a). This distinction is potentially important
453 because higher digestive efficiency in conventional compared to germ-free crickets has been
454 attributed to microbial colonization of the ileum in the orthopteran *A domesticus* (Kaufman &
455 Klug 1991).

456 Metabolism of the specific complex carbohydrates attributed to bacteria in this study
457 were also identified in our metagenomic predictions and localized to the hindgut, as well as
458 enzymes involved in the production of the essential amino acid phenylalanine via the shikimate
459 pathway. Phenylalanine is a precursor for tyrosine, which is required to stabilize the cuticle
460 during molting (Bernays & Woodhead 1984) and in phenoloxidase synthesis, an important

461 component of the insect immune system (González-Santoyo & Córdoba-Aguilar 2012).
462 Tradeoffs between allocation of tyrosine to immune function and cuticle formation during
463 development in Mormon crickets (Srygley 2012) might be impacted by microbial contributions
464 to amino acid production if sufficient quantities of these amino acids are not obtained directly
465 from their diet.

466 Of the three enteric bacteria represented in this study, *P agglomerans* was common to
467 both field and lab individuals and increased in abundance in the hindgut. *Pantoea* are known
468 plant pathogens and have been associated with a variety of medical conditions in humans
469 (Walterson & Stavrinides 2015). In insects, however, *Pantoea* have been shown to have
470 mutualistic associations with their host. They are required for the completion of development in
471 stinkbugs (Hosokawa *et al.* 2016; but see Dillon & Charnley 2002), produce compounds that
472 attract insects to their host plants in flies (Robacker, Lauzon & He 2004; Maccollom *et al.* 2009),
473 and in the orthopteran *Schistocerca gregaria*, produce a key component of the locust aggregation
474 pheromone (Dillon, Vennard & Charnley 2000, 2002) and reduce susceptibility to microbial
475 pathogens through the production of phenols (Dillon & Charnley 1986, 1995).

476 Our metagenomics predictions suggest that enteric bacteria in Mormon crickets might be
477 capable of producing at least two of the antimicrobial phenols identified in *S. gregaria*, although
478 *P. agglomerans* was not identified as an important contributor in our study. This illustrates a
479 limitation of PICRUSt, as genomes in the IMG database used to make inferences about gene
480 content may miss important among-strain variation in metabolic capabilities. *P. agglomerans*
481 derived from *S. gregaria* are likely to have acquired this capability independently, unless the
482 metabolic pathway is different from the one analyzed here or the taxonomic designation reported
483 by Dillon and Charnley (1986, 1995) is incorrect.

Metagenomic analyses are also dependent upon annotation of the relevant pathways in the KEGG database. We were unable, for example, to assess the potential for the Mormon crickets microbiome to produce bacteriocins or the aggregation pheromone guaiacol, a bacterial metabolite produced by *P. agglomerans* in *S. gregaria* (Dillon *et al.* 2000, 2002), because the predicted pathways are not annotated in the KEGG database. The role of the gut microbiome in protecting Mormon crickets from their own pathogens (MacVean & Capinera 1991) and its influence on aggregation behavior (MacVean 1987; Simpson *et al.* 2006) is thus an important direction for future research.

Conclusion

Variation in morphology and physiology is thought to differentiate niches within the gut that influence the organization of the microbiome. Our study describes at high resolution how bacterial communities vary among gut regions, and suggests that host-microbe and/or microbe-microbe interactions have a role in how microbial communities are assembled and maintained. While the metagenomics predictions gleaned from our study suggests that some of these bacteria might benefit Mormon cricket nutrition, immunity, and perhaps even modulate social behavior, experiments are needed to evaluate this possibility. Our establishment of methods for culturing Mormon cricket gut bacteria will enable experimental and comparative genomic approaches in the future to infer the ecological and evolutionary consequences of host-microbe symbiosis.

Data Accessibility

Sequences are archived under NCBI BioProject PRJNA362233.

Acknowledgements

We thank Laura Senior, USDA-ARS, for help with field collections and rearing laboratory insects. We also thank Bruce Shambaugh and the Cheyenne office of USDA-APHIS, PPQ for assisting with field collections, and Nancy Moran, UT-Austin, for providing the standard for the qPCR experiment. Work was funded by National Science Foundation award DEB-1354666 and the W.M. Wheeler Lost Pines Endowment from the University of Texas at Austin to UGM.

References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2013) *lme4: Linear Mixed-Effects Models Using S4 Classes*.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bernays, E. & Woodhead, S. (1984) The need for high levels of phenylalanine in the diet of *Schistocerca gregaria* nymphs. *Journal of Insect Physiology*, **30**, 489–493.

Bignell, D.E. (1984) The arthropod gut as an environment for microorganisms. *Invertebrate-Microbial Interactions* (eds J.M. Anderson, A.D.M. Rayner & D.W.H. Walton), pp. 205–227. Cambridge University Press, Cambridge.

Bistolas, K.S.I., Sakamoto, R.I., Fernandes, J.A.M. & Goffredi, S.K. (2014) Symbiont polyphyly, co-evolution, and necessity in pentatomid stinkbugs from Costa Rica. *Microbial Symbioses*, **5**, 349.

Blum, J.E., Fischer, C.N., Miles, J. & Handelsman, J. (2013) Frequent replenishment sustains the beneficial microbiome of *Drosophila melanogaster*. *mBio*, **4**, e00860-13.

Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M., Boone, D.R., De Vos, P., Goodfellow, M., Rainey, F.A. & Schleifer, K.-H. (eds). (2005) *Bergey's Manual® of Systematic Bacteriology*. Springer US, Boston, MA.

Bridson, E.Y. (ed). (1998) *The Oxoid Manual*, 8th ed. Oxoid LTD, Basingstoke, United Kingdom.

Buchon, N., Broderick, N.A. & Lemaitre, B. (2013) Gut homeostasis in a microbial world: insights from *Drosophila melanogaster*. *Nature Reviews Microbiology*, **11**, 615–626.

Bugg, T.D., Ahmad, M., Hardiman, E.M. & Singh, R. (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Current Opinion in Biotechnology*, **22**, 394–400.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13**, 581–583.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

545 Castex, M., Chim, L., Pham, D., Lemaire, P., Wabete, N., Nicolas, J.-L., Schmidely, P. &
546 Mariojouls, C. (2008) Probiotic *P. acidilactici* application in shrimp *Litopenaeus*
547 *stylirostris* culture subject to vibriosis in New Caledonia. *Aquaculture*, **275**, 182–193.

548 Castex, M., Lemaire, P., Wabete, N. & Chim, L. (2009) Effect of dietary probiotic *Pediococcus*
549 *acidilactici* on antioxidant defences and oxidative stress status of shrimp *Litopenaeus*
550 *stylirostris*. *Aquaculture*, **294**, 306–313.

551 Cho, I. & Blaser, M.J. (2012) The human microbiome: at the interface of health and disease.
552 *Nature Reviews Genetics*, **13**, 260–270.

553 Cintas, L.M., Casaus, M.P., Herranz, C., Nes, I.F. & Hernández, P.E. (2001) Review:
554 Bacteriocins of Lactic Acid Bacteria. *Food Science and Technology International*, **7**,
555 281–305.

556 Dillon, R. & Charnley, A. (1986) Inhibition of *Metarhizium anisopliae* by the gut bacterial flora
557 of the desert locust, *Schistocerca gregaria*: Evidence for an antifungal toxin. *Journal of*
558 *invertebrate pathology*, **47**, 350–360.

559 Dillon, R.J. & Charnley, A.K. (1988) Inhibition of *Metarhizium anisopliae* by the gut bacterial
560 flora of the desert locust: characterisation of antifungal toxins. *Canadian Journal of*
561 *Microbiology*, **34**, 1075–1082.

562 Dillon, R.J. & Charnley, A.K. (1995) Chemical Barriers to Gut Infection in the Desert Locust: In
563 Vivo Production of Antimicrobial Phenols Associated with the Bacterium Pantoea
564 agglomerans. *Journal of Invertebrate Pathology*, **66**, 72–75.

565 Dillon, R. & Charnley, K. (2002) Mutualism between the desert locust *Schistocerca gregaria*
566 and its gut microbiota. *Research in Microbiology*, **153**, 503–509.

567 Dillon, R.J. & Dillon, V.M. (2004) The gut bacteria of insects: nonpathogenic interactions.
568 *Annual review of entomology*, **49**, 71–92.

569 Dillon, R.J., Vennard, C.T. & Charnley, A.K. (2000) Pheromones: Exploitation of gut bacteria in
570 the locust. *Nature*, **403**, 851–851.

571 Dillon, R. j., Vennard, C. t. & Charnley, A. k. (2002) A Note: Gut bacteria produce components
572 of a locust cohesion pheromone. *Journal of Applied Microbiology*, **92**, 759–763.

573 Douglas, A.E. (2013) Alimentary canal, digestion and absorption. *The Insects: Structure and*
574 *Function*, 5th ed (eds S.. J. Simpson, A.E. Douglas & R.F. Chapman), pp. 46–80.
575 Cambridge University Press, Cambridge.

576 Douglas, A.E. (2015) Multiorganismal Insects: Diversity and Function of Resident
577 Microorganisms. *Annual Review of Entomology*, **60**, 17–34.

578 Drawz, S.M. & Bonomo, R.A. (2010) Three decades of β-Lactamase inhibitors. *Clinical*
579 *Microbiology Reviews*, **23**, 160–201.

580 Engel, P. & Moran, N.A. (2013) The gut microbiota of insects – diversity in structure and
581 function. *FEMS Microbiology Reviews*, **37**, 699–735.

582 Erkosar, B., Storelli, G., Mitchell, M., Bozonnet, L., Bozonnet, N. & Leulier, F. (2015) Pathogen
583 virulence impedes mutualist-mediated enhancement of host juvenile growth via inhibition
584 of protein digestion. *Cell Host & Microbe*, **18**, 445–455.

585 Faraway, J.J. (2006) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and*
586 *Nonparametric Regression Models*. Chapman & Hall/CRC, New York.

587 Forsgren, E., Olofsson, T.C., Vásquez, A. & Fries, I. (2010) Novel lactic acid bacteria inhibiting
588 *Paenibacillus* larvae in honey bee larvae. *Apidologie*, **41**, 99–108.

589 González-Santoyo, I. & Córdoba-Aguilar, A. (2012) Phenoloxidase: a key component of the
590 insect immune system. *Entomologia Experimentalis et Applicata*, **142**, 1–16.

591 Gwynne, D.T. (1984) Sexual selection and sexual differences in Mormon crickets (Orthoptera:
592 Tettigoniidae, *Anabrus simplex*). *Evolution*, **38**, 1011–1022.

593 Herrmann, K.M. & Weaver, L.M. (1999) The shikimate pathway. *Annual review of plant*
594 *biology*, **50**, 473–503.

595 Hosokawa, T., Ishii, Y., Nikoh, N., Fujie, M., Satoh, N. & Fukatsu, T. (2016) Obligate bacterial
596 mutualists evolving from environmental bacteria in natural insect populations. *Nature*
597 *Microbiology*, **1**, 15011.

598 Hothorn, T., Bretz, F. & Westfall, P. (2008) Simultaneous inference in general parametric
599 models. *Biometrical Journal*, **50**, 346–363.

600 Hunt, J. & Charnley, A.K. (1981) Abundance and distribution of the gut flora of the desert
601 locust, Schistocerca gregaria. *Journal of Invertebrate Pathology*, **38**, 378–385.

602 Kaufman, M.G. & Klug, M.J. (1991) The contribution of hindgut bacteria to dietary
603 carbohydrate utilization by crickets (Orthoptera: Gryllidae). *Comparative Biochemistry*
604 *and Physiology Part A: Physiology*, **98**, 117–123.

605 Kikuchi, Y., Meng, X.-Y. & Fukatsu, T. (2005) Gut symbiotic bacteria of the genus burkholderia
606 in the broad-headed bugs *Riptortus clavatus* and *Leptocorisa chinensis* (Heteroptera:
607 Alydidae). *Applied and Environmental Microbiology*, **71**, 4035–4043.

608 Köhler, T., Dietrich, C., Scheffrahn, R.H. & Brune, A. (2012) High-resolution analysis of gut
609 environment and bacterial microbiota reveals functional compartmentation of the gut in
610 wood-feeding higher termites (*Nasutitermes* spp.). *Applied and Environmental*
611 *Microbiology*, **78**, 4691–4701.

612 Lanan, M.C., Rodrigues, P.A.P., Agellon, A., Jansma, P. & Wheeler, D.E. (2016) A bacterial
613 filter protects and structures the gut microbiome of an insect. *The ISME Journal*.

614 Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A.,
615 Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G. &
616 Huttenhower, C. (2013) Predictive functional profiling of microbial communities using
617 16S rRNA marker gene sequences. *Nature Biotechnology*, **31**, 814–821.

618 Lehane, M.J. & Billingsley, P.F. (eds). (1996) *Biology of the Insect Midgut*. Springer
619 Netherlands, Dordrecht.

620 Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R. & Gordon, J.I. (2008) Worlds within
621 worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, **6**, 776–
622 788.

623 Maccollom, G.B., Lauzon, C.R., Sjogren, R.E., Meyer, W.L. & Olday, F. (2009) Association and
624 attraction of blueberry maggot fly curran (Diptera: Tephritidae) to *Pantoea*
625 *(enterobacter) agglomerans*. *Environmental Entomology*, **38**, 116–120.

626 MacVean, C.. (1987) Ecology and management of the Mormon cricket, *Anabrus simplex*.
627 *Integrated Pest Management on Rangeland: A Shortgrass Prairie Perspective*, pp. 116–
628 136. Westview Press, Boulder, CO.

629 MacVean, C.M. & Capinera, J.L. (1991) Pathogenicity and transmission potential of Nosema
630 locustae and Vairimorpha n. sp. (Protozoa: Microsporida) in Mormon crickets (Anabrus
631 simplex; Orthoptera: Tettigoniidae): A laboratory evaluation. *Journal of Invertebrate*
632 *Pathology*, **57**, 23–36.

633 McFrederick, Q.S., Cannone, J.J., Gutell, R.R., Kellner, K., Plowes, R.M. & Mueller, U.G.
634 (2013) Specificity between *Lactobacilli* and hymenopteran hosts is the exception rather
635 than the rule. *Applied and Environmental Microbiology*, **79**, 1803–1812.

636 McMurdie, P.J. & Holmes, S. (2013) Phyloseq: an R package for reproducible interactive
637 analysis and graphics of microbiome census data. *PLoS ONE*, **8**, e61217.

638 Mead, L.J., Khachatourians, G.G. & Jones, G.A. (1988) Microbial ecology of the gut in
639 laboratory stocks of the migratory grasshopper, *Melanoplus sanguinipes* (fab.)
640 (Orthoptera: Acrididae). *Applied and Environmental Microbiology*, **54**, 1174–1181.

641 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson,
642 G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2015) *Vegan: Community Ecology*
643 *Package*.

644 Powell, J.E., Martinson, V.G., Urban-Mead, K. & Moran, N.A. (2014) Routes of acquisition of
645 the gut microbiota of the honey bee *Apis mellifera*. *Applied and Environmental*
646 *Microbiology*, **80**, 7378–7387.

647 R Core Development Team. (2013) *R: A Language and Environment for Statistical Computing*.
648 R Foundation for Statistical Computing, Vienna, Austria.

649 Raskin, I. (1992) Role of salicylic acid in plants. *Annual review of plant biology*, **43**, 439–463.

Robacker, D.C., Lauzon, C.R. & He, X. (2004) Volatiles production and attractiveness to the Mexican fruit fly of *Enterobacter agglomerans* isolated from apple maggot and Mexican fruit flies. *Journal of Chemical Ecology*, **30**, 1329–1347.

Ryu, J.-H., Ha, E.-M. & Lee, W.-J. (2010) Innate immunity and gut–microbe mutualism in *Drosophila*. *Developmental & Comparative Immunology*, **34**, 369–376.

Servin, A.L. (2004) Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiology Reviews*, **28**, 405–440.

Simpson, S.J., Sword, G.A., Lorch, P.D. & Couzin, I.D. (2006) Cannibal crickets on a forced march for protein and salt. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 4152–4156.

Smith, C.C., Srygley, R.B., Dietrich, E.I. & Mueller, U.G. (2016) Partitioning the effect of mating and nuptial feeding on the microbiome in gift-giving insects. *Environmental Microbiology Reports*.

Srygley, R.B. (2012) Ontogenetic changes in immunity and susceptibility to fungal infection in Mormon crickets *Anabrus simplex*. *Journal of Insect Physiology*, **58**, 342–347.

Srygley, R.B. & Lorch, P.D. (2011) Weakness in the band: nutrient-mediated trade-offs between migration and immunity of Mormon crickets, *Anabrus simplex*. *Animal Behaviour*, **81**, 395–400.

Srygley, R.B., Lorch, P.D., Simpson, S.J. & Sword, G.A. (2009) Immediate protein dietary effects on movement and the generalised immunocompetence of migrating Mormon crickets *Anabrus simplex* (Orthoptera: Tettigoniidae). *Ecological Entomology*, **34**, 663–668.

Storelli, G., Defaye, A., Erkosar, B., Hols, P., Royet, J. & Leulier, F. (2011) *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metabolism*, **14**, 403–414.

Tai, V., Carpenter, K.J., Weber, P.K., Nalepa, C.A., Perlman, S.J. & Keeling, P.J. (2016) Genome evolution and nitrogen fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. *Applied and Environmental Microbiology*, **82**, 4682–4695.

Ulrich, R.G., Buthala, D.A. & Klug, M.J. (1981) Microbiota associated with the gastrointestinal tract of the common house cricket, *Acheta domestica*. *Applied and Environmental Microbiology*, **41**, 246–254.

Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*, Fourth. Springer, New York.

Walterson, A.M. & Stavrinides, J. (2015) Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiology Reviews*, fuv027.

685 Woodring, J. & Lorenz, M.W. (2007) Feeding, nutrient flow, and functional gut morphology in
686 the cricket *Gryllus bimaculatus*. *Journal of Morphology*, **268**, 815–825.

687 Xu, W.-H., Liang, Q., Zhang, Y.-J. & Zhao, P. (2015) Naturally occurring arbutin derivatives
688 and their bioactivities. *Chemistry & Biodiversity*, **12**, 54–81.

689 Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z. & Forney, L.J. (2012) Evaluation of methods for the
690 extraction and purification of dna from the human microbiome. *PLoS ONE*, **7**, e33865.

691

692

693 **Figures and Tables**

694

695 Table 1. Analysis of deviance comparing alpha diversity between source populations (wild or laboratory) and among tissue types
696 (foregut, midgut, and hindgut). Values represent the F-statistic (p-value) for each term. Statistically significant terms ($p<0.05$) are
697 indicated in bold. Degrees of freedom were estimated using the Kenward-Rogers approximation. In the reduced dataset, two
698 individuals from the laboratory-reared population were removed (see Methods).

	Reduced Dataset			**Full Dataset**		
	Species Richness	Chao1	Shannon-Weiner	Species Richness	Chao1	Shannon-Weiner
Source	**13.9 (0.003)**	**11.0 (0.007)**	**9.17(0.01)**	**14.0(0.002)**	**10.5 (0.006)**	**8.22 (0.013)**
Tissue type	**5.85 (0.010)**	**4.79 (0.02)**	**7.07 (0.005)**	**6.77 (0.004)**	**5.68 (0.008)**	**8.44 (0.001)**
Interaction	0.51 (0.61)	1.02 (0.38)	0.98 (0.39)	0.66 (0.53)	1.15 (0.33)	1.28 (0.29)

Table 2. Permutation tests of distance-based redundancy analyses of DADA2 16S rRNA sequence variants and PICRUSt metagenomic predictions. Source population (wild or laboratory) and tissue type (foregut, midgut and hindgut) were entered as factors.

	Reduced Dataset		**Full Dataset**	
	F	*P*	*F*	*P*
DADA2				
Source	**8.99**	**<0.001**	**7.75**	**<0.001**
Tissue type	**9.85**	**<0.001**	**5.49**	**<0.001**
Interaction	0.26	0.47	0.52	0.84
PICRUSt				
Source	0.23	0.76	**4.31**	**0.04**
Tissue type	**1.61**	**0.003**	**8.09**	**0.002**
Interaction	0.16	0.35	0.92	0.43

Table 3. Likelihood ratio tests from GLMMs fitting the abundance of sequence variants to source population (wild or laboratory) and tissue type (foregut, midgut or hindgut). Values are Chi-square (p-value). LAC1=*Lactobacillus sp.*, PED=*Pediococcus*, PAG=*Pantoea agglomerans*.

	Reduced Dataset			**Full Dataset**		
	LAC1	PED	PAG	LAC1	PED	PAG
Source	1.32 (0.25)	0.80 (0.37)	2.36 (0.12)	0.18 (0.66)	0.51 (0.48)	0.01 (0.96)
Tissue type	**16.3 (<0.001)**	**9.25 (0.01)**	**20.7 (<0.001)**	**12.7 (0.002)**	**16.1 (<0.001)**	**41.9 (<0.001)**
Interaction	0.36 (0.84)	0.84 (0.66)	3.86 (0.15)	0.3 (0.86)	0.95 (0.62)	2.67 (0.26)

Table 4. Permutation test of distance-based redundancy analysis of PICRUSt metagenomics predictions between source populations (wild or laboratory) and among tissue types (foregut, midgut and hindgut).

	Reduced Dataset		**Full Dataset**	
	F	*P*	*F*	*P*
Illumina 16S rRNA				
Source	**8.99**	**<0.001**	**7.75**	**<0.001**
Tissue type	**9.85**	**<0.001**	**5.49**	**<0.001**
Interaction	0.26	0.47	0.52	0.84
KEGG Pathways				
Source	0.23	0.76	**4.31**	**0.04**
Tissue type	**1.61**	**0.003**	**8.09**	**0.002**
Interaction	0.16	0.35	0.92	0.43

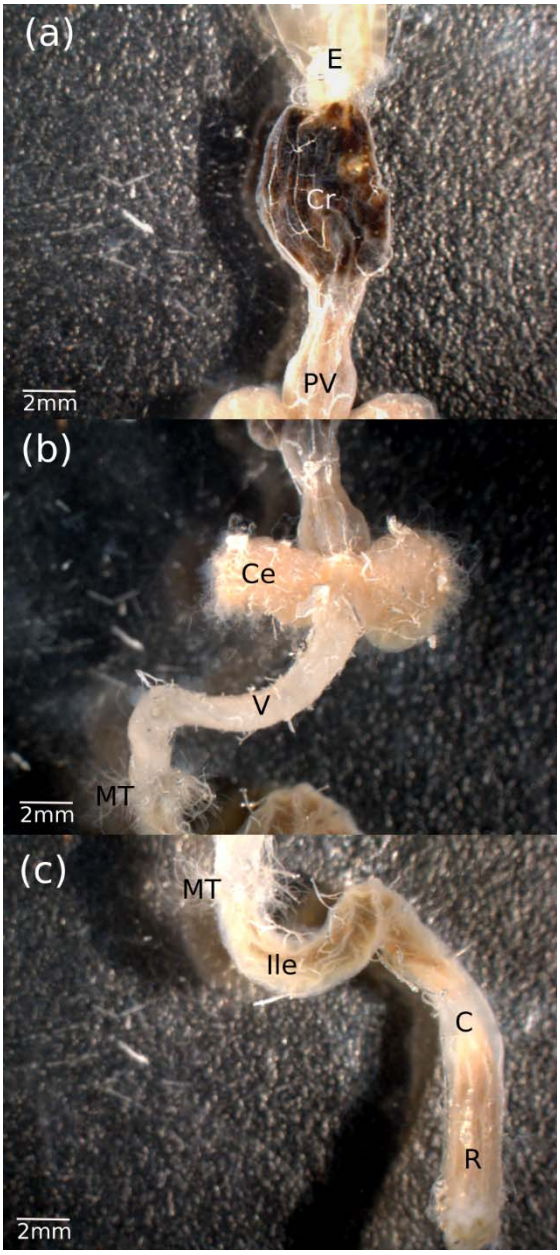

Figure 1. External morphology of the (a) foregut, (b) midgut, and (c) hindgut in the Mormon cricket. E=esophagus, Cr=crop, PV=proventriculus, Ce=cecum, V=ventriculus, MT=Malphigian tubules, Ile=ileum, C=colon, R=rectum. Malphigian tubules have been trimmed to illustrate their entry point into the hindgut.

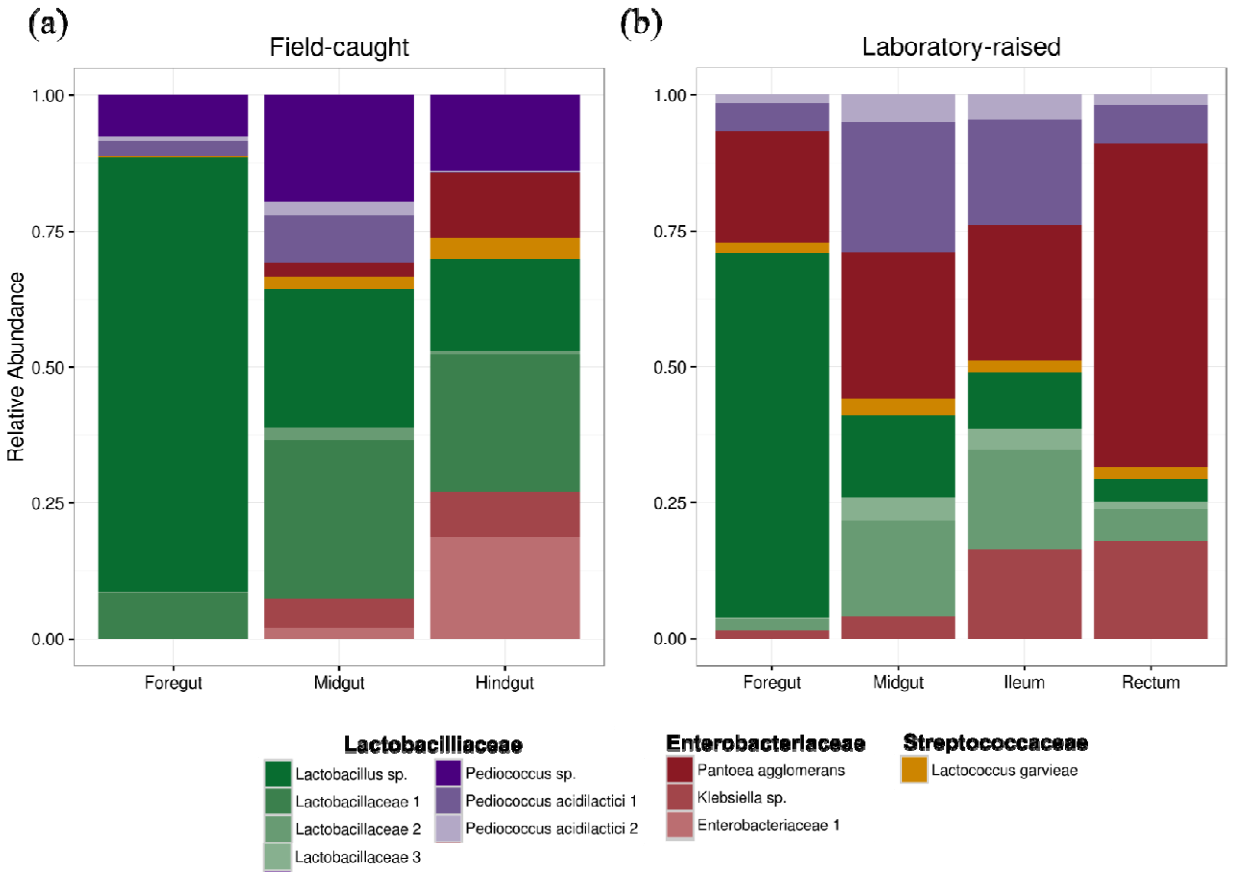

Figure 2. Mean relative abundance of the eleven dominant sequence variants from (a) field-caught and (b) laboratory-raised Mormon crickets from 16S rRNA Illumina sequencing.

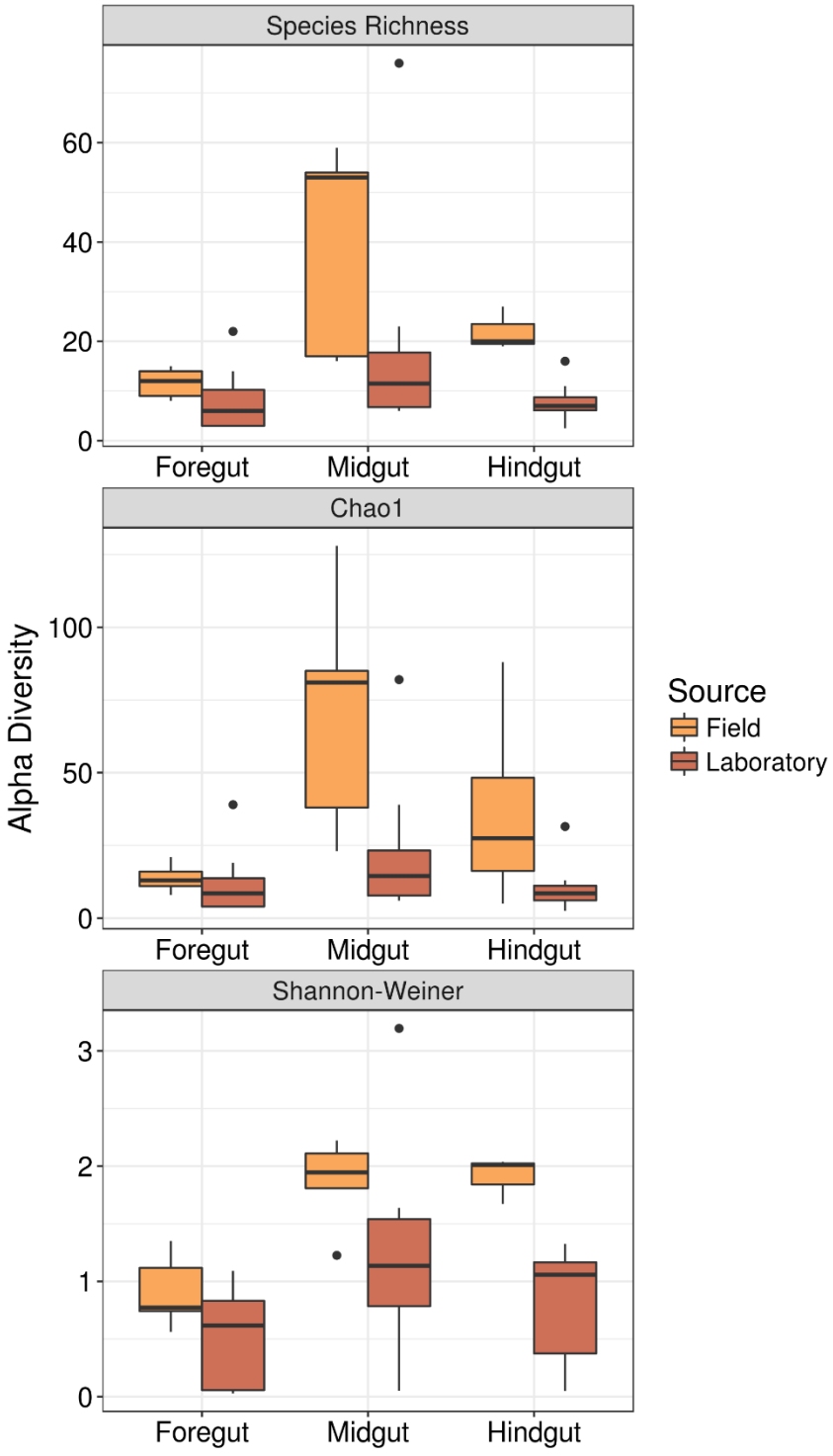

Figure 3. Alpha diversity in field-caught and laboratory-raised Mormon crickets.

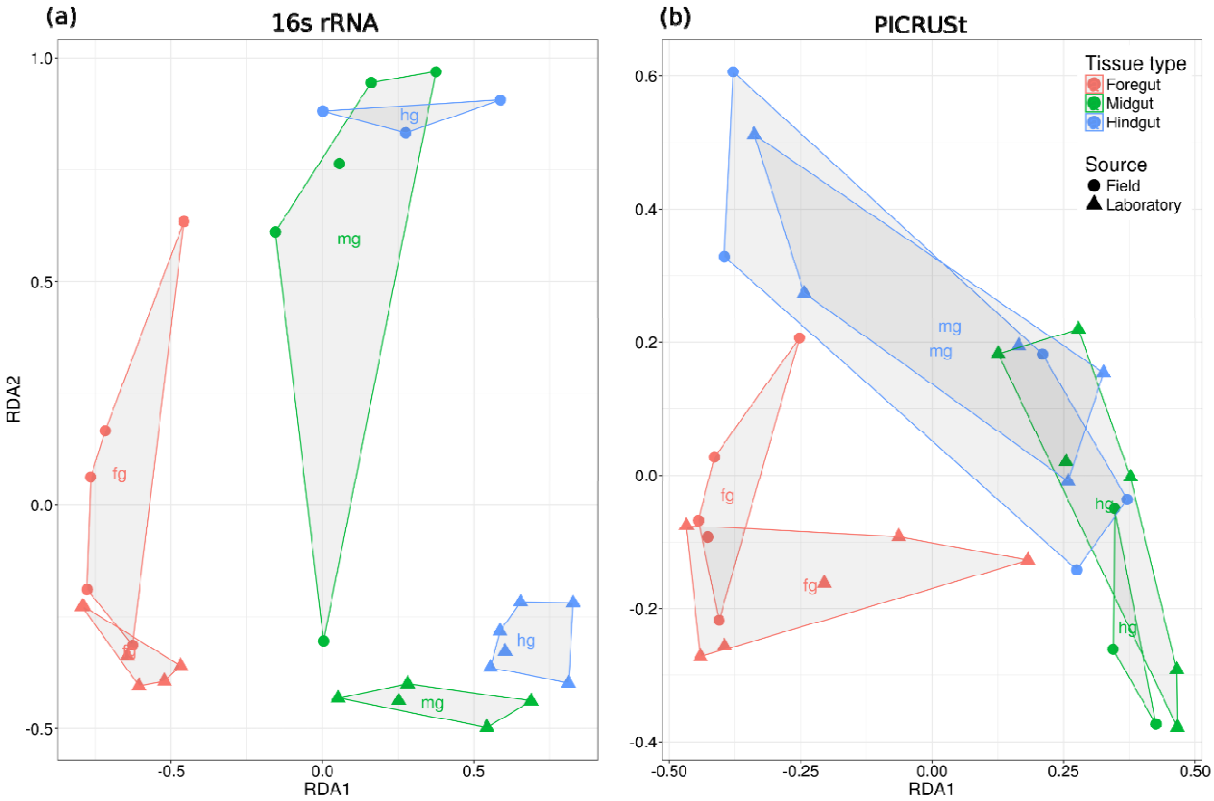

Figure 4. Ordination of sample scores from the db-RDA of the reduced dataset for (a) Illumina 16S sequencing and (b) PICRUSt metagenomic predictions.

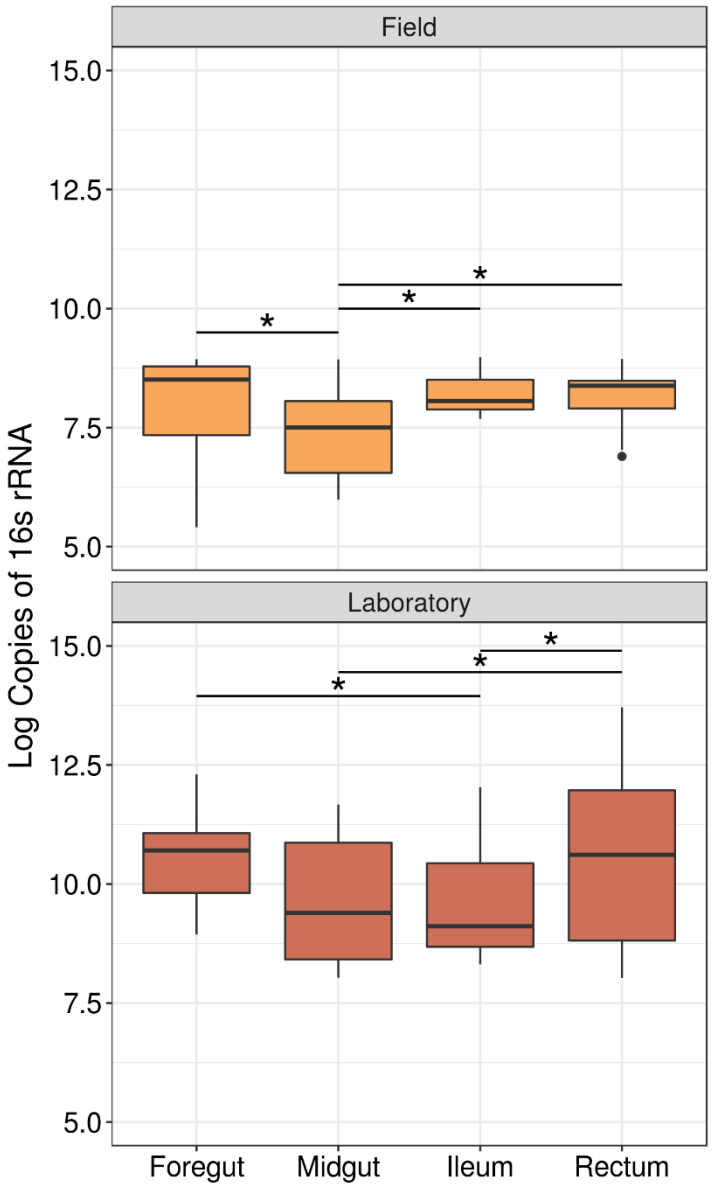

Figure 5. Abundance of bacterial 16S rRNA genes in the Mormon crickets gut. Bars indicate significant differences between regions (* $p<0.05$).

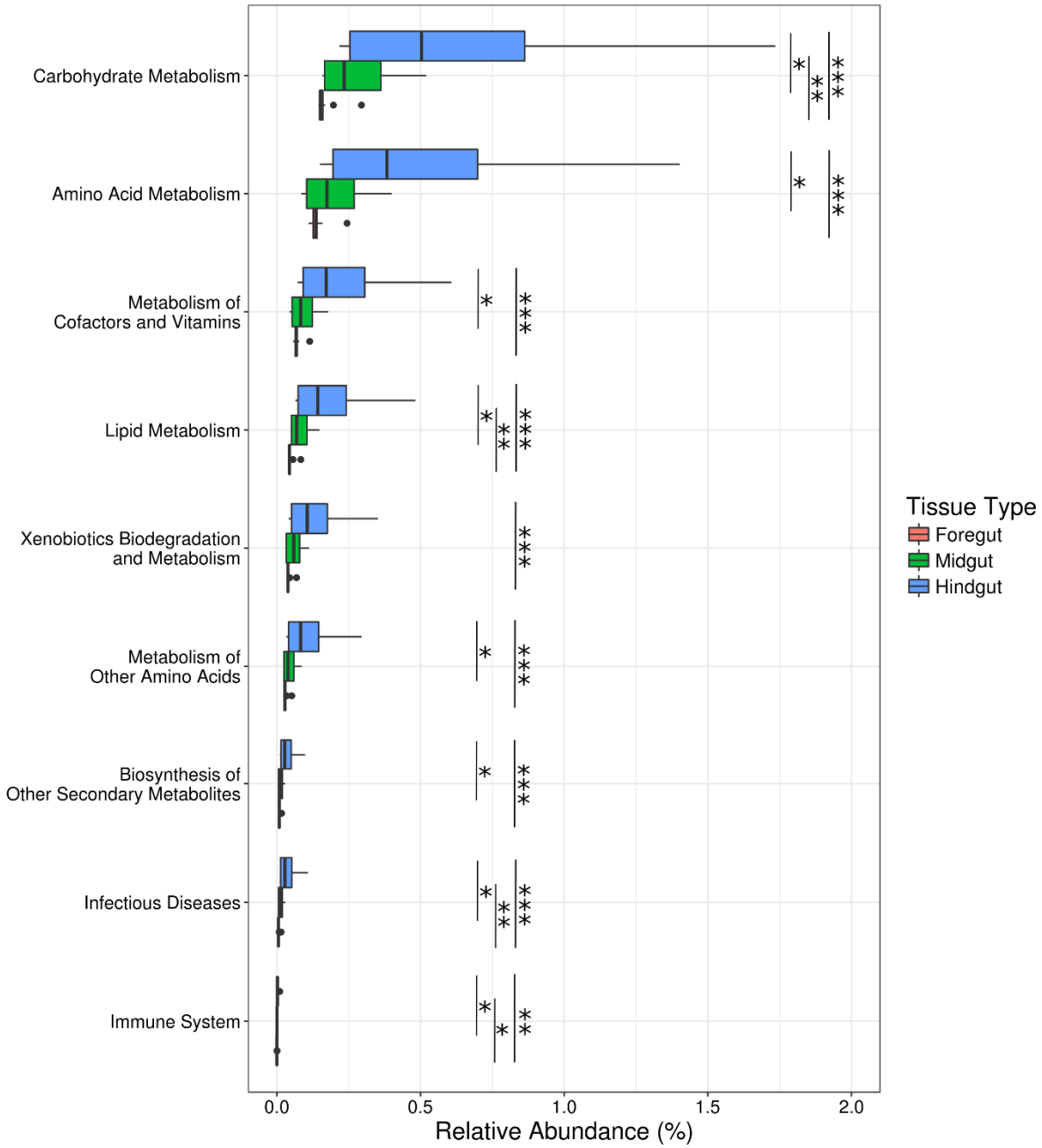

Figure 6. Relative abundance of KEGG pathways related to nutrition, immunity, xenobiotic degradation, and secondary metabolite production among tissue types.

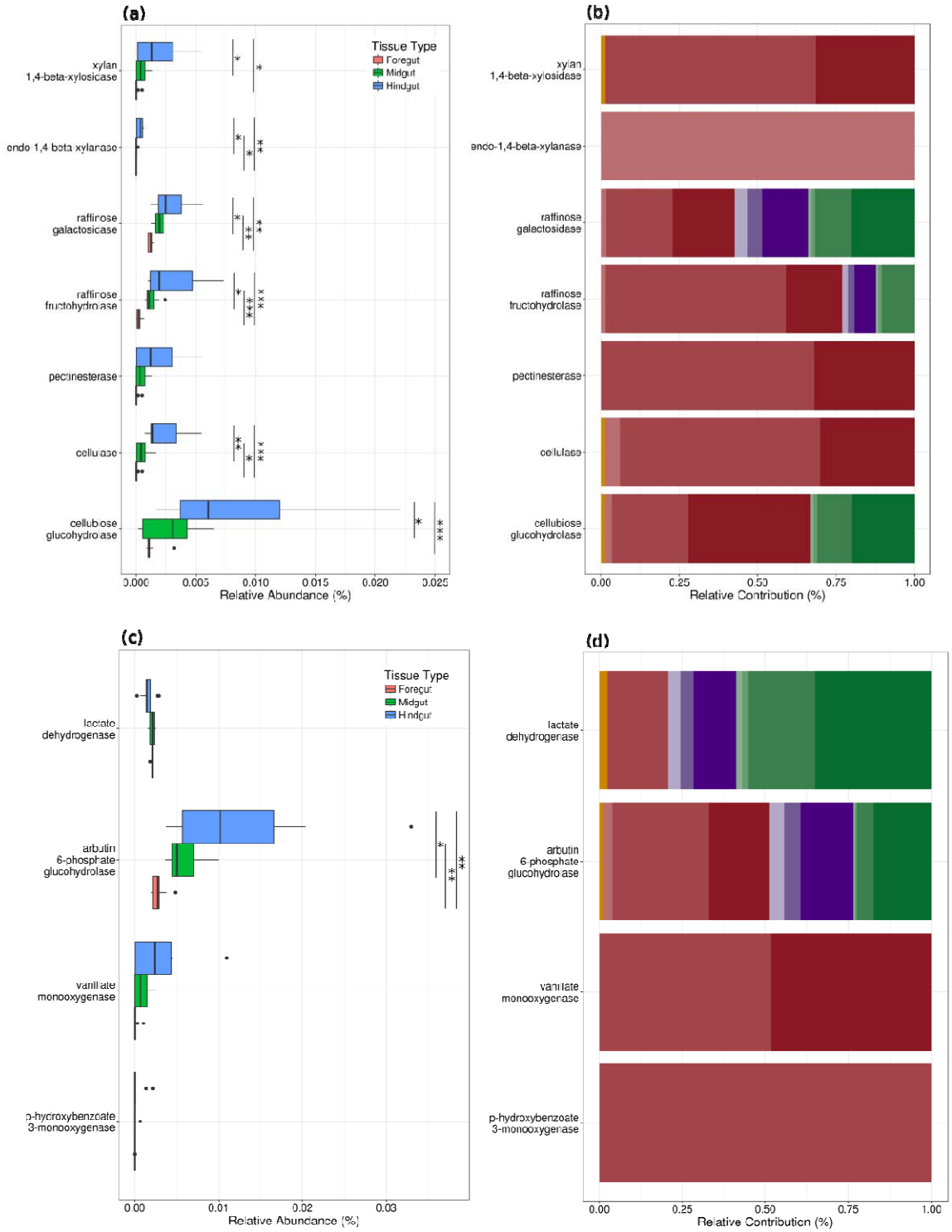

Figure 7. Abundance of KEGG orthologs related to (a) carbohydrate metabolism and (b) antimicrobial compound production among tissue types, and the contributions of taxonomic groups (c,d) to ortholog abundance. Key for colors representing taxonomic groups are in Figure 2.

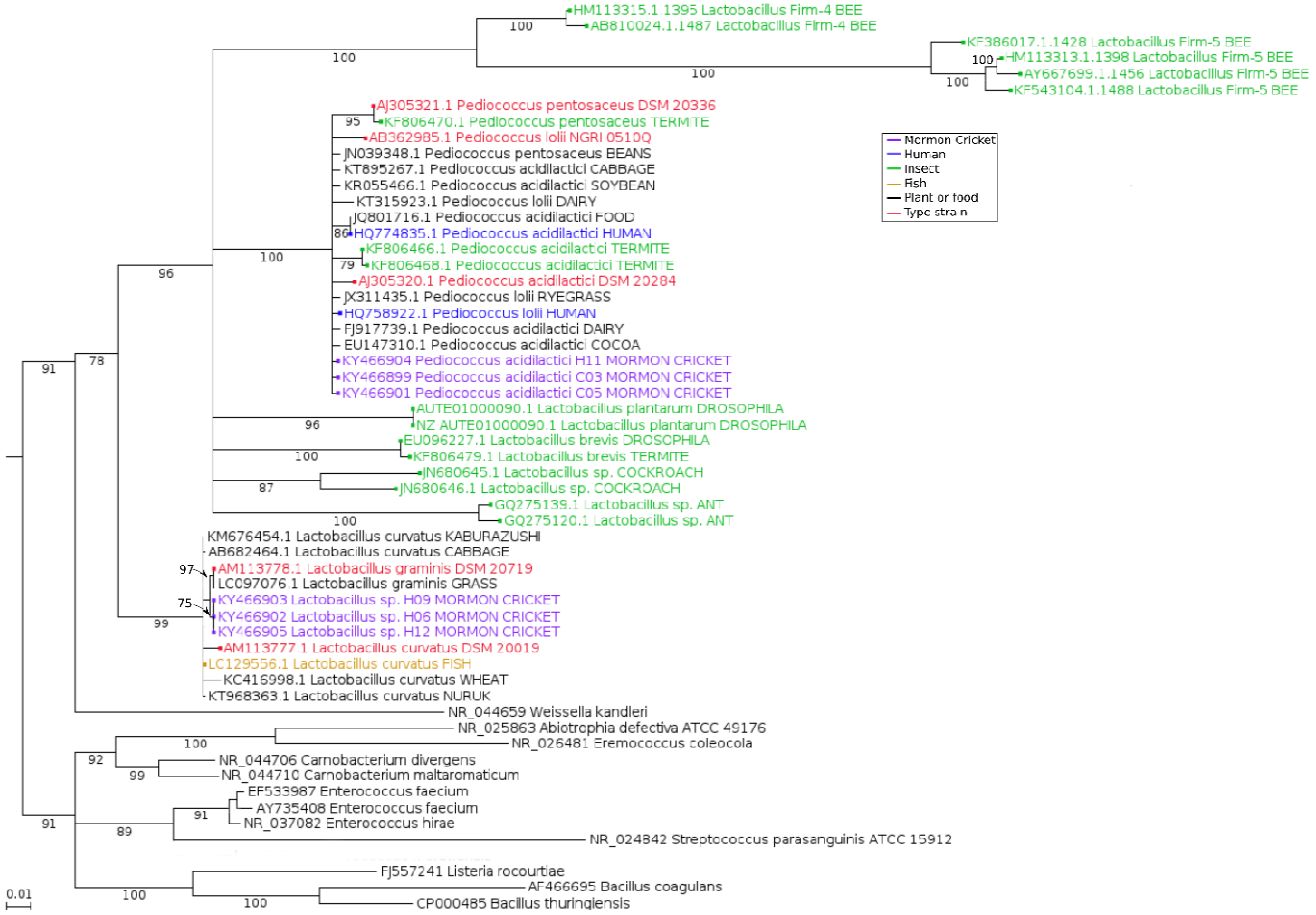

Fig 8. Maximum likelihood estimation of phylogenetic relationships among lactic-acid bacteria 16S rRNA sequences from Mormon cricket gut isolates and their relatives. Branches with bootstrap support <75% are collapsed.

Figure 9. Maximum likelihood estimation of phylogenetic relationships among enteric bacteria 16S rRNA sequences from Mormon cricket gut isolates and type strains. Branches with bootstrap support <50% are collapsed.