Geographical Distribution and Diversity of Bacteria Associated with Natural Populations of *Drosophila melanogaster*\(^\dagger\)

Vanessa Corby-Harris, \(^*\) Ana Clara Pontaroli, Lawrence J. Shimkets, Jeffrey L. Benetzen, Kristin E. Habel, and Daniel E. L. Promislow

Department of Genetics \(^1\) and Department of Microbiology, \(^2\) University of Georgia, Athens, Georgia

Received 7 September 2006/Accepted 24 March 2007

* * *

*Drosophila melanogaster* is one of the most widely used model systems in biology. However, little is known about its associated bacterial community. As a first step towards understanding these communities, we compared bacterial 16S rRNA gene sequence libraries recovered from 11 natural populations of adult *D. melanogaster*. Bacteria from these sequence libraries were grouped into 74 distinct taxa, spanning the phyla *Proteobacteria*, *Bacteroidetes*, and *Firmicutes*, which were unevenly spread across host populations. Summed across populations, the distribution of abundance of genera was closely fit by a power law. We observed differences among host population locations both in bacterial community richness and in composition. Despite this significant spatial variation, no relationship was observed between species richness and a variety of abiotic factors, such as temperature and latitude. Overall, bacterial communities associated with adult *D. melanogaster* hosts are diverse and differ across host populations.

Insects harbor diverse microbial communities (11, 29, 34, 65), and interactions between hosts and their microbes can range from mutualistic, such as the interaction between termites and their gut microbes (8, 65), to parasitic, such as the interaction of the bacterium *Paenibacillus larvae* (American foulbrood) in honeybees (62). Some of these interactions are relatively well characterized, owing to their economic importance or because of their remarkable biology. However, the exact nature of many other potentially interesting and experimentally tractable insect-microbe interactions, specifically those between microbes and the major insect model systems, remains poorly understood.

In addition to the immediate association between insect hosts and the bacterial communities they harbor, the bacteria that insects carry can also associate with and affect the fitness of other hosts through vector transmission. The most common vector-borne zoonotic inflammatory disease in the United States, Lyme disease (caused by *Borrelia burgdorferi*), is transmitted by the deer tick, *Ixodes scapularis*, and infected more than 23,000 people in 2002 (26). In addition, *Erwinia carotovora*, responsible for soft rot in many species of plants and for significant economic losses, can be vector transmitted by a variety of insects, including *Drosophila melanogaster* (38). Clearly, vector-borne bacterial infections can have large economic and health impacts and are important determinants of fitness for a variety of potential hosts.

It is estimated that approximately 99% of the bacteria in nature are unculturable (3). With the advent of molecular techniques, such as PCR and genome sequencing, and metagenomic approaches, researchers have uncovered an astonishing level of microbial diversity in natural habitats, ranging from soil (7, 54, 69) and marine environments (23, 64, 69, 70) to the human gut (25). The same techniques are currently being applied to understanding the microbiota of a range of insects (11, 29, 34, 47, 53, 55, 56, 65). For example, using such sequence-based approaches, Dunn and Stabb (20) found that the ant lion, *Myrmeleon mobilis*, harbors a relatively simple microbial community, represented mostly by *Enterobacteriaceae* and *Wolbachia*-like microorganisms. In contrast, Campbell et al. (11) found a more diverse microbiota in the biting midge, *Culicoides sonorensis*, comprised of genera from five different bacterial divisions. From the literature on the microbial communities of insects, two salient points emerge. First, these communities differ greatly among host species. Second, researchers are only beginning to understand the taxa comprising these microbial communities and how these microbes interact with their hosts. Considering the estimated 30 million insect species worldwide (21), the potential for uncovering new species of bacteria and for understanding interesting features of these insect-microbe interactions is staggering.

Despite their widespread use in the laboratory, relatively little is known about the interaction between model laboratory insect species and their associated microbial communities. Due to its experimental and genetic tractability, the fruit fly, *Drosophila melanogaster*, provides an ideal system for studying these interactions more closely. Recently, there have been numerous studies on *Drosophila* immunity, ranging from understanding the molecular basis of resistance (28, 32, 41) to the evolutionary ecology of the immune response (57, 61, 63). Some of these studies use bacteria isolated from natural populations of *Drosophila* (42, 71). However, in many other cases, researchers use bacteria that are not yet known to naturally cooccur with or infect *Drosophila*. One plausible reason that naturally occurring bacteria are not used to study immunity is that, to date, there are few comprehensive studies of the bacteria that associate with natural *D. melanogaster* populations (but see reference 19). Knowledge of the actual interactions...
taking place between insect hosts and their microbial communities is critical to those studying insect immunity. However, before characterizing these interactions, we must characterize the bacterial communities and identify taxa that are of potential ecological interest.

Specific species of microbes, as well as traits of the microbial population overall, such as community richness or composition, have the potential to greatly affect the ecology and evolution of their Drosophila hosts. Here we characterize this composition and richness across 11 natural populations of D. melanogaster. We first identify the bacterial microbes present in host populations collected across a latitudinal cline by using a sequence-based approach. We then compare microbial community richness among these host populations and ask whether richness is associated with latitude or climate. This is the first study to characterize microbial communities associated with D. melanogaster hosts within and among natural host populations. These data will provide an important first step in understanding host-microbe interactions in this widely studied model system.

MATERIALS AND METHODS

Fly collections. Flies were collected from 11 sites along a latitudinal transect on the east coast of the United States between June and September 2005 (Table 1), using a combination of sweep netting behind fruit stands and sweep netting over fruit bucket traps. Details of the 11 collection sites are presented in Table 1.

Following collection, flies were anesthetized over ice. Male D. melanogaster flies were separated from other drosophilid species based on the morphological characteristics, such as size, color, body patterning, wing shape, and genital morphology (4, 45). These individuals were set aside and preserved in groups of five in 70% ethanol. After our return to the laboratory from the collection sites, the ethanol-preserved samples were kept at −80°C. In preparation for DNA isolation, the ethanol-preserved flies were shaken before being removed from the ethanol. While this method will not remove all surface-associated bacteria, those that are loosely associated will likely wash off.

DNA isolation. Total DNAs from flies and bacteria were isolated from one group of five male flies at each of the 11 different locations. Flies were removed from the ethanol and homogenized using a pestle in 200 μl of STE buffer (10 mM Tris·HCl [pH 8.0], 1 mM EDTA, and 150 mM NaCl), with lysozyme added (final concentration, 4 mg/ml). The samples were then incubated for 30 min at 37°C. Following incubation, 20 μl of 10% sodium dodecyl sulfate and proteinase K (final concentration, 0.2 mg/ml) were added. These samples were then vortexed and incubated at 55°C overnight. After the overnight incubation, RNase A was added (final concentration, 0.1 mg/ml), and samples were incubated for 1 hour at 37°C. The samples were then extracted with equal volumes of phenol-chloroform and chloroform-isoamyl alcohol (24:24:1) and chloroform-isoamyl alcohol (24:1), and the DNAs were ethanol precipitated. The DNA pellets were resuspended in 50 μl of Tris-EDTA (10 mM Tris·HCl [pH 8.0], 1 mM EDTA [pH 8.0]).

PCR amplification. Bacterial 16S rRNA gene sequences were selectively PCR amplified from the isolated DNA samples for the construction of clone libraries. Each reaction tube contained 50 to 100 μg/ml of template genomic DNA, forward primer 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′), reverse primer 1522r (5′-AAG GAG GTG ATC CAG CCG CA-3′), and one Ready-to-Go PCR bead (GE Healthcare Life Science). The PCR program was as follows: 9-min at 95°C, 15 cycles of 1 min at 95°C, 1 min at 55°C, and 2 min at 72°C, and a final extension step of 60°C for 10 min. PCR products were resolved on a 1% agarose gel, and the gel was stained with SYBR green (Invitrogen) in order to visualize the relatively weak bands on a dark red transilluminator. The approximately 1.5-kb 16S rRNA gene fragment was extracted from the agarose gel by using a Quick Gel extraction kit (QIAAGEN) according to the manufacturer’s directions and was eluted in 30 μl of sterile distilled water.

Construction of clone libraries. Clone libraries of bacterial PCR products were constructed using a TOPO TA cloning kit (Invitrogen) with TOPO One Shot electrocompetent cells. Successful transformants were plated onto Luria-Bertani plates containing kanamycin (final concentration, 50 μg/ml). Plasmids were extracted from the bacteria by using standard techniques (59).

Sequencing of 16S rRNA genes. Partial-length 16S rRNA gene fragments were sequenced in one direction, using an ABI 3700 capillary sequencer with T7 forward and reverse primer (Table 1). The proportion of OTUs observed out of the estimated total number of OTUs in the population. As an estimate of the total number of OTUs, the Chao1 estimate of community richness (13, 15), using 3% sequence divergence, ± standard error around the estimate (14, 15), and the ACE estimate of community richness (14, 15), using 3% sequence divergence, ± standard error around the estimate. Calculations for standard errors were done as described previously (13, 15).

<table>
<thead>
<tr>
<th>Site name</th>
<th>Population size</th>
<th>Latitude</th>
<th>Date collected</th>
<th>Ambient temp (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Oakland</td>
<td>500</td>
<td>39.86</td>
<td>25 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>B Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>C Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>D Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>E Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>F Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>G Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>H Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>I Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>J Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>K Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of 11 collection sites and 16S rRNA gene sequence clone libraries.

<table>
<thead>
<tr>
<th>Site name</th>
<th>Population size</th>
<th>Latitude</th>
<th>Date collected</th>
<th>Ambient temp (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Oakland</td>
<td>500</td>
<td>39.86</td>
<td>25 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>B Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>C Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>D Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>E Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>F Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>G Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>H Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>I Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>J Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
<tr>
<td>K Eau Claire</td>
<td>250</td>
<td>45.02</td>
<td>13 August 2005</td>
<td>23.2</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of 11 collection sites and 16S rRNA gene sequence clone libraries.

The proportion of OTUs observed out of the estimated total number of OTUs in the population. As an estimate of the total number of OTUs, the Chao1 estimate of community richness (13, 15), using 3% sequence divergence, ± standard error around the estimate (14, 15), and the ACE estimate of community richness (14, 15), using 3% sequence divergence, ± standard error around the estimate. Calculations for standard errors were done as described previously (13, 15).
primers and ABI Big Dye Terminator chemistry. Sequences were examined visually, and vector and low-quality bases (20-bp window with an average PHRED score for quality of <16) were trimmed from the libraries by using LUCY (16). After trimming of these regions, the average sequence length was 827 base pairs.

Chimera detection. All clones were checked for chimeras, using both Chimera Check from the Ribosomal Database Project II (RDP II) (18) and Bellerophon (33). All sequences resembling chimeras by either program were removed from further analyses.

Determination of OTUs. Sequences were aligned separately for each clone library by using ClustalW (68) in BioEdit, version 7.0.5, with the default settings, with a gap-opening penalty of 10.0 and gap extension penalties of 0.1 and 0.2, for pairwise and multiple alignments, respectively. As points of reference, published sequences for Escherichia coli 16S rRNA (GenBank accession number L10328, base position numbers 131193 to 133733), Borellia burgdorferi (accession number X85189), and Bacillus subtilis (accession number AY553095) were used in aligning the sequences. After the alignments were performed, the sequences were truncated at the 5′ and 3′ ends, and ambiguous areas of the alignment were removed manually. The number of common bases for each was as follows: GA 441, 617 bp; Hillsborough, 440 bp; Horticulture Farm, 578 bp; Ikenberry, 371 bp; Inwood, 318 bp; Layman, 550 bp; Macom, 502 bp; Oakland, 413 bp; Raleigh, 661 bp; Thomas, 546 bp; and Woodstock, 248 bp. Distance matrices were constructed using the DNADIST program within BioEdit, version 7.0.5, using the Jukes-Cantor correction for multiple substitutions. Using the DOTUR software package (60) with the default settings, operational taxonomic unit (OTU) groupings were determined at 97% sequence identity. In further phylogenetic analyses, a consensus sequence was generated to represent each OTU.

Phylogenetic analyses of 16S rRNA gene sequences. Using the BLAST (2) and Seqmatch (18) tools, the RDP II (18) and GenBank sequence databases were screened for published sequences that closely matched the 74 consensus sequences generated to represent each of the OTUs isolated from the host populations. All analyses were performed between June 2006 and January 2007. These published sequences (see Fig. 2 and 3), along with the 74 OTU consensus sequences, were then used in a series of phylogenetic analyses designed to understand the relationship that the 16S rRNA gene sequences we recovered from D. melanogaster shared both with published 16S rRNA sequences and with each other. To begin, sequences were aligned using Clustal W (68) in BioEdit, version 7.0.5, with the default settings, with a gap-opening penalty of 10.0 and gap extension penalties of 0.1 and 0.2, for pairwise and multiple alignments, respectively. After the alignments were performed, ambiguous areas of alignment were removed using Gblocks (12), yielding 649 common bases. A distance matrix was constructed using the DNADIST program within BioEdit, version 7.0.5, using gap-opening penalty of 10.0 and gap extension penalties of 0.1 and 0.2, for pairwise and multiple alignments, respectively. After the alignments were performed, ambiguous areas of alignment were removed using Gblocks (12), yielding 649 common bases. A distance matrix was constructed using the DNADIST program within BioEdit, version 7.0.5, using the Jukes-Cantor correction for multiple substitutions. Using the DOTUR software package (60) with the default settings, operational taxonomic unit (OTU) groupings were determined at 97% sequence identity. In further phylogenetic analyses, a consensus sequence was generated to represent each OTU.

Statistical test for clinal variation in richness. To determine if there was significant clinal variation in microbial species richness, we tested for significant correlations between latitude and the Chao1 estimate of species richness and between latitude and the ACE estimate at 97% sequence identity. All data were analyzed using the Spearman nonparametric correlation test in JMP, version 5.0.1a (SAS, Cary, NC).

Statistical test for correlation between richness and climatic factors. Because latitude is not the only factor that may determine differences in microbial community richness, we also tested for a significant correlation between the Chao1 and ACE estimates of richness and several other climatic factors, namely, mean annual temperature, monthly temperature range (defined as the 12-month average of the differences between the monthly mean maximum and minimum temperatures), mean annual precipitation, and mean January low temperature. These climatic data were published by the National Oceanic and Atmospheric Administration and represent 30-year averages recorded between 1971 and 2000 at various locations across the United States (48–52). In cases where the exact collection site location was not listed in the report, the listed location that was closest to the collection site was used. All data were analyzed using the Spearman nonparametric correlation test in JMP, version 5.0.1a (SAS, Cary, NC).
OTUs formed a monophyletic group with supergroup A (Fig. 1). Two of the Wolbachia OTUs, OTUs 1 and 10, were widespread, while the remaining OTUs were limited to specific populations (Fig. 1).

The remaining 275 sequences were grouped into 64 OTUs (see Fig. 2 through 4; the supplemental material). Forty-two OTUs grouped with the Proteobacteria (15 Alphaproteobacteria, 5 Betaproteobacteria, 21 Gammaproteobacteria, and 1 Epsilonproteobacteria OTU) (Fig. 2). Seventeen of these OTUs grouped within the Firmicutes (Fig. 3). The remaining five OTUs grouped within the Bacteroidetes (Fig. 4). Bootstrap values for the neighbor-joining trees varied widely, with many OTUs assigned to clades defined by a published sequence with high bootstrap support (Fig. 2, 3, and 4).

Our analyses suggest that we have not sampled all of the taxa present in these Drosophila hosts, because the rarefaction curves do not plateau as more sequences are sampled from the library (Fig. 5). Using the Chao1 estimator of richness for comparison, the average coverage was 65.7% ± 6.06% (mean ± standard error) across all 11 libraries (Table 1). The Oakland population had the highest coverage (100%) (Table 1), whereas the Hillsborough population had the lowest coverage (26.5%) (Table 1).

Species richness varied across host locations, as measured by both the Chao1 and ACE estimators (Table 1). The two estimators of richness were positively correlated with each other (Spearman’s rho value = 0.76; P = 0.006). Despite spatial variation in richness, using either of the two estimators, there...
was no clear relationship between microbial species richness and latitude for the Chao1 (Spearman’s rho value = 0.06; \( P = 0.87 \)) or ACE (Spearman’s rho value = −0.05; \( P = 0.89 \)) estimate of species richness.

Species richness across host populations was not correlated with climate, using either of the two richness estimators. There was no correlation between the Chao1 estimate of richness and mean annual temperature (Spearman’s rho value = −0.05; \( P = 0.89 \)), monthly temperature range (Spearman’s rho value = 0; \( P = 1.0 \)), mean annual precipitation (Spearman’s rho value = −0.21; \( P = 0.56 \)), or mean January low temperature (Spearman’s rho value = 0; \( P = 1.0 \)). There was also no significant correlation between the ACE estimate of species richness and mean annual temperature (Spearman’s rho value = 0.13; \( P = 0.82 \)).

FIG. 2. Phylogenetic trees representing the taxonomic positions of proteobacterial OTUs isolated from 11 D. melanogaster host populations. (a) Alphaproteobacteria; (b) Betaproteobacteria; (c) Gammaproteobacteria; (d) Deltaproteobacteria. Phylogenies were inferred using the neighbor-joining method and were bootstrapped for 10,000 replicates. The number of bases used for analysis was (a) 670, (b) 427, (c) 253, and (d) 690. Numbers above branch points represent bootstrap values of ≥50%. Numbers below branches indicate branch lengths (nucleotide substitutions per site) of greater than zero. Trees are rooted with the 16S rRNA gene sequence for Synechococcus elongata (AF132930), a member of the phylum Cyanobacteria. Letters in parentheses to the right of each OTU indicate the D. melanogaster host populations where that OTU was observed. A key for these letters is presented in Table 1.
0.71), monthly temperature range (Spearman’s rho value = 0.02; \( P = 0.95 \)), mean annual precipitation (Spearman’s rho value = 0.20; \( P = 0.59 \)), or mean January low temperature (Spearman’s rho value = 0.11; \( P = 0.76 \)).

The frequency distribution of species abundance appeared to approximately fit a power law distribution, with a \( g \) value of −1.46. This value is not far from the range of values for \( g \) reported by Gans et al. (\( g = -1.96 \) to \(-2.11 \)) (24), although we used genera, as opposed to their use of species.

**DISCUSSION**

We used a sequence-based approach to study the microbial communities within natural host populations of *Drosophila melanogaster*. Our data suggest that there are many species of bacteria present in these *Drosophila* hosts, including a large number of *Wolbachia* species. Most of these species of bacteria were unevenly distributed among the host populations. The bacterial species richness of these microbial communities differed among host populations. However, despite significant spatial variation in microbial community richness, there was no clear relationship between latitude or climate and microbial species richness.

Our interpretation of these data comes with three caveats. First, our method of DNA extraction does not allow us to discern between bacteria associated with the inside or outside of the host. Second, because microbial DNAs were isolated from whole bodies of flies, we cannot draw conclusions about the tissue specificity of the microbes observed in these libraries. Last, we observed a large proportion of chimeric sequences in these libraries. This pattern could be due to our thorough
methods of identifying chimeras or to inappropriate concentrations of MgCl\textsubscript{2} or deoxynucleoside triphosphates in the PCR beads used to amplify the DNAs. With these three caveats in mind, we highlight several interesting characteristics of these libraries.

**Bacterial phyla present.** Three phyla—Proteobacteria, Firmicutes, and Bacteroidetes—were present in these samples. OTUs falling within the Proteobacteria phylum were represented considerably more than OTUs grouping within the Firmicutes or Bacteroidetes, even after subtracting the highly prevalent gram-negative bacterium Wolbachia. The overabundance of proteobacteria we observed could be due simply to the limitations of using lysozyme, which can be ineffective against gram-positive anaerobic cocci (22), instead of bead beating to lyse the bacterial cells. Alternatively, the overabundance of proteobacteria in these samples could be due to the ecology of the host. Our findings are consistent with work using the deer tick, *Ixodes scapularis* (6). Benson et al. (6) noticed a high prevalence of proteobacteria even after subtracting intracellular bacteria such as *Wolbachia*. They hypothesized that this pattern could be due to the humid environments that ticks prefer, which are more permissive conditions for desiccation-sensitive microbes such as proteobacteria. The prevalence of proteobacteria has also been shown for several other species of insects, including *Culicoides sonorensis*, an orbivirus vector (11); the honeybee *Apis mellifera* (34); and the ant lion, *Myrmeleon mobilis* (20). In contrast, bacterial communities associated with certain species of wood- and soil-feeding termites tend to be biased towards gram-positive microorganisms (29, 65). In *D. melanogaster*, it is possible that abiotic and biotic factors, such as climate or the availability of certain food sources, affect the proportion of proteobacteria or bacteria from other phyla in the host.

**Non-Wolbachia genera.** Aside from *Wolbachia*, the libraries contained sequences from diverse bacterial communities. Many of these sequences have not been found in a cultured organism and may represent novel genera. A phylogenetic approach was used to classify many of the OTUs isolated from the *D. melanogaster* hosts. For many of the OTUs, low bootstrap support precludes taxonomic identification. However, tentative classifications could be made for many other cases, and some suggest potentially interesting host-bacterium inter-
actions taking place in this system that might be studied more rigorously in future experiments.

Most of the OTUs isolated from *D. melanogaster* host populations belonged to four classes of the *Proteobacteria*, a diverse phylum containing upwards of 460 genera and 1,619 species (36). These OTUs were located primarily within the *Alpha*-, and *Gammaproteobacteria*, a feature consistent with both their ubiquity in nature (36) and their presence in many species of insect hosts (6, 11, 34, 65). Many OTUs were highly similar to taxa that interact with animals and plants in interesting ways, suggesting a possible role for these bacteria in the ecology of their *D. melanogaster* hosts and for *D. melanogaster* in mediating interactions between these bacteria and alternative hosts. For example, within the *Alphaproteobacteria*, many of the OTUs grouped closely within the *Glucosacetobacter* genus. Species in this genus are found primarily in sugary, acidic, and alcoholic habitats, such as flowers, fruits, plant tissues, and plant rhizospheres (37), and associate with insects such as the pink sugarcane mealybug (37) and honeybees (34). Three OTUs in the *Betaproteobacteria* grouped closely with published sequences from the genera *Acidovorax* and *Borde-tella*, which are implicated in both plant (5) and animal (72) diseases. OTU 18 closely matched a gammaproteobacterium isolated from the intestine of the honeybee and could represent a generalist capable of cross-species horizontal transmission. OTUs 51, 68, and 8 grouped closely with the insect pathogens *Providencia rettgeri* (33) and *Pseudomonas fluorescens* (40, 62) and could therefore be pathogenic to *D. melanogaster* in nature.

Seventeen of the 74 OTUs were members of the gram-positive *Firmicutes* phylum. Many of these OTUs showed high similarity with published sequences from the *Lecunostoc* genus. Since some members of the *Lecunostoc* genus ferment fructose (43), these OTUs may play a role in host digestion of fruits or other plant materials or may live communally in the host gut. OTU 58 was highly similar to the published sequences for members of the *Lactobacillus* genus. Although members of this genus are sometimes pathogenic to plants and animals (27), others are part of the normal nonpathogenic floras of plants, insects, and vertebrates (27) and have been shown to increase the life span in laboratory strains of *Drosophila* (10).

**Wolbachia.** One of the more striking characteristics of the 16S rRNA clone libraries was the large number of *Wolbachia* sequences. Because PCR can be biased, the frequency of Wolbachia sequences we observed cannot be a direct measure of the frequency of *Wolbachia* species in nature. However, the relatively strong bias towards *Wolbachia* species in our libraries probably reflects an abundance of these microbes relative to other bacteria. Although the exact frequency of *Wolbachia* infections in natural *D. melanogaster* populations is unclear and likely variable, it is estimated that approximately 30% to 75% of the *D. melanogaster* stocks housed at *Drosophila* stock centers are infected with this intracellular parasite (17, 46). Furthermore, preliminary studies in our lab suggest that 55 to 60% of wild-caught *D. melanogaster* isofemale lines, including lines derived from the populations used in this study, are infected with *Wolbachia* (V. Corby-Harris, unpublished data). Studies of other arthropods suggest that within-species infection rates range from 2 to 83% for *Solenopsis invicta* (66), 5 to 100% for *Acruea* species (35), and 25 to 100% for fig wasps (67). When sequences were grouped based on 3% sequence divergence, the *Wolbachia* and *Wolbachia*-like sequences were grouped into 10 distinct OTUs that were unevenly distributed across host locations. This pattern suggests distinct species or lineages of *Wolbachia* across host locations, a pattern similar to that demonstrated for *Wolbachia* in geographically distinct populations of *S. invicta* (1).

To understand how the *Wolbachia* sequences we isolated were related to each other and to published *Wolbachia* 16S rRNA gene sequences, we constructed a phylogeny consisting only of *Wolbachia*-like gene sequences we isolated and published *Wolbachia* 16S rRNA gene sequences. The phylogeny we constructed agrees qualitatively with previous work (44) in that the published sequences formed five distinct supergroups (A through F). However, the bootstrap support values varied widely, with many below 70%, and the relationships of the supergroups to each other were inconsistent with those in previous studies (44, 46). Such inconsistencies could be due to the slowly evolving nature of the 16S rRNA molecule, which may not provide adequate resolution between the clades (44). Indeed, Lo et al. (44) suggested that more rapidly evolving sequences, such as ftsZ, are more appropriate for understanding the phylogenetics of the *Wolbachia* genus. Nonetheless, the phylogeny we constructed suggests that there are two distinct lineages of *Wolbachia* in natural populations of *D. melano-gaster* hosts. Both of these lineages appear to be monophyletic with published sequences from supergroup A, which is found in a variety of arthropod hosts, including drosophilids, *Tri-bolium*, and *Nasonia* (44). This monophyly is consistent with one or a few origins of *Wolbachia* in *D. melanogaster* hosts found in nature.

**Species richness and composition.** Aside from identifying species of bacteria associated with natural populations of *D. melanogaster*, we also aimed to characterize the richness of these microbial communities overall. Here it is important to point out that we sampled only five flies per location (pooled into one sample), and our estimates of species richness and composition could change with increased sampling effort. In addition, our ability to identify trends based on these data is weakened by the lack of multiple samples from each collection site. With these limitations in mind, however, there are none-theless some intriguing features of the communities that were sampled.

The shapes of the rarefaction curves suggest that the taxonomically diverse populations present in these 11 host popula-tions have not been sampled completely. Many taxa were found only once or twice. The fact that we were able to uncover these rare taxa with this sampling effort suggests that many unidentified and ecologically important species of bacteria living in natural populations of *D. melanogaster* have yet to be identified.

We observed that microbial species richness varied across host populations. Motivated by previous studies of plants and animals that demonstrated a negative correlation between species richness and latitude (9, 58), we sought to explain this variation in richness among locations by using latitudinal data from each location. In addition, because latitude is only one of the many characteristics of a geographic location, we also tested for a significant relationship between climatic factors and richness. While microbial species richness varied across
the host locations that we sampled, there was no evidence of a relationship between richness and latitude. The lack of a clear relationship between richness and latitude in these 11 host populations could simply be due to the fact that the range we sampled (approximately 10 degrees latitude) was not large enough to observe an effect. There was also no evidence for a significant correlation between richness and climatic factors, such as mean annual temperature, monthly temperature range, mean annual precipitation, and mean January low temperature, using either the Chao1 or ACE richness estimator. One reason for these nonsignificant results is that bacterial populations are responding to climatic factors over relatively short time scales or within microenvironments that the 30-year climate averages cannot adequately represent. Alternatively, since many biotic and abiotic factors account for the distributions of microbes in the environment (30), it is possible that other unidentified and ecologically important factors, or interactions between these factors, may account for the differences in richness that we observed across populations. Factors such as the types of fruit present in these habitats, host genetic structure, or the presence of alternative insect hosts could account for such variation and need to be studied empirically.

Cox and Gilmore recently completed a survey of bacteria isolated from wild and laboratory-reared D. melanogaster flies (19). Several features of their study agree with the data presented here. First, they showed that wild D. melanogaster flies harbor a wide range of bacterial species from the Proteobacteria, Firmicutes, and Bacteroidetes phyla (19). Additionally, although the Gammaproteobacteria are the most diverse group isolated from flies, the Alphaproteobacteria are the most abundant, after the highly prevalent Wolbachia species are excluded (19). Last, Cox and Gilmore’s estimates of species richness (19) fall well within the range of species richness estimates in the present study. Together with Cox and Gilmore’s work, the data presented here move us towards a better understanding of the bacteria interacting with Drosophila.

Our study sheds light on the composition and richness of microbial communities present in natural populations of D. melanogaster hosts and highlights several important features of these communities. Although our findings have important consequences for how researchers understand the ecology and evolution of Drosophila hosts in nature and the dynamics of insect-associated bacterial communities, more work must be done to explicitly test hypotheses regarding the nature of the interactions between Drosophila and the microbes we identified. Fortunately, D. melanogaster is an experimentally tractable model organism that lends itself beautifully to such in-depth studies.

ACKNOWLEDGMENTS

We thank Judith Man, DeEtte Walker, Justine Lyons, Mary Ann Moran, Chih-Horng Kuo, and Chris Lasher for technical assistance, Tracy Leskey, H. T. Melvin’s Place, the Sokol family, Cynthia Marston, the Grove family, and Robert Cheves for access to collection sites, and members of the Promislov lab for comments on a previous version of the manuscript.

This work was supported by an NSF DDIG grant (DEB-0508785) to V.C.-H., an Alton Fellowship in Genetics to V.C.-H., a Senior Scholar Award to D.E.L.P., from the Ellison Medical Foundation, and the Georgia Research Alliance Wired Professorship to J.L.B.

REFERENCES

BACTERIA ASSOCIATED WITH DROSOPHILA MELANOGASTER


