THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BLOWFLY METAGENOMES AS A TOOL TO ASSESS THE PRESENCE OF MICROBIAL PATHOGENS ON FARMS

CAYLIE HAKE
SPRING 2013

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Animal Sciences
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Stephan Schuster
Principal Investigator, Schuster Lab
Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Ming Tien
Professor of Biochemistry
Honors Adviser in the Department of Biochemistry and Molecular Biology

Scott Selleck
Professor and Head,
Department of Biochemistry and Molecular Biology

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Flies can act as vectors for various diseases, and the close association between flies, livestock and poultry, and humans warrants closer study to develop an understanding of disease transmission. In this study, we aimed to identify bacterial species carried by flies collected from several animal facilities local to Penn State University. We used Illumina MiSeq and HiSeq sequencers to determine the fly species and the entire metagenome of the samples. After determining which bacterial species were most prominent, we chose four to investigate further using MLST schemes. We successfully amplified genes from *Acinetobacter baumannii*, *Helicobacter cinaedi*, *Proteus mirabilis*, and *Escherichia coli* directly from the original DNA samples. With sequences from the PCRs, we were able to verify the presence of these pathogenic species in some samples and in other samples verified the genera. Methods such as those used in this study could prove beneficial to agriculture, veterinary medicine, and public health as tools to determine the best route toward disease prevention.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thanks to Dr. Schuster, who welcomed me into the lab in the spring of 2011 and has been encouraging me ever since, and to the entire Schuster Lab team, who taught me about the technical side of research and the newest technologies available. Special thanks to Dr. Bodo Linz, who possesses incredible patience, without which I probably wouldn't have come this far.

Chapter 1

Introduction and Background

Introduction to the importance of microbial pathogens in the livestock industry

In recent years, with the increased efficiency of DNA and RNA sequencing technologies, scientists have been able to embark on metagenomic studies, such as our present study of the microbiome within and on fly species in the presence of livestock and poultry.

Determining the microbial species present could be significant in both veterinary and human medicine. Zoonoses, or zoonotic diseases, are transferred from a non-human to a human host, and some cause serious illnesses in animals and humans. Non-zoonotic disease pathogens may survive in favorable environments surrounding animals and infect humans who come into contact. Moreover, both zoonotic and non-zoonotic diseases impact human populations economically by negatively affecting livestock production, food production, and athletic and labor performance. Such diseases may be caused by viruses, prions, fungi, bacteria, protozoans, and other parasites. At this time, our work focuses on bacterial pathogens that could be carried from one host to the next or from the environment to humans by flies that thrive around several livestock species. We include deer in this study as a representative wildlife species that is prominent in the region as well as to incorporate an increasingly important farmed food animal.

Human travel and trade play an important role in the spread of pathogens and their vectors or original hosts (Kruse et al., 2004). West Nile virus, a recently-emerged zoonosis that originated in wild birds, was introduced in the United States in 1999 and now causes disease in humans and equines. Kruse also points out bovine tuberculosis as an example of a bacterial disease that spread to wildlife species after human-mediated movement of infected cattle. Now,

Mycobacterium bovis resides in many wildlife host species and is therefore able to continue moving to new areas and threatening human and animal health (Kruse et al., 2004).

To improve the effectiveness of disease control and prevention methods, disease vectors must be identified and transmission mechanisms understood. In the northeastern U.S., for instance, Lyme disease control methods are still widely debated, despite our present understanding of *Borrelia burgdorferi* transmission from animal host to tick vector to human host (Wood and Lafferty, 2013). Thus, research in the area of disease transmission and vector ecology is extremely important and can be advanced with today's technologies.

Flies are at the center of much scrutiny as likely vectors of many disease-causing microorganisms that affect economically-significant livestock. For example, *Campylobacter fetus* infects sheep and cattle as well as humans via opportunistic infections. Two subspecies have been described – *C. fetus* subsp. *fetus* (*Cff*) and *C. fetus* subsp. *venerealis* (*Cfv*). *Cff* is a normal part of ovine and bovine intestinal tract microflora, but when ingested, it can cause abortions in both sheep and cattle. In fact, *Cff* is considered the leading cause of abortions in sheep in New Zealand, where in 2012, sheep meat accounted for over 10% of the nation's agricultural production (Mannering et al., 2006; New Zealand Government, 2013). *Cfv* thrives in the bovine reproductive tract, where it can be passed from one animal to the next during natural breeding. This bacterium can cause bovine genital campylobacteriosis, which in turn leads to infertility, early embryonic loss, or abortion (Zhao et al., 2010). The bacteria cause septic abortion in human hosts, as well (Steinkraus and Wright, 1994).

Salmonella enterica and *Escherichia coli* receive a lot of attention in the field of food safety, and thus relate to animal production, management, and overall health. According to three studies conducted by the USDA National Animal Health Monitoring System, *Salmonella* identified in bovine fecal samples increased over time. In 1996, only 20% of dairy herds had

Salmonella-positive cows, but in 2007, 39.7% of the operations had *Salmonella*-positive cattle (USDA, 2009).

Escherichia coli is a constituent of the normal human gut microflora, as it is in many animals. A strain of particular significance to global health is *E. coli* O157, renowned as a foodborne pathogen. However, this strain is also infectious through environmental exposure such as contact with animals, especially cattle, and this bacterium is thus earning a name as a zoonotic pathogen (Chase-Topping et al., 2008). It can experimentally infect chickens, but studies of chickens in production have isolated *E. coli* O157 only rarely (Esteban et al., 2007).

As carriers of many bacterial species, flies and other insects may prove indispensable in disease prevention as a way to detect pathogens in the environment. Detection is the first step toward choosing how to limit disease risk. In production animals today, antibiotics are administered regularly for both disease prevention and animal growth, but concerns about antibiotic resistance increase with our knowledge of the pathogens around us. With flies indicating the presence of certain pathogenic species, more selective measures for antibiotic administration can be implemented.

Working toward improved animal and public health, we gathered flies (mostly blowfly species but also houseflies, flesh flies, and tachina flies) from State College area animal facilities in the hopes that we could determine the microbial pathogens carried by the flies. We hypothesized that the flies associated with different livestock species and poultry would carry different microbes in their gut and on their bodies.

About blowflies and myiasis

Blowfly species have a global distribution and varying economic significance as well as human and animal health impacts. For example, the blowflies *Lucilia sericata* and *Lucilia*

cuprina are major problems in Australia, where sheep are an important economic resource. Myiasis, breech flystrike in particular, has been a management problem for years, and persists today as veterinarians, farmers, and scientists search for better prevention methods. Breech flystrike occurs when a blowfly deposits eggs in the soiled area near a sheep's tail. When the eggs hatch, the larvae burrow into the tissue, leaving behind them a lesion that welcomes infectious bacteria. Blowflies can cause myiasis in humans, as well, and is in fact a significant problem in some developing regions and in individuals of low socioeconomic status or poor hygiene (Francesconi and Lupi, 2012).

Blowflies and disease transmission

Past studies have found that blowflies are certainly suspect in transmission of many pathogenic bacteria, some of which affect humans. Blowflies have been found to be more effective carriers of enteric bacteria than other types of flies, according to a 1998 study of flies in Greyhound dog kennels in Kansas. 63% of the blowflies collected from the kennels contained *Proteus*, *Salmonella*, *Pseudomonas*, or *Providentia* species (Urban and Broce, 1998). Similarly, research indicates that the blowfly *C. megacephala* is significantly more likely than the house fly *M. domestica* to carry bacteria (Sukontason et al., 2007). Within blowfly species, body size and local environment quality seem to play an important role in the ability to carry and transmit pathogens (Maldonado and Centeno, 2003). Species *Calliphora vicina* and *Lucilia sericata* have been shown to harbor pathogenic mycobacteria on their bodies which they can subsequently transmit to other animals (Fischer et al., 2004a).

The past and present of metagenomic sequencing

Microbiome studies have improved our understanding of the world's ecosystems and, in some cases, the flora within us. For example, in a study of the microbial populations of various soil samples, meant to determine the main ammonia-oxidizing species, DNA and RNA were extracted from pooled soil samples and then used for PCRs. Successful results indicated that crenarchaeota may be the most abundant ammonia-oxidizing organisms in soil ecosystems (Leininger et al., 2006). Similarly, environmental samples from the Sargasso Sea were collected and the whole genome was sequenced to shed some light on the biology of open-water microorganisms (Venter et al., 2004). Another study found that bacterial species acquired during birth depended on whether a child was delivered vaginally or by cesarean section (Dominguez-Bello et al., 2010). Metagenomic sequencing has a seemingly endless range of potential uses, each of which offer us a clearer perspective of the invisible bacterial world around us.

Chapter 2

Materials and Methods

Fly collection

Over the course of several months, we collected blowflies from multiple Penn State animal facilities the Poultry Education and Research Center, Dairy Barns, Horse Barns, Deer Research Center, and Swine Center as well as from the sheep farm a short distance from campus at Spring Creek Park. For greater geographic variation, we also collected flies from a pasture near Meyer Dairy.

Equipment included large insect nets and wide-mouth collection jars with screw-on lids, which contained pieces of rotting salmon head to attract blowflies. Caught flies were transferred from the net to 50 mL Falcon tubes containing a small amount of dry ice that immobilized and quickly killed them. The rapid freezing of the flies helped preserve DNA and RNA quality. The flies were subsequently transferred to individual 2 mL tubes and stored on dry ice throughout the sampling process and at -80°C in the laboratory. Care was taken during collection and transfer from tube to tube that flies remained intact (i.e. lost no legs or antennae).

Macrophotography of flies

For many of our original fly specimens, we took photographs to document the morphology and to send to an expert for visual identification of the species. After taking photographs of individual flies, each fly was immediately used for DNA extraction.

DNA extraction and construction of Illumina sequencing libraries

For DNA extraction from the samples, we used the Qiagen DNeasy Blood and Tissue Kit. We chose 12 flies from each sampling location from which to extract DNA. We chose those that looked like blowflies *for the most part, but in the cases where we needed to use non-blowfly species, we took ones that seemed morphologically diverse* as well as intact. Each fly was transferred to a clean microcentrifuge tube and ground thoroughly using a certified RNase, DNase, and DNA-free pestle for 1.5 mL microcentrifuge tubes. Lysis buffer and proteinase K were added to the tubes, and after 10 min. incubation at 56°C and regular vortexing, our samples were incubated at 56°C overnight. The following day, the samples were transferred to the columns, and the DNA was isolated according to the manufacturer's protocol. We collected two separate DNA elutions, which were stored at -20°C. The first elution was subsequently used to make libraries for Illumina MiSeq and HiSeq sequencing. The second elution was stored for PCR investigation (see below).

DNA sequencing and analysis

Genomic DNA extracted from 12 flies from each of the six sampling locations was chosen for DNA sequencing. Illumina paired-end libraries [150 bp × 150 bp] were constructed, and pooled libraries were sequenced on Illumina MiSeq as well as on Illumina HiSeq machines. The sequence reads from Illumina MiSeq runs were used to validate the quality of the libraries and also to extract the complete mtDNA genomes of the flies. To this end, the reads were aligned against the mtDNA reference genome of the screwworm fly *Chrysomya putoria* and subsequently assembled. We then used the sequence of the cytochrome oxidase 1 (CO1) gene of the

assemblies to determine the fly species using BLAST searches (blastn). As a control, the complete mtDNA genomes were used in BLAST searches.

MiSeq returned about 600,000 reads per fly, except for one fly that was overrepresented, and HiSeq returned about 50 million reads per fly. To begin analysis of the flies' metagenomes, the total data per fly was used in blastx, which translates the sequence in all six reading frames and compares the resulting proteins against the nr protein database. The results of BLAST searches were then analyzed in MEGAN 4 (MEtaGenome ANalyzer) (Huson et al., 2011). The total number of reads was normalized across all of the fly samples so that the number of bacterial reads in our analysis indeed reflected the presence of bacteria instead of being biased by differences in the number of total reads in each fly.

Choosing bacterial species of interest

From the 50 most abundant bacterial species in our flies, we chose the following four species for further investigation: *Proteus mirabilis*, *Acinetobacter baumannii, Escherichia coli*, and *Helicobacter cinaedi*. For *A. baumannii*, *E. coli*, and *H. cinaedi*, MLST schemes were already created, and the available primer sequences were modified to prevent PCR amplification from distantly related species. No MLST scheme was available for *P. mirabilis*, so primers for four genes were designed by Bodo Linz (Appendix B).

PCR protocol and PCR product analysis

We diluted the DNA samples 1:10. For each PCR, we used 2 µL DNA, 1 µL each of the primers (concentration 10 pmol/µL), 2 µL of 2 mmol dNTPs, buffer with MgOAc, and 0.4 µL rTth DNA polymerase (Applied Biosystems). The PCR conditions were as follows:

Initial denaturation at 94°C for 5 min

Denaturation at 94°C for 15s
Annealing at 58°C for 15s
Elongation at 68°C for 45s
} 38 cycles

Final elongation at 68°C for 5min

PCR products were analyzed on a 1% agarose gel and visualized under UV light using ethidium bromide. PCR amplicons of the correct size were purified using the Qiagen QIAquick PCR Purification Kit (protocol in Appendix A) and submitted for sequencing at the Penn State Nucleic Acid Facility, University Park, PA. The sequencing reads were analyzed, evaluated, trimmed using Staden Package (Staden et al., 1998).

Chapter 3

Results

Sampling

We sampled flies on several days between June 19 and September 16, 2012, using insect nets and rotting salmon head as bait to attract blowflies. Our first collection visit was dedicated to testing and optimizing the fly-catching methods. The flies that were attracted to the bait were caught in insect nets and then transferred to 50 mL tubes which contained a small piece of dry ice to immobilize and kill them. Then each of the flies was placed into individual 2 mL tubes and stored on dry ice. During this first expedition on June 19, 2012, we collected just 16 flies from the horse barn (Table 1, Fig. 1) and returned for another collection in September, when we collected 50 additional samples. This facility is surrounded by about 50 acres of pastures where the horses spend the majority of their time.

The Swine Center, about 1.74 km from the horse barn, is located on many acres of pasture that are separated into paddocks where the boars are held individually and the sows are held in small groups. Here, we also made two samplings, one near a boar and one near sows, gathering a total of 79 flies. The Penn State sheep operation is combined with beef cattle, so to collect flies that associated with sheep alone, we chose to go farther away from all of the other animal facilities, thereby adding some geographic variety. We collected 84 flies from a sheep farm near Spring Creek Park, about 2.5 km from the horse barn and 1.75 km from the Swine Center (Table 1, Fig. 1).

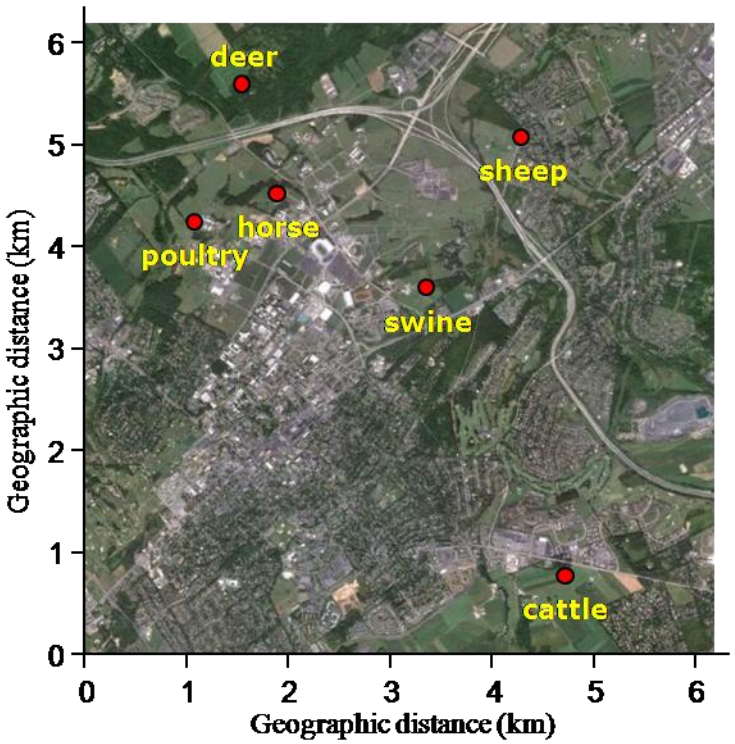

Figure 1. Geographic distribution of the sampling locations.

Table 1. Fly collection.

Animal host	No of flies	Date of collection
horse	16	6/19/2012
swine	23	6/21/2012
cattle	21	7/15/2012
deer	46	9/12/2012
poultry (inside barn)	30	9/12/2012
poultry (outside barn)	18	9/12/2012
cattle	40	9/14/2012
sheep	84	9/16/2012
horse	50	9/16/2012
swine	56	9/16/2012

To further increase geographic variety, we chose to collect our cattle samples near a farm in the vicinity of Meyer Dairy. With a distance of 3.2 to 5.8 km from any other animal facility, this location was the most geographically separated (Fig. 1). Again, the animals were held on

large pastures. In addition to the 21 flies we collected here, we collected 40 flies on a later trip (Table1, Fig. 1).

The 46 fly samples from deer were caught at the Deer Research Center, which is centrally located on 22 wooded acres, in contrast to the facilities on pastures. Since this facility is located farthest north and the deer are grouped in forested outdoor paddocks, we assumed that there was no or only limited exchange of flies between the deer facility and the other sampling points (Table 1, Fig. 1).

The poultry buildings are fully enclosed, except for ventilation, and the only flies present under the birds' cages were black flies, later found to be houseflies. Therefore, we also sampled flies outside of the buildings, even though their contact to the poultry was limited at best. Still, we collected 30 flies from inside and 18 from outside the building (Table 1, Fig.1).

DNA was extracted from 21 flies from the Swine Center (S), 16 from the Horse Barns (H), 18 from outside the poultry barn (PO), 12 from inside the poultry barn (PI), 32 from cattle (C), 21 from the Deer Research Center (D), and 20 from the sheep farm near Spring Creek Park (O). The samples with the highest DNA concentration (Table C1) were used for DNA library construction for sequencing. In addition, we aimed for maximal biodiversity and took into consideration the identified species, determined from the photographs.

Metagenome analysis

The sequencing results from the Illumina MiSeq and HiSeq machines contained information about the fly species as well as information about microbes associated with the flies. The MiSeq sequence reads were used to assemble the mtDNA genome of each fly in order to gain insight about the intra- and interspecies diversity of blowflies. Using the mitochondrial genome of the screwworm fly, *Chrysomya putoria*, as a reference, we extracted the respective mtDNA

reads of each fly and assembled the individual mtDNA genomes. We then used the CO1 gene sequence, coding for cytochrome oxidase 1, to determine the fly species. Of the 72 flies, most were *Phormia regina* (39), *Lucilia sericata* (17), and *Musca domestica* (4), but a few samples were *Stomoxys calcitrans* (2), *Winthemia rufoptica* (1), *Cochliomyia macellaria* (1), *Boettcheria bisetosa* (1), *Boettcheria latistema* (1), *Eudasyphora canadiana* (1), *Lucilia coeruleiviridis* (2), *Lucilia illustris* (1), *Muscina levida* (1), and *Hydrotaea* sp. (1). *Phormia regina*, *Lucilia sericata*, *Lucilia coeruleiviridis*, and *Cochliomyia macellaria* are all blowflies (Calliphoridae); *Stomoxys calcitrans*, *Eudasyphora canadiana*, *Muscina levida*, *Hydrotaea* sp., and *Musca domestica* are house flies (Muscidae); *Winthemia rufoptica* is a member of the Tachinidae family, which makes up part of the true flies; *Boettcheria bisetosa* and *Boettcheria latistema* are flesh flies (Sarcophagidae) (Fig. 2, Table C2).

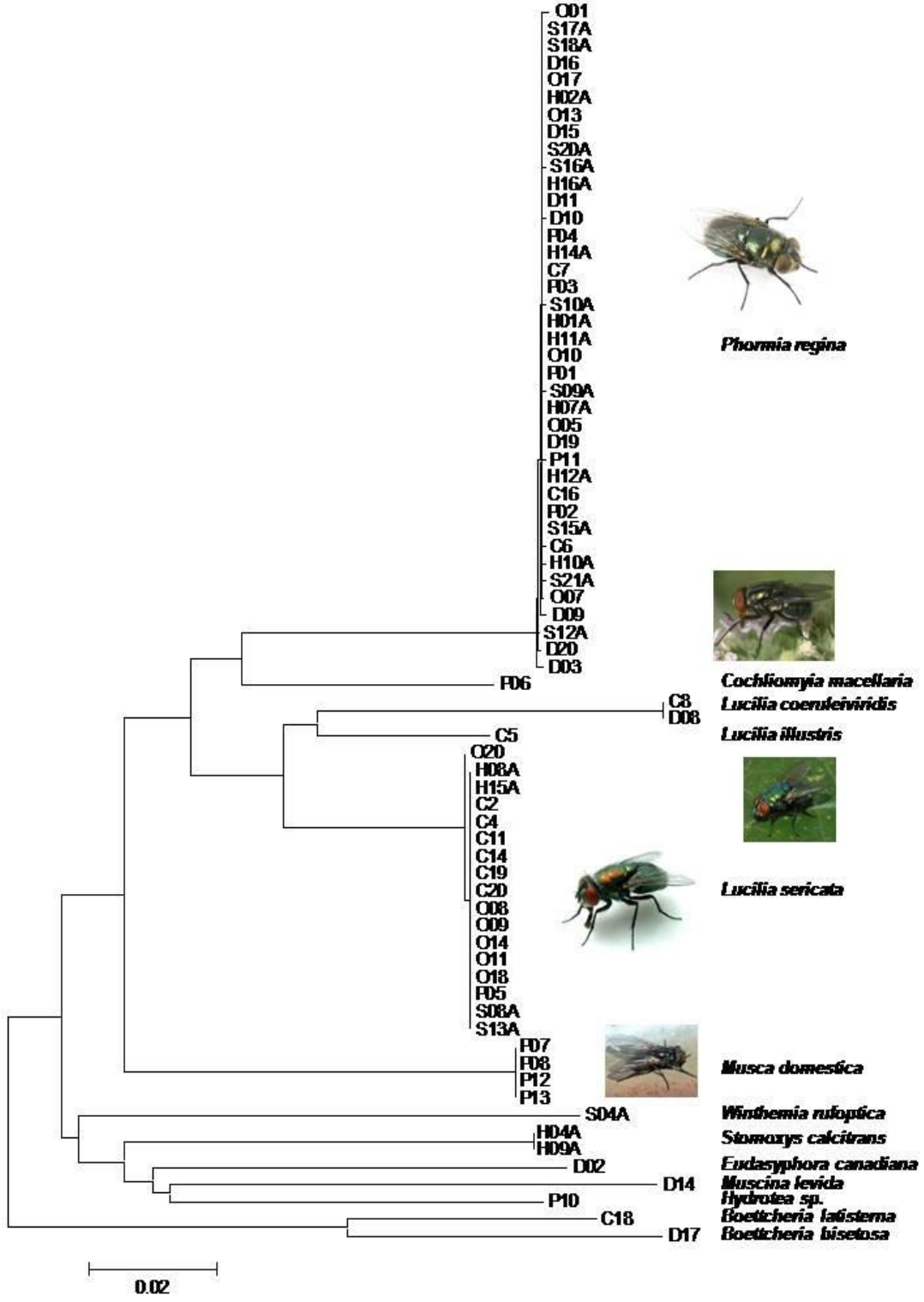

Figure 2. Phylogenetic tree of 72 flies based on the mitochondrial CO1 sequence.

The combined sequence reads from the MiSeq and HiSeq sequencer runs were analyzed using blastx to identify the bacterial species associated with the flies. 69 of the 72 flies were successfully sequenced on the HiSeq sequencer and were used for further analysis. The blastx output was subsequently examined in MEGAN, a computer program designed to interpret and visualize metagenome analyses. The number of identified bacterial reads was normalized against the total number of reads to ensure comparability between individual flies.

We identified 50 bacterial species that were commonly present in many of the fly samples (Fig. 3). The bacteria represented a variety of different lifestyles, from symbionts (light green in Fig. 3) and commensal bacteria (brown) to plant and animal pathogens (dark blue and red, respectively), as well as environmental bacteria (turquoise). While some bacteria such as *Wolbachia* and *Escherichia coli* were identified in almost all samples, others were found in very few samples. For example, *Helicobacter cinaedi* was found in only three fly samples, two collected from the sheep farm and one from the Swine Center. Various *Wolbachia* species are known to be endosymbionts in other insects, suggesting that flies might also harbor *Wolbachia* endosymbionts.

The bacterial load differed significantly between the individual flies, from almost no bacteria, such as in blowfly sample O20 (*Lucilia sericata*) from sheep, to high bacterial loads of many species as well as many reads per species, such as blowfly sample H02 (*Phormia regina*) from horses. Overall, the bacterial species present in the blowfly samples varied with mammalian host. Bacteria were most populous within or on the blowflies collected at the horse barn, and the most abundant species included *Myroides odoratimimus, E. coli*, and *Salmonella enterica*, as well as numerous *Pseudomonas* species: *P. mendocina, P. entomophila, P. aeruginosa, P. chlororaphis, P. stutzeri, P. putida*, and *P. fluorescens*. *Acinetobacter* species were also prominent: *A. baumannii, A. johnsonii*, and *A. lwoffii*. The samples also contained a significant amount of plant pathogens that were also very frequent among the deer fly samples, namely

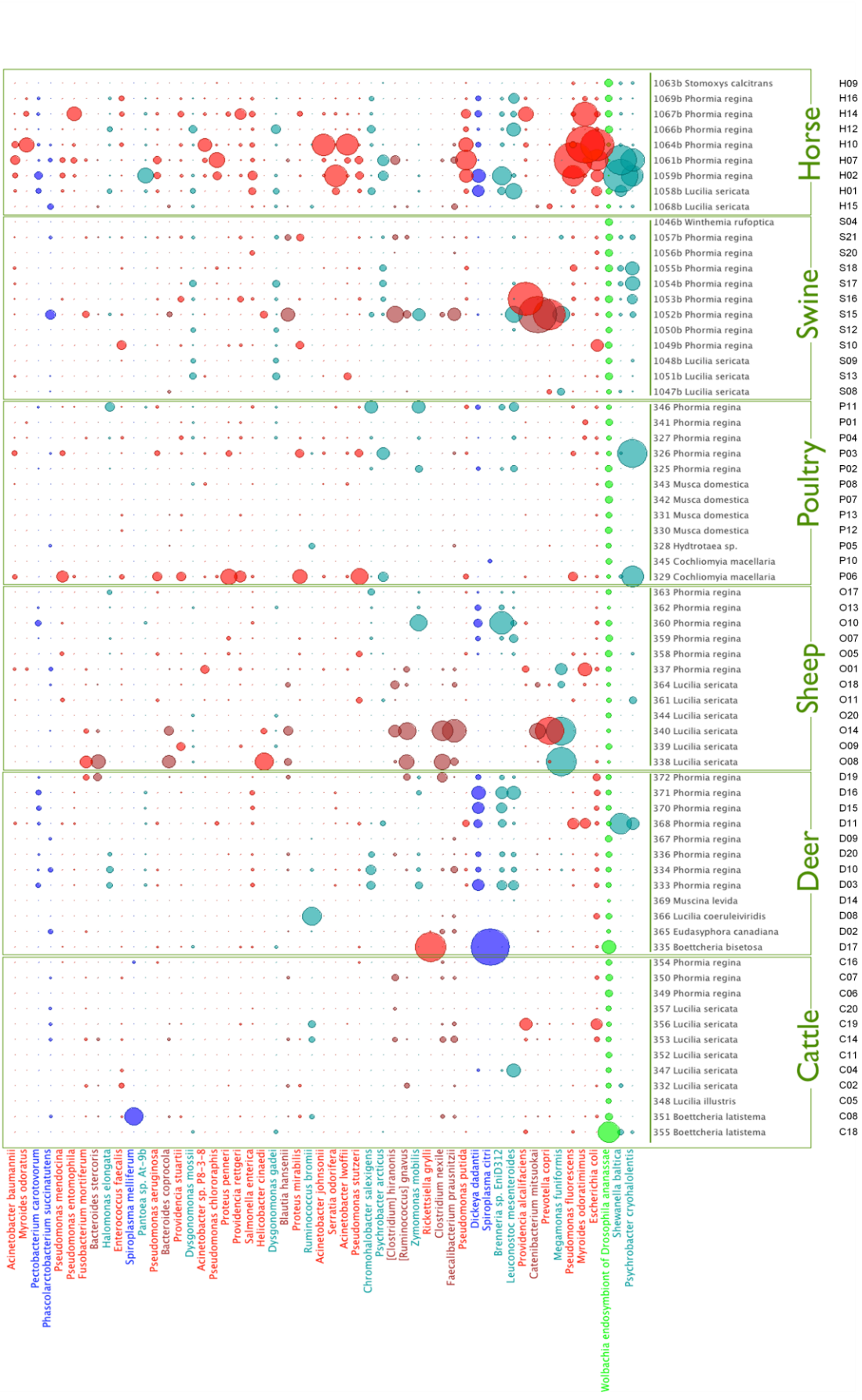
Horse
1063b Stomoxys calcitrans H09
1069b Phormia regina H16
1067b Phormia regina H14
1066b Phormia regina H12
1064b Phormia regina H10
1061b Phormia regina H07
1059b Phormia regina H02
1058b Lucilia sericata H01
1068b Lucilia sericata H15
Swine
1046b Winthemia rufoptica S04
1057b Phormia regina S21
1056b Phormia regina S20
1055b Phormia regina S18
1054b Phormia regina S17
1053b Phormia regina S16
1052b Phormia regina S15
1050b Phormia regina S12
1049b Phormia regina S10
1048b Lucilia sericata S09
1051b Lucilia sericata S13
1047b Lucilia sericata S08
Poultry
346 Phormia regina P11
341 Phormia regina P01
327 Phormia regina P04
326 Phormia regina P03
325 Phormia regina P02
343 Musca domestica P08
342 Musca domestica P07
331 Musca domestica P13
330 Musca domestica P12
328 Hydrotaea sp. P05
345 Cochliomyia macellaria P10
329 Cochliomyia macellaria P06
Sheep
363 Phormia regina O17
362 Phormia regina O13
360 Phormia regina O10
359 Phormia regina O07
358 Phormia regina O05
337 Phormia regina O01
364 Lucilia sericata O18
361 Lucilia sericata O11
344 Lucilia sericata O20
340 Lucilia sericata O14
339 Lucilia sericata O09
338 Lucilia sericata O08
Deer
372 Phormia regina D19
371 Phormia regina D16
370 Phormia regina D15
368 Phormia regina D11
367 Phormia regina D09
336 Phormia regina D20
334 Phormia regina D10
333 Phormia regina D03
369 Muscina levida D14
366 Lucilia coeruleiviridis D08
365 Eudasyphora canadiana D02
335 Boettcheria bisetosa D17
Cattle
354 Phormia regina C16
350 Phormia regina C07
349 Phormia regina C06
357 Lucilia sericata C20
356 Lucilia sericata C19
353 Lucilia sericata C14
352 Lucilia sericata C11
347 Lucilia sericata C04
332 Lucilia sericata C02
348 Lucilia illustris C05
351 Boettcheria latistema C08
355 Boettcheria latistema C18
Acinetobacter baumannii
Myroides odoratus
Pectobacterium carotovorum
Phascolarctobacterium succinatutens
Pseudomonas mendocina
Pseudomonas entomophila
Fusobacterium mortiferum
Bacteroides stercoris
Halomonas elongata
Enterococcus faecalis
Spiroplasma melliferum
Pantoea sp. At-9b
Pseudomonas aeruginosa
Bacteroides coprocola
Providencia stuartii
Dysgonomonas mossii
Acinetobacter sp. P8-3-8
Pseudomonas chlororaphis
Proteus penneri
Providencia rettgeri
Salmonella enterica
Helicobacter cinaedi
Dysgonomonas gadei
Blautia hansenii
Proteus mirabilis
Ruminococcus bromii
Acinetobacter johnsonii
Serratia odorifera
Acinetobacter lwoffii
Pseudomonas stutzeri
Chromohalobacter salexigens
Psychrobacter arcticus
[Clostridium] hiranonis
[Ruminococcus] gnavus
Zymomonas mobilis
Rickettsiella grylli
Clostridium nexile
Faecalibacterium prausnitzii
Pseudomonas putida
Dickeya dadantii
Spiroplasma citri
Brennaria sp. EniD312
Leuconostoc mesenteroides
Providencia alcalifaciens
Catenibacterium mitsuokai
Prevotella copri
Megamonas funiformis
Pseudomonas fluorescens
Myroides odoratimimus
Escherichia coli
Wolbachia endosymbiont of Drosophila ananassae
Shewanella baltica
Psychrobacter cryohalolentis

Figure 3. Top fifty bacterial species identified in the fly samples. The fly samples are ordered by sampling location and fly species. Light green – symbionts, brown – commensal bacteria, turquoise – environmental bacteria, dark blue – plant pathogens, red – animal pathogens.

Dickeya dadantii, *Pectobacterium carotovorum*, and *Phascolarctobacterium succinatutens*. Several environmental bacteria were also frequent in the fly samples from deer; however, the number of potential animal pathogens was very low in those samples, with the exception of *E. coli* and *Salmonella enterica*. The samples from the Deer Research Center showed greater uniformity than the samples from other locations and form a pattern that is unique to these flies. Besides the presence of plant pathogens and the virtual lack of animal pathogens, certain environmental bacteria were frequently found in most samples. For example, *Brenneria* sp. and *Leuconostoc mesenteroides* were uniformly present in the *Phormia regina* samples associated with deer.

Few bacteria were identified in flies associated with cattle. Most of the flies collected from sheep, swine, and poultry facilities contained intermediate amounts of bacteria, with the exception of two *Lucilia sericata* from sheep, one *Phormia regina* from swine, and one *Cochliomyia macellaria* from poultry, all of which contained a large amount of bacteria. The latter contained a particularly large number of animal pathogens, whereas the two from sheep contained mostly commensals with the exception of *H. cinaedi*, *Fusobacterium mortiferum*, and *Prevotella copri*.

Metagenome differences between fly families

There was a remarkable difference between the bacterial loads of the different fly families. While the 57 samples of true blowflies (genera *Lucilia* and *Phormia* of family Calliphoridae) carried a substantial amount and variety of bacteria, the 8 true house flies (family

Muscidae, genera *Musca*, *Muscina*, *Eudasyphora*, *Stomoxys*, and *Hydrotaea*) carried very few bacteria. Likewise, the bacterial load of the flesh flies of genus *Boettcheria* was low with the exception of D17 from deer. That sample contained a large number of reads from the plant pathogens *Spiroplasma citri* and *Rickettsiella grylli*, an intracellular pathogen of aquatic and terrestrial arthropods (Leclerque, 2008).

Validating the metagenome analysis results using multi-locus sequence typing (MLST)

We attempted to verify the presence of the potential animal pathogens *E. coli*, *Proteus mirabilis*, *A. baumannii*, and *H. cinaedi* by direct PCR amplification of bacterial housekeeping genes from fly DNA samples. We used the existing MLST schemes for *E. coli*, *A. baumannii*, and *H. cinaedi*, but modified the published primers to prevent PCR amplification of related genera and/or species. For *P. mirabilis*, we developed primers for four housekeeping genes because no MLST scheme was yet available. From diluted fly DNA samples, we successfully amplified PCR fragments from all four species (Fig. 4). Those fragments that appeared to be the correct size were purified and sequenced. Some of the returned sequences contained numerous double peaks, indicating PCR amplification from multiple bacteria that belong to different but

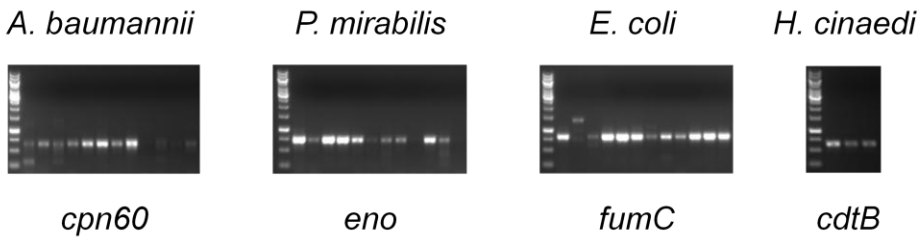

Figure 4. PCR amplifications from fly DNA samples using primers for *A. baumannii*, *P. mirabilis*, *E. coli*, and *H. cinaedi*. The gene amplified in each species is shown below the gel photograph.

closely-related species, such as *A. baumannii* and *A. calcoaceticus*, and/or multiple strains of one species (Table 2). The sequences were analyzed using phylogenetic trees and blastn searches against the NCBI nt (nucleotide) and wgs (whole genome shotgun) databases.

Table 2. PCR amplification and NCBI blastn search results of *Acinetobacter* genes from fly samples.

	BLASTn search result for housekeeping gene fragment					
Fly Sample	*cnp60* (600 bp)[1]	*gdhB* (618 bp)	*gltA* (684 bp)	*gyrB* (570 bp)	*recA*	*rpoD* (573 bp)
D11	PCR[2]			PCR		
H02	PCR				PCR	
H07	*Acinetobacter* sp.[3]		PCR	*Acinetobacter* sp.	PCR	*Acinetobacter* sp.
H10	*Acinetobacter* sp.	*A. calcoaceticus* (97%)	*Acinetobacter* sp.	PCR	PCR	PCR
H11	*Acinetobacter* sp.			PCR	PCR	*A. calcoaceticus* (98%)
O01	PCR	*A. oleivorans* (97%)	*Acinetobacter* sp.	PCR	PCR	*A. oleivorans* (98%)
P03	PCR				PCR	
P04	*A. baumannii* (100%)		*A. baumannii* (100%)	*A. baumannii* (99%)	PCR	*A. baumannii* (100%)
P06	PCR				PCR	
S17	*A. baumannii* (100%)					

[1] Size of the analyzed sequence.

[2] PCR indicates an amplicon of the correct size but with frequent double peaks in the sequence, suggesting PCR amplification from multiple *Acinetobacter* species and/or strains.

[3] *Acinetobacter* sp. showed 82 to 87% similarity to multiple *Acinetobacter* species.

The sequence analysis using both BLAST searches and phylogenetic trees revealed the presence of several *Acinetobacter* species in the fly samples, including *A. baumannii*, *A. calcoaceticus*, *A. oleivorans*, as well as currently unknown species (Table 2, Fig. 5).

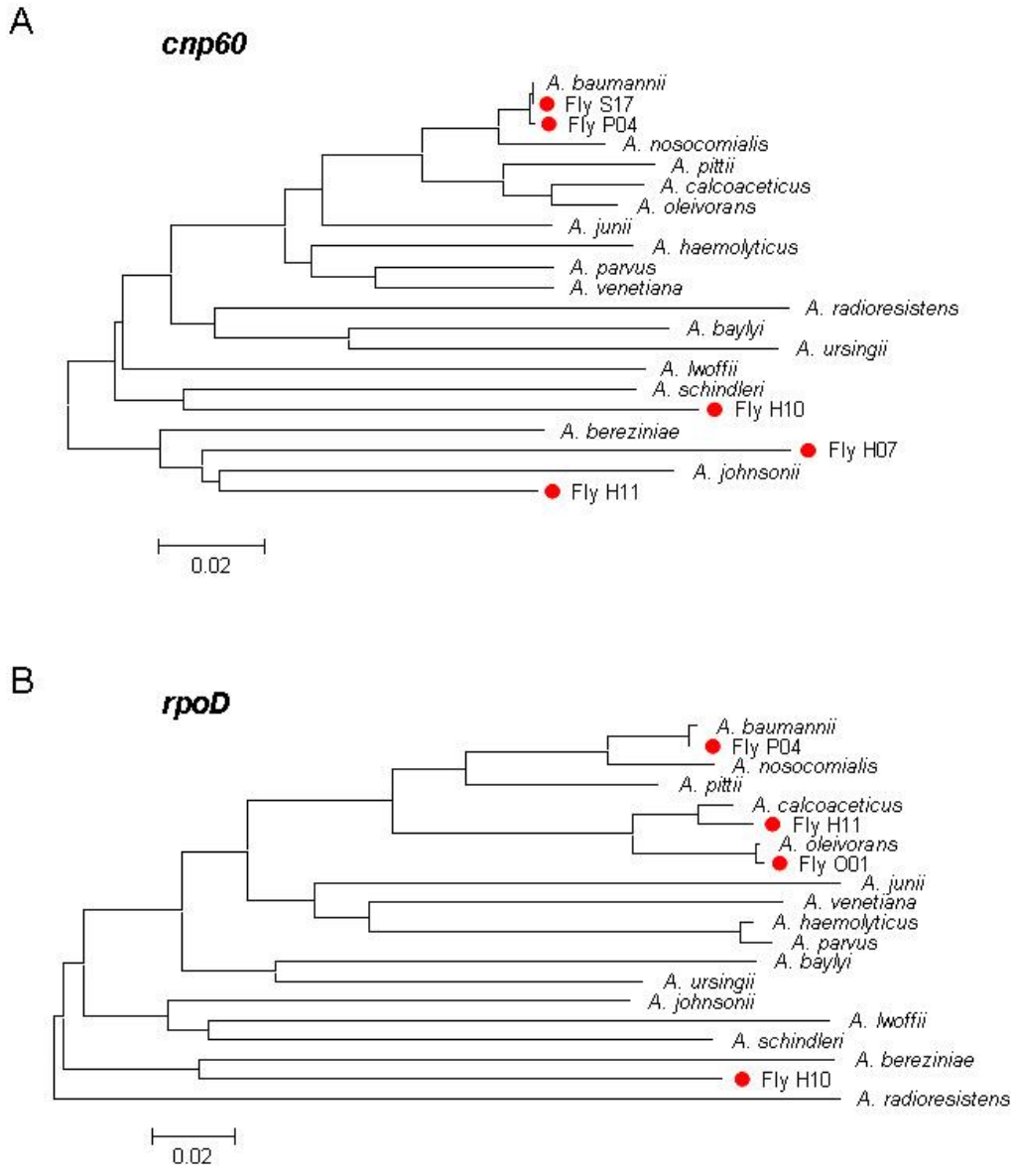

Figure 5. Neighbor-joining trees of two Acinetobacter genes, *cnp60* (A) and *rpoD* (B). The genes of the different Acinetobacter species were extracted from the respective reference genomes available in GenBank.

Likewise, we identified two *Proteus* species, *P. mirabilis* and *P. penneri* as well as currently unknown *Proteus* species (Table 3). Indeed, several species of both *Acinetobacter* and *Proteus* had been identified in the metagenome analysis (Fig. 3). *Acinetobacter* species included *A. baumannii*, *A. lwoffii*, and *A. johnsonii*, as well as *Acinetobacter* sp. P8-3-8. *Proteus* species included *P. mirabilis* and *P. penneri.*

Table 3. PCR amplification and NCBI blastn search results of *Proteus* genes from fly samples.

	BLASTn search result for housekeeping gene fragment			
Fly Sample	*atpA* (576 bp)[1]	*eno* (618 bp)	*fabH* (708 bp)	*trpC* (420 bp)
C02	*P. mirabilis* (100%)	PCR[2]	*P. mirabilis* (100%)	*P. mirabilis* (99%)
H08	PCR	PCR	PCR	PCR
O05	*P. penneri* (96%)	*P. penneri* (93%)	Proteus sp.[3]	Proteus sp.
O07	*P. penneri* (98%)	*P. penneri* (96%)	*P. penneri* (89%)	*P. penneri* (90%)
P03	PCR	PCR	PCR	PCR
P04	PCR	PCR	PCR	PCR
P06	*P. penneri* (96%)	*P. penneri* (93%)	Proteus sp.	Proteus sp.
S10	*P. mirabilis* (100%)	*P. mirabilis* (100%)	*P. mirabilis* (100%)	*P. mirabilis* (100%)
S21	*P. mirabilis* (100%)	*P. mirabilis* (100%)	*P. mirabilis* (100%)	*P. mirabilis* (100%)

[1] Size of the analyzed sequence.
[2] PCR indicates an amplicon of the correct size but frequent double peaks in the sequence, suggesting PCR amplification from multiple *Proteus* species and/or strains.
[3] *Proteus* sp. showed 77 to 86% similarity to both *P. mirabilis* and *P. penneri*.

Like in *Acinetobacter*, the sequence traces of several *Proteus* PCR fragments contained double peaks, probably due to simultaneous amplification of the same gene from two or more strains or species. Inspection of the gene traces revealed that the double peaks were likely due to mixed sequences from both *P. mirabilis* and *P. penneri*, since the observed pattern in all four genes

matched the pure sequences found in both species (Fig. 6). With this knowledge, we hypothesized that the same was true for *Acinetobacter* samples.

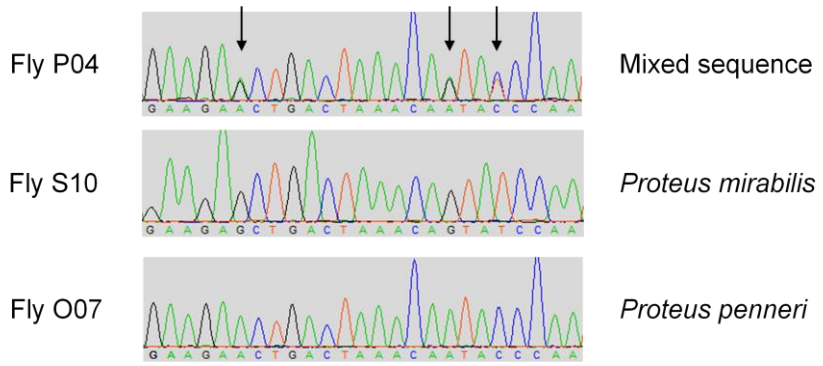

Figure 6. Sections of the sequencing chromatograms of the *eno* gene amplified from three fly samples. Fly sample P04 contained mixed traces that likely resulted from amplification of the gene from both *P. mirabilis*, as in fly S10, and *P. penneri*, as found in fly O07. The arrows indicate the double peaks shown in this chromatogram.

Similarly, we amplified and sequenced housekeeping gene fragments from *Helicobacter cinaedi* from the three fly samples that contained noticeable amounts of this bacterium, samples O08, O14, and S15. Sequences confirmed the presence of *H. cinaedi*.

Though many samples contained *E. coli*, we picked six for characterization and sequenced all seven housekeeping genes of the MLST scheme, *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*. All sequences confirmed that we successfully amplified *E. coli*. Amplicons from one fly sample again contained mixed traces, apparently amplified from various *E. coli* clones. The sequences of the other five samples were assigned the allele numbers of the MLST scheme. The majority of these sequences were known alleles; only six were new alleles that deviated by one nucleotide each from previously identified alleles. The allele combination was new for all five *E. coli* samples from the flies; however, we identified closely related isolates in the *E. coli* database that differed by only two or three out of the seven alleles. None of the related isolates from the database were known pathogens, but rather represented commensal *E. coli*. Since we were

unable to amplify all genes from all fly samples, however, it is still possible that other samples contained virulent *E. coli* strains.

Chapter 4

Discussion

Sampling methods

The animal facilities from which we obtained our fly samples are quite close together (Fig. 1). Depending on the daily flight radius of the flies, the barns could well have shared a fly population that travelled from one place to the next, carrying whatever bacteria they picked up. The time of year in which we collected flies may have affected the bacterial loads of our samples. The abundance of some microbes may follow a cyclic pattern throughout the year or perhaps fluctuates with ambient temperature, humidity, or day length. Similarly, the ability of flies to carry bacterial species could change with the seasons. Other points of unaddressed variability include fly age, size, and reproductive ability. The livestock with which the flies associate may demand more attention in future studies as they, too, may carry bacteria in different amounts during different parts of the year or experience a change in carrying capacity or microbe constituency over the course of its life.

Our sampling equipment served its purpose well, enabling us to efficiently capture the flies we needed. However, we used the rotting salmon head and jars repeatedly and at multiple locations, so further analysis of our data must include metagenomic sequencing of samples of the salmon. This is expected to indicate whether the presence or abundance of some species was affected by the bait. The salmon was stored at -80°C when not in use, which would kill many of the bacteria present, there might have been some psychrophilic species, such as *Psychrobacter cryohalolentis*, capable of surviving the extreme cold.

The effects of sampling location on bacterial flora

Despite the close proximity to each other, the sampling locations yielded noticeable differences in flies' bacterial loads, and this is presumably due to variations in the local habitat. For example, fly samples from the horse barn contained more bacteria, and more animal pathogens, than any other samples. The horses are taken off farm property more often than any of the other animals, which may have impacted their bacterial loads and, in turn, the bacterial loads of the flies. The Deer Research Center is segregated from the other facilities by the wooded acres on and around the property, and the clear pattern of bacterial species within each sample, absent in the samples from other locations, seems to reflect this unique characteristic. Blowfly samples from the poultry building reveal relatively limited amounts of bacteria and bacterial species. The poultry buildings are fully enclosed except for the ventilation system. Thus, the difference in samples obtained inside the building versus the samples gathered outdoors at the other locations is to be expected to some degree. The sheep farm was farthest from the other sampling locations, and this is reflected in the apparent bacterial population. More commensal species were identified in the sheep samples than in any other sampling group. Blowflies from near Meyer Dairy were sparsely populated with bacteria. This may be due to our distance from the cows if blowflies do not travel far from their food source. Alternatively, the size of the dairy's property and the movement of the cows through large fields during the day could cause such a low bacterial presence by effectively diluting the dairy's bacterial population. Finally, in the case of swine, metagenomic analysis revealed sporadic occurrence of several animal pathogens.

Importance of the chosen pathogenic bacteria of interest

Acinetobacter baumannii is an opportunistic pathogen of humans and the most common organism of the *Acinetobacter* genus to cause human infections. It is responsible for cases of hospital-acquired pneumonia and meningitis, urinary tract infection, and skin, soft tissue and bone infections (Cerqueira and Peleg, 2011). It is still unclear whether the hospital strains are of animal origin, but *Acinetobacter* has been isolated from swine and cattle feces and skin, nostril, and ear swabs (Hamouda et al., 2011). The presence of the bacteria on the animals' skin and areas that tend to attract flies, such as the ears, increases the potential of bacterial uptake/spread by flies that visit those areas. In our study, our most certain sample containing *A. baumannii* came from poultry (Table 2). After viewing the sequence traces from MLST PCRs, we saw that samples from horses contain *Acinetobacter* but not necessarily *A. baumannii*.

In humans, *Helicobacter cinaedi* is the most commonly isolated enterohepatic *Helicobacter* species. It is currently unknown whether *H. cinaedi* is part of the normal microflora of the human gastrointestinal tract, but it has indeed been identified in asymptomatic infections (Oyama, 2012). Infections can cause fever, nausea, diarrhea, recurrent cellulitis, endocarditis, and arthralgia (Kiehlbauch, 1994).

Various *Helicobacter* species have been identified in birds, humans, and several mammals, including rodents, ferrets, woodchucks, cats, dogs, and non-human primates. *H. cinaedi* was found in rhesus macaques, which may therefore be one reservoir for human infection (Fernandez et al., 2002). Additionally, *Helicobacter* species have been isolated from cetaceans and livestock (Whary and Fox, 2004). In the context of this work, the *Helicobacter* species infecting swine, cattle, sheep, and chickens are the most compelling due to the potential for transfer to the humans who work near them. A study of the prevalence of *H. heilmannii* in communities of varied demographics and geography indicated that infections occurred most often

in individuals from rural areas, where contact with domestic livestock and other animals is more common than in urbanized regions (Švec et al., 2000).

Swine may be infected with *H. suis*, which was initially considered a strain of *H. heilmannii*, *Candidatus H. bovis* colonizes the abomasum of cattle, and *H. rappini* was found to cause liver necrosis in sheep fetuses, suggesting that infection can be vertically transmitted. *H. pullorum* is isolated from the liver, duodenum, and cecum of chickens but was also isolated from humans with gastroenteritis, evidencing the potential for *H. pullorum* to act as a zoonotic agent (Whary and Fox, 2004). From our flies, two sheep and one swine sample contained *H. cinaedi*, as confirmed by sequenced PCR fragments.

Proteus mirabilis is an opportunistic pathogen, typically found in the human gastrointestinal tract or free-living in soil or water, but occasionally the bacterium causes urinary tract infections and, more rarely, pneumonia or endocarditis. *P. mirabilis* can cause mastitis in cows, making it a rare, opportunistic pathogen with economic significance. Not only can a cow lose milk-producing function in the affected portion of the udder, but the milk produced during the time she is infected must be discarded (Phiri et al., 2010). Multidrug-resistant strains have been isolated on meat products, predominantly poultry, and from processing facilities (Kim et al., 2005). Though normally only an opportunistic pathogen, antibiotic resistance could increase the virulence of this bacterium. We successfully amplified *P. mirabilis* from a few of our samples as well as *P. penneri*, also a rare but serious pathogen.

The variation in bacterial loads between flies

The impressive diversity of bacterial species and the amounts of certain microbial species in livestock and poultry could be due to the substrates available to the flies that associate with them. Each animal has a distinct fecal composition, with varied amounts of water and solids.

Blowflies prefer dead animals or decaying meat for laying eggs, but dead animals were not readily available to the flies at the facilities we visited. The next best substrate for blowflies to deposit their eggs is fecal material of the local animals.

Given the large amounts of manure produced by dairy cattle, one would perhaps expect large blowfly populations and significant bacterial carriage, but it is possible that the fluid feces of the cattle do not offer enough nutritional support for high numbers of blowflies. Stoffolano et al. studied the effects of various fecal substrates on *Phormia regina* sexual maturation and reproduction and found that male and female flies can both survive and successfully reach sexual maturity by feeding solely on feces from one species, but the time spent feeding to reach maturity varied based on the protein content of the dung (1995).

Our horse barn samples contained abundant *Pseudomonas* and *Acinetobacter* species in relatively high quantities. Additionally, *Salmonella enterica* and *E.coli* appeared to some degree in almost every sample. This may be particularly important due to the close interactions between horses and the humans who work with them. Whereas the swine, sheep, cattle, and deer are used strictly for production or research and receive limited direct contact with their caretakers, the horses are used individually for recreational or competitive purposes, as well. This close association leads to increased risk of disease transmission between humans and horses. Blowfly association with horses adds another dimension to the risk analysis of equine activities.

The bacterial loads in samples from the Deer Research Center are intriguing due to the large number of reads from the plant pathogens *Dickeya dadantii* and *Spiroplasma citri*. If the blowflies obtained these bacteria from the deer, then deer can probably spread the bacteria through fecal shedding and may thereby introduce the pathogens to plants during typical movements through their environment. Deer are abundant enough in Pennsylvania that this could become a threat to food crops. Alternatively, the increase in plant pathogens may simply correlate with the increased vegetation of the woods.

The difference in the bacterial loads from one family of flies to the next can be attributed to the lifestyles of the flies. Members of Calliphoridae, the blowflies, are attracted to smells of rotting meat and thus visit carrion to eat and possibly lay eggs. Blowflies also consume fecal matter, which inevitably contains the shed bacteria of the animal that excreted it. House flies are attracted to dung as ovipositional sites, as well, but may also select decaying organic matter such as human food scraps. These dietary differences could influence the bacterial populations that the flies carry.

Within the poultry samples, the flies with the highest bacterial load belonged to family Calliphoridae. The house flies, on the other hand, were nearly devoid of bacteria. During sampling, we collected our indoor samples from the ground below the birds' cages. The majority of this fly population belonged to the family Muscidae. Our *Musca domestica* samples were nearly devoid of bacteria. Because the chickens are kept indoors throughout their lives, they have little opportunity to become infected with any bacteria that weren't either vertically transmitted or introduced to the chickens by human workers. Thus, it makes sense that the flies within the poultry building would contain limited bacterial flora. This does not necessarily indicate that house flies are not vectors of pathogens, however. A study of the ability of *Musca domestica* to carry and transmit *H. pylori* to other animals or humans found that *H. pylori* can be isolated from the flies' body surface for up to 12 hours post-bacterial contact. Moreover, viable bacteria could be isolated from the alimentary tracts and feces/vomitus for up to 30 hours post-bacterial consumption (Grübel et al., 1997).

Our MLST primers occasionally amplified the genes of two different bacterial species within one sample, as revealed by the double peaks in the sequence traces in *E. coli*, *P. mirabilis*, and *A. baumannii*. Our ability to amplify all fragments from each of our samples may have been limited by our high annealing temperature, which we used in order to prevent cross-amplification

of genes from other genera. As seen in the cases of *Proteus* and *Acinetobacter*, we still frequently amplified from multiple species and/or clones.

Overall, the metagenome analysis methods presented here were effective at detecting the presence of many bacterial species and the relative amounts of each. Additionally, the use of MLST schemes to verify the presence of certain bacteria successfully provided proof of the presence of the species or genera in question. The knowledge of the genera present may still be enough to aid the development of specialized or targeted disease prevention plans. Analysis of sequence traces from PCR fragments also enabled us to identify fly samples that contained either more than one strain of a species or multiple closely-related species.

Further study correcting for the close proximity of the animal facilities to one another as well as the differences in housing livestock and poultry may offer further insight as to correlations between mammal or avian hosts, fly species, and bacterial loads. Additionally, it would be very interesting to study more closely the bacteria associated with swine, especially gastrointestinal populations, since pigs are quite similar to humans

Appendix A

Detailed protocols

DNA extraction using Qiagen DNeasy Blood and Tissue kit

Choose flies of the desired species or diversity from which to extract DNA, and work efficiently so that the flies do not thaw completely. Transfer each chosen fly into a separate, clean, 2 ml microcentrifuge tube, and grind them thoroughly using a pestle. To the ground fly add 360 µL of lysis buffer ATL, vortex, and then grind the sample again. Next add 40 µL proteinase K and vortex. The samples should be incubated on a heat block at 56°C for 10 min. and then vortexed again. Store the samples in an incubator at 56°C overnight.

The next day, the fly tissues should be totally lysed (except the exoskeleton). Add 400 µL of buffer AT to the microcentrifuge tube, vortex the mixture, and then incubate for 10 min. at 56°C. Add 400µL of ethanol (96-100%) and mix by vortexing. Centrifuge for 1 min. at 6000×*g* (8000 rpm). Transfer 600 µL of the supernatant to a DNeasy Mini spin column placed in a 2 ml collection tube and centrifuge at 6000×*g* (8000 rpm) for 1 min. Discard the flow-through and return the Mini spin column to its collection tube before adding an additional 500 µL of the fly tissue supernatant to the appropriate column. Centrifuge again at 6000×*g* (8000 rpm) for 1 min., then discard the flow-through and collection tube.

With the Mini spin column in a fresh collection tube, add 500 µL buffer AW1 and centrifuge at 6000×*g* (8000 rpm) for 1 min. Discard the collection tube and flow-through. Place the column in another clean collection tube and add 500 µL buffer AW2. Centrifuge at 20,000×*g* (14,000 rpm) for 3 min., and discard the flow-through and collection tube.

Transfer the column to a 1.5 mL microcentrifuge tube, add 100 µL buffer AE to the center of the column membrane, incubate at room temperature for 1 min., and centrifuge for 1 min. at 6000×*g* (8000 rpm) to obtain the first elution. Transfer the column to a new 1.5 mL microcentrifuge tube, add 50 µL buffer to the column membrane, incubate for 1 min., and then centrifuge at 6000×*g* (8000 rpm). This is the second elution.

PCR product purification protocol – use Qiagen QIAquick PCR Purification Kit

To begin, add 5 volumes of buffer PB to 1 volume of sample and mix. For example, our sample volumes post-gel electrophoresis were approximately 22 µL, so 110 µL Buffer PB was combined with each sample. The entirety of the mixture is then transferred to a QIAquick spin column in a 2 ml collection tube and centrifuged for 1 min. at 14,000 rpm (g?). Discard the flow-through but keep the collection tube. Add 750 µL Buffer PE and centrifuge for another 1 min. Discard the flow-through and centrifuge the samples for an extra minute. Discard the collection tube and place the column in a 1.5 mL microcentrifuge tube. To elute the DNA, add Buffer EB to the center of the column membrane and incubate for 1 min. The volume of Buffer EB added will depend on the desired end DNA concentration; published protocol recommends 50 µL, but we often used less for a more concentrated product.

Appendix B

Multi-locus sequence typing primers for chosen pathogenic bacteria

Acinetobacter baumannii

Locus	Sequence	Amplicon size	Fragment size	AnnealingTemp	Melting Temp
gltA	TTACAGTGGCACATTAGGTCC	717	484	58	62
	GATACCAGCAGAGATACACG				60
gyrB	TGAAGGCGGCTTATCTGAGT	594	457	58	60
	GCTGGGTCTTTTTCCTGACA				60
gdhB	GCTACTTTTATGCAACAGAGCC	774	344	58	62
	TTGAGTTGGCGTATGTTGTGC				62
recA	CCTGAATCTTCYGGTAAAACTAC	425	371	58	62
	TTCTGGGCTGCCAAACATTAC				62
cnp60	GGTGCTCAACTTGTTCGTGA	640	421	58	60
	CACCGAAACCAGGAGCTTTA				60
gpi	AAATTTCCGGAGCTCACAAAAC	456	305	58	62
	TCAGGAGCAATACCCCACTC				62
rpoD	ACCCGTGAAGGTGAAATCAG	672	513	58	60
	TTCAGCTGGAGCTTTAGCAAT				60

Helicobacter cinaedi

Locus	Sequence	Amplicon size	Fragment size	AnnealingTemp	Melting Temp
ppa	CTCAAAAGTATCAGTAGGCGA	514	411	58	60
	GCCCTTGTAGGCTTTGATTG				60
aspA	GGCGGCTCTAGCAAATAATG	650	532	58	60
	CCGTATCTTGTGTCGCTTCA				60
aroE	CGCACATTCTAAATCCCCAC	688	572	58	60
	TAAGGCTAGGGCTGCTTGAT				60
atpA	TGTGGTTGGACGCGTTATTAA	646	536	58	60
	TGGCAATGCTGTAAGTGAGC				60
tkt	AATCTGCTTCACTAGCCGGA	665	562	58	60
	CCTGTGGAAAATCGCCTTCA				60
cdtB	GGTGTAGCATTTGGTGCGAT	635	535	58	60
	TCAAGTATGCCTCCGCTTCT				60

Escherichia coli

Locus	Sequence	Amplicon size	Fragment size	AnnealingTemp	Melting Temp
adk	GCGTATCATTCTGCTTGGCG	590	536	58	62
	CCGTCAACTTTCGCGTATTT				58
fumC	GCCAGCGCTTCAAATTTGTTC	806	469	58	62
	GTACGCAGCGAAAAAGATTCG				62
gyrB	TCGGCGACACGGATGACGGC	911	460	58	64
	ATCAGGCCTTCACGCGCATC				64
icd	GAAAGTAAAGTAGTTGTTCCGG	878	518	58	62
	GGACGCAGCAGGATCTGTT				60
mdh	AGCGCGTTCTGTTCAAATGC	932	452	58	60
	GAACTCTCTCTGTATGATATCG				62
purA	CGCGCTGATGAAAGAGATGA	816	478	58	60
	CATACGGTAAGCCACGCAGA				62
recA	GCAGTCGCATTCGCTTTACC	780	510	58	62
	TCGTCGAAATCTACGGACCG				62

Proteus mirabilis

Locus	Sequence	Amplicon size	Melting/Annealing Temp °C	Genome position	
eno	GGTGCAAACGCAATCCTAGC	733	62 / 58	262848	264149
	GTCAGTGAACCGATTTGGTTG		60		
fabH	GCCGATTTAGAAAAAAATGGTTG	811	62 / 58	949259	950212
	CTCTTACAGCTTCATCTAATGC		62		
trpC	GCTAATGGCTTTTTAGTGGG	533	58 / 58	1424761	1426134
	CCACCTGCAAGTAATGCTTT		58		
atpA	CTCACCTGTTGAGATGATTGC	671	62 / 58	3364028	3365569
	CGGTAATAGAGATTACGTTCG		60		

Appendix C

Additional figures and tables

Table C1. DNA concentrations of all samples.

First elution		Second elution	
Sample	**DNA concentration (ng/µL)**	**Sample**	**DNA concentration (ng/µL)**
Swine			
S01a	7.82	**S01b**	8.75
S02a	5.71	**S02b**	7.64
S03a	7.27	**S03b**	12.1
S04a	11.1	**S04b**	11.7
S05a	5.66	**S05b**	6.59
S06a	7.32	**S06b**	12.2
S07a	5.22	**S07b**	6.02
S08a	10.2	**S08b**	10.3
S09a	11.9	**S09b**	13.6
S10a	15.8	**S10b**	22
S11a	9.56	**S11b**	9.34
S12a	16.3	**S12b**	18.2
S13a	10.2	**S13b**	10.2
S14a	9.59	**S14b**	7.16
S15a	12.6	**S15b**	9.89
S16a	12.8	**S16b**	11.4
S17a	11.1	**S17b**	11.5
S18a	14.5	**S18b**	14.5
S19a	8.67	**S19b**	7.39
S20a	15.2	**S20b**	13.4
S21a	17.2	**S21b**	16
Horses			
H01a	3.63	**H01b**	4.52
H02a	15.7	**H02b**	10.8
H03a	2.09	**H03b**	1.97

First elution		Second elution	
Sample	**DNA concentration (ng/µL)**	**Sample**	**DNA concentration (ng/µL)**
Horses			
H04a	9.06	**H04b**	9.83
H05a	3.63	**H05b**	3.43
H06a	3.23	**H06b**	3.76
H07a	6.81	**H07b**	5.48
H08a	7.62	**H08b**	4.53
H09a	5.71	**H09b**	7.17
H10a	12.8	**H10b**	9.04
H11a	18.1	**H11b**	4.66
H12a	9.11	**H12b**	5.93
H13a	1.04	**H13b**	0.77
H14a	7.31	**H14b**	6.75
H15a	13.2	**H15b**	9.92
H16a	6.62	**H16b**	5.65
Poultry - outside			
PO01a	269	**PO01b**	183
PO02a	283	**PO02b**	233
PO03a	265	**PO03b**	224
PO04a	288	**PO04b**	226
PO05a	69.4	**PO05b**	59
PO06a	59.8	**PO06b**	61.4
PO07a	91.4	**PO07b**	60.7
PO08a	88.3	**PO08b**	2.037
PO09a	71.5	**PO09b**	57
PO10a	50.3	**PO10b**	64
PO11a	54.9	**PO11b**	34.8

First elution		Second elution	
Sample	**DNA concentration (ng/µL)**	**Sample**	**DNA concentration (ng/µL)**
Poultry - outside			
PO12a	44.7	**PO12b**	33.6
PO13a	61.4	**PO13b**	43.5
PO14a	49.3	**PO14b**	110
PO15a	45.7	**PO15b**	41.4
PO16a	40.9	**PO16b**	35
PO17a	45.9	**PO17b**	34.6
PO18a	46.8	**PO18b**	39.9
Poultry - inside			
PI01a	159	**PI01b**	130
PI02a	206	**PI02b**	155
PI03a	219	**PI03b**	122
PI04a	307	**PI04b**	204
PI05a	292	**PI05b**	203
PI06a	235	**PI06b**	164
PI07a	324	**PI07b**	212
PI08a	235	**PI08b**	104
PI09a	219	**PI09b**	136
PI10a	219	**PI10b**	94.6
PI11a	190	**PI11b**	126
PI12a	181	**PI12b**	126
Dairy cattle			
C01a	0.7	**C01b**	0.721
C02a	22.8	**C02b**	14
C03a	1	**C03b**	0.791
C04a	6.2	**C04b**	4.21

First elution		Second elution	
Sample	**DNA concentration (ng/µL)**	**Sample**	**DNA concentration (ng/µL)**
Dairy cattle			
C05a	3.1	**C05b**	0
C06a	6.3	**C06b**	0.685
C07a	7.1	**C07b**	1.95
C08a	2.8	**C08b**	1.18
C09a	0.6	**C09b**	0.674
C10a	1.3	**C10b**	0.89
C11a	3	**C11b**	1.14
C12a	2.6	**C12b**	1.2
C13a	1.7	**C13b**	1.42
C14a	3.5	**C14b**	1.44
C15a	0.5	**C15b**	0
C16a	9	**C16b**	6.41
C17a	1.5	**C17b**	2.05
C18a	4.4	**C18b**	2.39
C19a	4.4	**C19b**	1.93
C20a	11.1	**C20b**	1.82
C21a	20.3	**C21b**	10.4
C22a	16.2	**C22b**	10.1
C23a	26.6	**C23b**	15.3
C24a	12.5	**C24b**	15.8
C25a	8.29	**C25b**	6.89
C26a	17.4	**C26b**	13.8
C27a	21.8	**C27b**	14.1
C28a	9.37	**C28b**	4.04
C29a	15.4	**C29b**	4.79

First elution		Second elution	
Sample	**DNA concentration (ng/µL)**	**Sample**	**DNA concentration (ng/µL)**
Dairy cattle			
C30a	10.6	**C30b**	5.88
C31a	8.2	**C31b**	3.1
C32a	15.4	**C32b**	9.11
Deer			
D01a		**D01b**	0.72
D02a		**D02b**	13.3
D03a		**D03b**	8.05
D04a		**D04b**	1.72
D05a		**D05b**	5.13
D06a		**D06b**	2.91
D07a		**D07b**	2.81
D08a	7.68	**D08b**	3.06
D09a	10.1	**D09b**	3.66
D10a	13.5	**D10b**	4.84
D11a	6.97	**D11b**	4.25
D12a	6.29	**D12b**	5.03
D13a	3.64	**D13b**	2.45
D14a		**D14b**	2.73
D15a		**D15b**	8.68
D16a		**D16b**	6.14
D17a		**D17b**	13.5
D18a		**D18b**	2.38
D19a		**D19b**	11.4
D20a		**D20b**	11.8
D21a		**D21b**	18.7

First elution		Second elution	
Sample	**DNA concentration (ng/µL)**	**Sample**	**DNA concentration (ng/µL)**
Sheep			
O01a		**O01b**	16.7
O02a		**O02b**	1.85
O03a		**O03b**	1.41
O04a		**O04b**	2.29
O05a		**O05b**	4.12
O06a		**O06b**	2.28
O07a		**O07b**	1.99
O08a		**O08b**	15.4
O09a		**O09b**	11.2
O10a		**O10b**	1.89
O11a		**O11b**	4.04
O12a		**O12b**	3.92
O13a		**O13b**	4.4
O14a		**O14b**	11.2
O15a		**O15b**	1.52
O16a		**O16b**	2.29
O17a		**O17b**	9.36
O18a		**O18b**	5.44
O19a		**O19b**	1.58
O20a		**O20b**	11.8

Figure C1. Top 50 bacterial species identified in fly samples from non-ruminants.

Figure C2. Top 50 bacterial species identified from fly samples in ruminants.

Table C2. Fly sample identifications using morphological traits and mtDNA.

Sample ID	Sample Name	Fly Family	*Morphological ID*	*Molecular ID*	Gender	Location
355	C18	Sarcophagidae	*Sarcophagidae family*	*Boettcheria latistema*	NA	Cattle
351	C08	Calliphoridae	*Lucilia sp.*	*Lucilia coeruleiviridis*	F	Cattle
348	C05	Calliphoridae	*Lucilia sp.*	*Lucilia illustris*	F	Cattle
332	C02	Calliphoridae	*Lucilia sp.*	*Lucilia sericata*	NA	Cattle
347	C04	Calliphoridae	*Lucilia sp.**	*Lucilia sericata*	F	Cattle
352	C11	Calliphoridae	*Lucilia sp.**	*Lucilia sericata*	F	Cattle
353	C14	Calliphoridae	*Lucilia sp.*	*Lucilia sericata*	F	Cattle
356	C19	Calliphoridae	*Lucilia sp.*	*Lucilia sericata*	F	Cattle
357	C20	Calliphoridae	*Lucilia sp.*	*Lucilia sericata*	F (??)	Cattle
349	C06	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Cattle
350	C07	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Cattle
354	C16	Calliphoridae	*Phormia regina*	*Phormia regina*	M	Cattle
335	D17	Sarcophagidae	*Sarcophagidae family*	*Boettcheria bisetosa*	NA	Deer
365	D02	Muscidae	*??*	*Eudasyphora canadiana*	NA	Deer
366	D08	Calliphoridae	*Lucilia sp.*	*Lucilia coeruleiviridis*	F	Deer
369	D14	Muscidae	*Muscidae or Sarcophagidae?*	*Muscina levida*	NA	Deer
333	D03	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Deer
334	D10	Calliphoridae	*Phormia regina*	*Phormia regina*	M	Deer
336	D20	Calliphoridae	*Phormia regina*	*Phormia regina*	M	Deer
367	D09	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Deer
368	D11	Calliphoridae	*??*	*Phormia regina*	F	Deer
370	D15	Calliphoridae	*Phormia regina*	*Phormia regina*	M	Deer
371	D16	Calliphoridae	*Phormia regina*	*Phormia regina*	M	Deer
372	D19	Calliphoridae	*Phormia regina*	*Phormia regina*	M	Deer
338	O08	Calliphoridae	*Lucilia sp.**	*Lucilia sericata*	M	Sheep
339	O09	Calliphoridae	*Lucilia sp.**	*Lucilia sericata*	M	Sheep
340	O14	Calliphoridae	*Lucilia sp.*	*Lucilia sericata*	F	Sheep
344	O20A	Calliphoridae	*Lucilia sp.*	*Lucilia sericata*	M	Sheep
361	O11A	Calliphoridae	*Lucilia sp.**	*Lucilia sericata*	F	Sheep
364	O18A	Calliphoridae	*Lucilia sp.**	*Lucilia sericata*	F	Sheep
337	O01	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Sheep
358	O05A	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Sheep
359	O07A	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Sheep
360	O10A	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Sheep
362	O13A	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Sheep
363	O17A	Calliphoridae	*Phormia regina*	*Phormia regina*	F	Sheep
329	P06	Calliphoridae	*Cochliomyia macellaria*	*Cochliomyia macellaria*	NA	Poultry

Sample ID	Sample Name	Fly Family	*Morphological ID*	*Molecular ID*	Gender	Location
345	P10		*Hydtrotaea sp.*	*Hydtrotaea sp.*	NA	Poultry
328	P05	Calliphoridae	*Lucilia sericata*	*Lucilia sericata*	NA	Poultry
330	P12	Muscidae	*Musca domestica*	*Musca domestica*	NA	Poultry
331	P13	Muscidae	*Musca domestica*	*Musca domestica*	NA	Poultry
342	P07	Muscidae	*Musca domestica*	*Musca domestica*	NA	Poultry
343	P08	Muscidae	*Musca domestica*	*Musca domestica*	NA	Poultry
325	P02	Calliphoridae	*Phormia regina*	*Phormia regina*	NA	Poultry
326	P03	Calliphoridae	*Phormia regina*	*Phormia regina*	NA	Poultry
327	P04	Calliphoridae	*Phormia regina*	*Phormia regina*	NA	Poultry
341	P01	Calliphoridae	*Phormia regina*	*Phormia regina*	NA	Poultry
346	P11	Calliphoridae	*Phormia regina*	*Phormia regina*	NA	Poultry
1047	S08a	Calliphoridae	*Lucilia sericata*	*Lucilia sericata*		Swine
1051	S13a	Calliphoridae	*Lucilia sericata*	*Lucilia sericata*		Swine
1048	S09a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1049	S10a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1050	S12a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1052	S15a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1053	S16a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1054	S17a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1055	S18a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1056	S20a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1057	S21a	Calliphoridae	*Phormia regina*	*Phormia regina*		Swine
1046	S04a	Tachinidae	*Winthemia rufoptica*	*Winthemia rufoptica*		Swine
1062	H08a	Calliphoridae	*Lucilia sericata*	*Lucilia sericata*		Horse
1068	H15a	Calliphoridae	*Lucilia sericata*	*Lucilia sericata*		Horse
1058	H01a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1059	H02a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1061	H07a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1064	H10a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1065	H11a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1066	H12a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1067	H14a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1069	H16a	Calliphoridae	*Phormia regina*	*Phormia regina*		Horse
1060	H04a	Muscidae	*Stomoxys calcitrans*	*Stomoxys calcitrans*		Horse
1063	H09a	Muscidae	*Stomoxys calcitrans*	*Stomoxys calcitrans*		Horse

REFERENCES

Boughammoura, A., D. Expert, and T. Franza. 2012. Role of *Dickeya dadantii* Dps protein. Biometals. 25:423-233.

Cerqueira, G.M. and A.Y. Peleg. 2011. Insights into *Acinetobacter baumannii* Pathogenicity. IUBMB Life. 63: 1055-60.

Chase-Topping, M., D. Gally, C. Low, L. Matthews, and M. Woolhouse. 2008. Super-shedding and the link between human infection and livestock carriage of *Escherichia coli* O157. Nat. Rev. Microbiol. 6: 904-912.

Dominguez, M.G., E.K. Costello, M. Contreras, M. Magris, G. Hidalgo, N. Fierer, and R. Knight. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. U.S.A. 107: 11971-5.

Esteban, J.I., B. Oporto, G. Aduriz, R.A. Juste, and A. Hurtado. 2007. A survey of food-borne pathogens in free-range poultry farms. Int. J. Food Microbiol. 123: 177-182.

Fernandez, K.R., L.M. Hansen, P. Vandamme, B.L. Beaman, and J.V. Solnick. 2002. Captive Rhesus Monkeys (*Macaca mulatta*) Are Commonly Infected with *Helicobacter cinaedi*. J. Clin. Microbiol. 40: 1908-12.

Fischer, O.A., L. Mátlová, L. Dvorská, P. Švástová, J. Bártl, R.T. Weston, and I. Pavlík. (2004*a*) Blowflies *Calliphora vicina* and *Lucilia sericata* as passive vectors of *Mycobacterium avium* subsp. *avium*, *M. a. paratuberculosis* and *M. a. hominissuis*. Med. Vet. Entomol. 18: 116-122.

Francesconi, F. and O. Lupi. 2012. Myiasis. Clin. Microbiol. Rev. 25:79-105.

Grübel, P., J.S. Hoffman, F.K. Chong, N.A. Burstein, C. Mepani, and D.R. Cave. 1997. Vector Potential of Houseflies (*Musca domestica*) for *Helicobacter pylori*. J. Clin. Microbiol. 53: 1300-3.

Hamouda, A., J. Findlay, L.A. Hassan, and S.G.B. Amyes. 2011. Epidemiology of *Acinetobacter baumannii* of animal origin. Int. J. Antimicrob. Agents. 38: 314-8.

Huson, D.H., S. Mitra, H-J. Ruscheweyh, N. Weber, S. C. Schuster. 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 21: 1552-1560.

Keihlbauch, J.A., R.V. Tauxe, C.N. Baker, and I.K. Wachsmuth. 1994. *Helicobacter cinaedi*-associated Bacteremia and Cellulitis in Immunocompromised Patients. Ann. Intern. Med. 121: 90-93.

Kruse, H., A-M. Kirkemo, and K. Handeland. 2004. Wildlife as Source of Zoonotic Infections. Emerging Infect. Dis. 10: 2067-2072.

Leclerque, A. 2008. Whole genome-based assessment of the taxonomic position of the arthropod pathogenic bacterium *Rickettsiella grylli*. FEMS Microbiol. Lett. 283: 117-27.

Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature. 442: 806-9.

Maldonado, M.A. and N. Centeno. 2003. Quantifying the Potential Pathogens Transmission of the Blowflies (Diptera: Calliphoridae). Mem. Inst. Oswaldo Cruz. 98: 213-216.

Mannering, S.A., D.M. West, S.G. Fenwick, R.M. Marchant, and K. O'Connell. 2006. Pulsed-field gel electrophoresis of *Campylobacter jejuni* sheep abortion isolates. Vet. Microbiol. 115: 237-242.

New Zealand Government. 2013. New Zealand Economic and Financial Overview 2013. www.treasury.govt.nz/economy/overview/2013/nzefo-13.pdf. (Accessed 31 March 2013.)

Kim, S., C. Wei, and H. An. 2005. Molecular Characterization of Multidrug-Resistant *Proteus mirabilis* Isolates from Retail Meat Products. J. Food. Prot. 68: 1408-1413.

Staden, R., K.F. Beal, and J.K. Bonfield. 1998. The Staden package, 1998. Methods Mol. Biol. 132: 115-30.

Steinkraus, G.E. and B.D. Wright. 1994. Septic Abortion with Intact Fetal Membranes Caused by *Campylobacter fetus* subsp. *fetus*. J. Clin. Microbiol. 32: 1608-9.

Stoffolano, J.G., M. Li, J.A. Sutton, and C. Yin. 1995. Faeces feeding by adult *Phormia regina* (Diptera: Calliphoridae): impact on reproduction. Med. Vet. Entomol. 9:388-92.

Sukontason, K.L., M. Bunchoo, B. Khantawa, S. Piangjai, Y. Rongsriyam, and K. Sukontason. 2007. Comparison Between *Musca Domestica* and *Chrysomya megacephala* as Carriers of Bacteria in Northern Thailand. Southeast Asian J. Trop. Med. Public Health. 38: 38-44.

Švec, A., P. Kordas, Z. Pavlis, and J. Novotný. 2000. High Prevalence of *Helicobacter heilmannii*-associated Gastritis in a Small, Predominantly Rural Area: Further Evidence in Support of a Zoonosis? Scand. J. Gastroenterol. 35: 925-928.

Oyama, K., S. Khan, T. Okamoto, S. Fujii, K. Ono, T. Matsunaga, J. Yoshitake, T. Sawa, J. Tomida, Y. Kawamura, and T. Akaike. 2012. Identification of and Screening for Human *Helicobacter cinaedi* Infections and Carriers via Nested PCR. J. Clin. Microbiol. 50: 3893-3900.

Phiri, A.M., W. Muleya, and K.E. Mwape. 2010. Management of chronic gangrenous mastitis in a 3-year-old cow using partial (quarter) mastectomy. 42: 1057-61.

Urban, J.E. and A. Broce. 1998. Flies and Their Bacterial Loads in Greyhound Dog Kennels. Curr. Microbiol. 36: 164-170.

USDA. 2007/2009??? Prevalence of *Salmonella* and *Listeria* in bulk tank milk and in-line filters on U.S. dairies. UDSA-APHIS-VS, CEAH, Fort Collins, CO.

Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D. Wu, I. Paulson, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. Rogers, and H.O. Smith. 2004. Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science. 304: 66-74.

Whary, M.T. and J.G. Fox. 2004. Natural and Experimental *Helicobacter* Infections. Comp. Med. 54: 128-158.

Wood, C.L., K.D. Lafferty. 2013. Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. Trends Ecol. Evol. 28: 239-247.

Zhao, H., H. Liu, Y. Du, S. Liu, H. Ni, Y. Wang, C. Wang, W. SI, J. Yang, and J. Ling. 2010. Development and evaluation of an indirect enzyme-linked immunosorbent assay for the detection of antibodies against *Campylobacter fetus* in cattle. Res. Vet. Sci. 88: 446-451.

ACADEMIC VITA

Caylie Hake

cayliehake@gmail.com

Education

B.S., Animal Sciences, May 2013, Pennsylvania State University, University Park, PA

Honors and Awards

- President's Freshman Award, 2010
- American Society of Animal Science undergraduate award
- College of Agricultural Sciences Honor Roll, Fall 2009 and Spring 2010

Association Memberships/Activities

- Collegiate Horseman's Association, 2009 through 2013
- Presidential Leadership Academy, 2010 through 2013

Research Interests

I am interested in veterinary medicine and large animal medicine as they apply to management. I enjoy developing an understanding of research relevant to current issues in agriculture and veterinary care.