11-29-2022

Genetic and Environmental Drivers of Microbial Colonization in Wild Birds

Sage Rohrer
University of Missouri-St. Louis, sdrq5f@mail.umsl.edu

Genetic and Environmental Drivers of Microbial Colonization in Wild Birds

Sage D. Rohrer

M.S. in Biology, University of Missouri-St. Louis, 2019

B.S. in Biology, University of Missouri-St. Louis, 2017

A Dissertation Submitted to The Graduate School at the University of Missouri-St. Louis

in partial fulfillment of the requirements for the degree

Doctor of Philosophy in Biology with an emphasis in Ecology, Evolution and Systematics

December

2022

Advisory Committee

Patricia G. Parker, Ph.D.

Lon M. Chubiz, Ph.D.

Robert J. Marquis, Ph.D.

Peter O. Dunn, Ph.D.

TABLE OF CONTENTS

ABSTRACT

Wild birds are teeming with microorganisms, ranging from commensal bacteria to eukaryotic parasites. These microbes impact host health in diverse ways; some long-term residents may aid digestion or provide immune system training, while others at times may induce disease. However, the factors driving the varying colonization patterns seen across taxa are still not fully understood, and wild avian populations are particularly understudied when compared to mammalian or domesticated systems. This work examines the relative importance of genetic and environmental factors driving microbiome composition and function in wild birds. We collected fecal and blood samples from Galapagos penguins (*Spheniscus mendiculus*) in Ecuador, and from Eurasian tree sparrows (*Passer montanus*) in the United States, screened all samples for blood parasites, and used a combination of targeted amplicon sequencing and whole genome sequencing to characterize the composition and function of the gut microbiome. We additionally used targeted amplicon sequencing to characterize major histocompatibility complex diversity in the Eurasian tree sparrow. Results showed that: (1) environmental and demographic factors drive microbial community structure in terms of alpha diversity, beta diversity, and differential abundance, particularly in the sparrow system with high environmental heterogeneity, (2) both gut pathogens and blood parasites are associated with significant shifts in microbial community structure, although putative bacterial pathogens in the gut are associated with the strongest variation in both community structure and function, and (3) genetic variation at immune system loci is associated with the Eurasian tree sparrow gut microbiome in terms of

MHC diversity, but this relationship is generally outweighed by other drivers in this system. Using a metagenomic approach across two distinct systems, this work provides greater insight into the role the gut microbiome plays in avian health by demonstrating significant associations with disease status and immunogenetics.

ACKNOWLEDGMENTS

I want to express my deep gratitude to the many people who were an essential part of this work, particularly to my committee. My co-advisor, Dr. Parker, has inspired me for many years through her extensive research on avian disease ecology and her commitment to conservation. She has been a steady guide, finding creative ways to overcome the inevitable roadblocks and encouraging me to take advantage of every opportunity. Dr. Chubiz, my co-advisor, has been a wonderful mentor through his excellent science and commitment to his students. His insightful feedback has shaped the way I think about scientific questions, always encouraging me to consider my research from new perspectives. Dr. Marquis and Dr. Dunn have both been vital members of my committee – their extensive knowledge and unerringly constructive feedback have been invaluable. Thank you all for your advice and support.

I also wish to thank the UMSL Biology Department, St. Louis Zoo, and collaborators in Galapagos and Missouri for their support and research assistance. The graduate students and faculty members in this department make UMSL a lovely place, and I am grateful for the knowledge I have gained and the close friendships I have made here. The members of the Parker Lab and Chubiz Lab have been endlessly helpful by critiquing my research, troubleshooting unexpected problems, and assisting with data collection in the field and lab – in particular, I owe a big thank you to Dr. Maricruz Jaramillo and Briana Robertson. The St. Louis Zoo has provided generous financial support and collaboration, and I am very grateful to Jane Merkel and Kimberly Aubuchon for their

enthusiastic efforts sampling penguins in Galapagos. I am also deeply grateful to Gustavo Jiménez-Uzcátegui, the Charles Darwin Foundation, the Galapagos National Park, and the Galapagos Biosecurity Agency for their collaboration and inspirational work.

Finally, I would like to thank my dear family – my partner Jose Antonio, my parents Lori & Rex, and my brother Kai – for their constant support during this challenging period in my life. You were always there when I needed you, and this work is for you.

CHAPTER 1.

Composition and function of the Galapagos penguin gut microbiome vary with age, location, and a putative bacterial pathogen

Sage D. Rohrer[1], Gustavo Jiménez-Uzcátegui[2], Patricia G. Parker[1,3], & Lon M. Chubiz[1]

[1]*Department of Biology and Whitney R. Harris World Ecology Center, University of Missouri-St. Louis, One University Blvd., St. Louis, MO 63121, USA*

[2]*Charles Darwin Research Station, Charles Darwin Foundation, Galapagos, Ecuador*

[3]*WildCare Institute, Saint Louis Zoo, One Government Drive, St. Louis, MO 63110, USA*

Abstract:

Since microbial colonization plays a direct role in host health, understanding the ecology of the resident microbial community for a given host species is an important step for detecting population vulnerabilities like disease. However, the idea of integrating microbiome research into conservation is still relatively new, and wild birds have received considerably less attention than mammals or domesticated animals. Here we examine the composition and function of the gut microbiome of the endangered Galapagos penguin (*Spheniscus mendiculus*) with the goal of characterizing the normal

microbial community and resistome, identifying likely pathogens, and testing hypotheses of structuring forces for this community based on demographics, location, and infection status. We collected fecal samples from wild penguins in 2018 and performed 16S rRNA gene sequencing and whole genome sequencing (WGS) on extracted DNA. 16S sequencing revealed that the phyla *Fusobacteria*, *Epsilonbacteraeota*, *Firmicutes*, and *Proteobacteria* dominate the community. Functional pathways were computed from WGS data, showing genetic functional potential primarily focused on metabolism. WGS samples were each screened for antimicrobial resistance, characterizing a resistome made up of nine antibiotic resistance genes. Samples were screened for potential enteric pathogens using virulence factors as indicators; *Clostridium perfringens* was revealed as a likely pathogen. Overall, three factors appear to be shaping the alpha and beta diversity of the microbial community: penguin developmental stage, sampling location, and presence or absence of *C. perfringens*. We found that juvenile penguins have significantly lower alpha diversity than adults based on three metrics, as well as significantly different beta diversity compared to adults. Location effects are slight, but one site has significantly lower Shannon diversity than the other primary sites. Finally, when samples were grouped by *C. perfringens* virulence factors, we found dramatic changes in beta diversity based on operational taxonomic units, protein families, and functional pathways. This study provides a baseline microbiome for an endangered species, implicates both penguin age and the presence of a potential bacterial pathogen as the primary factors

associated with microbial community variance, and reveals widespread antibiotic resistance genes across the population.

Keywords: Avian microbiome, gut pathogens, conservation, Galapagos penguin.

Original Article under review.

Introduction:

Resident microbes have a diverse impact on animal health. Some long-term members of the microbiome may facilitate digestion or provide immune system training for the host (Hills et al. 2019). Colonization resistance is an important benefit, as mutualistic microbes may out-compete a pathogenic invader for space and nutrients or even release toxins as a deterrent (Lazar et al. 2018). Some community members may also be opportunistically pathogenic, occupying an inconspicuous position in the microbiome until a disruption occurs and then infecting the host (Brown et al. 2012, Krismer et al. 2017).

Characterizing the normal microbiome for species of concern provides an essential baseline from which to measure changes. This has the potential to help zoos improve their level of care for collection animals as well as to increase reintroduction success by mimicking a species' normal microbiome (Bahrndorff et al. 2016, Trevelline et al. 2019). However, microbial community assessments can also improve management of wild populations in a number of ways. Dysbiosis in the microbiome (i.e. disruption of the normal bacterial community) can be an indicator of health problems and is associated with a number of diseases (Hills et al. 2019, Videvall et al. 2020). Metagenomic assessments can reveal potential pathogens that may pose a threat to the population (Gu et al. 2021). Sequencing data can additionally be used to assess the level of antimicrobial resistance in the community, known as the "resistome," which at high levels can indicate a potentially dangerous level of connectivity between humans and

wildlife (Wheeler et al. 2012). Thus, microbiome research provides a critical perspective for conservation efforts.

Factors structuring these resident microbial communities are complex and can vary substantially between taxa. Diet is known to be a primary driver of microbiome composition and taxa with atypical diets tend to have distinct microbiomes (Roggenbuck et al. 2014, Carmody et al. 2015, Waite and Taylor 2015). For instance, the vampire finch (*Geospiza septentrionalis*) supplements its diet in an unusual way by eating eggs, guano, and the blood of larger birds, and the species also exhibits a unique microbiome profile compared to the other Darwin's finches (Michel et al. 2018). These dietary changes can be rapid – one study of human diet found that switching from a plant- to animal-based diet resulted in reduced carbohydrate fermentation and increased protein fermentation by gut microbes in a matter of days, with corresponding abundance changes in bacteria associated with those activities (David et al. 2014). Environmental factors such as season and habitat are known to affect gut microbiomes as well, though these differences are largely attributed to changes in food availability between seasons (Tasnim et al. 2017, Hu et al. 2018). However, diet is a better predictor for microbiome composition and function in mammals than it is for birds, and the effect of diet is weakest in the microbiomes of both bats and flying birds (Song et al. 2020).

Demographic factors such as the host's sex or developmental stage can also play a role, as males and females of some species have different gastrointestinal microbiomes (Escallón et al. 2019, Vemuri et al. 2019). One study found that cloacal microbiomes differed between male and female rufous-collared sparrows during breeding seasons,

with the male microbiome becoming more diverse at the onset of the breeding season (Escallón et al. 2019). Hormonal differences or immune variation between sexes may be responsible to some degree (Markle et al. 2013, Bolnick et al. 2014b, Escallón et al. 2019). Many taxa also undergo microbiome changes throughout development; for example, little penguins (*Eudyptula minor*) exhibit increased abundances of *Firmicutes* and *Bacteroidetes* as they mature (Dewar et al. 2017). Microbial community differences based on sex and age are not consistent across taxa, and in some cases the effect size is quite small compared to other factors, indicating a continued need to assess community drivers on a case-by-case basis in wild populations (Bolnick et al. 2014a, Taylor et al. 2019).

This study examines the gut microbiome of the Galapagos penguin (*Spheniscus mendiculus*), a highly range-restricted species which occurs only in the Galapagos Islands (Vargas et al. 2005). This penguin faces regular population bottleneck events when the nutrient-rich Cromwell Current is disrupted periodically by warm El Niño weather patterns (Vargas et al. 2006). The species is classified as Endangered by the International Union for Conservation of Nature (IUCN) due to the severe declines associated with these events as well as their highly restricted range (Vargas et al. 2005, 2006). The frequent population bottlenecks are likely the reason for the genetic homogeneity found in this species – both microsatellite markers and major histocompatibility complex (MHC) sequences demonstrate a low degree of genetic variation (Akst et al. 2002, Bollmer et al. 2007, Nims et al. 2008).

The low genetic variation may leave the penguin population vulnerable to introduced diseases (Sommer 2005, Levin et al. 2009). Previous studies have found evidence of prior infections by *Chlamydophila psittaci* and *Toxoplasma gondii* (Travis et al. 2006, Deem et al. 2010), as well as infections by microfilariae (species unknown) and a lineage of *Plasmodium* (Lineage A) (Merkel et al. 2007, Levin et al. 2009). However, microbiome characterization and a thorough assessment of enteric pathogens using next-generation sequencing tools has not been completed and would provide valuable insight into the health of this species. Furthermore, widespread antibiotic resistance genes have been found previously in the Galapagos Islands (Wheeler et al. 2012, Overbey et al. 2015, Nieto-Claudin et al. 2019), but the extent to which antibiotic resistance is associated with the Galapagos penguin was unknown as we began this study.

This research thus has four goals: (1) establish a baseline for gut microbiome taxonomy and function in the Galapagos penguin, (2) identify putative enteric pathogens, (3) characterize the resistome; and (4) explore drivers of community structure. We hypothesized that hormonal variation and contrasting foraging habits may lead to distinctive microbiomes, and so microbial community structure would vary depending on both sex and age, respectively. Due to the largely homogeneous environment of the western coast of Isabela Island, we hypothesized that different locations may be a minimal factor in the gastrointestinal communities, particularly since the penguin population shows significant movement between sites and the penguins tend to forage near the coast (Steinfurth et al. 2008, Nims et al. 2008). Finally, we hypothesized that

gut pathogens -if found- may be associated with community changes facilitated through either disruptive or opportunistic invasion.

Methods:

All Galapagos penguin fecal samples used in this study were collected in July of 2018 from Isabela Island and the Marielas Islets in Galapagos, Ecuador. To collect samples, wild penguins were safely captured at each site using long-handled nets and brought to a large boat for processing. Processing included morphological measurements and opportunistic fecal collection. The fecal collection procedure involved harvesting feces from clean plastic sheets placed beneath the penguins during transport and handling, immediately following capture. Fecal samples were preserved in 95% ethanol at room temperature (Song et al. 2016). Males and females were identified based on bill depth measurements and size (Cappello and Boersma 2018, Jiménez-Uzcátegui et al. 2021). Fecal DNA was later extracted using Qiagen PowerFecal extraction kits and stored at -20°C.

Targeted Sequencing:

Sequencing of the 16S rRNA gene V4 region was performed with Illumina MiSeq on 40 of the DNA samples, excluding samples with insufficient DNA (<20ng), at the University of Michigan Medical School Microbiome Core. The V4 region was selected because its length of 250bp allows forward and reverse reads to fully overlap when sequenced with the cost-efficient Illumina platform, reducing the error rate (Kozich et al. 2013). Method standardization also facilitates comparisons between microbiome studies, and the V4 region has widespread use, notably through the Earth Microbiome Project and the

Mothur Standard Operating Procedure (SOP) (Kozich et al. 2013, Thompson et al. 2017). Negative sampling and extraction controls were included to assess contamination during processing, and a water sample and a mock community (ZymoBIOMICS Microbial Community Standard) were added by the Microbiome Core prior to library preparation as negative and positive PCR amplification and sequencing controls, respectively. Several penguin samples were later resequenced alongside a ZymoBIOMICS Gut Microbiome Standard, which had been stored in 95% ethanol prior to extraction to serve as a positive control for the preservation/extraction methods. The resequenced samples were examined to ensure bacterial composition was comparable to the originally sequenced samples; however, all data included in the downstream analysis was generated from the first sequencing run to avoid any potential batch effects. The ZymoBIOMICS Microbial Community Standard was also added to this second sequencing run by the Microbiome Core as a sequencing positive control.

Read analysis was conducted in Mothur, following the Schloss MiSeq standard operating procedure (Schloss et al. 2009, Kozich et al. 2013). Reads were filtered by base quality and length, aligned with the SILVA 132 reference database (Yilmaz et al. 2014) and filtered for alignment quality and chimeras. Samples with less than 10,000 reads following data cleaning steps were excluded, leaving 34 samples. Operational taxonomic unit (OTU) clustering was based on a 97% similarity threshold. Unclassified reads and reads that matched mitochondria, chloroplasts, or archaea were filtered out. Filtered reads were subsampled in Mothur to match the sample with the lowest read count (11,295 reads). Community data were imported into R version 3.6.3 to analyze alpha

diversity (within individuals) and beta diversity (between individuals) (R Core Team 2020).

Whole Genome Sequencing:

Whole genome sequencing (WGS) was performed on 20 of the extracted fecal samples by the University of Michigan Microbiome Core using the Illumina Nextera DNA Flex kit (Franzosa et al. 2018). Several of these samples were later resequenced with the ZymoBIOMICS Gut Microbiome Standard included in the same run as a positive control to ensure comparable results – the resequenced samples were examined to make sure the bacterial compositions were similar to the original sequence results, but downstream analysis included only data generated in the initial sequencing run. Quality control was performed by trimming reads using Trimmomatic with a sliding window of 4, minimum quality score of 20, and minimum length of 70 (Bolger et al. 2014). These cleaned reads were used directly for read mapping with KMA (k-mer alignment) against two databases, VFDB (Virulence Factor Database) and CARD (Comprehensive Antibiotic Resistance Database), on the platform PATRIC (Pathosystems Resource Integration Center), to identify likely pathogens by virulence factors and to assess probable antibiotic resistance genes (Liu et al. 2019, Alcock et al. 2020, Davis et al. 2020). The cleaned reads were categorized taxonomically using the k-mer based program Kraken 2, also available on PATRIC (Wood et al. 2019).

The trimmed reads were also assembled with SPAdes, using the recommended K-mer lengths of 21, 33, 55, 77, 99, and 127, with the BayesHammer module for error correction enabled (Bankevich et al. 2012, van der Walt et al. 2017). SPAdes output was

assessed with MetaQuast (Mikheenko et al. 2016). The assembly quality was not consistent across samples. N50 ranged from ~600 - ~2300, misassemblies ranged from 1-79, and longest contigs per sample ranged from 3375bp - 489281bp. Compared to single genome assemblies, lower quality is expected for metagenomic assemblies, but two samples with the worst assemblies were ultimately excluded from much of the functional analysis due to poor downstream annotation results; alternate tools relying on reads rather than contigs were similarly problematic for those two samples. The contigs from the SPAdes assembly were classified taxonomically using Kraken 2 on PATRIC.

Finally, parallel read-based and assembly-based approaches were used to achieve a robust picture of the functional profile in these communities. Cleaned reads were categorized functionally using the program HUMAnN 3.0, while contigs were annotated using the program MetaErg (Franzosa et al. 2018, Dong and Strous 2019). Pfam (Protein Family Database) annotations were obtained through both pipelines, which notably annotated virulence-associated proteins in addition to the virulence factors (pathogen indicators) obtained via PATRIC (Mistry et al. 2021). Pathways were only computed for a minority of gene families in both HUMAnN 3.0 and MetaErg, leaving a majority unclassified. Protein families and metabolism-related KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways detected from the metagenomic assemblies were imported to R to examine differences in metabolic profiles between groups (R Core Team 2020). Protein family abundances and pathway abundances were subsampled to match the lowest sample (22,259 and 19,684 reads, respectively), after excluding one

sample with insufficient protein families and two samples with insufficient pathway results, to allow comparisons between samples without depth-biased results.

Control Assessment:

Since negative controls contained ≤30 reads following batch filtering steps for the 16S dataset, and <15 classified fragments following Kraken 2 classification of trimmed shotgun sequencing reads, potential contamination (from sampling, extraction kits, sequencing crossover, etc.) was considered negligible. 16S sequences from the Zymo preservation/extraction control (ZymoBIOMICS Gut Microbiome Standard) were mapped to manufacturer-provided reference sequences using the "seq.error" function in Mothur, which calculated an error rate of 0.0079%. Sequences from the Zymo sequencing control (ZymoBIOMICS Microbial Community Standard) were mapped to 16S reference sequences provided by the MBC, with an error rate of 0.015% and 0.005% for the first and second sequencing runs. All expected bacterial species with theoretical abundances >0.01% were detected in the mock communities following processing; some abundance skew was apparent, but this is unlikely to affect the reported results as all study samples were processed in the same batch and comparable to each other. Shotgun sequences of the ZymoBIOMICS Gut Microbiome Standard were directly mapped to the whole genome reference sequences provided by the manufacturer using Bowtie2 (Langmead and Salzberg 2012). Reads successfully mapped to all expected bacteria, archaea, and yeasts.

Diversity Calculations and Statistical Analysis:

Alpha diversity was calculated for each sample from the subsampled 16S rRNA gene sequencing data using the R package phyloseq for three metrics: observed richness, Simpson's Index, and Shannon Diversity Index (McMurdie and Holmes 2013). Significant differences based on alpha diversity values were assessed between groups using Kruskal Wallis tests (R Core Team 2020). Jaccard distance was used to calculate beta diversity in the phyloseq package for both 16S and WGS datasets, and differences in beta diversity were assessed using permutational multivariate analysis of variance (PERMANOVA) in the R package vegan (Anderson and Walsh 2013, McMurdie and Holmes 2013, Oksanen et al. 2019). The dispersion assumption for PERMANOVA, PERMDISP, was tested for significant values using the vegan package (Anderson 2001, Anderson and Walsh 2013, Oksanen et al. 2019). Kruskal Wallis tests were used to assess significant differences between relative abundances of metabolic pathways and groups from the WGS dataset, and false discovery correction was applied with the Benjamini-Hochberg method (Benjamini and Hochberg 1995).

Results:

The primary bacterial phyla found in the Galapagos penguin fecal samples are *Fusobacteria*, *Epsilonbacteraeota*, *Firmicutes*, and *Proteobacteria*, and the most common families are *Fusobacteriaceae*, *Helicobacteraceae*, *Clostridiaceae*, *Pasteurellaceae*, and *Peptostreptococcaceae* (Fig. 1). WGS samples were profiled functionally using level II of the BRITE hierarchy for classifying KEGG pathways. Most identified KEGG pathways were involved in metabolic activity, primarily amino acid

metabolism, carbohydrate metabolism, energy metabolism, and nucleotide metabolism (Fig. 1).

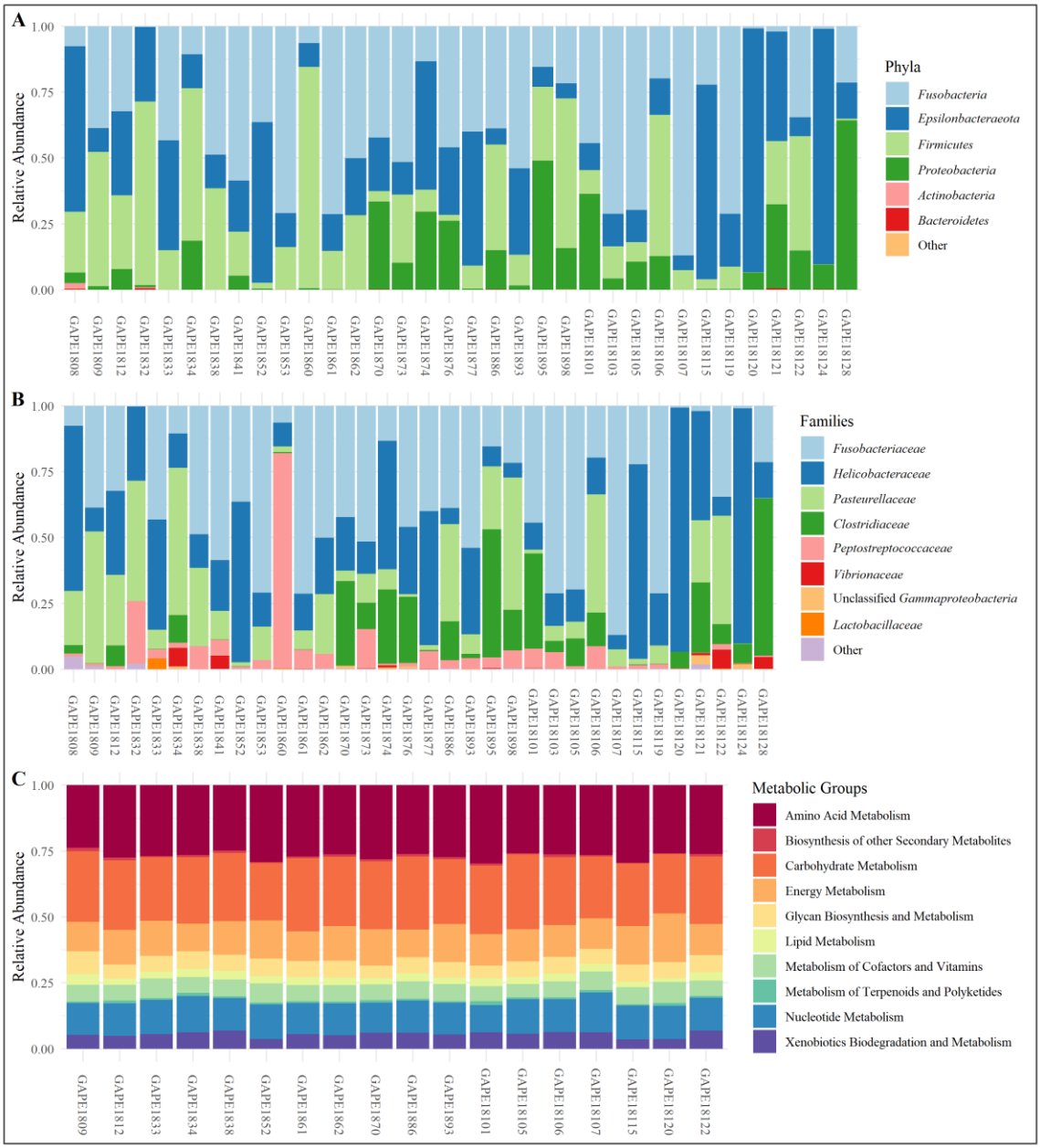

Figure 1. Bacterial phyla and families detected in 16S rRNA sequencing data (A, B), and metabolic KEGG pathway groups computed from shotgun sequencing data (C).

Alpha and beta diversity were calculated to assess diversity within and between communities, respectively. Overall, alpha diversity was low; no raw sample exceeded 100 identified OTUs and most contained fewer than 50 OTUs. Alpha diversity was significantly lower in juveniles compared to adults in each of the three diversity measures used (Fig. 2, observed richness P = 0.0068, Simpson's P = 0.0046, Shannon P = 0.0071). Beta diversity significantly differed between age classes after controlling for location in the model (PERMANOVA, R^2 = 0.06615, P = 0.016; PERMDISP, P = 0.816). Adult penguins did not vary by sex when comparing either alpha or beta diversity. No functional differences based on age or sex were apparent in the WGS dataset with either metabolic pathways or protein families.

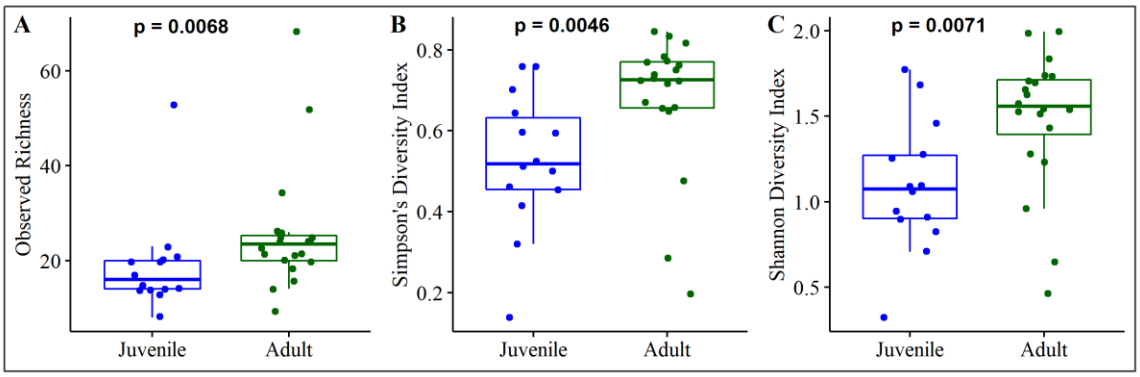

Figure 2. Alpha diversity is significantly lower in juveniles compared to adults when measured as observed richness (A), Simpson's Diversity Index (B), and Shannon Diversity Index (C).

We also examined the importance of location for microbiome composition. Due to the uneven sample distribution across sites, single samples from El Muñeco (at the northern end of Isabela Island) and Playa Perros (near Puerto Pajas) were first excluded. The remaining three sites – Caleta Iguana, Puerto Pajas, and Marielas – represent the largest

penguin colonies. Caleta Iguana is only represented by four samples following filtering steps, constraining our ability to compare between all three sites.

A Principal Coordinates Analysis (PCoA) indicates slight clustering by location (Fig. 3). Compared to the more exposed sample sites, Marielas is slightly different (PERMANOVA, 999 permutations, $R^2 = 0.05970$, P = 0.038; PERMDISP, P = 0.293). However, when the sites are limited to Marielas and Puerto Pajas there is no significant difference (PERMANOVA, 999 permutations, $R^2 = 0.06588$, P = 0.061). Shannon Diversity Index values significantly differ between the three sites (P = 0.03857), with the lowest Shannon Diversity seen at Marielas, but observed richness and Simpson's Index do not significantly vary (Fig. 3).

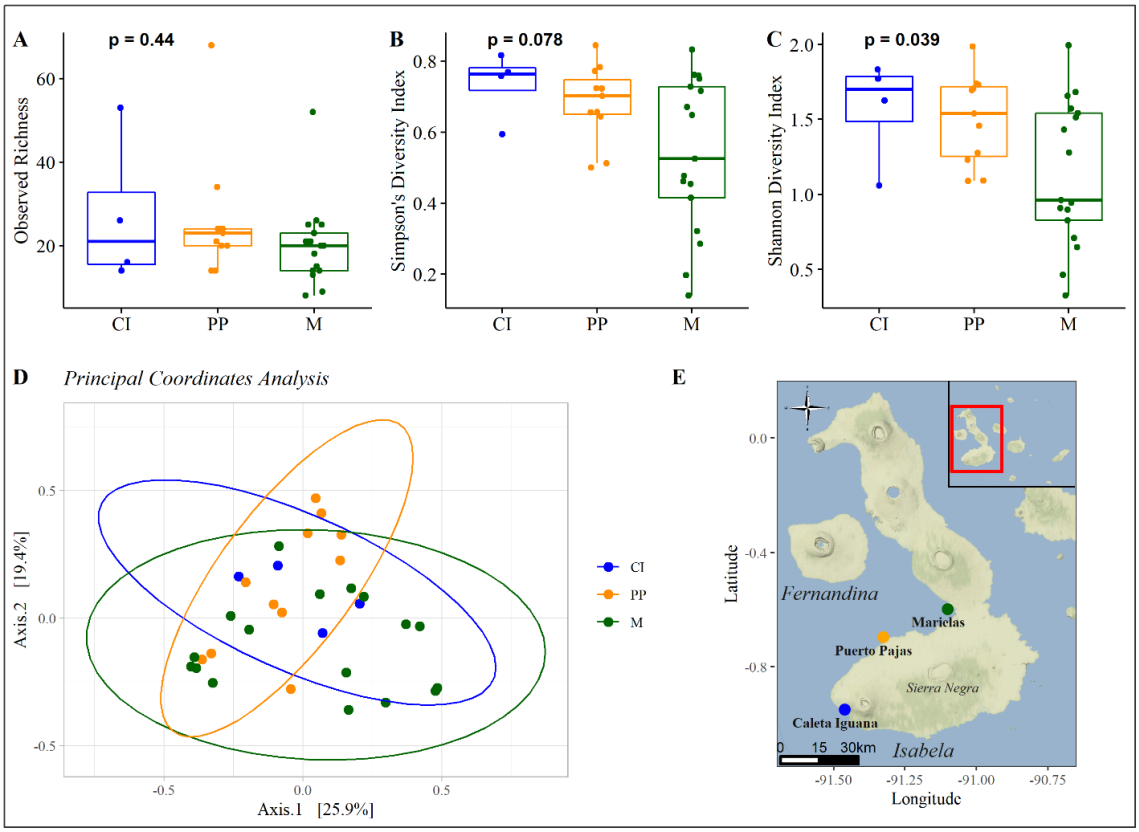

Figure 3. Alpha and beta diversity show minor variation between samples from the three primary sampling sites on Isabela Island: Caleta Iguana (CI), Puerto Pajas (PP), and Marielas (M). Observed richness and Simpson's Index are similar between sites (A, B), but Shannon Diversity Index significantly differs (C). A PCoA illustrates slight clustering between locations (D). Sample sites are shown along the western coast of Isabela Island, with the Galapagos Archipelago in the map inset (E).

Antibiotic Resistance Screening:

Screening shotgun sequencing reads from each sample with the CARD database revealed a total of nine putative antibiotic resistance genes (Table 1). Two resistance genes corresponding to *Helicobacter pylori* reference genomes were particularly widespread, occurring in almost all samples (19/20); these genes confer resistance to Tetracyclines and Macrolides. Genes resistant to Aminoglycosides and Lincosamides were also common in this small sample set, and one Peptide-resistant gene corresponding to *C. perfringens* SM101 was found in six samples.

Table 1. Antibiotic resistance genes detected in Galapagos penguin samples

CARD Accession	Function	# Samples	Reference Genome	Antibiotic Class
ARO:3003510	*Helicobacter pylori* 16S rRNA mutation conferring resistance to tetracycline	19	*Helicobacter pylori* 26695	Tetracycline
ARO:3004134	*Helicobacter pylori* 23S rRNA with mutation conferring resistance to clarithromycin	19	*Helicobacter pylori*	Macrolide
ARO:3003493	*Pasteurella multocida* 16S rRNA mutation conferring resistance to spectinomycin	9	*Pasteurella multocida* 36950	Aminoglycoside
ARO:3004149	*Escherichia coli* 23S rRNA with mutation conferring resistance to clindamycin	9	*Escherichia coli* CFT073	Lincosamide
ARO:3003773	*Clostridium perfringens* mprF	6	*Clostridium perfringens* SM101	Peptide
ARO:3001219	*Clostridium difficile* EF-Tu mutants conferring resistance to elfamycin	6	*Clostridium difficile*	Elfamycin
ARO:3003512	*Salmonella enterica* serovar Typhimurium 16S rRNA mutation in the rrsD gene conferring resistance to spectinomycin	5	*Salmonella enterica* subsp. *salamae*	Aminoglycoside
ARO:3004058	*Staphylococcus aureus* 23S rRNA with mutation conferring resistance to linezolid	1	*Staphylococcus aureus*	Oxazolidinone
ARO:3003403	*Escherichia coli* 16S rRNA mutation in the rrsB gene conferring resistance to paromomycin	1	*Escherichia coli* K-12	Aminoglycoside

Pathogen Screening:

WGS samples were screened for virulence factors using three tools with two databases: PATRIC (VFDB), HUMAnN 3.0 (Pfam), and MetaErg (Pfam). Virulence factors across all three tools were associated with a single bacterium, *Clostridium perfringens.* Most of the reference genome matches from the VFDB were from strain 13, which is classified as type A (Table 2). *C. perfringens* virulence factors from the VFDB were detected in 12/20 penguins, though the taxonomic search using Kraken 2 showed the presence of the bacterium in an additional seven penguins (19/20). The virulence-associated BrkB protein family was also classified with *C. perfringens* in both HUMAnN 3.0 and MetaErg (BrkB Pfam accession: PF0361), though to varying degrees – read-based HUMAnN 3.0 detected BrkB with *C. perfringens* in 8/20 samples, while contig-based MetaErg detected it in 10/20 samples.

Table 2. *Clostridium perfringens* virulence factors detected in PATRIC

Template	Gene	Product	Virulence Factor	# Samples	Reference Genome
VFDB|VFG002274	*plc*	phospholipase C	alpha-toxin	9	*C. perfringens* str. 13
VFDB|VFG002277	*nagH*	hyaluronidase	mu-toxin	8	*C. perfringens* str. 13
VFDB|VFG002276	*colA*	collagenase	kappa-toxin	7	*C. perfringens* str. 13
VFDB|VFG002284	*nanJ*	exo-alpha-sialidase	sialidase	7	*C. perfringens* str. 13
VFDB|VFG002285	*nanH*	sialidase	sialidase	7	*C. perfringens* ATCC 13124
VFDB|VFG002275	*pfoA*	perfringolysin O	theta-toxin	6	*C. perfringens* str. 13
VFDB|VFG002282	*cloSI*	alpha-clostripain	alpha-clostripain	6	*C. perfringens* str. 13
VFDB|VFG002283	*nanI*	exo-alpha-sialidase	sialidase	6	*C. perfringens* str. 13
VFDB|VFG002278	*nagI*	hyaluronidase	mu-toxin	6	*C. perfringens* str. 13
VFDB|VFG002279	*nagJ*	hyaluronidase	mu-toxin	6	*C. perfringens* str. 13
VFDB|VFG002280	*nagK*	hyaluronidase	mu-toxin	5	*C. perfringens* str. 13
VFDB|VFG002281	*nagL*	hyaluronidase	mu-toxin	5	*C. perfringens* str. 13
VFDB|VFG002286	*cpe*	enterotoxin Cpe	CPE (*C. perfringens* enterotoxin)	2	*C. perfringens* SM101

Additional bacterial taxa were associated with virulence factors in Pfam, though none appeared in the VFDB search. *Cetobacterium ceti, Clostridium baratii, Clostridium thermobutyricum, Paeniclostridium sordellii,* and *Photobacterium damselae* were

detected with virulence-associated protein families by both HUMAnN 3.0 and MetaErg. The contig-based approach, MetaErg, found an additional 28 virulence-associated species; the most common were *Helicobacter brantae*, *Gallibacterium anatis, Helicobacter* sp. 002287135*, Helicobacter* sp. 001693335, and *Fusobacterium* sp. 900015295. Most taxa were associated with the Pfam virulence factor BrkB, but others matched with a haemolysin (SMP_2), virulence-associated protein E, or virulence protein RhuM family, among others.

Since *C. perfringens* was the only bacterium consistently implicated as a pathogen by all pipelines, the presence of *C. perfringens* virulence factors was examined as a potential structuring force for the microbial communities. Samples grouped by *C. perfringens* virulence factors from the VFDB revealed signals of dysbiosis (Fig. 4). Beta diversity was calculated from protein families and from metabolic pathway abundances, and in both cases samples with virulence factors clustered away from samples without virulence factors on a PCoA. After controlling for location and including age and sex in the models, both protein families (PERMANOVA, 999 permutations, $R^2 = 0.14156$, $P = 0.020$; PERMDISP, $P = 0.076$) and metabolic pathways (PERMANOVA, 999 permutations, $R^2 = 0.22465$, $P = 0.014$; PERMDISP, $P = 0.081$) were significantly different based on the presence of *C. perfringens* virulence factors. Three metabolic groups and 17 metabolic KEGG pathways significantly differed between samples divided by virulence factors (adjusted P-values <0.05). This pattern held true in the 16S dataset, revealing taxonomic clustering based on the presence of *C. perfringens* virulence factors (PERMANOVA, 999 permutations, $R^2 = 0.17409$, $P = 0.001$; PERMDISP, $P = 0.27$). However, alpha diversity

did not significantly vary between groups. No significant relationship was found between *C. perfringens* status with age, sex, or location.

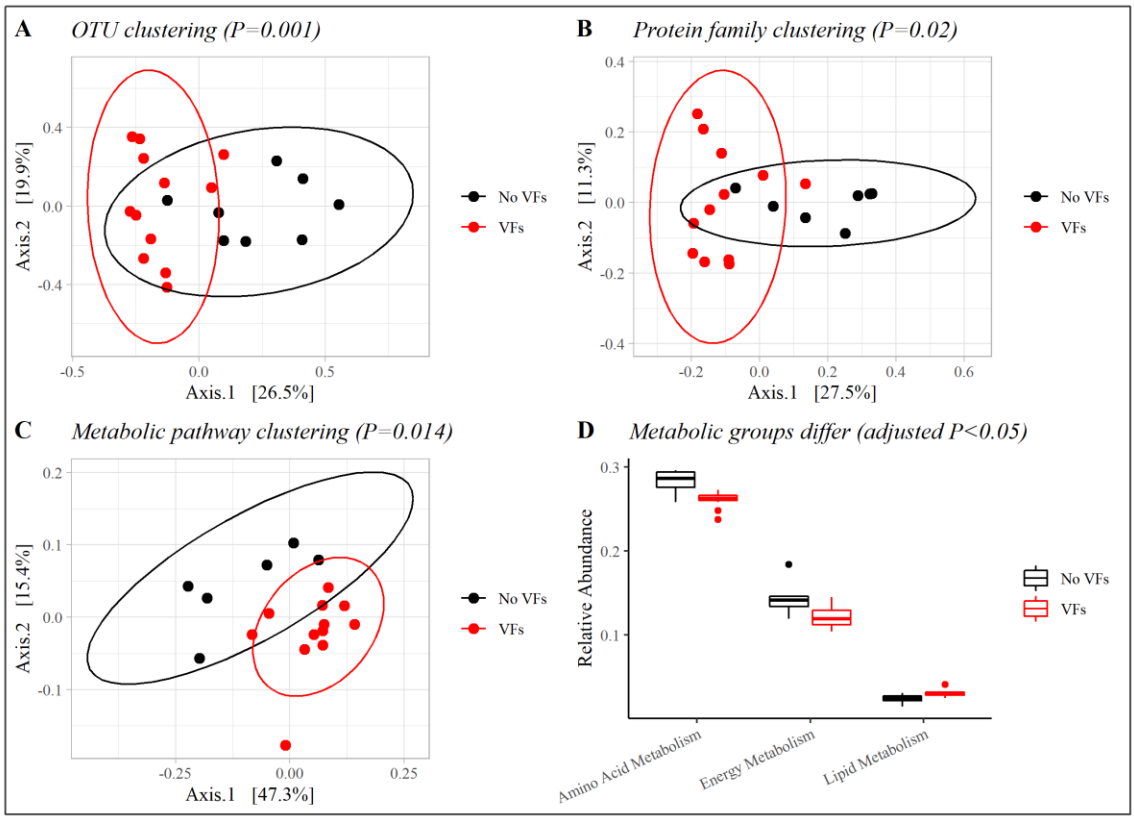

Figure 4. The penguin microbiome varies with *C. perfringens* virulence factors (VFs). Principal Coordinates Analysis shows distinct clustering based on OTUs from16S rRNA gene sequencing data (A), protein families from shotgun sequencing data (B), and KEGG metabolic pathways from shotgun sequencing data (C) when separated by Jaccard distance and sorted by the presence of *C. perfringens* virulence factors from the VFDB. Relative abundance of three metabolic groups significantly varies in the presence of *C. perfringens* (D).

Discussion:

Overall, these results indicate that developmental stage, location, and pathogen presence are structuring the gut microbiome in this species. The taxonomic profile of

this community appears similar to previously published penguin microbiomes, though the gut microbiome of the Galapagos penguin notably appears dominated by *Fusobacteria* and lacking in *Bacteroidetes* (Dewar et al. 2013, 2017, Lee et al. 2019). We found adult Galapagos penguins have significantly higher alpha diversity (observed richness, Shannon Diversity, and Simpson's Index) in their gut microbiomes compared to juvenile penguins. Differences in foraging behaviors between adults and juveniles may be an explanatory factor, as juveniles often return to the nest area as they gradually learn how to forage and adults occasionally demonstrate extended parental care during this learning period by feeding fully-fledged juveniles (Borboroglu and Boersma 2013, Boersma et al. 2017). Hormonal differences could also play a role since some adults were in breeding condition – this is known to increase alpha diversity in males in other avian species – but surprisingly, no sex-based differences were found (Escallón et al. 2019). The lack of sex-based differences may again be related to foraging habits, as males and females exhibit similar foraging behaviors (Steinfurth et al. 2008). It is important to understand how factors such as developmental stage can influence the avian microbiome, as this could indicate a varied degree of vulnerability to disease depending on host age.

Though location does not appear to be a strong force structuring the microbiome, both alpha and beta diversity show some differences at the Marielas Islets compared to the other primary sampling sites. The similar environment along the coast of Isabela and the movement of penguins between sites are likely factors behind the general homogeneity of microbiomes between locations; for example, during our short sampling period we

recaptured some of the same individuals at multiple sites. The small differences observed may be explained by unique diets between sites or exposure to different environmental microbes. Perhaps significantly, the Sierra Negra shield volcano on Isabela Island was erupting during the 2018 sampling trip. Lava flowed down the volcano's northwestern flank and reached the sea near the Marielas sampling site (Vasconez et al. 2018). This contributed to warmer water temperatures at that site and likely altered pH levels. The proximity to volcanic activity may have led to the slight variation in microbial signatures found between sites. Differing amounts of environmental heavy metal may also play a role, as this is known to alter microbiome compositions in some systems – a previous study found variation in heavy metal concentrations in Galapagos penguin feathers from different sites, with significantly higher levels of lead in feathers from Marielas (Jiménez-Uzcátegui et al. 2017, Xia et al. 2018). An additional factor may be the relatively exposed position of the southern sites compared to the more sheltered location of the Marielas away from the primary current.

We determined that the putative resistome for this species contains at least nine resistance genes. Two resistance genes associated with *Helicobacter pylori* were almost ubiquitous, detected in all but one penguin. Antibiotic resistant genes have been previously found in the Galapagos Islands in marine water, tortoise feces, and both land iguana and marine iguana feces, but this is the first time to our knowledge that they have been detected in Galapagos penguins (Wheeler et al. 2012, Overbey et al. 2015, Nieto-Claudin et al. 2019). Some antibiotic resistance occurs naturally, in areas as

remote as Antarctica, and resistance genes could potentially be found in Galapagos even in the absence of human activity (Marcelino et al. 2019). However, the increasing amount of antibiotic resistance found in the wild is broadly attributed to selection from the heavy use of antibiotics in modern agricultural and clinical settings (Ventola 2015). In one example of likely anthropogenic effects, a study in Galapagos found that proximity to humans (e.g. ports, towns) was generally associated with the antibiotic resistance found in seawater and reptile feces, with increased resistance detected in most populated areas and no resistance detected at certain isolated sites (Wheeler et al. 2012).

The exchange of antibiotic resistance genes happens readily among bacteria through horizontal gene transfer, making it challenging to prevent resistance genes from spreading. Antibiotic resistance is found at high levels in bacteria from human waste, and even after waste treatment sewage remains a potent source of antibiotics or resistance genes (Grehs et al. 2021). Resistance genes can even be transferred by wildlife across large distances, such as through resident flora of migratory birds (Marcelino et al. 2019). In Galapagos, the release of sewage into the ocean from towns and boats is a likely way in which antibiotics and/or bacteria with resistance genes could be introduced into the environment (which can also be detrimental to human health), though resistance genes may also arrive from other sources (Overbey et al. 2015). Wastewater control is thus an essential factor to limit the spread of antibiotic resistance in wild communities. Finding antibiotic resistance consistently in the penguin samples indicates that it is present even in more isolated areas along Isabela Island, emphasizing

a need for further investigation into the extent of resistance genes associated with anthropogenic activity (such as cruise ships) and whether wastewater management changes should be considered in the islands.

Finally, our pathogen screening highlighted a few potential enteric pathogens – *Clostridium perfringens, Paeniclostridium sordelli, Clostridium baratii, Gallibacterium anatis,* and *Photobacterium damselae* were among the most common bacteria linked with virulence-associated protein families in this microbiome. The class *Clostridia* contains some of the primary agents of enteric disease in birds, and *C. perfringens, P. sordellii* and *C. baratii* have all been associated with enteric disease (Crespo et al. 2013, Cooper et al. 2013). *Gallibacterium anatis* (previously *Pasteurella anatis*) has also been implicated as a pathogen in several avian species and is considered an emerging poultry disease (Varan Singh and R Singh 2015). However, the detection of virulence-associated factors is by itself no guarantee that a microbe is actually pathogenic to the host, and microbes such as *Gallibacterium anatis* are also commonly found as members of the normal flora (Persson and Bojesen 2015). Further, several bacteria in the microbiome were associated with virulence proteins but are unlikely to be pathogenic to penguins – for example, *Photobacterium damselae* was linked to several different virulence-associated protein families in this microbiome, though this bacterium is recognized as a pathogen in taxa such as fish and marine mammals rather than birds (Rivas et al. 2013).

Clostridium perfringens was the only putative pathogen detected with the Virulence Factor Database in addition to the Protein Family Database (Pfam). *C. perfringens* is an extremely widespread bacterium and a normal member of many microbiomes, but it is

also notorious for causing necrotic enteritis in poultry and other birds (as well as enteric diseases in humans, dogs, and a number of other taxa) (Petit et al. 1999, Kiu and Hall 2018). In poultry, *C. perfringens* injects toxins into the intestines, resulting in intestinal lesions and a range of clinical signs including lethargy, loss of appetite, and mortality (though lethal cases may also occur without any observable symptoms) (Cooper et al. 2013). Outbreaks of *C. perfringens* leading to fatalities have also been documented in captive penguin populations (Penrith et al. 1994). The pathogenicity of the detected strain of *C. perfringens* in Galapagos penguins is unknown, and we notably did not detect the *netB* gene encoding a pore-forming toxin which is associated with most occurrences of necrotic enteritis in poultry (Lepp et al. 2010). However, the detected virulence factors correspond to type A and the genes *plc* and *cpe*, which are associated with toxin production and avian enteric disease (Petit et al. 1999). Coupled with the apparent dominance of *C. perfringens* in the observed microbial communities and the strong structural changes observed in the presence of virulence factors, this suggests a level of pathogenicity at the time of sampling. Resampling the population is necessary to shed light on the role *C. perfringens* plays in this species' microbiome.

Conclusions:

This work establishes a baseline microbiome for an Endangered penguin, identifies two primary drivers of microbial community structure, and emphasizes the importance of minimizing interaction between wildlife and humans. Even in a place as remote and well-protected as the Galapagos Islands, human influence is still visible through factors such as antibiotic resistance genes. The human-inhabited islands also have some

domesticated animals, which increases the possibility of disease spillover occurring between domestic and wild species – the apparent pathogenicity of *C. perfringens* found in Galapagos penguins is concerning when considering the proximity of the penguin population to domestic chickens (Gottdenker et al. 2005). Thus, monitoring and limiting anthropogenic effects on wildlife is critical to the continued long-term preservation of Galapagos endemic species.

Acknowledgments:

We are very grateful to our collaborators and assistants who helped in this research, particularly to Jane Merkel and Dr. Maricruz Jaramillo for their valuable contributions through sample collection and laboratory work. We thank the Ecuadorian Ministry of the Environment and Water and the Galapagos National Park Directorate for granting research permissions. We are grateful to the Charles Darwin Foundation and to the Agency for the Regulation and Control of Biosecurity and Quarantine for Galapagos for their support. This research was supported in part by the University of Michigan Medical School Microbiome Core. This publication is contribution number 2405 of the Charles Darwin Foundation for the Galapagos Islands.

Funding:

Funding support for this project was generously provided by the NSF Graduate Research Fellowship #1945995, Saint Louis Zoo Field Research for Conservation Grant #2018.02, Saint Louis Zoo WildCare Institute, Des Lee Collaborative Vision, NIH Grant #AI137984, Harris World Ecology Center, and by the donors of the Galapagos Penguin, Flightless Cormorant and Waved Albatross Project at the Charles Darwin Foundation: The Leona M. and Harry B. Helmsley Charitable Trust, Penguin Fund of Japan, and Mr. Seishi Sakamoto.

Ethics approval:

Sampling procedures and sample export were approved by UMSL´s Institutional Animal Care and Use Committee (Protocol #1211796), USDA (Permit #47418), Galapagos

National Park Directorate (#PC-05-18), the Agency for the Regulation and Control of Biosecurity and Quarantine for Galapagos (#ABG-CT-2019-0019-O), and the Ecuadorian Ministry of the Environment and Water (Contract #MAE-DNB-CM-2016-0043, Export Authorization #128-2019-EXP-CM-MBI-DNB/MA).

References:

Akst, E. P., Boersma, P. D. and Fleischer, R. C. 2002. A comparison of genetic diversity between the Galapagos Penguin and the Magellanic Penguin. – Conserv Genet. 375–383.

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., Faltyn, M., Hernandez-Koutoucheva, A., Sharma, A. N., Bordeleau, E., Pawlowski, A. C., Zubyk, H. L., Dooley, D., Griffiths, E., Maguire, F., Winsor, G. L., Beiko, R. G., Brinkman, F. S. L., Hsiao, W. W. L., Domselaar, G. V. and McArthur, A. G. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. – Nucleic Acids Res 48: D517–D525.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. – Austral Ecol 26: 32–46.

Anderson, M. J. and Walsh, D. C. I. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? – Ecol Monogr 83: 557–574.

Bahrndorff, S., Alemu, T., Alemneh, T. and Lund Nielsen, J. 2016. The microbiome of animals: implications for conservation biology. – Int J Genomics <https://doi.org/10.1155/2016/5304028>.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. – J Comput Biol 19: 455–477.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. – J R Stat Soc Series B Stat Methodol 57: 289–300.

Boersma, P. D., Cappello, C. D. and Merlen, G. 2017. First observations of post-fledging care in Galapagos Penguins (*Spheniscus mendiculus*). – Wilson J Ornithol 129: 186–191.

Bolger, A. M., Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. – Bioinformatics 30: 2114–2120.

Bollmer, J. L., Vargas, F. H. and Parker, P. G. 2007. Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*). – Immunogenetics 59: 593–602.

Bolnick, D. I., Snowberg, L. K., Caporaso, J. G., Lauber, C., Knight, R. and Stutz, W. E. 2014a. Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. – Mol Ecol 23: 4831–4845.

Bolnick, D. I., Snowberg, L. K., Hirsch, P. E., Lauber, C. L., Org, E., Parks, B., Lusis, A. J., Knight, R., Caporaso, J. G. and Svanbäck, R. 2014b. Individual diet has sex-dependent effects on vertebrate gut microbiota. – Nat Commun 5: 4500.

Borboroglu, P. G. and Boersma, P. D.Eds. 2013. Penguins: Natural History and Conservation. – University of Washington Press.

Brown, S. P., Cornforth, D. M. and Mideo, N. 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. – Trends Microbiol 20: 336–342.

Cappello, C. and Boersma, P. 2018. Sexing Galápagos penguins *Spheniscus mendiculus* by morphological measurements. – Endang. Species. Res. 35: 169–173.

Carmody, R. N., Gerber, G. K., Luevano, J. M., Gatti, D. M., Somes, L., Svenson, K. L. and Turnbaugh, P. J. 2015. Diet dominates host genotype in shaping the murine gut microbiota. – Cell Host Microbe 17: 72–84.

Cooper, K. K., Songer, J. G. and Uzal, F. A. 2013. Diagnosing clostridial enteric disease in poultry. – J Vet Diagn 25: 314–327.

Crespo, R., Franca, M. and Shivaprasad, H. L. 2013. Ulcerative enteritis-like disease associated with *Clostridium sordellii* in quail. – Avian Dis 57: 698–702.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J. and Turnbaugh, P. J. 2014. Diet rapidly and reproducibly alters the human gut microbiome. – Nature 505: 559–63.

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., Olsen, G. J., Olson, R. D., Overbeek, J. C., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomas, C., VanOeffelen, M., Vonstein, V., Warren, A. S., Xia, F., Xie, D., Yoo, H. and Stevens, R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. – Nucleic Acids Res 48: D606–D612.

Deem, S. L., Merkel, J., Ballweber, L., Vargas, F. H., Cruz, M. B. and Parker, P. G. 2010. Exposure to *Toxoplasma gondii* in Galapagos Penguins (*Spheniscus mendiculus*) and Flightless Cormorants (*Phalacrocorax harrisi*) in the Galapagos Islands, Ecuador. – J Wildl Dis 46: 1005–1011.

Dewar, M. L., Arnould, J. P. Y., Dann, P., Trathan, P., Groscolas, R. and Smith, S. 2013. Interspecific variations in the gastrointestinal microbiota in penguins. – MicrobiologyOpen 2: 195–204.

Dewar, M. L., Arnould, J. P. Y., Allnutt, T. R., Crowley, T., Krause, L., Reynolds, J., Dann, P. and Smith, S. C. 2017. Microbiota of little penguins and short-tailed shearwaters during development. – PLoS ONE 12: e0183117.

Dong, X. and Strous, M. 2019. An integrated pipeline for annotation and visualization of metagenomic contigs. – Front. Genet. <https://doi.org/10.3389/fgene.2019.00999>.

Escallón, C., Belden, L. K. and Moore, I. T. 2019. The cloacal microbiome changes with the breeding season in a wild bird. – Integr Org Biol <https://doi.org/10.1093/iob/oby009>.

Franzosa, E. A., Knight, R., Rahnavard, G., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N. and Huttenhower, C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. – Nat Methods 15: 962–968.

Gottdenker, N. L., Walsh, T., Vargas, H., Merkel, J., Jiménez, G. U., Miller, R. E., Dailey, M. and Parker, P. G. 2005. Assessing the risks of introduced chickens and their pathogens to native birds in the Galápagos Archipelago. – Biol Conserv 126: 429–439.

Grehs, B. W. N., Linton, M. A. O., Clasen, B., de Oliveira Silveira, A. and Carissimi, E. 2021. Antibiotic resistance in wastewater treatment plants: understanding the problem and future perspectives. – Arch Microbiol 203: 1009–1020.

Gu, W., Deng, X., Lee, M., Sucu, Y. D., Arevalo, S., Stryke, D., Federman, S., Gopez, A., Reyes, K., Zorn, K., Sample, H., Yu, G., Ishpuniani, G., Briggs, B., Chow, E. D., Berger, A., Wilson, M. R., Wang, C., Hsu, E., Miller, S., DeRisi, J. L. and Chiu, C. Y. 2021. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. – Nat Med 27: 115–124.

Hills, R. D., Pontefract, B. A., Mishcon, H. R., Black, C. A., Sutton, S. C. and Theberge, C. R. 2019. Gut microbiome: profound implications for diet and disease. – Nutrients <https://doi.org/10.3390/nu11071613>.

Hu, X., Liu, G., Li, Y., Wei, Y., Lin, S., Liu, S., Zheng, Y. and Hu, D. 2018. High-throughput analysis reveals seasonal variation of the gut microbiota composition within forest musk deer (*Moschus berezovskii*). – Front. Microbiol. 9: 1674.

Jiménez-Uzcátegui, G., Vinueza, R. L., Urbina, A. S., Egas, D. A., García, C., Cotín, J. and Sevilla, C. 2017. Lead and cadmium levels in Galapagos penguin *Spheniscus mendiculus*, flightless cormorant *Phalacrocorax harrisi*, and waved albatross *Phoebastria irrorata*. – Mar Ornithol 45: 159–163.

Jiménez-Uzcátegui, G., Carrera-Játiva, P. D., Sarzosa, S., Encalada, E., Rodríguez-Hidalgo, R. and Sevilla, C. 2021. Normal physical parameters characterizing Galapagos marine birds. – Mar Ornithol 49: 257–263.

Kiu, R. and Hall, L. J. 2018. An update on the human and animal enteric pathogen *Clostridium perfringens*. – Emerg Microbes Infect <https://doi.org/10.1038/s41426-018-0144-8>.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina Sequencing Platform. – Appl. Environ. Microbiol. 79: 5112–5120.

Krismer, B., Weidenmaier, C., Zipperer, A. and Peschel, A. 2017. The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. – Nat Rev Microbiol 15: 675–687.

Langmead, B. and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. – Nat Methods 9: 357–359.

Lazar, V., Ditu, L.-M., Pircalabioru, G. G., Gheorghe, I., Curutiu, C., Holban, A. M., Picu, A., Petcu, L. and Chifiriuc, M. C. 2018. Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. – Front. Immunol. 9: 1830.

Lee, W. Y., Cho, H., Kim, M., Tripathi, B. M., Jung, J.-W., Chung, H. and Kim, J.-H. 2019. Faecal microbiota changes associated with the moult fast in chinstrap and gentoo penguins. – PLoS ONE 14: e0216565.

Lepp, D., Roxas, B., Parreira, V. R., Marri, P. R., Rosey, E. L., Gong, J., Songer, J. G., Vedantam, G. and Prescott, J. F. 2010. Identification of novel pathogenicity loci in *Clostridium perfringens* strains that cause avian necrotic enteritis. – PLoS One <https://doi.org/10.1371/journal.pone.0010795>.

Levin, I., Outlaw, D., Vargas, F. and Parker, P. 2009. *Plasmodium* blood parasite found in endangered Galapagos penguins (*Spheniscus mendiculus*). – Biol Conserv 142: 3191–3195.

Liu, B., Zheng, D., Jin, Q., Chen, L. and Yang, J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. – Nucleic Acids Res 47: D687–D692.

Marcelino, V. R., Wille, M., Hurt, A. C., González-Acuña, D., Klaassen, M., Schlub, T. E., Eden, J.-S., Shi, M., Iredell, J. R., Sorrell, T. C. and Holmes, E. C. 2019. Meta-transcriptomics reveals a diverse antibiotic resistance gene pool in avian microbiomes. – BMC Biology 17: 31.

Markle, J. G. M., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K. D., Macpherson, A. J. and Danska, J. S. 2013. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. – Science 339: 1084–1088.

McMurdie, P. J. and Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. – PLoS ONE 8: e61217.

Merkel, J., Jones, H. I., Whiteman, N. K., Gottdenker, N., Vargas, H., Travis, E. K., Miller, R. E. and Parker, P. G. 2007. Microfilariae in Galàpagos penguins (*Spheniscus mendiculus*) and flightless cormorants (*Phalacrocorax harrisi*): genetics, morphology, and prevalence. – J Parasitol 93: 495–503.

Michel, A. J., Ward, L. M., Goffredi, S. K., Dawson, K. S., Baldassarre, D. T., Brenner, A., Gotanda, K. M., McCormack, J. E., Mullin, S. W., O'Neill, A., Tender, G. S., Uy, J. A. C., Yu, K., Orphan, V. J. and Chaves, J. A. 2018. The gut of the finch: uniqueness of the gut microbiome of the Galápagos vampire finch. – Microbiome 6: 167.

Mikheenko, A., Saveliev, V. and Gurevich, A. 2016. MetaQUAST: evaluation of metagenome assemblies. – Bioinformatics 32: 1088–1090.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D. and Bateman, A. 2021. Pfam: The protein families database in 2021. – Nucleic Acids Res 49: D412–D419.

Nieto-Claudin, A., Esperón, F., Blake, S. and Deem, S. L. 2019. Antimicrobial resistance genes present in the faecal microbiota of free-living Galapagos tortoises (*Chelonoidis porteri*). – Zoonoses Public Health 66: 900–908.

Nims, B. D., Vargas, F. H., Merkel, J. and Parker, P. G. 2008. Low genetic diversity and lack of population structure in the endangered Galápagos penguin (*Spheniscus mendiculus*). – Conserv Genet 9: 1413–1420.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H. 2019. vegan: Community Ecology Package.

Overbey, K. N., Hatcher, S. M. and Stewart, J. R. 2015. Water quality and antibiotic resistance at beaches of the Galápagos Islands. – Front. Environ. Sci. <https://doi.org/10.3389/fenvs.2015.00064>.

Penrith, M., Huchzermeyer, F. W., De Wet, S. C. and Penrith, M. J. 1994. Concurrent infection with *Clostridium* and *Plasmodium* in a captive king penguin *Aptenodytes patagonicus*. – Avian Pathol 23: 373–380.

Persson, G. and Bojesen, A. M. 2015. Bacterial determinants of importance in the virulence of *Gallibacterium anatis* in poultry. – Vet Res <https://doi.org/10.1186/s13567-015-0206-z>.

Petit, L., Gibert, M. and Popoff, M. R. 1999. *Clostridium perfringens*: toxinotype and genotype. – Trends Microbiol 7: 104–110.

R Core Team. 2020. R: A language and environment for statistical computing. – Foundation for Statistical Computing, Vienna, Austria.

Rivas, A. J., Lemos, M. L. and Osorio, C. R. 2013. *Photobacterium damselae* subsp. *damselae*, a bacterium pathogenic for marine animals and humans. – Front Microbiol <https://doi.org/10.3389/fmicb.2013.00283>.

Roggenbuck, M., Bærholm Schnell, I., Blom, N., Bælum, J., Bertelsen, M. F., Pontén, T. S., Sørensen, S. J., Gilbert, M. T. P., Graves, G. R. and Hansen, L. H. 2014. The microbiome of New World vultures. – Nat Commun 5: 5498.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. and Weber, C. F. 2009. Introducing mothur: open-source, platform-

independent, community-supported software for describing and comparing microbial communities. – Appl Environ Microbiol 75: 7537–7541.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. – Front Zool 2: 16.

Song, S. J., Amir, A., Metcalf, J. L., Amato, K. R., Xu, Z. Z., Humphrey, G. and Knight, R. 2016. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. – mSystems 1: e00021-16.

Song, S. J., Sanders, J. G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., Mazel, F., Lutz, H. L., Winker, K., Graves, G. R., Humphrey, G., Gilbert, J. A., Hackett, S. J., White, K. P., Skeen, H. R., Kurtis, S. M., Withrow, J., Braile, T., Miller, M., McCracken, K. G., Maley, J. M., Ezenwa, V. O., Williams, A., Blanton, J. M., McKenzie, V. J. and Knight, R. 2020. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. – mBio 11: e02901-19.

Steinfurth, A., Vargas, F., Wilson, R., Spindler, M. and Macdonald, D. 2008. Space use by foraging Galápagos penguins during chick rearing. – Endang. Species. Res. 4: 105–112.

Tasnim, N., Abulizi, N., Pither, J., Hart, M. M. and Gibson, D. L. 2017. Linking the gut microbial ecosystem with the environment: does gut health depend on where we live? – Front Microbiol <https://doi.org/10.3389/fmicb.2017.01935>.

Taylor, M. J., Mannan, R. W., U'Ren, J. M., Garber, N. P., Gallery, R. E. and Arnold, A. E. 2019. Age-related variation in the oral microbiome of urban Cooper's hawks (*Accipiter cooperii*). – BMC Microbiology 19: N.PAG-N.PAG.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Zech Xu, Z., Jiang, L., Haroon, M. F., Kanbar, J., Zhu, Q., Jin Song, S., Kosciolek, T., Bokulich, N. A., Lefler, J., Brislawn, C. J., Humphrey, G., Owens, S. M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J. A., Clauset, A., Stevens, R. L., Shade, A., Pollard, K. S., Goodwin, K. D., Jansson, J. K., Gilbert, J. A. and Knight, R. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. – Nature 551: 457–463.

Travis, E. K., Vargas, F. H., Merkel, J., Gottdenker, N., Miller, R. E. and Parker, P. G. 2006. Hematology, serum chemistry, and serology of Galápagos penguins (*Spheniscus mendiculus*) in the Galápagos Islands, Ecuador. – J Wildl Dis 42: 625–632.

Trevelline, B. K., Fontaine, S. S., Hartup, B. K. and Kohl, K. D. 2019. Conservation biology needs a microbial renaissance: a call for the consideration of host-associated microbiota in wildlife management practices. – Proc. R. Soc. B 286: 20182448.

van der Walt, A. J., van Goethem, M. W., Ramond, J.-B., Makhalanyane, T. P., Reva, O. and Cowan, D. A. 2017. Assembling metagenomes, one community at a time. – BMC Genomics 18: 521.

Varan Singh, S. and R Singh, B. 2015. *Gallibacterium anatis*: An emerging pathogen of poultry birds and domiciled birds. – J Vet Sci Technol <https://doi.org/10.4172/2157-7579.1000324>.

Vargas, H., Lougheed, C. and Snell, H. 2005. Population size and trends of the Galápagos Penguin *Spheniscus mendiculus*. – Ibis 147: 367–374.

Vargas, F. H., Harrison, S., Rea, S. and Macdonald, D. W. 2006. Biological effects of El Niño on the Galápagos penguin. – Biol Conserv 127: 107–114.

Vasconez, F. J., Ramón, P., Hernandez, S., Hidalgo, S., Bernard, B., Ruiz, M., Alvarado, A., Femina, P. L. and Ruiz, G. 2018. The different characteristics of the recent eruptions of Fernandina and Sierra Negra volcanoes (Galápagos, Ecuador). – Volcanica 1: 127–133.

Vemuri, R., Sylvia, K. E., Klein, S. L., Forster, S. C., Plebanski, M., Eri, R. and Flanagan, K. L. 2019. The microgenderome revealed: sex differences in bidirectional interactions between the microbiota, hormones, immunity and disease susceptibility. – Semin Immunopathol 41: 265–275.

Ventola, C. L. 2015. The Antibiotic Resistance Crisis. – P & T 40: 277–283.

Videvall, E., Song, S. J., Bensch, H. M., Strandh, M., Engelbrecht, A., Serfontein, N., Hellgren, O., Olivier, A., Cloete, S., Knight, R. and Cornwallis, C. K. 2020. Early-life gut dysbiosis linked to juvenile mortality in ostriches. – Microbiome 8: 147.

Waite, D. W. and Taylor, M. W. 2015. Exploring the avian gut microbiota: current trends and future directions. – Front Microbiol <https://doi.org/10.3389/fmicb.2015.00673>.

Wheeler, E., Hong, P.-Y., Bedon, L. C. and Mackie, R. I. 2012. Carriage of antibiotic-resistant enteric bacteria varies among sites in Galápagos reptiles. – J Wildl Dis 48: 56–67.

Wood, D. E., Lu, J. and Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. – Genome Biol 20: 257.

Xia, J., Jin, C., Pan, Z., Sun, L., Fu, Z. and Jin, Y. 2018. Chronic exposure to low concentrations of lead induces metabolic disorder and dysbiosis of the gut microbiota in mice. – Sci Total Environ 631–632: 439–448.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F. O. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. – Nucleic Acids Res. 42: D643–D648.

CHAPTER 2.

Seasonal shifts in the function and composition of the Eurasian tree sparrow gut microbiome

Department of Biology and Whitney R. Harris World Ecology Center, University of Missouri-St. Louis, One University Blvd., St. Louis, MO 63121, USA

Abstract:

Vertebrates are extensively colonized by bacteria, forming microbial communities which can be closely tied to host health. The importance of host-associated microbial communities is increasingly recognized and has been studied extensively in recent years, but many questions remain about how these communities form in wild avian populations and how stable they remain over time. Using a locally abundant introduced species, the Eurasian tree sparrow (*Passer montanus*), we collected fecal samples from birds at two sites in Missouri to assess variation in the gut microbiome from spring through fall over two years. We used a combination of targeted amplicon sequencing and whole genome sequencing to examine bacterial community variation over time. Alpha diversity results showed a non-significant downward trend in Shannon diversity from spring to fall at both sites. Beta diversity based on both ASVs and predicted metabolic pathways significantly clustered between seasons, indicating differing

composition and function in the gut community based on season. Differential abundance analysis showed a high number of ASVs significantly decreasing in fall compared to spring, including many members of *Bacilli* and *Alphaproteobacteria.* Finally, we detected a large number of virulence-associated genes in samples across all seasons, indicating that the sparrow microbiome contains a relatively large number of likely pathogenic bacteria throughout much of the year. These results shed light on temporal stability in the wild avian microbiome, indicating that microbial communities in birds are dynamic and subject to structural changes across seasons.

Keywords: Gut microbiome, avian microbiome, seasonal variation, sparrows

Original Article to be submitted.

Introduction:

The microbiome has been linked to host health in many vertebrates (Hills et al. 2019). Resident microbes often carry out key functions for the host such as synthesizing vitamins or preventing pathogen colonization (Rowland et al. 2018). Microbial diversity is an important factor, typically measured in terms of species richness and evenness within a given microbial community, and high diversity within a host-associated microbiome is often associated with better health outcomes for the host (Videvall et al. 2020). However, whether resident microbial communities remain stable over time is still unclear in many wild systems. Seasonal shifts in host-associated microbiomes have been studied the most in mammals, revealing diet-associated patterns occurring in seasonal cycles across several mammalian species, but similar research in other taxa such as birds lags far behind (Bergmann et al. 2015, Gong et al. 2021). Given the importance of microbial mutualists to health-related processes (Hills et al. 2019, Sun et al. 2022), understanding community stability over time is an essential foundational step in wild microbiomes.

Life history changes are responsible for some of the cyclical changes seen in certain taxa (Markle et al. 2013, Cao et al. 2020, Si et al. 2022). For example, the hormonal changes associated with reproductive cycles can influence microbial composition (Markle et al. 2013, Escallón et al. 2019). One study found that microbial diversity increased in the cloacal microbiomes of male rufous-collared sparrows (*Zonotrichia capensis*) as they went into breeding condition, with diversity returning to normal after the reproductive season, a pattern which correlated with testosterone levels (Escallón et al. 2019).

Migration is also associated with changes in microbial composition, as animals are exposed to different environments and dietary sources (Cao et al. 2020, Obrochta et al. 2022). For example, migratory Canada geese (*Branta canadensis*) have lower gut microbiome diversity but higher abundances of putatively beneficial microbes compared to urban geese (Obrochta et al. 2022).

Other cyclical changes in vertebrate-associated microbial communities may be simply due to the differences in diet which accompany changing seasons, although microbial patterns are inconsistent (Hu et al. 2018, Gong et al. 2021). Some studies have found that changes in diet can prompt significant alterations in microbial communities of wild hosts such as the North American bison (*Bison bison*) (Bergmann et al. 2015) and yaks (*Bos grunniens*) (Guo et al. 2021). Season-associated patterns have been identified in mammals such as forest musk deer (*Moschus berezovskii*), which have higher bacterial diversity in winter and spring when they consume dry leaf diets than they do in summer and fall with fresh leaf diets (Hu et al. 2018). In great evening bats (*Ia io*), dietary shifts from insects in the summer to birds in the fall lead to microbiomes clustering by season (an indication of compositional changes) and significantly higher abundances of certain microbes; however, microbial diversity levels remain similar between summer and fall (Gong et al. 2021). More work is needed to determine how prevalent these seasonal patterns of microbial variation are in avian microbiomes and to understand how predictable these patterns might be across systems.

For this study, we examined seasonal variation from spring to fall (April to November) in a locally abundant avian species, the Eurasian tree sparrow (*Passer montanus*). The

species was introduced to St. Louis approximately 150 years ago and now has a patchy distribution extending North of the St. Louis area along the Mississippi River (Jackson 2003). Many of our predictions rely on alpha diversity, which is measured as the richness and/or evenness of microbial species within each sample, or beta diversity, which measures the community similarities between samples. We expected to see the highest alpha diversity in summer due to greater foraging opportunities, as a more diverse diet generally leads to a more diverse microbiome (Heiman and Greenway 2016), and we predicted seasonal clustering based on community composition. We expected to see differentially abundant bacteria between seasons. Finally, since several studies have found higher rates of bacterial infections in younger birds (Benskin et al. 2009, Minich et al. 2022), we expected putative pathogens would be significantly more prevalent in young birds, and thus would be more prevalent during summer when juvenile tree sparrows are common.

Methods:

Field sampling:

We captured juvenile and adult Eurasian tree sparrows using mist nets at two sites during spring (April – mid June), summer (late June – early September), and fall (late September – November) in 2020 and 2021 (n = 100). The sites are both in Saint Louis County, Missouri, with one at the Columbia Bottom Conservation Area (Columbia Bottom CA) visitor center and the other approximately 20 miles away at a residential site in Webster Groves. Bird feeders with similar feed mixes were present at each

sample site, but the Columbia Bottom CA site is adjacent to undeveloped forests and fields while the Webster Groves site is entirely residential, likely resulting in different food sources between sampling sites. While efforts were made to sample evenly across seasons with a buffer period between each season, factors such as site access, uneven capture rates, and unpredictable DNA yields resulted in uneven sample numbers between seasons. This was most notable at Columbia Bottom CA, where we were unable to collect samples earlier than June in either year. We banded each bird to identify recaptured individuals and collected basic morphological measurements including wing chord, mass, and bill length. We collected feces from each bird by placing them individually into bags with alcohol-swabbed double floors, which allows feces to fall through the wire mesh floor onto the clean bag floor below (Knutie and Gotanda 2018). Samples were placed on ice immediately and taken to a freezer after field sampling was complete for each day.

16S rRNA gene sequencing and data processing:

Fecal sample DNA was extracted with QIAamp PowerFecal Pro DNA Kits following manufacturer instructions and quantified using a Qubit fluorometer. Samples containing at least 20ng of DNA were sequenced in three separate runs by the University of Michigan Medical School Microbiome Core with 16S rRNA gene sequencing, using the methods and V4 region primers described in detail in Kozich et al. 2013 (Kozich et al. 2013). Negative sampling, extraction, and sequencing controls were included with the runs. Positive sequencing controls were included with each run by the sequencing

company (Zymo Microbial Community Standard), and positive extraction controls were also included in the second and third runs (Zymo Gut Microbiome Standard).

Sequencing reads were imported into R (version 4.2.0) using the program Dada2, then filtered and classified following the Dada2 pipeline (Callahan et al. 2016 p. 2). Recommended Dada2 steps for quality filtering and trimming were followed, and merged reads <250bp and >256bp long were discarded. SILVA v138 was used as a reference to assign taxonomies to amplicon sequence variants (ASVs) (Yilmaz et al. 2014). Using the package phyloseq, further read filtering was performed to remove any reads that matched mitochondria, chloroplasts, or Archaea (McMurdie and Holmes 2013). Controls revealed negligible contamination and all expected bacteria >1% were detected by the pipeline following processing. Sample reads were filtered for likely contaminants using the package decontam (Davis et al. 2018), and two ASVs were flagged as contaminants and removed from all samples. A minimum threshold of 4000 reads was then applied, which excluded six samples and left a sample size of 100.

To examine community patterns in alpha and beta diversity, filtered reads were then subsampled (rarefied) to 4108 reads using phyloseq to match the sample with the lowest read count following filtering steps. Simple relative abundances were calculated from the rarefied data by dividing the total read number of each taxon per group by the total reads in the sample (4108) using the R packages phyloseq and tidyverse, then used to create plots with phyla and families to visualize community diversity. Phyloseq was used to calculate two common metrics of alpha diversity: ASV-level observed richness and Shannon diversity for each individual. Beta diversity was also measured in phyloseq

by calculating Bray-Curtis dissimilarity between each sample pair from the rarefied data to create a distance matrix. Alpha and beta diversity were visualized using phyloseq and ggplot2 (Wickham et al. 2019). A Shapiro-Wilk test was used to assess whether Shannon diversity and observed richness deviated from a normal distribution; the test was not significant for Shannon diversity ($P = 0.2993$) but was significant for observed richness ($P < 0.05$), so only Shannon diversity was used for parametric statistical tests. Shannon diversity variation between individuals grouped by season was tested using an ANOVA in R with sequencing run, age, and location included as confounding variables. PERMANOVAs in the R package vegan were used to test beta diversity variation between seasons using the previously described distance matrix while controlling for sequencing run, age, and location (Anderson 2001). PERMDISP was used for any significant PERMANOVA results to check whether dispersion was confounding the PERMANOVA test (Anderson 2017, Oksanen et al. 2019).

Differential abundance analysis was also performed using the R package DESeq2 to examine whether fall samples contain any specific ASVs that are more or less abundant compared to spring (Love et al. 2014 p. 2). Summer samples were excluded to test how samples later in the year (fall) differed from samples early in the year (spring), and the previously filtered (but not rarefied) fall and spring samples were pruned by excluding any taxa found fewer than 3 times in fewer than 5% of the samples. DESeq2 was run with default settings using spring as the reference level. The Benjamini and Hochberg method was used to perform false discovery correction, including only results with adjusted *p* values >0.01 to reduce false positives (Benjamini and Hochberg 1995).

Metagenomic sequencing and data processing:

A subset of samples was selected from each season at the two sites to use for metagenomic whole genome sequencing, with the goal of estimating potential metabolic activity and identifying likely pathogens using virulence factors. These samples were chosen based on higher DNA yield (required for metagenomic sequencing), sampling location, season, age, and sex. Approximately equal groups were obtained for each category with a total sample size of 28. The DNA was sent to the University of Michigan Medical School Microbiome Core for whole genome sequencing alongside negative and positive controls. Following sequencing, quality filtering and adapter removal was performed on the raw WGS sequences with the University of Missouri Cluster using the program BBDuk (settings: minlen=50, qtrim=rl, trimq=20, k=23, mink=11, ktrim=r) ("BBTools" in press). Contaminating host sequences and PhiX sequences were removed with BBSplit using whole-genome references obtained from GenBank. Cleaned reads were then mapped with KMA (k-mer alignment) against the Virulence Factor Database on the platform PATRIC (Pathosystems Resource Integration Center) to obtain an estimation of how many putative pathogens were present in the bacterial communities based on the presence of virulence factors (Davis et al. 2018, Liu et al. 2019, "BBTools" in press). Kraken 2 was used for taxonomic classification of WGS reads on PATRIC (Lu and Salzberg 2020).

To obtain functional annotations of the metagenomic data, reads were assembled using MetaSPAdes with default settings (Nurk et al. 2017). Reads were then mapped to contigs and resulting bam files were sorted using BBMap and Samtools (Li et al. 2009,

"BBTools" in press). Depth files were created for each sample based on sorted bam files using MetaBat, and contigs were annotated with MetaErg with depth files included (Kang et al. 2015, Dong and Strous 2019). MetaCyc pathway annotations generated by MetaErg were converted to an abundance table, normalized using HUMANn2, and imported into R for further analysis using phyloseq and vegan (Caspi et al. 2014, Franzosa et al. 2018, Dong and Strous 2019). Kruskal Wallis tests followed by the Benjamini-Hochberg correction were used to assess differential abundance in metabolic pathways between seasons, and Principal Coordinates Analysis (PCoA) based on MetaCyc pathway abundances was generated using phyloseq to visualize beta diversity variation between groups. PERMANOVA was used to assess significant variation in metabolic pathway beta diversity between seasons (Caspi et al. 2014, Anderson 2017).

Samples with parallel 16S rRNA gene and whole genome sequencing data were examined to determine whether the presence of virulence factors (pathogenic indicators) is associated with community shifts in terms of alpha or beta diversity. Binary logistic regressions were used to test whether the number of infected birds is significantly influenced by season, age, or location.

Results:

The tree sparrow microbiome was largely composed of *Firmicutes, Proteobacteria,* and *Actinobacteriota* (Fig. 1). A small number of samples had high proportions of unclassified bacteria, but for most samples the number of unclassified reads was

negligible (<1%). The gut microbiome for this species was highly diverse, and eight phyla were represented in proportions greater than 1%.

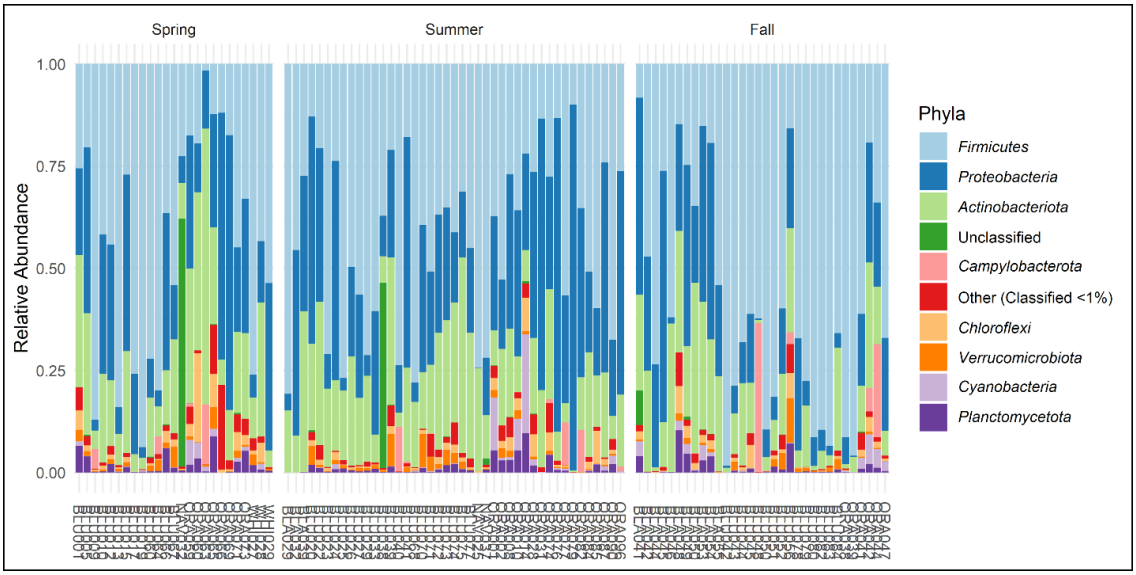

Figure 1. Bacterial composition across seasons (relative abundances).

Alpha diversity was high in this sample set, with observed richness of approximately 100-200 ASVs per individual in samples rarefied to 4108 reads, but no significant patterns were apparent based on season. Shannon diversity trended downward across seasons but did not vary significantly (Fig. 2). PERMANOVAs based on ASVs revealed significant clustering between seasons across both sites, although the results were somewhat confounded by significantly different dispersion between seasons (PERMANOVA, 9999 permutations, P = 0.0001, R^2 = 0.0593; PERMDISP, P = 0.0139; Fig. 2). The other significant variables included in the model were sex (R^2 = 0.01375, P = 0.0446), age (R^2 = 0.02726, P = 0.0001), and location (R^2 = 0.02864, P = 0.0002). When limited to the 28 samples of more balanced design used for both 16S rRNA gene sequencing and whole genome sequencing, the dispersion effect for season

disappeared, leaving a clear seasonal pattern of varying beta diversity for both sites combined (PERMANOVA, 9999 permutations, P = 0.0001, R^2 = 0.12813; PERMDISP, 9999 permutations, P = 0.6926; Fig. 3).

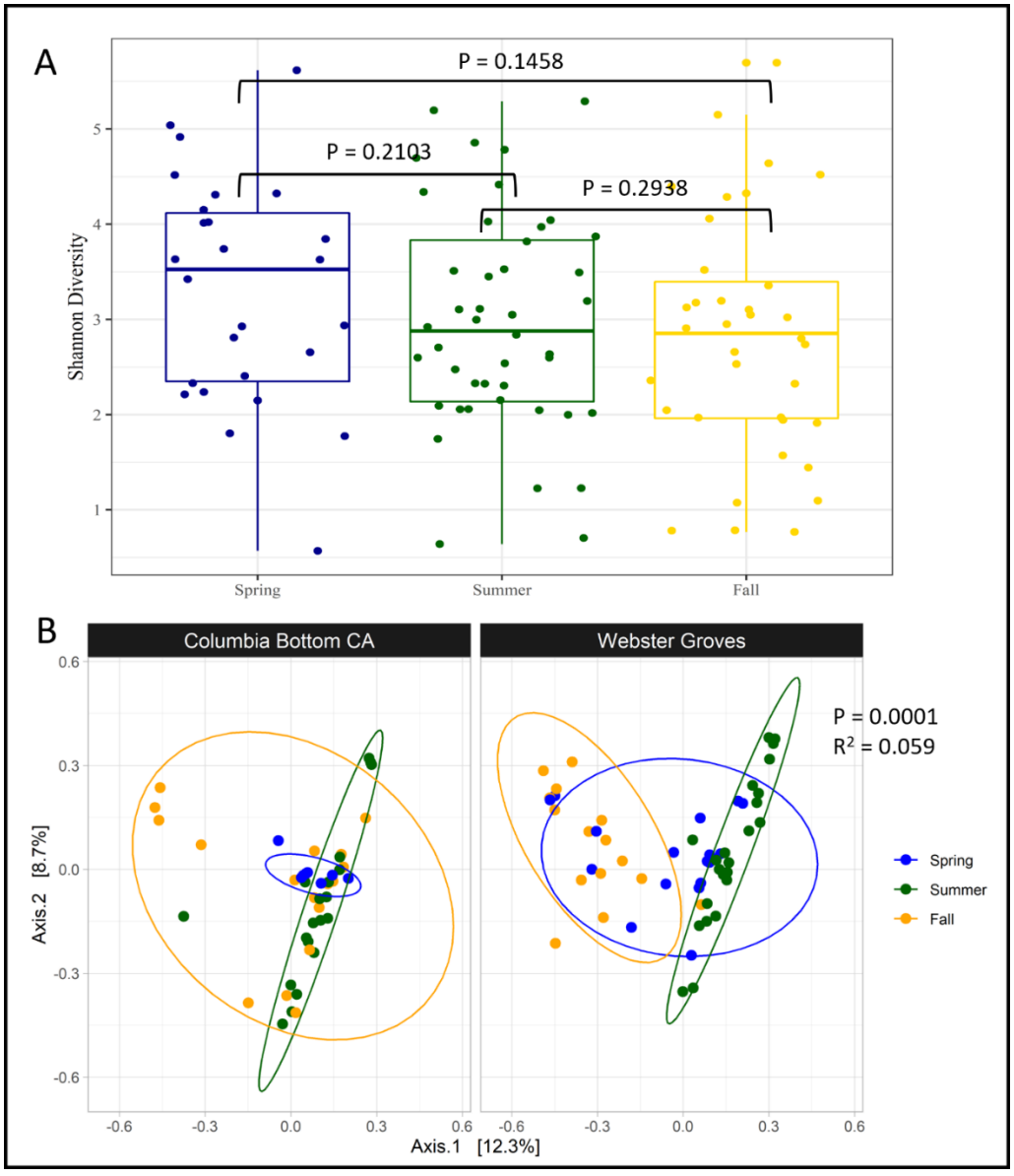

Figure 2. Alpha diversity trended downward from spring to fall, but ANOVAs were not significant (A). Beta diversity (Bray-Curtis dissimilarity) showed clustering in PCoAs based on season at each of the two sites, Columbia Bottom Conservation Area and Webster Groves (B). Beta diversity significantly varied based on

season (PERMANOVA, 9999 permutations, P = 0.0001, R^2 = 0.059), but dispersion was also significantly different (PERMDISP, 9999 permutations, P = 0.0139, B).

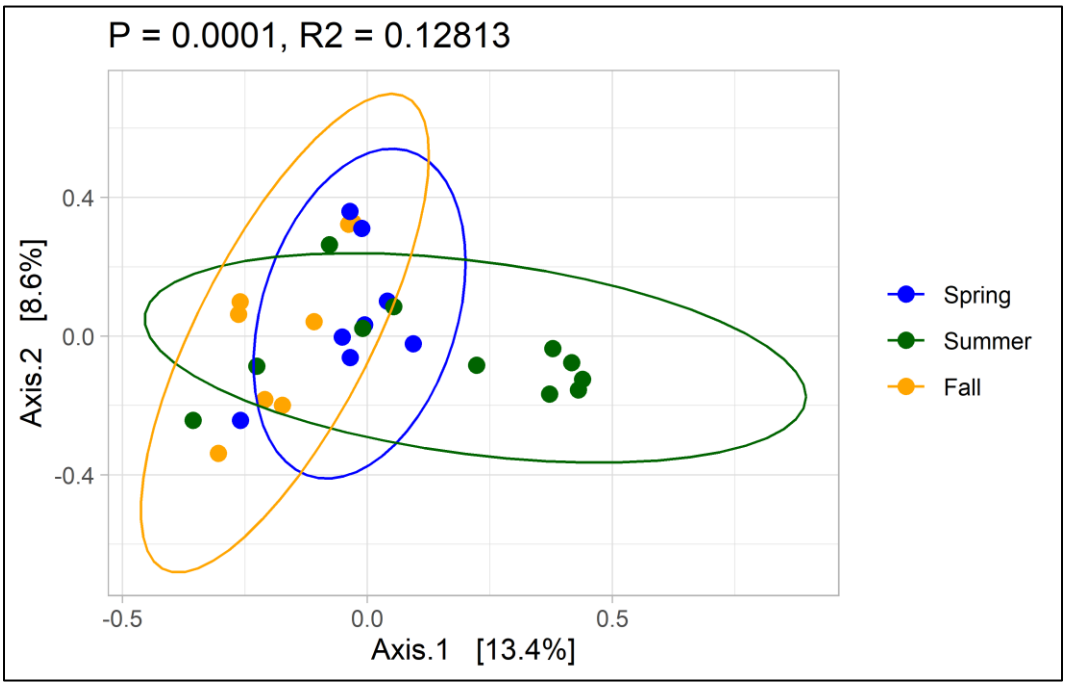

Figure 3. When limited to the 28 samples of balanced design which were used for both 16S rRNA gene sequencing and whole genome sequencing, seasons were clearly distinguished at both sites combined (PERMANOVA P = 0.0001) based on beta diversity from 16S data with no confounding dispersion effect (PERMDISP P > 0.05).

DESeq2 analysis revealed a high number of ASVs that were differentially abundant between spring and fall (Fig. 4). Of the 46 ASVs that significantly varied following false discovery correction, 13 belonged to the class *Bacilli,* 10 to *Alphaproteobacteria*, eight to *Gammaproteobacteria,* and seven to *Actinobacteria*. The remaining ASVs belonged to the classes *Clostridia*, *Planctomycetes, Thermoleophilia, Acidimicrobiia, TK10,* and *Chloroflexia.*

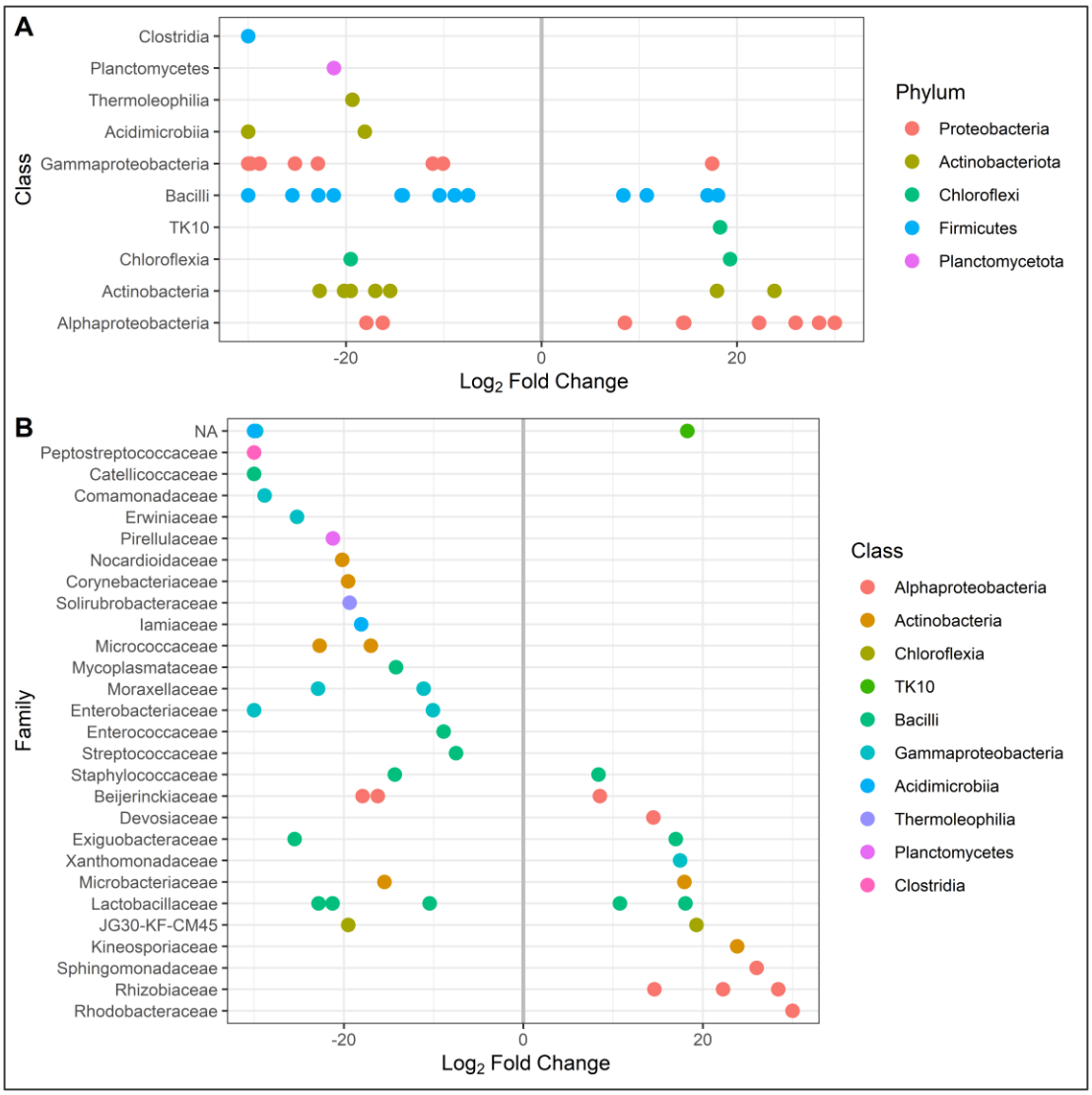

Figure 4. DESeq2 analysis revealed many ASVs that were differentially abundant in fall compared to spring. These ASVs belonged to five phyla (A) and ten classes (B). ASVs to the right of the grey bar were significantly more abundant in fall, while ASVs to the left of the grey bar were significantly less abundant in fall.

Virulence factors were detected using the VFDB in 16/27 samples (59%). The virulence factors were distributed roughly evenly across seasons. 73% of juveniles had detected virulence factors compared to 46% of adults, and 71% of birds from Webster Groves had

detected virulence factors compared to 50% at Columbia Bottom CA; however, binary logistic regression results were not significant for any variable ($P > 0.05$). These assorted virulence factors matched several reference genomes: *Escherichia coli, Mycobacterium tuberculosis, Brucella melitensis, Klebsiella pneumoniae, Salmonella enterica, Shigella dysenteriae, Staphylococcus aureus, Enterococcus faecalis,* and *Yersinia pestis*. Most of these genome matches corresponded to bacteria which were also identified using Kraken 2, but two appeared misclassified: the sample with a virulence factor matching *Yersinia pestis* had a Kraken 2 match for fellow *Yersiniaceae* member *Serratia marcescens,* while individuals with virulence factors matching *Mycobacterium tuberculosis* had Kraken 2 matches for *M. avium.* Metabolic pathways computed from whole genome sequencing data revealed clustered seasonal beta diversity similar to the seasonal beta diversity patterns shown in ASVs from 16S sequencing data (Fig. 5). Beta diversity based on metabolic pathways significantly varied between seasons (PERMANOVA, 9999 permutations, $P = 0.0305$, $R2 = 0.14450$; PERMDISP, 9999 permutations, $P = 0.1141$).

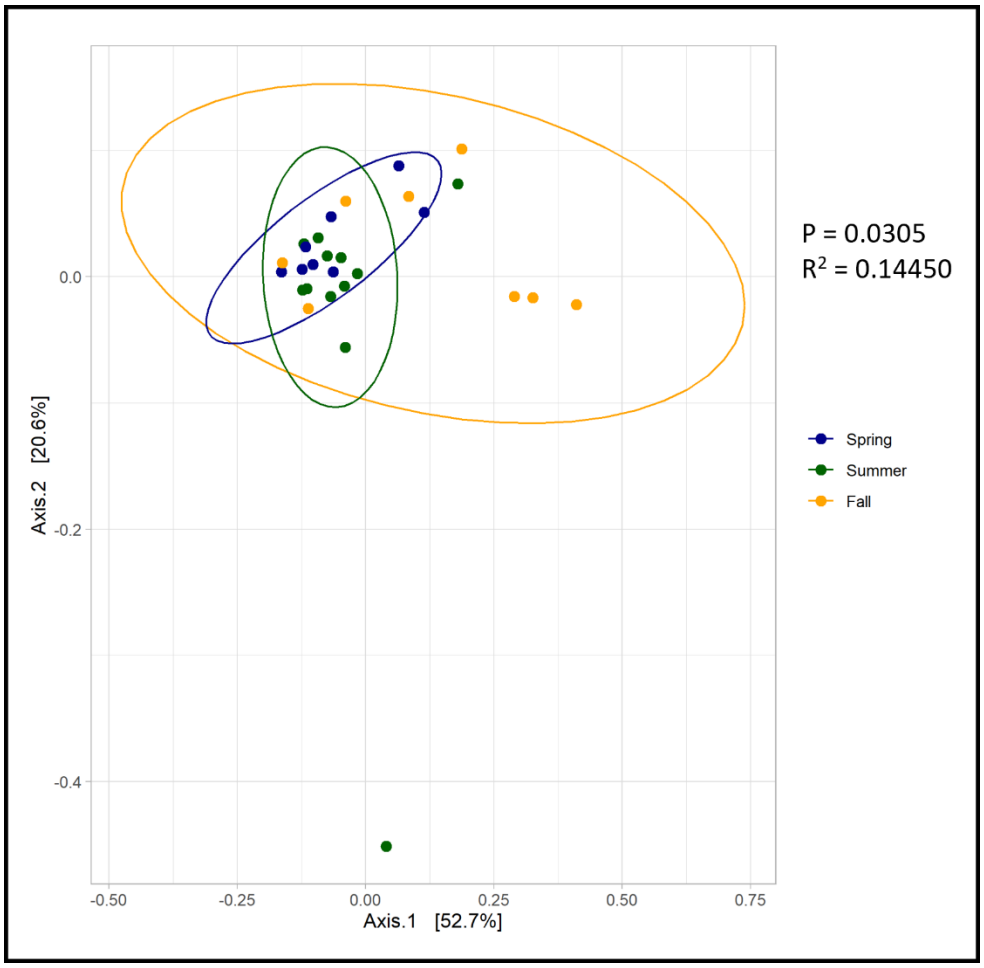

Figure 5. A PCoA based on Bray-Curtis dissimilarity (beta diversity) from predicted MetaCyc pathways revealed a season-associated clustering pattern which was similar to the seasonal clustering pattern found in the ASV data. Beta diversity significantly varied between seasons (PERMANOVA, 9999 permutations, P = 0.0305, R^2 = 0.14450; PERMDISP, 9999 permutations, P = 0.1141).

When examining ASV-based variation between groups with virulence factors versus groups without, alpha diversity was not significantly different but beta diversity varied (P = 0.0090, R2 = 0.05312). A PCoA based on ASVs visualized clustering between birds with putative bacterial pathogens and those without, although substantial overlap still appears (Fig. 6).

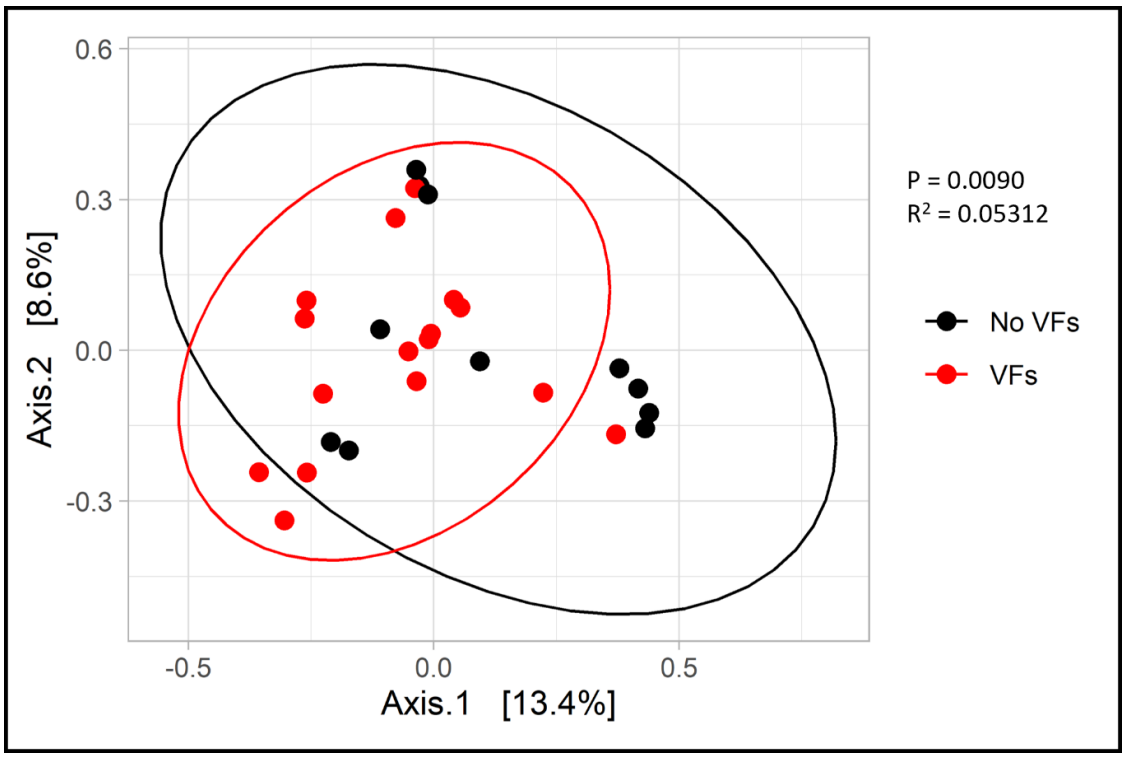

Figure 6. A PCoA based on Bray-Curtis dissimilarity (beta diversity) visualized clustering between birds with virulence factors detected (VFs) versus those with no detected virulence factors. Beta diversity between these groups significantly varied ($P = 0.0090$, $R^2 = 0.05312$).

Discussion:

In this system with high environmental heterogeneity, the gut microbiome exhibited temporal shifts at several levels of the community. Alpha diversity did not significantly vary, although Shannon diversity generally decreased from spring to fall. However, beta diversity based on ASVs significantly varied between seasons, and PCoA visualization of beta diversity revealed a clustering pattern that appears consistently within each individual site as well as in the combined sites. The discrepancy between alpha and beta diversity suggests that while the total number of microbial taxa within each sample may remain relatively consistent across seasons, differences exist between samples from

each season in terms of microbial identities and abundances. Between the two sampling sites, beta diversity clustering by season was more pronounced at Webster Groves, where we were able to begin sampling earlier in the year and to leave a larger buffer period between sampling Spring, Summer, and Fall. Many bacteria were also differentially abundant in fall compared to spring, most notably members of *Bacilli, Alphaproteobacteria, Gammaproteobacteria,* and *Actinobacteria.* Members of *Bacilli* are often associated with better health outcomes for hosts (Arif et al. 2021), and so the general decrease in abundance of *Bacilli* members in fall hints at a possibly less beneficial microbiome toward the end of the year. While the availability of feeders at each site provided some diet stability, seasonal changes in food availability from environmental sources is likely responsible for the seasonal variation seen here (Hu et al. 2018, Gong et al. 2021).

Functional microbiome results echoed the same patterns shown with community composition data. Beta diversity based on computed metabolic pathways showed similar seasonal clustering when visualized with PCoA. Somewhat surprisingly, however, the prevalence of likely pathogenic bacteria appeared relatively consistent across seasons, and no metadata factors emerged as significant drivers for pathogenicity in the microbiome. Our prediction of more gut pathogens in juveniles during the summer was thus unsupported. A relationship between pathogenic bacteria and the overall bacterial community was still seen, albeit to a lesser extent than seasonal variation, and beta diversity based on ASVs significantly differed between birds with detected virulence factors compared to seemingly uninfected birds.

Together, these results show that community composition changes in consistent ways across sites throughout the year, that many putatively beneficial microbes are present in reduced abundances toward the end of the year, and that metabolic activity in the microbiome changes throughout the year. Further work is needed to determine to what extent host diet is responsible for this variation. In all, this work provides useful insight into the temporal stability of the avian microbiome.

Acknowledgments:

We appreciate the help provided by field and lab assistants, particularly Briana Robertson. We are grateful to the Missouri Department of Conservation and Columbia Bottom Conservation Area for providing research permissions and sampling site access. The computation for this work was performed on the high performance computing infrastructure provided by Research Computing Support Services and in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO.

Funding:

Funding for this research was generously provided by the NSF Graduate Research Fellowship #1945995, the Harris World Ecology Center, Sigma Xi Grant in Aid of Research #G2020100193156089, NIH Grant #AI137984, and the Des Lee Collaborative Vision.

Ethics approval:

Avian handling protocols were approved by UMSL's Institutional Animal Care and Use Committee (Protocol # 1568558). Sampling permission was provided by the Missouri Department of Conservation (Wildlife Collector's Permit #18682 and #19163) and by a Columbia Bottom Conservation Area Special Use Permit.

References:

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. – Austral Ecol 26: 32–46.

Anderson, M. J. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). – Wiley StatsRef: Statistics Reference Online. American Cancer Society, pp. 1–15.

Arif, M., Akteruzzaman, Md., Tuhin-Al-Ferdous, Islam, S. S., Das, B. C., Siddique, M. P. and Kabir, S. M. L. 2021. Dietary supplementation of Bacillus-based probiotics on the growth performance, gut morphology, intestinal microbiota and immune response in low biosecurity broiler chickens. – Veterinary and Animal Science 14: 100216.

BBTools. – DOE Joint Genome Institute.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. – J R Stat Soc Series B Stat Methodol 57: 289–300.

Benskin, C. M. H., Wilson, K., Jones, K. and Hartley, I. R. 2009. Bacterial pathogens in wild birds: a review of the frequency and effects of infection. – Biological Reviews 84: 349–373.

Bergmann, G. T., Craine, J. M., Robeson, M. S. and Fierer, N. 2015. Seasonal Shifts in Diet and Gut Microbiota of the American Bison (Bison bison). – PLoS ONE 10: e0142409.

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

Cao, J., Hu, Y., Liu, F., Wang, Y., Bi, Y., Lv, N., Li, J., Zhu, B. and Gao, G. F. 2020. Metagenomic analysis reveals the microbiome and resistome in migratory birds. – Microbiome 8: 26.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., Holland, T. A., Keseler, I. M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., Weerasinghe, D., Zhang, P. and Karp, P. D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. – Nucl. Acids Res. 42: D459–D471.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. and Callahan, B. J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. – Microbiome 6: 226.

Dong, X. and Strous, M. 2019. An integrated pipeline for annotation and visualization of metagenomic contigs. – Front. Genet. <https://doi.org/10.3389/fgene.2019.00999>.

Escallón, C., Belden, L. K. and Moore, I. T. 2019. The cloacal microbiome changes with the breeding season in a wild bird. – Integr Org Biol <https://doi.org/10.1093/iob/oby009>.

Franzosa, E. A., Knight, R., Rahnavard, G., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N. and Huttenhower, C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. – Nat Methods 15: 962–968.

Gong, L., Liu, B., Wu, H., Feng, J. and Jiang, T. 2021. Seasonal Dietary Shifts Alter the Gut Microbiota of Avivorous Bats: Implication for Adaptation to Energy Harvest and Nutritional Utilization. – mSphere 6: e00467-21.

Guo, N., Wu, Q., Shi, F., Niu, J., Zhang, T., Degen, A. A., Fang, Q., Ding, L., Shang, Z., Zhang, Z. and Long, R. 2021. Seasonal dynamics of diet–gut microbiota interaction in adaptation of yaks to life at high altitude. – npj Biofilms Microbiomes 7: 38.

Heiman, M. L. and Greenway, F. L. 2016. A healthy gastrointestinal microbiome is dependent on dietary diversity. – Mol Metab 5: 317–320.

Hills, R. D., Pontefract, B. A., Mishcon, H. R., Black, C. A., Sutton, S. C. and Theberge, C. R. 2019. Gut microbiome: profound implications for diet and disease. – Nutrients <https://doi.org/10.3390/nu11071613>.

Hu, X., Liu, G., Li, Y., Wei, Y., Lin, S., Liu, S., Zheng, Y. and Hu, D. 2018. High-throughput analysis reveals seasonal variation of the gut microbiota composition within forest musk deer (*Moschus berezovskii*). – Front. Microbiol. 9: 1674.

Jackson, J. 2003. The Eurasian Tree Sparrow. – Missouri Department of Conservation.

Kang, D. D., Froula, J., Egan, R. and Wang, Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. – PeerJ 3: e1165.

Knutie, S. A. and Gotanda, K. M. 2018. A Non-invasive Method to Collect Fecal Samples from Wild Birds for Microbiome Studies. – Microbial Ecology 76: 851–855.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina Sequencing Platform. – Appl. Environ. Microbiol. 79: 5112–5120.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. – Bioinformatics 25: 2078–2079.

Liu, B., Zheng, D., Jin, Q., Chen, L. and Yang, J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. – Nucleic Acids Res 47: D687–D692.

Love, M. I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. – Genome Biol 15: 550.

Lu, J. and Salzberg, S. L. 2020. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. – Microbiome 8: 124.

Markle, J. G. M., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K. D., Macpherson, A. J. and Danska, J. S. 2013. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. – Science 339: 1084–1088.

McMurdie, P. J. and Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. – PLoS ONE 8: e61217.

Minich, D., Madden, C., Navarro, M. A., Glowacki, L., French-Kim, K., Chan, W., Evans, M. V., Soares, K., Mrofchak, R., Madan, R., Ballash, G. A., LaPerle, K., Paul, S., Vodovotz, Y., Uzal, F. A., Martinez, M., Hausmann, J., Junge, R. E. and Hale, V. L. 2022. Gut microbiota and age shape susceptibility to clostridial enteritis in lorikeets under human care. – Animal Microbiome 4: 7.

Nurk, S., Meleshko, D., Korobeynikov, A. and Pevzner, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. – Genome Res 27: 824–834.

Obrochta, S., Savo Sardaro, M. L., Amato, K. R. and Murray, M. H. 2022. Relationships Between Migration and Microbiome Composition and Diversity in Urban Canada Geese. – Frontiers in Ecology and Evolution.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H. 2019. vegan: Community Ecology Package.

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I. and Tuohy, K. 2018. Gut microbiota functions: metabolism of nutrients and other food components. – Eur J Nutr 57: 1–24.

Si, J., Vázquez-Castellanos, J. F., Gregory, A. C., Decommer, L., Rymenans, L., Proost, S., Centelles Lodeiro, J., Weger, M., Notdurfter, M., Leitner, C., Santer, P., Rungger, G., Willeit, J., Willeit, P., Pechlaner, R., Grabherr, F., Kiechl, S., Tilg, H. and Raes, J. 2022. Long-term life history predicts current gut microbiome in a population-based cohort study. – Nat Aging 2: 885–895.

Sun, F., Chen, J., Liu, K., Tang, M. and Yang, Y. 2022. The avian gut microbiota: Diversity, influencing factors, and future directions. – Frontiers in Microbiology.

Videvall, E., Song, S. J., Bensch, H. M., Strandh, M., Engelbrecht, A., Serfontein, N., Hellgren, O., Olivier, A., Cloete, S., Knight, R. and Cornwallis, C. K. 2020. Early-life gut dysbiosis linked to juvenile mortality in ostriches. – Microbiome 8: 147.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. 2019. Welcome to the Tidyverse. – JOSS 4: 1686.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F. O. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. – Nucleic Acids Res. 42: D643–D648.

CHAPTER 3.

Gut microbiome composition associated with *Plasmodium* infection in the Eurasian tree sparrow

Sage D. Rohrer[1], Briana Q. Robertson[1], Lon M. Chubiz[1], & Patricia G. Parker[1,2]

[1]*Department of Biology and Whitney R. Harris World Ecology Center, University of Missouri-St. Louis, One University Blvd., St. Louis, Missouri 63121, USA*

[2] *WildCare Institute, Saint Louis Zoo, One Government Drive, St. Louis, Missouri 63110, USA*

Abstract:

Recent expansion of microbiome research has uncovered connections between resident microbial communities and blood parasite risk, establishing the potential for microbial disease treatments such as probiotics in the future. However, this field has largely focused on humans and model organisms, leaving much unknown about how microbial communities might directly or indirectly impact parasite infection in wild populations and non-mammals. To contribute to this knowledge base in wild birds, we collected fecal and blood samples from wild Eurasian tree sparrows (*Passer montanus*) in the

United States to test for associations between blood parasite infection and the gut microbiome. We used a widespread molecular approach to test 81 samples from peripheral blood for *Plasmodium* and *Haemoproteus*, and we characterized the gut microbiome using fecal samples as a proxy. Neither alpha nor beta diversity significantly varied with detected *Plasmodium* infection. However, differential abundance analysis highlighted a number of significantly varying bacteria, with the greatest representation within the phyla *Proteobacteria* and *Firmicutes* in *Plasmodium*-infected birds. These differentially abundant taxa offer a starting point for experimental work establishing the relationship between microbial abundance and *Plasmodium* infection.

Keywords: Avian microbiome, gut microbiome, *Plasmodium,* blood parasites, sparrows

Original Article in press at the Journal of Avian Biology, DOI: 10.1111/jav.03027.

Introduction:

Avian malarial parasites are globally distributed, with the mosquito-vectored parasites from the genus *Plasmodium* found on every continent other than Antarctica (Fecchio et al. 2021). *Plasmodium* species and other haemosporidian parasites can cause high mortality among avian species that are immunologically naïve, both in captive settings and in the wild (Hernandez-Colina et al. 2021). For example, when mosquito vectors of *Plasmodium relictum* were introduced to Hawaii, the introduction contributed to dramatic population reductions and extinctions of many native Hawaiian birds (Warner 1968, LaPointe et al. 2012, Liao et al. 2017). Avian species with a long evolutionary history with malaria parasites generally exhibit a higher tolerance to infection, but negative effects including decreased fitness and mortality are still observed (Marzal et al. 2005, Lachish et al. 2011, Ilgūnas et al. 2019). The global importance of this infectious disease underscores the need to understand factors influencing susceptibility and resistance to infection.

Malaria parasites undergo complex life stages within avian hosts. Following injection of parasite sporozoites into the host from a mosquito vector, the sporozoites invade skin and tissue cells and go through multiple phases of preerythrocytic asexual reproduction to produce merozoites (Valkiunas 2005). Some merozoites then invade red blood cells and asexual reproduction continues simultaneously in erythrocytes (erythrocytic merogony) and in host tissues (exoerythrocytic merogony) (LaPointe et al. 2012, Valkiūnas and Iezhova 2017). Gametogony (production of gametocytes) occurs within some infected erythrocytes, and gametocytes remain inside the red blood cells until

they are taken up by a vector to continue the sexual reproduction stage of the parasite's life cycle (Valkiunas 2005).

Infection of an avian host typically occurs over two stages: acute and chronic (van Riper et al. 1986, LaPointe et al. 2012, Asghar et al. 2012). The highest numbers of parasites in the blood (i.e. parasitemia levels) occur during the primary acute phase, typically within about two weeks, although the initial dose amount and a number of host-specific factors can vary the timeline of infection (van Riper et al. 1986, Atkinson et al. 1995). Birds undergoing the acute phase often suffer from severe anemia, and necropsies of deceased birds typically reveal swollen and abnormally colored livers and spleens (van Riper et al. 1986, LaPointe et al. 2012). In surviving birds, the acute stage is often followed by a chronic infection, which can persist for years and is associated with long-term fitness effects (Manwell 1934, LaPointe et al. 2012, Asghar et al. 2015). Even after parasitemia has decreased and the chronic stage has begun, exoerythrocytic invasion can lead to sudden mortality by blocking brain capillaries, resulting in cerebral ischemia (Ilgūnas et al. 2016).

A bird's risk of malaria infection depends on many factors. Habitat and/or temperature can have a strong influence, as warmer temperatures at lower elevations can be optimal for the vector stage of parasite development (LaPointe et al. 2010). Behavioral changes such as migration and nesting can affect infection rates; one recent study found that some birds with high levels of nest care exhibit substantially higher rates of haemosporidian infection than brood parasites in the same area, particularly in birds with open nests (Ganser et al. 2020). This cost of parental care could be a result of the

physiological changes which accompany nesting, or due to reduced ability to avoid vectors while nesting (Ganser et al. 2020). Host genetics are also important predictors of disease risk, and many studies have demonstrated significant associations between major histocompatibility complex genes and haemosporidian parasite infection or infection intensity (Sommer 2005, Westerdahl et al. 2005, Bonneaud et al. 2006, Loiseau et al. 2008).

In addition to these many predictors of infection risk, several recent studies have demonstrated connections between the host microbiome and susceptibility to malaria, primarily in mammalian systems within laboratory settings. For example, mice resistant to *Plasmodium yoelii* exhibit differential expression of bacterial genes when compared to genetically similar susceptible mice (Stough et al. 2016). Resistant mice also show significant enrichment of *Lactobacillus* and *Bifidobacterium* in the gut microbiome (Villarino et al. 2016). In a rare field example, human stool samples collected longitudinally through a high *Plasmodium falciparum* transmission season reveal bacterial community profiles that correlate with malaria infection risk – resistant individuals possess higher proportions of *Bifidobacterium*, *Streptococcus,* and *Enterobacteriaceae Escherichia/Shigella* (Yooseph et al. 2015). Infection by *Plasmodium* parasites may in turn cause dysbiosis in the microbiome, and *Plasmodium berghei* infection in mice is accompanied with physical intestinal changes (including intestinal shortening and increased permeability) as well as altered bacterial communities (Taniguchi et al. 2015).

In some cases, microbiome-associated resistance is inducible in animals that were previously susceptible. For example, injecting mice with *Lactobacillus casei* can confer resistance to *Plasmodium chabaudi*, reducing the parasite load and shortening infection periods (Martínez-Gómez et al. 2006). Transplanting the cecal microbiome from susceptible or resistant mice to new germfree mice can similarly transfer risk or resistance (Villarino et al. 2016). Treating susceptible mice with antibiotics followed by yogurt probiotics containing *Lactobacillus* and *Bifidobacterium* can also result in a lower parasite infection intensity (Villarino et al. 2016). The precise mechanism of resistance is unclear in these examples, but increased levels of host immune response correlate with decreased infection severity in mice (Villarino et al. 2016), hinting at an interactive relationship between the microbiome, host immune system, and pathogen infection.

One mechanistic explanation for microbiome-*Plasmodium* associations may be through bacterial priming of the immune system prior to parasite invasion. Pathogens commonly express a glycan epitope called alpha-gal (α-gal), which provokes an immune response in primates in general, Old World monkeys in particular (Galili et al. 1988), and other non-mammal vertebrates including birds (Yilmaz et al. 2014b, Mateos-Hernández et al. 2020). Production of α-gal has been confirmed in several species of *Plasmodium*, expressed on the surfaces of sporozoites (Yilmaz et al. 2014b). Antibodies specific to α-gal are higher in humans with greater resistance to *Plasmodium* infection, although antibody levels are not predictive of febrile malaria outcome (Yilmaz et al. 2014b). Notably, when “human-like” mice (lacking the ability to express α-gal) are treated with antibiotics and inoculated with an *E. coli* strain O86:B7 that exhibits high expression of

α-gal, they demonstrate increased α-gal antibody production and increased resistance to colonization by *Plasmodium* through antibody-mediated blockage of sporozoites (Yilmaz et al. 2014b). More work is needed to demonstrate other specific mechanisms linking the microbiome to host immune response and *Plasmodium* risk.

While evidence thus points to indirect effects between the microbiome and *Plasmodium* infection in mammals, it remains unclear whether this interaction occurs in avian systems. Mammalian malaria and avian malaria share many similarities in pathogenesis, although tissue infection is more extensive in avian malaria and disease tends to be more severe (Ilgūnas et al. 2016, Valkiūnas and Iezhova 2017, 2018), and it is plausible for microbial influence on parasite invasion to exist in birds as well as mammals. One of the few studies examining this potential relationship focused on the uropygial gland microbiome in house sparrows (*Passer domesticus*) (Videvall et al. 2021). No correlation was observed between alpha or beta diversity in microbiome samples and infection status, but several uropygial gland bacteria were differentially abundant between infected and uninfected birds (Videvall et al. 2021). Whether or not the gut microbiome influences blood parasite infection in birds, or vice versa, remains an open question.

Here we tested whether several diversity and abundance measures of the gut microbiome correlate with *Plasmodium* infection in wild Eurasian tree sparrows (*Passer montanus*). This locally abundant species was introduced to St. Louis, Missouri in 1870 with a founding population of approximately 20 individuals (Louis and Barlow 1988). Range expansion has occurred northward to a limited extent, and this sparrow is now present primarily in northern Missouri, Illinois, and Iowa (Burnett et al. 2017). A

previous study reported haemosporidian parasite infections in the local population of Eurasian tree sparrows as high as 60% (Lee et al. 2006). This introduced bird presents an opportunity to examine associations between the gut microbiome and blood parasite infection in a species of little conservation concern with a known history of infection with malaria parasites.

We hypothesized that the community composition or presence of specific members of the Eurasian tree sparrow microbiome may confer resistance or susceptibility to blood parasite infection. We expected to see lower alpha diversity in individuals infected with *Plasmodium* and significant clustering based on beta diversity between birds grouped by infection status. Based on results from previous studies in mammals, we predicted differentially abundant bacteria between infected and uninfected birds in groups such as *Lactobacillus, Bifidobacterium, Streptococcus,* or *Escherichia/Shigella* (Yooseph et al. 2015, Villarino et al. 2016).

Methods:

Adult and juvenile Eurasian tree sparrows (n = 81) were captured using mist nets at two sites in Saint Louis County, Missouri – a residential site in Webster Groves and the visitor center of Columbia Bottom Conservation Area (directly adjacent to residential areas). The sites are approximately 20 miles apart. Feces was collected from each bird using bags with double floors that had been swabbed with alcohol prior to sampling; the fecal sample falls through the wire floor at the bottom of the bag onto a clean surface immediately below (Knutie and Gotanda 2018). Birds were banded to identify

individuals, and basic measurements were taken including wing chord and body mass. Small blood samples (<50uL) were taken from the brachial vein and preserved in Longmire's lysis solution (Longmire et al. 1997) and thin blood smears. Slides were fixed with methanol within one hour and later stained with Giemsa. All birds were released after handling. Fecal samples were kept on ice during sampling and immediately placed in frozen storage following return from the field.

Blood Parasite Testing:

Blood sample DNA was extracted with a standard phenol-chloroform method. DNA samples were then tested in triplicate for *Plasmodium*/*Haemoproteus* blood parasites using a nested PCR protocol. This protocol amplifies a region of *cytochrome B* using the primer pair HAEMNF and HAEMNR2 for the first reaction and the primers HAEMF and HAEMR2 for the second reaction (Waldenström et al. 2004). All positive amplicons were sent for Sanger sequencing at Eurofins Genomics LLC, with forward and reverse reads obtained for each positive amplicon as well as for duplicate amplicons when applicable (up to three duplicates per individual bird).

High-quality forward and reverse trace files and duplicates were assembled using the "Pearl" tool on GEAR Genomics (https://www.gear-genomics.com/) (Rausch et al. 2020). Consensus sequences obtained from the assembly were matched to existing parasite sequences using the BLAST tool on the MalAvi database (Bensch et al. 2009). Assembled traces generally matched to the same lineages as the raw traces, but using consensus sequences substantially improved the percent match of the lineage in most cases. Any

samples with lineage matches <100% were re-amplified and sequenced again. All samples included in the analysis had a 100% match to a MalAvi lineage.

16S rRNA Gene Sequencing and Processing:

DNA was extracted from the frozen fecal samples using the Qiagen Power Fecal Pro DNA Kit following manufacturer instructions. Extracted DNA samples with at least ~205ng of DNA (measured by a Qubit fluorometer) were then sent to the University of Michigan Medical School Microbiome Core for 16S rRNA gene sequencing of the V4 region, using the V4 primers and methods described in Kozich et al. 2013 (Kozich et al. 2013). Negative controls for the sampling and extraction protocols were also sent for the sequencing runs, and an extra negative sequencing control (water) and positive sequencing control (ZymoBIOMICS Microbial Community Standard) were added by the Microbiome Core for each of the two sequencing runs. A positive extraction control (ZymoBIOMICS Gut Microbiome Standard) was extracted alongside fecal samples and included in the second sequencing run. The resulting sequencing reads were processed using the Dada2 pipeline (Callahan et al. 2016). Quality filtering and trimming were performed using recommended Dada2 filtering steps (keeping only merged reads between 250-256bp long) and taxonomies were assigned to amplicon sequence variants (ASVs) using SILVA v138 as a reference (Yilmaz et al. 2014a). Reads that matched mitochondria, chloroplasts, or Archaea were removed. Following Dada2 filtering steps, all expected genera were detected in the two microbial community standards, and unclassified reads and/or potential contamination represented ~0% in the first run and ~0.15% in the second run. All genera with expected abundances ≥0.97% (based on

manufacturer guidelines) were also apparent in the more complex gut microbiome standard, but low-abundance *Salmonella* (expected abundance 0.009%) appears misclassified by the pipeline as the fellow *Enterobacteriaceae* member *Aquamonas,* and the two bacteria with the lowest expected abundances, *Enterococcus* (0.0009%) and *Clostridium* (0.0002%), were not detected, indicating that more abundant community members are reliably detected by this pipeline but that the rarest community members may not be well-represented. Unclassified reads or potential contamination represented ~0.05% of the gut microbiome standard.

Filtered reads were imported into R for analysis using the R package phyloseq (McMurdie and Holmes 2013, R Core Team 2020). Negative controls were used to determine probable contaminant ASVs with the R package decontam, and the two detected contaminant ASVs were then removed from all samples (Davis et al. 2018). Samples with fewer than 4000 reads were excluded, a threshold based on minimum read depth used in other studies and the distribution of read depth in this dataset. Following filtering steps, 81 samples were included in the analysis, including 14 birds with *Plasmodium* infection evident in the peripheral blood and 67 with no detected infections.

Differential Abundance Analysis:

Filtered reads were further pruned using the R package phyloseq to limit taxa to those found at least 3 times in at least 5% of the samples. The phyloseq object was then converted to DESeq2 format, and DESeq2 was run using default settings (which includes a data normalization step) to simultaneously test for differential taxa based on

Plasmodium status while controlling for sequencing run, age, season, and location in the model (Love et al. 2014). Uninfected individuals were set as the reference level for DESeq2. False discovery correction was performed using the Benjamini and Hochberg method, and adjusted *p* values were filtered using alpha = 0.01 to reduce the probability of false positives (Benjamini and Hochberg 1995). Only ASVs with adjusted *p* values < 0.01 are reported here.

Alpha and Beta Diversity:

Filtered samples (without the additional DESeq2 pruning) were subsampled to 4005 reads per sample in phyloseq to match the sample with the lowest read count above 4000. Observed richness and Shannon Diversity were calculated in phyloseq based on the rarefied ASV tables. Significant variation in alpha diversity between groups was tested in R using ANOVAs with sequencing run, location, season, and age included in the models. Beta diversity was calculated from the rarefied ASV tables using Bray–Curtis dissimilarity and visualized using phyloseq and ggplot2 (Wickham et al. 2019). Beta diversity significance between infected and uninfected groups was assessed using PERMANOVAs that controlled for sequencing run and included season, location, and age in the model, using the R package vegan (Anderson and Walsh 2013, Anderson 2017, Oksanen et al. 2019). Significant PERMANOVA results were checked with PERMDISP to determine whether dispersion was a confounding factor (Anderson and Walsh 2013, Oksanen et al. 2019).

Results:

All parasite sequences were classified as *Plasmodium* (Table 1). Two lineages, SEIAUR01 and WW3, were detected at both Webster Groves and Columbia Bottom Conservation Area. A third lineage, PADOM11, was only detected at Columbia Bottom. Detected *Plasmodium* prevalence in these two sites was 14% at Webster Groves (7/47) and 20% at Columbia Bottom CA (7/34), for a total prevalence of 17% in this sample set (14/81). Since sample size per lineage is low, we did not examine potentially varying effects based on *Plasmodium* lineage but instead considered all *Plasmodium* infections as one group.

Sampling Site	**Total Samples**	**Total Positives**	***SEIAUR01***	***WW3***	***PADOM11***
Webster Groves	47	7	3	4	0
Columbia Bottom CA	34	7	3	2	2

Table 1. Sampling sites and *Plasmodium* lineages.

16S rRNA gene sequencing showed that the primary bacterial phyla in these communities are *Firmicutes, Proteobacteria,* and *Actinobacteria* (Fig. 1). Bacterial community composition at a lower taxonomic level was highly varied, and 17 bacterial families were detected in proportions greater than 1%. Among the most abundant families were *Catellicoccaceae*, *Enterobacteriaceae*, and *Staphylococcaceae* (Fig. 1).

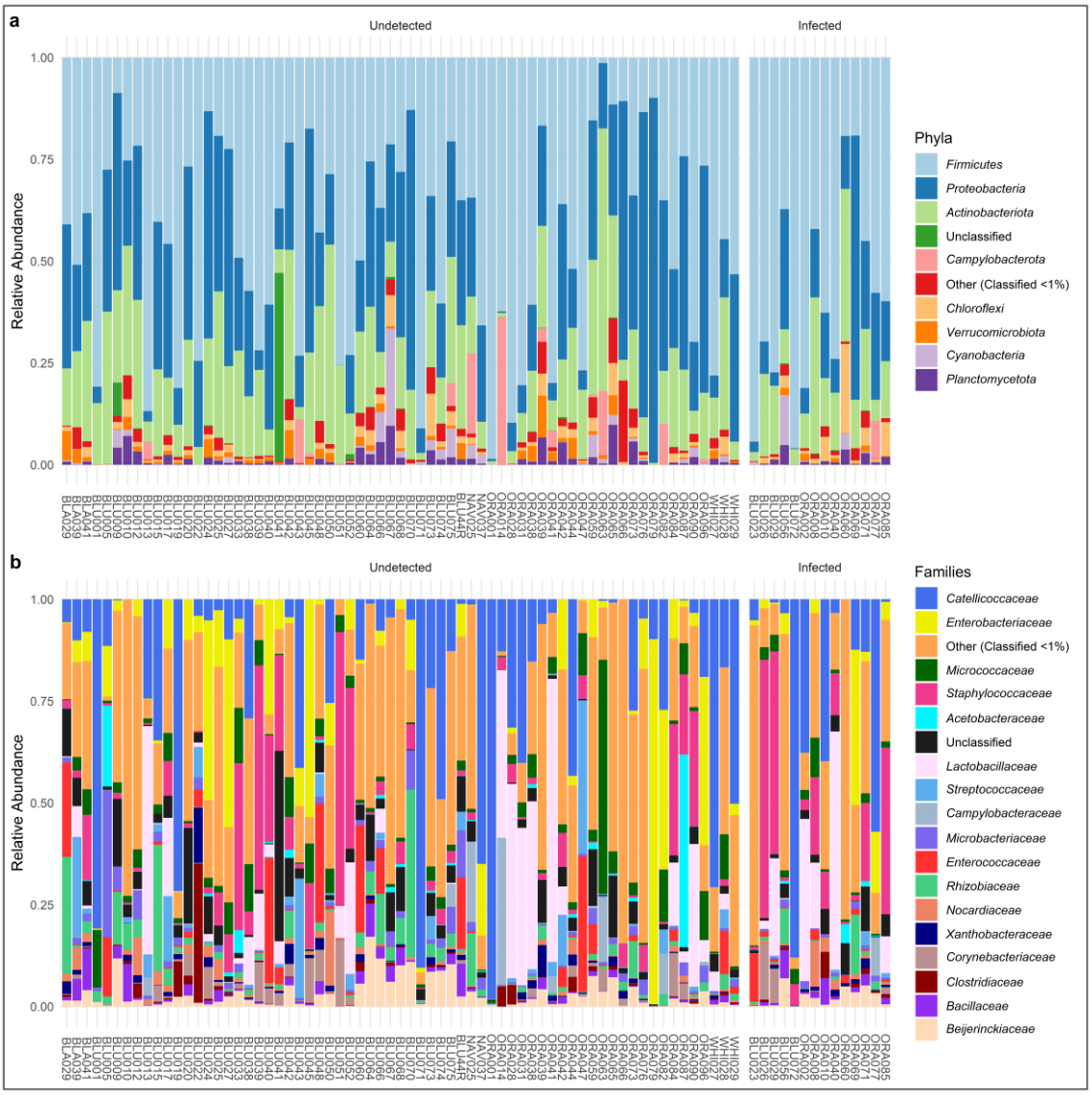

Figure 1. Relative diversity at phylum (a) and family (b) levels across individuals with no detected blood parasites vs. those infected with *Plasmodium.*

Alpha diversity did not significantly vary between infection states or any other variable when measured as observed richness or Shannon Diversity with ANOVAs that controlled for the potential confounding variables of sequencing run, location, season, and age. Principal Coordinates Analysis based on Bray–Curtis dissimilarity showed substantial overlap between the two groups (Fig. 2). PERMANOVAs testing for variation in beta

diversity between groups were not significant at either site based on infection status when including sequencing run, season, and age in that order in separate models for each site, or in a combined model controlling for location (Columbia Bottom, PERMANOVA, R^2 = 0.02702, P = 0. 5658; Webster Groves, PERMANOVA, R^2 = 0.02197, P = 0. 2396; combined, PERMANOVA, R^2 = 0.01329, P = 0.2110). In the combined model (Table 2), significant beta diversity variation was observed for location (PERMANOVA, P = 0.0001, R^2 = 0.03678; PERMDISP, P = 0.655) and season (PERMANOVA, P = 0.0001, R^2 = 0.07285; PERMDISP, P = 0.002), but not for sequencing run (PERMANOVA, P = 0.3375, R^2 = 0.01212) or age (PERMANOVA, P = 0.0591, R^2 = 0.01605).

	Df	Sum Of Sqs	R2	F	Pr(>F)
Sequencing Run	1	0.406	0.0121	1.0569	0.3375
Location	1	1.231	0.0368	3.2065	0.0001***
Season	2	2.439	0.0729	3.1752	0.0001***
Age	1	0.537	0.0161	1.3993	0.0591
Infection Status	1	0.445	0.0133	1.1587	0.211
Residuals	74	28.421	0.8489		
Total	80	33.479	1		

Table 2: PERMANOVA results showing beta diversity differences based on Bray-Curtis similarity for combined samples grouped by infection status. The PERMANOVA controlled for sequencing run, location, season, and age. Location and age both significantly varied in this model.

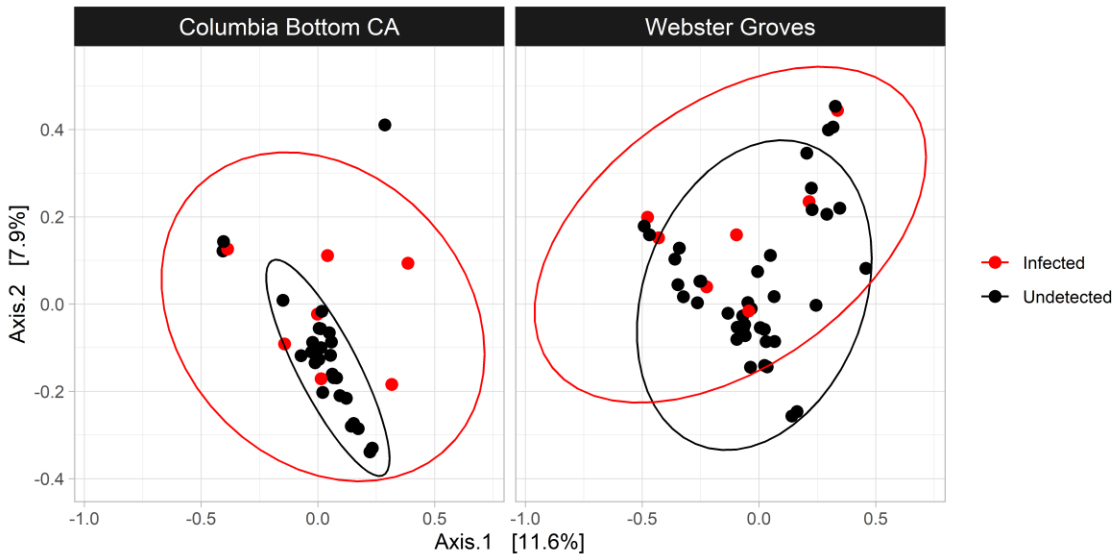

Figure 2. PCoA based on Bray–Curtis dissimilarity of infected birds vs. birds with no detected *Plasmodium* infections.

DESeq2 analysis identified several bacterial taxa that significantly differed between infected birds and birds with no detected parasites when controlling for sequencing run, age, season, and location (Fig. 3). Twenty-seven ASVs in ten bacterial classes significantly varied based on adjusted P values. Six ASVs belonged to the class *Bacilli*, six to the class *Gammaproteobacteria,* five to the class *Actinobacteria,* and four to the class *Alphaproteobacteria*, with the remaining classes represented by a single ASV each. Four of the seven ASVs with the highest calculated base mean (>15) were from the class *Bacilli*, from the genera *Streptococcus* (97.362 base mean), *Ligilactobacillus* (40.189 base mean), *Staphylococcus* (30.11 base mean), and *Weissella* (18.363 base mean). The other ASVs with high base mean values were from the classes *Gammaproteobacteria* (*Shimwellia*, 250.24 base mean, and *Escherichia-Shigella,* 18.78 base mean) and *Alphaproteobacteria* (*Candidatus Tokpelaia,* 54.177 base mean).

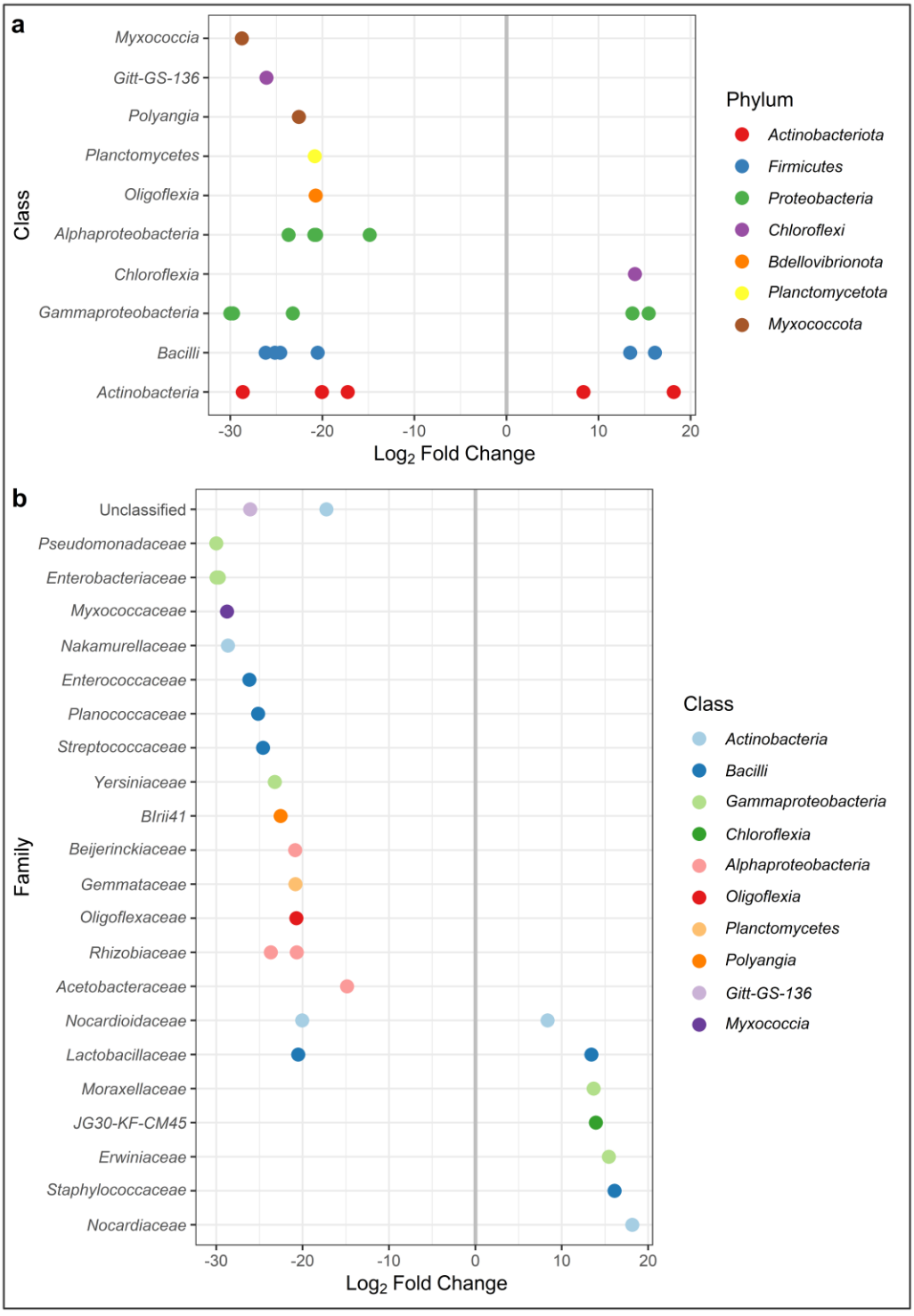

Figure 3. DESeq2 results showing $\log_2$ fold change in significantly different bacterial classes (a) and families (b) with uninfected samples as a reference. Each dot represents an ASV. Taxa with positive $\log_2$

fold change values are significantly more abundant in individuals infected by *Plasmodium*; negative values are significantly less abundant in infected individuals.

Discussion:

These results point to ASV-level associations between the gut microbiome and *Plasmodium* infection, but not to large-scale changes to the broader community structure. DESeq2 analysis highlighted 27 bacterial ASVs (out of a total of 373 ASVs following filtering steps) that were differentially abundant in infected birds while controlling for confounding effects of sequencing run, season, location, and bird age. However, alpha and beta diversity did not vary with *Plasmodium* infection while controlling for confounding variables, and infection status explained only ~1.33% of the beta diversity for the two sites combined. The significantly varying bacterial taxa may be indicative of a small-scale relationship between members of the gut microbiome and avian malaria infection. While a relationship between the gut microbiome and blood parasite infection has been identified previously in mammalian systems (Yooseph et al. 2015, Villarino et al. 2016), this is one of the first studies demonstrating similar associations in an avian species (Videvall et al. 2021).

Of the 27 differentially abundant ASVs, six belonged to the bacterial class *Bacilli* and six to the class *Gammaproteobacteria,* followed by five from *Alphaproteobacteria* and four from *Actinobacteria*. The classes *Chloroflexia, Oligoflexia, Planctomycetes, Polyangia, Gitt-GS-136*, and *Myxococcia* were each represented by a single ASV. Results at a genus level are highly varied, containing bacteria from putatively beneficial groups such as *Ligilactobacillus* (formerly *Lactobacillus*) as well as a number of genera with pathogenic

member species such as *Escherichia-Shigella, Pseudomonas, Shimwellia, Enterococcus, Streptococcus, Staphylococcus* and *Serratia* (Devriese et al. 1994, Benskin et al. 2009). Somewhat surprisingly, most of the members of potentially pathogenic groups were present at significantly lower abundances in birds infected with *Plasmodium* (except for the significantly enriched *Staphylococcus*), while the likely beneficial *Ligilactobacillus* was significantly enriched in infected birds. These results include some bacterial groups that have been found to significantly differ in mammals infected with *Plasmodium*, such as *Streptococcus, Lactobacillus,* and *Escherichia-Shigella* (Yooseph et al. 2015, Villarino et al. 2016).

Interpretation of these results is complicated by the impossibility of knowing which uninfected birds were naïve and which had previously cleared or undetectable infections. Additionally, most of the birds with infections detected in this study likely represented chronic infections, because birds suffering acute infections are caught less frequently in mist nets (Lachish et al. 2011). Further work is therefore needed to ascertain directionality of the associations between bacterial taxa and *Plasmodium* infection found in this study, and a larger proportion of infected birds in the sample set or a larger overall sample size would be beneficial. The differentially abundant ASVs may have already varied prior to infection, thus indicating a potential susceptibility factor, or blood parasite infection could have prompted a change in the microbial communities. Those outcomes are not mutually exclusive, and both microbial influence on *Plasmodium* susceptibility and *Plasmodium* influence on microbial composition have

been found previously in experimental mouse studies (Taniguchi et al. 2015, Villarino et al. 2016).

Future research that includes experimental inoculation of captive birds with *Plasmodium* would help to understand how initial gut microbiome structure and community members may predict outcomes and change with infection, although differences between captive and wild microbiomes in many species limit inferences about wild microbiomes from captive studies (Alberdi et al. 2021, San Juan et al. 2021). Longitudinal studies with high site fidelity among avian focal species could clarify the directionality of this relationship in wild populations by, ideally, sampling the same individuals before and after parasite exposure. The bacterial taxa highlighted by this study could also be used as a starting point for future work on indirect interactions between the microbiome and blood parasites, such as examining whether treatment with this combination of bacteria measurably influences bird susceptibility to malaria under controlled conditions.

The rapidly advancing field of animal microbiomes is beginning to turn toward the use of probiotics for prevention or treatment of infectious diseases in animals (McKenzie et al. 2018, Stedman et al. 2020). For example, probiotics with antifungal properties are being explored in endangered amphibians as a potential defense against the emerging fungal disease chytridiomycosis (which has contributed to the extinction of several amphibian species) (Woodhams et al. 2016, Harrison et al. 2020), although interactive effects with the host immune system complicate potential treatments (Woodhams et al. 2020). Additional studies are exploring probiotics across varied systems, including as a

treatment to increase trout survival from *Aeromonas salmonicida* infection (Balcázar et al. 2007) and to prevent diseases such as American foulbrood in honeybees (Audisio 2017, Daisley et al. 2020). Probiotics made up of commensal bacteria (such as *Lactobacillus* or *Bifidobacterium*) are already being used in poultry feed to reduce severity of bacterial infections in commercial birds (Redweik et al. 2020). However, much work remains to be done to fully understand both the relationship between the microbiome and disease, as well as how specific probiotics might be used to improve outcomes in avian species.

Several studies in mammals have already demonstrated the potential for microbial-based treatments to modulate the severity of blood parasite infections. Under controlled conditions, *Plasmodium* infection intensity has been reduced by various treatments such as injecting mice with probiotics prior to parasite infection (Martínez-Gómez et al. 2006), transplanting cecal microbiomes from resistant mice to susceptible mice (Villarino et al. 2016), or treating mice with yogurt probiotics (Villarino et al. 2016). However, in avian systems, very little is known about interactive effects between the microbiome and blood parasites. The results of this study indicate several bacterial taxa with differential abundances in infected wild sparrows, which can be used as a foundation for future studies examining the relationship between the gut microbiome and *Plasmodium* infection in wild birds.

Acknowledgments:

We are grateful to the field and lab assistants who provided valuable contributions to this work, particularly to members of the UMSL Parker Lab who assisted in data collection and provided feedback on the project. We thank the Missouri Department of Conservation and Columbia Bottom Conservation Area for providing permissions and site access. This research was supported in part by the University of Michigan Medical School Microbiome Core. Additional sequencing was performed by Eurofins Genomics, LLC (Louisville, Kentucky).

Funding:

Funding for this research was generously provided by the NSF Graduate Research Fellowship #1945995, the Harris World Ecology Center, Sigma Xi Grant in Aid of Research #G2020100193156089, NIH Grant #AI137984, and the Des Lee Collaborative Vision.

Ethics approval:

Avian handling protocols were approved by UMSL's Institutional Animal Care and Use Committee (Protocol # 1568558). Sampling permission was provided by the Missouri Department of Conservation (Wildlife Collector's Permit #18682 and #19163) and by a Columbia Bottom Conservation Area Special Use Permit.

References:

Alberdi, A., Martin Bideguren, G. and Aizpurua, O. 2021. Diversity and compositional changes in the gut microbiota of wild and captive vertebrates: a meta-analysis. – Sci Rep 11: 22660.

Anderson, M. J. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). – Wiley StatsRef: Statistics Reference Online. American Cancer Society, pp. 1–15.

Anderson, M. J. and Walsh, D. C. I. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? – Ecol Monogr 83: 557–574.

Asghar, M., Westerdahl, H., Zehtindjiev, P., Ilieva, M., Hasselquist, D. and Bensch, S. 2012. Primary peak and chronic malaria infection levels are correlated in experimentally infected great reed warblers. – Parasitology 139: 1246–1252.

Asghar, M., Hasselquist, D., Hansson, B., Zehtindjiev, P., Westerdahl, H. and Bensch, S. 2015. Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. – Science 347: 436–438.

Atkinson, C. T., Woods, K. L., Dusek, R. J., Sileo, L. S. and Iko, W. M. 1995. Wildlife disease and conservation in Hawaii: Pathogenicity of avian malaria (*Plasmodium relictum*) in experimentally infected Iiwi (*Vestiaria coccinea*). – Parasitology 111: S59–S69.

Audisio, M. C. 2017. Gram-positive bacteria with probiotic potential for the *Apis mellifera* L. Honey Bee: the experience in the northwest of Argentina. – Probiotics & Antimicro. Prot. 9: 22–31.

Balcázar, J. L., De Blas, I., Ruiz-Zarzuela, I., Vendrell, D., Gironés, O. and Muzquiz, J. L. 2007. Enhancement of the immune response and protection induced by probiotic lactic acid bacteria against furunculosis in rainbow trout (*Oncorhynchus mykiss*). – FEMS Immunology & Medical Microbiology 51: 185–193.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. – J R Stat Soc Series B Stat Methodol 57: 289–300.

Bensch, S., Hellgren, O. and Pérez-Tris, J. 2009. MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. – Molecular Ecology Resources 9: 1353–1358.

Benskin, C. M. H., Wilson, K., Jones, K. and Hartley, I. R. 2009. Bacterial pathogens in wild birds: a review of the frequency and effects of infection. – Biological Reviews 84: 349–373.

Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O. and Sorci, G. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. – Evolution 60: 383–389.

Burnett, J., Roberts, C., Allen, C., Brown, M. and Moulton, M. 2017. Range expansion by Passer montanus in North America. – Biological Invasions 19: 5–9.

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

Daisley, B. A., Pitek, A. P., Chmiel, J. A., Al, K. F., Chernyshova, A. M., Faragalla, K. M., Burton, J. P., Thompson, G. J. and Reid, G. 2020. Novel probiotic approach to counter *Paenibacillus larvae* infection in honey bees. – ISME J 14: 476–491.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. and Callahan, B. J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. – Microbiome 6: 226.

Devriese, L. A., Haesebrouck, F., Herdt, P. de, Dom, P., Ducatelle, R., Desmidt, M., Messier, S. and Higgins, R. 1994. *Streptococcus suis* infections in birds. – Avian Pathology 23: 721–724.

Fecchio, A., Clark, N. J., Bell, J. A., Skeen, H. R., Lutz, H. L., De La Torre, G. M., Vaughan, J. A., Tkach, V. V., Schunck, F., Ferreira, F. C., Braga, É. M., Lugarini, C., Wamiti, W., Dispoto, J. H., Galen, S. C., Kirchgatter, K., Sagario, M. C., Cueto, V. R., González-Acuña, D., Inumaru, M., Sato, Y., Schumm, Y. R., Quillfeldt, P., Pellegrino, I., Dharmarajan, G., Gupta, P., Robin, V. V., Ciloglu, A., Yildirim, A., Huang, X., Chapa-Vargas, L., Álvarez-Mendizábal, P., Santiago-Alarcon, D., Drovetski, S. V., Hellgren, O., Voelker, G., Ricklefs, R. E., Hackett, S. J., Collins, M. D., Weckstein, J. D., Wells, K. and Kamath, P. 2021. Global drivers of avian haemosporidian infections vary across zoogeographical regions. – Global Ecol. Biogeogr. 30: 2393–2406.

Galili, U., Shohet, S. B., Kobrin, E., Stults, C. L. and Macher, B. A. 1988. Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. – Journal of Biological Chemistry 263: 17755–17762.

Ganser, C., Monadjem, A., McCleery, R. A., Ndlela, T. and Wisely, S. M. 2020. Is it best on the nest? Effects of avian life-history on haemosporidian parasitism. – International Journal for Parasitology: Parasites and Wildlife 13: 62–71.

Harrison, X. A., Sewell, T., Fisher, M. and Antwis, R. E. 2020. Designing Probiotic Therapies With Broad-Spectrum Activity Against a Wildlife Pathogen. – Front. Microbiol. 10: 3134.

Hernandez-Colina, A., Gonzalez-Olvera, M., Eckley, L., Lopez, J. and Baylis, M. 2021. Avian malaria affecting penguins in zoological gardens, aquariums and wildlife parks in the UK. – Veterinary Record 189: e511.

Ilgūnas, M., Bukauskaitė, D., Palinauskas, V., Iezhova, T. A., Dinhopl, N., Nedorost, N., Weissenbacher-Lang, C., Weissenböck, H. and Valkiūnas, G. 2016. Mortality and pathology in birds due to Plasmodium (Giovannolaia) homocircumflexum infection, with emphasis on the exoerythrocytic development of avian malaria parasites. – Malaria Journal 15: 256.

Ilgūnas, M., Palinauskas, V., Platonova, E., Iezhova, T. and Valkiūnas, G. 2019. The experimental study on susceptibility of common European songbirds to Plasmodium elongatum (lineage pGRW6), a widespread avian malaria parasite. – Malar J 18: 290.

Knutie, S. A. and Gotanda, K. M. 2018. A Non-invasive Method to Collect Fecal Samples from Wild Birds for Microbiome Studies. – Microbial Ecology 76: 851–855.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina Sequencing Platform. – Appl. Environ. Microbiol. 79: 5112–5120.

Lachish, S., Knowles, S. C. L., Alves, R., Wood, M. J. and Sheldon, B. C. 2011. Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. – Journal of Animal Ecology 80: 1196–1206.

LaPointe, D. A., Goff, M. L. and Atkinson, C. T. 2010. Thermal constraints to the sporogonic development and altitudinal distribution of avian malaria *Plasmodium relictum* in Hawai'i. – The Journal of Parasitology 96: 318–324.

LaPointe, D. A., Atkinson, C. T. and Samuel, M. D. 2012. Ecology and conservation biology of avian malaria: Ecology of avian malaria. – Annals of the New York Academy of Sciences 1249: 211–226.

Lee, K. A., Martin II, L. B., Hasselquist, D., Ricklefs, R. E. and Wikelski, M. 2006. Contrasting adaptive immune defenses and blood parasite prevalence in closely related Passer sparrows. – Oecologia 150: 383–392.

Liao, W., Atkinson, C. T., LaPointe, D. A. and Samuel, M. D. 2017. Mitigating Future Avian Malaria Threats to Hawaiian Forest Birds from Climate Change. – PLoS ONE 12: e0168880.

Loiseau, C., Zoorob, R., Garnier, S., Birard, J., Federici, P., Julliard, R. and Sorci, G. 2008. Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. – Ecology Letters 11: 258–265.

Longmire, J. L., Baker, R. J., Maltbie, Mary., and Texas Tech University. 1997. Use of "Lysis Buffer" in DNA isolation and its implication for museum collections. – Museum of Texas Tech University.

Louis, V. L. S. and Barlow, J. C. 1988. Genetic Differentiation Among Ancestral and Introduced Populations of the Eurasian Tree Sparrow (passer Montanus). – Evolution 42: 266–276.

Love, M. I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. – Genome Biol 15: 550.

Manwell, R. B. 1934. THE DURATION OF MALARIAL INFECTION IN BIRDS*. – American Journal of Epidemiology 19: 532–538.

Martínez-Gómez, F., Ixta-Rodríguez, O., Aguilar-Figueroa, B., Hernández-Cruz, R. and Monroy-Ostria, A. 2006. *Lactobacillus casei* ssp. *rhamnosus* enhances non specific protection against *Plasmodium chabaudi* AS in mice. – Salud pública Méx 48: 498–503.

Marzal, A., de Lope, F., Navarro, C. and Møller, A. P. 2005. Malarial Parasites Decrease Reproductive Success: An Experimental Study in a Passerine Bird. – Oecologia 142: 541–545.

Mateos-Hernández, L., Risco-Castillo, V., Torres-Maravilla, E., Bermúdez-Humarán, L. G., Alberdi, P., Hodžić, A., Hernández-Jarguin, A., Rakotobe, S., Galon, C., Devillers, E., de la Fuente, J., Guillot, J. and Cabezas-Cruz, A. 2020. Gut microbiota abrogates anti-α-gal IgA response in lungs and protects against experimental *Aspergillus* infection in poultry. – Vaccines (Basel) 8: 285.

McKenzie, V. J., Kueneman, J. G. and Harris, R. N. 2018. Probiotics as a tool for disease mitigation in wildlife: insights from food production and medicine. – Annals of the New York Academy of Sciences 1429: 18–30.

McMurdie, P. J. and Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. – PLoS ONE 8: e61217.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H. 2019. vegan: Community Ecology Package.

R Core Team. 2020. R: A language and environment for statistical computing. – Foundation for Statistical Computing, Vienna, Austria.

Rausch, T., Fritz, M. H.-Y., Untergasser, A. and Benes, V. 2020. Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. – BMC Genomics 21: 230.

Redweik, G. A. J., Jochum, J. and Mellata, M. 2020. Live Bacterial Prophylactics in Modern Poultry. – Frontiers in Veterinary Science.

San Juan, P. A., Castro, I. and Dhami, M. K. 2021. Captivity reduces diversity and shifts composition of the Brown Kiwi microbiome. – anim microbiome 3: 48.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. – Front Zool 2: 16.

Stedman, A., van Vliet, A. H. M., A. Chambers, M. and Gutierrez-Merino, J. 2020. Gut commensal bacteria show beneficial properties as wildlife probiotics. – Annals of the New York Academy of Sciences 1467: 112–132.

Stough, J. M. A., Dearth, S. P., Denny, J. E., LeCleir, G. R., Schmidt, N. W., Campagna, S. R. and Wilhelm, S. W. 2016. Functional characteristics of the gut microbiome in C57BL/6 mice differentially susceptible to *Plasmodium yoelii*. – Front Microbiol 7: 1520.

Taniguchi, T., Miyauchi, E., Nakamura, S., Hirai, M., Suzue, K., Imai, T., Nomura, T., Handa, T., Okada, H., Shimokawa, C., Onishi, R., Olia, A., Hirata, J., Tomita, H., Ohno, H., Horii, T. and Hisaeda, H. 2015. *Plasmodium berghei* ANKA causes intestinal malaria associated with dysbiosis. – Sci Rep 5: 15699.

Valkiunas, G. 2005. Avian malaria parasites and other Haemosporidia. – Boca Raton: CRC Press.

Valkiūnas, G. and Iezhova, T. A. 2017. Exo-erythrocytic development of avian malaria and related haemosporidian parasites. – Malar J 16: 101.

Valkiūnas, G. and Iezhova, T. A. 2018. Keys to the avian malaria parasites. – Malar J 17: 212.

van Riper, C., van Riper, S. G., Goff, M. L. and Laird, M. 1986. The Epizootiology and Ecological Significance of Malaria in Hawaiian Land Birds. – Ecological Monographs 56: 327–344.

Videvall, E., Marzal, A., Magallanes, S., Fleischer, R. C., Espinoza, K. and García-Longoria, L. 2021. The uropygial gland microbiome of house sparrows with malaria infection. – Journal of Avian Biology <https://doi.org/10.1111/jav.02686>.

Villarino, N. F., LeCleir, G. R., Denny, J. E., Dearth, S. P., Harding, C. L., Sloan, S. S., Gribble, J. L., Campagna, S. R., Wilhelm, S. W. and Schmidt, N. W. 2016. Composition of the gut microbiota modulates the severity of malaria. – Proc Natl Acad Sci USA 113: 2235–2240.

Waldenström, J., Bensch, S., Hasselquist, D. and Östman, Ö. 2004. A New Nested Polymerase Chain Reaction Method Very Efficient in Detecting Plasmodium and Haemoproteus Infections From Avian Blood. – Journal of Parasitology 90: 191–194.

Warner, R. E. 1968. The Role of Introduced Diseases in the Extinction of the Endemic Hawaiian Avifauna. – The Condor 70: 101–120.

Westerdahl, H., Waldenström, J., Hansson, B., Hasselquist, D., von Schantz, T. and Bensch, S. 2005. Associations between malaria and MHC genes in a migratory songbird. – Proc Biol Sci 272: 1511–1518.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. 2019. Welcome to the Tidyverse. – JOSS 4: 1686.

Woodhams, D. C., Bletz, M., Kueneman, J. and McKenzie, V. 2016. Managing Amphibian Disease with Skin Microbiota. – Trends in Microbiology 24: 161–164.

Woodhams, D. C., Rollins-Smith, L. A., Reinert, L. K., Lam, B. A., Harris, R. N., Briggs, C. J., Vredenburg, V. T., Patel, B. T., Caprioli, R. M., Chaurand, P., Hunziker, P. and Bigler, L. 2020. Probiotics Modulate a Novel Amphibian Skin Defense Peptide That Is Antifungal and Facilitates Growth of Antifungal Bacteria. – Microb Ecol 79: 192–202.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F. O. 2014a. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. – Nucleic Acids Res. 42: D643–D648.

Yilmaz, B., Portugal, S., Tran, T. M., Gozzelino, R., Ramos, S., Gomes, J., Regalado, A., Cowan, P. J., d'Apice, A. J. F., Chong, A. S., Doumbo, O. K., Traore, B., Crompton, P. D., Silveira, H. and Soares, M. P. 2014b. Gut microbiota elicits a protective immune response against malaria transmission. – Cell 159: 1277–1289.

Yooseph, S., Kirkness, E. F., Tran, T. M., Harkins, D. M., Jones, M. B., Torralba, M. G., O'Connell, E., Nutman, T. B., Doumbo, S., Doumbo, O. K., Traore, B., Crompton, P. D. and Nelson, K. E. 2015. Stool microbiota composition is associated with the prospective risk of *Plasmodium falciparum* infection. – BMC Genomics <https://doi.org/10.1186/s12864-015-1819-3>.

CHAPTER 4.

Major histocompatibility complex diversity associated with gut microbiome composition in a wild sparrow

Department of Biology and Whitney R. Harris World Ecology Center, University of Missouri-St. Louis, One University Blvd., St. Louis, MO 63121, USA

Abstract:

The major histocompatibility complex (MHC) plays a key role in the detection of pathogens by the host immune system, and surprisingly high MHC diversity is maintained across most vertebrate taxa. Balancing selection due to pathogen pressure is thought to be an important force driving this high genetic diversity. At a population level, higher MHC diversity can provide a wider arsenal of responses to invading pathogens, while heterozygote advantage and divergent allele advantage have been proposed as mechanisms in which higher MHC diversity can improve individual fitness. However, MHC variation may also be maintained as a way to modulate the gut microbiome, most likely through immune recognition of antigens aided by MHC molecules. Using an introduced species, the Eurasian tree sparrow (*Passer montanus*), we collected fecal and blood samples to characterize MHC diversity across sites in the Saint Louis region and to test for correlations with the gut microbiome. Results show

that MHC effects on the microbiome in this system are overshadowed by environmental effects such as season, but small associations are evident in differentially abundant bacteria. Individuals with higher MHC diversity have several differentially abundant taxa compared to individuals with lower MHC diversity, providing slight support for the divergent allele advantage hypothesis. These results indicate that the major histocompatibility complex does play a role in shaping the microbiome, but it is not a primary driver in this system.

Keywords: MHC, avian microbiome, gut microbiome, immunogenetics

Original Article to be submitted.

Introduction:

One of the most polygenic and polymorphic regions of vertebrate genomes is the major histocompatibility complex (MHC). However, the evolutionary forces maintaining that strikingly high diversity across taxa are not entirely clear (Garrigan and Hedrick 2003, Sommer 2005). MHC proteins are essential for pathogen recognition by the host immune system, with MHC Class I genes encoding proteins that help identify antigens from intracellular pathogens while Class II genes are associated with extracellular pathogens (Minias et al. 2018). Since each MHC protein can recognize a unique range of antigens, higher MHC variation is associated with greater flexibility in population-level responses to new pathogens (Sommer 2005, Borg et al. 2011). Thus, balancing selection due to pathogen pressure is thought to be an important force maintaining high genetic variability in MHC genes across most vertebrates (Garrigan and Hedrick 2003, Sommer 2005).

Even at an individual level, higher MHC variation may promote host fitness. The hypothesis of "heterozygote advantage" suggests that individuals who are heterozygous for MHC genes will have greater resistance to pathogens than homozygotes. Support for this idea is mixed (Doherty and Zinkernagel 1975, Wakeland et al. 1990, Penn et al. 2002). The "divergent allele advantage" hypothesis counters that simple heterozygosity matters less than the actual level of polymorphism between alleles of a heterozygote (Wakeland et al. 1990, Froeschke and Sommer 2012). In other words, heterozygotes with highly varying alleles may be more resilient to pathogen pressure than heterozygotes with similar alleles (Froeschke and Sommer 2012). Specific alleles can

additionally make a host more or less likely to be parasitized by a specific microorganism, or they can simply have a modulating effect on infection severity (Bonneaud et al. 2006, Loiseau et al. 2008, 2011, Westerdahl et al. 2013, Karlsson et al. 2015). Wild populations may exhibit a mixture of these patterns – for example, one study examining susceptibility of great reed warblers (*Acrocephalus arundinaceus*) to three species of malarial parasites found that either possessing one specific MHC I allele or having a greater number of alleles in general made a host less likely to contract a severe case of malaria (Westerdahl et al. 2013). More work is needed in wild populations to determine how widely applicable these patterns are across taxa.

However, a newer hypothesis posits that high MHC diversity is also evolutionarily maintained to regulate resident microbial communities by contributing to antigen recognition of the host immune system (Kubinak et al. 2015, Khan et al. 2019). This benefit of higher MHC diversity would thus be an extension of the divergent allele advantage hypothesis (Sommer 2005, Bolnick et al. 2014). Laboratory studies in mice and rats have established that differences in MHCII genes between individuals under highly controlled conditions result in markedly different microbiomes (Toivanen et al. 2001, Lin et al. 2014, Kubinak et al. 2015, Khan et al. 2019). Field studies on this topic are few, but specific MHC motifs have been associated with variations in bacterial family abundances between the gut microbiomes of wild three-spined sticklebacks (*Gasterosteus aculeatus*), although diversity levels did not correlate (Bolnick et al. 2014). A recent study in wild Seychelles warblers (*Acrocephalus sechellensis*) found a significant association between two MHCIIB alleles and gut microbiome composition, although the

study used only total allele numbers to measure MHC diversity (rather than quantifying variation between alleles) and did not find a relationship between MHC diversity and the microbiome (Davies et al. 2022). Additionally, a significant association between MHC IIB heterozygosity and uropygial gland microbiota has been noted in male Leach's storm petrels (*Oceanodroma leucorhoa*), though the sample size was quite small (*n*=8) (Pearce et al. 2017). In blue petrels (*Halobaena caerulea*), higher MHC IIB diversity has been associated with decreased diversity in feather microbiota (*Halobaena caerulea*), matching study predictions that higher MHC diversity would lead to lower microbiome diversity due to greater antigen recognition capacity (Leclaire et al. 2019). Whether the MHC-microbiome relationship is present across other taxa in the wild is unknown, and the importance of divergent alleles to shape the gut microbiome remains unclear.

We tested associations between MHC diversity, specific alleles, and the gut microbiome in an introduced species, the Eurasian tree sparrow (*Passer montanus*). Passerines have one of the highest degrees of polymorphism at MHC loci of all major vertebrate groups (Bollmer et al. 2010, Minias et al. 2018). However, the North American population of the Eurasian tree sparrow underwent a founding event when ~20 individuals were introduced to St. Louis from Germany in 1870, and since then the species has only slightly expanded its range beyond the St. Louis region to neighboring states (Burnett et al. 2017). Previous studies have established that this introduction led to a founder effect in the population (Louis and Barlow 1988), but balancing selection on MHC loci is likely still acting on this population to some degree (Borg et al. 2011). This population provides the opportunity to test these widely applicable patterns in a species which is

not of conservation concern, while also shedding light on how MHC diversity is maintained after bottleneck events.

The aim of this work was to further characterize the genetic diversity of this introduced species using neutral markers and MHC loci, to examine microbial variation at a fine geographic scale across several sites in the region, and to test the hypothesis of divergent allele advantage in the context of the gut microbiome. We predicted that certain alleles would be associated with microbial community metrics. We expected that higher MHC diversity would correlate with lower alpha diversity in the microbial community due to increased pathogen recognition, we anticipated beta diversity clustering between individuals with high versus low MHC diversity, and we expected differential abundances of microbial taxa between individuals with varying MHC diversity.

Methods:

We captured Eurasian tree sparrows (N = 135) using mist nets at six sites in Saint Louis County and Saint Charles County (Missouri, USA) in 2020 and 2021. Birds were banded with numbered color bands to identify recaptures, wing chord and body mass were measured, and fecal samples were collected using bags with false floors (Knutie and Gotanda 2018). Age was recorded based on the plumage variation visible between juveniles (prior to first molt) and adults (after first molt). Small blood samples (<60ul) were collected from each bird via the brachial vein to preserve in Longmire's solution (Longmire et al. 1997). Thin blood smears were also collected, fixed in methanol within one hour, and stained with Giemsa within one month. Fecal samples were immediately

placed on ice in the field and taken to frozen storage later that day. All birds were released in the same capture site after sampling. This handling protocol is minimally invasive (Sheldon et al. 2008), and an UMSL IACUC protocol was approved prior to sampling (#1568558). A Missouri Department of Conservation Wildlife Collector's Permit was also obtained (#18682).

Microbiome Characterization

DNA was extracted from fecal samples using the QiaAMP Power Fecal Pro DNA Kit using manufacturer instructions. A negative sampling control, negative extraction control, and positive extraction control (ZymoBIOMICS Gut Microbiome Standard) were extracted with the fecal samples using the same protocols. Extracted DNA was measured by a Qubit fluorometer and samples with more than 200ng of DNA were sent to the University of Michigan Medical School Microbiome Core. To characterize the gut microbiome, we used sequences of the V4 region of the 16S rRNA gene (Kozich et al. 2013). Sequencing was performed on an Illumina MiSeq by the Microbiome Core for a total of three separate sequencing runs in 2020 and 2021. The negative and positive sampling and extraction controls were included for the sequencing runs, and the sequencing core added a negative sequencing control (water) and positive sequencing control (ZymoBIOMICS Microbial Community Standard) to each run. The Dada2 pipeline was used to process the 16S sequencing reads (Callahan et al. 2016). Recommended parameters were used for Dada2 quality filtering and trimming, merged reads outside of the 250-256bp range were excluded, and SILVA v. 138 was used to assign taxonomies to amplicon sequence variants (ASVs) (Yilmaz et al. 2014). Further filtering was performed

in R v. 4.2.0 using the package phyloseq to remove all reads matching mitochondria, chloroplasts, or Archaea (McMurdie and Holmes 2013). The package decontam was then used with the negative controls to identify ASVs which were likely contaminants in R; this process identified two ASVs as putative contaminants and they were removed from each sample (Davis et al. 2018). A minimum threshold of 3900 reads was applied and all samples with read counts below that number were excluded.

Microsatellites:

DNA from blood samples was extracted at the University of Missouri-St. Louis (UMSL) using a standard phenol-chloroform protocol. Nine microsatellite loci were selected from the literature based on their previous use with the Eurasian tree sparrow and their high number of previously observed alleles: Pamo1, Pamo7, Pamo8, Pamo12 (Izumi et al. 2009), Pdo3, Pdo5, Pdo9, (Griffith et al. 2007, Seress et al. 2007), ZC02, and Ctc105 (Tarvin 2006, Poesel et al. 2009, Yang et al. 2020). A multiplex PCR was performed for each sample using the Qiagen Type-it Microsatellite PCR Kit with fluorescently labeled primers (with each locus assigned to a unique dye and fragment size combination) and an annealing temperature of 57°C. Amplicons were diluted 1:10 with purified water and sent to the University of Missouri Genomics Technology Core for fragment analysis. A sample was amplified as a control for each PCR to verify that results were consistent between reactions. One locus, Pdo9, amplified poorly in the multiplex and was excluded from the analysis. The program Geneious Prime v. 2022.2.2 was used to call peaks and classify alleles (https://www.geneious.com). Pairwise F_{st} values were calculated from

the microsatellite data in Gene Pop v. 4.7.5 to examine potential population differentiation.

MHC Genotyping:

We isolated and amplified a 159bp segment of the classical MHC IIB region (Burri et al. 2014) using barcoded primers developed in Dr. John Eimes' lab (Sungkyunkwan University, South Korea) to target the MHC IIB region in Eurasian tree sparrows (unpublished data). The MHC IIB region was chosen based on previous studies which focused on class II due to its potential for interaction with extracellular bacteria (Bolnick et al. 2014). The barcoded amplicons were pooled and sent to the University of Missouri Genomics Technology Core for Illumina sequencing for a total of two separate sequencing runs. Raw Illumina reads were merged using BBmerge and quality trimming was performed with BBduk (minlen=200, qtrim=rl, trimq=30) using the Lewis computing cluster at the University of Missouri ("BBTools" in press). The cleaned and merged reads were then demultiplexed and genotyped using the clustering approach offered in the AmpliSAS pipeline, which provided a degree of normalization by using 5000 reads per sample (Sebastian et al. 2016). Amplicons with fewer than 5000 reads prior to genotyping were excluded. Following the methods described in Biedrzycka et al. 2017, we set the "minimum dominant frequency" parameter to 10%, which keeps variants which are similar to each other but excludes many high-frequency artefacts (Biedrzycka et al. 2017). The "maximum number of alleles per amplicon" parameter was set to 30. This high-throughput method allows for more effective MHC characterization, since passerines typically have highly complex MHC regions with many loci (Minias et al. 2018,

O'Connor et al. 2019). We first ran the clustering program on the preliminary data set from the first sequencing run at regularly spaced thresholds of "minimum amplicon frequency" (0%, 0.1%, 0.2%, 0.3%, etc.) to determine a reasonable variant frequency cutoff for this system. We ultimately set 0.2% as the "minimum amplicon frequency", based on the largest drop in the number of unique variants at that threshold, as described in other studies (Biedrzycka et al. 2017, Rekdal et al. 2018). This was near the value chosen for other studies in highly variable passerine MHC loci, such as 0.4% in Biedrzycka et al. 2017 and 0.2% in Rekdal et al. 2018. MHCII alleles for *Passer montanus* and *Passer domesticus* were taken from Genbank to use as allele references for the clustering pipeline.

Allelic Divergence:

Several measures of allelic divergence exist which generally correlate with each other, but we chose to use Grantham distance because a previous study showed it is the most consistent predictor of peptide binding capacity (Pierini and Lenz 2018). Grantham distance accounts for chemical properties of amino acid residues to create an estimate of bound peptides (Grantham 1974). We used the R package MHCtools to calculate average Grantham distances from the allelic nucleotide sequences corresponding to each sample (Roved et al. in press). For one part of the analysis, the sample with the median Grantham distance was identified, and 20 samples immediately above and below the median were excluded to create two distinct groups of individuals with higher Grantham distances and individuals with lower Grantham distances (N=88). This allowed

us to test the tails of the Grantham score distribution across the sample set. This smaller dataset will be referred to as the divergence subset.

MHC Diversity and the Microbiome:

Microbiome samples were rarefied to 3908 reads per sample using phyloseq. Relative abundances were then calculated from the rarefied dataset by dividing each taxon read count by the total number of reads (3908), and these relative abundances were used to visualize community composition for each individual. Rarefied ASV tables were used to calculate alpha diversity of each individual in the R package vegan, using three common alpha diversity metrics of Shannon diversity, observed richness, and Simpson's diversity (Oksanen et al. 2019). As Shannon diversity was normally distributed, ANOVAs were used to test variation in Shannon diversity between groups. Non-parametric Kruskal Wallis tests were used to assess statistical differences between individuals based on observed richness and Simpson's diversity. Beta diversity was measured for each sample pair using Bray-Curtis dissimilarity to create a distance matrix, and the results were visualized via Principal Coordinates Analysis (PCoA). Statistical significance in beta diversity between locations, MHC allele groups, and MHC diversity groups was tested using Permutational Multivariate Analysis of Variance (PERMANOVA) (Anderson 2001, 2017). PERMDISP was used to check significant PERMANOVA results for the potentially confounding effect of dispersion (Anderson and Walsh 2013).

Thirteen alleles with frequencies between 5% and 95% were tested for associations with alpha and beta diversity (Loiseau et al. 2011). For each allele, the confounding variables

of sequencing run, location, season, and age were included first in the model. Linear regressions were used to test whether Grantham distances correlate with microbial alpha diversity.

Using the divergence subset (described above), groups with high versus low MHC diversity were tested for associations with alpha and beta diversity as well as differentially abundant bacteria. Clean reads (without rarefaction) were used with the R package DESeq2 to calculate differential abundances of ASVs based on MHC diversity (Love et al. 2014). Phyloseq was first used to trim the dataset to taxa found at least 3 times in at least 5% of samples. DESeq2 was then used with default settings, including confounding variables of sequencing run, location, season, and age first in the model. Individuals with lower MHC diversity (Grantham distance averages) were set as the reference level for the model. The Benjamini and Hochberg method was used for false discovery correction, and a lower adjusted *p* value threshold of 0.01 was used to reduce false positives (Benjamini and Hochberg 1995).

Results:

Gut microbiome geographic variation:

Gut microbiome composition was highly variable across individuals both between and within sampling sites (Fig. 1). While most taxa belonged to the same three phyla (*Firmicutes*, *Proteobacteria*, and *Actinobacteriota*), the taxonomic makeup at lower levels was much more diverse; at a family level, 17 families were found in abundances

greater than 1%. The most abundant family was *Catellicoccaceae*, followed by *Staphylococcaceae, Enterobacteriaceae, Lactobacillaceae,* and *Micrococcaceae.*

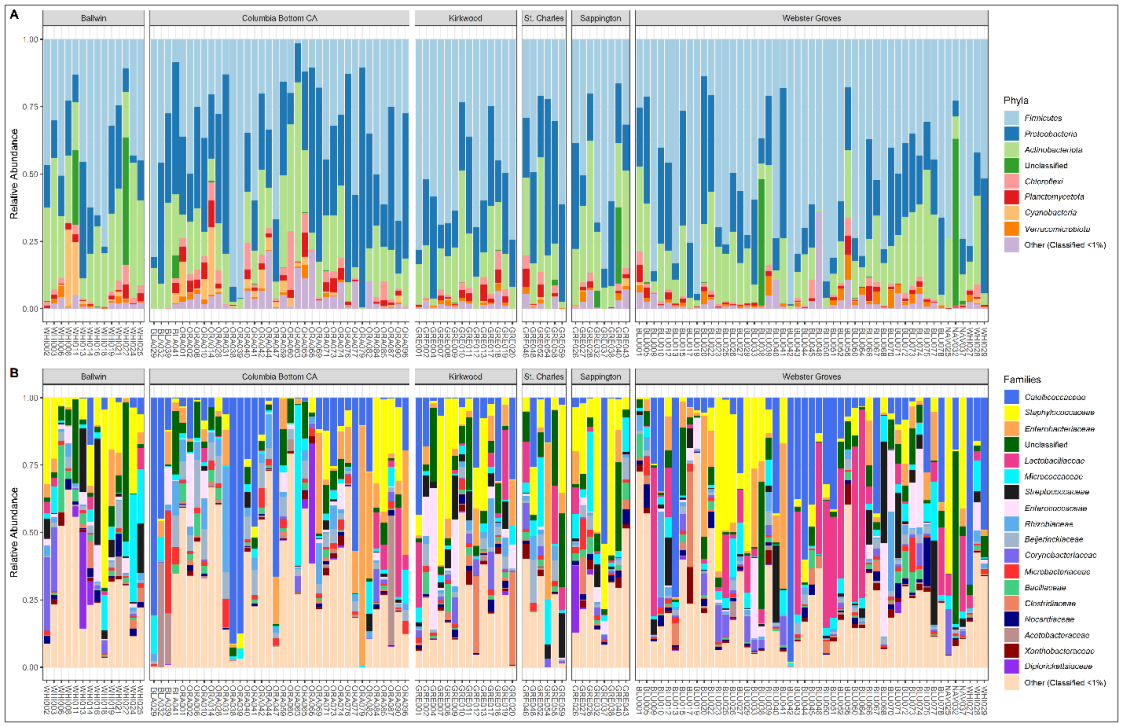

Figure 1. Microbial diversity shown as relative abundance at phylum (A) and family (B) levels across six sampling sites, with columns representing individual birds.

Alpha diversity was high in these communities, with Shannon diversity averaging between 2.5 and 3.5 for each sampling site (Fig. 2A). Shannon diversity did not significantly differ between sites, however, as an ANOVA controlling for confounding variables was not significant (Fig. 2A). A PCoA visualizing beta diversity demonstrated substantial overlap between community composition at each site (Fig. 2B). Several of the sites with smaller sample sizes overlapped almost completely. PERMANOVA was used to test significance between different sites based on beta diversity, and location significantly varied when first controlling for sequencing batch, age, and season in the

model (9999 permutations; $R^2 = 0.06726$, P = 0.001; Table 1). However, PERMDISP was also significant for location (9999 permutations, P = 0.0003), indicating that the significant PERMANOVA result may be due to dispersion variance. The other variables which significantly differed based on beta diversity were age ($R^2 = 0.01741$, P = 0.001) and season ($R^2 = 0.04324$, P = 0.001). Sequencing run and sex were not significant.

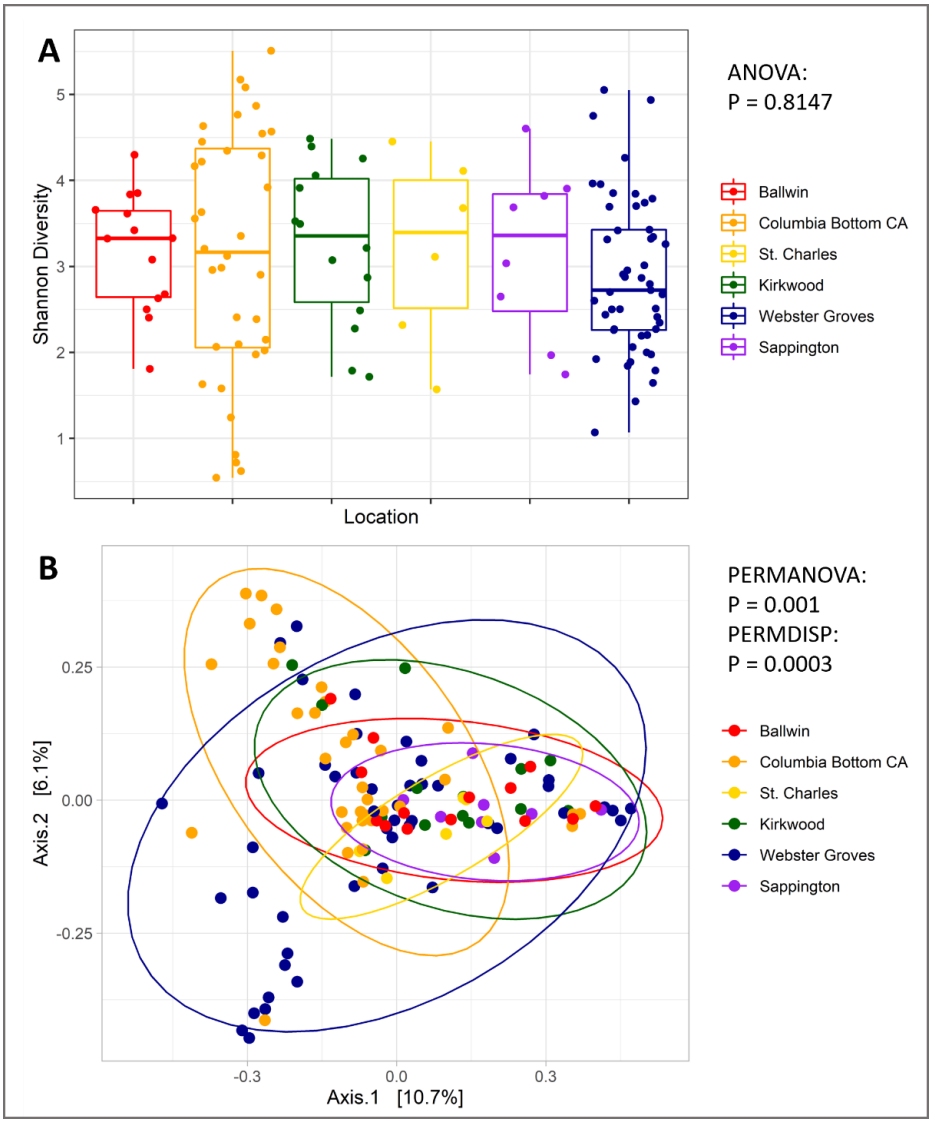

Figure 2. Shannon diversity remained broadly consistent across all sampling sites (P = 0.8147) (A). A PCoA visualizing beta diversity showed substantial overlap between sample sites (B). A

PERMANOVA was significant for location (9999 permutations, $R^2 = 0.06726$, P = 0.001), but PERMDISP was also significant (P = 0.0003), indicating confounding dispersion effects.

	Df	Sum of Sqs	R2	F	Pr(>F)
Sequencing Run	2	0.884	0.01683	1.1425	0.15
Sex	1	0.446	0.0085	1.1544	0.214
Age	1	0.914	0.01741	2.3645	0.001***
Season	2	2.27	0.04324	2.936	0.001***
Location	5	3.532	0.06726	1.8268	0.001***
Residual	115	44.466	0.84677		
Total	126	52.513	1		

Table 1. Significant PERMANOVA results for location using Bray-Curtis dissimilarity, controlling for sequencing run, sex, age, and season. Both age and season also significantly varied.

Genetic variation:

Putatively neutral markers (microsatellites) were used to estimate Eurasian tree sparrow population differentiation between sites in the St. Louis region. Pairwise F_{st} values calculated from microsatellites are shown for each locus and for combined loci in Table 2. The F_{st} results were very low, with all values less than 0.1 and most values approximately 0, indicating high levels of gene flow between sites.

	Webster Groves	Kirkwood	Columbia Bottom CA	Ballwin	Sappington
	Pamo8				
Kirkwood	0.0160				
Columbia Bottom CA	-0.0032	0.0213			
Ballwin	-0.0027	0.0221	-0.0088		
Sappington	-0.0152	-0.0103	-0.0233	-0.0432	
Saint Charles	0.0485	0.0626	0.0112	0.0203	-0.0006

	Pamo12				
Kirkwood	-0.0144				
Columbia Bottom CA	0.0097	-0.0055			
Ballwin	-0.0031	-0.0216	0.0260		
Sappington	0.0359	0.0356	0.0719	0.0285	
Saint Charles	-0.0241	-0.0309	-0.0329	-0.0111	0.0026

	Pamo1				
Kirkwood	0.0298				
Columbia Bottom CA	-0.0033	0.0203			
Ballwin	-0.0028	-0.0083	-0.0112		
Sappington	-0.0017	0.0726	0.0120	0.0052	
Saint Charles	-0.0246	0.0159	-0.0235	-0.0331	-0.0112

	Pdo5				
Kirkwood	0.0128				
Columbia Bottom CA	0.0150	-0.0127			
Ballwin	0.0096	-0.0007	0.0047		
Sappington	-0.0101	-0.0074	0.0043	-0.0180	
Saint Charles	0.0353	-0.0147	0.0055	0.0302	0.0127

	Pdo3				
Kirkwood	-0.0067				
Columbia Bottom CA	0.0109	0.0177			
Ballwin	-0.0047	-0.0022	-0.0089		
Sappington	-0.0062	0.0060	0.0330	0.0157	
Saint Charles	-0.0004	0.0009	0.0013	-0.0052	0.0043

	ZC02				
Kirkwood	0.0363				
Columbia Bottom CA	0.0092	0.0123			
Ballwin	0.0165	0.0566	0.0066		
Sappington	0.0082	0.0867	0.0067	0.0550	
Saint Charles	0.0136	0.0433	0.0357	0.0917	0.0373

	Pamo7				
Kirkwood	0.0369				
Columbia Bottom CA	-0.0006	0.0660			
Ballwin	-0.0036	0.0748	0.0008		
Sappington	-0.0124	0.0480	-0.0343	-0.0128	
Saint Charles	-0.0211	0.0804	-0.0260	-0.0442	-0.0271

	Ctc105				
Kirkwood	-0.0151				
Columbia Bottom CA	0.0015	-0.0101			
Ballwin	-0.0034	-0.0131	0.0052		
Sappington	-0.0025	-0.0211	0.0100	0.0234	
Saint Charles	0.0810	0.0550	0.0387	0.1178	0.0545

	Combined Loci				
Kirkwood	**0.0123**				
Columbia Bottom CA	**0.0049**	**0.0134**			
Ballwin	**0.0009**	**0.0138**	**0.0017**		
Sappington	**-0.0011**	**0.0251**	**0.0104**	**0.0071**	
Saint Charles	**0.0165**	**0.0284**	**0.0031**	**0.0244**	**0.0105**

Table 2. F_{st} values calculated from microsatellite data indicated little population structuring.

A total of 35 MHC IIB alleles were detected in the 135 samples included in this analysis (Fig. 3). Six alleles were fixed (100%) or nearly fixed (>91%) in the population - PadoIIB_02, PadoIIB_22, PadoIIB_25, PadoIIB_09, PamoIIB_26 and PadoIIB_10. The number of total alleles per individual was high, ranging from 5-13 with an average of eight. Fourteen alleles occurred in less than 5% of the samples.

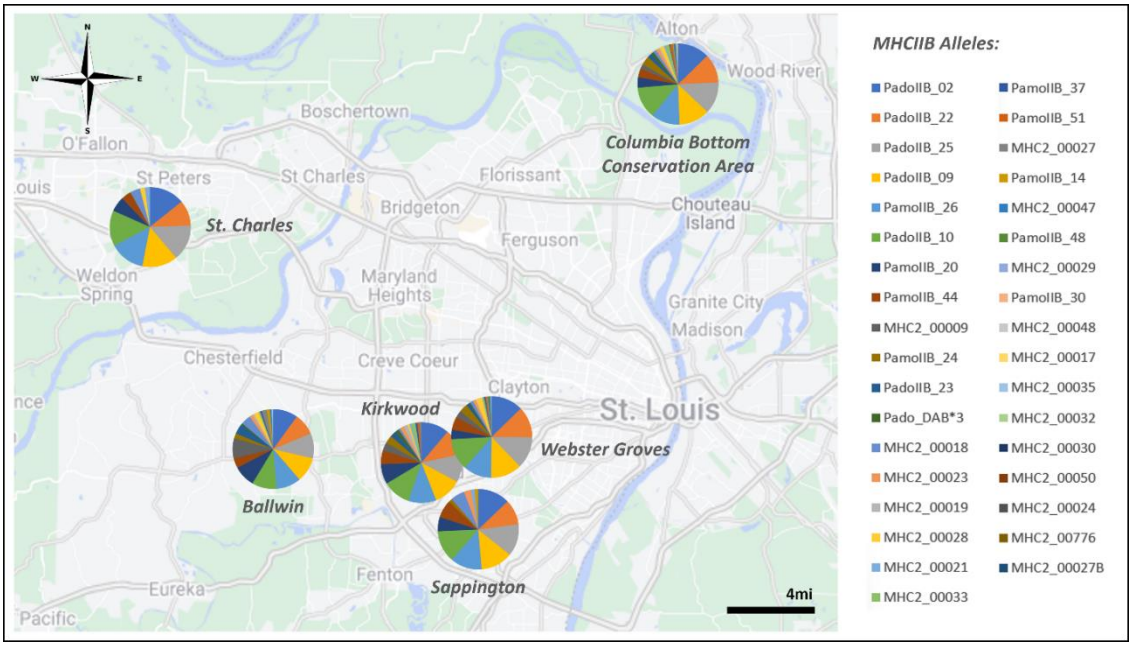

Figure 3. MHC IIB allelic variation across six sampling sites.

Allelic variation and the microbiome:

Specific MHC alleles did not significantly vary with the microbiome when controlling for potentially confounding variables. All alleles were non-significant when tested with Shannon diversity (alpha diversity) using ANOVAs, and no other variable significantly varied based on alpha diversity – sequencing run, location, season, and age were all non-significant. All alleles were similarly non-significant when tested with beta diversity, and no significant clustering patterns were observed.

High vs. low MHC diversity:

Relative abundances for the microbiome composition of individuals with higher Grantham distances (high diversity) vs. individuals with lower Grantham distances (low

diversity) from the divergence subset dataset are visualized in Figure 4 at a phylum (A) and family (B) level.

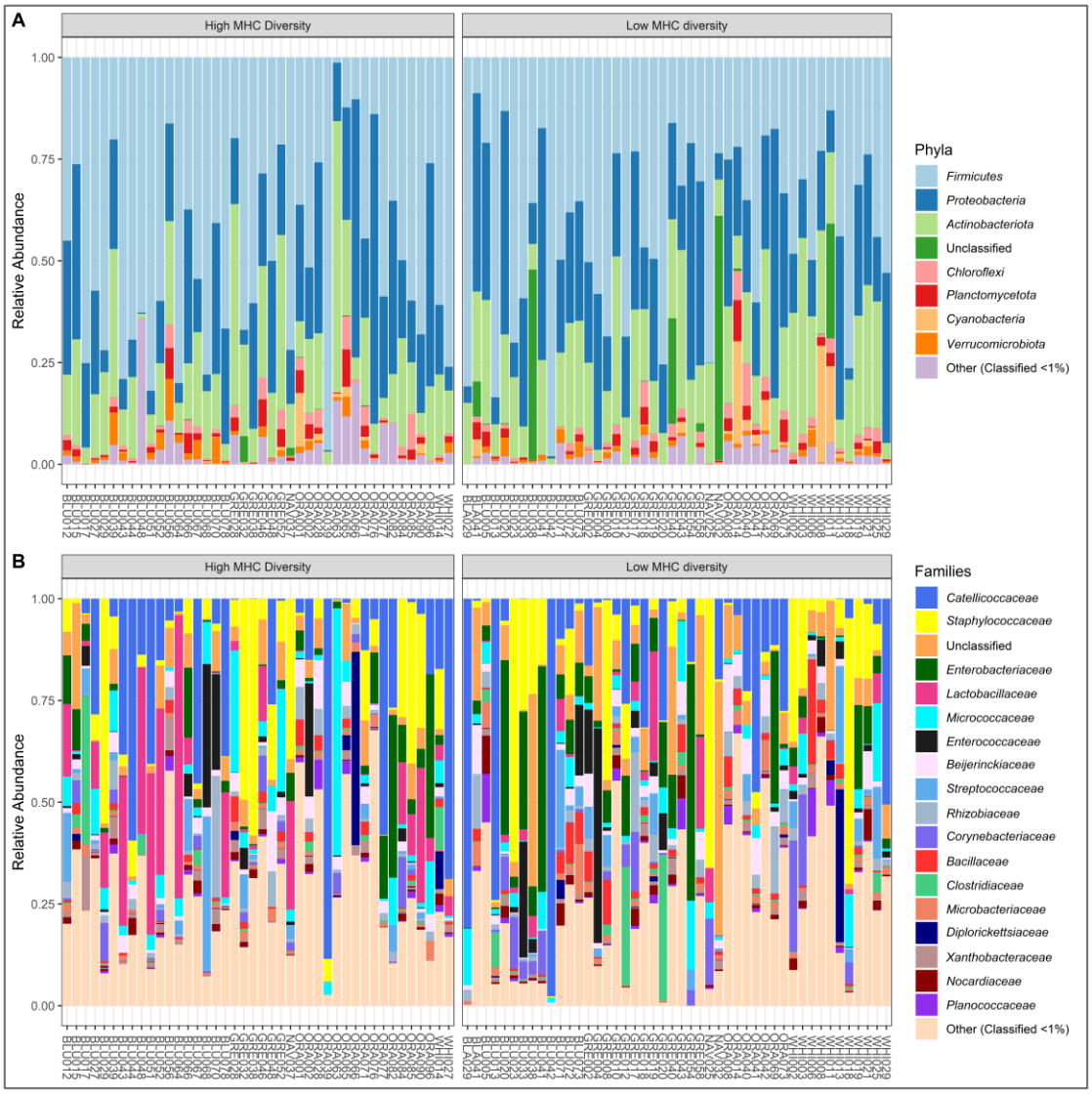

Figure 4. Phyla representing >1% of the communities (A) and families representing >1% of the communities (B) visualized in samples with high versus low MHC diversity (samples near median excluded).

Alpha diversity was measured as observed richness, Shannon diversity, and Simpson's diversity for each group of high and low MHC diversity (Fig. 5). An ANOVA testing

Shannon diversity variation between MHC diversity groups was not significant (Table 3). Kruskal Wallis tests assessing observed richness and Simpson's diversity between groups were not significant.

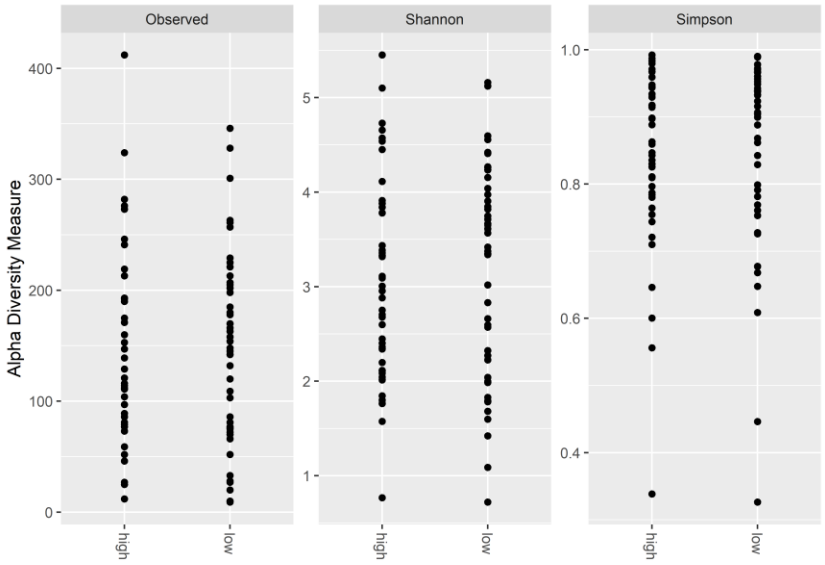

Figure 5. Alpha diversity did not vary between samples with high versus low MHC diversity (P > 0.05), measured as Grantham distances.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Sequencing Run	2	0.46	0.2312	0.182	0.834
Location	5	5.4	1.0796	0.85	0.519
Season	2	0.98	0.491	0.387	0.681
Age	1	0	0.0021	0.002	0.968
MHC IIB Diversity	1	0.02	0.0239	0.019	0.891
Residual	76	96.53	1.2701		

Table 3. ANOVA results for Shannon diversity between groups with high vs. low MHC diversity.

Microbial beta diversity was also examined between groups with high versus low MHC diversity. A PCoA made with Bray-Curtis dissimilarity demonstrated extensive overlap between the groups (Fig. 6). A PERMANOVA confirmed that beta diversity did not

significantly differ in groups with high MHC diversity (R^2 = 0.00971, P = 0.6342; Table 4), although location and season were still significant (Table 4).

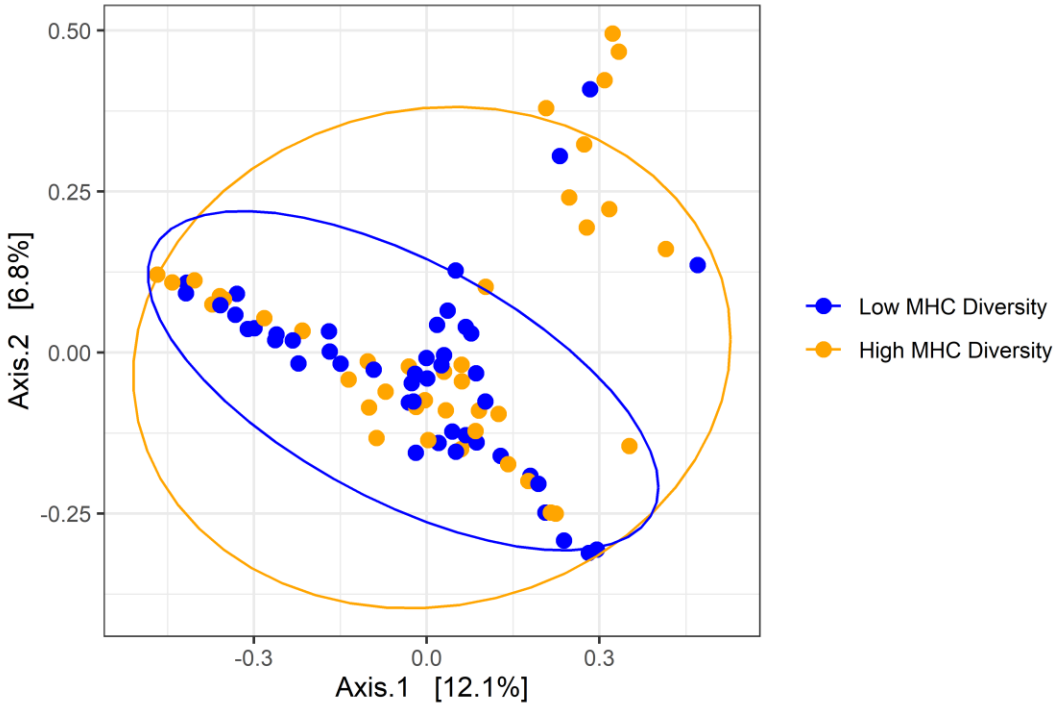

Figure 6. Principal Coordinates Analysis visualizing microbiome beta diversity of individuals with low MHC diversity compared to those with high diversity (PERMANOVA, P = 0.6342, R^2 = 0.00971). Confidence ellipses indicate the variance for each group and demonstrate a high degree of overlap.

	Df	Sum of Sqs	R^2	F	Pr(>F)
Sequencing Run	2	0.795	0.02212	1.0243	0.383
Location	5	3.166	0.0881	1.6319	0.0001***
Season	2	1.703	0.0474	2.1949	0.0001***
Age	1	0.433	0.01204	1.1152	0.2565
MHC IIB Diversity	1	0.349	0.00971	0.8997	0.6342
Residual	76	29.486	0.82062		
Total	87	35.931	1		

Table 4. PERMANOVA results for beta diversity (Bray-Curtis dissimilarity) of the subset of samples divided by high and low Grantham distances.

Finally, DESeq2 analysis identified 18 different ASVs (from a total of 419 ASVs used for the analysis) with significantly different abundances in birds with high MHC diversity compared to those with low diversity (Fig. 7). Six ASVs were from the class *Alphaproteobacteria*, four were from the class *Bacilli*, three were from the class *Actinobacteria,* and the rest were from the classes *Chloroflexia, Polyangia, KD4-96, Planctomycetes,* and *Gammaproteobacteria.* The significantly varying taxa with the largest base means were the genera *Ligilactobacillus* (family *Lactobacillaceae*) with a base mean of 63, *Lactococcus* (*Streptococcaceae*) with a base mean of 55, and *Glutamicibacter* (*Micrococcaceae*) with a base mean of 43, all of which were significantly more abundant in individuals with higher MHC diversity.

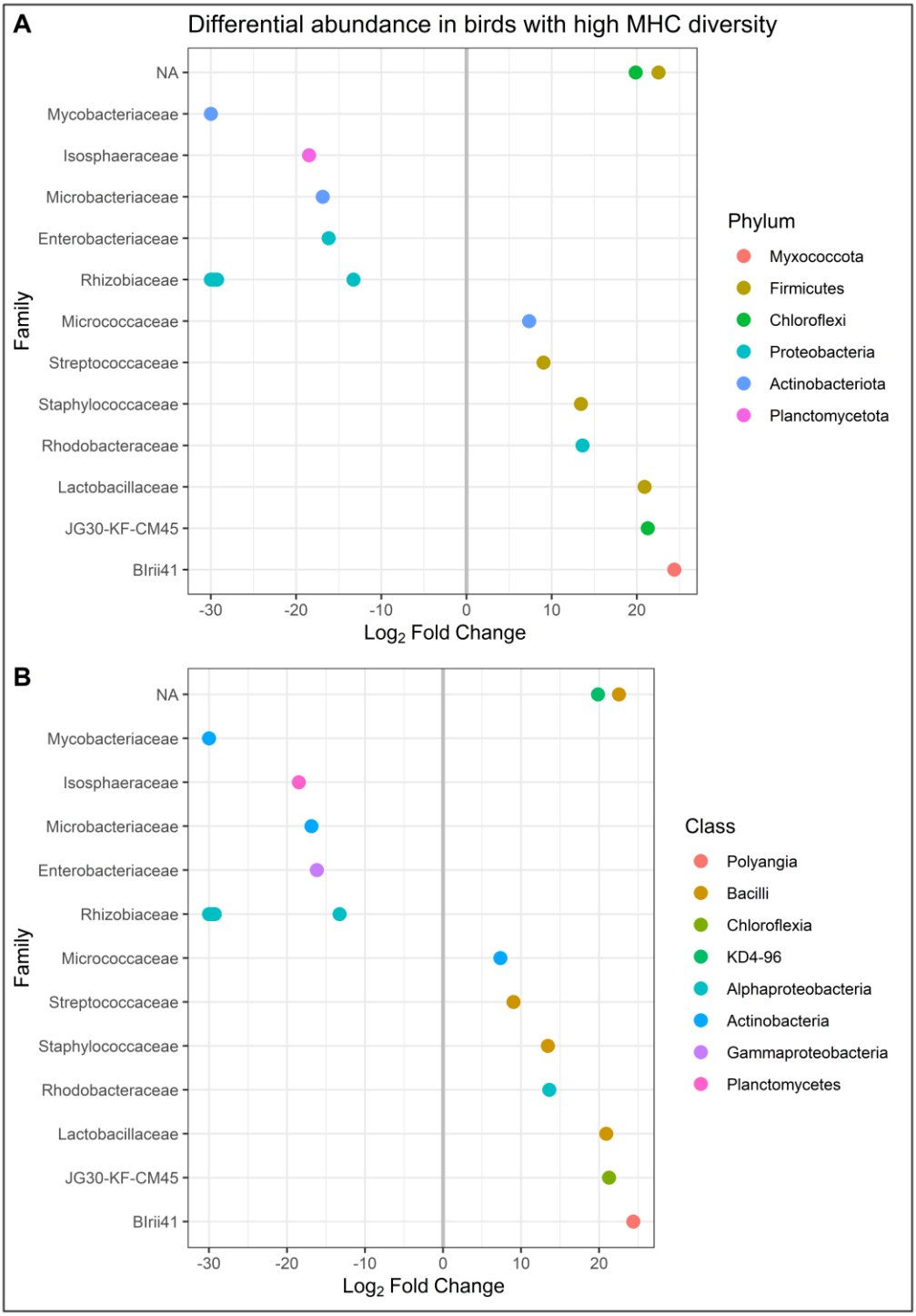

Figure 7. DESeq2 analysis showed ASVs that were differentially abundant in individuals with higher MHC diversity (measured as higher average Grantham distance). Each dot represents a

single ASV, categorized at a phylum level (A) and class level (B). All ASVs to the right of the grey bar were significantly more abundant in birds with higher MHC diversity, while all ASVs to the left were significantly less abundant in birds with higher MHC diversity. Two ASVs with significantly higher abundances lacked family-level classifications in the taxonomic database used, and they are indicated by "NA" in the figure.

Discussion:

These results demonstrated a consistently diverse gut microbiome in Eurasian tree sparrows across the St. Louis region. Alpha diversity values were high, but no significant relationship was observed between alpha diversity levels at different sampling sites. Beta diversity appeared to overlap substantially when visualized with a PCoA; a significant PERMANOVA result for location signaled some compositional differences between sites, but dispersion was also significant, confounding the initial result by violating the PERMANOVA assumption of homogeneous dispersion (Anderson and Walsh 2013). The most prevalent bacteria in the gut microbiomes of the Eurasian tree sparrows sampled for this study are strikingly different from the gut microbes found in captive Eurasian tree sparrows in a recent study in China (Jiang et al. 2020); at the broadest scale, the wild tree sparrows in our study had much higher abundances of *Firmicutes* but very low abundances of *Bacteroidetes* compared to the captive tree sparrows in China. These differences may simply be due to the captivity of the tree sparrows used for the other study, as captive animals are well known to have significantly varying microbiomes compared to their wild counterparts (Gibson et al. 2019, Alberdi et al. 2021).

From a genetic perspective, we found high levels of diversity at MHC IIB loci with no evidence for population differentiation in neutral markers. Microsatellite data revealed consistently low F_{st} values across sites. Since the sampling sites were within 30 miles of each other, high gene flow is unsurprising, although we never observed or recaptured banded individuals at any location other than their original capture site (the recapture rate was quite low at ~5% across all sampling sites). The tree sparrows exhibited the high MHC diversity often associated with passerines (Bollmer et al. 2010, Minias et al. 2018), ranging from 5-13 alleles per individual, with a mean and median of ~8 alleles.

We also detected a small but significant relationship between gut microbiome composition and the major histocompatibility complex in this species, although it was not reflected at higher community levels. Alpha diversity did not vary significantly between individuals with specific alleles or groups with high versus low MHC diversity (measured as Grantham distance averages), and beta diversity also did not significantly vary based on either specific allele presence or MHC diversity. However, differential abundance analysis revealed 18 bacterial ASVs with significantly higher and/or lower abundances in individuals with higher MHC diversity. The genera with the largest base means were *Ligilactobacillus* and *Lactococcus*, both members of the class *Bacilli.* These two genera were both significantly more abundant in individuals with high MHC diversity. This may indicate a more beneficial microbiome in the groups with higher MHC diversity, as members of *Bacilli* (and specifically *Ligilactobacillus*) are generally associated with better host health and are often targeted for probiotic use in environments such as poultry farms (Dec et al. 2021, Arif et al. 2021). However, the

varying abundances of the other bacterial groups are more difficult to interpret, since potential benefits or costs are less understood in those groups in the context of the avian microbiome and because base means were lower.

The association between high MHC diversity and putatively beneficial bacteria provides some support for the divergent allele advantage hypothesis (Davies et al. 2022). Differences in microbial community composition caused by MHC genotype could logically translate to direct fitness benefits or detriments for the hosts, as the microbiome is involved in varied tasks such as digestion, nutrient acquisition, and immune system training (Clemente et al. 2012). One study using laboratory mice found that the gut microbiomes of heterozygotes have more diversified functional capacity, greater metabolic capability, and increased resistance to invasion compared to homozygotes (Khan et al. 2019). This implies that higher MHC diversity may promote host fitness via the gut microbiome rather than incur costs, though a major caveat to this study was that their conclusions were drawn from functional predictions based on 16S rRNA gene sequencing rather than observations using shotgun metagenomic sequencing of bacterial communities (Khan et al. 2019). Our findings of putatively beneficial bacteria at higher abundances in birds with high MHC diversity are an interesting indication that MHC diversity may have measurable impacts on health-associated components of microbiomes in wild populations.

However, the scale of the associations found here indicates that MHC influence is substantially less important than environmental drivers of community variance, as we found significant differences and higher R^2 values in beta diversity when comparing

individuals across sampling sites and seasons. In some systems, genotype can be an important driver of microbial colonization patterns, and a study using wild sticklebacks found that MHC genotype had a stronger effect over microbiome composition than diet or sex, accounting for approximately 10% of microbial variation (Bolnick et al. 2014). However, in the tree sparrow system, genotype appears overshadowed by environmental heterogeneity observed in terms of geographic and seasonal differences. More work is needed to understand how much MHC diversity may shape the microbiomes of other taxa, and a functional microbiome perspective would be a useful addition. In all, these results show small-scale support for the divergent allele advantage hypothesis in the context of the microbiome, with implications for the maintenance of MHC diversity and host health in other vertebrate groups.

Acknowledgments:

We are grateful for the field and lab assistants who helped to collect samples and carry out laboratory analyses necessary for this work. We thank the Missouri Department of Conservation and Columbia Bottom Conservation Area for providing their permission and support for this project. The computation for this work was performed on the high performance computing infrastructure provided by Research Computing Support Services and in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO.

Funding:

This research was generously funded by the NSF Graduate Research Fellowship #1945995, the Harris World Ecology Center, Sigma Xi Grant in Aid of Research #G2020100193156089, NIH Grant #AI137984, and the Des Lee Collaborative Vision.

Ethics approval:

UMSL's Institutional Animal Care and Use Committee approved all bird handling protocols (Protocol # 1568558). The Missouri Department of Conservation provided sampling permissions through Wildlife Collector's Permits #18682 and #19163 and a Columbia Bottom Conservation Area Special Use Permit.

References:

Alberdi, A., Martin Bideguren, G. and Aizpurua, O. 2021. Diversity and compositional changes in the gut microbiota of wild and captive vertebrates: a meta-analysis. – Sci Rep 11: 22660.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. – Austral Ecol 26: 32–46.

Anderson, M. J. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). – Wiley StatsRef: Statistics Reference Online. American Cancer Society, pp. 1–15.

Anderson, M. J. and Walsh, D. C. I. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? – Ecol Monogr 83: 557–574.

Arif, M., Akteruzzaman, Md., Tuhin-Al-Ferdous, Islam, S. S., Das, B. C., Siddique, M. P. and Kabir, S. M. L. 2021. Dietary supplementation of Bacillus-based probiotics on the growth performance, gut morphology, intestinal microbiota and immune response in low biosecurity broiler chickens. – Veterinary and Animal Science 14: 100216.

BBTools. – DOE Joint Genome Institute.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. – J R Stat Soc Series B Stat Methodol 57: 289–300.

Biedrzycka, A., Sebastian, A., Migalska, M., Westerdahl, H. and Radwan, J. 2017. Testing genotyping strategies for ultra-deep sequencing of a co-amplifying gene family: MHC class I in a passerine bird. – Molecular Ecology Resources 17: 642–655.

Bollmer, J. L., Dunn, P. O., Whittingham, L. A. and Wimpee, C. 2010. Extensive MHC Class II B Gene Duplication in a Passerine, the Common Yellowthroat (Geothlypis trichas). – Journal of Heredity 101: 448–460.

Bolnick, D. I., Snowberg, L. K., Caporaso, J. G., Lauber, C., Knight, R. and Stutz, W. E. 2014. Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. – Mol Ecol 23: 4831–4845.

Bonneaud, C., Pérez-Tris, J., Federici, P., Chastel, O. and Sorci, G. 2006. Major histocompatibility alleles associated with local resistance to malaria in a passerine. – Evolution 60: 383–389.

Borg, Å. A., Pedersen, S. A., Jensen, H. and Westerdahl, H. 2011. Variation in MHC genotypes in two populations of house sparrow (*Passer domesticus*) with

different population histories: Variation in MHC Genotypes of House Sparrow. – Ecology and Evolution 1: 145–159.

Burnett, J., Roberts, C., Allen, C., Brown, M. and Moulton, M. 2017. Range expansion by Passer montanus in North America. – Biological Invasions 19: 5–9.

Burri, R., Promerová, M., Goebel, J. and Fumagalli, L. 2014. PCR-based isolation of multigene families: lessons from the avian MHC class IIB. – Molecular Ecology Resources 14: 778–788.

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

Clemente, J. C., Ursell, L. K., Parfrey, L. W. and Knight, R. 2012. The Impact of the Gut Microbiota on Human Health: An Integrative View. – Cell 148: 1258–1270.

Davies, C. S., Worsley, S. F., Maher, K. H., Komdeur, J., Burke, T., Dugdale, H. L. and Richardson, D. S. 2022. Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. – Microbiome 10: 41.

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. and Callahan, B. J. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. – Microbiome 6: 226.

Dec, M., Stępień-Pyśniak, D., Puchalski, A., Hauschild, T., Pietras-Ożga, D., Ignaciuk, S. and Urban-Chmiel, R. 2021. Biodiversity of Ligilactobacillus salivarius Strains from Poultry and Domestic Pigeons. – Animals (Basel) 11: 972.

Doherty, P. C. and Zinkernagel, R. M. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex | Nature. – Nature 256: 50–52.

Froeschke, G. and Sommer, S. 2012. Insights into the Complex Associations Between MHC Class II DRB Polymorphism and Multiple Gastrointestinal Parasite Infestations in the Striped Mouse. – PLoS ONE 7: e31820.

Garrigan, D. and Hedrick, P. W. 2003. Perspective: Detecting Adaptive Molecular Polymorphism: Lessons from the Mhc. – Evolution 57: 1707–1722.

Gibson, K. M., Nguyen, B. N., Neumann, L. M., Miller, M., Buss, P., Daniels, S., Ahn, M. J., Crandall, K. A. and Pukazhenthi, B. 2019. Gut microbiome differences between wild and captive black rhinoceros – implications for rhino health. – Sci Rep 9: 7570.

Grantham, R. 1974. Amino Acid Difference Formula to Help Explain Protein Evolution. – Science 185: 862–864.

Griffith, S. C., Dawson, D. A., Jensen, H., Ockendon, N., Greig, C., Neumann, K. and Burke, T. 2007. Fourteen polymorphic microsatellite loci characterized in the house sparrow Passer domesticus (Passeridae, Aves). – Molecular Ecology Notes 7: 333–336.

Izumi, H., Hasegawa, O. and Higashi, S. 2009. Isolation of polymorphic microsatellite markers in the tree sparrow (Passer montanus). – Molecular Ecology Resources 9: 245–247.

Jiang, J., He, Y., Kou, H., Ju, Z., Gao, X. and Zhao, H. 2020. The effects of artificial light at night on Eurasian tree sparrow (Passer montanus): Behavioral rhythm disruption, melatonin suppression and intestinal microbiota alterations. – Ecological Indicators 108: 105702.

Karlsson, M., Schroeder, J., Nakagawa, S., Smith, H. G., Burke, T. and Westerdahl, H. 2015. House sparrow Passer domesticus survival is not associated with MHC-I diversity, but possibly with specific MHC-I alleles. – Journal of Avian Biology 46: 167–174.

Khan, M. A. W., Stephens, W. Z., Mohammed, A. D., Round, J. L. and Kubinak, J. L. 2019. Does MHC heterozygosity influence microbiota form and function? – PLoS ONE 14: e0215946.

Knutie, S. A. and Gotanda, K. M. 2018. A Non-invasive Method to Collect Fecal Samples from Wild Birds for Microbiome Studies. – Microbial Ecology 76: 851–855.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina Sequencing Platform. – Appl. Environ. Microbiol. 79: 5112–5120.

Kubinak, J. L., Stephens, W. Z., Soto, R., Petersen, C., Chiaro, T., Gogokhia, L., Bell, R., Ajami, N. J., Petrosino, J. F., Morrison, L., Potts, W. K., Jensen, P. E., O'Connell, R. M. and Round, J. L. 2015. MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. – Nat Commun 6: 8642.

Leclaire, S., Strandh, M., Dell'Ariccia, G., Gabirot, M., Westerdahl, H. and Bonadonna, F. 2019. Plumage microbiota covaries with the major histocompatibility complex in blue petrels. – Molecular Ecology 28: 833–846.

Lin, P., Bach, M., Asquith, M., Lee, A. Y., Akileswaran, L., Stauffer, P., Davin, S., Pan, Y., Cambronne, E. D., Dorris, M., Debelius, J. W., Lauber, C. L., Ackermann, G., Baeza, Y. V., Gill, T., Knight, R., Colbert, R. A., Taurog, J. D., Van Gelder, R. N. and

Rosenbaum, J. T. 2014. HLA-B27 and Human β2-Microglobulin Affect the Gut Microbiota of Transgenic Rats. – PLoS ONE 9: e105684.

Loiseau, C., Zoorob, R., Garnier, S., Birard, J., Federici, P., Julliard, R. and Sorci, G. 2008. Antagonistic effects of a Mhc class I allele on malaria-infected house sparrows. – Ecology Letters 11: 258–265.

Loiseau, C., Zoorob, R., Robert, A., Chastel, O., Julliard, R. and Sorci, G. 2011. *Plasmodium relictum* infection and MHC diversity in the house sparrow (*Passer domesticus*). – Proc. R. Soc. B 278: 1264–1272.

Longmire, J. L., Baker, R. J., Maltbie, Mary., and Texas Tech University. 1997. Use of "Lysis Buffer" in DNA isolation and its implication for museum collections. – Museum of Texas Tech University.

Louis, V. L. S. and Barlow, J. C. 1988. Genetic Differentiation Among Ancestral and Introduced Populations of the Eurasian Tree Sparrow (passer Montanus). – Evolution 42: 266–276.

Love, M. I., Huber, W. and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. – Genome Biol 15: 550.

McMurdie, P. J. and Holmes, S. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. – PLoS ONE 8: e61217.

Minias, P., Pikus, E., Whittingham, L. A. and Dunn, P. O. 2018. A global analysis of selection at the avian MHC. – Evolution 72: 1278–1293.

O'Connor, E. A., Westerdahl, H., Burri, R. and Edwards, S. V. 2019. Avian MHC Evolution in the Era of Genomics: Phase 1.0. – Cells 8: 1152.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H. 2019. vegan: Community Ecology Package.

Pearce, D. S., Hoover, B. A., Jennings, S., Nevitt, G. A. and Docherty, K. M. 2017. Morphological and genetic factors shape the microbiome of a seabird species (Oceanodroma leucorhoa) more than environmental and social factors. – Microbiome 5: 146.

Penn, D. J., Damjanovich, K. and Potts, W. K. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. – Proceedings of the National Academy of Sciences 99: 11260–11264.

Pierini, F. and Lenz, T. L. 2018. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. – Molecular Biology and Evolution 35: 2145–2158.

Poesel, A., Gibbs, H. L. and Nelson, D. A. 2009. Twenty-one novel microsatellite DNA loci isolated from the Puget Sound white-crowned sparrow, Zonotrichia leucophrys pugetensis. – Molecular Ecology Resources 9: 795–798.

Rekdal, S. L., Anmarkrud, J. A., Johnsen, A. and Lifjeld, J. T. 2018. Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-Experience from a passerine bird. – Ecol Evol 8: 1680–1692.

Roved, J., Hansson, B., Stervander, M., Hasselquist, D. and Westerdahl, H. MHCtools – an R package for MHC high-throughput sequencing data: Genotyping, haplotype and supertype inference, and downstream genetic analyses in non-model organisms. – Molecular Ecology Resources <https://doi.org/10.1111/1755-0998.13645>.

Sebastian, A., Herdegen, M., Migalska, M. and Radwan, J. 2016. amplisas: a web server for multilocus genotyping using next-generation amplicon sequencing data. – Molecular Ecology Resources 16: 498–510.

Seress, G., Szabó, K., Nagy, D. and Liker, A. 2007. EXTRA-PAIR PATERNITY OF TREE SPARROW (PASSER MONTANUS) IN A SEMI-URBAN POPULATION.

Sheldon, L. D., Chin, E. H., Gill, S. A., Schmaltz, G., Newman, A. E. M. and Soma, K. K. 2008. Effects of blood collection on wild birds: an update. – Journal of Avian Biology 39: 369–378.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. – Front Zool 2: 16.

Tarvin, K. A. 2006. Polymorphic microsatellite loci from the American goldfinch (Carduelis tristis) and their cross-amplification in a variety of passerine species. – Molecular Ecology Notes 6: 470–472.

Toivanen, P., Vaahtovuo, J. and Eerola, E. 2001. Influence of Major Histocompatibility Complex on Bacterial Composition of Fecal Flora. – Infect Immun 69: 2372–2377.

Wakeland, E. K., Boehme, S., She, J. X., Lu, C.-C., McIndoe, R. A., Cheng, I., Ye, Y. and Potts, W. K. 1990. Ancestral polymorphisms of MHC class II genes: Divergent allele advantage. – Immunol Res 9: 115–122.

Westerdahl, H., Stjernman, M., Råberg, L., Lannefors, M. and Nilsson, J.-Å. 2013. MHC-I Affects Infection Intensity but Not Infection Status with a Frequent Avian Malaria Parasite in Blue Tits. – PLoS ONE 8: e72647.

Yang, W., Ding, J., Wang, S., Yang, Y., Song, G. and Zhang, Y. 2020. Variation in genetic diversity of tree sparrow (Passer montanus) population in long-term environmental heavy metal polluted areas. – Environmental Pollution 263: 114396.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. and Glöckner, F. O. 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. – Nucleic Acids Res. 42: D643–D648.

CONCLUSIONS: THE AVIAN MICROBIOME AND DISEASE

Vertebrates are extensively colonized by bacteria, creating microbial assemblages that are collectively called the microbiome (Gilbert et al. 2018). These microbial communities are vast, with a recent estimate placing the total number of bacteria on a typical human at about 39 trillion (Sender et al. 2016). Microbiomes are highly variable, with distinct communities found between host species and even between individuals of the same species, and varied factors such as developmental stage, sex, and diet are known to cause variation in the microbiome (Gilbert et al. 2018, Bodawatta et al. 2022). Host-associated microbiomes carry out many activities that are necessary for host health; for example, some mutualists process essential vitamins for the host or aid in immune function (Rowland et al. 2018, Hills et al. 2019). Alternatively, some microbial community members may function as commensals or even pathogens rather than mutualists, with bacteria out-competing beneficial members of the microbiome or inducing disease (Gilbert et al. 2018, Hills et al. 2019).

Host-associated microbiomes have been studied extensively in humans and some domestic mammals, reaching an exciting point where microbiomes may be harnessed to develop more effective diagnostics or even microbial treatments (Gilbert et al. 2018). However, parallel research in other systems such as wild birds still lags behind, despite the potential for microbiome research to contribute to conservation goals (Bahrndorff et al. 2016, Bodawatta et al. 2022). For this dissertation, I addressed several gaps in avian microbiome research related to avian health and disease risk, including

associations with disease state, community stability and pathogenicity over time, and the influence of immunogenetic variation on the microbiome.

Gut Pathogens and the Microbiome

Enteric pathogens are closely intertwined with gut microbiome dynamics, with some pathogenic invaders out-competing members of the microbiome to colonize the host while other opportunistic pathogens exist as normal members of the microbiome until they take advantage of perturbations in the community (Hills et al. 2019). Factors such as microbial diversity (i.e. the number of species in a community) can be important – for example, one study using juvenile captive ostriches found that juveniles with initially low bacterial diversity were more likely to develop pathogen-associated dysbiosis and later succumb to enterocolitis mortality compared to juveniles with high initial diversity (Videvall et al. 2020). Despite the apparent conservation benefits of understanding pathogen-associated dysbiosis and identifying risk factors in vulnerable species, little is known about this relationship in wild birds, and bacterial pathogens tend to be identified using targeted screening approaches rather than examining community-wide patterns (Smith et al. 2020). Furthermore, many avian microbiome studies rely solely on 16S rRNA gene sequencing, which provides little indication of whether detected bacteria are pathogenic or not (Smith et al. 2020).

We examined microbial community-level associations with enteric pathogens in two avian species with a combination of 16S rRNA gene sequencing and metagenomic whole genome sequencing (WGS), using genetic virulence factors detected in the WGS

datasets as indicators that certain bacteria were likely pathogenic (Chapters 1 & 2). In the low-diversity microbiome of the Galapagos penguin, only *Clostridium perfringens* was identified as a likely pathogen (Chapter 1). The presence of *C. perfringens* virulence factors was associated with striking microbial community changes: community composition, protein family composition, and putative metabolic activity all significantly varied between infected and uninfected birds, and metabolic activity was significantly decreased in birds with *C. perfringens* virulence factors. In the higher-diversity microbiome of the Eurasian tree sparrow, nine bacteria were identified as likely pathogens, and birds with virulence-associated bacteria had significantly different community composition in their microbiomes compared to birds without detected virulence factors (Chapter 2). However, no significant predictors such as age or sex were discernible in either system. Significant temporal variation was detected in the sparrow microbiome across seasons in terms of community composition, but no seasonal associations with pathogenic bacteria were found from spring to fall (Chapter 2). More work is needed to determine several key points, such as the prevalence of enteric pathogens across wild bird species, the directionality of common enteric pathogen infections and microbial perturbations, and the factors rendering certain birds susceptible to enteric infections (Smith et al. 2020).

Blood Parasites and the Microbiome

While enteric pathogens have an obvious relationship with the normal gut microbiome, other disease associations may be less direct. Recent studies have indicated a fascinating link between microbial community composition and *Plasmodium* infection

risk, demonstrating that captive mice exhibit varying levels of susceptibility to blood parasite infection based solely on microbiome composition and that transferring microbiomes to naïve mice also transfers the degree of susceptibility (Villarino et al. 2016). One mechanism explaining this relationship is that certain microbes may prime the immune system, allowing it to react more rapidly and efficaciously to *Plasmodium* invasion (Yilmaz et al. 2014). This relationship is increasingly established in mammals, but whether it exists similarly in birds remains unclear.

We tested associations between the gut microbiome and *Plasmodium* infection status in the Eurasian tree sparrow (Chapter 3; Rohrer et al. *in press*). We did not find large-scale associations between community structure and disease; neither alpha nor beta diversity significantly varied in the presence of *Plasmodium* infection. However, we found several microbes which were differentially abundant in birds infected with *Plasmodium.* Many of the bacteria with significantly lower abundances in infected birds belonged to genera with many pathogenic members such as *Pseudomonas, Shimwellia, Enterococcus,* and *Serratia*, while a putative mutualist – *Ligilactobacillus* – was present in significantly higher abundances in infected birds. Unfortunately, it is not possible to identify causality from this study, as we could not determine key factors such as whether uninfected birds were naïve or had been previously exposed to *Plasmodium*, whether infected birds represented acute or chronic infections, and whether the detected microbial variations were present prior to infection or were caused by *Plasmodium* invasion. Longitudinal studies in wild birds with high recapture rates would be an ideal way to assess causality (for example, using nest boxes), as the microbial alterations typically caused by captivity

limit inferences from captive studies (Alberdi et al. 2021, San Juan et al. 2021). Probiotic treatments or inoculation with the set of bacteria highlighted in our study would clarify the directionality of this *Plasmodium*-microbiome relationship in wild birds.

Immunogenetics and the Microbiome

Adding to the complexity of the microbiome relationship with disease risk is the potential influence of the host immune system. Genome wide association studies have identified some immune-related markers associated with altered microbiome composition, but one of the most promising targets is the major histocompatibility complex (MHC) (Bolnick et al. 2014, Goodrich et al. 2014, 2016). The highly polymorphic MHC is a key part of microbial recognition by the adaptive immune system, and some studies have found associations between variation at MHC loci and microbial community structure (Sommer 2005, Bolnick et al. 2014). Studies in captive mice, wild three-spined sticklebacks, and Seychelles warblers have found specific alleles associated with altered community composition in the gut microbiome (Bolnick et al. 2014, Kubinak et al. 2015, Davies et al. 2022). This variation may in turn affect disease risk; for example, one study found that captive-reared mice with different MHC genotypes had contrasting levels of susceptibility to *Salmonella enterica Typhimurium,* and when the microbiomes from those mice were transferred to germfree mice with identical genotypes, the mice with the transplanted microbiomes exhibited similar susceptibility or resistance to *S. e. Typhimurium* infection compared to the original mice (Kubinak et al. 2015).

It remains unclear whether MHC variation shapes the microbiome only through host-specific associations with certain alleles, or whether there might be more generalizable patterns related to the degree of MHC diversity present in an individual. The heterozygote advantage and divergent allele advantage hypotheses have been proposed as a way to explain the astonishing diversity present at MHC loci of many birds; essentially, these hypotheses posit that greater allelic diversity allows a wider range of responses to pathogenic invaders at an individual level (Sommer 2005, O'Connor et al. 2016). In a microbiome context, higher MHC variation could be maintained as a way to modulate the microbial community (Bolnick et al. 2014, Khan et al. 2019).

We examined the MHC-microbiome relationship in the Eurasian tree sparrow, testing microbial associations with specific MHC IIB alleles as well with the degree of allelic diversity found within each individual (Chapter 4). No large-scale microbiome associations were evident, and alpha and beta diversity were not associated with specific alleles or with MHC diversity. However, higher MHC diversity was associated with higher abundances of certain putatively beneficial bacteria such as *Ligilactobacillus*. These results indicate that higher MHC diversity may lead to higher abundances of beneficial microbes in wild birds, providing some support for the divergent allele advantage hypothesis. However, much remains unclear about how widespread these associations may be and whether these differences are substantial enough to influence disease risk in wild populations. Future studies in other systems should examine how

these significant MHC-microbiome associations influence host fitness and disease outcomes.

Conclusions:

The host-associated microbiome can influence disease risk in many ways, but much work remains to be done in wild populations to achieve generalizable patterns across taxa. Understanding microbiome dynamics is key to determining susceptibility factors and designing microbiome-based treatments (Hills et al. 2019). Some early examples already exist of potential microbial treatments for disease using probiotics or fecal transplants, indicating exciting potential outcomes for this relatively new field – for example, probiotic additives are already in widespread use in poultry feed to prevent bacterial infections (Redweik et al. 2020). Future avian microbiome studies in wild populations should focus on identifying causal relationships between the microbiome, host immune system, and disease.

References:

Alberdi, A., Martin Bideguren, G. and Aizpurua, O. 2021. Diversity and compositional changes in the gut microbiota of wild and captive vertebrates: a meta-analysis. – Sci Rep 11: 22660.

Bahrndorff, S., Alemu, T., Alemneh, T. and Lund Nielsen, J. 2016. The microbiome of animals: implications for conservation biology. – Int J Genomics <https://doi.org/10.1155/2016/5304028>.

Bodawatta, K. H., Hird, S. M., Grond, K., Poulsen, M. and Jønsson, K. A. 2022. Avian gut microbiomes taking flight. – Trends in Microbiology 30: 268–280.

Bolnick, D. I., Snowberg, L. K., Caporaso, J. G., Lauber, C., Knight, R. and Stutz, W. E. 2014. Major Histocompatibility Complex class IIb polymorphism influences gut microbiota composition and diversity. – Mol Ecol 23: 4831–4845.

Davies, C. S., Worsley, S. F., Maher, K. H., Komdeur, J., Burke, T., Dugdale, H. L. and Richardson, D. S. 2022. Immunogenetic variation shapes the gut microbiome in a natural vertebrate population. – Microbiome 10: 41.

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V. and Knight, R. 2018. Current understanding of the human microbiome. – Nat Med 24: 392–400.

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. T., Spector, T. D., Clark, A. G. and Ley, R. E. 2014. Human Genetics Shape the Gut Microbiome. – Cell 159: 789–799.

Goodrich, J. K., Davenport, E. R., Waters, J. L., Clark, A. G. and Ley, R. E. 2016. Cross-species comparisons of host genetic associations with the microbiome. – Science 352: 532–535.

Hills, R. D., Pontefract, B. A., Mishcon, H. R., Black, C. A., Sutton, S. C. and Theberge, C. R. 2019. Gut microbiome: profound implications for diet and disease. – Nutrients <https://doi.org/10.3390/nu11071613>.

Khan, M. A. W., Stephens, W. Z., Mohammed, A. D., Round, J. L. and Kubinak, J. L. 2019. Does MHC heterozygosity influence microbiota form and function? – PLoS ONE 14: e0215946.

Kubinak, J. L., Stephens, W. Z., Soto, R., Petersen, C., Chiaro, T., Gogokhia, L., Bell, R., Ajami, N. J., Petrosino, J. F., Morrison, L., Potts, W. K., Jensen, P. E., O'Connell, R. M. and Round, J. L. 2015. MHC variation sculpts individualized microbial

communities that control susceptibility to enteric infection. – Nat Commun 6: 8642.

O'Connor, E. A., Strandh, M., Hasselquist, D., Nilsson, J.-Å. and Westerdahl, H. 2016. The evolution of highly variable immunity genes across a passerine bird radiation. – Molecular Ecology 25: 977–989.

Redweik, G. A. J., Jochum, J. and Mellata, M. 2020. Live Bacterial Prophylactics in Modern Poultry. – Frontiers in Veterinary Science.

Rohrer, S.D., Robertson, B.Q., Chubiz, L.M., and Parker, P.G. Gut microbiome composition associated with *Plasmodium* infection in the Eurasian tree sparrow. In press. J of Avian Biol DOI: 10.1111/jav.03027.

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I. and Tuohy, K. 2018. Gut microbiota functions: metabolism of nutrients and other food components. – Eur J Nutr 57: 1–24.

San Juan, P. A., Castro, I. and Dhami, M. K. 2021. Captivity reduces diversity and shifts composition of the Brown Kiwi microbiome. – Animal Microbiome 3: 48.

Sender, R., Fuchs, S. and Milo, R. 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. – PLoS Biol 14: e1002533.

Smith, O. M., Snyder, W. E. and Owen, J. P. 2020. Are we overestimating risk of enteric pathogen spillover from wild birds to humans? – Biological Reviews of the Cambridge Philosophical Society 95: 652.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. – Front Zool 2: 16.

Videvall, E., Song, S. J., Bensch, H. M., Strandh, M., Engelbrecht, A., Serfontein, N., Hellgren, O., Olivier, A., Cloete, S., Knight, R. and Cornwallis, C. K. 2020. Early-life gut dysbiosis linked to juvenile mortality in ostriches. – Microbiome 8: 147.

Villarino, N. F., LeCleir, G. R., Denny, J. E., Dearth, S. P., Harding, C. L., Sloan, S. S., Gribble, J. L., Campagna, S. R., Wilhelm, S. W. and Schmidt, N. W. 2016. Composition of the gut microbiota modulates the severity of malaria. – Proc Natl Acad Sci USA 113: 2235–2240.

Yilmaz, B., Portugal, S., Tran, T. M., Gozzelino, R., Ramos, S., Gomes, J., Regalado, A., Cowan, P. J., d'Apice, A. J. F., Chong, A. S., Doumbo, O. K., Traore, B., Crompton, P. D., Silveira, H. and Soares, M. P. 2014. Gut microbiota elicits a protective immune response against malaria transmission. – Cell 159: 1277–1289.