Spring 2022

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washingto State bats

Dana E. Colley

INVESTIGATING HOW BAT ECTOPARASITES INFLUENCE THE SKIN MICROBIOME DIVERSITY AND COMPOSITION IN WASHINGTON STATE BATS

A Thesis

Presented To

Eastern Washington University

Cheney, Washington

In Partial Fulfillment of the Requirements

for the Degree

Master of Science in Biology

By

Dana E. Colley

Spring 2022

THESIS OF DANA E. COLLEY APPROVED BY

DATE 6/13/22

DR. KRISZTIAN MAGORI, CHAIR, GRADUATE STUDY COMMITTEE

DATE 6/14/2022

DR. JENIFER WALKE, GRADUATE STUDY COMMITTEE

DATE 6/14/2022

DR. ASHLEY LAMM, GRADUATE STUDY COMMITTEE

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Eastern Washington University, I agree that the JFK Library shall make copies freely available for inspection. I further agree that copying of this project in whole or in part is allowable only for scholarly purposes. It is understood, however that any copying or publication of this thesis for commercial purposes, or for financial gain, shall not be allowed without my written permission.

Signature

6-16-22

Date

INVESTIGATING HOW BAT ECTOPARASITES INFLUENCE THE SKIN MICROBIOME DIVERSITY AND COMPOSITION IN WASHINGTON STATE BATS

by

Dana E. Colley

Spring 2022

White-Nose Syndrome (WNS), caused by the psychrophilic pathogenic fungus *Pseudogymnoascus destructans* (Pd), has killed millions of bats in the eastern United States since its initial introduction in 2006 and recent expansion into the western U.S. Understanding factors that contribute to the spread of Pd and risk of infection is crucial for management of WNS. Bat ectoparasites, including bat mites and bat flies, are omnipresent in bat populations, yet the relationship between these ectoparasites and bat health is still unknown. We examined the relationship between bat ectoparasites and the skin microbiome in relation to WNS infection risk in Washington State bats. We hypothesized that bats with ectoparasites would have a decreased skin microbiome diversity thus increasing their susceptibility to Pd infection. We collaborated with Washington Department of Fish and Wildlife during Spring 2021, sampling 147 bats representing five different species across 10 bat roosts in Washington State. We found that certain bat species were more likely to be infested with ectoparasites than others, especially *Eptesicus fuscus* ($p = 0.0429$) and *Myotis volans* ($p = 0.0094$). Using next-gen sequencing techniques, we found that ectoparasite infestation did not decrease the skin microbiome diversity of Washington bats ($p = 0.965$), although bat species ($p = 0.006$)

and roost location ($p = 0.001$) significantly influenced the skin microbiome diversity. Using culturing methods, we identified 20 species of culturable bacteria from bat skin with four isolates belonging to genera known to possess antifungal properties. These isolates could be used to develop probiotic therapies for local colonies to prevent and treat WNS in the future.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Krisztian Magori for his advice, support, and unfailing good humor throughout the course of this research project. I must also thank my other advisor Dr. Jeni Walke for her kindness, generosity, and expertise for the molecular and culturing lab work of this research, in addition to patiently helping me navigate my data analyses. I must also thank my third committee member Dr. Ashley Lamm for her support and time. I would also like to thank Abby Tobin from the Washington Department of Fish and Wildlife for allowing me to collect my samples during her White-Nose Syndrome surveillance in Spring 2021, and for sharing her wealth of bat knowledge with me. Thank you to all the other Washington Department of Fish and Wildlife staff that allowed me to shadow their White-Nose Syndrome surveillance sessions and learn from them. I would also like to thank the Department of Natural Resources and the Bureau of Land Management for allowing me to sample at our sites in Thurston and Klickitat Counties. Many thanks to the Spruell Lab for the generous use of their lab space and refrigerator for incubating my culture plates. I must also thank my friends and lab-mates Tiffany Jordan and Sarah DeShazer for their support, sanity, and assistance with proofreading countless drafts of prospectuses, grants, and this thesis, in addition to sitting through countless practice presentations. Lastly, I would like to thank Eastern Washington University's Biology Department and the Northwest Scientific Association for funding this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Importance of the Microbiome:

Nearly all animal orders have co-evolved with and continue to engage in symbiotic relationships with microbes [1-3]. Microbes form diverse and functionally active communities in and on their animal hosts, forming distinct communities on the skin surfaces [1,3,4], the gastrointestinal tract, the respiratory tracts, and the urogenital tract [1-3,5,6]. This collection of microbial communities is collectively referred to as the microbiome, which includes bacteria, archaea, fungi, protists, bacteriophages, and viruses, in addition to their respective genes and metabolites [1-3]. The microbiome performs a myriad of biologically significant functions that are critical for survival and overall host health [3-8], including digestion and diet supplementation [2,3,5,7,9-11], immune function and protection from pathogens [1-3,12,13] and developmental cues [2,6,8,10,11].

The skin microbiome is an important subset of the host's microbiome and is the first line of defense against invading pathogens, as it is the only microbiome in constant contact with the external environment [1,3,14]. The primary role of the skin microbiome is to modulate immune function by preventing the colonization of pathogenic microbes by outcompeting invading pathogens for physical space and resources [3,9,15]. Additionally, many microbial taxa on the skin produce metabolites that inhibit the growth of pathogenic microbes [3,9,15], further protecting the host from cutaneous infections. The protective role of the skin microbiome makes it an important area of focus in overall health of the host, especially skin health [3].

The composition of the skin microbiome directly influences its ability to function as an immune modulator [4,16], with more diverse skin microbiomes often having an increased ability to defend the skin surface from pathogenic invaders [4,16-18]. Maintaining diverse communities of skin microbes is a more complete use of niche space on the skin surface, increasing the probability of containing functional microbes that can actively inhibit pathogens through their metabolic activity [4]. Additionally, diverse skin microbial communities have been found to alter their metabolites in the presence of other microbes, working more effectively together to prevent pathogen invasions through enhanced antimicrobial metabolites than they would as monocultural isolates [4,19-21].

Maintaining the diversity of the skin microbiome is therefore crucial in preventing cutaneous infections that compromise the health of the host [4,15-17]. Dysbiosis, or microbial imbalance, is the result of decreased microbial diversity where proportions of microbial taxa become altered, allowing pathogens to more easily colonize and dominate surfaces of the host [15]. Several cutaneous diseases have been documented to be caused or exasperated by dysbiosis of the skin microbiome, including atopic dermatitis in domestic dogs [22] and humans [18,23], bovine digital dermatitis [13], sarcoptic mange [24], *Batrachochytrium dendrobatidis* infection (chytrid fungus) in amphibians [4,14,25] and White-Nose Syndrome in bats [26-30]. It is important to evaluate the diversity of the skin microbiome as a key indicator when determining risk for cutaneous diseases, especially for vulnerable populations of wildlife facing mass population declines from cutaneous infections [4,14,25-30].

The importance of skin microbial diversity in preventing cutaneous infections [4,16-18] has recently been applied to the control and management of cutaneous wildlife

diseases, especially with the amphibian chytrid fungus [4,14,17] and White-Nose-Syndrome in bats [14,26,27,31-33]. Augmenting the diversity of the skin microbiomes of at-risk animals with antimicrobial microbes are a successful therapy for inhibiting the growth of the chytrid fungus and the White-Nose Syndrome pathogen in controlled laboratory settings [14,17,26,27,31,32] and in small-scale field trials [33]. Although additional probiotic-therapy studies need to be conducted on scales closer to actual wildlife management scenarios, the positive effects of probiotics as a remedy for cutaneous infections in wildlife are promising [14,26,27,31,32]. While there are many host-associated factors that influence the skin microbiome composition and diversity [3,15,16], other extraneous factors, such as ectoparasites, can also alter the important microbial composition and diversity of the skin microbiome [24], which can have associated health consequences for the host.

The Importance of Ectoparasites in Wildlife:

Ectoparasites are a widespread and diverse group of parasites that utilize the outer surfaces of their hosts, subsisting on specialized diets of blood meals, skin scrapings, or host secretions [34]. These surface-specialists often form obligate relationships with microbes to supplement their poor diets and parasitic lifestyles [5,7,35]. Nearly all orders of animals are parasitized by ectoparasites, including mammals [24,34,36-38], birds [10,39-41], lizards [42-45], amphibians [45], fish [46], and insects [47]. Ectoparasites include flukes, lampreys, leeches, and vampire bats, but the largest and most successful group of ectoparasites are the arthropods (crustaceans, insects, and arachnids) [34].

Common arthropod ectoparasites include lice, mites, and ticks [7,8,10,34,37,48,49], and most ectoparasites never or only briefly leave their hosts during their life cycle [50].

Ectoparasites negatively influence host fitness both directly and indirectly [34,39,43]. Ectoparasites directly decrease the fitness of their host through the depletion of host blood and other fluids, and by the creation of irritating sores from biting or chewing activity that disrupt the skin's barrier, facilitating secondary cutaneous infections [4,15,34]. Indirectly, ectoparasites decrease host fitness by initiating energetically costly immune defenses in response to heavy ectoparasite infestations [10,39]. Additionally, heavy ectoparasite loads act as a nuisance to the host, prompting the reallocation of energy budgets towards excessive grooming instead of foraging or searching for mating opportunities [38,50,51]. Heavy ectoparasite loads can also decrease mating success by reducing the attractiveness of an individual to potential mates, commonly observed in birds [10,39-41].

Most importantly, ectoparasites pose a threat to their hosts by acting as vectors of disease [7,8,10,40,52-54]. Ticks and flies especially are known vectors of pathogens including bacteria (e.g., *Rickettsia* spp., responsible for Rocky Mountain Spotted Fever), endoparasites (e.g., *Plasmodium* spp., responsible for malaria, trypanosomes), and fungal pathogens (e.g., *Pseudogymnoascus destructans,* responsible for White-Nose Syndrome) [7,8,36,48,52-54]. Ectoparasites as vectors of disease can have serious consequences for both human and wildlife populations, resulting in immediate mortalities or reduced fitness associated with the long-term costs of disease recovery [52,55,56].

Recently, ectoparasite infestation has been reported to decrease skin microbiome diversity of their hosts. For example, *Sarcoptes scabiei* mite infestation has been

negatively associated with the skin microbiome diversity in three species of canids, increasing the prevalence of opportunistic pathogens and the severity of sarcoptic mange infection [24]. However, this phenomenon is understudied in the disease ecology of other animal systems, such as bats. Bat ectoparasites, such as Hippoboscidae bat flies and *Spinturnix* bat mites, are obligate and omnipresent parasites of bat colonies [57]. Many of these parasites are species-specific to their hosts [58] and can act as vectors of the fungal pathogen responsible for White-Nose Syndrome [54]. However, it is unknown whether these common parasites influence the bat skin microbiome and subsequent severity of White-Nose Syndrome in their host.

White-Nose Syndrome Disease Ecology:

White-Nose Syndrome (WNS), caused by the psychrophilic, pathogenic fungus *Pseudogymnoascus destructans* (Pd), continues to decimate bat populations in North America since its initial introduction in the eastern United States and recent expansion into the western United States [59,60] (Fig. 1). Pd is a slow-growing Ascomycete fungus that reproduces primarily through distinctive crescent-shaped conidia (Fig. 2), fungal spores that are produced asexually and are an exact clone of the parent fungus [61,62]. Pd primarily acts as a pathogen but is also capable of saprotrophic utilization of organic and nitrogenous substrates found in hibernacula (i.e., mammal hair, chitin, guano, etc.) allowing it to persist in hibernacula when bats are not present [63,64].

Phylogenetic studies indicate that Pd was introduced into North America from Europe [65,66]. While Pd results in mass mortalities in North American bats as a novel pathogen, colonies in Europe do not experience mass die-offs or severe infection

symptoms compared to North American bats [67], indicating that European bat species have evolved tolerance mechanisms to attenuate the harmful effects of Pd [67-69]. The ongoing invasions of Pd in North American bat colonies is almost exclusively clonal, as populations of North American Pd exhibit very little genetic variation [70], that would be indicative of the genetic recombination generated during sexual reproduction. While it is known that Pd has the potential to reproduce sexually with a heterothallic mating system [71], Pd reproduces primarily via asexual conidia [61,62,71] and only one of the two mating types required for sexual reproduction has been detected in North American isolates [65]. However, the potential for the sexual reproduction of Pd in the future could introduce adaptation and diversity in North American strains, introducing the potential to adapt and become more virulent in North America [71].

Pd grows on the muzzle and delicate wing membranes of bats, disrupting gas and fluid exchanges across the tissues and creating scarring cupping erosions [59]. In addition to damaging the protective epidermal barrier, Pd increases evaporative water loss across the damaged tissues, causing bats to arouse from torpor periodically. Bats search for water and unavailable insect prey after arousal, depleting valuable fat reserves in the process and eventually leading to starvation and death [59]. Although Pd thrives in colder temperatures found in winter hibernacula [72], Pd is a hardy fungus that can survive and remain viable far above its ideal thermal range of 12.5-15.8°C, and has been reported to survive and remain viable in temperatures up to 30°C [72-74], growing and persisting on multiple types of substrates [63,75]. This allows Pd to have a year-round potential for transmission and mortalities [72-74]. Not all bats exhibit visible WNS symptoms

following Pd infection [76], but still retain the ability to spread viable Pd spores and conidia to other bats.

While there have been documented recoveries of *Myotis lucifugus* colonies from WNS [55,77,78], bats that recover from WNS still suffer long-term fitness consequences [55,56,78]. The scarring wing damage resulting from Pd infection compromises flight ability and reduces foraging efficiency [55,56], reducing the ability to store up valuable fat reserves for winter hibernation [59]. Additionally, recovering from Pd infection has associated energy costs of recovery, reallocating energy budgets from other activities and body processes to fight infection [55]. Pd infection has also been studied to change the community composition and ecological roles of recovering bat colonies [78], having long-term repercussions for the population dynamics of recovering colonies.

WNS in North America:

WNS was first detected in Albany, New York in the winter of 2006, rapidly spreading throughout bat colonies in the eastern United States, resulting in the mortalities of millions of bats since its initial introduction [59]. In 2016, WNS was detected in the western United States [60]. Since its initial detection in western Washington, WNS has continued to spread throughout Washington state, increasing the prevalence of bat and colony-wide mortalities [79] (WDFW 2020, unpublished data). Few WNS studies in North America [60,80] have been conducted outside of the eastern United States, so the extent of differences between the eastern and western WNS invasion dynamics are largely unknown. Documented differences between eastern and western WNS invasions include differences in affected bat host species between the eastern and western United

States [81], and differences in roosting behavior and utilized hibernacula between the different bat species [80,81], which could influence how Pd spreads and invades its hosts, influencing the success of current management strategies designed for the eastern invasion.

It is well-established that maintaining a healthy and diverse skin microbiome is crucial for supporting overall health and preventing cutaneous infections [4,16-21]. Recent efforts have been made to describe the factors influencing the health and composition of the bat skin microbiome [26-30] and how the skin microbiome differs between species, locations, and different areas of the body [26-30] in response to WNS. Naturally occurring microbes in bat skin microbiomes in the eastern United States have been documented to possess antifungal properties [31-33,82], especially *Pseudomonas* spp. and *Rhodococcus* spp. These bacteria have been tested as successful probiotic therapies for WNS management in the eastern United States [26,27,31-33,82], reducing Pd abundance in the bat skin microbiome and reducing bat mortalities. Pd infection changes the composition of the skin microbiome [31-33,83], reducing the diversity and abundance of these antifungal bacteria, although the magnitude of reduced diversity and tolerance toward Pd infection differs among bat host species [27,29]. Augmenting the skin microbiome with these naturally occurring antifungal bacteria may serve as a useful management tool for WNS in the future [26,27,31-33,82].

Bats provide humans with multiple ecological services, including millions of dollars in pest control, seed dispersal, and guano for fertilizer [84,85]. WNS is the largest and fastest spreading threat to ecologically valuable bat populations in North America [59,78,84]. It is critical to study factors influencing WNS disease ecology, such as

ectoparasites, that are nearly universal in bat colonies but are likely to differ between eastern and western WNS invasions as bat species and life histories differ. As WNS becomes more pervasive throughout the western United States [59,60], it is important to understand local factors influencing WNS disease ecology to make effective management decisions such as monitoring and surveillance, vaccines, and probiotic therapies, as strategies may need to be modified to the region [31-33,86-88]. This work will determine whether common bat ectoparasites influence the risk of contracting and spreading WNS in the western United States by altering the skin microbiome, which can be applied to other colonies across North America. This study will also provide knowledge of culturable bat skin microbes of western bats that could be used as probiotic therapies in the future, as these are largely understudied in the western United States [26,27,31-33,82].

Objectives and Hypotheses:

The objectives of this study were 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats and to determine their relative abundances in the sequenced skin microbiome. We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured [89-91], we also hypothesized that the culturable bacterial isolates from western bats would not be members of abundant taxa when compared to the sequencing data.

METHODS

Study Sites:

Vespertilionidae microbats were sampled from ten maternity colonies across nine counties in Washington state (Fig. 3) with the Washington Department of Fish and Wildlife (WDFW) as part of their annual WNS surveillance from April – June 2021. Bat roosts were selected by the WDFW according to their need for WNS surveillance activities across Washington state. The names and coordinates of sampling sites have been withheld due to the sensitive nature of bat roost locations, according to Appendix A of Policy-5210 from the WDFW.

Bat and Ectoparasite Sampling:

Mist nets and harp traps were used to capture 20 bats from each study site after sunset, following WDFW sample sizes [81]. Bats were gently untangled from mist nets or collected from harp traps, and each individual bat was placed in their own sterilized cloth sack. Bat sacks were held in small plastic tubs with hand warmers until bats were processed [81]. Bats were weighed on a digital-read scale, and each bat was handled after sterilizing nitrile gloves with 95% ethanol between bats to avoid cross-contamination between individuals. Equipment and clothing decontamination between study sites followed WDFW methods, and all personnel were required to wear Tyvek™ suits during sampling to prevent the transmission of Pd conidia or spores from site to site [81]. Bat wing membranes were examined for fungal growth and damage (i.e., lesions, scarring, tears, holes) with a UV light and a lightbox (Figs. 4,5), using a standardized wing damage index on a scale of zero (no damage) to three (severe damage), or “P” for physical

damage (i.e., tears or holes not relating to Pd infection) [81]. Ear length and forearm length were measured with a millimeter ruler. Species, sex, calcar type, tragus type, and age were assessed according to WDFW protocols [81] and recorded for each individual. Because *M. lucifugus* and *M. yumanensis* are essentially morphologically identical in the field, each potential *M. lucifugus* or *yumanensis* species had their alarm call recorded during post-processing release to identify species by the frequency of the call afterwards [81]. Guano was collected opportunistically for Pd screening by the WDFW but was not incorporated into this study.

Two DNA-free water-moistened rayon swabs were taken from each bat. One swab was rolled back and forth on the muzzle five times and the forearm five times and was repeated for the other swab using the other forearm, following the standard protocol of other bat studies [76,88] (Fig. 6). The microbial DNA from one swab was extracted and sequenced, while the collected microbes from the other swab were cultured and isolated (Fig. 7). Sequencing swabs were placed in sterile 1.5 mL centrifuge tubes, placed on ice in the field, and then stored at -80°C until processed. Culture swabs were placed in sterile 1.5 mL centrifuge tubes with 200 µL of Trypticase Soy-Yeast Extract (TSYE) with 20% glycerol in each tube to preserve bacteria until culturing [32,76,82]. Air temperature at each site was recorded, and after a 45-minute ambient incubation, culturing swabs were placed on ice in the field and then stored at -80°C until processed [76,82]. An additional moistened swab was exposed to the roost atmosphere of each site during bat processing as a field control to ensure sterile handling practices during culturing of bacteria. Bats were screened for ectoparasites by examining the wing membranes, ears, and body fur, and were collected using sterile forceps [92]. Forceps

were sterilized between bats with 95% ethanol. Ectoparasites from each individual were collected in sterile 1.5 mL centrifuge tubes, placed on ice in the field, and frozen upon return to the lab for identification. Ectoparasites were later stored in 70% ethanol for long-term preservation [92]. The proportions of bats with ectoparasites were compared using general linear mixed modeling in R version 4.1.2 (Appendix A).

To prevent the transmission of SARS-CoV-2 to bats and other research personnel, all personnel wore N-95 grade masks and plastic face shields during bat handling, in addition to other PPE. Our handling of bats and sample collection were covered under the WDFW's existing IACUC permits and approvals. Only WDFW personnel handled bats directly to prevent the risk of rabies transmission to unvaccinated personnel. Sampling activities in Thurston County and Klickitat County were conducted on Department of Natural Resources (DNR) and Bureau of Land Management (BLM) land respectively. Appropriate sampling permits were applied for and approved to sample bats at these sites with the WDFW.

Molecular Methods:

DNA was extracted from sequencing swabs using the Qiagen DNEasy Blood and Tissue kit, following the manufacturer's instructions with lysozyme pre-treatment and final elution in sterile molecular-grade water (Appendix B). Qiagen DNEasy kits have been documented to have high extraction efficiencies and are ideal when DNA amounts are in low abundances [75], as with bat skin bacteria [26,28,29,82,93]. During DNA extractions of the sequencing swabs, kit controls were taken with sterile DNA-free water in place of DNA to check extraction reagents for potential contamination for each DNA

extraction batch, resulting in nine kit controls total. DNA was also extracted from the 10 field controls to ensure sterile field sampling methods. A subset of 96 of the extracted samples were randomly selected to be amplified and sequenced, incorporating samples of differing sexes, species, roosts, and ectoparasite presence to include a spread of potential confounding variables [30] (Table 1).

Bacterial community composition was determined by amplifying the hypervariable V4 and V5 regions of the bacterial 16S rRNA gene [26,27,29,30,82], as it is the most reliable region to represent the total length of the 16S rRNA gene and provides robust phylogenetic resolution [94]. Bacterial DNA was amplified using primers 515f + barcode and 926r, following the methods and thermocycler settings of the Earth Microbiome Protocol [95] and Walke et al. (Appendix C). Bovine serum albumin (BSA) was incorporated into the PCR reagents to increase DNA yield, as bacterial DNA from PCR runs without BSA had DNA levels that were too low for detection during gel electrophoresis [95] (Appendix C). Sterile DNA-free water was run through PCR for each sample as a negative control in all trials, with samples being randomly selected and processed in batches of eight to avoid confounding results by study site or PCR trial. The 10 extracted field controls were pooled into a single sample and amplified with PCR, and the nine extracted kit controls were pooled into a single sample and amplified with PCR. These control samples were sequenced as part of our total 96 samples. Any contamination present in these samples were removed from our bat skin swab samples bioinformatically [95,96] (Appendices D, E). The Illumina sequencing method we used pools all 96 samples into a single sample. Barcoded forward primers were used to identify individual samples after the pooled sample was sequenced.

After amplification, PCR products were run on 1.5% agarose gels to check amplification success and to ensure sterile practices during PCR. The amount of DNA in each PCR product in ng/µL were quantified with a Qubit fluorometer and pooled in equimolar DNA concentrations into a single sample and cleaned with the Qiagen QIAquick purification kit to remove excess primers and non-target DNA [95] (Appendix C). The cleaned, pooled sample was sent to Harvard University's Dana Farber Cancer Institute to be sequenced using Illumina Mi-Seq with a 250 bp single-end strategy [95,97].

Sequence data was entered into the Quantitative Insights Into Microbial Ecology (QIIME2) pipeline version 2021.11 to be processed [98] (Appendices D, E). Raw sequence data was demultiplexed, quality scored, filtered, aligned, and trimmed using deblur in QIIME2 for single-end sequencing strategies [99,100]. Processed sequence data was then assigned taxonomy using the Silva database in QIIME2 [101]. QIIME2 and Silva are the most curated and most frequently updated databases used in microbial studies and are widely used for bioinformatics analyses [98-101]. Processed sequence data was aligned using MAFFT [102] and was used to construct phylogenetic trees using fasttree2 [103] in QIIME2. This allowed us to infer evolutionary relationships between identified amplicon sequence variants (ASVs, bacterial taxa) sequenced from bat skin microbiomes. We rarefied our filtered sequence data at a sequencing depth of 1,850 sequences per sample to standardize sampling effort (Appendices D, E). DNA extraction kit contamination in the reagents (*Pseudoalteromonas* spp., *Vibrio* spp., *Halomonas* spp., *Idiomarina* spp., *Marinobacter* spp., *Marinomonas* spp., *Salinisphaera* spp., *Salinarimonas* spp.) were removed from our sequence data before conducting alpha and

beta diversity analyses [28,30,95,96]. All three control samples (DNA extraction kit control, field control, sequencing run control) were removed from our sequence data after contamination was identified and removed.

Alpha diversity (the number of different bacterial species on a single individual) of bat microbiomes was measured using common alpha diversity metrics including Shannon's diversity, Faith's phylogenetic diversity, evenness, and observed features [28, 29]. Shannon diversity incorporates the number of species present (richness) and the abundance of each species [104], while Faith's phylogenetic diversity measures the relatedness of bacterial taxa present [105]. Evenness measures the abundance of each species present, while observed features measures the richness of bacterial taxa (the number of different bacterial taxa present). These diversity metrics were compared between bats using Kruskal-Wallis tests for categorical independent variables (ectoparasite presence, bat species, roost location), and Spearman's rank correlations for continuous independent variables (body mass, wing damage index). Because total body mass is confounded by bat species (i.e., Yuma myotis on average weigh ~5 grams whereas big brown bats are much larger and weigh 11-23 grams) [106,107], we compared alpha and beta diversity within single bat species only.

Beta diversity (the bacterial diversity between different bats) was determined using common microbial beta diversity metrics including Bray-Curtis dissimilarity, Jaccard similarity, and unweighted and weighted UniFrac distances [26,28,29]. Jaccard similarity measures the presence or absence of bacterial taxa as a comparison of microbial composition [108] while Bray-Curtis dissimilarity is a measure of the abundances of particular bacterial taxa [109]. Unweighted UniFrac distances are a

qualitative measure of bacterial presence/absence in a microbiome, while weighted UniFrac distances are quantitative and more comprehensive, measuring the abundance of bacterial taxa and the relatedness of the taxa [110].

Alpha and beta diversity metrics were calculated from our sequence data using QIIME2 [101] (Appendices D, E). Beta diversity was compared between bats using a permutational ANOVA (PERMANOVA) for our categorical variables [28], and Mantel tests for our continuous variables. We tested for differences in taxonomic abundances of bacteria between ectoparasite and non-ectoparasite bats, different bat species, and different roost locations using the indicspecies package in R [111,112] (Appendix A). Indicspecies is a permutational analysis that examines the statistical significance of the presence/absence or abundance of taxa between different groups [111,112]. In our case, this analysis examined the statistical strength and significance of the presence or absence of bacterial taxa between bats with differing ectoparasite presence statuses, between different bat species, and between different roost sites. Nonparametric methods (i.e., Kruskal-Wallis tests, PERMANOVAs, permutational indicspecies analysis) were used to compare dependent variables (the calculated alpha and beta diversity metrics) between different independent variables, as these data were not normally distributed and had many double zeros, consistent with presence-absence data (the presence or absence of bacterial taxa) [28].

Culturing Methods:

A subset of 10 randomly selected bat skin swab samples stored in TYSE + 20% glycerol were thawed and plated, incorporating samples from female bats of the most

common species observed with ectoparasites (*M. lucifugus*), incorporating individuals with and without ectoparasites from six different counties. We originally planned to select our samples from five counties so we could select two samples from each site to culture, one sample with and one sample without ectoparasites, keeping location and species consistent. However, only four counties had both *M. lucifugus* and ectoparasite presence, so we selected two additional *M. lucifugus* samples from two counties that had *M. lucifugus* but no ectoparasites to keep species consistent among samples. The six counties represented in our culturing subset (Thurston, Lewis, Chelan, Klickitat, Lincoln, Pend Oreille) represent a geographic spread across Washington state, including two counties each from western Washington, central Washington, and eastern Washington. Keeping species and sex consistent and providing a spread of roosts avoided biasing our culturing results [30]. The two samples from Lewis County did not initially have any microbial growth after three weeks of incubation, so an additional two samples from Lewis County were plated for a total of 12 mass-culture plates in our culturing subset.

Culturing swabs were too dry after nearly a year of freezer storage to use the standard method of plating samples by pipetting and spreading 100 µL of the TSYE + 20% glycerol and bat skin bacteria from each sample on Reasoner's 2A (R2A) agar. We instead plated each sample on R2A agar by gently rolling swabs on the agar in an "M" shape, inverting and incubating the plates at 9°C for three weeks according to methods described in Hoyt et al. [32] and McArthur et al. [113]. Mass-colony plates were checked once a week during the initial three-week incubation period to monitor growth. Distinct bacterial colonies were picked and plated for isolation during each periodic check to prevent plate overgrowth [32,113]. Any fungal growth on R2A was excluded from our

analyses. All morphologically distinct bacterial colonies were isolated from the mass-colony plates by examining each colony's color, shape, margin, elevation, and surface texture [114] (Fig. 8). Distinct colonies were streaked for isolation on fresh R2A agar using sterile toothpicks and inoculating loops and were incubated for 2-5 days at 9°C [32,113]. At the end of the 2–5-day growth period, a colony from each plated isolate were then transferred to sterile screwcap 1.5 mL cryotubes with 1 mL of TSYE + 20% glycerol. This preserves bacteria and prevents evaporation of TSYE + glycerol during long-term archival storage at -80°C.

All morphologically distinct bacterial colonies from each mass-culture plate had their DNA extracted using the freeze-thaw method [115] (Appendix F). The bacterial DNA from these samples were amplified at the 16S rRNA gene with PCR using primers 8f and 1492r and were sent to Genewiz to be sequenced with Sanger sequencing to determine colony identities [116,117] (Appendix G). Sanger sequencing data was entered into Geneious software version 2019.1.1 to be trimmed and assembled into consensus sequences with de novo assembling [118,119] (Appendix H). Geneious trims out low-quality bases from sequence data and assembles consensus sequences with mapping algorithms, aligning forward and reverse sequence reads so that only one nucleotide is allowed to be off during alignment [118,119]. Processed sequenced data was then entered into the NCBI BLAST database for identification of bacterial species.

We searched for our identified culturable bacterial isolates in the sequenced skin microbiome to determine their relative abundances on bat skin. We matched our Sanger sequences from our cultured isolates to our reference sequences (our Illumina sequences) by aligning and assembling consensus sequences in Geneious using 'Map to Reference'

with the standard Geneious assembler. We then examined the frequency table of bacterial ASVs (bacterial "species") used to calculate the alpha and beta diversity metrics for our sequenced skin microbiome. This frequency table reports the read counts of each specific ASV in the skin microbiome for each bat sample. For each bat sample, we summed up the counts of all ASVs found in that bat sample and divided this value by the total number of sequences in our sequenced skin microbiome (1,850 sequences after rarefaction) and multiplied this number by 100% to get the relative abundance of each ASV found in that bat sample. This was repeated for all 90 bat samples. We then searched for the ASV corresponding to the identified isolate in the frequency table (i.e., searching for the specific ASV in the frequency table that corresponds to e.g., the cultured isolate *Pseudarthrobacter equi*). Once the corresponding ASV was found in the frequency table, we calculated the average relative abundance of that ASV across all bat samples. This allowed us to determine how abundant that specific isolate was in the total sequenced skin microbiome across all bat samples.

RESULTS

Bat and Ectoparasite Sampling:

We sampled a total of 147 individuals representing five different species within the microbat family Vesperontilidae across Washington state. The majority of bats sampled were females (n = 142), and all individuals sampled were adults. Of the bat species sampled, the most common species was the Yuma myotis (*Myotis yumanensis*), followed by the little brown bat (*M. lucifugus*), big brown bat (*Eptesicus fuscus*), long-legged myotis (*M. volans*), and the Townsend's big-eared bat (*Corynorhinus townsendii*)

(Table 2, Fig. 9). Because *M. yumanensis* and *M. lucifugus* are morphologically identical in the field, some individuals were unable to be identified to species and were placed in their own combined group. *Myotis yumanensis, M. lucifugus,* and the *Myotis* spp. combined groups were widespread throughout Washington state, being found almost every sampled county, whereas *E. fuscus* and *C. townsendii,* were only observed in two counties each (*E. fuscus* in Mason and Klickitat, *C. townsendii* in Lincoln and Pend Oreille). *Myotis volans* was the only species observed in a single county, which was Lewis County in Western Washington (Fig. 10).

All bat species sampled had ectoparasites (Fig. 11) except for *C. townsendii.* Four types of ectoparasites were observed during our sampling season (Fig. 12), and 28.8% of all sampled bats had ectoparasites. The most common ectoparasite observed were *Spinturnix* bat mites, either orange *Spinturnix* eggs around ears and tragi or adult mites attached to the wing membranes (Fig. 13). Hippoboscidae bat flies were less common, only observed in Thurston and Lewis Counties. *Cimex* bat bugs were observed and collected in Klickitat County but were not actively parasitizing the bat we sampled from that roost. One individual in Lewis County was observed to have a generalist flea.

From our general linear mixed model, we determined that certain bat species are at a greater risk of ectoparasite infestation that other species, regardless of ectoparasite type (Fig. 14, $p = 8.53\times 10^{-5}$)**.** We found that *M. volans* was significantly more likely to have ectoparasites compared to *M. yumanensis* ($\chi^2 = 26.101$; $p = 0.0015$), *M. lucifugus* ($\chi^2 = 26.101$; $p = 0.0094$), and the combined *Myotis* spp. group ($\chi^2 = 26.101$; $p = 0.0479$). *Eptesicus fuscus* was also found to be significantly more likely to have ectoparasites

compared to *M. yumanensis* (χ^2= 26.101; p = 0.0125) and *M. lucifugus* (χ^2= 26.101; p = 0.0429).

Characteristics of the Sequenced Bat Skin Microbiome:

Before rarefaction of our sequence data, we had 93 bat samples with an average of 3,905.37 bacterial sequences per individual bat (range: 425 – 11,726 sequences). After rarefaction at a depth of 1,850 sequences, we lost three samples giving us a total of 90 bat samples to use in our alpha and beta diversity analyses. Our post-rarefaction dataset had a total of 692 ASVs (bacterial "species") across all samples, representing gamma diversity, the diversity of bacteria across all bat samples. There was an average of 95 ASVs on each bat. Across nearly all sampled individuals in our post-rarefaction dataset, *Pseudomonas* spp. was the most abundant taxon in the sampled bat microbiomes, comprising an average of 27% of each bat's skin microbiome. The next most abundant taxon in our sampled microbiomes was *Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium* spp., comprising on average 12% of the bat microbiome, followed by the family Rhizobiaceae, comprising on average 8% of the bat microbiome (Fig. 15).

Ectoparasite Presence and the Skin Microbiome:

Across all four alpha diversity metrics (Shannon's diversity, Faith's phylogenetic diversity, evenness, observed features) bat ectoparasite presence did not influence the skin microbiome diversity or composition of bats (Tables 3,4). Similarly, ectoparasite presence did not influence skin microbiome diversity and composition of bats across all four beta diversity metrics (Bray-Curtis dissimilarity, Jaccard similarity, unweighted

UniFrac distance, weighted UniFrac distance) (Table 5, Fig. 16). Across both alpha and beta diversity metrics, ectoparasite presence also did not influence the skin microbiome diversity and composition within individual bat species for *M. lucifugus* ($p = 0.339$, Pseudo-F = 1.07), *M. yumanensis* ($p = 0.567$, Pseudo-F = 0.555), the *Myotis* spp. combined group ($p = 0.694$, Pseudo-F = 0.764), *M. volans* ($p = 0.261$, Pseudo-F = 1.217) and *E. fuscus* ($p = 0.963$, Pseudo-F = 0.63) (Tables 6,7).

From our indicspecies relative abundance analyses, we found that one bacterial genus, *Micrococcus* spp., was significantly more abundant in bats without ectoparasites, while four genera (*Nocardioides* spp., *Pelagibacterium* spp., *Bergeyella* spp. *Alkalibacterium* spp.) were significantly more abundant in bats with ectoparasites (Table 8, Figs. 17-19).

Bat Species and the Skin Microbiome:

We found that bat species significantly influenced the skin microbiome diversity and composition of bats across all alpha diversity metrics (Tables 9,10). Bat species also significantly influenced the skin microbiome diversity of bats across all beta diversity metrics (Table 11). Across all alpha and beta diversity metrics, *M. yumanensis* had the highest skin microbiome diversity, while *E. fuscus* had the lowest skin microbiome diversity. Additionally, *M. yumanensis* and *M. lucifugus* had the most similar skin microbiome compositions and diversities to each other (Table 12, Fig. 20).

We found that four bacterial genera were significantly more abundant in *C. townsendii,* four bacterial genera were significantly more abundant in *E. fuscus,* two bacterial genera were significantly more abundant in *M. lucifugus,* and three bacterial

genera were significantly more abundant in the *M. yumanensis* and *M. lucifugus* combined group compared to the skin microbiota of all other bat species (Table 13, Figs. 21-25).

Roost Location and the Skin Microbiome:

Roost location also significantly influenced the bat skin microbiome diversity and composition of bats across all alpha (Tables 14,15) and beta diversity metrics (Table 16). Across all alpha and beta diversity metrics, Lincoln County bats had the highest skin microbial diversity while Okanogan and Klickitat County bats had the lowest skin microbial diversity (Tables 14-16). Spokane and Thurston Counties had the most similar bat skin microbiome diversities and compositions, while other counties were distinct bat skin microbiome diversities and compositions (Table 17, Fig. 26).

From our indicspecies analysis, we found that one bacterial genus was significantly more abundant in Chelan County, three genera in Klickitat County, seven genera in Lincoln County, three genera in Mason County, three genera in Okanogan County, one genus in Pend Oreille County, and two genera in Thurston County compared to all other sampled counties (Table 18, Figs. 27-35).

Other Variables and the Skin Microbiome:

We found that overall, body mass of bats within a single species did not influence the skin microbiome diversity and composition of bats for all alpha and beta diversity metrics (Tables 19,20), except for the *Myotis* spp. combined group for Shannon diversity

($p = 0.0137$, Shannon's $H = -0.8571$) and *E. fuscus* for Jaccard similarity ($p = 0.001$, Spearman's rho = 0.531).

Wing damage index overall did not influence the skin microbiome diversity and composition of bats across all alpha and beta diversity metrics (Tables 21,22), except for *M. lucifugus* for Jaccard similarity ($p = 0.04$, Spearman's rho = 0.263778) when comparing wing damage between individuals within a single species (Table 23).

Culturing Results:

From our subset of 12 mass culture plates (Table 24, Fig. 36), we identified 20 species of culturable bacteria from the subset of our sampled bats (Table 25, Fig. 37). The samples from Lincoln County had the greatest species richness of culturable bacteria with 10 species between the two Lincoln County samples, while the sample from Chelan County had the lowest species richness of culturable bacteria with one colony on the mass culture plate (Table 24, Fig. 36). All bacteria that grew on the mass culture plates were slow-growing, and 67% of the mass culture plates had at least one fungus growing on the plates, while 17% of the mass culture plates had only fungal growth and no bacterial growth. The mean bacterial richness per mass culture plate was two morphologically distinct colonies. When comparing the culturable bacteria to the sequenced skin microbiome, we found that all culturable isolates were found in the sequenced skin microbiome, although in very low relative abundances, with an average relative abundance of 0.32% (Table 25). From our 20 isolates, four (*Streptomyces laculatispora, Rhodococcus corynebacterioides, R. tukisamuensis, R. oryzae*) were from genera known to possess antifungal and specific anti-Pd properties (Fig. 38).

DISCUSSION

Novel Bat Microbiome Study in the Pacific Northwest Region:

To date, no bat microbiome studies have been conducted in the Pacific Northwest region of the United States [28,120,121], as the majority of North American bat microbiome studies have been conducted in the eastern United States [28,29,32,33,82], eastern Canada [27,31,93] and central Canada [26]. Microbiome studies focusing on western bat species specifically have only spanned as far westwards as Colorado [28], Arizona, and New Mexico [120,121]. Our bat microbiome findings are therefore novel to the western coastal United States (California, Oregon, Washington), and novel to the Pacific Northwest region especially.

Bat and Ectoparasite Sampling:

In this study, we sampled a total of 147 bats representing five different species, plus the combined *Myotis* spp. group. All bats sampled were in the microbat family Vesperontilidae, the only bat family found in Washington state [106,107,122-125]. Bats in this family are primarily insectivorous, feeding by echolocation, and engaging in hibernation [125]. Of the species observed, *M. yumanensis, M. volans,* and *C. townsendii* are native to the western United States [106,123,124] while *M. lucifugus* and *E. fuscus* are widely distributed throughout North America [107,122]. All of our sampled species are consistent with the 15 bat species known to inhabit Washington State [126-128]. Of our five sampled species, two species (*M. yumanensis* and *M. lucifugus*) are documented with confirmed cases of WNS in Washington State [79]. During our sampling, we found one *M. lucifugus* female infected with the Pd fungus in Chelan County (Fig, 4), although

this sample was excluded from sequencing analyses to avoid skewing results as Pd is known to alter the diversity and composition of the bat skin microbiome [26,27].

Our sampled bats were primarily collected from maternity colonies established in man-made structures (i.e., abandoned outbuildings, bat condos, bat boxes, etc.), similar to the findings of other bat surveys in the Pacific Northwest [126]. Consistent with the findings of the WDFW, we found that *M. yumanensis* and *M. lucifugus* were the most common species throughout Washington State while *C. townsendii* was rarer throughout Washington State [127]. In our sampled counties, two (Mason County, Chelan County) have confirmed cases of WNS [79,129].

All of our sampled species had individuals with ectoparasites except for *C. townsendii*. Other studies [106,107,122-125] have documented bat-specific ectoparasites across all of our sampled bat species, including *C. townsendii* [123]. However, in this study, the sample size of *C. townsendii* was very small (n = 6), consistent with its rarity in Washington State, and we expect that with a larger sample size we would have observed incidences of ectoparasites within this species.

The ectoparasites we observed during bat sampling were common ectoparasites of our sampled species [106,107,122-125]. A quantitative assessment of bat ectoparasites by Poissant and Broders [92] found that 22% of all bats they sampled in Nova Scotia had at least one ectoparasite, and *Spinturnix* bat mites were the most common bat ectoparasite observed [92]. These findings are consistent with our study, as *Spinturnix* bat mites were also the most common ectoparasite we observed across Washington State bat roosts. This prevalence is attributed to the life cycle of *Spinturnix* bat mites, which spend their entire life cycles on their bat hosts while other bat-specific ectoparasites (i.e., *Cimex* bat bugs)

only spend a portion of their life cycles on their bat hosts [35,37,54,57,92,130-133]. In contrast to the Poissant and Broders [92] study, we found that a greater percentage of our sampled bats had ectoparasites (28.8%) compared to the percentage of infested individuals in the Poissant and Broder study (22%). It is known that bat ectoparasite distribution is influenced by a range of factors, including environmental variables such as temperature and humidity [51,132], host age [50,131], bat group size [50], reproductive condition of the host [50,131,132], and sex of the host [50,131,132]. Nearly all bats with ectoparasites in our study were females, and all roost sites except the sites in Spokane and Chelan Counties had incidences of ectoparasites. Poissant and Broder's study found that females had greater incidences of ectoparasites than males, consistent with the findings of other bat ectoparasite studies [38,50,131,132]. Perhaps the nature of our sampling, conducted on maternity colonies comprised almost entirely of female bats in each roost, explains the greater percentages of individuals with ectoparasites in our study, whereas the Poissant and Broders study did not focus their sampling on maternity colonies [92].

Characteristics of the Sequenced Bat Skin Microbiome:

We found that across our 90 bat samples, *Pseudomonas* spp. was the most abundant bacterial taxon in the sampled bat microbiomes, followed by *Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium* spp. and the family Rhizobiaceae. Avena et al. [28] sampled eight different bat species from New York, Virginia, and Colorado, focusing on the same region of the bacterial 16S rRNA gene as our study. Consistent with our findings, Avena et al. [28] also found that *Pseudomonas* spp. were the most abundant taxa on bat skin [28], a genus known to possess antifungal and anti-Pd properties

[26,27,31-33]. *Pseudomonas* is commonly found in the environment, especially in soils, which could be acquired by bats during foraging activities or migration [28]. However, in the Avena et al. [28] study, *Pseudomonas* spp. comprised on average 9% of the total bat microbiome, whereas in our study *Pseudomonas* spp. was three times more abundant in the Washington bat microbiome, comprising on average 27% of the bat skin microbiome across all samples.

Lemieux-Labonté et al. [27] also found *Pseudomonas* spp. to be one of the more abundant taxa in bat skin microbiomes, although Lemieux-Labonté et al. [27] only found *Pseudomonas* spp.to be the third most abundant species in their sampled bat microbiome, while in our study, *Pseudomonas* spp. was the most abundant genus across all bat samples, and across all bat species. However, the Lemieux-Labonté study was conducted on bats in Eastern Canada [27] and location is known to influence skin microbiome diversity and composition of bats [15,28,20]. Perhaps this difference in *Pseudomonas* spp. abundance could be attributed to regional differences between the studies [15,28,30], especially since *Pseudomonas* spp. is commonly found in the environment [28] and is likely to differ in abundnace by geographic location [15,28,30].

Ectoparasite Presence and the Skin Microbiome:

Contrary to our hypothesis, we found that bat ectoparasites did not decrease the skin microbiome diversity of their bat hosts. We also compared the diversity of bat skin microbiomes of bats with and without ectoparasites within bat species (i.e., comparing *M. lucifugus* to other *M. lucifugus*) to avoid confounding results by any bat-species effects. However, ectoparasite presence also did not influence skin microbiome diversity when

compared within individual species. Ectoparasites alter skin microbiome diversity by facilitating secondary infections from burrowing activities or by secreting immune inhibitors into burrows and lesions, resulting in dysbiosis [135-137]. Our findings are interesting given that the one other study examining the relationship between the diversity of the skin microbiome and ectoparasites, the study by DeCandia et al. [24], found that *Sarcoptes scabiei* mite infestation significantly decreased the skin microbiome diversity across three different species of canids [24]. However, there are marked differences in our study and the study by DeCandia et al. [24], including the host species of interest (canids vs. bats) and the number and types of body sites swabbed for skin bacteria (the external ear, dorsal flank, axilla, groin, and outer back leg in the DeCandia et al. study, as opposed to the muzzle and forearm in our study). However, the most important difference in our study and the DeCandia et al. study is the ectoparasite of interest. The DeCandia et al. [24] study focused on the *S. scabiei* mite, a successful generalist ectoparasite that parasitizes a wide range of mammals from humans to small mammals, exploiting a diverse array of niches [138,139].

In contrast, the *Spinturnix* bat mites, Hippoboscidae bat flies, and *Cimex* bat bugs examined in this study are highly specialized, bat-obligate ectoparasites that have a long co-evolutionary history with their bat hosts and are unable to survive without their respective hosts [35,37,54,57,92,130-133,130-143]. Perhaps this long co-evolutionary history [143] between these bat-specific and bat-obligate ectoparasites may explain why we did not detect any differences in skin microbiome diversity and composition in bats with and without these ectoparasites. The virulence transmission trade-off hypothesis predicts that over the course of evolutionary time, specialized parasites co-evolve to be

less virulent so as not to kill the host that they depend on for survival and reproduction, which would be maladaptive for the parasite [144,145]. This hypothesis has been tested in multiple types of macroparasites, especially schistosomes, flatworm endoparasites with life cycles that utilize specific mammalian and mollusk hosts [144,145]. These studies found that schistosomes have evolved to be less virulent in their hosts over time, thus increasing the longevity of the parasite [144,145]. Although this hypothesis has not yet been examined for ectoparasites, perhaps our bat-obligate and bat-specific ectoparasites are following a similar co-evolutionary path, evolving attenuated virulence (via not disrupting the protective skin microbiome) to promote ectoparasite longevity on the bat host. In contrast, the virulence attenuation evolutionary path would not be adaptive for a generalist ectoparasite such as *S. scabiei,* which is not limited to one type of host to survive and reproduce [144,145].

Although we found that bat ectoparasites did not influence fitness of our bats by decreasing the skin microbiome diversity, the bat-specific ectoparasites in this study do result in other fitness costs for their bat hosts [57,133]. Bat-specific ectoparasites are common members of bat communities that directly influence skin health by disrupting the skin barrier through biting and chewing activity, facilitating secondary cutaneous infections [34, 57,133]. Bat ectoparasites also reduce fitness in their hosts by depleting the host of blood [34, 57,133,138,139] and initiating energetically costly immune defenses [10,39]. It would be fascinating to study these bat-specific ectoparasites further to determine if the virulence transmission trade-off hypothesis is applicable to these specialized ectoparasites, and if so, what virulence factors are being attenuated

specifically (i.e., the alteration of the skin microbiome, triggering of host immune defenses, etc.).

Although we did not find any significant relationship between the skin microbiome diversity of bats and the presence or absence of bat ectoparasites, we did find significant differences in the compositions of the skin microbiota in bats with and without ectoparasites. Bats without ectoparasites had significantly greater abundances of *Micrococcus* spp., a Gram-positive, non-motile bacterium understood to be strictly aerobic and capable of surviving in extreme environments [146]. While this genus was more abundant in bats without ectoparasites, on average this taxon was not an abundant group in the total bat skin microbiome, comprising 0.011% of the total sequenced bat skin microbiome. However, *Micrococcus* spp. was one of the bacterial taxa that was more abundant in one of our outlier samples from an individual in Pend Oreille County (Fig. 15). This individual did not have ectoparasites and the skin microbiome was composed of 76.3% *Micrococcus* spp., while most bat samples appeared to have comparably low abundances of *Micrococcus* spp. in their skin microbiota. Perhaps this outlier sample played a role in driving the significant abundance of *Micrococcus* spp. found in bats without ectoparasites, so this must be taken into consideration before drawing conclusions about the ecological role of *Micrococcus* spp. in bats without ectoparasites.

In comparison, four genera of bacteria were significantly more abundant in bats with ectoparasites (*Nocardioides* spp., *Pelagibacterium* spp., *Bergeyella* spp. *Alkalibacterium* spp.), although the frequency of each taxon in the bat skin microbiome was small (Fig. 19). *Nocardiodies* spp. is a common bacterium of parasitic arthropods [147] explaining its increased abundance on bats infested with bat ectoparasites.

However, the other three genera are not closely associated with ectoparasites. *Pelagibacterium* spp. has previously been isolated from seawater [148], *Bergeyella* spp. includes many pathogenic members [149], and *Alkalibacterium* spp. is alkaliphilic and has been isolated from lakes, soils, and feces [150]. It is possible that these microbes may be freely present in the environment and are transferred to the bat ectoparasites, but further studies must be conducted to determine the relevance of these taxa to bat ectoparasites and their hosts.

Although we did not find any significant relationship between the skin microbiome diversity of bats and bat-specific ectoparasites, we now understand that the negative effect of ectoparasites on skin microbiome diversity is not consistent across all groups of ectoparasites, and may be contingent on the type of ectoparasite, especially in regard to generalist vs. specialist life histories. Further studies need to be done to fully understand the relationship between these different types of ectoparasites and their influence on the skin microbiomes of their respective hosts.

Bat Species and the Skin Microbiome:

We found that bat species significantly influenced the skin microbiome diversity and composition of bats, and this relationship was consistent across all alpha and beta diversity metrics. Specifically, we found that *C. townsendii*, *E. fuscus*, *M. yumanensis*, and the *Myotis* spp. combined group had significantly different abundances of specific bacterial taxa in their skin microbiome assemblages compared to all other sampled bat species, although these specific bacterial taxa were found in low abundances in the total bat skin microbiome.

It is well-established that skin microbiome composition and diversity differs between different host species [27-29,93,151], from amphibians [151] to mammals [27-29,93]. There have been a myriad of recent studies aiming to describe the North American bat skin microbiome and to determine factors that influence its diversity and composition, as the skin microbiome is a crucial host-associated factor affecting the susceptibility of North American bats to WNS [26-29,93].

Ange-Stark et al. [29] compared the skin microbiome diversities and compositions of North American bat species *M. lucifugus*, *E. fuscus,* and *Perimyotis subflavus,* the tricolored bat, in regard to susceptibility to Pd infection. Consistent with our study, Ange-Stark et al. found that specific bacterial taxa significantly differed in abundance across the different bat species [29]. Specifically, both the Ange-Stark et al. study [29] and our findings found that the order Bacilliales was significantly more abundant in the *E. fuscus* skin microbiome, which is striking since the Ange-Stark et. al. [29] data was collected from *E. fuscus* distributed in 22 eastern states in the U.S., while our *E. fuscus* samples were collected from western Washington State exclusively. This indicates that the order Bacilliales may be a core bacterial taxon across *E. fuscus,* as it is present on *E. fuscus* regardless of geographical location in North America [28,29]. In contrast to Ange-Stark et al. [29], we did not find any significantly different abundances of specific bacterial taxa on *M. lucifugus*, and *P. subflavus* is not in the western United States [127].

Avena et al. [28] and numerous studies by Lemieux-Labonté et al. [26,27,93] also found that the skin microbiome diversity and composition differed between different bat species. Consistent with our study, Avena et al. found bat species to be a significant

driver of bat skin microbiome diversity and composition, although Avena et al. found significantly different abundances of bacterial taxa in *M. lucifugus and M. volans,* whereas we did not find any significantly different abundances of bacterial taxa associated with either species [28]. The other bat species in the Avena et al. [28] study (*M. sodalis,* the Indiana bat, *P. subflavus, M. septentrionalis,* the Northern Myotis) found to have distinct abundances of specific bacterial taxa are not found in the western United States.

Also consistent with our findings are the Lemieux-Labonté studies [26,27,93], which found that bat species significantly influenced the skin microbiome composition and diversity of bats. The Lemieux-Labonté et al. study [27] focusing on the skin microbiome of *E. fuscus* exclusively found that *E. fuscus* had significantly greater relative abundances of *Clostridium* spp. in the skin microbiome [27], consistent with our findings for *E. fuscus*. Another Lemieux-Labonté et al. study [26] comparing the skin microbiomes between *M. lucifugus* and *E. fuscus* found that skin microbiome diversities and composition differed significantly between the different bat species [26], consistent with our study. However, Lemieux-Labonté et al. [26] found distinct taxa of bacteria associated with the species *M. lucifugus,* whereas we did not. Another Lemieux-Labonté et al. study [93] concluded that bat species was a significant driver of skin microbiome diversity and composition in bats, consistent with our study. However, that Lemieux-Labonté et al. [93] study focused on the skin microbiomes of frugivorous neotropical megabats in captivity, which are not found in North America outside of captivity [93].

We found that the skin microbiome diversity and composition of bats was significantly influenced by the bat species, consistent with other bat studies, including

studies focusing on North American bats [26-29] and neotropical bats [93]. It is fascinating that although our study is the first bat microbiome study of its kind for the coastal western United States (California, Oregon, Washington) and the Pacific Northwest region specifically, we still found bat species to be a consistent driver of bat skin microbiome diversity and composition.

Roost Location and the Skin Microbiome:

We found that geographical location of bat colonies significantly influenced the skin microbiome diversity and composition of bats, and this was consistent across all alpha and beta diversity metrics. We also found that the bats in seven out of our nine sampled counties had significantly greater abundances of specific bacterial taxa in the skin microbiomes compared to the other sampled counties (Table 18). However, all of the bacterial taxa associated with each county were found in low abundances in the total bat skin microbiome (Figs. 28-35).

Geographical location is a well-documented driver of skin microbiome diversity and microbial composition [15]. This relationship has also been observed within bat studies [28,30], and many microbes are transferred to the bat skin microbiome from the environment [28,30]. Consistent with our findings, Avena et al. [28] and Lutz et al. [30] found that the location of bat roosts significantly influenced the skin microbiome diversity and composition of sampled bats. Avena et al. [28] sampled 11 bat species across three states in the U.S., New York, Virginia, and Colorado. Avena et al. found that the bat species sampled from the two eastern states were more similar to each other than the bats sampled in Colorado, and that bacterial richness and abundance was significantly

different between individual roost sites when comparing the same species (*M. lucifugus*) across different roost locations [28]. Many dominant members of the bat skin microbiomes in the Avena et al. study were commonly found in the environment, especially soils, indicating that perhaps one explanation for the site-specificity of bat skin microbiomes is partly due to the microbes bats accumulate from the environment during foraging and migration activities [28].

Lutz et al. [30] sampled 31 afrotropical bat species across eastern Africa. Lutz et al. found that geographic location was also a strong influencer of microbial diversity and composition on bat skin, and that roost elevation was positively correlated with increasing bacterial richness [30]. In our study, we did not incorporate specific roost variables other than temperature and weather, so perhaps future bat skin microbiome studies should incorporate elevation data to understand how elevation influences the skin microbiome of North American bat species. Lutz et al. also found that the most abundant taxa on afrotropical bat skin was the class Proteobacteria [30]. We also found Proteobacteria in our sequenced bat skin microbiome, although it was one of the least abundant groups in our bat skin microbiomes.

In addition to geographical location, specific geographical conditions and microclimates are also widely documented to strongly influence skin microbiome diversity and composition, as different climatic conditions facilitate the growth of different groups of microbes [152-155]. Different taxa of bacteria have specific growth requirements, including salinity, air or water temperature, oxygen, pH, and metabolic substrates [156,157]. The presence or absence of these growth requirements therefore influence the types of bacteria that will be present in the environment or on the host

[156,157]. The relationship between climatic conditions and skin microbiome composition and diversity has been widely documented, from wildlife [152,154,155] to humans [153]. In our study, we observed significant differences in the skin microbiome diversity and composition of bats between the different roost sites. Perhaps the specific microclimate conditions of our study sites (i.e., the roosts located in the wet regions of western Washington vs. the roosts located in the more arid regions of eastern Washington) could be partially responsible for these regional differences in the skin microbiota.

Other Variables and the Skin Microbiome:

We found that overall, body mass of bats within a single species did not influence the skin microbiome diversity and composition of bats, with the exception of the *Myotis* spp. combined group for Shannon diversity and *E. fuscus* for Jaccard similarity. Because the *Myotis* spp. combined group contained two different bat species (*M. yumanensis* and *M. lucifugus*) we attribute this significant difference in bacterial richness and abundance to the confounded differences between the species. It is well-documented that the skin microbiome composition differs between different bat species [26-29,93], consistent with our findings in this study. Additionally, in regard to body mass, *M. lucifugus* are slightly larger than *M. yumanensis* [106,122]. Both of these factors could explain why we found significant differences in body mass and skin microbiota. We also found that body mass was negatively correlated with skin microbiome diversity for the *Myotis* spp. combined group. However, this was a singular trend for Shannon diversity in the *Myotis* spp. group

and cannot be used to draw conclusions between body mass and the skin microbiome diversity of bats.

It is known that body mass of insectivorous bats is an effective indicator of body condition, and bats that are heavier than their average species weight have greater amounts of energy stores [158]. These individuals are healthier, as they are more likely to survive and reproduce with larger fat stores [159]. It was interesting that we found body mass to significantly influence the presence or absence of bacterial taxa in the *E. fuscus* skin microbiome. One *E. fuscus* individual from Mason County was suspect for the Pd fungus after examining the wing membranes with a UV light. Perhaps that individual had an altered body condition due to Pd infection, which is known to directly influence body mass [59] and the skin microbiota [26,29]. This, coupled with the small sample size of *E. fuscus* in this study (n = 11), would be expected to bias results when comparing body mass within the *E. fuscus* group. We also found a positive correlation between pairwise Jaccard similarity distances and body mass for *E. fuscus*, meaning that individuals with larger differences in body mass (larger pairwise differences) also had larger differences in the diversity and composition of their skin microbiomes. However, this correlation is not based on direct diversity measures, so it is difficult to draw biological conclusions from this trend.

We found that overall, wing damage index did not influence the skin microbiome diversity and composition of bats, with the exception of *M. lucifugus* in regard to the presence or absence of specific bacterial taxa when comparing wing damage within *M. lucifugus*. We also found a positive correlation between pairwise Jaccard similarity distances and wing damage for *M. lucifugus*, where individuals with larger differences in

wing damage indices also had larger differences in the diversity and composition of their skin microbiome. Again, this correlation is not based on direct diversity measures and is therefore difficult to draw biological conclusions from. Most individuals in this study (90%) had a wing damage index of zero, which correlates to observing no scarring or necrotic tissue on a wing membrane [56,81]. Within *M. lucifugus,* ~13% of individuals within this species had a wing damage index (WDI) of one, with all WDI = 1 individuals from Thurston County. Wing damage correlates to physical damage (i.e., tears and holes acquired during regular foraging activities) or scarring from necrotic tissue following Pd infection [56,81]. It is possible that the WDI = 1 for these individuals was the result of scarring and necrotic tissue following Pd infection, which would be logical since Thurston County is surrounded by three other counties in Western Washington State documented to have WNS cases or Pd present (Mason, Lewis, and Pierce Counties) [79,129], where individuals could be infected through migration or foraging activities. Additionally, *M. lucifugus* are especially vulnerable to Pd infection [26,27,29], and Pd is known to alter the composition of the bat skin microbiome [26,29]. Therefore any *M. lucifugus* that were exposed to Pd would be more likely to have an altered skin microbiome, which could bias results when comparing wing damage within the *M. lucigufus* group.

Indicspecies Analyses from the Sequenced Bat Skin Microbiome:

From our indicspecies analyses, we found significantly greater abundances of certain bacterial taxa in bats with and without ectoparasites, in four different bat species, and in seven different Washington State counties. Across these findings, the frequency of

all significant bacterial taxa in those groups were low in the overall bat skin microbiome. There are many factors that can result in some bacterial taxa being low in abundance, or rare, in a microbiome [160]. For example, certain bacterial taxa may be rare in a microbiome due to bacterial taxa being newly integrated members in a microbial community, belonging to taxa requiring narrow environmental niches, or due to fitness trade-off constraints, such as possessing increased resistance to stressors at the cost of decreased growth rates [160].

However, many low-abundance or rare bacterial taxa in a microbiome have been found to perform important ecological roles in microbial communities, and many of these rare taxa are keystone species, disproportionately contributing to the function and stability of their ecosystem [160-166]. Keystone species are important members in multiple ecological settings, from marine habitats with the purple sea star (*Pisaster ochraceus*), to terrestrial environments with gray wolves (*Canis lupus*) [166,167], and bacterial taxa in mammalian microbial communities [161] and soil microbial communities [162,165]. In microbial ecosystems, the important ecological roles of rare bacterial taxa have been documented in a range of systems, from rare bacterial taxa maintaining the health of the human colon [161] or driving the production of bio-hydrogen in anaerobic environments [162], to rare bacterial taxa increasing the suppression of fungal pathogens in soil and plant communities [165]. Although we did not examine the roles of the taxa found in the indicspecies analyses, future studies could follow the methods of other rare bacterial studies [161,162,165] to determine the roles of these rare bacteria in the bat skin microbiome.

Culturable Bacteria from Bat Skin:

During the growth period of our 12 culturing samples, we observed that most samples took longer than the standard three-week incubation period at 9°C [32,113] to get any bacterial growth on the media, and that most mass-culture plates had low culturable richness. Perhaps this is due to the cold incubation conditions in established bat skin bacteria culturing protocols [32,113]. These conditions are presumably designed to mimic the cold conditions of bat roosts and bat hibernacula in the Eastern United States, where these studies were conducted [32,113]. However, these published bat probiotic studies [32,113] have established culturing protocols with the goal of isolating culturable bacteria from bat species and bat populations located in the Eastern United States, where bats primarily roost and hibernate in caves.

In contrast, western bat species and western bat populations do not roost in caves, roosting instead in man-made structures such as abandoned outbuildings and bat condos, or more exposed natural environments such as trees and rocky crevices [80,81], where the ambient temperatures of these environments are warmer than the ambient temperatures of a subterranean cave environment [80,168-170]. For example, one study examining the abiotic conditions of cave and mine bat roosts in the Eastern United States found that ambient roost temperatures ranged from 0°C to 13.9°C [170]. Few studies have examined the specific roost conditions of the man-made or exposed natural roosts found in the Western United States, although the temperatures of these roosts are expected to fluctuate with the ambient air temperatures [80]. For bacteria cultured from western bat skin, perhaps the previously established culturing protocols in bat probiotic studies are inadequate and may be too cold for most culturable isolates from these bats to

be grown on media. This alteration of culturing protocols for western bat species was utilized by Hamm et al. [120], who cultured bacteria from the skin of southwestern bat species and populations [120]. Although Hamm et al. did not develop anti-Pd probiotic therapies, they did culture bacteria from the skin of southwestern bats at warmer temperatures than the standard protocol established by probiotic studies conducted in the Eastern United States [32,113], incubating their samples at 20°C [120]. Perhaps bat skin bacteria culturing protocols need to be adjusted and standardized for the culturing of bacteria from western bat species and populations, altering the incubating temperature to better align with the warmer ambient temperatures of western bat roosts following Hamm et al. [120]. Culturing our western bat skin bacteria at the low temperatures described in probiotic therapy protocols designed for the Eastern United States may explain the prolonged growth period and low culturable richness observed for our culturing samples in this study.

From our subset of 12 culturing samples, we isolated and identified a total of 20 culturable species of bacteria from bat skin. Consistent with our hypothesis, all culturable isolates were found in very low relative abundances in the sequenced skin microbiome, as no isolate comprised more than 2% of the sequenced skin microbiome (Table 23). Our findings are opposite to the findings Walke et al. [116], who conducted a study with similar methods examining the skin microbiomes of amphibians. Walke et al. determined that the most abundant taxa in amphibian skin microbiomes were also culturable, being found in high relative abundances in the sequenced skin microbiome [116].

Pseudomonas spp. was our most abundant taxon in the sequenced bat skin microbiome, on average making up 27% of the bacterial bat skin microbiota across all bat

samples. Surprisingly, none of our culturable isolates included members of *Pseudomonas,* although it is known that *Pseudomonas* spp. are able to be cultured on the media we used [116], and that other bat skin microbiome studies have isolated culturable *Pseudomonas* from bat skin [27,31-33,120]. Perhaps this phenomenon is due to the diversity of the genus *Pseudomonas*, especially the diversity present at the species and strain levels [171,172]. This variation can result in different metabolic requirements for particular species or strains, limiting the culturing potential of the bacterium [171,172]. It is possible that the species and strains of *Pseudomonas* in our region require specific metabolic parameters that make them difficult to culture, as opposed to the culturable species and strains of *Pseudomonas* observed in previous bat studies mainly conducted in the Eastern United States [27,31-33], with the exception of the Hamm et al. study conducted in Arizona and New Mexico [120]. This would explain why *Pseudomonas* spp. were so abundant in the sequenced bat skin microbiome, but absent from the culturable isolates. It would be interesting to conduct additional bat microbiome studies in the Pacific Northwest region to determine if *Pseudomonas* spp. are abundant but unculturable in those bat skin microbiomes as well, which could be indicative of a broader region-specific difference in bat skin microbiota.

Although none of the culturable isolates included the well-known antifungal genus *Pseudomonas* that was so abundant in our sequenced skin microbiome, of the isolated culturable species, four (*Streptomyces laculatispora, Rhodococcus corynebacterioides, R. tukisamuensis, R. oryzae*) are members of genera with known antifungal properties [173-175]. We found that 20% of culturable isolates identified from our subset were from known antifungal genera (Fig. 38) [173-175], and all isolates from

these antifungal genera were cultured from bats sampled in eastern Washington (Lincoln County, Pend Oreille County). These genera have previously been isolated and cultured from bat skin in other studies conducted in the eastern United States [32,33], southwestern United States [120], and Canada [26,27,31]. Both *Streptomyces* spp. and *Rhodococcus* spp. are members of the phylum Actinobacteriota, a phylum composed of aerobic, Gram-positive bacteria with high guanine-cytosine contents, with most members engaging in differentiated life cycles [176].

Streptomyces spp. are widely abundant in soils, and their production of a large variety of secondary metabolites make them incredibly valuable for use in agriculture and ecosystem management [173]. Most importantly for bat ecology, most members of the genus *Streptomyces* are documented to possess antifungal properties [174,175], which could be incorporated in probiotic therapies to prevent or treat Pd infections [26,27,31-33,120]. Most members of the genus *Streptomyces* are antifungal via the production of chitin-digesting enzymes known as chitinases [174], digesting the main component in the rigid cell walls of fungi such as Pd [174]. Some members of *Streptomyces* also possess antifungal properties by targeting ergosterol in the cell membranes of fungi, lipids that reinforce the structure of the fungal cell membrane [175].

Similar to our findings, culturable *Streptomyces* spp. were previously isolated from bat skin by Hamm et al. in the Southwestern United States and tested for Pd-inhibitive properties using co-culture challenge assays [120]. Hamm et al. found that ~89% of all culturable isolates from bat skin with Pd-inhibitive properties were members of *Streptomyces*, indicating its potential for use as a probiotic to treat and prevent Pd infection [120]. In contrast to our study, Hamm et al. found that *Streptomyces* spp. was

the dominant genus of culturable bacteria found in their sequenced bat skin microbiomes, while in our study *Streptomyces* spp. was less common, only representing one bacterial isolate (*S. laculatispora*) from our culturing subset and on average only making up 0.004% of the sequenced bat skin microbiome. However, Hamm et al. [120] cultured their bat bacteria samples on actinobacterium-selective media. Perhaps if we plated our samples on actinobacterium-selective media instead of R2A, we would have found a greater proportion of *Streptomyces* spp. in our culturing subset. As *Streptomyces* spp. are commonly found in the environment in soils [173], it is possible that the small proportion of *S. laculatispora* in our sequenced bat skin microbiome could be due to regional differences between the region in our study (Pacific Northwest) and the region in the Hamm et al. study (Southwest) [120]. This would be a logical prediction as both geographic location and climate are well-documented to influence the types of microbes found in the environment and the skin microbiomes of the regional animal hosts [15,28-30,152-155].

We also isolated and identified three members of the genus *Rhodococcus* from our culturing sample subset: *R. corynebacterioides, R. tukisamuensis,* and *R. oryzae.* Overall, the genus *Rhodococcus* was one of the more abundant taxa in our sequenced bat skin microbiome, on average making up 1.63% of the bat skin microbiome in this study. However, the specific culturable species of *Rhodococcus* isolated in this study were not abundant in the total skin microbiome (Table 21). *Rhodococcus* is another genus of bacteria that is widely documented to possess antifungal properties via the production of antifungal tetrapeptides that break down cellulose and chitin in fungal cell walls [120,176,177-179]. *Rhodococcus* spp. are ubiquitous in the environment and have been

isolated from a diverse range of ecosystems and hosts, from soils and marine environments [180] to insects [181]. Many *Rhodococcus* spp. also produce pigments (i.e., carotenoids) for photo-oxidative protection [176].

Consistent with our findings, *Rhodococcus* spp. have been commonly cultured from bat skin from bat colonies in Costa Rica [182] Canada [26,29], and the southwestern United States [120], although this study is the first documentation of *Rhodococcus* spp. isolated from bats in the Pacific Northwest. Recent studies have isolated *Rhodococcus* spp. from bat skin and challenged *Rhodococcus* isolates with Pd to test for Pd-inhibitive properties [120,177]. Hamm et al. found that 13 *Rhodococcus* species isolated from the skin of bats successfully inhibited Pd *in vitro* during co-culture challenge assays [120], while Cornelison et al. found that *R. rhodochrous* inhibited Pd growth contact-independent of the Pd fungus, inhibiting the growth of Pd when sharing the same closed-air space as the fungal pathogen [120]. Field studies have also found that bats surviving post-Pd infections have altered skin microbiota with greater proportions of antifungal bacteria in their skin microbiomes, including members of *Rhodococcus* [26]. The clear anti-Pd properties of *Rhodococcus* spp. supports its potential as a key bacterial taxon for use in probiotic therapies to treat WNS in bat colonies [26,120,177].

From our culturing analyses, we identified species from two genera of bacteria (*Streptomyces* spp. and *Rhodococcus* spp.) previously documented to have both antifungal and specific Pd-inhibitive properties [120,174,175,177]. Although most published probiotic studies have focused on culturable *Pseudomonas* spp. as the antifungal bacterium of choice to isolate from bat skin and incorporate into anti-Pd probiotic therapies [26,27,31-33], the documented anti-Pd properties of *Streptomyces*

spp. and *Rhodococcus* spp. establish their potential and value for use in future probiotic therapy development [120,174,175,177]. While we isolated four species of bacteria from groups known to possess members with antifungal properties [120,174,175,177], Pd-inhibition assays would need to be conducted on our isolates to determine if these species are indeed antifungal and useful for use as probiotics.

However, even if these isolates are found to be anti-Pd, these isolates would need to be examined further for any potentially negative effects to the host species, especially as the abundances of those isolates in the skin microbiome would be increased with the application of the probiotic. Most notably, one of our isolates, *R. corynebacterioides*, has been documented to cause sepsis in humans [183]. If this taxon is found to be anti-Pd and is therefore a candidate for incorporation into probiotic therapies, *R. corynebacterioides* would need to be studied further to determine effects on non-target organisms and whether increasing its abundance in the bat skin microbiome would be deleterious to bat health, overriding any beneficial anti-Pd properties. The potential hazards to mammal and bat health specifically for our three remaining isolates are poorly understood, as all three have been previously isolated from soils but have not been examined in health applications [184-186].

Limitations of This Study:

In this study, we were limited by unbalanced representation of different bat species, ectoparasite presence statuses, and uneven numbers of sampled individuals from each roost due to the unpredictable nature of observational field studies. The unbalanced representation of variables in this study reduced the statistical power of our alpha and

beta diversity metric comparisons, resulting in higher p-values than a study with completely balanced variables [187]. Additionally, while we had a large sample size overall, the unbalanced observations of different bat species resulted in some groups having small sample sizes (i.e., n = 6 for *C. townsendii*) while other groups were more abundant (i.e., n = 57 for *M. yumanensis*), reducing the statistical power of our comparisons. Future studies could employ stratified sampling methods to ensure more balanced representation of different variables influencing the skin microbiota of bats.

We did not find any detectable influence of bat ectoparasite presence on the diversity and composition of the skin microbiome. However, we did find that bat species was a strong influencer of skin microbiome diversity and composition, consistent with other bat microbiome studies [26-29,93]. As previously mentioned, the representation of the different species sampled in this study was unbalanced. It is possible that the large influence of bat species on skin microbiome diversity and composition overrode any detectable difference in the skin microbiomes of bats with ectoparasites. Perhaps with larger sample sizes for each of our observed bat species, we would find a small effect of ectoparasite infestation on the diversity and composition of the skin microbiome.

During the analysis of our sequenced microbiome data, we found bacterial contamination present in our DNA extraction kit reagents, which is a common problem encountered in sensitive microbial studies [28,30,95,96]. We examined the taxonomy list of our sequenced bat skin microbiome data to identify the contaminants present in our sequenced controls, and other suspicious bacterial taxa that are unlikely to be present in large abundances in the bat skin microbiome (i.e., *Halomonas* spp., a deep-sea bacterium that would be found in large abundances in high-salinity buffer solutions, but not the bat

skin microbiome) [29,30]. Although it is difficult to ensure that a factory-direct DNA extraction kit is contamination-free before use [95,96], the common occurrence of reagent contamination solidifies the value of including control samples (i.e., DNA extraction kit controls, environmental controls from the field, etc.) when sequencing microbiomes in order to confidently remove contamination bioinformatically during subsequent data processing.

From the culturing portion of this study, we isolated and identified 20 species of culturable bacteria from bat skin, with four species from genera known to possess antifungal and anti-Pd properties [120,174,175,177]. While this knowledge is useful for future development of bat probiotic therapies [26,27,31-33,177], the sample size of our culturing subset was small (n = 12), so it is difficult to draw statistical conclusions from this subset of culturing data (i.e., how culturable species richness or certain taxa differ by the roost locations, ectoparasite presence, etc.). Future studies primarily focusing on the culturable bacteria from bat skin would need to use larger sample subsets in order to answer biologically relevant study questions.

Future Directions:

We found that bat ectoparasites did not influence the skin microbiome diversity and composition of our sampled bats. We hypothesize that these findings are attributed to the highly specialized life histories of bat ectoparasites [49,50,57,92,130-133,140-143] as opposed to the generalist life history of previously studied ectoparasites such as *S. scabiei* [24]. More skin microbiome studies relating to ectoparasites must be conducted to

determine if there are different effects on the host skin microbiota between generalist and specialists ectoparasites.

The 20 identified bacterial isolates and remaining culturing swabs collected for the culturing portion of this study can be used to test culturable bat skin bacteria for antifungal and specific Pd-inhibitive properties, by challenging each isolate with Pd and other fungal pathogens such as *Aspergillus* species [116,117]. This work would determine if naturally occurring bacteria in the skin of western bats could potentially be used as probiotic therapies for local colonies [116,117]. Future studies could follow the methods of Walke et al. [116,117] and McArthur et al. [113], culturing isolates from bat skin in lysogeny broth and filtering out bacterial cells from the broths, to be left with a cell-free supernatants containing the metabolites from the bacterial isolates [116,117]. Cell-free supernatants could then be challenged with the Pd fungus to identify isolates with anti-Pd properties [31,177]. Pd-inhibitive isolates could be potential candidates for the development of future probiotic therapies, following the applications of other anti-Pd probiotic studies [26,27,31-33,177].

There have been many studies examining the role of bacterial communities specifically in the protective bat skin microbiome as WNS becomes more pervasive throughout North America [28,29,31-33,82,93,120,121]. However, few North American studies have focused on the skin mycobiomes of bats – the naturally occurring communities of fungi living on bat skin, in relation to Pd [83,188]. In this study, we only focused on the bacterial communities of the bat skin microbiome, although the entire skin microbiota also includes fungi, archaea, protists, bacteriophages, and viruses [1-3]. Although the fungal communities of bat skin were not examined in this study, during our

culturing analyses we did observe that most of our mass-culture plate subset (67%) had at least one fungus growing on the media (Fig. 36). Because fungi were so abundant on our mass culture plates, we suspect that fungal communities may also be abundant and biologically relevant taxa in the skin microbiomes of our sampled bats. Further examination of these naturally occurring fungi, following the studies of Vanderwolf et al. [83,188] would be a valuable complement to the work done in this study.

Recent bat skin mycobiome studies by Vanderwolf et al. have found similar relationships between the skin mycobiome and risk of Pd as those observed between the bacterial communities of the skin microbiome and risk of Pd infection [83,188]. For example, Vanderwolf et al. [83] found that bat species with less diverse skin mycobiomes are at a greater risk of becoming infected with Pd, as lower-diversity communities, whether bacterial or fungal, are easier for a pathogen such as Pd to invade and proliferate [4,188]. Similar to the findings of bacterial probiotic studies examining anti-Pd bacteria from bat skin [26,27,31-33,177], a recent study by Vanderwolf et al. found that some groups of fungi, especially some strains of yeasts, also possess anti-Pd properties that could be used as fungal alternatives to traditional bacterial probiotics [188]. The important findings of these few mycobiome studies [83,188] indicate that the bat skin mycobiome is a promising area of future research for the ongoing and evolving management of WNS in North American bats.

Conclusions:

We found that bat ectoparasites did not decrease the skin microbiome diversity of bats, indicating that the previously hypothesized ectoparasite and skin microbiome

relationship (i.e., that mammalian ectoparasites reduce the skin microbiome diversity in mammals) may be contingent on the type of ectoparasite (i.e., generalist vs. specialist) and its relation to the host. Additionally, this indicates that bat ectoparasite infestation is unlikely to increase the risk of Pd infection in bats via an altered skin microbiome. However, consistent with other bat microbiome studies, we found that differences between bat species [26-29,93] and roost locations [28,30] significantly influenced the skin microbiome diversities and compositions of bats, although this is the first microbiome study conducted on bats in the Pacific Northwest region of the United States. We determined that overall, body mass within individual bat species and wing damage did not influence the diversity or composition of the skin microbiome, with a few exceptions that are likely due to variation between species and specific individuals sampled in this study.

We also identified 20 culturable species of bacteria from a subset of our samples, and consistent with our hypothesis, these culturable isolates were found in low abundances in the sequenced skin microbiome. Additionally, four of our culturable isolates were members of genera known to possess valuable antifungal and anti-Pd properties [120,174,175,177]. These isolates especially may be useful to incorporate into probiotic therapies as a means to prevent and treat WNS in the field for our region specifically, although further work still needs to be done to assess antifungal potential of these isolates.

This work provides valuable insight into the extent of which host-related factors such as ectoparasites influence the skin microbiome composition and diversity of North American bats. We stress that additional studies are needed to fully understand the

relationship between ectoparasites and the skin microbiota of bats, especially in regard to the type of ectoparasite. Understanding factors that influence the protective skin microbiome of bats is crucial in determining susceptibility of bats to Pd infection, and when managing colonies as WNS continues to spread throughout the Western United States. This work also provides knowledge of culturable bat skin microbes that may have potential as antifungal bacteria that could be incorporated into probiotic therapies in the future.

LITERATURE CITED

1. Hoffmann, A.R. et al. (2016) The Microbiome: the trillions of microorganisms that maintain health and cause disease in humans and companion animals. Vet. Pathol. 53, 10-21

2. McFall-Ngai, M., et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. USA. 110, 3229-3236

3. Hoffmann, A.R. (2017) The cutaneous ecosystem: the roles of the skin microbiome in health and its association with inflammatory skin conditions in humans and animals. In *Advances in Veterinary Dermatology* (8^{th} edn) (Torres, S.M.F. and Roudebush, P., eds), pp. 71–83, John Wiley & Sons, Ltd

4. Piovia-Scott, J. et al. (2017) Greater species richness of bacterial skin symbionts better suppresses the amphibian fungal pathogen *Batrachochytrium dendrobatidis*. Microb. Ecol. 74, 217–226

5. Dillon, R.J. and Dillon, V.M. (2004) The gut bacteria of insects: nonpathogenic interactions. Annu. Rev. Entomol. 49, 71–92

6. Coon, K.L. et al. (2014) Mosquitoes rely on their gut microbia for development. Mol. Ecol. 23, 2727–2739

7. Swei, A. and Kwan, J.Y. (2016) Tick microbiome and pathogen acquisition altered by host blood meal. ISME J. 1, 813–816

8. Dennison, N.J. et al. (2014) The mosquito microbiota influences vector competence for human pathogens. Curr. Opin. Insect Sci. 1, 6–13

9. Findley, K. and Grice, E.A. The skin microbiome: a focus on pathogens and their association with skin disease. PLOS Pathog. 10, 1–3

10. McCabe, R.A. et al (2020) Characterizing the microbiome of ectoparasitic louse flies feeding on migratory raptors. PLoS One. 15, 1–14

11. Weiss, B. and Aksoy, S. (2011) Microbiome influences on insect host vector competence. Trends Parisitol. 11, 514–522

12. Bahrndorff, S. et al. (2016) The microbiome of animals: implications for conservation biology. Int. J. Genomics. 2016, 1–17

13. Zinicola, M. et al. (2015) Altered microbiomes in Bovine Digital Dermatitis lesions, and the gut as a pathogen reservoir. PLoS One. 10, 1–23

14. Walke, J.B. and Belden, L.K. (2016) Harnessing the microbiome to prevent fungal infections: lessons from amphibians. PLoS Pathog. 12, 6–11

15. Ross, A.A. et al. (2019). The skin microbiome of vertebrates. Microbiome. 7, 1–14

16. Harrison, X.A. et al. (2019) Diversity-stability dynamics of the amphibian skin microbiome and susceptibility to a lethal viral pathogen. Front. Microbiol. 10, 1–13

17. Bates, K.A. et al. (2018) Amphibian chytridiomycosis outbreak dynamics are linked with host skin bacterial community structure. Nat. Commun. 9, 1–11

18. Bjerre, R.D. et al. (2017) The role of the skin microbiome in atopic dermatitis: a systematic review. Br. J. Dermatol. 177, 1272–1278

19. Mallon, C.A. et al. (2015) Microbial invasions: the process, patterns, and mechanisms. Trends Microbiol. 23, 719–729

20. Case, T.J. (1990) Invasion resistance arises in strongly interacting species-rich model competition communities. Proc. Natl. Acad. Sci. 87, 9610–9614

21. Kennedy, T.A. et al. (2002) Biodiversity as a barrier to ecological invasion. Nature. 417, 636–638

22. Shrestha, D. et al. (2015) Prevalence of demodectic mange in canines of Kathmandu Valley having skin disorder and its associated risk factors. Int. J. Appl. Sci. Biotechnol. 3, 459–463

23. Hoffmann, A.R. et al. (2014) The skin microbiome in healthy and allergic dogs. PLoS One. 9, 1–12

24. DeCandia, A.L. et al. (2019) Of microbes and mange: consistent changes in the skin microbiome of three canid species infected with *Sarcoptes scabiei* mites. Parasites Vectors. 12, 1-10

25. Jani, A.J. and Briggs, C.J. (2014) The pathogen *Batrachochytrium dendrobatidis* disturbs the frog skin microbiome during a natural epidemic and experimental infection. Proc. Natl. Acad. Sci. 111, 5049–5058

26. Lemieux-Labonté, V. et al. (2017) Enrichment of beneficial bacteria in the skin microbiota of bats persisting with white-nose syndrome. Microbiome. 5, 1-14
27. Lemieux-Labonté, V. et al. (2020) Antifungal potential of skin microbiota of hibernating Big Brown Bats (*Eptesicus fuscus*) infected with the causal agent of white-nose syndrome. Front. Microbiol. 1, 1-13

28. Avena, C.V. et al. (2016) Deconstructing the bat skin microbiome: influences of the host and the environment. Front. Microbiol. 7, 1-14

29. Ange-Stark, M. et al. (2019) White-nose syndrome restructures bat skin microbiomes. bioRxiv. DOI: https://doi.org/10.1101/614842

30. Lutz, H.L. et al. (2019) Ecology and host identity outweigh evolutionary history in shaping the bat microbiome. mSystems. 4, 1-16

31. Cheng, T. L. et al. (2017) Efficacy of a probiotic bacterium to treat bats affected by the disease white-nose syndrome. J. Appl. Ecol. 54, 701-708

32. Hoyt, J.R. et al. (2015) Bacteria isolated from bats inhibit the growth of *Pseudogymnoascus destructans*, the causative agent of white nose syndrome. PLoS One. 10, 1-12

33. Hoyt, J.R. et al. (2019) Field trial of a probiotic bacteria to protect bats from white-nose syndrome. Sci. Rep. 9, 1–9

34. Hopla, C.E. et al. (1994). Ectoparasites and classification. Rev. Sci. Tech. 13, 985 –1017

35. Morse, S.F. et al. (2013) Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). AEM. 79, 2952–2961

36. de Sousa, K.C.M. et al. (2018) *Rickettsia* spp. among wild mammals and their respective ectoparasites in Pantanal wetland, Brazil. Ticks Tick. Borne. Dis. 9, 10–17

37. Dick, C.W. (2007) High host specificity of obligate ectoparasites. Ecol. Entomol. 32, 446–450

38. Webber, Q.M.R. et al. (2015) Host demographic predicts ectoparasite dynamics for a colonial host during pre-hibernation mating. Parisitol. 142, 1260-1269

39. de Lope, F. et al. (1993) Increased detrimental effects of ectoparasites on their bird hosts during adverse environmental conditions. Oecologia. 95, 234–240

40. Cerutti, F. et al. (2018) The microbiota of hematophagous ectoparasites collected from migratory birds. PLoS One. 13, 1–16

41. Chu, X. et al. (2019) The influence of host body size and food guild on prevalence and mean intensity of chewing lice (Phthiraptera) on birds in southern China. J. Parasitol. 105, 334–344

42. Jackson, L.N. and Bateman, H.L. (2018) Differing ectoparasite loads, sexual modes, and abundances of whiptail lizards from native and non-native habitats. Herpetol. Conserv. Biol. 13, 294–301

43. Argaez, V. et al. (2019) Do ectoparasites affect survival of three species of lizards of the genus *Sceloporus*? Zoology. 138, 1–9

44. de Oliveira, C.N. et al. (2019). Acari of lizards from Atlantic forest in northeastern Brazil. Neotrop. Biol. Conserv. 14, 109–116

45. Torres, A.C. et al. (2018). *Amblyomma* ticks infesting amphibians and Squamata reptiles from the lower Amazon region, Brazil. Exp. Appl. Acarol. 75, 399–407

46. Reverter, M. et al. (2020) Insights into the natural defenses of a coral reef fish against gill ectoparasites: integrated metabolome and microbiome approach. Metabolites 10, 1–18

47. Kanbar, G. and Engels, W. (2003) Ultrastructure and bacterial infection of wounds in honeybee (*Apis mellifera*) pupae punctured by Varroa mites. Parasitol. Res. 90, 349 – 354

48. Estrada-Peña, A. (2015) Ticks as vectors: taxonomy, biology and ecology. Rev. Sci. Tech. 34, 53–65

49. Baker, J.R. (1967) A review of the role played by the Hippoboscidae (Diptera) as vectors of endoparasites. J. Parisitol. 53, 412–418

50. Czenze, Z.J. and Broders, H.G. (2011) Ectoparasite community structure of two bats (*Myotis lucifugus* and *M. septentrionalis*) from the maritimes of Canada. J. Parasitol. Res. 2011, 1–10

51. Korine, C. (2017) The effect of water contamination and host-related factors on ectoparasite load in an insectivorous bat. Parasitol. Res. 116, 2517–2526

52. Saldaña, M.A. et al. (2017) Microbial control of arthropod-borne disease. Mem. Inst. Oswaldo Cruz. 112, 81–93

53. Jupatanakul, N. et al. (2014) The insect microbiome modulates vector competence for arboviruses. Viruses. 6, 4294–4313

54. Lučan, R.K. et al. (2016) Ectoparasites may serve as vectors for the white-nose syndrome fungus. Parasites Vectors. 9, 1-5

55. Fuller, N.W. et al. (2020) Disease recovery in bats affected by white-nose syndrome. J. Exp. Biol. 22, 1–172

56. Reichard, J.D. and Kunz, T.H. (2009) White-nose syndrome inflicts lasting injuries to the wings of Little Brown Myotis (*Myotis lucifugus*). BioOne. 11, 457–464

57. Dick, C.W. and Pospischil, R. (2015) Nycteribiidae (bat flies). In: Encyclopedia of Parasitology. (1st edn) (Mehlhorn ed) pp. 1-4, Springer

58. Reeves, W.K., et al. (2016) Ecology of bats, their ectoparasites, and associated pathogens on Saint Kitts Island. J. Med. Entomol. 53, 1218–1225

59. Frick, W. F. et al. (2016) White-nose syndrome in bats. In *Bats in the Anthropocene: conservation of bats in a changing world* (1st edn) (Voigt, C.C. and Kingston, T., eds), pp. 245-262, Springer

60. Lorch, J.M. et al. (2016) First detection of white-nose syndrome in Western North America. mShere. 1, 1-5

61. Gargas, A., et al. (2009) *Geomyces destructans* sp. nov. associated with bat white-nose syndrome. Mycotaxon. 108, 147 – 154

62. Meteyer, C.U., et al. (2009) Histopathologic criteria to confirm white-nose syndrome in bats. J. Vet. Diagn. Invest. 21, 411 – 414

63. Raudabaugh, D.B. and Miller, A.N. (2013) Nutritional capability of and substrate suitability for *Pseudogymnoascus destructans*, the causal agent of bat White-Nose Syndrome. PLoS One. 8, 1 – 9

64. Reynolds, H.T. and Barton H.A. (2014) Comparison of the White-Nose Syndrome agent *Pseudogymnoascus* to cave-dwelling relatives suggests reduced saprotrophic enzyme activity. PLoS One. 9, 1 – 11

65. Drees, K.P., et al. (2017) Phylogenetics of a fungal invasion: origins and widespread dispersal of White-Nose Syndrome. mBio. 8, 1 – 15

66. Thapa, V., et al. (2021) Phylogeographic analysis of *Pseudogymnoascus destructans* partitivirus-pa explains the spread dynamics of white-nose syndrome in North America. PLoS Pathog.17, 1 – 21

67. Wibbelt, G., et al. (2013) Skin lesions in European hibernating bats associated with *Geomyces destructans*, the etiologic agent of White-Nose Syndrome. PLoS One. 8, 1 – 10

68. Zukal, J., et al. (2016) White-nose syndrome without borders: *Pseudogymnoascus destructans* infection tolerated in Europe and Palearctic Asia but not in North America. Sci. Rep. 6, 1 – 17

69. Hecht-Höger, A.M., et al. (2020) Plasma proteomic profiles differ between European and North American myotid bats colonized by *Pseudogymnoascus destructans.* Mol. Ecol. 29, 1745 – 1755

70. Khankhet, J., et al. (2014) Clonal expansion of the *Pseudogymnoascus destructans* genotype in North America is accompanied by significant variation in phenotypic expression. PLoS One. 9, 1 – 9

71. Palmer, J.M., et al. (2014) Molecular characterization of a heterothallic mating system in *Pseudogymnoascus destructans*, the fungus causing White-Nose Syndrome of bats. G3-Genes Genom. Genet. 4, 1755 – 1763

72. Verant, M.L. et al. (2012) Temperature-dependent growth of *Geomyces destructans*, the fungus that causes bat white-nose syndrome. PLoS One 7, 1-7

73. Ballmann, A.E. et al. (2017) Dispersal hazards of *Pseudogymnoascus destructans* by bats and human activity at hibernacula in summer. J. Wildl. Dis. 53, 725–735

74. Campbell, L.J. et al. (2020) Long-term survival of *Pseudogymnoascus destructans* at elevated temperatures. J. Wildl. Dis. 56, 278–287

75. Urbina, J. et al. (2021) *Pseudogymnoascus destructans* growth in wood, soil and guano substrates. Sci. Rep. 11, 1–11

76. Langwig, K.E. et al. (2015) Host and pathogen ecology drive the seasonal dynamics of a fungal disease, white-nose syndrome. Proc. R. Soc. B. 282, 1-7

77. Meteyer, C.U. et al. (2011) Recovery of Little Brown Bats (*Myotis lucifugus*) from natural infection with *Geomyces destructans*, white-nose syndrome. J. Wildl. Dis. 47, 618-626
78. Faure-Lacroix, J. et al. 2020. Long-term changes in bat activity in Quebec suggest climatic responses and summer niche partitioning associated with white-nose syndrome. Ecol. Evol. 10, 5226-5239

79. Washington Department of Fish and Wildlife. (2022) Where is white-nose syndrome in Washington? https://wdfw.wa.gov/species-habitats/diseases/bat-white-nose#wns-where

80. Weller, T.J. et al. (2018) A review of bat hibernacula across the western United States: implications for white-nose syndrome surveillance and management. PLoS One. 13, 1–20

81. Washington Department of Fish and Wildlife. (2018) White-nose syndrome strategy plan. https://www.whitenosesyndrome.org/response-plans/washington-department-of-fish-and-wildlife-strategy-plan-april-2018

82. Grisnik, M. et al. (2020) The cutaneous microbiota of bats has *in vitro* antifungal activity against the white nose pathogen. FEMS Microbiol. Ecol. DOI:10.1093/femsec/fiz193

83. Vanderwolf, K.J. et al. (2020) Skin-fungal assemblages of bats vary based on susceptibility to white-nose syndrome. ISME J. 15, 909–920

84. Kunz, T.H. et al. (2011) Ecosystem services provided by bats. Ann. N. Y. Acad. Sci. 1223, 1–38

85. Kolkert, H. et al. (2021) Insectivorous bats provide significant economic value to the Australian cotton industry. Ecosyst. Serv. 49, 101280

86. Rocke, T.E., et al. (2019) Virally-vectored vaccine candidates against white-nose syndrome induce anti-fungal immune response in little brown bats (*Myotis lucifugus*). Sci. Rep. 9, 1 – 12

87. Hicks, L.L., et al. (2020) A statistical approach to white-nose syndrome surveillance monitoring using acoustic data. PLoS One. 15, 1 – 16

88. Janicki, A.F. et al. (2015) Efficacy of visual surveys for white-nose syndrome at bat hibernacula. PLoS One. 10, 1–12

89. Amann, R.I. et al. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. FEMS Microbiol. Rev. 59, 143–169

90. Ranjard, L. et al. (2000) Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. Res. Microbiol. 151, 167–177

91. Rhoads, D.D. et al. (2012) Clinical identification of bacteria in human chronic wound infections: culturing vs. 16S ribosomal DNA sequencing. BMC Infect. Dis. 12, 1–8

92. Poissant, J.A. and Broders H.G. (2008) Ectoparasite prevalence in *Myotis lucifugus* and *M. septentrionalis* (Chiroptera: Vespertilionidae) during fall migration at Hayes Cave, Nova Scotia. Northeast. Nat. 15, 515–522

93. Lemieux-Labonté, V. et al. (2016) Environment and host species shape the skin microbiome of captive neotropical bats. PeerJ. 4, e2430

94. Yang, B. et al. (2016) Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinform. 17, 1–8

95. Caporaso, J.G. et al. (2018) EMP 16S Illumina amplicon protocol. https://www.protocols.io/view/emp-16s-illumina-amplicon-protocol-nuudeww

96. Weyrich, L.S. et al. (2019) Laboratory contamination over time during low-biomass sample analysis. Mol. Ecol. Resour. 19, 982 – 96

97. Caporaso, J.G. et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624

98. Bolyen, E. et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857

99. Amir, A. et al. (2016) Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems. 2, 1–7

100. Bokulich, N.A. et al. (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods. 10, 1–7

101. Quast, C. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596

102. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780

103. Price, M.N. et al. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One. 5, 1–10

104. DeJong, T.M. (1975) A comparison of three diversity indices based on their components of richness and evenness. Oikos. 26, 222–227

105. Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10

106. Braun, J.K. et al. (2015) *Myotis yumanensis* (Chiroptera: Vespertilionidae). Mamm. Species. 47, 1 – 14

107. Kurta, A. and Baker, R.H. (1990) *Eptesicus fuscus*. Mamm. Species. No. 356, 1 – 10

108. Bag, S. et al. (2019) An efficient recommendation generation using relevant Jaccard similarity. Inf. Sci. 483, 53–64

109. Sørensen, T. (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biologiske Skrifter. 5, 1–34

110. Lozupone, C.A. et al. (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl. Environ. Microbiol. 73, 1576–1585

111. De Cáceres, M. and Legendre, P. (2009) Associations between species and groups of sites: indices and statistical inference. Ecology. 90, 3566–3574

112. De Cáceres, M. et al. (2010) Improving indicator species analysis by combining groups of sites. Oikos. 119, 1674–1684

113. McArthur, R.L. (2017) Improvement of protocols for the screening of biological control agents against white-nose syndrome. JEMI. 2, 1–7

114. Kołwzan, B. et al. (2011) Sanitary Biology. In *Environmental Quality Management* (1st edn) pp. 13, Wrocław University of Technology

115. Tsai, Y. and Olson, B.H. (1991) Rapid method for direct extraction of DNA from soil and sediments. Appl. Env. Microbiol. 57, 1070 – 1074

116. Walke, J.B. et al. (2015) Most of the dominant members of amphibian skin bacterial communities can be readily cultured. Appl. Environ. Microbiol. 81, 6589–6600

117. Walke, J.B. et al. (2017) Dominance-function relationships in the amphibian skin microbiome. Environ. Microbiol. 19: 3387–3397

118. Kearse, M. et al. (2012a) The Geneious 6.0.3 Read Mapper.

119. Kearse, M., et al. (2012b) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28, 1647–1649

120. Hamm, P.S. et al. (2017) Western bats as a reservoir of novel *Streptomyces* species with antifungal activity. Appl. Environ. Microbiol. 83, e03057–16

121. Winter, A.S. et al. (2017) Skin and fur bacterial diversity and community structure on American southwestern bats: effects of habitat, geography and bat traits. PeerJ. 5, e3944

122. Fenton, M.B., and Barclay, M.R.B. (1980) *Myotis lucifugus*. Mamm. Species. No. 142, 1 – 8

123. Kunz, T.H., and Martin, R.A. (1982) *Plecotus townsendii.* Mamm. Species. No. 175, 1 – 6

124. Warner, R.M., and Czaplewski, N.J. (1984) *Myotis volans*. Mamm. Species. No. 224, 1 – 4

125. Hoofer, S.R, et al. (2003) Molecular phylogenetics of the Chiropteran family Vespertilionidae. Acta Chiropt. 5, 1 – 63

126. Perkins, J.M. et al. (1990) Winter records of bats in Oregon and Washington. Northwest Nat. 71, 59–62

127. Washington Department of Fish and Wildlife (2022) Common bats of Washington. https://wdfw.wa.gov/species-habitats/living/species-facts/bats#types

128. Dalquest, W.W. (1938) Bats in the State of Washington. J. Mammal. 19, 211–213

129. Washington Department of Fish and Wildlife (2021) Bat-killing fungus that causes white-nose syndrome continues to spread in Washington with recent detections in Chelan, Mason, and Yakima counties. https://wdfw.wa.gov/news/bat-killing-fungus-causes-white-nose-syndrome-continues-spread-washington-recent-detections

130. Christie, P. et al. (2000) Variation in intensity of a parasitic mite (*Spinturnix myoti*) in relation to the reproductive cycle and immunocompetence of its bat host (*Myotis myotis*). Ecol. Lett. 3, 207–212

131. Lučan, R.K. (2006) Relationships between the parasitic mite *Spinturnix andegavinus* (Acari: Spinturnicidae) and its bat host, *Myotis daubentonii* (Chiroptera: Vespertilionidae): Seasonal, sex- and age-related variation in infestation and possible impact of the parasite on the host condition and roosting behavior. Folia Parasitol. 53, 147–152

132. Pearce, R.D.,and O'Shea, T. J. (2007) Ectoparasites in an urban population of Big Brown Bats (*Eptesicus fuscus*) in Colorado. J. Parasitol. 93, 518–530

133. Dick, C.W., and Patterson, B.D. (2006) Bat flies: obligate ectoparasites of bats. In: *Micromammals and Macroparasites* (1st edn) (Morand, S., B.R. Krasnov, R. Poulineds eds) pp.179-194, Springer

134. Grisnik, M. et al. (2021) The presence of *Pseudogymnoascus destructans*, a fungal pathogen of bats, correlates with changes in microbial metacommunity structure. Sci. Rep. 11, 11685

135. Shelley, W.B. et al. (1988) *Staphylococcus aureus* colonization of burrows in erythrodermic Norwegian scabies. A case study of iatrogenic contagion. J. Am. Acad. Dermatol. 19, 673–678

136. Mika, A. et al. (2012) complement inhibitors from Scabies mites promote Streptococcal growth – a novel mechanism in infected epidermis? PLoS Negl. Trop. Dis. 6, e1563

137. Swei, P.M., and Fischer, K. (2014) A Scabies mite serpin interferes with complement-mediated neutrophil functions and promotes Staphylococcal growth. PLoS Negl. Trop. Dis. 8, e2928

138. Arlian, L.G., and Morgan M.S. (2017) A review of *Sarcoptes scabiei*: past, present and future. Parasites Vectors. 10, 1–22

139. Browne, E. et al. (2022) Sustaining transmission in different host species: the emblematic case of *Sarcoptes scabiei.* BioScience. 72, 166–176

140. Frank, R. et al. (2015) Parasite diversity of European *Myotis* species with special emphasis on *Myotis myotis* (Microchiroptera, Vespertilionidae) from a typical nursery roost. Parasites Vectors, 8, 1–13

141. Goddard, J. et al. (2012) Bed bugs (*Cimex lectularius*) and bat bugs (several *Cimex* species): a confusing issue. Outlooks Pest Manag. 23, 125–127

142. Luong, N.T. et al. (2021) First record of *Eyndhovenia* (Mesostigmata: Gamasina: Spinturnicidae) from Vietnam. Parasitol. Int. 82, 102301

143. Giorgi, M.S. et al. (2004) Casual mechanisms underlying host specificity in bat ectoparasites. Oceologica. 138, 648–654

144. Davies, C.M. et al. (2003) Trade-offs in the evolution of virulence in an indirectly transmitted macroparasite. Proc. R. Soc. Lond. B. 268, 251–257

145. Le Clec'h, W. et al. (2019) Striking differences in virulence, transmission and sporocyst growth dynamics between two schistosome populations. Parasites Vectors. 12, 1–12

146. Greenblat, C.L. et al. (2004) *Micrococcus luteus* – survival in amber. Microb. Ecol. 48, 120–127

147. Murrell, A. et al. (2003) A survey of bacterial diversity in ticks, lice and fleas from Australia. Parasitol. Res. 89, 326–334

148. Xu, X. et al. (2011) *Pelagibacterium halotolerans* gen. nov., sp. nov. and *Pelagibacterium luteolum* sp. nov., novel members of the family Hyphomicrobiaceae. Int. J. Syst. Evol. 61, 1817–1822

149. Shukla, S.K. et al. (2004) Isolation of a fastidious *Bergeyella* species associated with cellulitis after a cat bite and a phylogenetic comparison with *Bergeyella zoohelcum* strains. J. Clin. Microbiol. 42, 290–93

150. Yumoto, I. et al. (2004) *Alkalibacterium psychrotolerans* sp. nov., a psychrotolerant obligate alkaliphile that reduces an indigo dye. Int. J. Syst. Evol. 54, 2379 – 2383

151. Muletz Wolz, C.R. et al. (2017) Effects of host species and environment on the skin microbiome of Plethodontid salamanders. J. Anim. Ecol. 87, 341–353

152. Woodhams, D.C. et al. (2020) Host-associated microbiomes are predicted by immune system complexity and climate. Genome Biol. 21, 1–20

153. van Mierlo, M.M.F. et al. (2019) The influence of treatment in alpine and moderate maritime climate on the composition of the skin microbiome in patients with difficult to treat atopic dermatitis. Clin. Exp. Allergy. 49, 1437–1445

154. Kueneman, J.G. et al. (2019) Community richness of amphibian skin bacteria correlates with bioclimate at the global scale. Nat. Ecol. Evol. 3, 381–389

155. Bierlich, K.C. et al. (2018) Temporal and regional variability in the skin microbiome of humpback whales along the Western Antarctic Peninsula. AEM. 84, e02574–17

156. Lozupone, C.A., and Knight, R. (2007) Global patterns in bacterial diversity. PNAS. 104, 11436–11440

157. Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. PNAS. 103, 626–631

158. McGuire, L.P. et al. (2018) Common condition indices are no more effective than body mass for estimating fat stores in insectivorous bats. J. Mammal. 99, 1065–1071

159. Lindström, A., and Piersma, T. (1993) Mass changes in migrating birds: the evidence for fat and protein storage re-examined. Ibis. 135, 70–78

160. Jousset, A. et al. (2017) Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. 11, 853–862

161. Ze, X. et al. (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. ISME J. 6, 1535–1543

162. Rafrafi, Y. et al. (2013) Sub-dominant bacteria as keystone species in microbial communities producing bio-hydrogen. Int. J. Hydrog. Energy. 38, 4975–4985

163. Shade, A. et al. (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio. 5, 1–9

164. Walke, J.B. et al. (2014) Amphibian skin may select for rare environmental microbes. ISME J. 8, 2207–2217

165. Hol, W.H.G. et al. (2015) Non-random species loss in bacterial communities reduces antifungal volatile production. Ecology. 96, 2042–2048

166. Hale, S.L., and Koprowski, J.L. (2018) Ecosystem-level effects of keystone species reintroduction: a literature review. Restor. Ecol. 26, 439–445

167. Menge, B.A. et al. (1994) The keystone species concept: variation in interaction strength in a rocky intertidal habitat. Ecol. Monogr. 64, 249–286

168. Hayes, M.A. et al. (2011) Hibernacula selection by Townsend's big-eared bat in Southwestern Colorado. J. Wildl. Manag. 75, 137–143

169. McClure, M.L. et al. (2020) Linking surface and subterranean climate: implications for the study of hibernating bats and other cave dwellers. Ecosphere. 11, e03274

170. Swezey, C.S., and Garrity, C.P. (2011) Geographical and geological data from caves and mines infected with White-Nose Syndrome (WNS) before September 2009 in the Eastern United States. J. Caves Karst Stud. 73, 125–157

171. Spiers, A.J. et al. (2000) The causes of *Pseudomonas* diversity. Micobiology. 146, 2345–2350

172. Sibley, M.W. et al. (2011) *Pseudomonas* genomes: diverse and adaptable. FEMS Microbiol. Rev. 35, 652–680

173. Olanrewaju, O.S., and Babalola, O.O. (2019) *Streptomyces*: implications and interactions in plant growth promotion. Appl. Microbiol. Biotechnol. 103, 1179–1188

174. Schrempf, H. (2001) Recognition and degradation of chitin by streptomycetes. Antonie van Leeuwenhoek. 79, 285–289

175. Seco, E.M. et al. (2005) Two polyene amides produced by genetically modified *Streptomyces diastaticus* var. 108. Chem. Biol. 12, 535–543

176. Elsayed, Y. et al. (2017) The Genus *Rhodococcus* as a source of novel bioactive substances: a review. J. Pharmacogn. Phytochem. 6, 83–92

177. Cornelison, C.T. et al. (2014) A preliminary report on the contact-independent antagonism of *Pseudogymnoascus destructans* by *Rhodococcus rhodochrous* strain DAP96253. BMC Microbiol. 14, 246

178. Kawato, H.C. et al. (2000) Synthesis and antifungal activity of Rhodopeptin analogues. 1. modification of the east and south amino acid moieties. Org. Lett. 2, 973–976

179. Chiba, H. et al. (1999) Rhodopeptins (Mer-N1033), novel cyclic tetrapeptides with antifungal activity from *Rhodococcus* sp. J. Antibiot. 52, 695–699

180. Choi, H. et al. (2016) Microbial community composition in the marine sediments of Jeju Island: next-generation sequencing surveys. J. Microbiol. Biotechnol. 26, 884–890

181. Berlanga, M. (2015) Functional symbiosis and communication in microbial ecosystems. The case of wood-eating termites and cockroaches. Int. Microbiol. 18, 159–169

182. Voigt, C.C. et al. (2005) Bats, bacteria, and bat small: sex-specific diversity of microbes in a sexually selected scent organ. J. Mammal. 86, 745–749

183. Kitamura, Y. et al. (2011) First report of sepsis caused by *Rhodococcus corynebacterioides* in a patient with Myelodysplastic Syndrome. J. Clin. Microbiol. 50, 1089–1091

184. Matsuyama, H. et al. (2003) *Rhodococcus tukisamuensis* sp. nov., isolated from soil. Int. J. Syst. Evol. Microbiol. 53, 1333–1337

185. Zucchi, T.D. et al. (2012) *Streptomyces brevispora* sp. nov. and *Streptomyces laculatispora* sp. nov., actinomycetes isolated from soil. Int. J. Syst. Evol. Microbiol. 62, 478–483

186. Li, C. et al. (2020) *Rhodococcus oryzae* sp. nov., a novel actinobacterium isolated from rhizosphere soil of rice (*Oryza sativa* L.). Int. J. Syst. Evol. Microbiol. 70, 3300–3308

187. Hector, A. et al. (2010) Analysis of variance with unbalanced data: an update for ecology & evolution. J. Anim. Ecol. 79, 308–316

188. Vanderwolf, K.J. et al. (2021) Mycobiome traits associated with disease tolerance predict many Western North American bat species will be susceptible to White-Nose Syndrome. Microbiol. Spectrum. 9, e00254–21

189. White-nose Syndrome Response Team. (2020) Spread map. https://www.whitenosesyndrome.org/

190. Sames, W., et al. (2008) Distribution of mosquitoes in Washington state. J. Am. Mosquito Contr. 23, 442–448

191. Kunz, T.H. and Anthony, E.L.P. (1982) Age estimation and post-natal growth in the bat *Myotis lucifugus*. J. Mammal. 63, 23–32

192. Fettig, S. et al. (2018) Survey of Little Brown Bats (*Myotis lucifugus*): observing general health and disease progression of White Nose Syndrome (*Pseudogymnoascus destructans*) in Lincoln County, Washington. Eastern

Washington University Research and Creative Works Symposium, Cheney, Washington

TABLES

Table 1: Metadata from the 90 individuals used in sequenced skin microbiome analyses.

Sample ID	Site	Ectoparasites	Species	Sex	WDI	Weight	Pd Status
RR.L.10	Lincoln	Y	MYYU	F	1	5	(-)
RR.L.11	Lincoln	N	MYYU	F	0	5.5	(-)
RR.L.12	Lincoln	N	MYYU	F	0	6	(-)
RR.L.13	Lincoln	Y	MYYU	F	0	6	(-)
RR.L.14	Lincoln	Y	MYYU	F	0	6	(-)
RR.L.15	Lincoln	N	MYYU	F	0	6.5	(-)
RR.L.16	Lincoln	N	COTO	M	0	8.5	(-)
RR.L.17	Lincoln	N	MYYU	F	0	5	(-)
RR.L.18	Lincoln	N	MYYU	F	0	6	(-)
RR.L.19	Lincoln	Y	MYYU	F	0	6	(-)
RR.L.20	Lincoln	N	MYYU	F	0	6.5	(-)
S.S.1	Spokane	N	MYYU	F	0	5.3	(-)
S.S.5	Spokane	N	MYYU	F	0	5	(-)
S.S.7	Spokane	N	MYYU	F	0	5.1	(-)
S.S.9	Spokane	N	MYYU	F	0	5.1	(-)
S.S.11	Spokane	N	MYYU	F	0	5	(-)
S.S.12	Spokane	N	MYYU	F	0	5.3	(-)
S.S.13	Spokane	N	MYYU	F	0	4.5	(-)
S.S.16	Spokane	N	MYYU	F	0	5	(-)
S.S.18	Spokane	N	MYYU	F	0	5	(-)
S.S.19	Spokane	N	MYYU	F	0	5.4	(-)
S.S.20	Spokane	N	MYYU	F	0	5	(-)
WB.T.1	Thurston	N	MYYU	F	1	5.2	(-)
WB.T.2	Thurston	N	MYYU	F	1	4.2	(-)
WB.T.3	Thurston	N	MYYU	F	0	5.6	(-)
WB.T.10	Thurston	N	MYLU	F	1	6.1	(-)
WB.T.12	Thurston	Y	MYLU	F	1	6.1	(-)
WB.T.13	Thurston	Y	MYLU	F	1	5.7	(-)
WB.T.15	Thurston	Y	MYYU	F	1	5.6	(-)
WB.T.16	Thurston	N	MYYU	F	0	5.3	(-)
WB.T.17	Thurston	Y	MYYU	F	0	6	(-)
WB.T.18	Thurston	Y	MYLU	F	0	6.2	(-)
POM.PO.1	Pend Oreille	Y	MYYU	F	0	5.4	(-)
POM.PO.2	Pend Oreille	N	COTO	F	0	9.5	(-)
POM.PO.4	Pend Oreille	Y	MYYU.LU	F	0	6.1	(-)
POM.PO.5	Pend Oreille	Y	MYYU.LU	F	0	6.9	(-)

Sample ID	Site	Ectoparasites	Species	Sex	WDI	Weight	Pd Status
POM.PO.8	Pend Oreille	N	COTO	F	0	9.3	(-)
POM.PO.10	Pend Oreille	N	COTO	F	0	9.8	(-)
POM.PO.13	Pend Oreille	Y	MYYU.LU	M	0	6	(-)
POM.PO.16	Pend Oreille	N	COTO	F	0	10.2	(-)
POM.PO.18	Pend Oreille	N	COTO	F	0	8.7	(-)
POM.PO.19	Pend Oreille	N	MYLU	F	0	7.1	(-)
POM.PO.20	Pend Oreille	N	MYYU.LU	F	0	5.8	(-)
HH.M.2	Mason	N	MYLU	F	0	5.5	(-)
HH.M.6	Mason	Y	EPFU	M	0	14.3	S
HH.M.7	Mason	N	MYLU	F	0	5.4	(-)
HH.M.8	Mason	N	MYLU	F	0	5.5	(-)
HH.M.10	Mason	N	MYLU	F	0	5.9	(-)
HH.M.12	Mason	N	EPFU	M	0	13.7	(-)
HH.M.13	Mason	N	MYLU	F	0	5.4	(-)
HH.M.14	Mason	N	MYLU	F	0	6.4	(-)
HH.M.17	Mason	Y	MYLU	F	0	5.4	(-)
HH.M.19	Mason	N	MYLU	F	0	6.5	(-)
HH.M.20	Mason	N	MYLU	F	0	5.1	(-)
CL.KL.1	Klickitat	N	EPFU	F	1	17.4	(-)
CL.KL.2	Klickitat	Y	EPFU	F	0	20.7	(-)
CL.KL.3	Klickitat	Y	EPFU	F	1	17.6	(-)
CL.KL.5	Klickitat	Y	EPFU	F	0	17.1	(-)
CL.KL.6	Klickitat	Y	EPFU	F	0	20.7	(-)
CL.KL.7	Klickitat	Y	EPFU	F	0	19.1	(-)
CL.KL.8	Klickitat	N	EPFU	F	0	18.6	(-)
CL.KL.9	Klickitat	Y	EPFU	F	0	17.4	(-)
CL.KL.10	Klickitat	N	EPFU	F	0	19.6	(-)
LC.C.1	Chelan	N	MYYU	F	0	4.8	(-)
LC.C.2	Chelan	N	MYLU	F	0	4.4	(-)
LC.C.3	Chelan	N	MYLU	F	0	5.8	(-)
LC.C.4	Chelan	N	MYLU	F	0	5.1	(-)
LC.C.6	Chelan	N	MYLU	F	0	5.6	(-)
LC.C.7	Chelan	N	MYLU	F	0	5.5	(-)
LC.C.8	Chelan	N	MYYU.LU	F	1	4.6	(+)
LC.C.1	Chelan	N	MYYU	F	0	4.8	(-)
TL.KL.1	Klickitat	N	MYYU	F	0	5.1	(-)
OR.LS.1	Lewis	N	MYVO	F	0	7.7	(-)
OR.LS.2	Lewis	Y	MYVO	F	0	7.2	(-)

Sample ID	Site	Ectoparasites	Species	Sex	WDI	Weight	Pd Status
OR.LS.3	Lewis	Y	MYVO	F	0	7	(-)
OR.LS.7	Lewis	Y	MYVO	F	0	7.1	(-)
OR.LS.8	Lewis	N	MYLU	F	0	5.6	(-)
OR.LS.10	Lewis	Y	MYVO	F	0	6.9	(-)
OR.LS.12	Lewis	Y	MYLU	F	0	5.8	(-)
OR.LS.13	Lewis	N	MYVO	F	0	7.7	(-)
OR.LS.14	Lewis	N	MYVO	F	0	7.1	(-)
OR.LS.17	Lewis	N	MYLU	F	0	5.4	(-)
OR.LS.18	Lewis	N	MYVO	F	0	6.7	(-)

Table 2: Bat species sampled in this study.

Bat Species	Distribution	Morphology and Life History	Sources
Myotis lucifugus 	Widespread throughout North America, also throughout central Mexico	Pointy tragus, forearm length 35 – 44 mm, no keeled calcar. Insectivorous. Mating occurs from Spring – early Fall, delayed fertilization until Spring, birth to pups in late Spring. Hibernate singly or in small groups in the Pacific Northwest, large groups in eastern U.S.	Fenton and Barclay 1980 [122] WDFW [81]
Myotis yumanensis 	Western North America, northwest Mexico, found as far east as western Texas	Pointy tragus, forearm length >35 – 38 mm, no keeled calcar. Higher frequency calls than *M. lucifugus.* Insectivorous. Mating occurs from Spring – early Fall, delayed fertilization until Spring, birth to pups in late Spring.	Braun et al. 2015 [106] WDFW [81]
Eptesicus fuscus 	Widespread throughout North America, Mexico, northern South America	Blunt tragus, forearm length 39 – 54 mm, no keeled calcar. Insectivorous, ovulation and fertilization delayed until Spring, birth to pups in late Spring. Hibernate singly or in small groups	Kurta and Baker 1990 [107] WDFW [81]
Myotis volans 	Western North America through central Mexico, found as far east as northwest Texas	Pointy tragus, forearm length 37 – 41.2 mm, keeled calcar, long forelegs. Insectivorous, especially moths. Ovulation and fertilization delayed until Spring, birth to pups from late June – August. Hibernate singly or in small groups.	Warner and Czaplewski 1984 [124] WDFW [81]
Corynorhinus townsendii 	Western North America through central Mexico, found as far east as western Texas	Pointy tragus, forearm length 39 – 47.6 mm, no keeled calcar, large pinnae. Low frequency echolocation calls. Insectivorous, especially moths. Ovulation and fertilization delayed until Spring, birth to pups in late Spring. Most hibernate singly or in small groups.	Kunz and Martin 1982 [123] WDFW [81]

Table 3: Alpha diversity results for ectoparasite presence in bats.

Alpha Diversity: Ectoparasite Presence		
Alpha Diversity Metric	**p-value**	**H Test Statistic**
Shannon's Diversity	0.493	0.47
Faith's Phylogenetic Diversity	0.336	0.924
Evenness	0.482	0.495
Observed Features	0.869	0.0272

Table 4: Summary statistics for ectoparasite presence alpha diversity metrics.

Alpha Diversity Summary Statistics: No Ectoparasites		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.313	0.5257593
Faith's Phylogenetic Diversity	7.81304	0.7365939
Evenness	0.6565895	0.07099988
Observed Features	95	12.97739
Alpha Diversity Summary Statistics: Ectoparasites		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.318	0.2941659
Faith's Phylogenetic Diversity	7.975324	0.6672077
Evenness	0.6572095	0.03210475
Observed Features	95.76923	13.95796

Table 5: Beta diversity results for ectoparasite presence in bats.

Beta Diversity: Ectoparasite Presence		
Beta Diversity Metric	**p-value**	**Pseudo-F Test Statistic**
Bray-Curtis Dissimilarity	0.509	0.906
Jaccard Similarity	0.058	1.281
Unweighted UniFrac Distance	0.158	0.495
Weighted UniFrac Distance	0.965	0.306

Table 6: Alpha diversity results for ectoparasite presence within individual bat species. *Corynorhinus townsendii* is excluded from this table because this species was not observed to have ectoparasites.

Alpha Diversity: Ectoparasite Presence Within Species			
Species	**Alpha Diversity Metric**	**p-value**	**H Test Statistic**
MYYU	Shannon's Diversity	0.36	0.839
	Faith's Phylogenetic Diversity	0.297	1.0889
	Evenness	0.241	1.372
	Observed Features	0.536	0.382
MYLU	Shannon's Diversity	0.145	2.123
	Faith's Phylogenetic Diversity	0.499	0.456
	Evenness	0.0816	3.0328
	Observed Features	0.695	0.154
MYYU/LU	Shannon's Diversity	0.289	1.125
	Faith's Phylogenetic Diversity	0.48	0.5
	Evenness	0.724	0.125
	Observed Features	0.721	0.127
MYVO	Shannon's Diversity	0.248	1.333
	Faith's Phylogenetic Diversity	0.386	0.75
	Evenness	0.148	2.0833
	Observed Features	0.148	2.0833
EPFU	Shannon's Diversity	0.345	0.893
	Faith's Phylogenetic Diversity	0.345	0.893
	Evenness	0.45	0.571
	Observed Features	0.295	1.095

Table 7: Beta diversity results for ectoparasite presence within individual bat species. *Corynorhinus townsendii* is excluded from this table because this species was not observed to have ectoparasites.

Beta Diversity: Ectoparasite Presence Within Species			
Species	**Beta Diversity Metric**	**p-value**	**Pseudo-F Test Statistic**
MYYU	Bray-Curtis Dissimilarity	0.264	1.123
	Jaccard Similarity	0.223	1.11
	Unweighted UniFrac Distance	0.297	1.0978
	Weighted UniFrac Distance	0.339	1.07
MYLU	Bray-Curtis Dissimilarity	0.468	0.82
	Jaccard Similarity	0.067	1.298
	Unweighted UniFrac Distance	0.18	1.262
	Weighted UniFrac Distance	0.567	0.555
MYYU/LU	Bray-Curtis Dissimilarity	0.355	1.103
	Jaccard Similarity	0.095	1.197
	Unweighted UniFrac Distance	0.143	1.316
	Weighted UniFrac Distance	0.694	0.764
MYVO	Bray-Curtis Dissimilarity	0.656	0.983
	Jaccard Similarity	0.17	1.144
	Unweighted UniFrac Distance	0.597	0.994
	Weighted UniFrac Distance	0.261	1.217
EPFU	Bray-Curtis Dissimilarity	0.714	0.685
	Jaccard Similarity	0.339	1.0548
	Unweighted UniFrac Distance	0.963	0.63
	Weighted UniFrac Distance	0.963	0.63

Table 8: Indicspecies analysis results for ectoparasite presence in bats.

Ectoparasites	Taxon	p-value	Maximum Preference Value Test Statistic
No ectoparasites	*Micrococcus* spp.	0.045	0.395
Ectoparasites	*Nocardioides* spp.	0.04	0.432
Ectoparasites	*Pelagibacterium* spp.	0.02	0.4
Ectoparasites	*Bergeyella* spp.	0.02	0.381
Ectoparasites	*Alkalibacterium* spp.	0.04	0.329

Table 9: Alpha diversity results for bat species.

Alpha Diversity: Bat Species		
Alpha Diversity Metric	**p-value**	**H Test Statistic**
Shannon's Diversity	0.00001	30.817
Faith's Phylogenetic Diversity	0.00397	17.299
Evenness	0.000175	24.486
Observed Features	0.0000651	26.705

Table 10: Summary statistics for bat species alpha diversity metrics.

Alpha Diversity Summary Statistics: COTO		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.166595	0.18383732
Faith's Phylogenetic Diversity	7.669598	0.5736866
Evenness	0.6518048	0.03249243
Observed Features	84.66667	9.521905
Alpha Diversity Summary Statistics: EPFU		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.051772	0.16566015
Faith's Phylogenetic Diversity	7.487326	0.4305137
Evenness	0.6366051	0.01910651
Observed Features	82.72727	8.331757
Alpha Diversity Summary Statistics: MYLU		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.16566	0.72928399
Faith's Phylogenetic Diversity	7.704279	0.7110886
Evenness	0.6336169	0.09887552
Observed Features	94.41667	14.261278
Alpha Diversity Summary Statistics: MYVO		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.295199	0.24948447
Faith's Phylogenetic Diversity	8.75417	0.6365075
Evenness	0.6389253	0.03169596
Observed Features	105.625	6.545173
Alpha Diversity Summary Statistics: MYYU		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.556329	0.29781257
Faith's Phylogenetic Diversity	7.897108	0.6844648
Evenness	0.6868898	0.03879206
Observed Features	99.82353	11.771745
Alpha Diversity Summary Statistics: MYYU/LU		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.21224	0.09401101
Faith's Phylogenetic Diversity	7.939588	0.7700855
Evenness	0.6461752	0.01423101
Observed Features	92.42857	11.19311

Table 11: Beta diversity results for bat species.

Beta Diversity: Bat Species		
Beta Diversity Metric	**p-value**	**Pseudo-F Test Statistic**
Bray-Curtis Dissimilarity	0.001	2.142
Jaccard Similarity	0.001	2.269
Unweighted UniFrac Distance	0.001	2.079
Weighted UniFrac Distance	0.006	2.527

Table 12: pairwise comparisons between different bat species for Weighted UniFrac Distances.

Weighted UniFrac Pairwise Comparisons by Bat Species		
Species Comparison	**p-value**	**Pseudo-F Test Statistic**
COTO vs. EPFU	0.02	3.386
COTO vs. MYYU	0.004	5.374
EPFU vs. MYVO	0.018	2.266
EPFU vs. MYYU	0.002	6.051
MYVO vs. MYYU	0.02	3.795
MYYU vs. MYYU/LU	0.008	4.293

Table 13: Indicspecies analysis results for bat species.

Bat Species	Taxon	p-value	Maximum Preference Value Test Statistic
COTO	*Brumimicrobium* spp.	0.005	0.697
COTO	Family Alcaligenaceae	0.035	0.581
COTO	Family Stappiaceae	0.045	0.408
COTO	*Klenkia* spp.	0.045	0.408
EPFU	Order Bacillales	0.005	0.922
EPFU	*Clostridium_sensu_stricto_1* spp.	0.005	0.793
EPFU	*Bergeyella* spp.	0.005	0.739
EPFU	*Ferruginibacter* spp.	0.04	0.426
MYYU	*Brachybacterium* spp.	0.015	0.812
MYYU	Family Staphylococcaceae	0.01	0.782
MYYU/LU	*Candidatus_Blochmannia* spp.	0.005	0.655
MYYU/LU	Family Verrucomicrobiaceae	0.015	0.535
MYYU/LU	*Flectobacillus* spp.	0.02	0.508

Table 14: Alpha diversity results for roost location.

Alpha Diversity: Roost Location		
Alpha Diversity Metric	**p-value**	**H Test Statistic**
Shannon's Diversity	6.07×10^{-9}	54.288
Faith's Phylogenetic Diversity	0.01	19.975
Evenness	3.52×10^{-8}	50.337
Observed Features	0.000342	28.81

Table 15: Summary statistics for roost location alpha diversity metrics.

Alpha Diversity Summary Statistics: Chelan		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.271949	0.1465881
Faith's Phylogenetic Diversity	7.873514	0.3890182
Evenness	0.6512929	0.02109436
Observed Features	94.57143	7.299706
Alpha Diversity Summary Statistics: Klickitat		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.067424	0.1676
Faith's Phylogenetic Diversity	7.496521	0.4894639
Evenness	0.6379996	0.01922312
Observed Features	83.5	9.958246
Alpha Diversity Summary Statistics: Lewis		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.3066	0.2198123
Faith's Phylogenetic Diversity	8.565582	0.6232268
Evenness	0.6430973	0.029135
Observed Features	103.81818	6.353238
Alpha Diversity Summary Statistics: Lincoln		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.110884	0.229099
Faith's Phylogenetic Diversity	8.133537	0.59751
Evenness	0.6945929	0.02663268
Observed Features	103.81818	9.600189
Alpha Diversity Summary Statistics: Mason		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.110884	0.7464708
Faith's Phylogenetic Diversity	7.687145	0.79136
Evenness	0.6239174	0.09871209
Observed Features	95.27273	15.395395
Alpha Diversity Summary Statistics: Okanogan		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.13947	0.1264838
Faith's Phylogenetic Diversity	7.937782	0.5607918
Evenness	0.6332946	0.01461873
Observed Features	93.25	9.437766

Alpha Diversity Summary Statistics: Pend Oreille		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	3.93894	0.7140115
Faith's Phylogenetic Diversity	7.646854	0.846017
Evenness	0.6157307	0.10263367
Observed Features	84	14.205633
Alpha Diversity Summary Statistics: Spokane		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.695434	0.2293452
Faith's Phylogenetic Diversity	7.638878	0.8641198
Evenness	0.7127893	0.02401398
Observed Features	97.09091	15.082802
Alpha Diversity Summary Statistics: Thurston		
Diversity Metric	**Mean**	**Standard Deviation**
Shannon's Diversity	4.58988	0.2147029
Faith's Phylogenetic Diversity	7.741898	0.5788112
Evenness	0.6912365	0.03163564
Observed Features	100.3	10.285373

Table 16: Beta diversity analysis results for roost location.

Beta Diversity: Roost Location		
Beta Diversity Metric	**p-value**	**Pseudo-F Test Statistic**
Bray-Curtis Dissimilarity	0.001	3.323
Jaccard Similarity	0.001	3.0136
Unweighted UniFrac Distance	0.001	2.682
Weighted UniFrac Distance	0.001	3.793

Table 17: pairwise comparisons between different roost locations for Weighted UniFrac Distances.

Weighted UniFrac Pairwise Comparisons by Roost Location		
Species Comparison	**p-value**	**Pseudo-F Test Statistic**
Chelan vs. Klickitat	0.001	5.253
Chelan vs. Lincoln	0.007	5.874
Chelan vs. Okanogan	0.012	2.887
Chelan vs. Spokane	0.002	11.129
Chelan vs. Thurston	0.005	8.499
Klickitat vs. Lewis	0.001	3.124
Klickitat vs. Lincoln	0.001	10.724
Klickitat vs. Spokane	0.001	11.129
Klickitat vs. Thurston	0.001	8.499
Lewis vs. Lincoln	0.001	9.725
Lewis vs. Mason	0.013	1.916
Lewis vs. Pend Oreille	0.035	0.755
Lewis vs. Spokane	0.001	9.082
Lewis vs. Thurston	0.001	7.947
Lincoln and Mason	0.002	3.72
Lincoln and Okanogan	0.001	9.837
Lincoln and Pend Oreille	0.035	3.121
Lincoln and Spokane	0.001	4.765
Lincoln and Thurston	0.002	4.769
Mason and Spokane	0.001	4.23
Mason and Thurston	0.001	4.221
Okanogan and Spokane	0.001	11.476
Okanogan and Thurston	0.002	8.723
Pend Oreille and Spokane	0.009	3.597
Pend Oreille and Thurston	0.014	3.691

Table 18: Indicspecies analysis results for roost location.

Site	Taxon	p-value	Maximum Preference Value Test Statistic
Chelan	*Mycobacterium* spp.	0.02	0.535
Klickitat	*Clostridium_sensu_stricto_1* spp.	0.005	0.825
Klickitat	*Bergeyella* spp.	0.005	0.775
Klickitat	*Ferruginibacter* spp.	0.045	0.447
Lincoln	Class Actinobacteria	0.005	0.846
Lincoln	Order Corynebacteriales	0.005	0.82
Lincoln	Family Planococcaceae	0.005	0.813
Lincoln	*Ornithinimicrobium* spp.	0.005	0.626
Lincoln	*Sporosarcina* spp.	0.025	0.571
Lincoln	*Nakamurella* spp.	0.005	0.57
Lincoln	*Arthrobacter* spp.	0.03	0.516
Mason	*Rickettsiella* spp.	0.025	0.723
Mason	*Lysobacter* spp.	0.005	0.701
Mason	Class Alphaproteobacteria	0.005	0.674
Okanogan	*Flectobacillus* spp.	0.005	0.707
Okanogan	*Aphanizomenon_NIES81* spp.	0.015	0.5
Okanogan	*Rickettsia* spp.	0.04	0.42
Pend Oreille	*Candidatus_Blochmannia* spp.	0.01	0.522
Thurston	*Nosocomiicoccus* spp.	0.005	0.939
Thurston	*Atopostipes* spp.	0.005	0.701

Table 19: Alpha diversity results for body mass, within each species.

Alpha Diversity: Body Mass Within Species			
Species	**Alpha Diversity Metric**	**p-value**	**Spearman's Correlation Statistic**
MYYU	Shannon's Diversity	0.66	0.0782
	Faith's Phylogenetic Diversity	0.982	0.004
	Evenness	0.745	0.0579
	Observed Features	0.4672	0.129
MYLU	Shannon's Diversity	0.476	0.153
	Faith's Phylogenetic Diversity	0.0675	-0.379
	Evenness	0.462	0.158
	Observed Features	0.536	-0.133
MYYU/LU	Shannon's Diversity	0.0137	-0.8571
	Faith's Phylogenetic Diversity	0.4821	-0.3214
	Evenness	0.7017	0.1786
	Observed Features	0.0897	-0.6847
MYVO	Shannon's Diversity	0.432	0.325
	Faith's Phylogenetic Diversity	0.6474	-0.193
	Evenness	0.329	0.398
	Observed Features	0.9098	0.0482
EPFU	Shannon's Diversity	0.8729	0.0548
	Faith's Phylogenetic Diversity	0.6004	0.1781
	Evenness	0.9045	-0.0411
	Observed Features	0.6759	0.1425
COTO	Shannon's Diversity	0.7040	-0.2000
	Faith's Phylogenetic Diversity	0.6228	-0.2571
	Evenness	0.3287	-0.4857
	Observed Features	0.9565	0.0290

Table 20: Beta diversity results for body mass, within each species.

Beta Diversity: Body Mass Within Species			
Species	**Beta Diversity Metric**	**p-value**	**Spearman's Rho Test Statistic**
MYYU	Bray-Curtis Dissimilarity	0.814	0.0219
	Jaccard Similarity	0.206	0.100
	Unweighted UniFrac Distance	0.517	0.060175
	Weighted UniFrac Distance	0.899	0.013
MYLU	Bray-Curtis Dissimilarity	0.317	0.144927
	Jaccard Similarity	0.123	0.168648
	Unweighted UniFrac Distance	0.985	-0.002367
	Weighted UniFrac Distance	0.221	0.162787
MYYU/LU	Bray-Curtis Dissimilarity	0.332	-0.250975
	Jaccard Similarity	0.332	-0.250975
	Unweighted UniFrac Distance	0.568	0.161846
	Weighted UniFrac Distance	0.641	-0.085148
MYVO	Bray-Curtis Dissimilarity	0.901	-0.038737
	Jaccard Similarity	0.503	-0.174153
	Unweighted UniFrac Distance	0.152	-0.355448
	Weighted UniFrac Distance	0.515	-0.173604
EPFU	Bray-Curtis Dissimilarity	0.537	0.112556
	Jaccard Similarity	0.001	0.531138
	Unweighted UniFrac Distance	0.73	0.061852
	Weighted UniFrac Distance	0.876	0.029464
COTO	Bray-Curtis Dissimilarity	0.684	-0.127013
	Jaccard Similarity	0.759	-0.090421
	Unweighted UniFrac Distance	0.084	0.420394
	Weighted UniFrac Distance	0.307	-0.264759

Table 21: Alpha diversity results for wing damage.

Alpha Diversity: Wing Damage		
Alpha Diversity Metric	**p-value**	**Spearman's Correlation Statistic**
Shannon's Diversity	0.501	0.0708
Faith's Phylogenetic Diversity	0.2891	-0.113
Evenness	0.2246	0.1293
Observed Features	0.5792	-0.0592

Table 22: Beta diversity for wing damage.

Beta Diversity: Wing Damage		
Beta Diversity Metric	**p-value**	**Spearman's Rho Test Statistic**
Bray-Curtis Dissimilarity	0.815	0.022
Jaccard Similarity	0.984	0.001426
Unweighted UniFrac Distance	0.294	-0.074
Weighted UniFrac Distance	0.936	0.006995

Table 23: Alpha and beta diversity results for wing damage within *M. lucifugus.*

Alpha Diversity: Wing Damage for *M. lucifugus*		
Alpha Diversity Metric	**p-value**	**Spearman's Correlation Statistic**
Shannon's Diversity	0.326	0.209
Faith's Phylogenetic Diversity	0.704	-0.0819
Evenness	0.213	0.264
Observed Features	0.469	-0.155
Beta Diversity: Wing Damage for *M. lucifugus*		
Beta Diversity Metric	**p-value**	**Spearman's Rho Test Statistic**
Bray-Curtis Dissimilarity	0.645	0.087655
Jaccard Similarity	0.04	0.263778
Unweighted UniFrac Distance	0.955	-0.007639
Weighted UniFrac Distance	0.824	0.043502

Table 24: Metadata for the 12 culturing samples used in this study.

Mass Culture Plate Metadata				
Sample ID	**County**	**Ectoparasites**	**Number of Bacterial Isolates**	**Fungal Growth?**
RR L-14	Lincoln	Y	7	Some fungi
RR L-11	Lincoln	N	4	N
WB T-18	Thurston	Y	1	N
WB T-5	Thurston	N	2	N
POM PO-5	Pend Oreille	Y	2	Some fungi
POM PO-20	Pend Oreille	N	0	Only fungi
LC C-1	Chelan	N	1	N
TL KL-1	Klickitat	N	0	Only fungi
OR LS-12	Lewis	Y	0	One fungus
OR LS-17	Lewis	N	1	One fungus
OR LS-7	Lewis	Y	1	One fungus
OR LS-18	Lewis	N	1	One fungus

Table 25: Bacterial isolates identified from the culturing subset. Sample OR LS-18_A has two potential identifications, as the sequences were not high-quality enough to distinguish the isolates to species.

Sample ID	Species ID	Max Score	Total Score	Query Cover	Percent ID	Accession Number	Relative Abundance in Sequenced Microbiome
LC C-1_A	*Pseudarthrobacter equi*	2531	2531	100%	99.43%	NR_117032.1	1.838%
POM PO-5_A	*Rhodococcus corynebacterioides*	2512	2512	99%	99.28%	NR_119107.1	0.0306%
POM PO-5_C	*Rhodococcus tukisamuensis*	1125	1125	99%	99.04%	NR_028629.1	0.179%
WB T-5_A	*Neobacillus cucumis*	255	255	100%	96.13%	NR_148626.1	0.0126%
WB T-5_B	*Arthrobacter silviterrae*	2307	2307	98%	96.61%	NR_159109.1	0.007%
WB T-18_A	*Subtercola boreus*	2340	2340	99%	97.05%	NR_115024.1	0.0348%
OR LS-7_A	*Microbacterium pumilum*	2551	2551	99%	99.78%	NR_041331.1	0.003%
OR LS-17_A	*Methylobacterium cerastii*	2425	2425	99%	99.11%	NR_117118.1	0.0018%
OR LS-18_A*	*Kocuria arsenatis*	1286	1286	100%	100.00%	NR_148610.1	0.0498%
OR LS-18_A*	*Kocuria rhizophila*	1286	1286	100%	100.00%	NR_026452.1	0.0498%
RR L-11_A	*Priestia megaterium*	2623	2623	99%	99.93%	NR_112636.1	0.008%
RR L-11_B	*Rhodococcus oryzae*	2444	2444	99%	98.55%	NR_170410.1	1.231%
RR L-11_C	*Pseudarthrobacter phenanthrenivorans*	2459	2459	99%	98.77%	NR_074770.2	1.838%

Sample ID	Species ID	Max Score	Total Score	Query Cover	Percent ID	Accession Number	Relative Abundance in Sequenced Microbiome
RR L-11_D	*Rhodococcus oryzae*	1144	1144	100%	98.17%	NR_170410.1	1.231%
RR L-14_A	*[Brevibacterium] frigoritolerans*	2601	2601	100%	99.51%	NR_115064.1	0.0126%
RR L-14_B	*Paenibacillus macquariensis subsp. defensor*	2621	2621	99%	99.72%	NR_041635.1	0.0162%
RR L-14_C	*Paenibacillus macquariensis subsp. defensor*	1205	1205	100%	99.85%	NR_041635.1	0.0162%
RR L-14_D	*Sporosarcina psychrophila*	2625	2625	99%	99.93%	NR_113752.1	0.0498%
RR L-14_E	*Sporosarcina psychrophila*	2617	2617	100%	99.86%	NR_113752.1	0.0498%
RR L-14_F	*Paenisporosarcina indica*	2556	2556	99%	99.02%	NR_108473.1	0.005%
RR L-14_G	*Streptomyces laculatispora*	2531	2531	99%	99.57%	NR_117082.1	0.004%

FIGURES

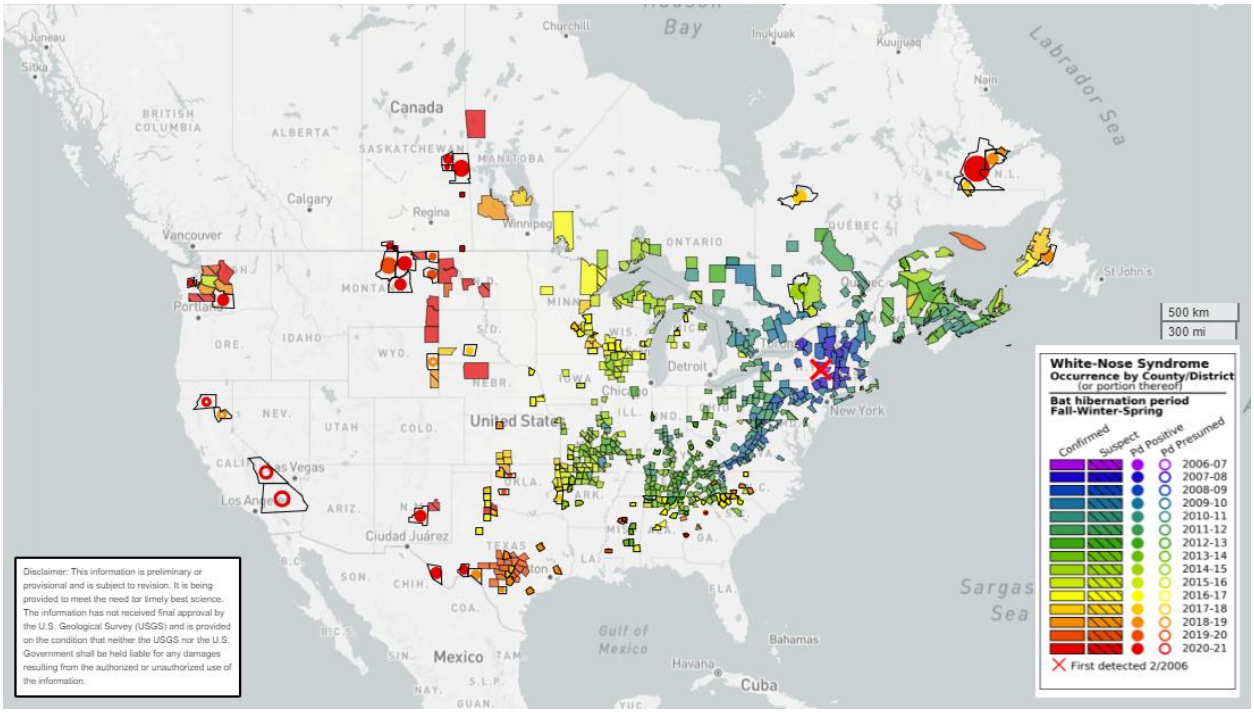


Fig. 1: WNS spread in North America since its initial introduction in Albany, New York in the winter of 2006 (site with an "X"). Note the recent and growing introductions into the western United States, with the first western detection in western Washington in 2016 [60]. From https://www.whitenosesyndrome.org/ [189].

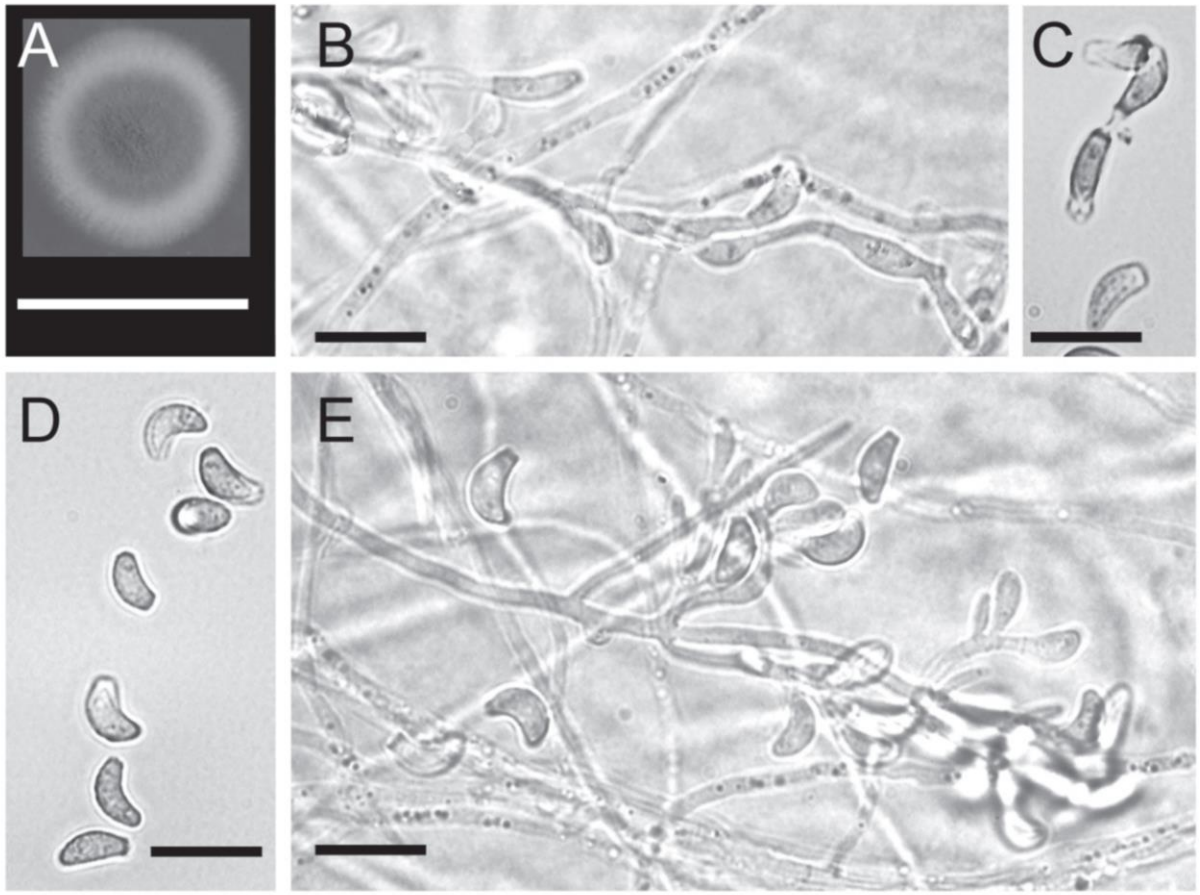


Fig. 2: Morphology of Pd colonies on media (A), Pd hyphae (B) and distinctive crescent-shaped Pd conidia (C – E). From Gargas et al. [61].

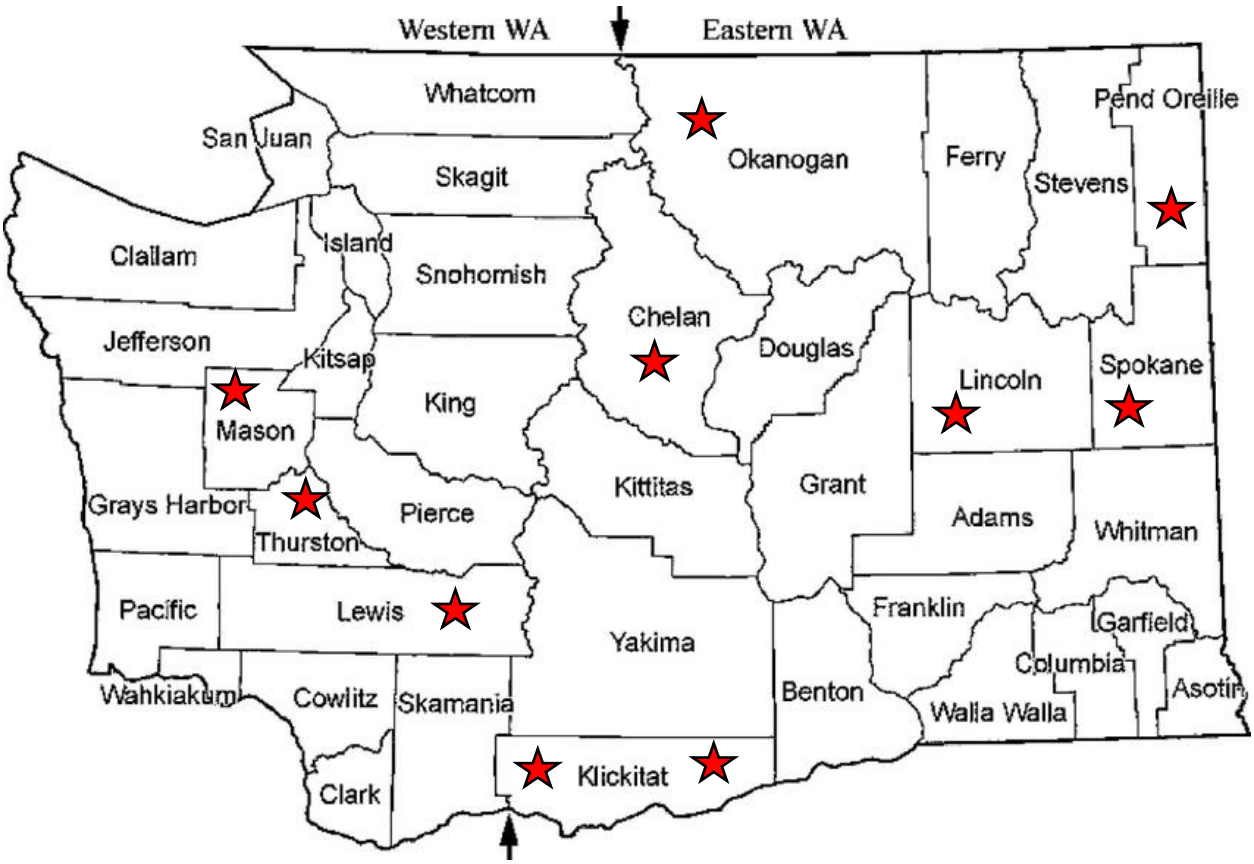


Fig. 3: Sampling sites across Washington state. Stars indicate which counties were sampled, but not exact roost locations. Blank map from Sames et al. [190].

Fig. 4: Visualizing Pd growth in the field using a UV light. Pd growth is indicated by the glowing patches on the wing membrane. As Pd hyphae penetrate the wing tissues, it creates long-lasting wing scarring, resulting in varying degrees of wing damage depending on infection severity. This compromises flight ability and foraging efficiency [54]. Photograph from Dana Colley.

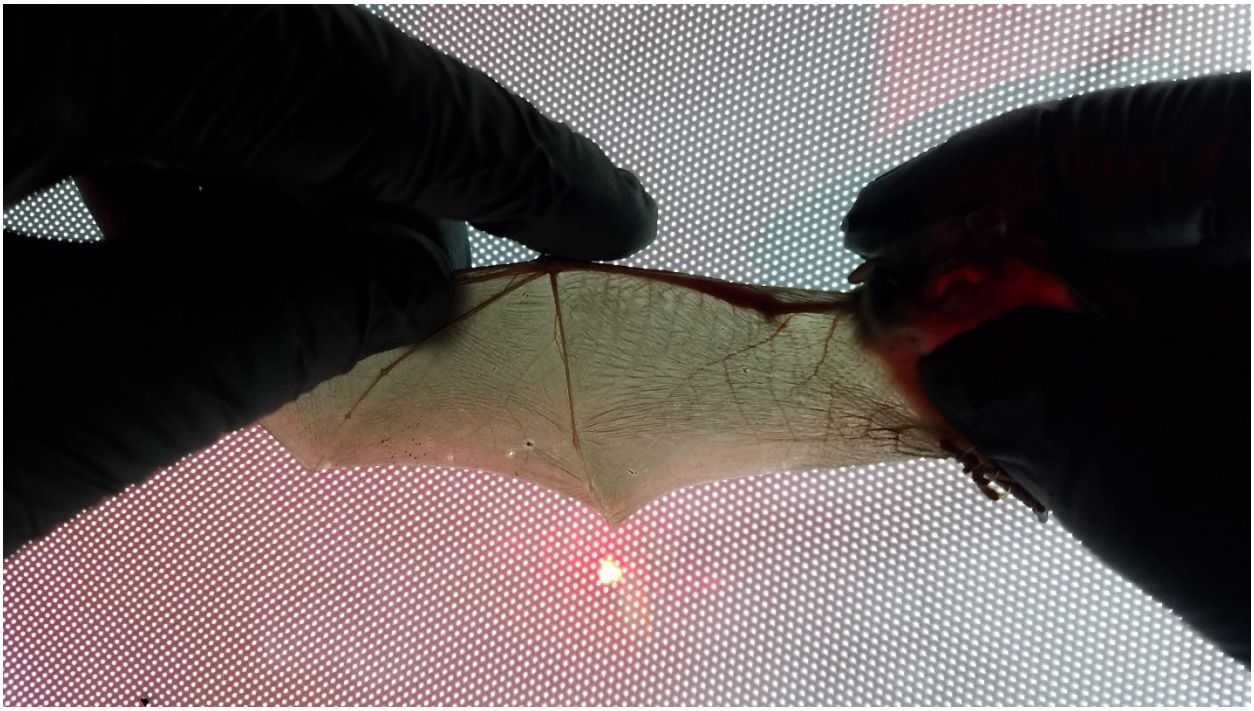


Fig. 5: Lightbox used to examine wing membranes for damage and assess age. Note the pinprick holes, minimal scarring, and the fully ossified epiphyseal joints, indicative of an adult bat [191] with a wing damage index of zero [81]. Photograph from Krisztian Magori.

Fig. 6: Swabbing sites on sampled bats. Bats were swabbed on the forearm five times and the muzzle five times. This was repeated for the other swab using the other forearm. Photograph from Dana Colley.

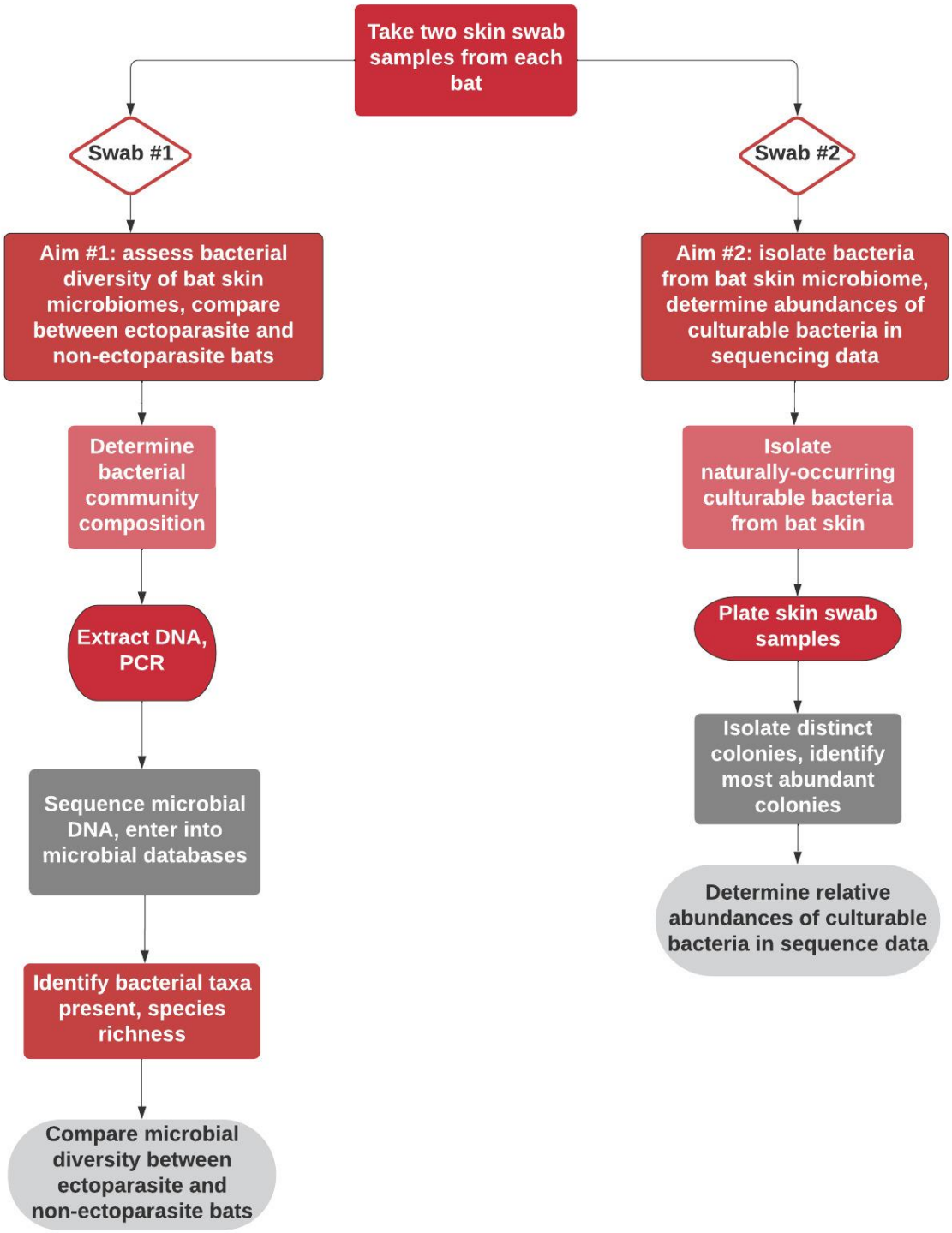


Fig. 7: methodological workflow of this study.

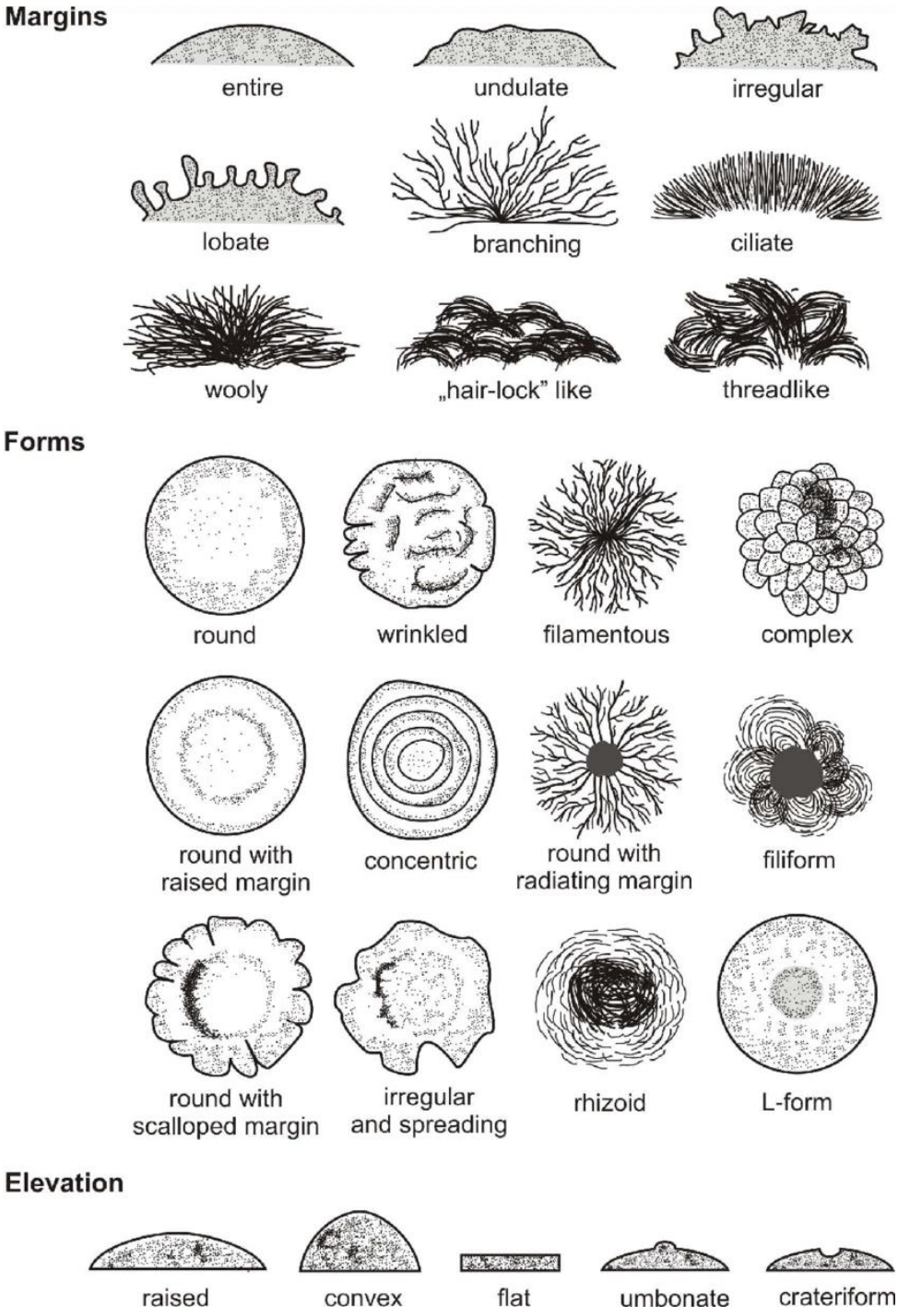
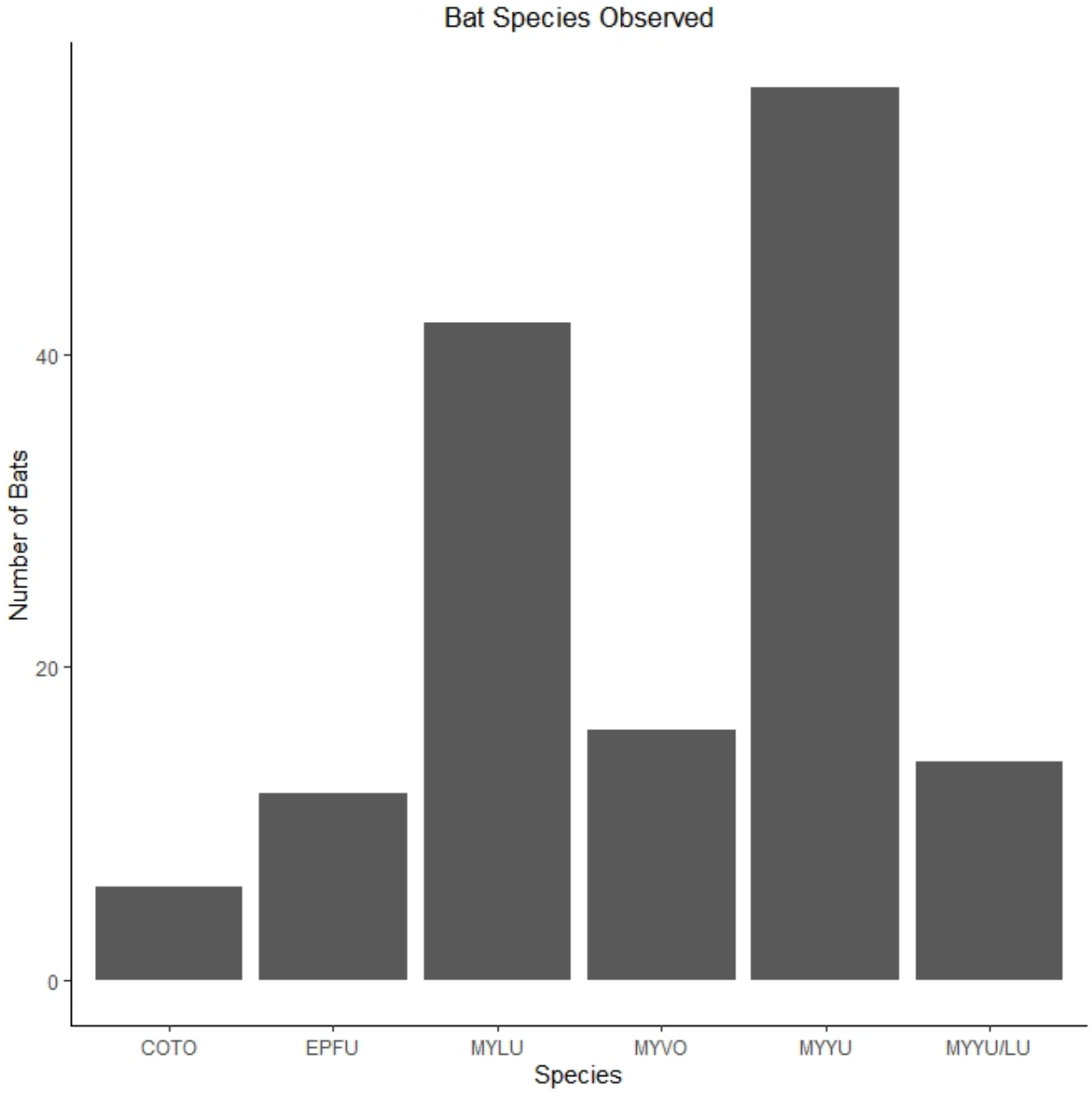
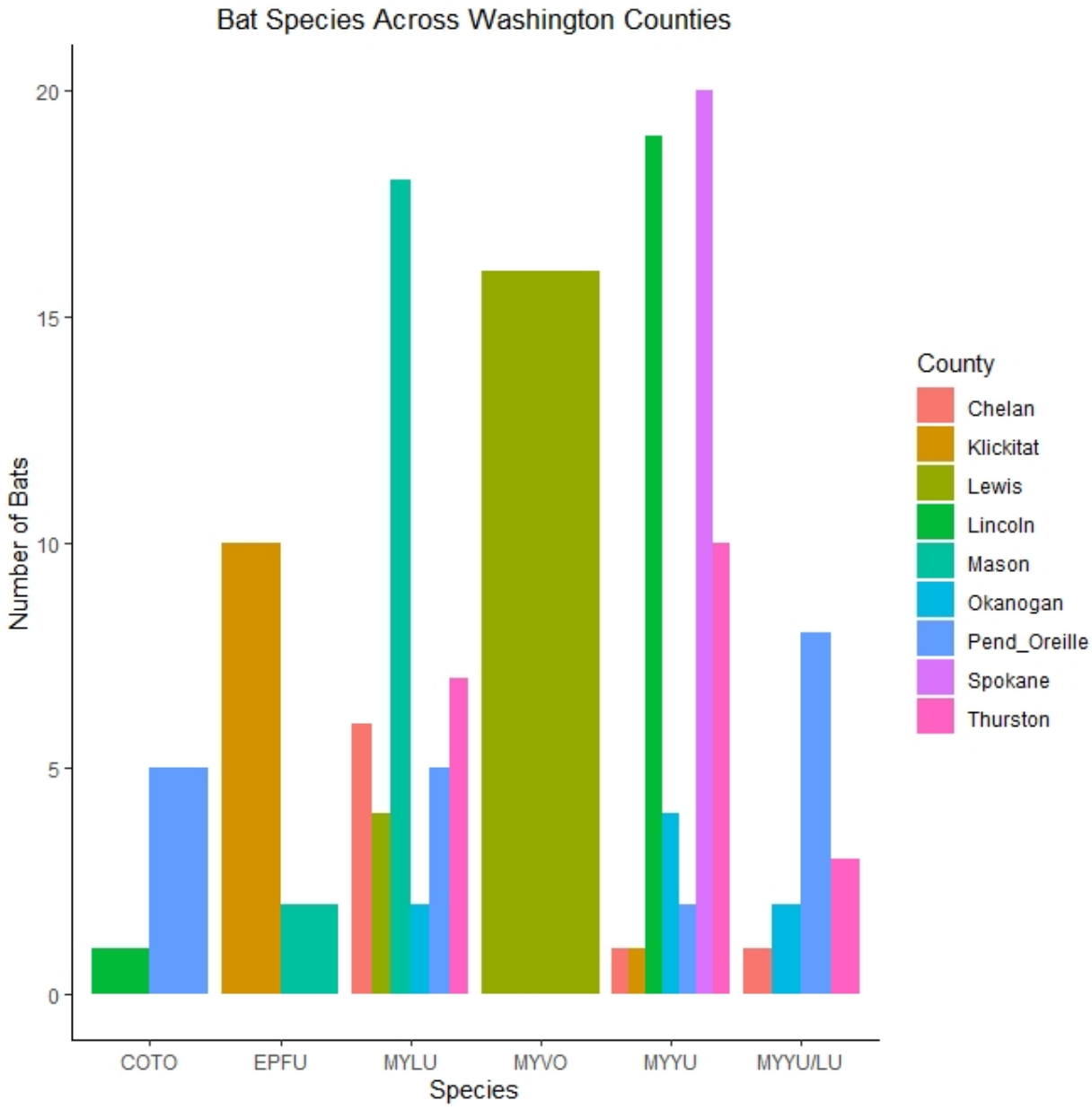


Fig. 8: Examples of bacterial colony morphologies. From Kolwzan et al. [114]



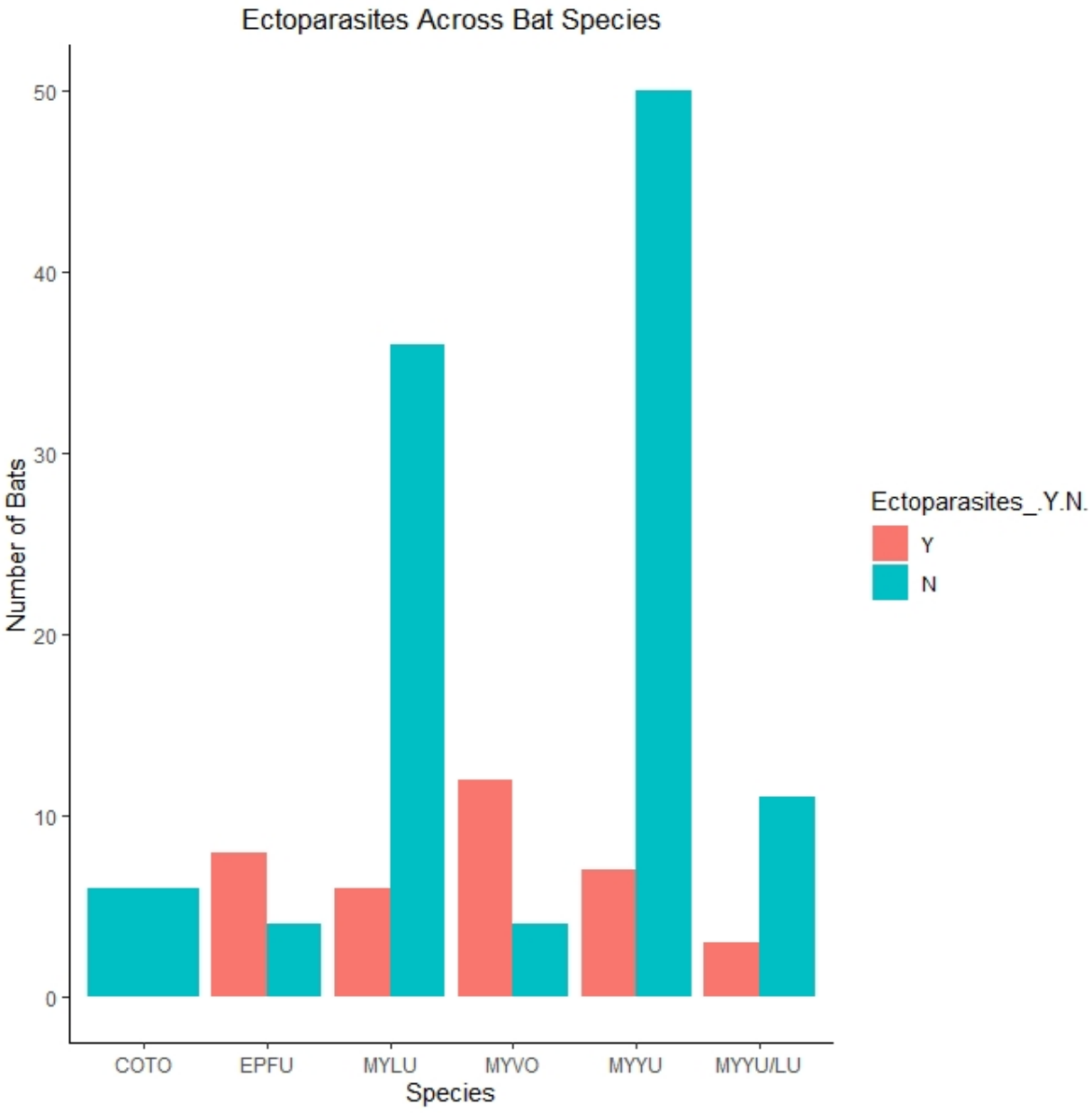
COTO: *Corynorhinus townsendii*
EPFU: *Eptesicus fuscus*
MYLU: *Myotis lucifugus*
MYVO: *Myotis volans*
MYYU: *Myotis yumanensis*
MYYU/LU: *Myotis* spp. combined group

Fig. 9: Number of bats sampled in this study for each species. *Myotis yumanensis* and *M. lucifugus* were the most abundant species sampled in this study, while *C. townsendii* was the least abundant species sampled.



COTO: *Corynorhinus townsendii*
EPFU: *Eptesicus fuscus*
MYLU: *Myotis lucifugus*
MYVO: *Myotis volans*
MYYU: *Myotis yumanensis*
MYYU/LU: *Myotis* spp. combined group

Fig. 10: Number of bats species sampled in this study in each county. *Myotis yumanensis*, *M. lucifugus*, and the *Myotis* spp. combined group were the most common across Washington State, being sampled in almost every county.

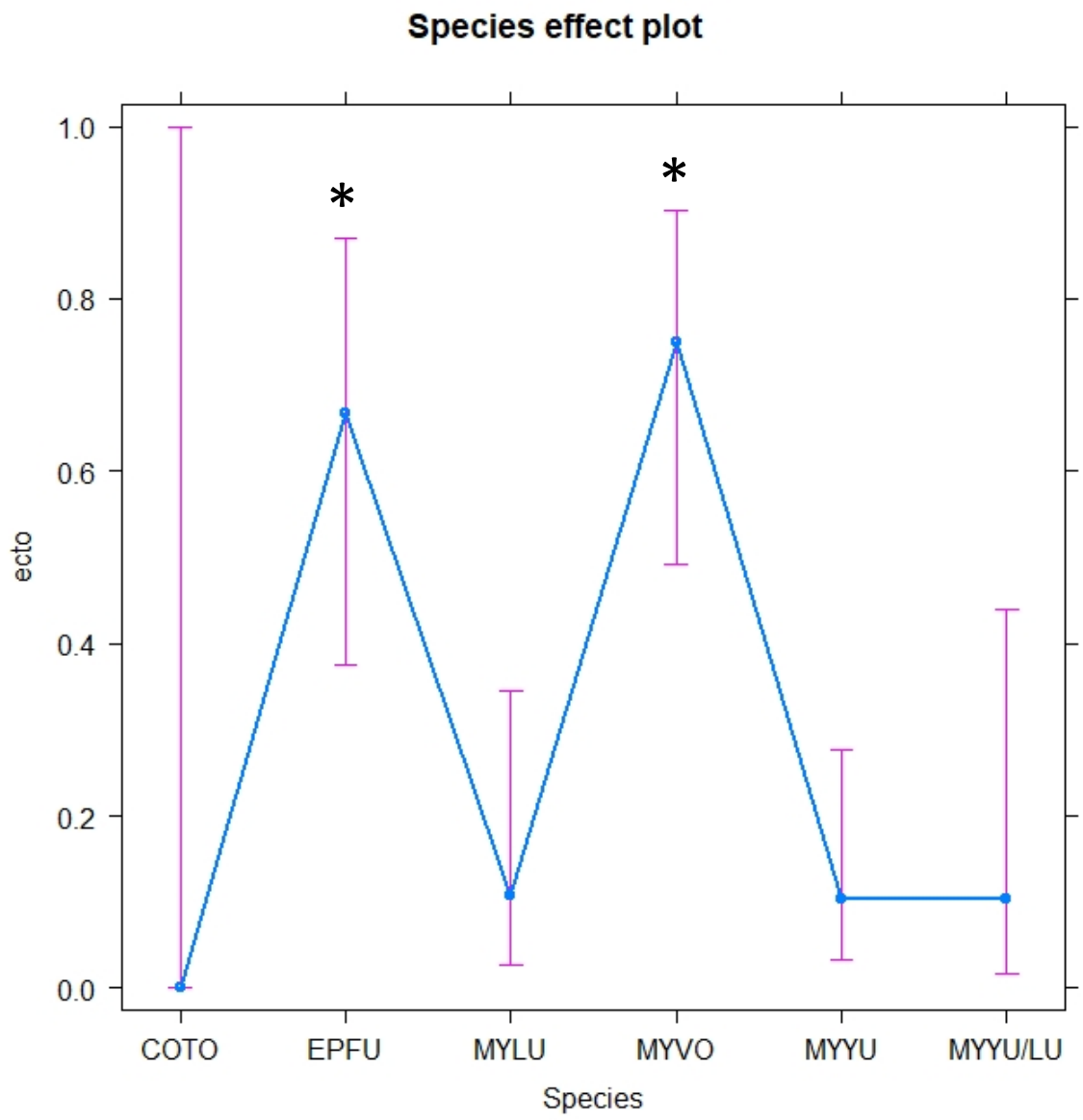


COTO: *Corynorhinus townsendii*
EPFU: *Eptesicus fuscus*
MYLU: *Myotis lucifugus*
MYVO: *Myotis volans*
MYYU: *Myotis yumanensis*
MYYU/LU: *Myotis* spp. combined group

Fig. 11: Counts of individuals with ectoparasites across the sampled bat species. *Myotis volans* and *E. fuscus* had the greatest incidences of ectoparasite presence. *Corynorhinus townsendii* was the only species that had zero individuals with ectoparasites.

Fig. 12: Bat ectoparasites observed in this study. A) adult *Spinturnix* bat mite, B) Hippoboscidae bat fly, C) *Cimex* bat bug, D) flea. *Spinturnix* bat mites were the most abundant bat ectoparasites observed. All photographs from Dana Colley, except for the bat fly photograph (B), from Fettig et al. [192].

Fig. 13: Observing *Spinturnix* bat mites on bats in the field. A) and B) are examples of *Spinturnix* mite eggs attached to the ears of *Myotis* spp., while C) is an example of adult *Spinturnix* mites attached to the wing membranes of an *E. fuscus* female. All photographs from Dana Colley.



COTO: *Corynorhinus townsendii*
EPFU: *Eptesicus fuscus*
MYLU: *Myotis lucifugus*
MYVO: *Myotis volans*
MYYU: *Myotis yumanensis*
MYYU/LU: *Myotis* spp. combined group

Fig. 14: Effects plot of ectoparasite presence as a function of bat species. *Eptesicus fuscus* was significantly more likely to have ectoparasites compared to *M. yumanensis* (χ^2= 26.101; p = 0.0125) and *M. lucifugus* (χ^2= 26.101; p = 0.0429), and *M. volans* was significantly more likely to have ectoparasites compared to *M. yumanensis* (χ^2= 26.101; p = 0.0015), *M. lucifugus* (χ^2= 26.101; p = 0.0094), and the *Myotis* spp. combined group (χ^2= 26.101; p = 0.0479).

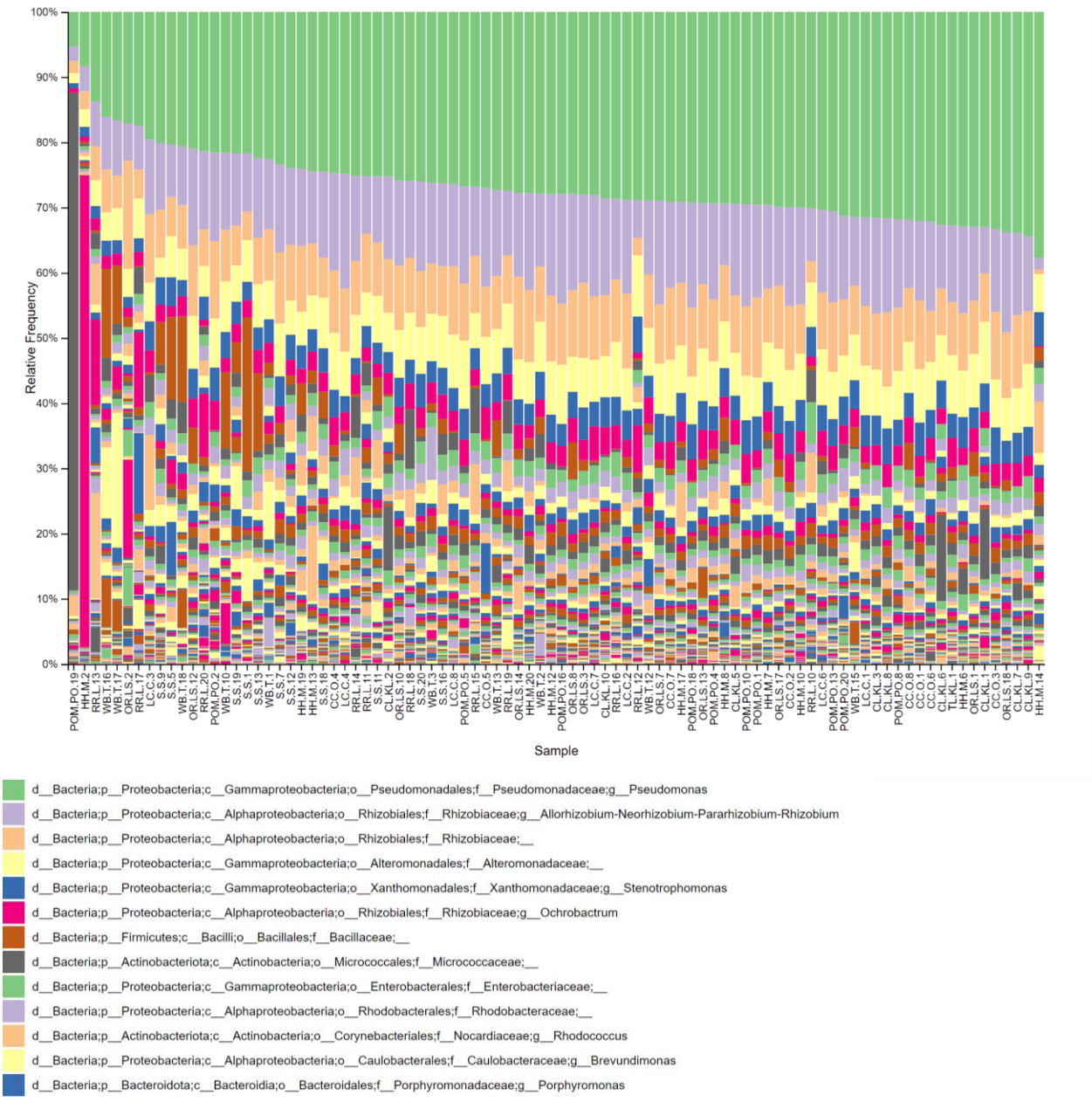


Fig. 15: Bacterial taxa bar plot across all 90 sequenced bat samples. Across all bat samples, *Pseudomonas* spp. was the most abundant taxon in the sampled bat microbiomes, comprising an average of 27% of each bat's skin microbiome. The next most abundant taxon was *Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium* spp., comprising on average 12% of the bat microbiome, followed by the family Rhizobiaceae, comprising on average 8% of the bat microbiome.

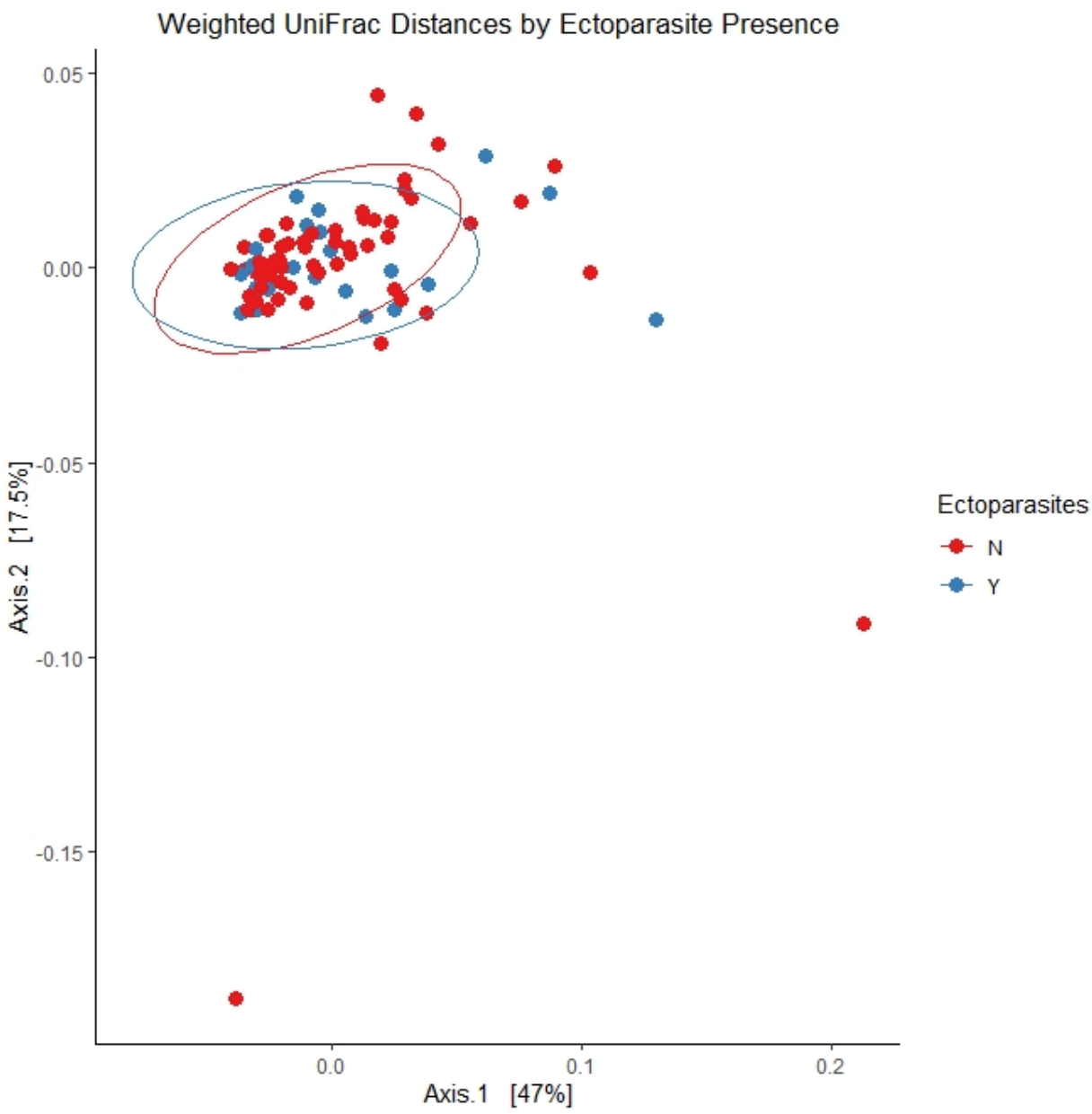


Fig. 16: Ordination plot of weighted UniFrac distances by ectoparasite presence (yes/no). Bat ectoparasites did not influence the skin microbiome diversity of bats (p = 0.965, pseudo-F = 0.306). Ellipses indicate the 95% confidence interval and are calculated in the R package ggord. Note the overlap of points and ellipses, indicative that the skin microbiome diversities and compositions between bats with and without ectoparasites are essentially the same.

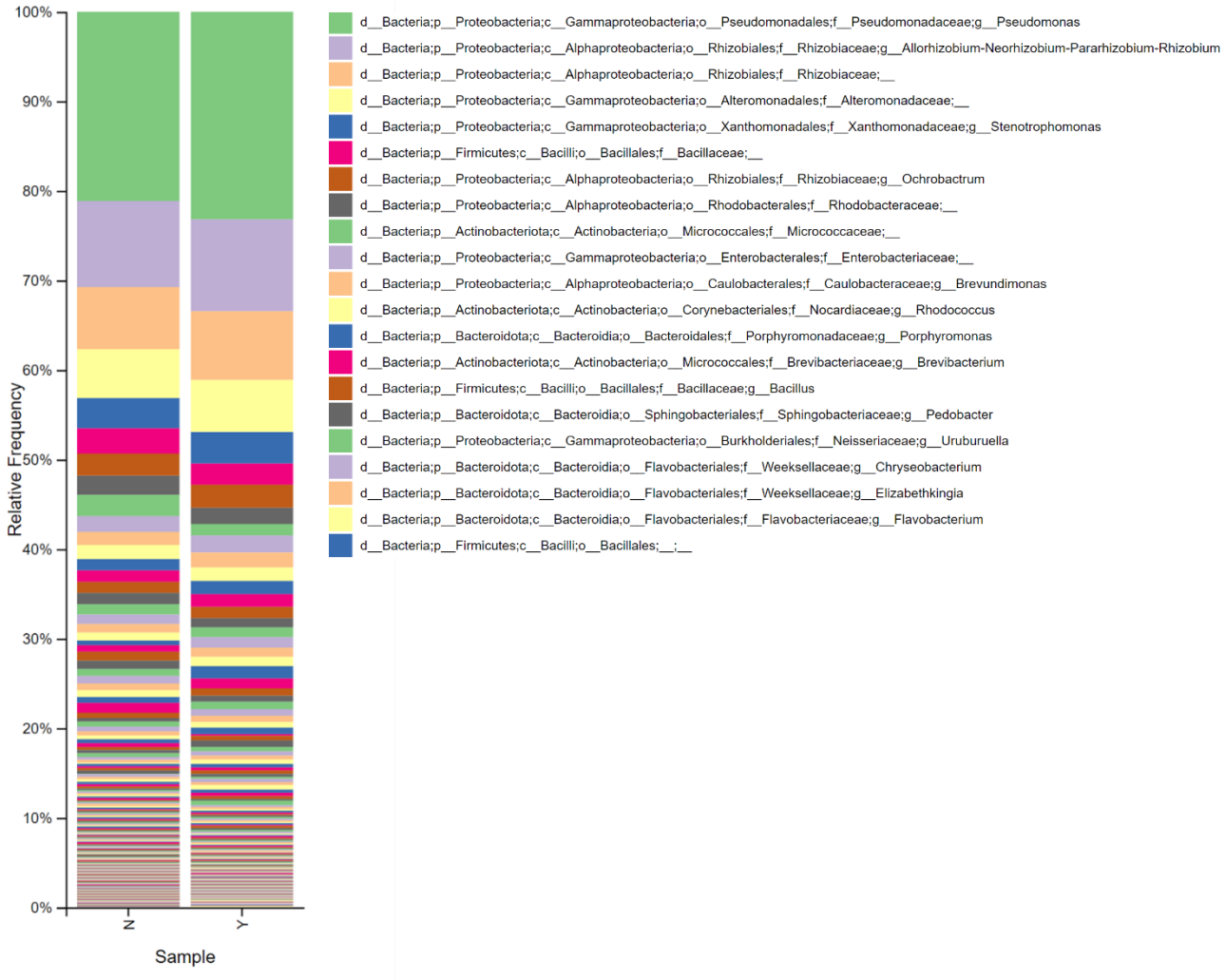


Fig. 17: Bacterial taxa bar plots illustrating relative abundances of bacterial taxa, grouped by ectoparasite presence status (yes or no) and averaged across all samples.

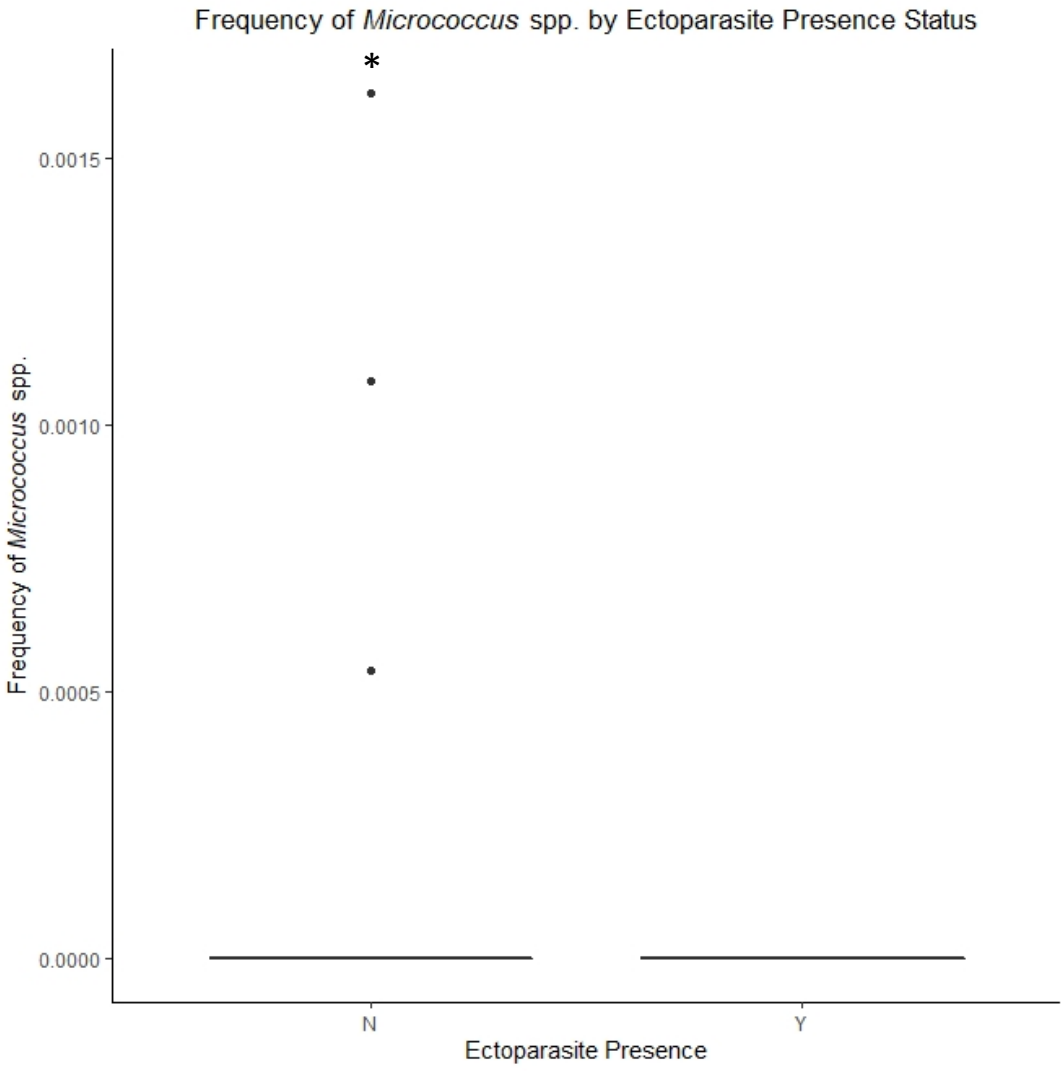


Fig. 18: Frequency of *Micrococcus* spp. in bats with and without ectoparasites. Bats without ectoparasites had significantly greater abundances of *Micrococcus* spp. than bats with ectoparasites (p = 0.045, maximum preference value test statistic = 0.395).

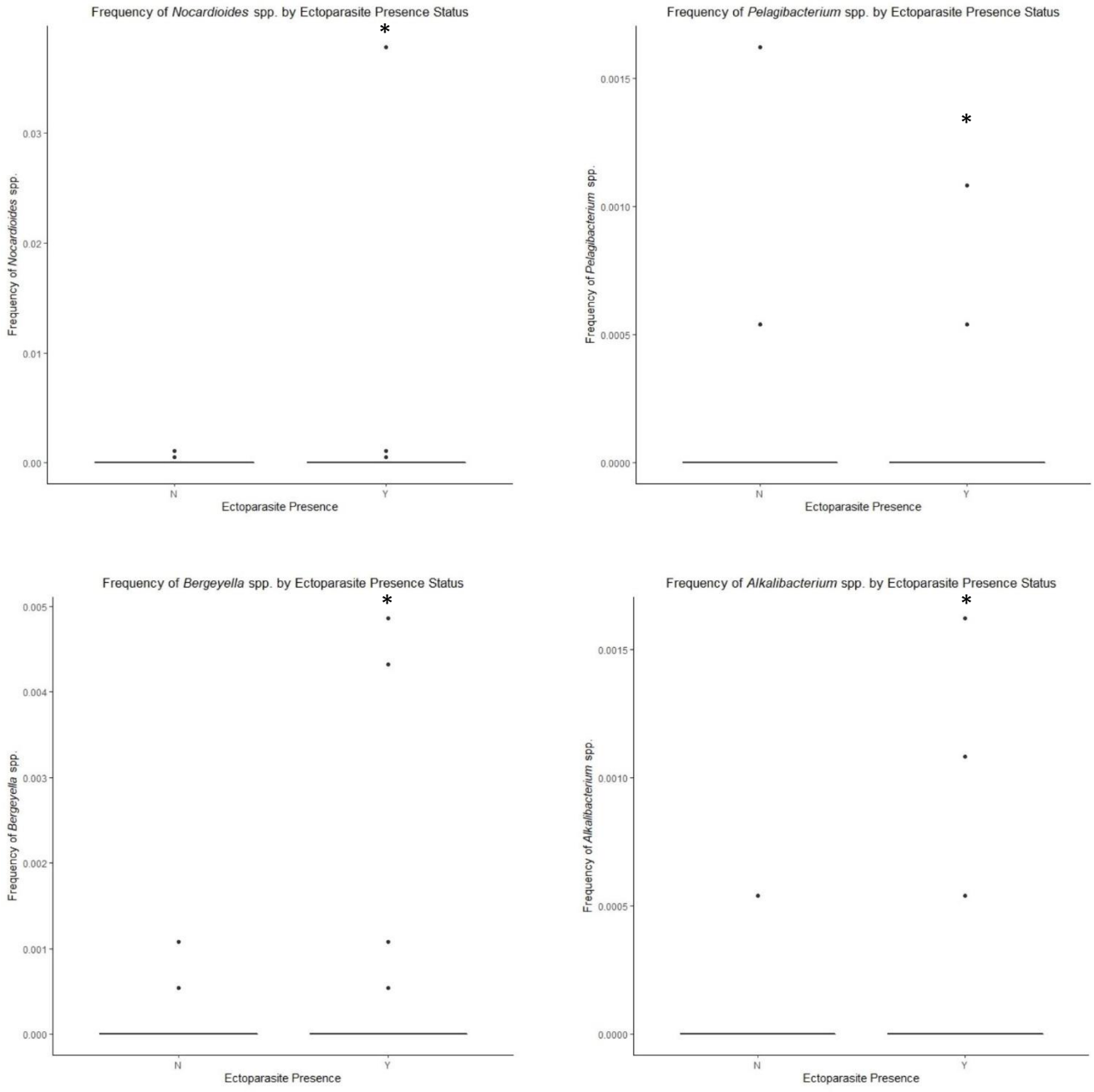


Fig. 19: Bats with ectoparasites had significantly greater abundances of *Nocardioides* spp. (p = 0.04, maximum preference value test statistic = 0.432), *Pelagibacterium* spp. (p = 0.02, maximum preference value test statistic = 0.4), *Bergeyella* spp. (p = 0.02, maximum preference value test statistic = 0.4), and *Alkalibacterium* spp. (p = 0.04, maximum preference value test statistic = 0.329).

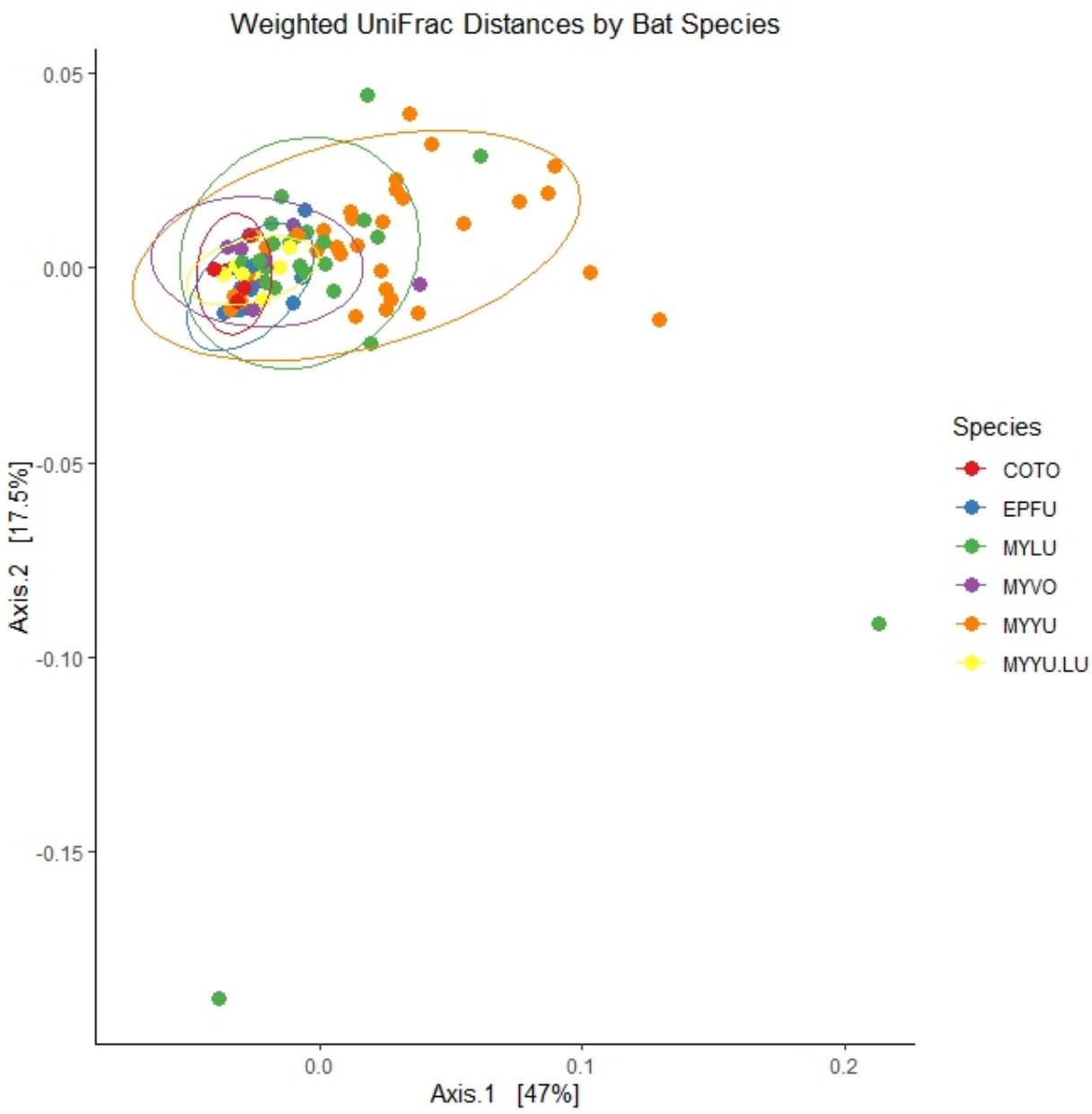


Fig. 20: Ordination plot of weighted UniFrac distances by bat species. Bat species significantly influenced the skin microbiome diversity and composition of bats (p = 0.006, pseudo-F statistic = 2.527). *Myotis yumanensis* had the highest skin microbiome diversity, while *E. fuscus* had the lowest skin microbiome diversity. *Myotis yumanensis* and *M. lucifugus* had skin microbiome compositions and diversities that were the most similar to each other. Ellipses indicate the 95% confidence interval and are calculated in the R package ggord.

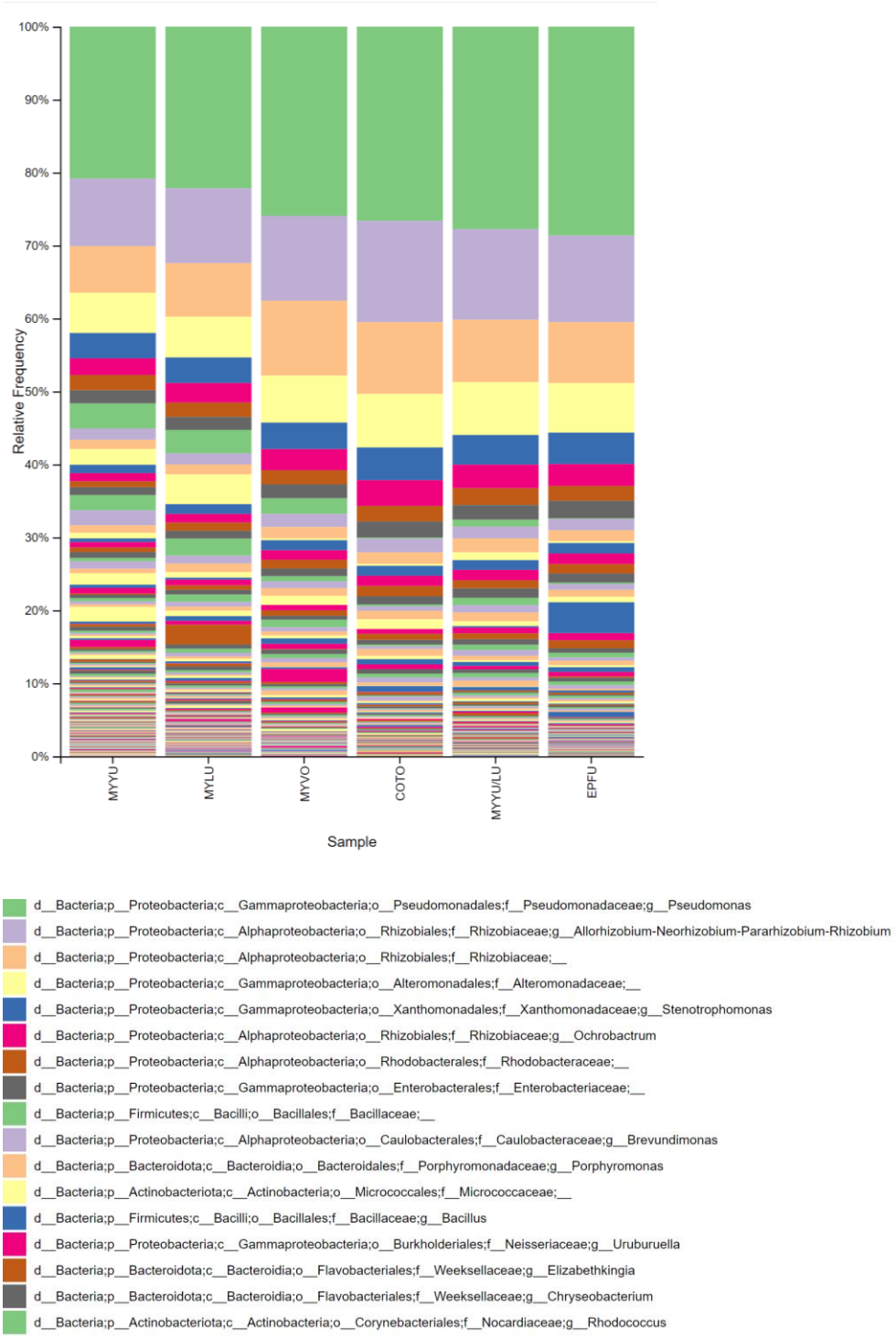


Fig. 21: Bacterial taxa bar plots illustrating relative abundances of bacterial taxa, grouped by bat species and averaged across all samples.

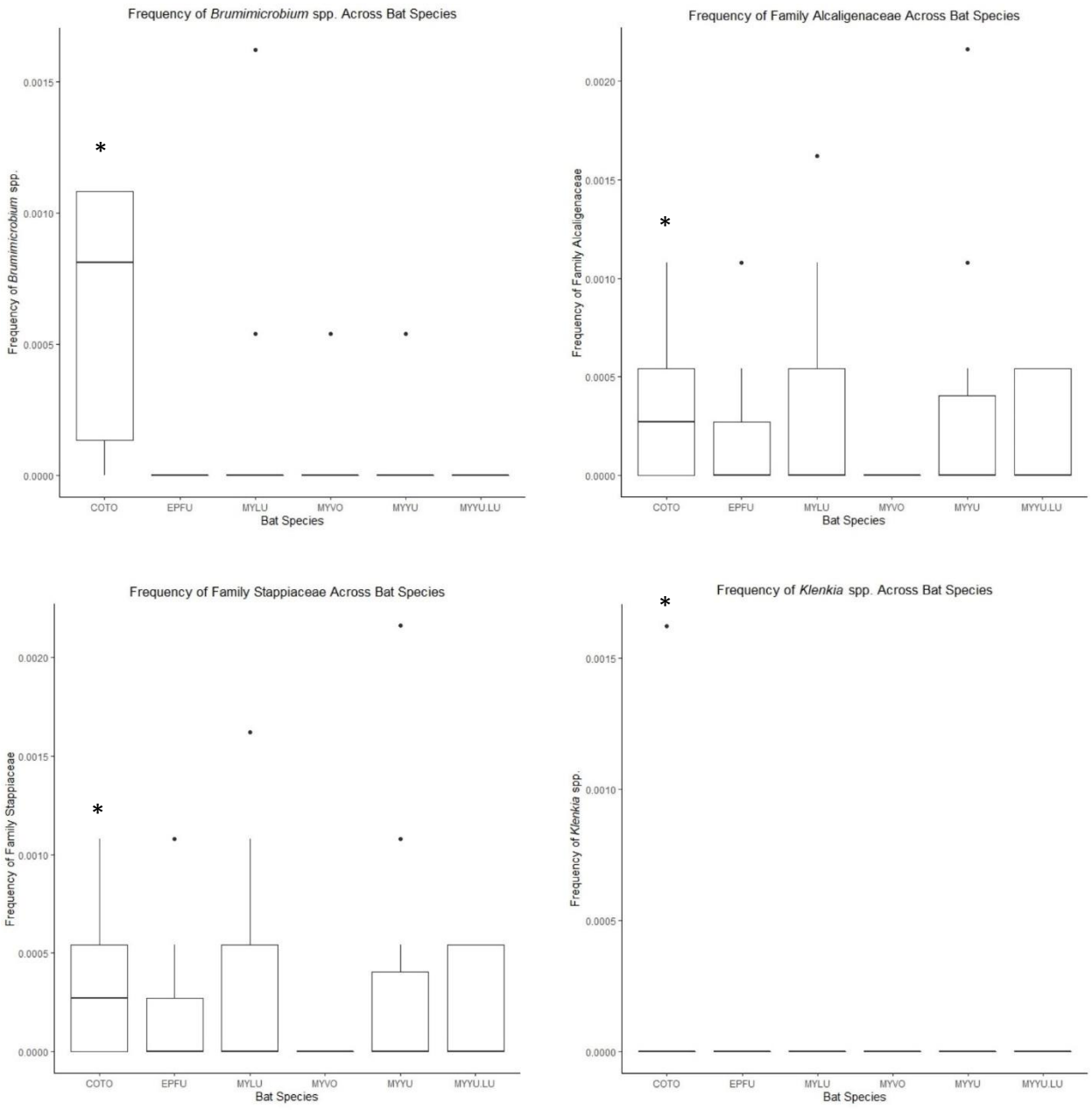


Fig. 22: *Corynorhinus townsendii* had significantly greater abundances of *Brumimicrobium* spp. (p = 0.005, maximum preference value test statistic = 0.697), family Alcaligenaceae (p = 0.035, maximum preference value test statistic = 0.581), family Stappiaceae (p = 0.045, maximum preference value test statistic = 0.408), and *Klenkia* spp. (p = 0.045, maximum preference value test statistic = 0.408).

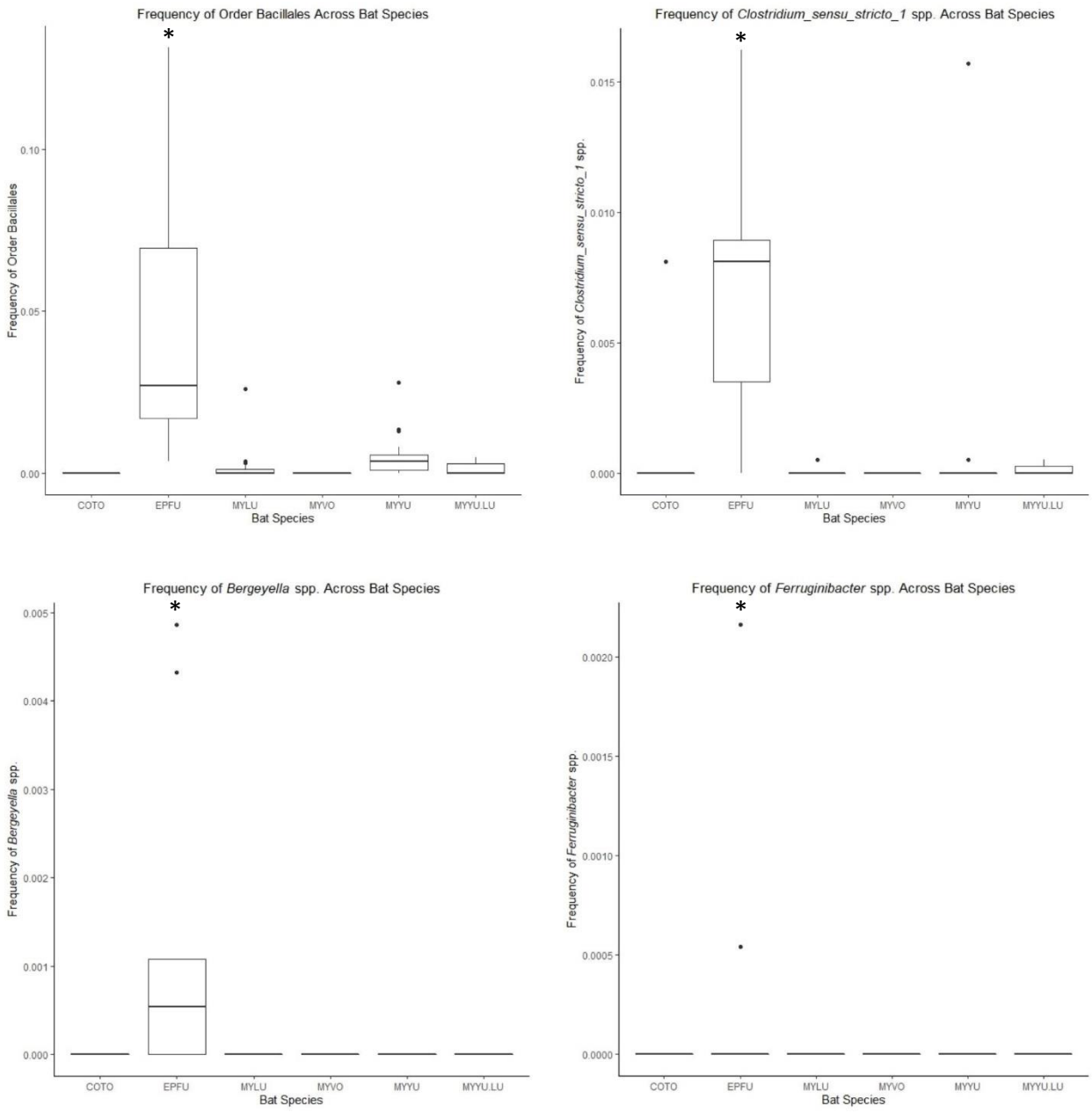


Fig. 23: *Eptesicus fuscus* had significantly greater abundances of the order Bacillales (p = 0.005, maximum preference value test statistic = 0.922), *Clostridium_sensu_stricto_1* spp. (p = 0.005, maximum preference value test statistic = 0.793), *Bergeyella* spp. (p = 0.005, maximum preference value test statistic = 0.739), and *Ferruginibacter* spp. (p = 0.04, maximum preference value test statistic = 0.426).

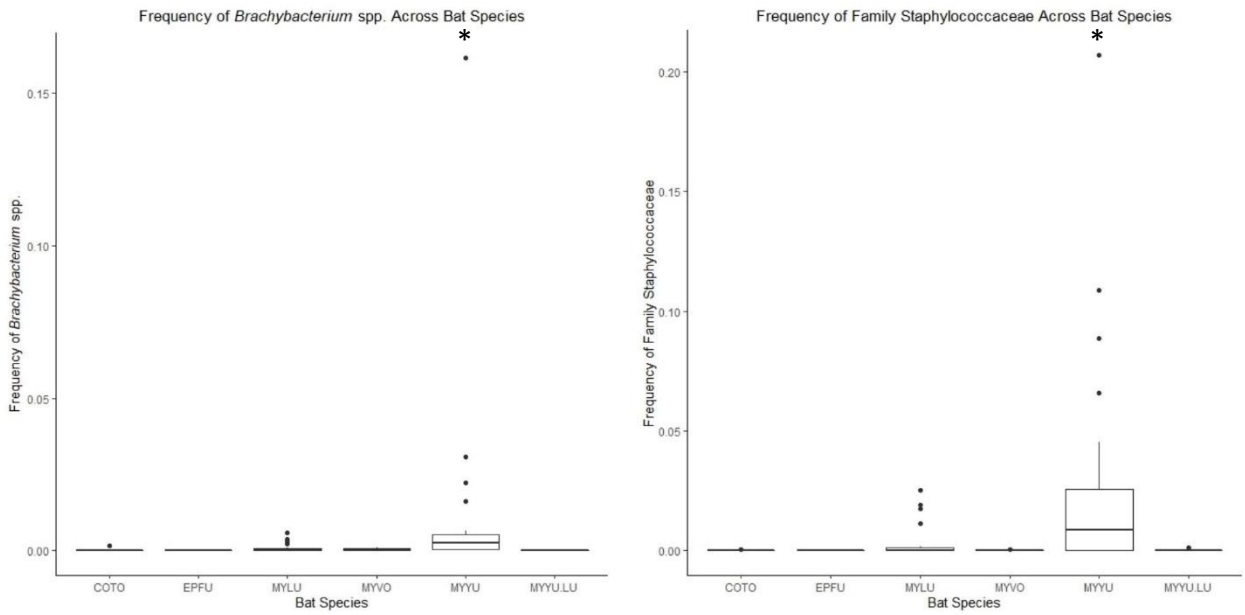


Fig. 24: *Myotis yumanensis* had significantly greater abundances of *Brachybacterium* spp. ($p = 0.015$, maximum preference value test statistic = 0.812) and the family Staphylococcaceae ($p = 0.01$, maximum preference value test statistic = 0.782).

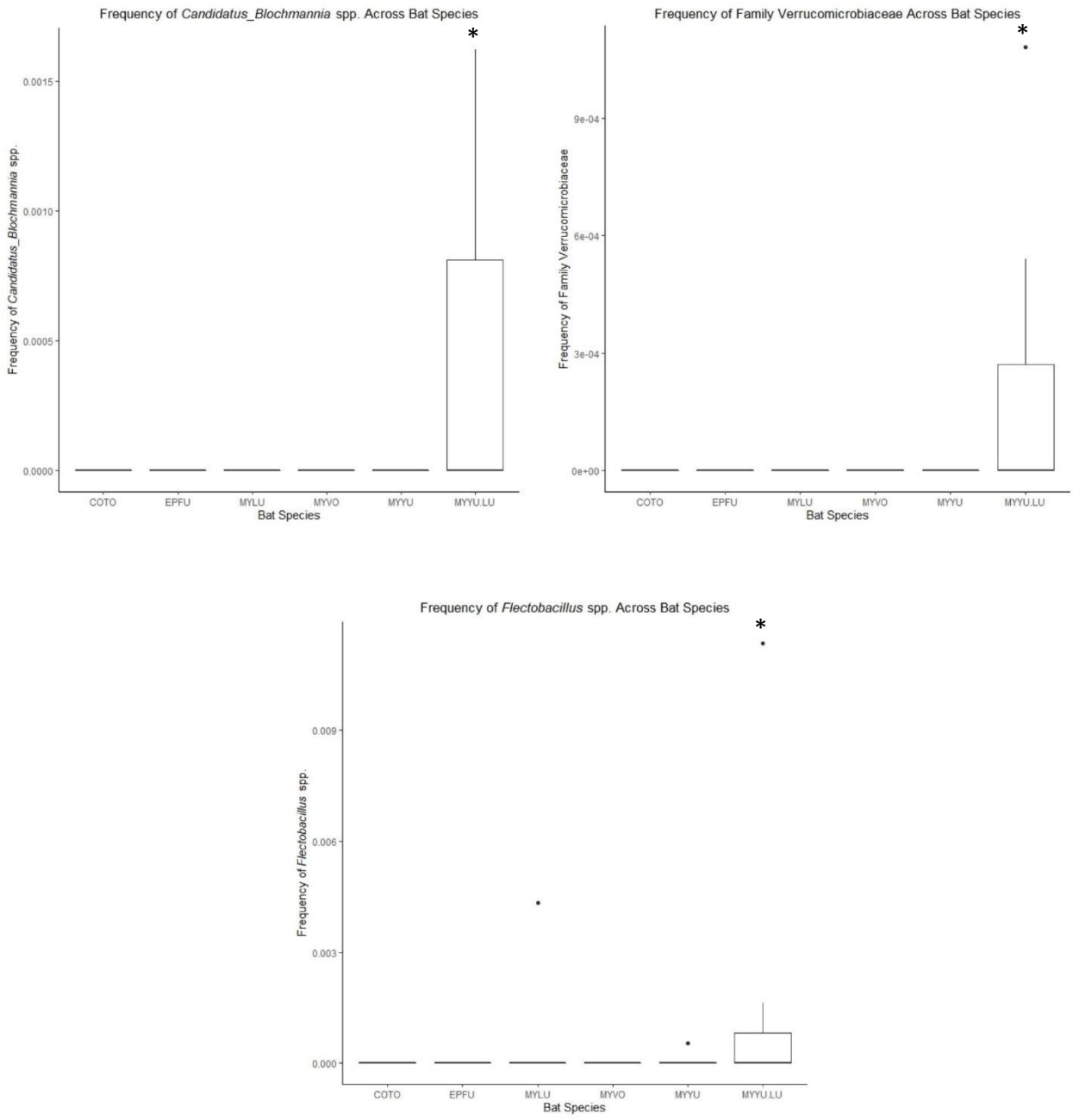


Fig. 25: The *Myotis* spp. combined group had significantly greater abundances of *Candidatus_Blochmannia* spp. (p = 0.005, maximum preference value test statistic = 0.655), family Verrucomicrobiaceae (p = 0.015, maximum preference value test statistic = 0.535), and *Flectobacillus* spp. (p = 0.02, maximum preference value test statistic = 0.508).

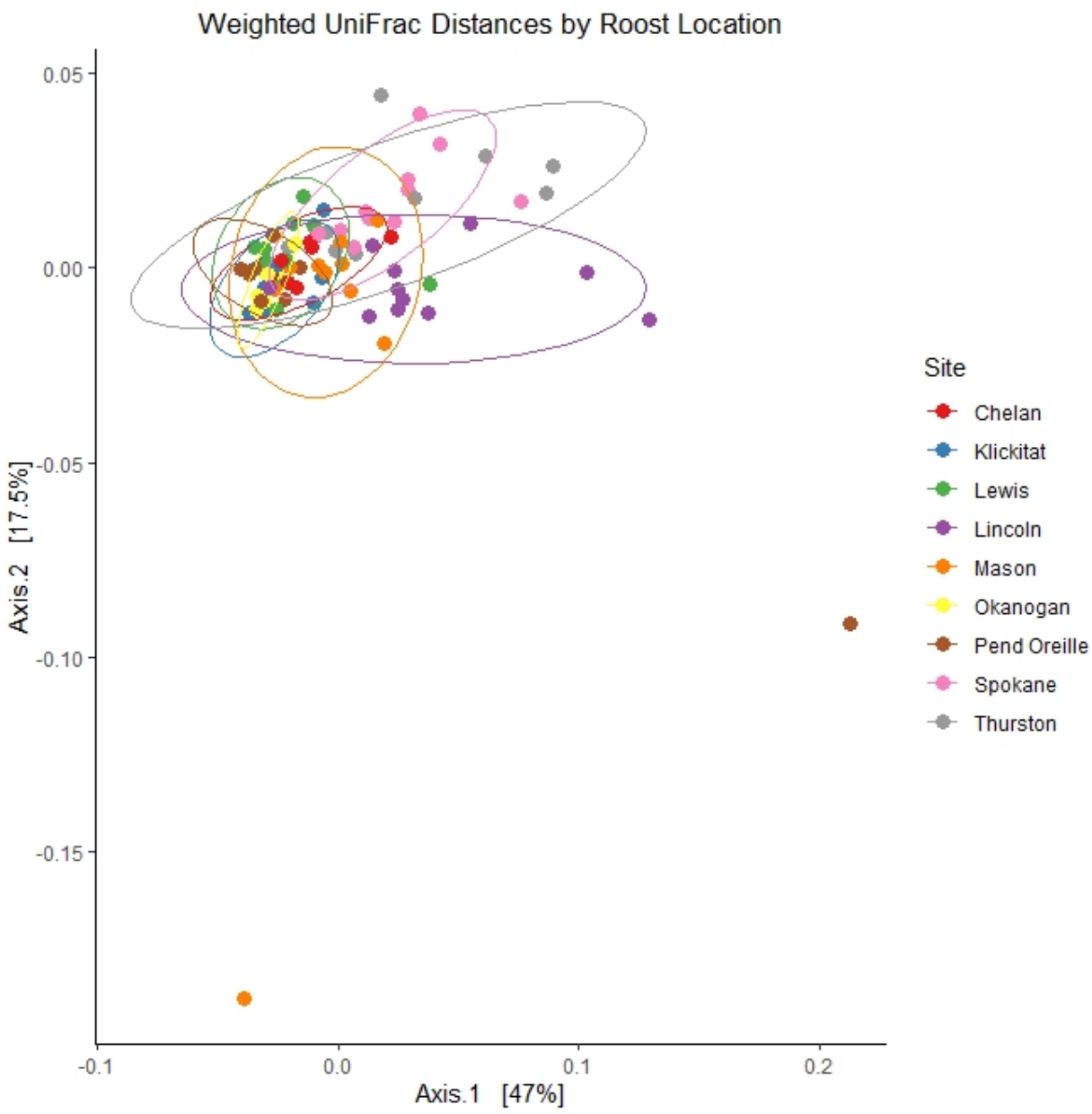


Fig. 26: Ordination plot of weighted UniFrac distances by roost location. Roost location significantly influenced the skin microbiome diversity and composition of bats ($p = 0.001$, pseudo-F statistic = 3.793). Lincoln County had the highest bat skin microbiome diversity, while Okanogan and Klickitat Counties had the lowest bat skin microbiome diversity. Spokane and Thurston Counties had bat skin microbiome compositions and diversities that were the most similar to each other. Ellipses indicate the 95% confidence interval and are calculated in the R package ggord.

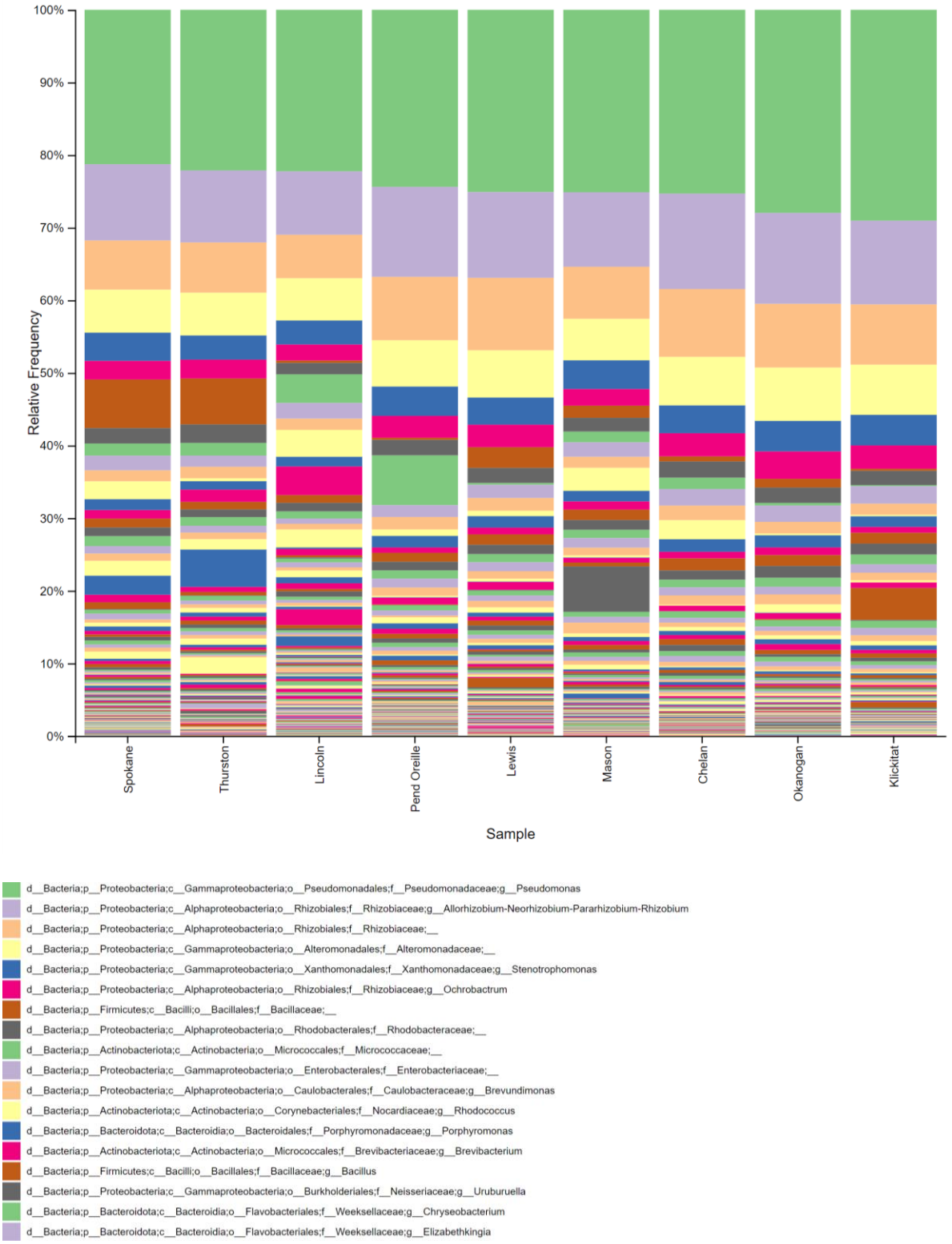


Fig. 27: Bacterial taxa bar plots illustrating relative abundances of bacterial taxa, grouped by roost location and averaged across all samples.

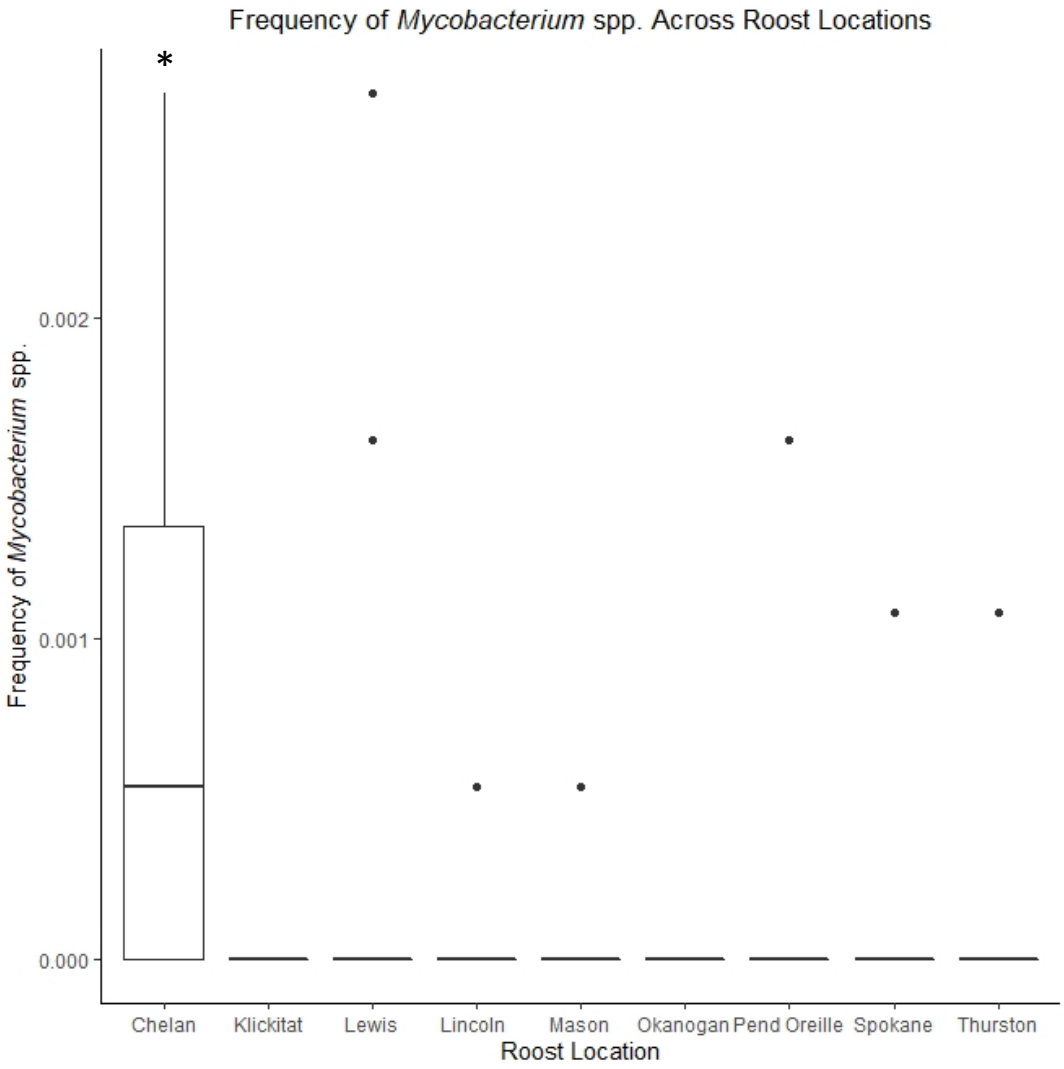


Fig. 28: Chelan County had significantly greater abundances of *Mycobacterium* spp. in the skin microbiomes of its bats (p = 0.02, maximum preference value test statistic = 0.535) compared to other counties.

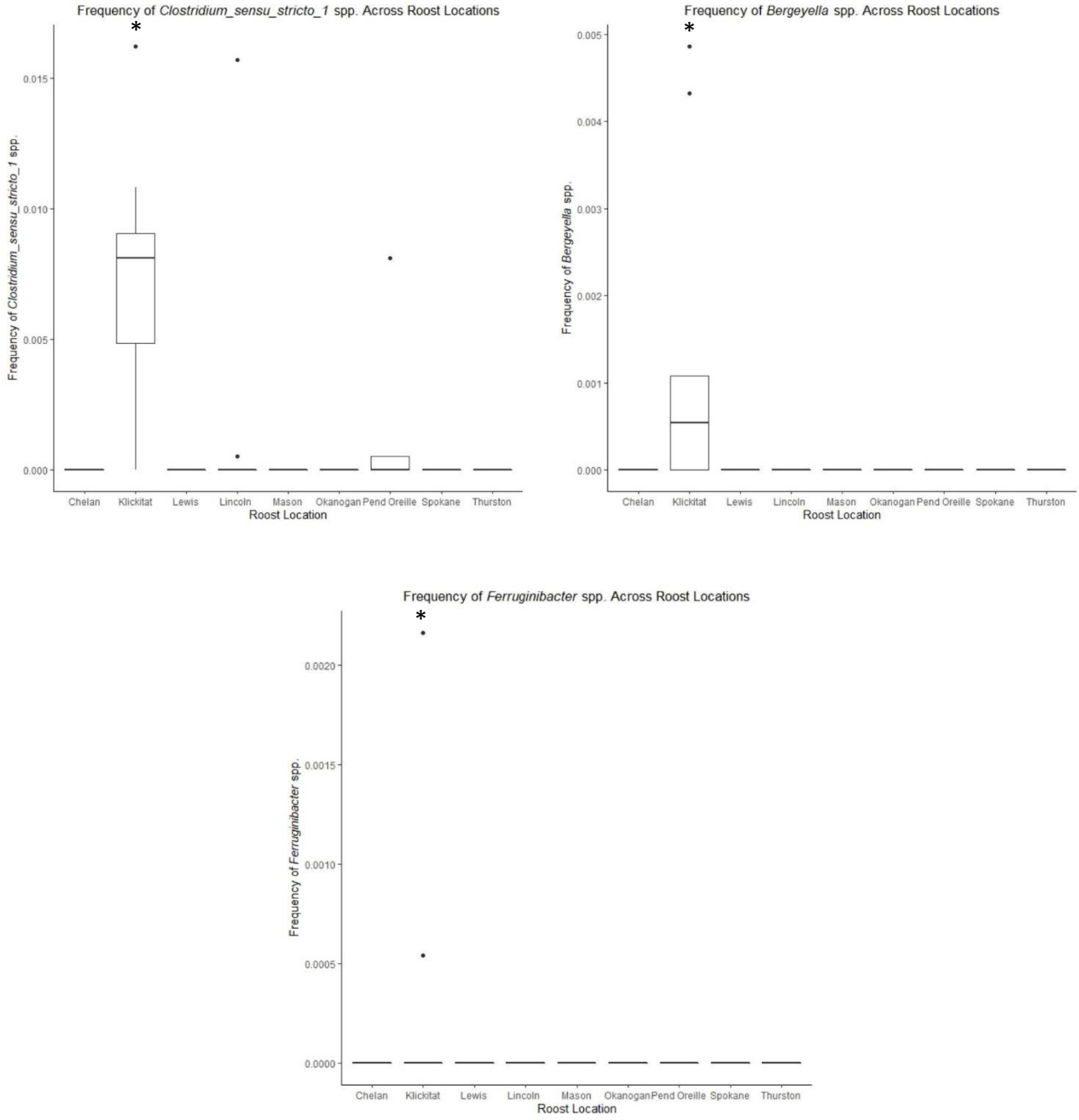


Fig. 29: Klickitat County had significantly greater abundances of *Clostridium_sensu_stricto_1* spp. (p = 0.005, maximum preference value test statistic = 0.825), *Bergeyella* spp. (p = 0.005, maximum preference value test statistic = 0.775), and *Ferruginibacter* spp. (p = 0.045, maximum preference value test statistic = 0.447) in the skin microbiomes of its bats compared to other counties.

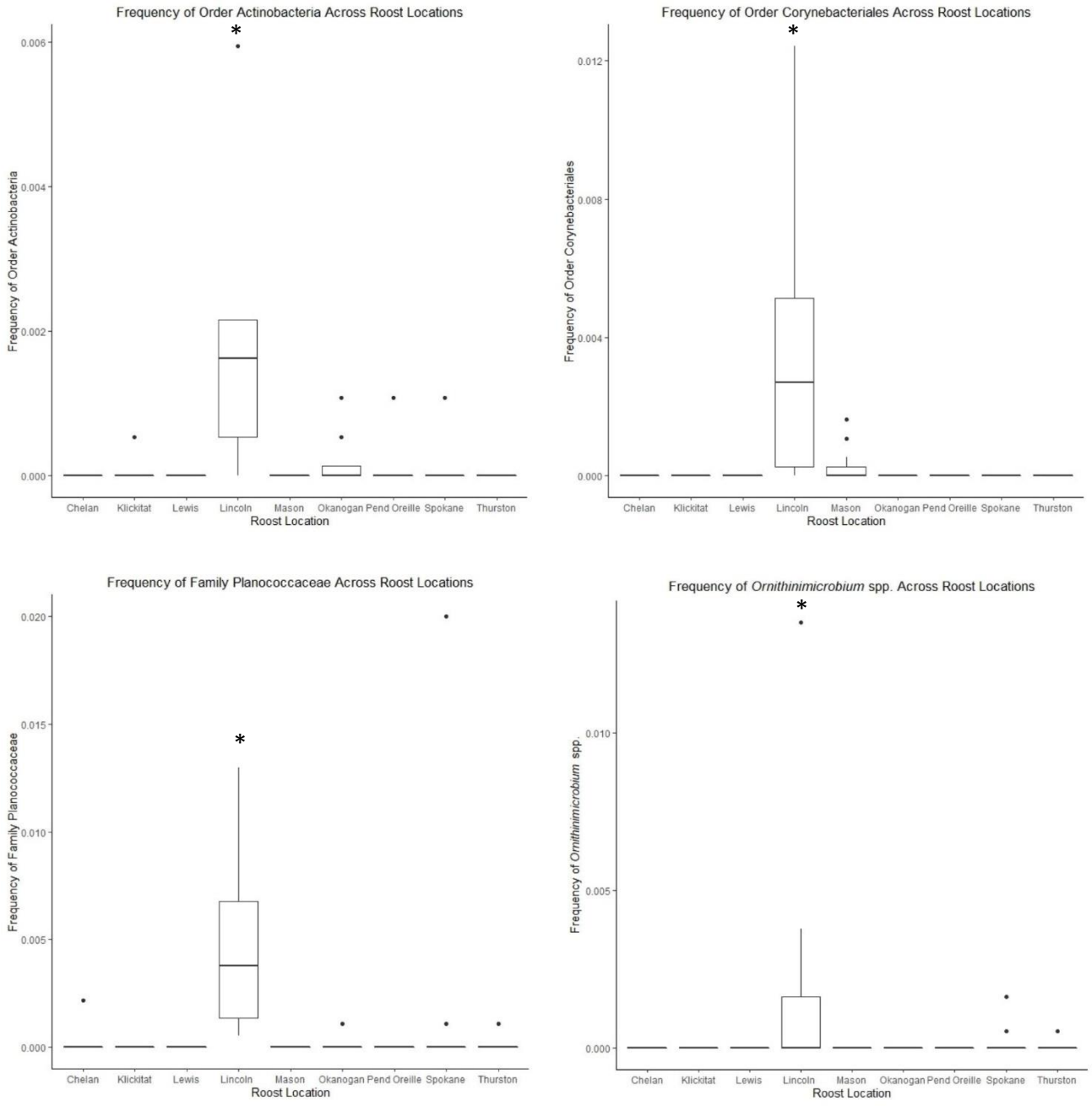


Fig. 30: Lincoln County had significantly greater abundances of class Actinobacteria (p = 0.005, maximum preference value test statistic = 0.846), order Corynebacteriales (p = 0.005, maximum preference value test statistic = 0.82), family Planococcaceae (p = 0.005, maximum preference value test statistic = 0.813), and *Ornithinimicrobium* spp. (p = 0.005, maximum preference value test statistic = 0.626) in the skin microbiomes of its bats compared to other counties.

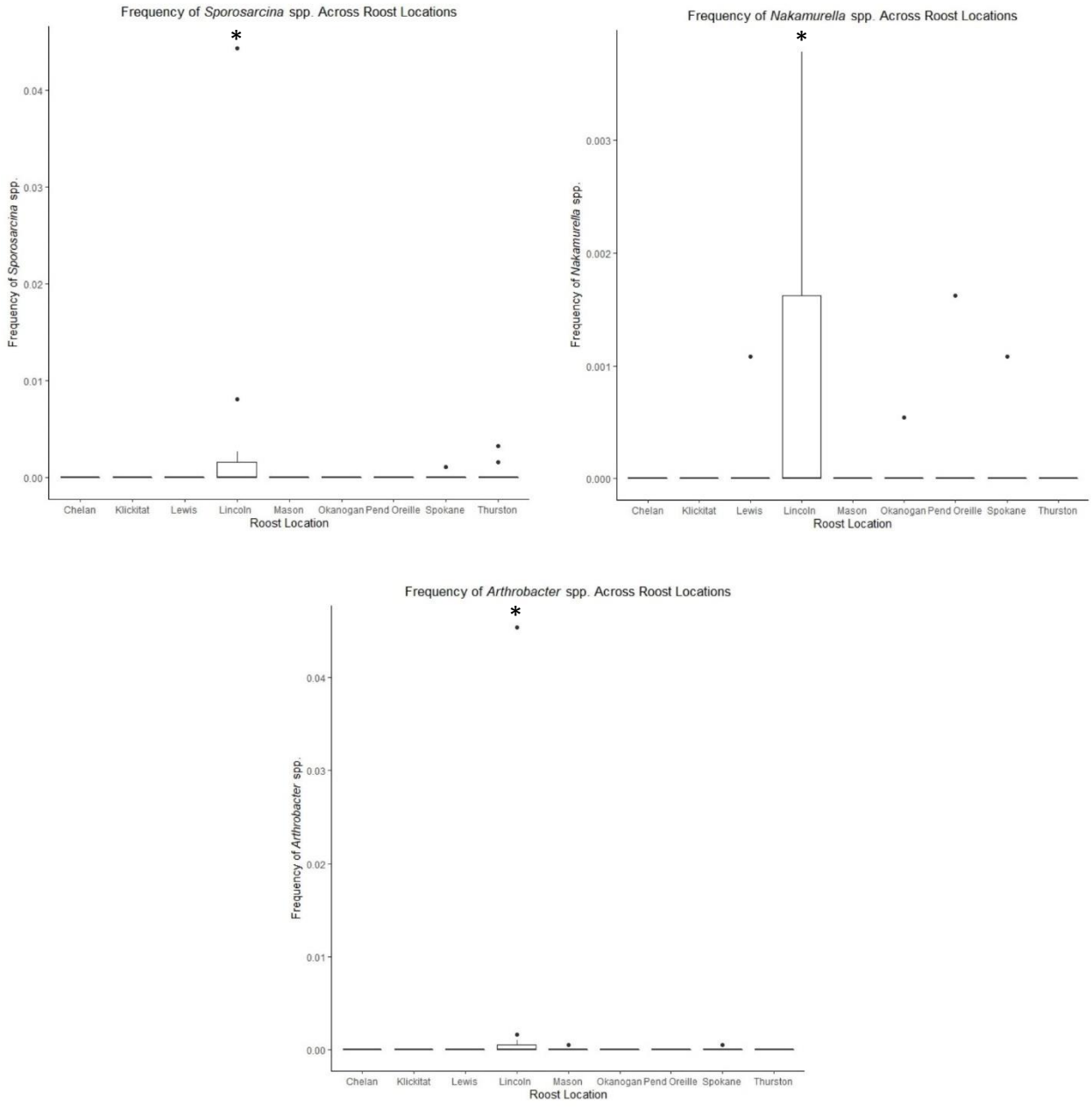


Fig. 31: Lincoln County also had significantly greater abundances of *Sporosarcina* spp. ($p = 0.025$, maximum preference value test statistic = 0.571), *Nakamurella* spp. ($p = 0.005$, maximum preference value test statistic = 0.57), and *Arthrobacter* spp. ($p = 0.03$, maximum preference value test statistic = 0.516) in the skin microbiomes of its bats compared to other counties.

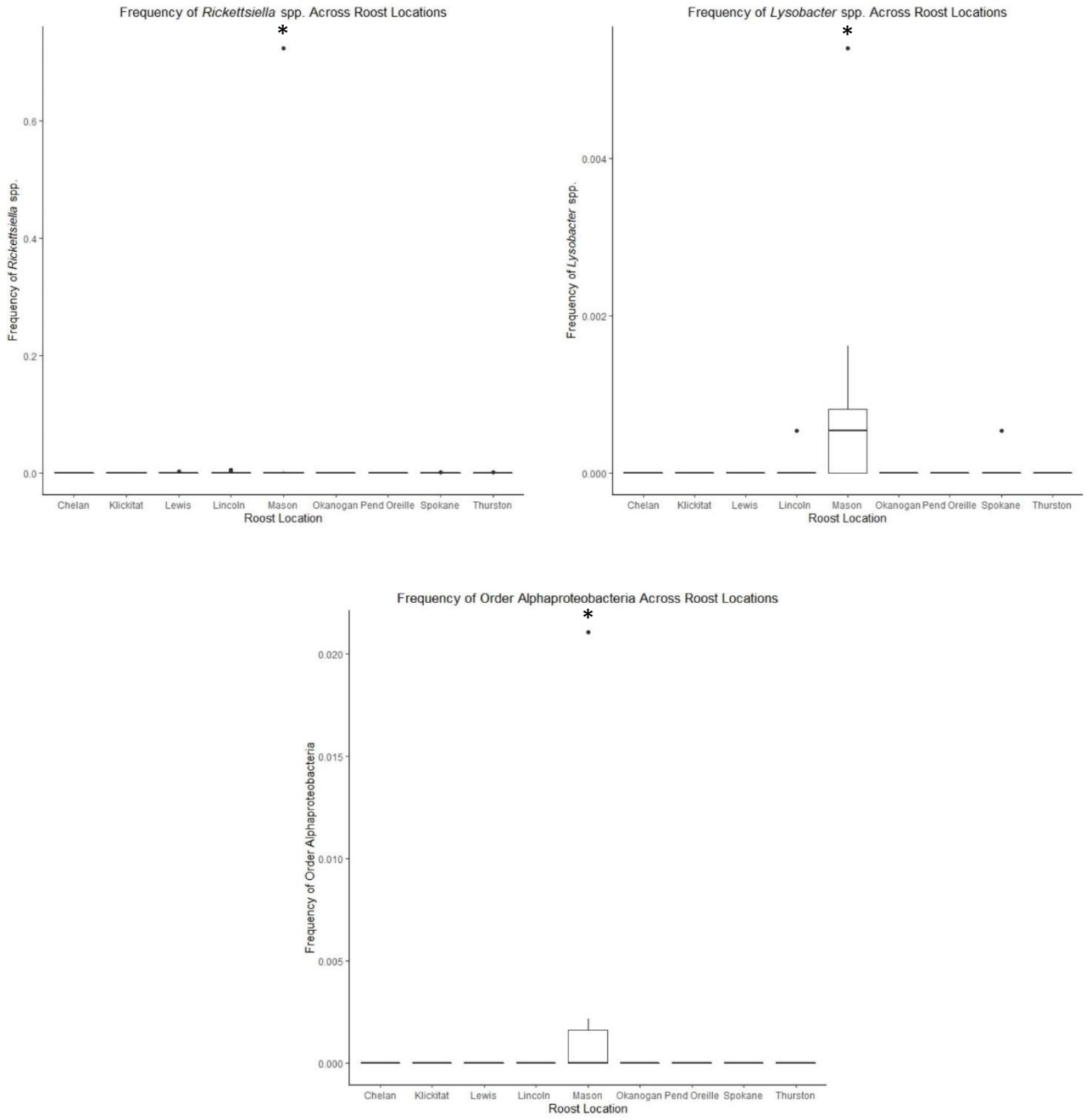


Fig. 32: Mason County had significantly greater abundances of *Rickettsiella* spp. (p = 0.025, maximum preference value test statistic = 0.723), *Lysobacter* spp. (p = 0.005, maximum preference value test statistic = 0.701), and class Alphaproteobacteria (p = 0.005, maximum preference value test statistic = 0.674) in the skin microbiomes of its bats compared to other counties.

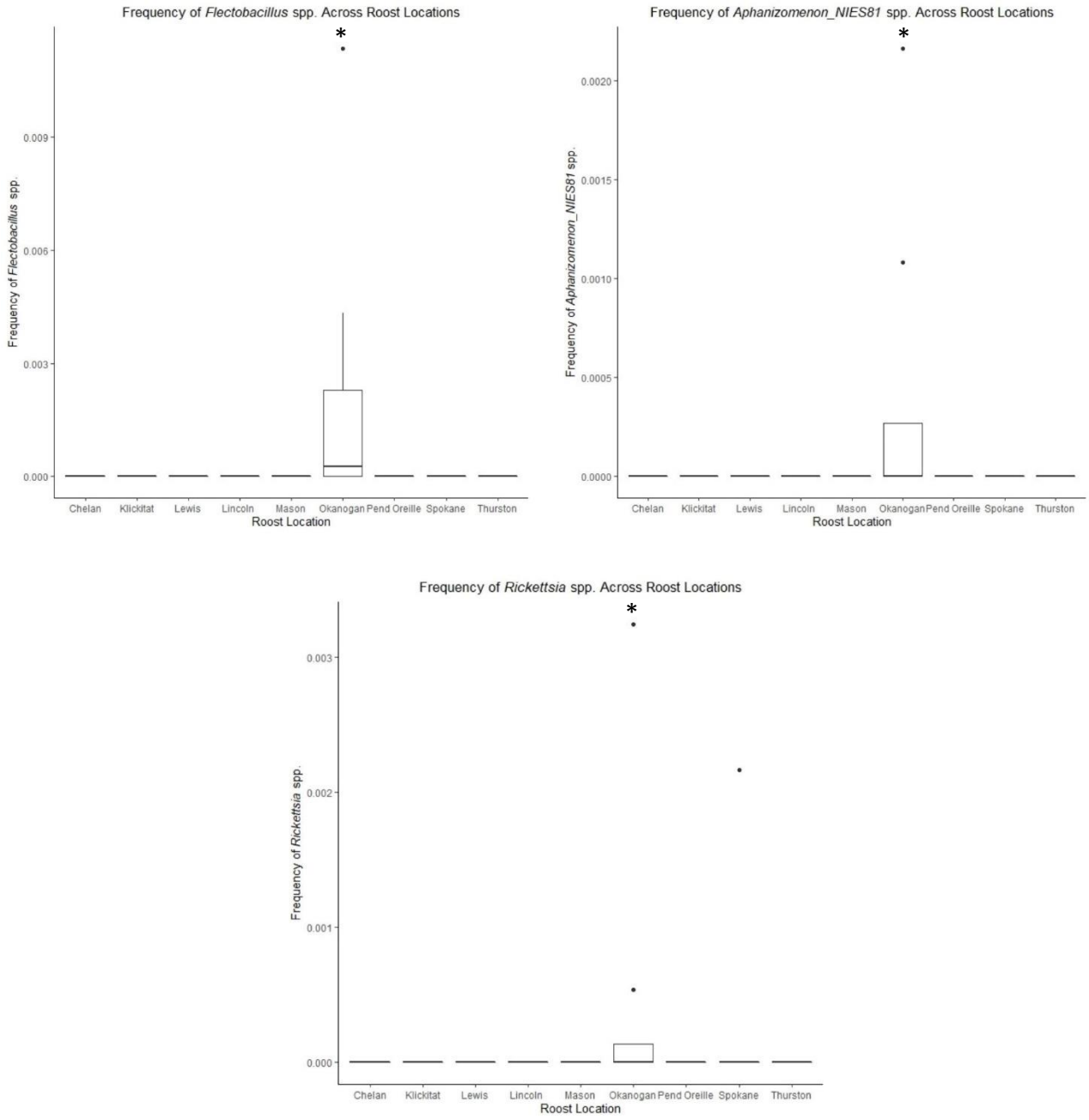


Fig. 33: Okanogan County had significantly greater abundances of *Flectobacillus* spp. (p = 0.005, maximum preference value test statistic = 0.707, *Aphanizomenon_NIES81* spp. (p = 0.015, maximum preference value test statistic = 0.5), and *Rickettsia* spp. (p = 0.04, maximum preference value test statistic = 0.42) in the skin microbiomes of its bats compared to other counties.

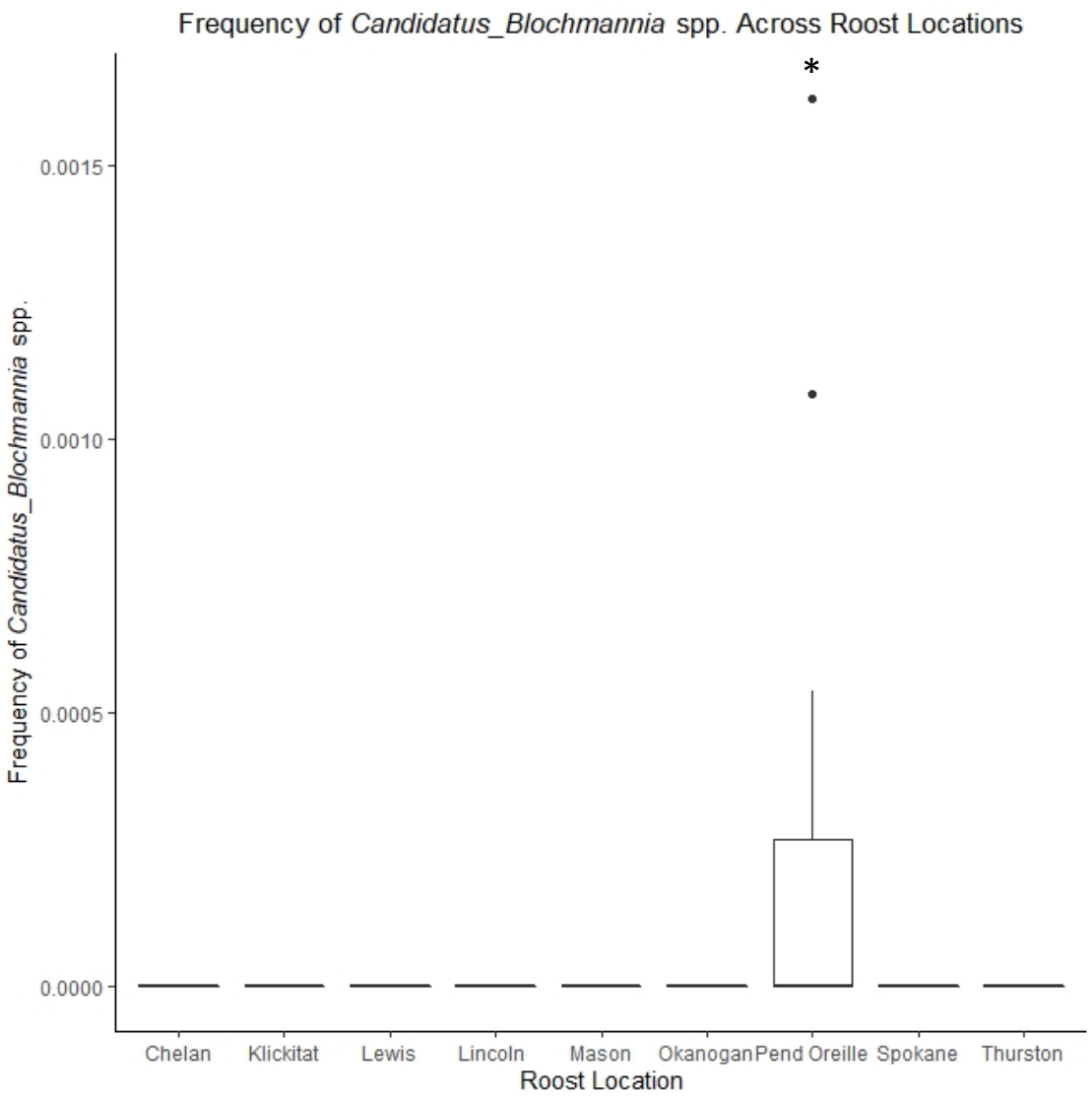


Fig. 34: Pend Oreille County had significantly greater abundances of *Candidatus_Blochmannia* spp. (p = 0.01, maximum preference value test statistic = 0.522) in the skin microbiomes of its bats compared to other counties.

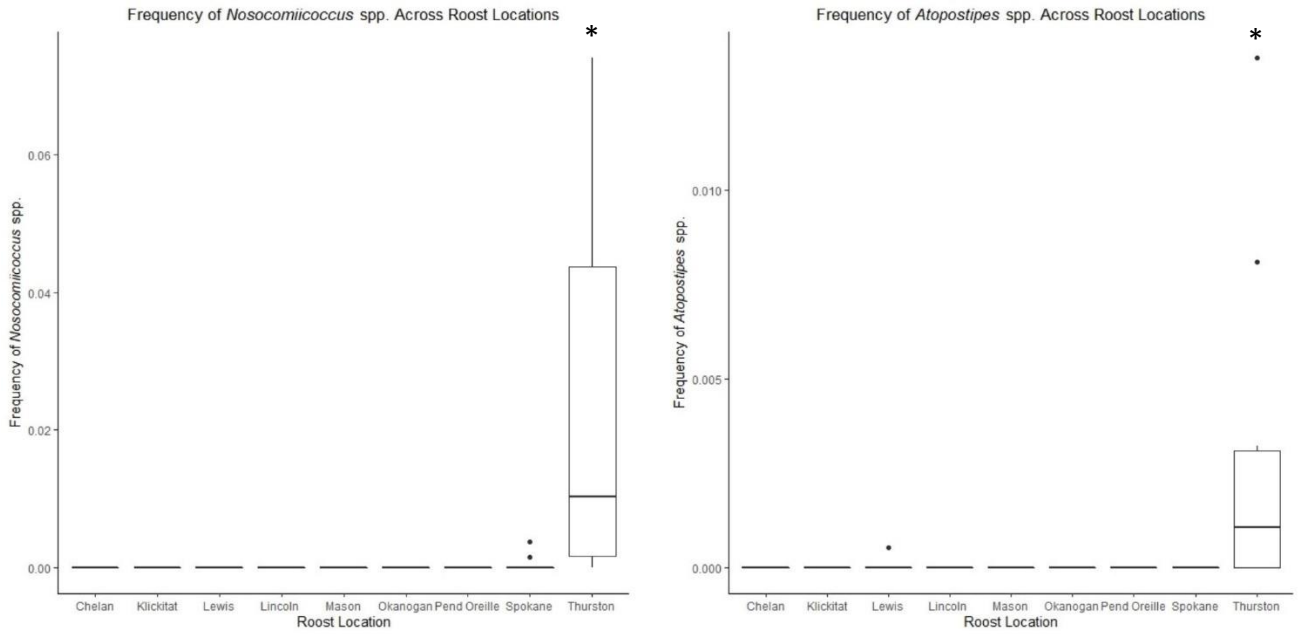


Fig. 35: Thurston County had significantly greater abundances of *Nosocomiicoccus* spp. ($p = 0.005$, maximum preference value test statistic = 0.939) and *Atopostipes* spp. ($p = 0.005$, maximum preference value test statistic = 0.701) in the skin microbiomes of its bats compared to other counties.

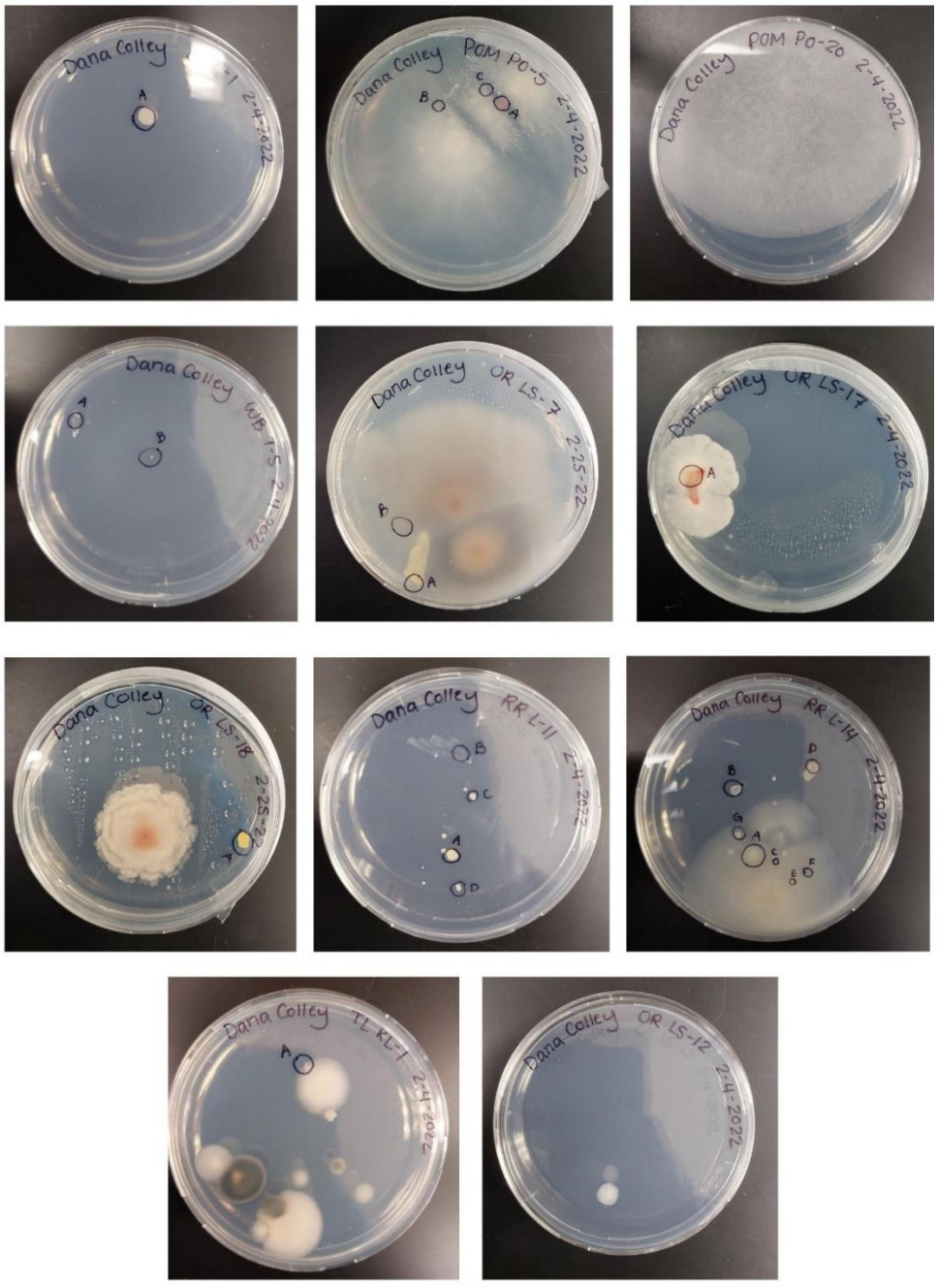


Fig. 36: 11 out of the 12 mass culture plates used in the culture sampling subset. The missing culture plate, sample OR LS-8 had no bacterial or fungal growth. Of all 12 mass-culture plates, 67% had at least one fungus. The two samples from Lincoln County had the highest bacterial richness, with 10 species between the two samples, while the sample from Chelan County had the lowest bacterial richness with just one colony. All photographs from Dana Colley.

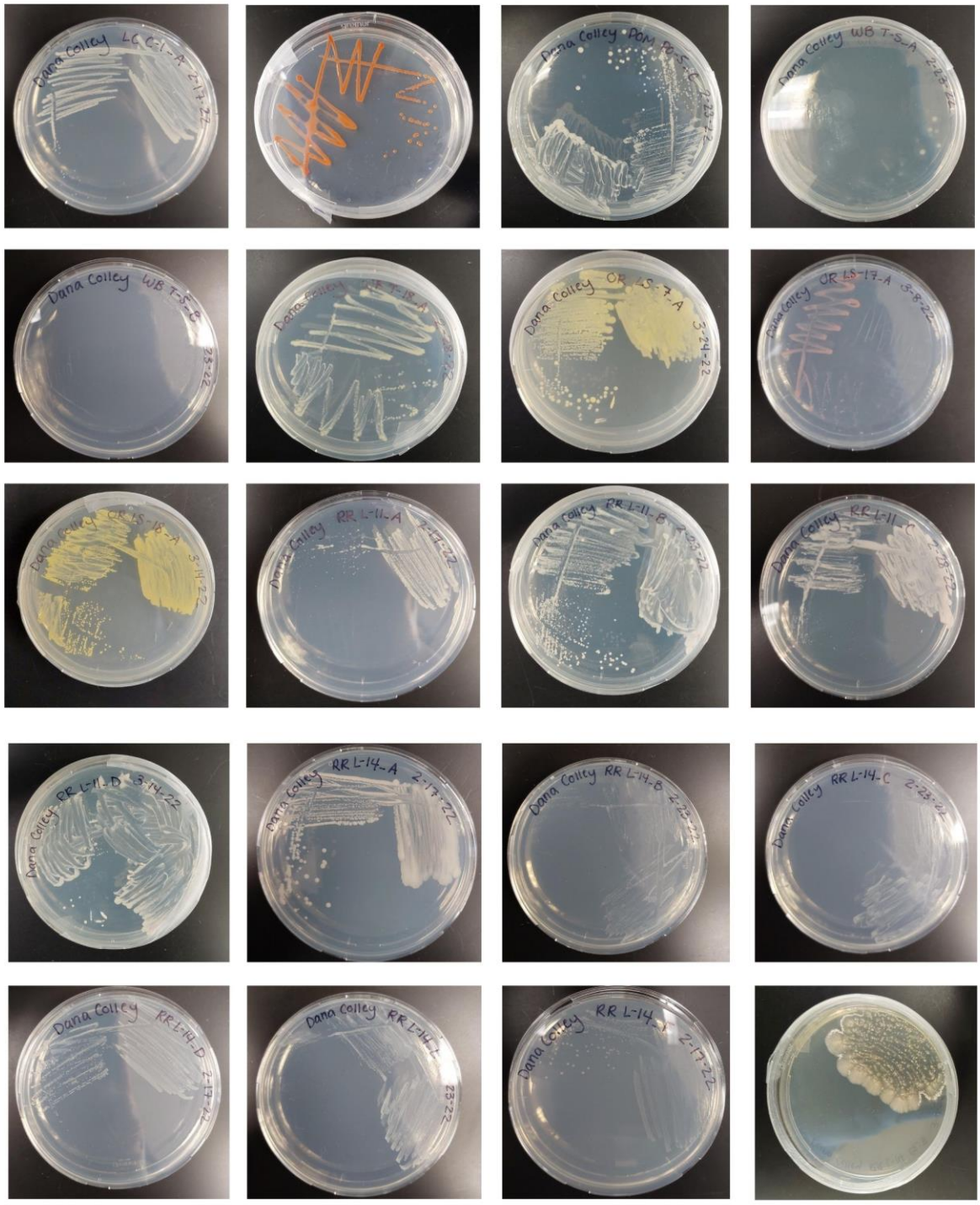


Fig. 37: The 20 bacterial isolates isolated from the mass culture plates. All culturable isolates were found in very low relative abundances in the sequenced bat skin microbiome, with an average relative abundance of 0.32% for any given isolate. All photographs from Dana Colley.

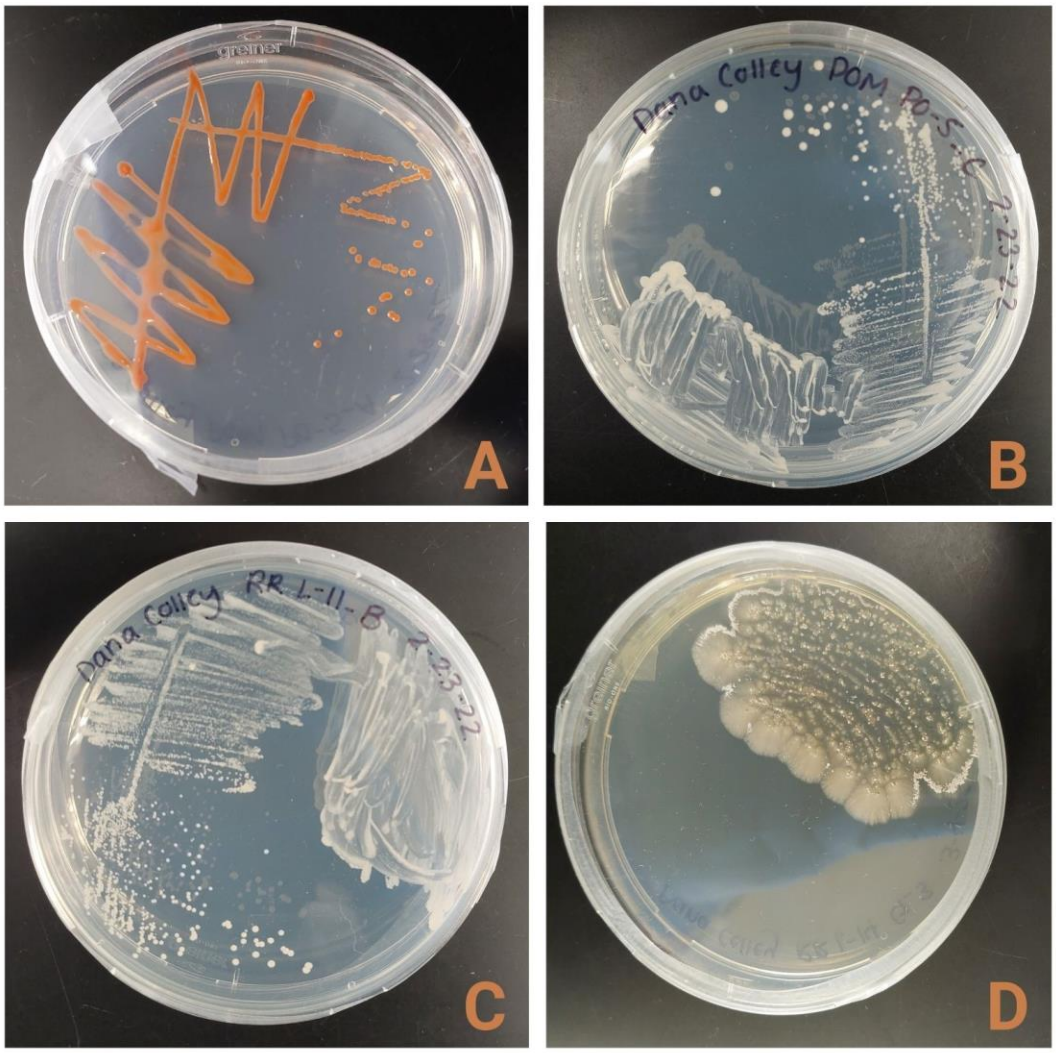


Fig. 38: The four isolates from genera known to possess antifungal properties. A) *Rhodococcus corynebacterioides*, B) *Rhodococcus tukisamuensis,* C) *Rhodococcus oryzae,* D) *Streptomyces laculatispora. Rhodococcus corynebacterioides* had a relative abundance of 0.0306% in the sequenced bat skin microbiome, while *R. tukisamuensis* had a relative abundance of 0.179%, and *R. oryzae* had a relative abundance of 1.231%. *Streptomyces laculatispora* had a relative abundance of 0.004% in the sequenced bat skin microbiome. All photographs from Dana Colley.

APPENDIX A

R Code Notebook for Data Analyses in R version 4.1.2

Investigating how bat ectoparasites influence the skin microbiome diversity and composition of Washington state bats

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify the most abundant bacteria from the skin of bats to compare them to the most abundant taxa from the sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the culturable bacterial isolates from western bats be in low relative abundances in our sequencing data.

1.) Cleaning up sampling metadata files in R and running preliminary data analyses:

- Basic information from the dataset: how many bats total, how many different bat species, sexes, how many ectoparasites by the different bat species?
- Run logistic regression (general linear mixed modeling) and an ANOVA to determine how ectoparasite presence differs by bat species (are some bat species more likely to have ectoparasites than others?)
- No microbiome data, just sampling data
- All outputs of statistical tests have been omitted in the code for sake of space.

```
#Set working directory to file with data you want to use
#Load in entire bat dataset - focusing on sampling, not microbiome
fullbat=read.csv("WDFW_EWU_Bat_Sampling_Data_Spring_2021.csv",header=T)

#Load in bat microbiome dataset
batmicrobiome=read.csv("Microbiome_WDFW_EWU_Bat_Sampling_Data_Spring_2021.csv
",header=T)

#How many different spp?
length(levels(fullbat$Species))

## [1] 0

#Set as a factor
fullbat$Species=factor(fullbat$Species)
length(levels(fullbat$Species))

## [1] 6
```

```r
#Look at frequency of diff spp, needs to be a data frame
speciestable=as.data.frame(table(fullbat$Species))


#Make a graph of bat spp observed
#load in ggplot2 to make pretty graphs
library(ggplot2)

#Bar graph with counts of each bat spp
ggplot(fullbat,aes(x=Species))+geom_bar()+theme_classic()+labs(y="Number of B
ats",title="Bat Species Observed",)

#grouped barplot, bat spp across the different roost sites (counties)
ggplot(fullbat,aes(x=Species,fill=County))+geom_bar(position="dodge")+theme_c
lassic()+labs(y="Number of Bats",title="Bat Species Observed")

#Look at ectoparasite presence across the bat spp
ggplot(fullbat,aes(x=Species,fill=Ectoparasites_.Y.N.))+geom_bar(position="do
dge")+theme_classic()+labs(y="Number of Bats",title="Bat Species Observed")

#Make a table of ectoparasite
speciesectotable=(table(fullbat$Ectoparasites_.Y.N.,fullbat$Species))
speciesectotable
##
##     COTO EPFU MYLU MYVO MYYU MYYU/LU
##   N    6    4   36    4   50      11
##   Y    0    8    6   12    7       3

ggplot(fullbat,aes(x=Ectoparasites_.Y.N.,fill=Species))+geom_bar(position="do
dge")+theme_classic()+labs(y="Number of Bats",title="Bat Species Observed")

#Run Chi-Square on Spp/Ecto table, null=no association btw spp and EP = same
proportion
chisq.test(speciesectotable)

#some #s really small so approx may be incorrect, less than 20% of #s in tabl
e should be less than 5

#run Fisher's, no assumption!
fisher.test(speciesectotable)

#some species have sig. diff portions of EPs, assuming those are representati
ve of the spp

#some spp more likely to have Eps than others

#which species specifically? Chi-sq table, big ones = diff; pairwise testing
with chisq.; turn into logistic regression, make ep the response var., sp as
predictor
#probability that bat has EP depending on what spp

#logistic regression
```

```
#make a new variable
fullbat$ecto=1
#finds bat w/o ecto, those = to 0
fullbat$ecto[fullbat$Ectoparasites_.Y.N.=="N"]=0
#run the logistic regression
model=glm(ecto~Species,data=fullbat,family="binomial")
summary(model)

#natural log of the odds (probability/not probability)in results table, coto
is intercept
#relative to coto, how sig diff?

#load in car library, run an anova
library(car)

## Loading required package: carData

Anova(model)

#make an effects plot
library(effects)

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

plot(allEffects(model),type="response")

#predicted probabilites of finding EP for each spp and 95% CI
#no variation in observed coto data, infinite uncertainty! Just b/c we didn't
find any?

#load in emmeans library (estimated means), will give us the p-values for eac
h comparison
library(emmeans)

#runs Tukey hsd for logistic regression, pairwise comparison
emmeans(model,pairwise~Species)

#compares spp to each other, gives p-vales, should match up with effects plot
#could be confounded by time and site, site as random effect

#mixed model, site as random effect? MYVO only on one site

#including site as a random effect
library(lme4)

## Loading required package: Matrix

#allows differences between spp to be different in diff counties
#making the least assumptions
model2=glmer(ecto~Species+(Species|County),data=fullbat,family=binomial)
```

```r
summary(model2)

#random effects, how much differences between counties, how much variation is
explained by the differences in counties
#similar numbers (lrg) = differences between counties

#run anova, makes 95% CI bigger, accounts for sites, use this for results!
#Uses proportions of ectoparasites, not counts = better representation
Anova(model2)
#Effects plot, ectoparasite presence by bat spp
plot(allEffects(model2),type="response")
#gives biger CI's
emmeans(model2,pairwise~Species)

#Number of sexes
table(fullbat$Sex_.M.F.)

##
##   F   M
## 142   5

#Number of sexes in each bat spp
table(fullbat$Sex_.M.F.,fullbat$Species)
##
##     COTO EPFU MYLU MYVO MYYU MYYU/LU
##   F    5   10   41   16   57      13
##   M    1    2    1    0    0       1

table(fullbat$Species)

##
##    COTO    EPFU    MYLU    MYVO    MYYU MYYU/LU
##       6      12      42      16      57      14

#Center the title on the graphs, add gg-title at the end of the code
#Make publication-worthy figures with centered titles

#Bar plot with number of bat species observed
ggplot(fullbat,aes(x=Species))+geom_bar()+theme_classic()+labs(y="Number of B
ats",title="Bat Species Observed")+ggtitle("Bat Species Observed") +
theme(plot.title = element_text(hjust = 0.5))

#Barplot with ectoparasites by bat species
ggplot(fullbat,aes(x=Species,fill=Ectoparasites_.Y.N.))+geom_bar(position="do
dge")+theme_classic()+labs(y="Number of Bats",title="Bat Species Observed")+g
gtitle("Ectoparasites Across Bat Species") +
  theme(plot.title = element_text(hjust = 0.5))

#Grouped bar plot, number of bats across counties
ggplot(fullbat,aes(x=Species,fill=County))+geom_bar(position="dodge")+theme_c
lassic()+labs(y="Number of Bats",title="Bat Species Observed")+ggtitle("Bat S
```

```
pecies Across Washington Counties") +
  theme(plot.title = element_text(hjust = 0.5))
```

2.) Calculating summary statistics from alpha diversity metrics.

- Dealing with our sequenced skin microbiome data.
- What are the means and standard deviations for the raw alpha diversity metric data?
    - Shannon's diversity, Faith's phylogenetic diversity, evenness, observed features
- I downloaded the raw TSV data files from the QIIME2 view website as I visualized the alpha diversity metrics. We then used those data files to calculate the means and standard deviations for all alpha diversity metrics for all variables (ectoparasite presence, bat species, roost location)

```
#Summary statistics for Shannon diversity results
#Read in the Shannon diversity alpha diversity results dataset
Shannon_bat=read.csv("Shannon_alpha_metadata.csv", header=T)

#Polish up the Shannon diversity dataset, remove unnecessary rows
names(Shannon_bat)=Shannon_bat[1,] #makes header names the true headers
Shannon_bat=Shannon_bat[-1,] #removes 1st row
Shannon_bat=Shannon_bat[-1,]

#Convert shannon entropy to numbers
Shannon_bat$shannon_entropy=as.numeric(Shannon_bat$shannon_entropy)

#Histogram
hist(Shannon_bat$shannon_entropy)

#ANOVA on Shannon entropy values
#Shannon entropy values by bat species
model1=aov(shannon_entropy~Species,data=Shannon_bat)
summary(model1)

#which bat species are different? TukeyHSD
TukeyHSD(model1)

library(effects)
plot(allEffects(model1)) #effects plot
plot(model1) #See all the different types of effects plots to check that the
data is normally distributed before moving on

#Shannon entropy values by roost location
model2=aov(shannon_entropy~Site,data=Shannon_bat)
summary(model2)

#which one, TukeyHSD
TukeyHSD(model2)
```

```
library(effects)
plot(allEffects(model2))
plot(model2)

#Shannon entropy values by ectoparasites
model3=aov(shannon_entropy~Ectoparasites,data=Shannon_bat)
summary(model3)

#which one, TukeyHSD
TukeyHSD(model3)

library(effects)
plot(allEffects(model3))
plot(model3)

summary(allEffects(model3)) #provides means, 95% CI not st.dev



#Shortcut - to just get means and standard deviations, us tapply
#Shannon summary stats by group, ectoparasites
tapply(Shannon_bat$shannon_entropy,Shannon_bat$Ectoparasites,summary)
tapply(Shannon_bat$shannon_entropy,Shannon_bat$Ectoparasites,sd)


#Shannon summary stats for bat species
tapply(Shannon_bat$shannon_entropy,Shannon_bat$Species,mean)
tapply(Shannon_bat$shannon_entropy,Shannon_bat$Species,sd)

#summary stats for site
tapply(Shannon_bat$shannon_entropy,Shannon_bat$Site,mean)
tapply(Shannon_bat$shannon_entropy,Shannon_bat$Site,sd)


#Evenness summary stats
#Read in evenness metadata - alpha diversity
Evenness_bat=read.csv("Evenness_alpha_metadata.csv", header=T)

#Polish evenness data
names(Evenness_bat)=Evenness_bat[1,]
Evenness_bat=Evenness_bat[-1,] #removes 1st row
Evenness_bat=Evenness_bat[-1,]

#Convert pielou's evenness to numbers so you can use it
Evenness_bat$pielou_evenness=as.numeric(Evenness_bat$pielou_evenness)

#Histogram
hist(Evenness_bat$pielou_evenness)

#Evenness summary stats by group, ectoparasites
tapply(Evenness_bat$pielou_evenness,Evenness_bat$Ectoparasites,mean)
```

```
tapply(Evenness_bat$pielou_evenness,Evenness_bat$Ectoparasites,sd)

#Evennesssummary stats for species
tapply(Evenness_bat$pielou_evenness,Evenness_bat$Species,mean)
tapply(Evenness_bat$pielou_evenness,Evenness_bat$Species,sd)
#Evenness summary stats for site
tapply(Evenness_bat$pielou_evenness,Evenness_bat$Site,mean)
tapply(Evenness_bat$pielou_evenness,Evenness_bat$Site,sd)


#Faith's phylogenetic diversity summary stats
#Read in faith's pd metadata - alpha diversity
Faith_pd_bat=read.csv("Faith_pd_alpha_metadata.csv", header=T)

#Clean up faith’s pd data
names(Faith_pd_bat)=Faith_pd_bat[1,]
Faith_pd_bat=Faith_pd_bat[-1,] #removes 1st row
Faith_pd_bat=Faith_pd_bat[-1,]

#Convert faith's pd to numbers
Faith_pd_bat$faith_pd=as.numeric(Faith_pd_bat$faith_pd)

#Histogram
hist(Faith_pd_bat$faith_pd)

#Faith pd summary stats by group, ectoparasites
tapply(Faith_pd_bat$faith_pd,Faith_pd_bat$Ectoparasites,mean)
tapply(Faith_pd_bat$faith_pd,Faith_pd_bat$Ectoparasites,sd)

#summary stats for species
tapply(Faith_pd_bat$faith_pd,Faith_pd_bat$Species,mean)
tapply(Faith_pd_bat$faith_pd,Faith_pd_bat$Species,sd)

#summary stats for site
tapply(Faith_pd_bat$faith_pd,Faith_pd_bat$Site,mean)
tapply(Faith_pd_bat$faith_pd,Faith_pd_bat$Site,sd)



#Observed features summary stats
#Read in obs. feat metadata - alpha diversity
Obs_feat_bat=read.csv("Observed_featured_alpha_metadata.csv", header=T)


#Polish observed features data
names(Obs_feat_bat)=Obs_feat_bat[1,]
Obs_feat_bat=Obs_feat_bat[-1,] #removes 1st row
Obs_feat_bat=Obs_feat_bat[-1,]
```

```
#Convert observed features to numbers
Obs_feat_bat$observed_features=as.numeric(Obs_feat_bat$observed_features)

#Histogram
hist(Obs_feat_bat$observed_features)

#summary stats by group, ectoparasites
tapply(Obs_feat_bat$observed_features,Obs_feat_bat$Ectoparasites,mean)
tapply(Obs_feat_bat$observed_features,Obs_feat_bat$Ectoparasites,sd)

#summary stats for species
tapply(Obs_feat_bat$observed_features,Obs_feat_bat$Species,mean)
tapply(Obs_feat_bat$observed_features,Obs_feat_bat$Species,sd)

#summary stats for site
tapply(Obs_feat_bat$observed_features,Obs_feat_bat$Site,mean)
tapply(Obs_feat_bat$observed_features,Obs_feat_bat$Site,sd)
```

3.) Running indicspecies analysis in R to determine how relative abundance of bacterial taxa differ by different variables (ectoparasite presence, bat species, roost location)

- How do the relative abundance of bacterial taxa differ between different groups (ectoparasites, bat species, roost location)?
- Using microbiome data, specifically the feature table from QIIME
- Be careful, make sure that you get the correct group for the correct variable
- We built boxplots for all of the indicspecies taxa since they were in such low abundances that they couldn't be seen on the grouped taxa bar plot

```
#Indicspecies Analysis
#Use the feature table from QIIME2, make sure you export it from QIIME2 and t
hen convert it to a .csv file
Bat = read.csv("Bat Feature Table_Excel.csv", header=T)

#Polish frequency table
#Trim extra taxonomy column from data table
Bat = Bat[,-92]
names(Bat)=Bat[2,]
Bat=Bat[-2,]
Ectoparasite=Bat[1,]#stores this data
Bat=Bat[-1,]
row.names(Bat)=Bat[,1]
Bat=Bat[,-1]#removes 1st row

for (i in 1:90) Bat[,i]=as.numeric(Bat[,i])#convert that one column


#Transpose dataset
```

```
Bat=t(Bat)

#Trim ectoparasite dataset
#Convert ectoparasite to character
Ectoparasite=as.character(Ectoparasite)#not a data frame
Ectoparasite=Ectoparasite[-1]
#Convert ectoparasite to 0s and 1s
#Convert to factor, then convert to numeric
Ectoparasite=as.factor(Ectoparasite)
Ectoparasite=as.numeric(Ectoparasite)
#2 = Y, 1 = N

#load in the indicspecies package
library(indicspecies)
#Run indicspecies analysis by ectoparasite presence (yes/no)

indval=multipatt(Bat,Ectoparasite)
summary(indval)

##  List of species associated to each combination:
##
##  Group 1  #sps.  1 #make sure you know which group corresponds to what
##
##  Group 2  #sps.  4
#Trying to figure out which groups of bacteria are associated with each group
, which bacteria associated with which group
#mixes up group randomly, compares randomly to actual, which is not like the
random
#only 5 spp that are associated with EP/non-EP group
#permutational test, larger = better?

#Indicspecies analysis by site (roost location)
#The frequency table file name is called species but it actually has site dat
a! Whoopsies.
Site = read.csv("Bat Feature Table_Species.csv", header=T)

#Polish bat species dataset
#Need to put bat species names to the proper sample, forgot to put this in th
e frequency table
Site = Site[,-92]
names(Site)=Site[2,]
Site=Site[-2,]
County=Site[1,]#stores this data
Site=Site[-1,]
row.names(Site)=Site[,1]
Site=Site[,-1]#removes 1st row #2 = EP, associated with EP bats! Just shows t
he sig. groups

for (i in 1:90) Site[,i]=as.numeric(Site[,i])#convert that one column
```

```
#Transpose dataset
Site=t(Site)

#Trim site dataset
#Convert site to character
County=as.character(County)#not a data frame
County=County[-1]

#Convert to factor, then convert to numeric
County=as.factor(County)
levels(County) #important so you know what group number corresponds to what c
ounty!
## [1] "Chelan"       "Klickitat"    "Lewis"        "Lincoln "     "Mason"
## [6] "Okanogan"     "Pend Oreille" "Spokane"      "Thurston"

County=as.numeric(County)
#1 = Chelan; 2 = Klickitat; 3 = Lewis; 4 = Lincoln; 5 = Mason; 6 = Okan. 7 =
PO; 8=Spokane; 9 = Thurston

#load in the indicspecies package
library(indicspecies)
indval_County=multipatt(Site,County)
summary(indval_County)



#Indicspecies analysis by bat species

#Polish frequency table with species info
#This freuqnecy table file actually ahs species data included
Species = read.csv("Bat_Species(Real).csv", header=T)
#compare sample IDs, which samples were dropped?
Bat = read.csv("Bat Feature Table_Excel.csv", header=T)

#Trim extra taxonomy column from data table
Bat = Bat[,-92]
Not_in=which(!Species$X.SampleID %in% Bat[2,])
Species=Species[-Not_in,]
names(Species)[1]="SampleID"

Bat=Bat[-1,]
row.names(Bat)=Bat[,1]
Bat=Bat[,-1]#removes 1st row

#Transpose dataset
Bat=t(Bat)
Bat=as.data.frame(Bat)
```

```
#Need to add in the proper bat species info with each sample in R since I for
got to put it in the frequency table
Bat_w_species=merge(Bat,Species,by="SampleID")



Species=Bat_w_species$Species
Bat_w_species=Bat_w_species[,-c(1,252)]

for (i in 1:250) Bat_w_species[,i]=as.numeric(Bat_w_species[,i])#convert that
one column

Species=as.factor(Species)
levels(Species)
## [1] "COTO"    "EPFU"    "MYLU"    "MYVO"    "MYYU"    "MYYU.LU"

Species=as.numeric(Species)


#Indicspecies analysis by bat species
indval_Species=multipatt(Bat_w_species,Species)
summary(indval_Species)



#Individual taxa barplots
#Just for species

#Load in ggplot library
library(ggplot2)
Species2=as.data.frame(Species)
Species2$Species=as.factor(Species2$Species)

#ggplot(Bat_w_species,aes_string(x=names(Bat_w_species)[1],y=names(Bat_w_spec
ies)[which(names(Bat_w_species)=="d__Bacteria;p__Bacteroidota;c__Bacteroidia;
o__Flavobacteriales;f__Crocinitomicaceae;g__Brumimicrobium")]))+geom_point()
boxplot(Bat_w_species[,which(names(Bat_w_species)=="d__Bacteria;p__Bacteroido
ta;c__Bacteroidia;o__Flavobacteriales;f__Crocinitomicaceae;g__Brumimicrobium"
)]~Species,ylab="Brumimicrobium")

#could change the column name for that species so easier to work with
#Polish bat species data
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Bacteroidota
;c__Bacteroidia;o__Flavobacteriales;f__Crocinitomicaceae;g__Brumimicrobium")]
="Brumimicrobium"

Species2$Species=as.character(Species2$Species)
Species2$Species[which(Species2$Species=="5")]="MYYU"
Species2$Species[which(Species2$Species=="4")]="MYVO"
```

```
Species2$Species[which(Species2$Species=="3")]="MYLU"
Species2$Species[which(Species2$Species=="2")]="EPFU"
Species2$Species[which(Species2$Species=="1")]="COTO"
Species2$Species[which(Species2$Species=="6")]="MYYU.LU"



#COTO Bacteria

#Check in frequency table with CTRL+F
#Klenkia spp.
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Actinobacter
iota;c__Actinobacteria;o__Frankiales;f__Geodermatophilaceae;g__Klenkia")]="Kl
enkia"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Klenkia))+geom_boxplot()+theme_
classic()+labs(y=expression(paste("Frequency of ",italic("Klenkia"),"
spp.")),x="Bat Species")+ggtitle(expression(paste("Frequency of
",italic("Klenkia"),"  spp. Across Bat Species"))) +
  theme(plot.title = element_text(hjust = 0.5))


#Family Stappiaceae
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Proteobacter
ia;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Stappia")]="Stappi
aceae"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Alcaligenaceae))+geom_boxplot()
+theme_classic()+labs(y="Frequency of Family Stappiaceae",x="Bat
Species")+ggtitle("Frequency of Family Stappiaceae Across Bat Species") +
  theme(plot.title = element_text(hjust = 0.5))

#Repeat for other abundant COTO bacteria from indicspecies analysis

#EPFU Bacteria
#Bergeyella spp.
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Bacteroidota
;c__Bacteroidia;o__Flavobacteriales;f__Weeksellaceae;g__Bergeyella")]="Bergey
ella"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Bergeyella))+geom_boxplot()+the
me_classic()+labs(y=expression(paste("Frequency of ",italic("Bergeyella"),"
spp.")),x="Bat Species")+ggtitle(expression(paste("Frequency of
",italic("Bergeyella"),"  spp. Across Bat Species"))) +
  theme(plot.title = element_text(hjust = 0.5))

#Order d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Firmicutes;c
__Bacilli;o__Bacillales;__;__")]="Bacillales"
```

```
ggplot(Bat_w_species,aes(x=Species2$Species,y=Bacillales))+geom_boxplot()+the
me_classic()+labs(y="Frequency of Order Bacillales",x="Bat
Species")+ggtitle("Frequency of Order Bacillales Across Bat Species") +
  theme(plot.title = element_text(hjust = 0.5))

#Repeat for other abundant EPFU bacteria from indicspecies analysis

#MYYU

#d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Dermab
acteraceae;g__Brachybacterium
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Actinobacter
iota;c__Actinobacteria;o__Micrococcales;f__Dermabacteraceae;g__Brachybacteriu
m")]="Brachybacterium"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Brachybacterium))+geom_boxplot(
)+theme_classic()+labs(y=expression(paste("Frequency of
",italic("Brachybacterium")," spp.")),x="Bat
Species")+ggtitle(expression(paste("Frequency of
",italic("Brachybacterium")," spp. Across Bat Species"))) +
  theme(plot.title = element_text(hjust = 0.5))

#d__Bacteria;p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcacea
e;__
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Firmicutes;c
__Bacilli;o__Staphylococcales;f__Staphylococcaceae;__")]="Staphylococcaceae"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Staphylococcaceae))+geom_boxplo
t()+theme_classic()+labs(y="Frequency of Family Staphylococcaceae",x="Bat
Species")+ggtitle("Frequency of Family Staphylococcaceae Across Bat Species")
+
  theme(plot.title = element_text(hjust = 0.5))



#MYYU/LU

#d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__
Morganellaceae;g__Candidatus_Blochmannia
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Proteobacter
ia;c__Gammaproteobacteria;o__Enterobacterales;f__Morganellaceae;g__Candidatus
_Blochmannia")]="Candidatus_Blochmannia"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Candidatus_Blochmannia))+geom_b
oxplot()+theme_classic()+labs(y=expression(paste("Frequency of
",italic("Candidatus_Blochmannia")," spp.")),x="Bat
Species")+ggtitle(expression(paste("Frequency of
",italic("Candidatus_Blochmannia")," spp. Across Bat Species"))) +
```

```
  theme(plot.title = element_text(hjust = 0.5))

#d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f
__Verrucomicrobiaceae;g__uncultured
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Verrucomicro
biota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae;g__unc
ultured")]="Verrucomicrobiaceae"

ggplot(Bat_w_species,aes(x=Species2$Species,y=Verrucomicrobiaceae))+geom_boxp
lot()+theme_classic()+labs(y="Frequency of Family Verrucomicrobiaceae",x="Bat
Species")+ggtitle("Frequency of Family Verrucomicrobiaceae Across Bat
Species") +
  theme(plot.title = element_text(hjust = 0.5))

#Repeat for other abundant MYYU/LU bacteria from indicspecies analysis


#Boxplots by ectoparasite presence

Bat = read.csv("Bat Feature Table_Excel_Site.csv", header=T)

#Trim extra taxonomy column from data table
Bat = Bat[,-92]

Bat=Bat[-1,]
row.names(Bat)=Bat[,1]
Bat=Bat[,-1]#removes 1st row
Bat=Bat[-1,]#removes 1st row


for (i in 1:90) Bat[,i]=as.numeric(Bat[,i])#convert that one column


#Transpose dataset
Bat=t(Bat)
Bat=as.data.frame(Bat)
Bat_w_species=Bat

#no ectoparasites
#d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Microc
occaceae;g__Micrococcus
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Actinobacter
iota;c__Actinobacteria;o__Micrococcales;f__Micrococcaceae;g__Micrococcus")]="
Micrococcus"

ggplot(Bat_w_species,aes(x=ectoparasite,y=Micrococcus))+geom_boxplot()+theme_
classic()+labs(y=expression(paste("Frequency of ",italic("Micrococcus"),"
spp.")),x="Ectoparasite Presence")+ggtitle(expression(paste("Frequency of
",italic("Micrococcus")," spp. by Ectoparasite Presence Status"))) +
```

```
  theme(plot.title = element_text(hjust = 0.5))


#ectoparasites
#d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Propionibacteriales;f__
Nocardioidaceae;g__Nocardioides
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Actinobacter
iota;c__Actinobacteria;o__Propionibacteriales;f__Nocardioidaceae;g__Nocardioi
des")]="Nocardioides"

ggplot(Bat_w_species,aes(x=ectoparasite,y=Nocardioides))+geom_boxplot()+theme
_classic()+labs(y=expression(paste("Frequency of ",italic("Nocardioides"),"
spp.")),x="Ectoparasite Presence")+ggtitle(expression(paste("Frequency of
",italic("Nocardioides"),"  spp. by Ectoparasite Presence Status"))) +
  theme(plot.title = element_text(hjust = 0.5))

#Repeat for other abundant ectoparsite bacteria from indicspecies analysis


#By site
Bat_Site = read.csv("Bat Feature Table_Species.csv", header=T)

as.character(Bat_Site[1,2:91])

#Klickitat County
#d__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae;g
__Clostridium_sensu_stricto_1
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Firmicutes;c
__Clostridia;o__Clostridiales;f__Clostridiaceae;g__Clostridium_sensu_stricto_
1")]="Clostridium_sensu_stricto_1"

ggplot(Bat_w_species,aes(x=as.character(Bat_Site[1,2:91]),y=Clostridium_sensu
_stricto_1))+geom_boxplot()+theme_classic()+labs(y=expression(paste("Frequenc
y of ",italic("Clostridium_sensu_stricto_1"),"  spp.")),x="Roost
Location")+ggtitle(expression(paste("Frequency of
",italic("Clostridium_sensu_stricto_1"),"  spp. Across Roost Locations"))) +
  theme(plot.title = element_text(hjust = 0.5))

#Linoln County
#d__Bacteria;p__Actinobacteriota;c__Actinobacteria;__;__;__
#labs(y="Frequency of Family Staphylococcaceae",x="Bat
Species")+ggtitle("Frequency of Family Staphylococcaceae Across Bat Species")
names(Bat_w_species)[which(names(Bat_w_species)=="d__Bacteria;p__Actinobacter
iota;c__Actinobacteria;__;__;__")]="Actinobacteria"

ggplot(Bat_w_species,aes(x=as.character(Bat_Site[1,2:91]),y=Actinobacteria))+
geom_boxplot()+theme_classic()+labs("Frequency of Order
Actinobacteria",x="Roost Location",y="Frequency of Order
```

```
Actinobacteria")+ggtitle("Frequency of Order Actinobacteria Across Roost
Locations") +
  theme(plot.title = element_text(hjust = 0.5))

#Repeat for other abundant bacteria in different counties from indicspecies
analysis
```

4.) Creating ordination plots in R using our data output from QIIME2. This is a way to interface QIIME in R, which can be useful for future QIIME2 data analyses.

- A little tricky!
- We used the rarefied sequence table from QIIME2, in addition to the rooted-tree file and taxonomy file. We were able to use the straight .qza files from QIIME, no exporting necessary.
- All ordination plots are already set up to have the ellipses and centered titles.

```
#PCoA plots in R using microbiome data from QIIME2

#Run an NMDS, need vegan library
library(vegan)
library(ggord)

#Install qiime2R with the commented chunks below
#if (!requireNamespace("devtools", quietly =
TRUE)){install.packages("devtools")}
#devtools::install_github("jbisanz/qiime2R") # current version is 0.99.20

#Load in qiime2R package
library(qiime2R)
library (devtools)

#Read in your rarefied sequence table directly from QIIME2
SVs<-read_qza("Bat_2021_rarefied-table.qza")
names(SVs) #this was supposed to read "ASVs", not "SVs"
## [1] "uuid"       "type"       "format"     "contents"   "version"
## [6] "data"       "provenance"

#Load in gplot2 and readr packages
library(ggplot2)
library(readr)

#Read in your .tsv metadata file with read_tsv
metadata=read_tsv("Bat_Metadata_R.tsv")
taxonomy=read_qza("Bat_2021_taxonomy.qza")
head(taxonomy$data) #make sure everything looks OK like it does below
##                          Feature.ID
## 1 a4a5cc927e391a59011c7e017f949dfd
```

```
## 2 9518639833147ad82db09d616f535f2b
## 3 0f268443b2b455b7c1f9dc9874a4c05f
## 4 2bf439a3c30945a7ef125b94d6d95929
## 5 3d09cd886d2a8b01c6ac64d2da2ebbfa
## 6 40f73c7caeb05aaed60dcfce2a90b099
##
Taxon
## 1 d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria;
o__Alteromonadales; f__Pseudoalteromonadaceae; g__Pseudoalteromonas
## 2                          d__Bacteria; p__Proteobacteria;
c__Gammaproteobacteria; o__Vibrionales; f__Vibrionaceae; g__Vibrio

#use qza_to_phloseq to load in more files that you made with or for QIIME
physeq<-qza_to_phyloseq(
  features="Bat_2021_rarefied-table.qza", #rarefied table
  tree="Bat_2021_rooted-tree.qza", #rooted tree
  "Bat_2021_taxonomy.qza", #taxonomy file
  metadata = "Header_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File
- Sheet1.tsv" #metadata file
)
physeq
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 692 taxa and 90 samples ]
## sample_data() Sample Data:       [ 90 samples by 23 sample variables ]
## tax_table()   Taxonomy Table:    [ 692 taxa by 7 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 692 tips and 691 internal nodes ]
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 759 taxa and 34 samples ]
## sample_data() Sample Data:       [ 34 samples by 10 sample variables ]
## tax_table()   Taxonomy Table:    [ 759 taxa by 7 taxonomic ranks ]
## phy_tree()    Phylogenetic Tree: [ 759 tips and 757 internal nodes ]

#Load in phloseq library
library(phyloseq)

#Run ordination, specify what type of UniFrac distance you want
#We used weighted UniFrac because it's the most comprehensive
physeq.ord.wuni <- ordinate(physeq, "PCoA", "unifrac", weighted=T)

#Weighted unifrac PCoA plot for ectoparasite presence
#Ellipse is 95% CI
b.div.wuni <- plot_ordination(physeq, physeq.ord.wuni, type= "samples",
color= "Ectoparasites") + geom_point(size=3)
b.div.wuni <- b.div.wuni + stat_ellipse() + ggtitle("Weighted UniFrac
Distances by Ectoparasite Presence")  + theme_classic() +
scale_color_brewer("Ectoparasites", palette = "Set1")+theme(plot.title =
element_text(hjust = 0.5))
print(b.div.wuni)
```

```
#Ordination by bat species
b.div.wuni <- plot_ordination(physeq, physeq.ord.wuni, type= "samples",
color= "Species") + geom_point(size=3)
b.div.wuni <- b.div.wuni + stat_ellipse() + ggtitle("Weighted UniFrac
Distances by Bat Species")  + theme_classic() + scale_color_brewer("Species",
palette = "Set1")+theme(plot.title = element_text(hjust = 0.5))
print(b.div.wuni)



#Ordination by roost location
#If you have more samples than colors in a color-scheme, use a different set,
check out this website for different color palettes: https://r-graph-
gallery.com/38-rcolorbrewers-palettes.html
b.div.wuni <- plot_ordination(physeq, physeq.ord.wuni, type= "samples",
color= "Site") + geom_point(size=3)
b.div.wuni <- b.div.wuni + stat_ellipse() + ggtitle("Weighted UniFrac
Distances by Roost Location")  + theme_classic() + scale_color_brewer("Site",
palette = "Set1")+theme(plot.title = element_text(hjust = 0.5))
print(b.div.wuni)
## Warning in MASS::cov.trob(data[, vars]): Probable convergence failure
```

APPENDIX B

Protocol: Dissection and DNA extraction of honeybee gut using Qiagen DNeasy Kit

Based off Koch and Schmid-Hempel 2011 Microbial Ecology [Updated June 2017]

Reagents:

5% bleach solution and Sterile MilliQ water for surface sterilization of bees

Sterile molecular water for DNA elution

Lysis buffer

Lysozyme (powder)

Proteinase K from the DNeasy kit (or a preparation of 20 mg/ml proteinase K using autoclaved diH_2O)

Buffer AL, AW1, AW2 included in the DNeasy kit. If this is the first time you are using the kit, make sure you add ethanol to the appropriate buffers as described in the manufacturer's instructions.

Preparing lysis buffer:

-Prepare and autoclave stock solutions of

- Tris-HCl pH 8 1M (calibrate pH with HCl)
- EDTA pH 8 0.5M (calibrate pH with NaOH pellets)

-Prepare and autoclave lysis buffer

- 20mM Tris-HCl pH 8
- 2mM EDTA pH 8
- 1.2% Triton-x-100

Preparing lysing solution: (do this immediately prior to sample collection)

-Measure out lysozyme into a sterile falcon tube; sterilize spatula (ethanol and flame) before transferring powder. Add appropriate amount of buffer (20mg lysozyme per 1ml lysis buffer). Vortex falcon tube.

For 12 samples, make enough for 13 tubes by adding 48 mg lysozyme to 2.4 ml lysis buffer.

For 24 samples, 90 mg lysozyme + 4.5 ml lysis buffer.

For 30 samples, 111.6 mg lysozyme + 5.58 ml lysis buffer.

Prior to extraction:

-Sterilize workspace, dissection tools, and pipettors with 10% bleach solution.

-Prepare sterile 1.5ml tubes with 180 ul of lysis buffer and lysozyme solution.

Dissection

1. To surface sterilize, soak whole bee in 5% bleach solution for 30 sec., followed by three 5 sec. rinses in sterile water (use fresh 1ml aliquots in 1.5ml tubes of bleach and water for each bee). Place bee onto sterile petri dish under dissecting microscope.

2. Remove gut from bee by:

A. Pulling off stinger with GI tract attached. Cut out midgut if interested in this section.

B. If necessary, remove abdomen, cut along sides with micro-scissors, remove ventral cuticle, remove whole or midgut.

**Flame or bleach sterilize tools between each sample.

3. Transfer gut(s) to 1.5ml tube containing 180ul of lysis solution (with lysozyme). (note: I have pooled 3-6 bees/sample, depending on study) [Preliminary studies: 3 bees. Cage/hive studies: 5 bees. Bee Biocide Expt 2016: 6 bees.]

DNA Extraction

-Set incubator to 37C.

4. Grind gut(s) in lysis solution with sterile pestle and mixer for 5 seconds to homogenize solution.

5. Incubate at 37C for 1 hour

6. Reset incubator or thermal block to 56 C.

Add 25 ul proteinase k to each tube.

Add 200 ul buffer AL.

Vortex each sample.

7. Incubate at 56 C for 30 minutes.

While waiting, set up and label filter/collection tubes (from DNeasy kit) and sterilized storage tubes. Place in racks that have been cleaned and bleached.

8. Turn off incubator.

Add 200 ul cold ethanol (100%; maintain in freezer) to each tube. (Ethanol binds to the DNA and prevents it from washing through the filter.)

Vortex 5-10 seconds.

For pools: Centrifuge at 13,000rpm for 2 minutes, transfer supernatant to spin column (next step).

9. Pipette solution into DNeasy mini spin columns (the special filter tubes) that have been placed in the 2 ml collection tubes. Use a large pipette set at ~800 ul. Discard pipette tip every time in between tubes.

   Centrifuge at 8500 rpm for 1 min.

   Discard the liquid in the collection tube, along with the tube. Retain the filter tube.

10. Place mini spin column in a new collection tube.

    Add 500 ul AW1 buffer.

    Centrifuge at 8500 rpm for 1 min; discard collection tube and liquid.

11. Place mini spin column in a new collection tube.

    Add 500 ul AW2 buffer.

    Centrifuge at 14000 rpm for 3 mins; discard liquid. Centrifuge again for 1 min. then discard liquid and collection tube, being careful not to splash liquid up onto filter.

12. Place mini spin column in a clean, sterile 1.5ml storage tube.

    Pipette 100 ul of sterile water directly onto membrane in tube (use 200ul for pools; 100ul for cage/hive studies). Discard pipette tip every time in between tubes.

    Let sit incubating at room temperature for 5 minutes.

    Centrifuge at 8500 rpm for 1 min. Be sure to position vials so caps don’t break.

13. Optional: To get more DNA, pass this 100ul back through the filter, let sit 5 minutes, and centrifuge again for 1 min. ***Do this for amphibian skin swabs, but not bee gut samples.

Store DNA at -20 or -80 for long term.

APPENDIX C

Illumina MiSeq Sequencing Protocol

Adapted from the Earth Microbiome Project (EMP, http://www.earthmicrobiome.org/emp-standard-protocols/)
Edited in August, 2017 by: Jeni Walke, Angie Estrada, Daniel Medina, Jessica Hernandez and Lisa Belden.

Reagents:
UltraClean PCR grade H2O

5 Prime Hot Master Mix

Forward primer IL 515F

Reverse primer + barcode IL 806R

Before beginning:

- Sterilize workspace with 10% bleach solution followed by 70% ethanol. If possible, perform in a hood dedicated to PCR set up. UV hood before using.
- Sterilize pipettors (use pipettors dedicated for PCR reagents and use a separate pipettor for the DNA) with bleach and ethanol or with DNA away.
- Clean and sterilize with bleach 1 large centrifuge tube rack and several small PCR tube racks. Rinse and allow to dry.
- Prepare new labels for all of your tubes if necessary.
- Locate samples and barcodes. Assign samples to barcodes. Keep both in fridge until ready to use.

Step 1: Make your PCR reactions
A) For each sample, you will run triplicate PCR reactions plus a negative control with just water = 4 PCR tubes per sample.

Per sample	5X volume (extra for pipetting)
12 ul UltraClean PCR grade H2O	65 ul
10 ul 5 Prime Hot Master Mix	50 ul
0.5 ul Forward primer IL 515F	2.5 ul
0.5 ul Reverse primer + barcode IL 806R	2.5 ul

B) For samples that might have LOW DNA CONCENTRATIONS, the PCR reactions could be prepared with the same method as above, but with a small change in the volume of the reagents and DNA, aditionally BSA could be added to increase PCR yield as follows:

Per sample		4X Volume
12 ul	Ultra Clean PCR grade H20	48 ul
10 ul	5Prime Hot Master Mix	40 ul
0.5 ul	Forward Primer IL 515F	2.0 ul
0.5 ul	Reverse Primer + barcode IL 806R	2.0 ul
2.0 ul	DNA	6.0 ul (in triplicate)

1. Add all reagents, except DNA, to the each PCR tube in the first row of the plate.

2. Pipette 23 ul from the first row of PCR tubes, with every reagent listed above except DNA, into the negative PCR tubes.

3. Add DNA (6ul) to first replicate. Vortex gently, then centrifuge briefly

4. Take 25 ul from the first row of PCR tubes and add into replicate rows #2 and #3.

5. Centrifuge each PCR tube briefly to eliminate any bubbles.

Step 2: Run reactions in thermocycler
1. Make sure machine is set for 25 ul samples.

2. Thermocycler conditions:

Temp	Time	
94°C	3 min	
94°C	45 sec	
50°C	1 min	35 cycles
72°C	1.5 min	
72°C	10 min	
4°C	hold	

You can maintain your PCR product in the fridge overnight if you need to wait until the next day to run your gel.

Step 3. Run gels to check amplification and negative controls

1. Combine your three separate PCR reactions into a single PCR tube. Use post-PCR pipettors and tips.

2. Make a 1% gel. Combine 100 ml 1X TBE and 1 g agarose in a small Erlenmeyer flask. Microwave until just boiling. Swirl. Continue boiling/swirling until solution is clear.

3. Once the solution has cooled slightly, add 10 ul gel red stain. Note: Gel red stain is light sensitive--keep away from light as much as possible.

4. Pour gel into mold and allow to cool completely.

5. Combine 4 ul PCR product and 2 ul loading dye Pipette up and down to combine.

6. Reset pipettor to 7 or 7.5 ul. Pipette each sample into gel well. As the amount of solution decreases (due to evaporation), you may need to reset your pipette ul setting. Avoid air bubbles in the pipette tip as this will cause the DNA to leak out. Gently pipette solution into wells.

7. Load your ladder. You can use a broad range 50-10,000 bp ladder.

8. Run gel at a voltage of ~160 for approximately 20 minutes, until dye is about halfway.

9. Visualize gels. Bands will be at ~ 300-350 bp. Sample bands may be a little smeary, but there should not be multiple bands. No bands should be visible for the negative controls.

NOTE: If sample bands are very faint (indicating too low or too high DNA content), try the following alternatives (see table):

a. Modify the starting DNA concentration with 1:10 or 1:50 dilutions. Or use ½ of the DNA volume. Dilute in PCR water.

b. Reduce de volume of water (for example: 4ml/sample) and replace with BSA which increases PCR yield (also usefull when bands are not amplyfing) .

c. If the previos does not work, is possible that DNA is too low in which case duplicate the volume of DNA samples (to 4ml) or try to duplicate DNA + BSA

	Original Reaction		BSA only		½ DNA		2XDNA + BSA	
	Per sample	4X Vol	Per sample	4X Vol	Per sample	4X Vol	Per sample	4X Vol
PCR grade H20	12	48	10	40	13	52	9	36
5Prime Hot MasterMix	10	40	10	40	10	40	10	40
Forward IL 515F	0.5	2	0.5	2	0.5	2	0.5	2
Reverse barcode	0.5	2	0.5	2	0.5	2	0.5	2

DNA (3x)	2	6	2	6	1	3	4	12
BSA			2	8			1	4
Total	25	98	25	98	25	98	25	98

NOTE: If there are bands in the negative control for a sample, redo the PCR

Store PCR products at -20 C until you've accumulated all of the samples that you are going to run on a single Illumina plate before moving on to Step 4.

Step 4: Quantifying the DNA
We use a Qubit 2.0 Florometer and the dsDNA High Sensitivity assay kit. Readings can be a bit fickle, so it is better to do all of your samples on the same day at the same time with the same working solution and standards. This can be done on the countertop. Use post-PCR pipettors and tips.

Before beginning:

- Organize your samples in a single PCR tube rack on ice.
- Label florometry tubes supplied by Qubit in a tube rack with sample names, in the same order as they occur in the PCR tube rack.

1. Combine in a 50 ml falcon tube:
Per sample (so multiply by the number of samples you are quantifying, plus your 2 standards, plus a little extra for pipetting)

1 ul Qubit reagent

199 ul Qubit buffer

Vortex. This is your working solution.

2. Make your standards. Combine 10 ul of each standard with 190 ul working solution. Make a separate solution for each standard and combine in the tubes supplied by Qubit.

3. For your samples: Combine 2-5 ul sample with 198-195 ul working solution. Total solution volume should be 200 ul. Make a separate solution for each sample and combine in the florometry tubes that you labeled already. To get the most accurate measurements, it is very important that you get the precise amount of your entire sample into the working solution. Try 2 ul of sample first. If the readings are too low (there's too little DNA), then redo, increasing the amount sample.

4. Vortex and briefly centrifuge all tubes. Drops of liquid stuck on the sides or lids of tubes can mess up the readings.

5. Incubate at room temperature for 2 min.

6. Read tubes in the Florometer. Specify the amount of sample you used (i.e., 2-5 ul). Record reading in ng/ul.

Step 5: Combine equal amounts of amplicons into a single tube
1. Based on the concentration determined by the Florometer, determine how much of each sample you need to add. The goal is to to add the same amount of ng of DNA per sample (~180 ng) into a single, 1.5 ml centrifuge tube.
Example: If Sample 1 has a concentration of 38 ng/ul, you should add 200/38 = 5.3 ul to the pool.

2. Add the appropriate volume of each sample to a single centrifuge tube. This is your pooled sample. Compute the volume of the pooled sample.

Step 6: Clean up pooled sample.
We use the Qiagen QIAquick PCR Clean Up Kit.
If this is the first time you are using the kit, make sure you add ethanol and the PH indicator to the appropriate buffers as described in the manufacturer's instructions.

1. Vortex the pooled sample to thoroughly mix it. Pipette 100 ul of the pooled sampled into a new, clean 1.5 ml centrifuge tube. **Store the remaining, uncleaned pooled sample in storage box in -20C.

2. Add 500 ul of Buffer PB to the 100 ul of your pooled sample. Vortex. Check that the color of the mixture is yellow. If the color of the mixture is orange or violet, add 10 µl of 3 M sodium acetate, pH 5.0, and mix. The color of the mixture will turn to yellow.

3. Place a Qiaquick spin column in a provided 2 ml collection tube.

4. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30–60 s at 13,000 rpm.

5. Discard flow-through. Place the QIAquick column back into the same tube.

6. Wash the pooled sample. Add 0.75 ml Buffer PE to the QIAquick column, let the buffer sit on the filter for 2 min, then centrifuge for 30–60 s at 13,000 rpm.

7. Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min at 13,000 rpm.

8. Place the QIAquick column in a new, clean 1.5 ml centrifuge tube.

9. To elute the DNA, add 50 ul Buffer EB to the QIAquick column, let the buffer sit on the filter for 3 min, then centrifuge for 1 min at 13,000 rpm.

10. Measure the concentration of the cleaned, pooled sample using the Qubit Florometer (as above, but with only one sample) and the 260/280 using the Nanodrop. 260/280 should be between 1.8-2.0.

Step 7: Add PhiX
For running these libraries in the MISeq and HiSeq, you may need to make your sample more complex by adding 30-50% PhiX to your run.

However, the sequencing facility may add PhiX for you. Check with the particular sequencing facility you are using for information about adding PhiX. The sequencing facility that we use (listed below) adds PhiX for you.

Step 8: Send for sequencing!
Keep cleaned, pooled sample frozen until ready to send. Send sample on dry ice.

Sequencing Facility and contact info:

Zach Herbert <zherbert@research.dfci.harvard.edu>
Molecular Biology Core Facilities
Dana Farber Cancer Institute at Harvard
http://mbcf.dfci.harvard.edu/

APPENDIX D

QIIME2 Simplified Workflow: 16S rRNA Sequence Processing

Before you start:

1.) Check and record the QIIME program and version: qiime info

- Copy-paste all information from the terminal after this command, report this in your methods and bioinformatics notebook

2.) Activate the conda environment if your version of QIIME requires this (i.e., after a recent update)

- In my case: conda activate qiime2-2021.11

3.) Set your working directory (the file you will be working from)

- cd <drag and drop your directory file here>

4.) Make sure that your manifest file is in the correct format, location, and that the file paths are all correct

- Importing the data via the manifest file was the fussiest step I encountered, and I think it had to do with using a PC and a Virtual Box to use QIIME, so proceed with caution here

5.) Include your project background at the beginning of you bioinformatics notebook

- Brief description, objectives/hypotheses

Part I – Sequence data import and preparations (importing data, filtering and trimming sequences, aligning sequences, importing mapping file for later analyses)

1.) Import your sequence data into QIIME

- This is often the most difficult step that requires the most troubleshooting
- Gives you **file-name_single-end_demux.qza**

2.) Visualize the imported sequence data in QIIME2 View: https://view.qiime2.org/

- Use the **file-name_single-end_demux.qza** from the import step (#1) in QIIME2 View, convert it to a qzv file (**file-name_single-end_demux.qzv**) and use that in QIIME2 View
- Gives you an idea of how high quality your samples are
- Record the following information from your interactive quality plot and the tables in the overview tab:
    - Total number of sequences
    - Average number of reads per sequence
    - Sample with the highest number of reads
    - Sample with the lowest number of reads
    - The sample with the first big jump in sequence read number
- Determine where to trim your sequences for future steps:
    - By using figaro (command line installation and code if you can get it to work): https://github.com/Zymo-Research/figaro
        - You will need various information on your primers for these commands
    - Or visually via the interactive quality plot: where does the sequence quality begin to drop off? Somewhat subjective, so give it careful thought.

3.) Filter your sequence data

- Using deblur filter by quality score, takes about 40 minutes to run
- Gives you **file_name_demux-filtered.qza** and **file_name_demux-filter-stats.qza**

4.) Visualize your filtered sequences in QIIME2 View

- Convert the **file-name_demux-filtered.qza** to **file-name_demux-filtered.qzv**, use qzv file in QIIME2 View
- Record the same information that you did for your raw unfiltered sequences (#2)
- Compare the filtered and unfiltered stats

    - How many sequences were lost in the filtering step? What is this value as a percentage?
    - If a huge number of sequences were lost, adjust the parameters to find the balance between good quality sequences and having enough sequences for later analyses. I just used the default parameters.
- If the filtered data looks good, use the demux-filtered.qza for future steps from now on

5.) Visualize the filtered stats in QIIME2 View

- Convert **file-name_demux-filter-stats.qzv** to **file-name_demux-filter-stats.qzv**, use qzv in QIIME2 View
- Not a make-or-break step, use this jointly with your visual interpretation of the filtered sequences quality plot (#4)
    - Allows you know how the filtering step worked, but I found it more useful to just look at the quality plot in #4

6.) Run deblur (for single-end sequence data only)

- Use **file-name_demux-filtered.qza** from deblur filter by quality score step (#3)
- This took roughly 24 hours to complete, because your computer is aligning millions of sequences
- Deblur groups and aligns based on sequence similarities and an algorithm, determines if a sequence is a real biological sequence or noise (based on how present that sequence is across the whole dataset)
    - Useful deblur info: https://awbrooks19.github.io/vmi_microbiome_bootcamp/rst/3_sequences_to_composition.html
- Deblur will align your sequences and trim them (if you specified that option)
- Gives you **file-name_rep-seqs-deblur.qza**, **file-name_deblur-stats.qza**, and **file-name_table-deblur.qza**

7.) Visualize the deblur stats from the deblur step – what did deblur do to my sequence data?

- Use the **file-name_deblur-stats.qza** from the deblur step (#6) and convert it to **file-name_deblur-stats.qzv** using the provided code
- Visualize the **file-name_deblur-stats.qzv** table in QIIME2 View
    - Look at the unique reads column – that acts as a pre-indicator of how many unique bacterial species you will most likely have

8.) Visualize the representative sequences from the deblur step

- Checks to see if deblur did what it was supposed to do (i.e., trimming)
- Use the **file-name_rep-seqs-deblur.qza** from deblur (#6) and convert it to **file-name_rep-seqs-deblur.qzv**
- Visualize **file-name_rep-seqs-deblur.qzv** in QIIME2 View
    - Look at sequence length to see if everything was trimmed correctly
    - Check a sequence to make sure the primer sequences were cut out

9.) Visualize and summarize the table from the deblur step

- Allows you to match your sample IDs and other information to the bacterial sequences for that sample. You will run your data analyses on this in later steps
- Use the **file-name_table-deblur.qza** from deblur (#6) and your **metadata-file.tsv**, the code will convert these to **file-name_table-deblur.qzv**
    - Your metadata file (mapping file) must be in a tab separated values (.tsv) file format, otherwise it won't work. I used Google Sheets to export my mapping file as a .tsv
- Use the **file-name_table-deblur.qzv** in QIIME2 View, record the following:
    - How many samples are in the dataset?
    - How many features (bacterial "species") are in the dataset?
    - What is the total frequency (total number of DNA sequences in the dataset)?

- What is the frequency per sample?
- What is the mean frequency per feature (the mean number of sequences assigned to a feature, a.k.a., bacterial "species")?

Part II – Analyses on your sequence data

a.) Preparing your sequences and assigning taxonomy to your sequence data with Silva

1.) Training feature classifier:

- This step creates a new pared-down reference database that QIIME will compare your sequences to.
  - The Silva database has the full 16S rRNA gene (with a lot of sequences and assigned identities), but we only need to look at the V4-V5 region, which is the region specified by our primers (515f, 926r). The database is pared down in the classifier step, so QIIME isn't trying to search the entire database.
- In our case, the training feature classifier step had to be custom-made by Dr. Walke and Shelby Fettig because of the primers we are using (515f, 926r), as QIIME only has the 806r primer pre-installed.
- The series of steps here will give you a qiime classifier file (i.e., the **Shelby Fettig - silva_99_138.1_qiime_classifer.qza** I used) that you will need to assign taxonomy to your sequences (#2).
- **Caution!** → the classifier must be trained using the same version of QIIME that you have installed, otherwise it won't work (because you are using an older version of the classifier that does not match the latest version of the software).
  - Also, the **qiime feature-classifier fit-classifier-naive-bayes** step generally can't be run on a Virtual Box because there isn't enough RAM on it. Dr. Walke ended up running this code on her Mac to get around this. There are ways to upgrade the memory on your Virtual Box, but I didn't want to mess around with that.

2.) Assign taxonomy to your sequences

- Identifies the bacterial "species" found in your samples

- Uses the **file-name_Silva138_515_926_classifier.qza** from the training feature classifier step (#1) and the **file-name_rep-seqs-deblur.qza** from the deblur step (#6, part I).
- This step gives you a **file-name_taxonomy.qza** to use in later steps.

3.) Visualize the taxonomy of your sequences:

- Use **file-name_taxonomy.qza** from taxonomy step (#2), converts it to a **file-name_taxonomy.qzv**, visualize in QIIME2 View
- Will be an interactive table, very useful!
- Use this to determine what you need to filter out in the filtering steps (i.e. chloroplasts, mitochondria, unassigned sequences, etc.) and how to format that in your code.
- Also useful for getting an idea of what types of bacteria are in your sample.

4.) Filter out mitochondria and chloroplasts from your table:

- Removes mitochondria and chloroplasts from your sequence table
- Uses the **file-name_table-deblur.qza** from the deblur step (#6, part I) and the **file-name_taxonomy.qza** from the taxonomy assignment step (#2)
- This step gives you a **file-name_filtered-table.qza** to use in later steps

5.) Check that the mitochondria and chloroplasts were filtered out from your table

- Uses **file-name_filtered-table.qza** the from the table filtering step (#3) and your **mapping/metadata file.tsv** that you used in the deblur visualization (#9, part I) converts it to a **file-name_filtered-table.qzv**
- Upload the **file-name_filtered-table.qzv** into QIIME2 View to visualize your data table, compare it with your unfiltered data table (the **file-name_table-deblur.qzv** from the deblur step):
    - Compare the number of features (ASV or bacterial “species”), make sure that there are fewer in the filtered table

6.) Filter the mitochondria and chloroplasts out of your sequences:

- Removes mitochondria and chloroplasts from your sequences themselves
- Use the **file-name_rep-seqs-deblur.qza** from the deblur step (#6, part I) and the **file-name_taxonomy.qza** from the taxonomy assignment step (#2)
- Gives you **file-name_filtered-rep-seqs.qza** to use in later steps

7.) Remove any contaminants from your sequences

- Examine controls and suspicious taxa, use similar code to #6, see notebook for details

8.) Remove control samples from your dataset

- Use **qiime feature-table filter-samples**, see notebook for details

9.) Visualize the microbiome using taxa bar plots

- Allows you to look at the relative abundance of bacteria on each sample, more useful than the table in #6
- Use the **file-name_filtered-table.qza** from the mitochondria and chloroplast filter table step (#3), the **file-name_taxonomy.qza** from the taxonomy step (#2) and the **mapping/metadata file.tsv** you uploaded earlier
- Gives you **file-name_taxa-bar-plots.qzv**, look at this in QIIME2 View

b.) Generate a multiple sequence alignment (MSA) using MAFFT (Multiple Alignment using Fast Fourier Transform), build phylogenetic tree using fastree, rarefy sequence data

1.) Make a phylogenetic tree for phylogenetic diversity analyses later on

- Aligns your filtered deblur sequences with MAFFT into a phylogenetic tree, will be used in alpha and beta diversity analyses later on. This code gives you many outputs!
- Use **file-name_filtered-rep-seqs.qza** from the mitochondria/chloroplast sequence filter step (#5,a)
- Gives you: **file-name_aligned-rep-seqs.qza**, **file-name_masked-aligned-rep-seqs.qza**, **file-name_unrooted-tree.qza**, and **file-name_rooted-tree.qza**

2.) Rarify your filtered data and phylogenetic tree

- Rarefaction is a way to standardize your tests so the number of sequences does not affect the results. This step will determine the sequence depth in later analyses (i.e., in calculating alpha and beta diversity)
- Use the **file-name_filtered-table.qza** from table filter step (#4,a), the **file-name_rooted-tree.qza** from the phylogenetic tree step (#1) and the **mapping/metadata file.tsv** you've been using

- Gives you **file-name_alpha-rarefaction.qzv**, examine in QIIME2 View and determine the sequence depth that you will use in later analyses
- **Caution! →** this step is extremely fussy about the formatting of your mapping file. Remove all weird characters and messy labels (see notebook for more details).

c.) Actual data analyses

1.) Calculate alpha and beta diversity of your samples

- Use **file-name_rooted-tree.qza** from phylogenetic tree step (#), the **file-name_filtered-table.qza** from table filter step (#1,b), and the **mapping/metadata file.tsv** that you've been using.
    - If you edited your mapping file for the rarefaction step, continue to use that edited mapping file.
- Gives you **file-name_rarefied-table.qza**, view this in QIIME2 View
- Also gives you a **file-name_core-metrics-results folder** (fancy!)
    - Output folder contains alpha diversity metrics (by default: Shannon's diversity index, observed OTU's, Faith's phylogenetic diversity, evenness) and beta diversity metrics (Jaccard, Bray-Curtis, unweighted UniFrac distances, weighted UniFrac distances)

2.) Calculate alpha diversity statistics with nonparametric Kruskal-Wallis tests for categorical variables:

- **Shannon's diversity:** number of species present (richness) and abundance of each species
    - Use **file-name_core-metrics-results/shannon_vector.qza from diversity metrics folder** (#1) and the **mapping/metadata file.tsv** that you've been using
    - Gives you **file-name_core-metrics-results/shannon_significance.qzv**, view in QIIME2 View
- **Observed Features:** number of different ASVs (bacterial "species") in a sample, the richness of different bacterial taxa present

    - Use **file-name_core-metrics-results/faith_pd_vector.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
    - Gives you **file-name_core-metrics-results/observed_otus_significance.qzv**, view in QIIME2 View
- **Faith's Phylogenetic Diversity:** relatedness of bacterial taxa present
    - Use **file-name_core-metrics-results/faith_pd_vector.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
    - Gives you **file-name_core-metrics-results/faith_pd_significance.qzv**, view in QIIME2 View
- **Evenness:** the abundances of each species present
    - Use **file-name_core-metrics-results/evenness_vector.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
    - Gives you **file-name_core-metrics-results/evenness_significance.qzv**, view in QIIME2 View

3.) Calculate alpha diversity statistics using Spearman's rank correlations for continuous variables:

- Use the alpha diversity vector file **(i.e., Bat_2021_MYVO_only_core-metrics-results/shannon_vector.qza) from diversity core metrics folder)**
    - Repeat for all diversity metrics (**Shannon, Faith's phylogenetic diversity, Evenness, Observed Features**)
- Gives you **Bat_2021_MYVO_only_core-metrics-results/shannon_correlation_Spearman.qzv**, visualize with QIIME2 View
- Output includes all continuous variables (i.e., weight, wing damage)

4.) Calculate beta diversity statistics using nonparametric PERMANOVAs for categorical variables:

Bray-Curtis dissimilarity: abundances of bacterial taxa

- Use **file-name_core-metrics-results/bray_curtis_distance_matrix.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
- Specify which column in your metadata file that you want to examine – the "treatment" that you are investigating (i.e., differences in bacterial diversity between site, Bd status, etc.). In my case, ectoparasite presence (the "Ectoparasites" column)
- Gives you **file-name_core-metrics-results/bray_curtis_site_significance.qzv**, view in QIIME2 View as the Principal Coordinates Ordination Plot (PCoA), each point is a sample and the closer the points are to each other, the more similar their microbiomes.

Jaccard similarity: presence or absence of bacterial taxa, comparing microbial composition

- Use **core-metrics-results/jaccard_distance_matrix.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
- Specify which column in your metadata file that you want to examine – in my case, ectoparasite presence (the "Ectoparasites(Y/N)" column)
- Gives you **file-name_core-metrics-results/jaccard_site_significance.qzv**, view in QIIME2 View

Unweighted UniFrac Distances: qualitative measure of bacterial presence/absence

- Use **file-name_core-metrics-results/unweighted_unifrac_distance_matrix.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
- Specify which column in your metadata file that you want to examine – in my case, ectoparasite presence (the "Ectoparasites(Y/N)" column)
- Gives you **file-name_core-metrics-results/unweighted_unifrac_site_significance.qzv**, view in QIIME2 View

Weighted UniFrac Distances: quantitative, more comprehensive, abundance and relatedness of bacterial taxa

- Use **file-name_core-metrics-results/weighted_unifrac_distance_matrix.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
- Specify which column in your metadata file that you want to examine – in my case, ectoparasite presence (the "Ectoparasites(Y/N)" column)
    - Need to repeat for other variables you want to examine

- Gives you **file-name_core-metrics-results/weighted_unifrac_site_significance.qzv**, view in QIIME2 View

5.) Calculate beta diversity statistics using Mantel tests for continuous variables

- Uses beta diversity distance matrix files from core-metrics results folder (i.e., **Bat_2021_core-metrics-results/bray_curtis_distance_matrix.qza**)
- Specify what column in your metadata file you want to examine (i.e., weight, wing damage)
- Gives you **Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qzv**, view in QIIME2 View
    - Repeat for all beta diversity metrics (**Bray-Curtis dissimilarity, Jaccard similarity, unweighted Unifrac distance, weighted UniFrac distance**)
- Visualize with the emperor ordination plots made for the beta diversity analyses
    - Weighted UniFrac plots are the most comprehensive overall, use weighed UniFrac distance matrix to make the ordination plots
    - Can make prettier ordination plots in R, see R notebook for details

6.) Run an indicspecies analysis or linear discriminant analysis to determine if the relative abundances of bacterial taxa differ between your variables

- I did indicspecies in R: see QIIME notebook and R code notebook for details
- If you can get the website to work, do an LDA at the Harvard website (http://huttenhower.sph.harvard.edu/galaxy/), defaults are OK)
- Visualize with grouped taxa bar plots, or if abundances are really small, make individual bar plots or boxplots

7.) Make grouped taxa bar plots to go along with the indicspecies analyses

- Use **qiime feature-table group** and **--p-mode 'mean-ceiling'** to get the average abundances of taxa across all samples, grouped by whatever variable you're interested in (i.e., ectoparasites, bat species, etc.), see notebook for details

APPENDIX E

QIIME2 16S rRNA Data Processing Bioinformatics Notebook

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington state bats

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats to compare their relative abundances sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the most abundant culturable bacterial isolates from western bats will differ from the most abundant bacteria in our sequencing data.

This sequencing data comes from maternity colonies sampled across Washington state from April – June 2021 with the WDFW.

QIIME2 Program and Version: qiime info
(qiime2-2021.11) qiime2@qiime2core2021-2:~$ **qiime info**

System versions
Python version: 3.8.12
QIIME 2 release: 2021.11
QIIME 2 version: 2021.11.0
q2cli version: 2021.11.0

Installed plugins
alignment: 2021.11.0
composition: 2021.11.0
cutadapt: 2021.11.0
dada2: 2021.11.0
deblur: 2021.11.0
demux: 2021.11.0
diversity: 2021.11.0
diversity-lib: 2021.11.0
emperor: 2021.11.0
feature-classifier: 2021.11.0
feature-table: 2021.11.0
fragment-insertion: 2021.11.0
gneiss: 2021.11.0
longitudinal: 2021.11.0
metadata: 2021.11.0
phylogeny: 2021.11.0
quality-control: 2021.11.0
quality-filter: 2021.11.0

sample-classifier: 2021.11.0
taxa: 2021.11.0
types: 2021.11.0
vsearch: 2021.11.0

Application config directory
/home/qiime2/miniconda/envs/qiime2-2021.11/var/q2cli

Getting help
To get help with QIIME 2, visit https://qiime2.org

To activate the latest version of QIIME before you start:
conda activate qiime2-2021.11

2/11/2022: I created my manifest file for our single-end sequencing data and tried to import my data into QIIME2.

Creating your manifest file:

- The headers must be the way they are below, with "sample-id" and "absolute-filepath"
    - Also, QIIME doesn't like underscores (_), dashes (-) or spaces in sample IDs, so I replaced all of those with periods
- To get the absolute file path, I dragged and dropped my zipped fasta files into the QIIME2 terminal, and then copy-pasted those into my manifest for the corresponding sample
- ***The manifest file I am using is in the directory I am working from (my "QIIME" folder) not a subfolder within that directory folder – this is important or else it won't work for the import step!***
- My manifest file was saved as a .csv, as this seems to work well for PCs. It seems that Mac users have some more wiggle room in regards to what file type they save their manifest file as (i.e., .txt), but I've had the best success with a .csv.

	A	B
1	sample-id	absolute-filepath
2	RR.L.10	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL10_JW9664_S1_R1_001.fastq.gz'
3	RR.L.11	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL11_JW9664_S2_R1_001.fastq.gz'
4	RR.L.12	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL12_JW9664_S3_R1_001.fastq.gz'
5	RR.L.13	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL13_JW9664_S4_R1_001.fastq.gz'
6	RR.L.14	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL14_JW9664_S5_R1_001.fastq.gz'
7	RR.L.15	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL15_JW9664_S6_R1_001.fastq.gz'
8	RR.L.16	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL16_JW9664_S7_R1_001.fastq.gz'
9	RR.L.17	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL17_JW9664_S8_R1_001.fastq.gz'
10	RR.L.18	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL18_JW9664_S9_R1_001.fastq.gz'
11	RR.L.19	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL19_JW9664_S10_R1_001.fastq.gz'
12	RR.L.20	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL20_JW9664_S11_R1_001.fastq.gz'
13	S.S.1	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS1_JW9664_S12_R1_001.fastq.gz'
14	S.S.5	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS5_JW9664_S13_R1_001.fastq.gz'
15	S.S.7	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS7_JW9664_S14_R1_001.fastq.gz'
16	S.S.9	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS9_JW9664_S15_R1_001.fastq.gz'
17	S.S.11	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS11_JW9664_S16_R1_001.fastq.gz'
18	S.S.12	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS12_JW9664_S17_R1_001.fastq.gz'
19	S.S.13	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS13_JW9664_S18_R1_001.fastq.gz'
20	S.S.16	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS16_JW9664_S19_R1_001.fastq.gz'
21	S.S.18	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS18_JW9664_S20_R1_001.fastq.gz'
22	S.S.19	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS19_JW9664_S21_R1_001.fastq.gz'
23	S.S.20	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS20_JW9664_S22_R1_001.fastq.gz'
24	WB.T.1	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_WBT1_JW9664_S23_R1_001.fastq.gz'
25	WB.T.2	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_WBT2_JW9664_S24_R1_001.fastq.gz'
26	WB.T.3	'/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_WBT3_JW9664_S25_R1_001.fastq.gz'

Getting started: it's QIIME for QIIME!

1. Activate your conda environment as needed: conda activate qiime2-2021.11
2. Set your working directory: cd <drag and drop your directory folder to the terminal>

Part I – Sequence data import and preparations (importing data, filtering and trimming sequences, aligning sequences, importing mapping file for later analyses)

1.) Importing Sequence Data into QIIME:

For single-end sequence data:

```
qiime tools import \
--type 'SampleData[SequencesWithQuality]' \
--input-path <dragged and dropped manifest file filepath> \
--output-path <file_name_single-end-demux.qza> \
--input-format SingleEndFastqManifestPhred33V2
```

The code I used:

```
qiime tools import \
--type 'SampleData[SequencesWithQuality]' \
--input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv' \
--output-path Bat_2021_single-end-demux.qza \
--input-format SingleEndFastqManifestPhred33V2
```

- QIIME didn't like the .csv. For some reason it was reading the headers as one long smushed word instead of as separate columns.
- Saving the manifest as different file types (text file, tab-delimited) straight from excel did not work. Dr. Magori thinks that when excel exports from a PC, some things are lost in translation and the Linux system on QIIME doesn't like that.

Code and error message:

(qiime2-2021.11) qiime2@qiime2core2021-2:/media/sf_QIIME/QIIME$ qiime tools import --type 'SampleData[SequencesWithQuality]' --input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv' --output-path Bat_2021_single-end-demux.qza --input-format SingleEndFastqManifestPhred33V2

There was a problem importing /media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv:

/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv is not a(n) SingleEndFastqManifestPhred33V2 file:

Found unrecognized ID column name 'sample-id,absolute-filepath' while searching for header. The first column name in the header defines the ID column, and must be one of these values:

Case-insensitive: 'feature id', 'feature-id', 'featureid', 'id', 'sample id', 'sample-id', 'sampleid'

Case-sensitive: '#OTU ID', '#OTUID', '#Sample ID', '#SampleID', 'sample_name'

NOTE: Metadata files must contain tab-separated values.

There may be more errors present in the metadata file. To get a full report, sample/feature metadata files can be validated with Keemei: https://keemei.qiime2.org

Find details on QIIME 2 metadata requirements here: https://docs.qiime2.org/2021.11/tutorials/metadata/

2/15/2022: Dr. Magori and I troubleshooted the manifest file problem in R, got the sequence data imported, and created the .qzv file for the visualization step.

Troubleshooting in R – the code we used:

```
#Load in the dataset without the 'filepath', use data without the single quotations
bat3=read.csv("DC_Bat_2021_Manifest_No_Quote.csv",header=T)
#Fix the column name so they have the dashes
names(bat3)[1]="sample-id"
names(bat3)[2]="absolute-filepath"
#Save the edited file as a tab delimited file from R
write.table(bat3,file="DC_Bat_2021_Manifest_Tab_Fixed.txt",sep="\t",row.names=F,quote=F)
```

For some reason, when exporting the manifest from a PC to the Linux system (the QIIME system), it messes up the formatting and takes away the dashes in the file names. We used R to edit the manifest file and saved it to a tab-delimed file.

- **First, we had to remove all single quotations (') around the file path names in excel. We think that these are added during the transfer between the two systems. So, the 'file paths' were changed to just file paths with no quotations**
- **Then, we edited the column names in R to fix the transfer problem**
- **We saved the manifest as a tab-delimited file. This file format seems to save the changes to the headers, so QIIME doesn't have a problem reading them**

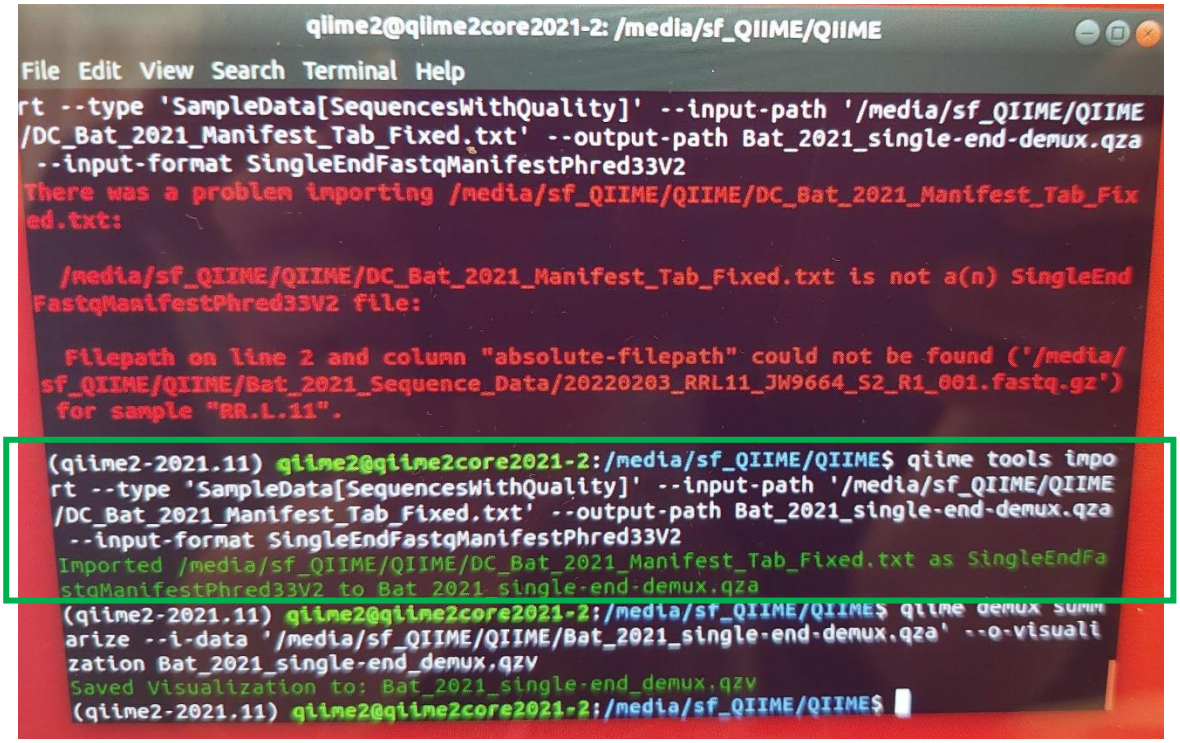
The manifest that worked:

	A	B
1	sample-id	absolute-filepath
2	RR.L.10	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL10_JW9664_S1_R1_001.fastq.gz
3	RR.L.11	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL11_JW9664_S2_R1_001.fastq.gz
4	RR.L.12	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL12_JW9664_S3_R1_001.fastq.gz
5	RR.L.13	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL13_JW9664_S4_R1_001.fastq.gz
6	RR.L.14	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL14_JW9664_S5_R1_001.fastq.gz
7	RR.L.15	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL15_JW9664_S6_R1_001.fastq.gz
8	RR.L.16	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL16_JW9664_S7_R1_001.fastq.gz
9	RR.L.17	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL17_JW9664_S8_R1_001.fastq.gz
10	RR.L.18	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL18_JW9664_S9_R1_001.fastq.gz
11	RR.L.19	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL19_JW9664_S10_R1_001.fastq.gz
12	RR.L.20	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL20_JW9664_S11_R1_001.fastq.gz
13	S.S.1	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS1_JW9664_S12_R1_001.fastq.gz
14	S.S.5	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS5_JW9664_S13_R1_001.fastq.gz
15	S.S.7	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS7_JW9664_S14_R1_001.fastq.gz
16	S.S.9	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS9_JW9664_S15_R1_001.fastq.gz
17	S.S.11	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS11_JW9664_S16_R1_001.fastq.gz
18	S.S.12	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS12_JW9664_S17_R1_001.fastq.gz
19	S.S.13	/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_SS13_JW9664_S18_R1_001.fastq.gz

- We removed the single quotations from the copy-pasted file paths, and then fixed the file in R and saved it as a tab-delimited file
- My manifest was saved in the directory file I was working from, not a subfolder within that directory

1.) Import the Bat 2021 sequence data into QIIME:

qiime tools import \
--type 'SampleData[SequencesWithQuality]' \
--input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt' \
--output-path Bat_2021_single-end-demux.qza \
--input-format SingleEndFastqManifestPhred33V2



Success!

2.) Visualize the sequence data in QIIME: ***use the .qza file from the import step here***

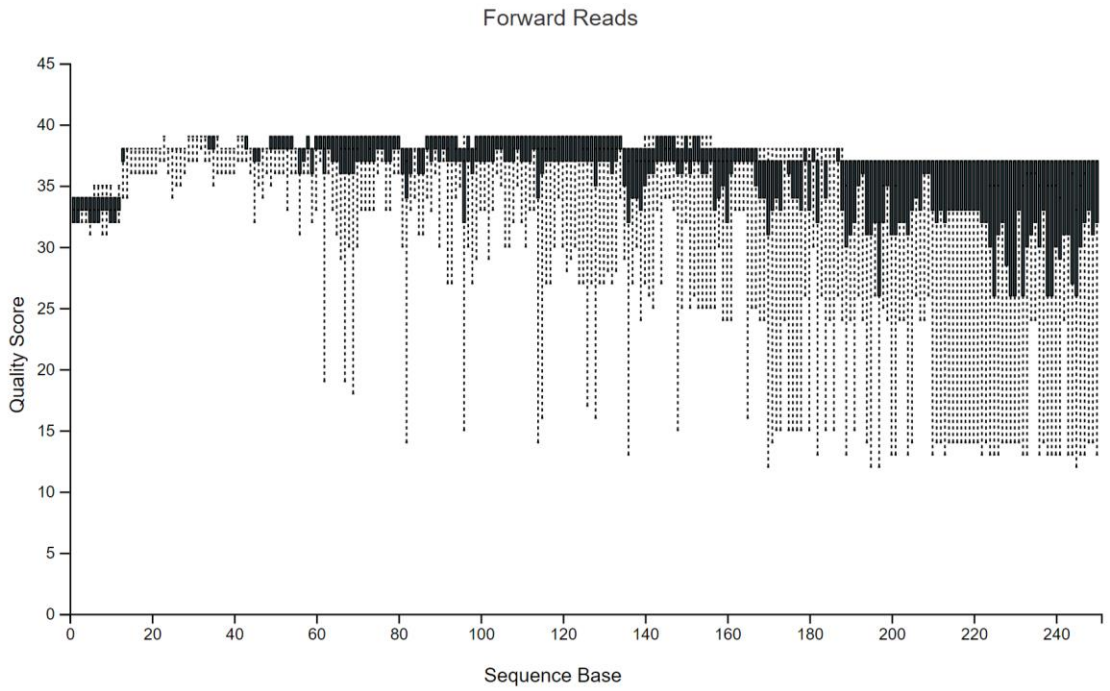
For single-end reads:
qiime demux summarize \
--i-data <drag and drop your single-end demux.qza file from import step here> \
--o-visualization <file_name_single-end_demux.qzv>

The code I used to visualize my single-end sequence data:
qiime demux summarize \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_single-end-demux.qza' \
--o-visualization Bat_2021_single-end_demux.qzv

Use the .qzv file from this step in qiime2view to visualize the sequence quality

3.) QZV File visualization via qiime2view: https://view.qiime2.org/

Quality plot: what a beauty!



- The sequence data quality drops off initially around 140 bases, although it drops off even more around 190 bases
- Total number of sequences: 8,374,027 sequences
- Number of reads per sequence: average of 87,229.45
- Sample with the highest number of reads: OR LS-18, with 154,469 reads
- Sample with the lowest number of reads: LC C-5, with 9,144 reads
    - 2nd lowest: E-C with 11,691 reads
    - 3rd lowest: S_13-19 with 26,453 reads, ~1,500 more sequences than E-C

Where should I trim the sequences for the deblur step?

- **Visually:** left (5') = 18, right (3') = 220
- **Figaro:** https://github.com/Zymo-Research/figaro → **use command line installation and code**
    - Amplicon length = 250 (250 bp long total)
    - Forward primer length (515f) = 19, have barcode tag as well (Parada)
    - Reverse primer length (926r) = 20 (Quince)
    - Path to output → drag and drop a folder for a location, give it a unique name, specify location
    - Figaro ended up not working so we didn't bother with it
- **Quality score below 25? Look at dark lines going down**

What is deblur and what does it do?
Deblur is a quality-control step used to trim and filter single end sequence data by their quality. Sequences that fall below a certain quality threshold are discarded, and low-quality reads are trimmed out.
2/22/2022: we gave up trying to mess with Figaro to figure out where to trim the sequences, so we decided to run the quality control step to see how many sequences were filtered out vs. how many we started with in the original qzv file from the import step.

After this, we will trim the sequences by sight (either no trimming at all because the quality of these sequences is not that bad, or after 220). By sight, we looked at the dark bands coming down from the curve. Most people use a quality score of 25 as the cut-off (which most of ours are above), although using 30 if possible is even better.

I updated my mapping file with all of the information we collected from bat sampling (species, sex, forearm length, etc.) and fixed any formatting issues (i.e., making sure to include periods in the sample ID names instead of spaces or underscores) by checking it with **Keemei** (https://keemei.qiime2.org/). The metadata file was supposedly good when I checked it with Keemei (make sure you copy-paste the contents of the excel sheet, Keemei won't be there unless it's a Google Sheet), but we'll see.

Useful info about deblur and QIIME processing workflow in general:
https://awbrooks19.github.io/vmi_microbiome_bootcamp/rst/3_sequences_to_composition.html

1.) Quality control using Deblur Filter by quality score: gets rid of poor quality sequences based on default parameters *(do this first!).* You can also change up these parameters as needed too. If the default removed too many sequences, go ahead and adjust the parameters (good quality sequences vs. enough sequences for analysis)
Note: the filtering step took around 40 minutes before it was completed

qiime quality-filter q-score \
--i-demux <drag and drop file-name_single-end-demux.qza from import data step> \
--o-filtered-sequences <file_name_demux-filtered.qza> \
--o-filter-stats <file_name_demux-filter-stats.qza> \

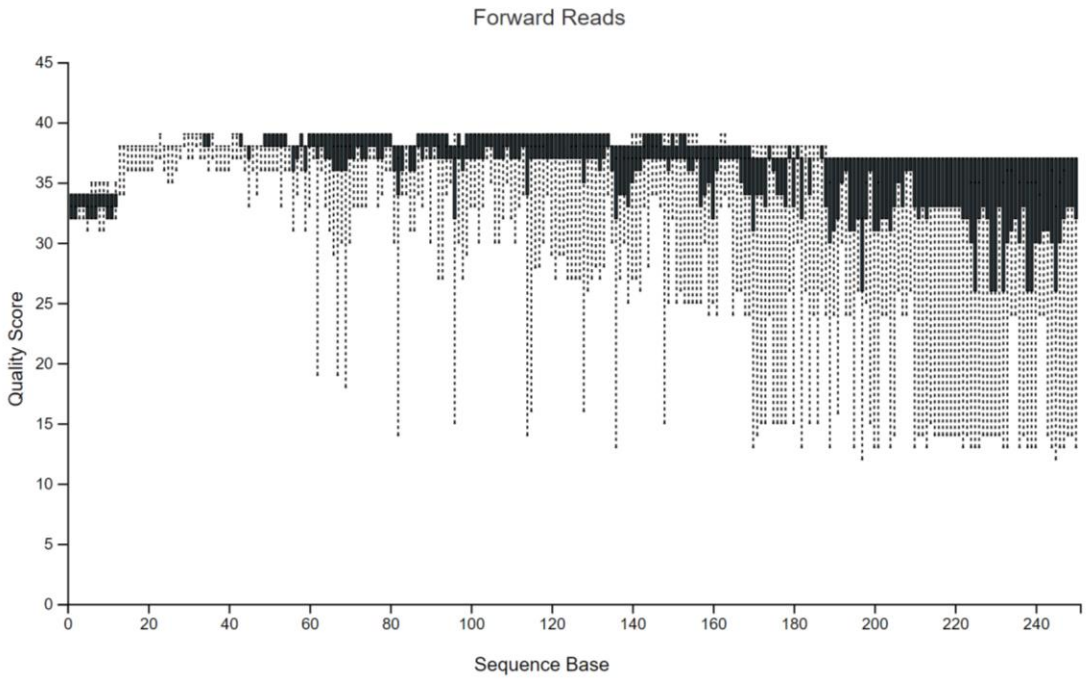
The code I used for deblur filter quality control:
qiime quality-filter q-score \
--i-demux Bat_2021_single-end-demux.qza \
--o-filtered-sequences Bat_2021_demux-filtered.qza \
--o-filter-stats Bat_2021_demux-filter-stats.qza

2.) Visualize the output from the filter quality control step:
qiime demux summarize \
--i-data <drag and drop file-name_demux-filtered.qza from deblur filter quality control step> \
--o-visualization <file-name_demux-filtered.qzv >

The code I used to visualize the demux-filtered qza output file: → ***never use the unfiltered qza file, only use filtered qza as the input for analyses from now on!***

qiime demux summarize \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_demux-filtered.qza' \
--o-visualization Bat_2021_demux-filtered.qzv

Filtered sequences (Bat_2021_demux-filtered.qzv) QZV visualization in qiime2 view:



- Total number of sequences: 8,372,516 reads (vs. 8,374,027 unfiltered)
  - The filter step filtered out 1,511 sequences (lost ~1.8 x 10^-4% of total sequences, kept 99.9% of sequences)
- Number of reads per sequence: average of 87,213.71 reads (vs. 87,229.45 unfiltered)
- Sample with the highest number of reads: OR LS-18, with 154,445 reads (vs. 154,469 unfiltered)
- Sample with the lowest number of reads: LC C-5, with 9,144 reads (same as unfiltered)
  - 2nd lowest: E-C with 11,686 reads (vs. 11,691 unfiltered)
  - 3rd lowest: S_13-19 with 26,449 reads (vs. 26,453 unfiltered), ~1,500 more sequences than E-C

No crazy amounts of sequences were removed, so it's OK to stick with using the filtered data made with the default settings (the demux-filtered.qza) from here on out

<u>3.) Visualize the filtered stats -</u> visualize the sample list, how quality filtering step worked (use this in addition to the quality plot comparisons)

qiime metadata tabulate \
--m-input-file <file-name_demux-filter-stats.qza> \
--o-visualization <demux-filter-stats.qzv> \

Note: QIIME doesn't like it when you try to turn the filter stats into a qzv file – it will not work

Quality filtering: total # of sequences that were filtered out (total # or as a percentage)

How to look up a command to see the options (parameters) and info:
<The command you're interested in> **--help**
To look up info on qiime deblur denoise:
qiime deblur denoise-16S --help

2/25/2022: I ran deblur on my sequence data, trimming out the first 12 bases to cut out that lower quality section in the beginning, which may be due to residual primers [not actual bat sequences]. I also visualized the qza output from deblur, checking to see that the areas we wanted trimmed actually got trimmed out.
Note: I started deblur around 7:30 PM, was still not done when I checked again around 10:30 PM
Update: I don't know if QIIME pauses when the computer sleeps or not, but regardless this process wasn't finished even after overnight. In the future, remember that this process will take a very long time so plan accordingly.
Update on the update: the deblur processed finished by 1:15 PM the following day (2/26/2022), so don't hold your breath for future deblur runs.

4.) Run the deblur process: **check sequence length summary in QIIME2 view**
Deblur groups based on sequence similarity and an algorithm (figures out if sequence is real biological sequence vs. an error based on how present it is across the whole dataset).
qiime deblur denoise-16S \
--i-demultiplexed-seqs <drag and drop your demux-filtered qza file> \
--p-trim-length 250 \
--p-left-trim-len 12 \ → **this code is optional, trims out the first 12 bases (5' end, the left side)**
--o-representative-sequences <file-name_rep-seqs-deblur.qza> \
--o-table <file-name_table-deblur.qza> \
--p-sample-stats \
--o-stats <file-name_deblur-stats.qza>

The code I used to run deblur on my filtered sequence data:
qiime deblur denoise-16S \
--i-demultiplexed-seqs '/media/sf_QIIME/QIIME/Bat_2021_demux-filtered.qza' \
--p-trim-length 250 \
--p-left-trim-len 12 \
--o-representative-sequences Bat_2021_rep-seqs-deblur.qza \
--o-table Bat_2021_table-deblur.qza \
--p-sample-stats \
--o-stats Bat_2021_deblur-stats.qza

Look at summary and quality plot again to verify that it did what we wanted it to do, extra primers on the left on explaining that drop?

3/1/2022: I visualized the deblur stats to see what the deblur step did to my sequence data via the deblur stats table. I also ran the qiime metadata tabulate on the filtered data just because. I imported my mapping (metadata) file after some troubleshooting. Turns out the secret lies with Google Sheets.

5.) Visualize the deblur stats – what did the deblur step do to my sequence data?
qiime deblur visualize-stats \
--i-deblur-stats <drag and drop the deblur-stats.qza file from previous step> \
--o-visualization <matched-file-name_deblur-stats.qzv>

The code I used to visualize the deblur stats:
qiime deblur visualize-stats \
--i-deblur-stats '/media/sf_QIIME/QIIME/Bat_2021_deblur-stats.qza' \
--o-visualization Bat_2021_deblur-stats.qzv

Deblur stats table in qiime2 view: visualize with the deblur-stats.qzv

	sample-id	reads-raw	fraction-artifact-with-minsize	fraction-artifact	fraction-missed-reference	unique-reads-derep	reads-derep	unique-reads-deblur	reads-deblur	unique-reads-hit-artifact	reads-hit-artifact	unique-reads-chimeric	reads-chimeric	unique-reads-hit-reference	reads-hit-reference	unique-reads-missed-reference	reads-missed-reference
0	S.13.19	26449	0.531135	0.000076	0.019319	1644	12403	213	7873	1	2	3	5	120	7545	7	152
1	S.S.7	48805	0.471673	0.000082	0.000145	2344	25789	175	13902	2	4	17	103	99	13682	1	2
2	E.C	11686	0.412117	0.000000	0.000843	807	6870	157	4759	0	0	7	13	93	4619	1	4
3	HH.M.14	56198	0.376188	0.000000	0.000511	2548	35057	179	19745	0	0	25	190	120	19460	1	10
4	WB.T.17	75594	0.375757	0.000079	0.003247	3722	47195	298	26439	3	6	36	259	148	25870	2	85
5	LC.C.6	93765	0.374937	0.000128	0.001641	4938	58621	291	29547	6	12	46	294	111	28952	3	48
6	S.S.16	61199	0.370513	0.000098	0.000743	2915	38530	180	21678	3	6	18	149	100	21389	1	16
7	CL.KL.4	37433	0.370128	0.000294	0.002069	2286	23589	220	13305	5	11	45	255	75	12850	1	27
8	POM.PO.5	69020	0.366517	0.000116	0.000000	3602	43731	200	23653	4	8	23	185	94	23287	0	0
9	CL.KL.5	94526	0.366174	0.000254	0.001929	4740	59937	305	31484	12	24	50	372	118	30784	3	60
10	CC.O.8	83761	0.365994	0.000143	0.002574	4021	53117	324	29515	6	12	52	377	115	28751	3	75
11	TL.KL.1	95882	0.365293	0.000188	0.000415	4737	60875	314	31784	9	18	65	427	126	31121	2	13
12	POM.PO.20	93994	0.365034	0.000319	0.000505	4775	59713	338	31947	15	30	38	266	133	31316	2	16
13	S.S.9	100995	0.364582	0.000139	0.007923	5011	64188	479	36552	7	14	82	455	176	35307	2	286
14	HH.M.8	86884	0.363312	0.000219	0.002559	4807	55337	336	28434	9	19	37	295	133	27765	4	72

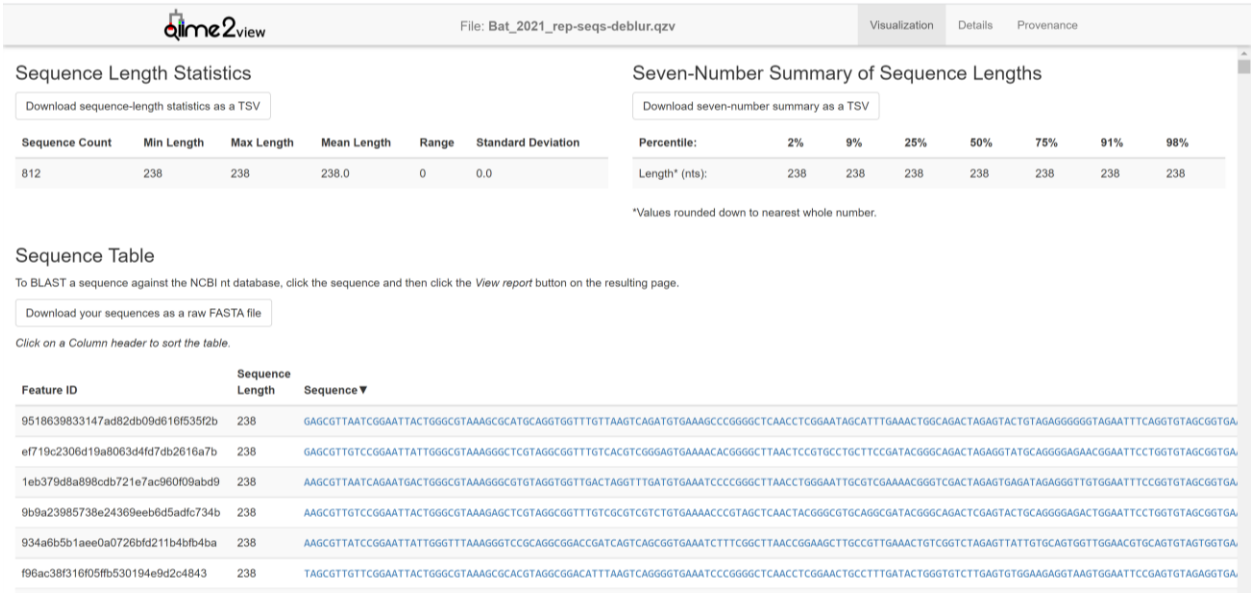
- The unique sequences can act as preliminary indicators of how many unique bacterial species each sample will probably have
- Chimeric sequences are sequence hybrids
- Deblur uses the quality info to do the deblur step, so you can't see the quality plot after deblur runs, you only have your straight sequences without the quality information

6.) Visualize the representative sequences from the deblur step: will check and see if deblur did what it was supposed to do
qiime feature-table tabulate-seqs \
--i-data <drag and drop the rep-seqs-deblur.qza from deblur step> \
--o-visualization <file-name_rep-seqs-deblur.qzv>

The code I used to visualize the representative sequences:
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \

--o-visualization Bat_2021_rep-seqs-deblur.qzv

Deblur representative sequences qza visualization:



qiime2view File: Bat_2021_rep-seqs-deblur.qzv — Visualization | Details | Provenance

Sequence Length Statistics

Download sequence-length statistics as a TSV

Sequence Count	Min Length	Max Length	Mean Length	Range	Standard Deviation
812	238	238	238.0	0	0.0

Seven-Number Summary of Sequence Lengths

Download seven-number summary as a TSV

Percentile:	2%	9%	25%	50%	75%	91%	98%
Length* (nts):	238	238	238	238	238	238	238

*Values rounded down to nearest whole number.

Sequence Table

To BLAST a sequence against the NCBI nt database, click the sequence and then click the *View report* button on the resulting page.

Download your sequences as a raw FASTA file

Click on a Column header to sort the table.

Feature ID	Sequence Length	Sequence▼
9518639833147ad82db09d616f535f2b	238	GAGCGTTAATCGGAATTACTGGGCGTAAAGCGCATGCAGGTGGTTTGTTAAGTCAGATGTGAAAGCCCGGGGCTCAACCTCGGAATAGCATTTGAAACTGGCAGACTAGAGTACTGTAGAGGGGGGTAGAATTTCAGGTGTAGCGGTGA
ef719c2306d19a8063d4fd7db2616a7b	238	GAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTTTGTCACGTCGGGAGTGAAAACACGGGGCTTAACTCCGTGCCTGCTTCCGATACGGGCAGACTAGAGGTATGCAGGGGAGAACGGAATTCCTGGTGTAGCGGTGA
1eb379d8a898cdb721e7ac960f09abd9	238	AAGCGTTAATCAGAATGACTGGGCGTAAAGGGCGTGTAGGTGGTTGACTAGGTTTGATGTGAAATCCCCGGGCTTAACCTGGGAATTGCGTCGAAAACGGGTCGACTAGAGTGAGATAGAGGGTTGTGGAATTTCCGGTGTAGCGGTGA
9b9a23985738e24369eeb6d5adfc734b	238	AAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCGTCTGTGAAAACCCGTAGCTCAACTACGGGCGTGCAGGCGATACGGGCAGACTCGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGA
934a6b5b1aee0a0726bfd211b4bfb4ba	238	AAGCGTTATCCGGAATTATTGGGTTTAAAGGGTCCGCAGGCGGACCGATCAGTCAGCGGTGAAATCTTTCGGCTTAACCGGAAGCTTGCCGTTGAAACTGTCGGTCTAGAGTTATTGTGCAGTGGTTGGAACGTGCAGTGTAGTGGTGA
f96ac38f316f05ffb530194e9d2c4843	238	TAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGACATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCTCGGAACTGCCTTTGATACTGGGTGTCTTGAGTGTGGAAGAGGTAAGTGGAATTCCGAGTGTAGAGGTGA

- Looking at the sequence length statistics, we see that deblur did indeed trim out the 12 base pairs at the 5' end like we wanted (350 – 12 = 238)
- We copy-pasted a sequence into a new word document to make sure that the following primer sequences were not included in the final sequence (via "find"):
    - **Illumina 5' adapter:** AATGATACGGCGACCACCGAGATCTACACGCT
    - **Forward primer pad:** TATGGTAATT
    - **515F forward primer (Parada):** GTGYCAGCMGCCGCGGTAA → Klaatu *parada* nikto

7.) Visualize and summarize the table from the deblur step: your metadata file (mapping file) will be imported in this step! This is what you will run your analyses on in later steps.

qiime feature-table summarize \
--i-table <drag and drop the table-deblur.qza file from the deblur step> \
--o-visualization <matched-file-name_table-deblur.qzv> \
--m-sample-metadata-file <drag and drop your metadata file in .tsv file format>

Note: your metadata file needs to be in a .tsv file format. Do this using Google Sheets

The code I used to (try and) visualize and summarize the deblur table:

qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--o-visualization Bat_2021_table-deblur.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_DC_Mapping-File_TSV.txt'

Error code: something is wrong with our mapping file format

There was an issue with loading the file /media/sf_QIIME/QIIME/Updated_DC_Mapping-File_TSV.txt as metadata:

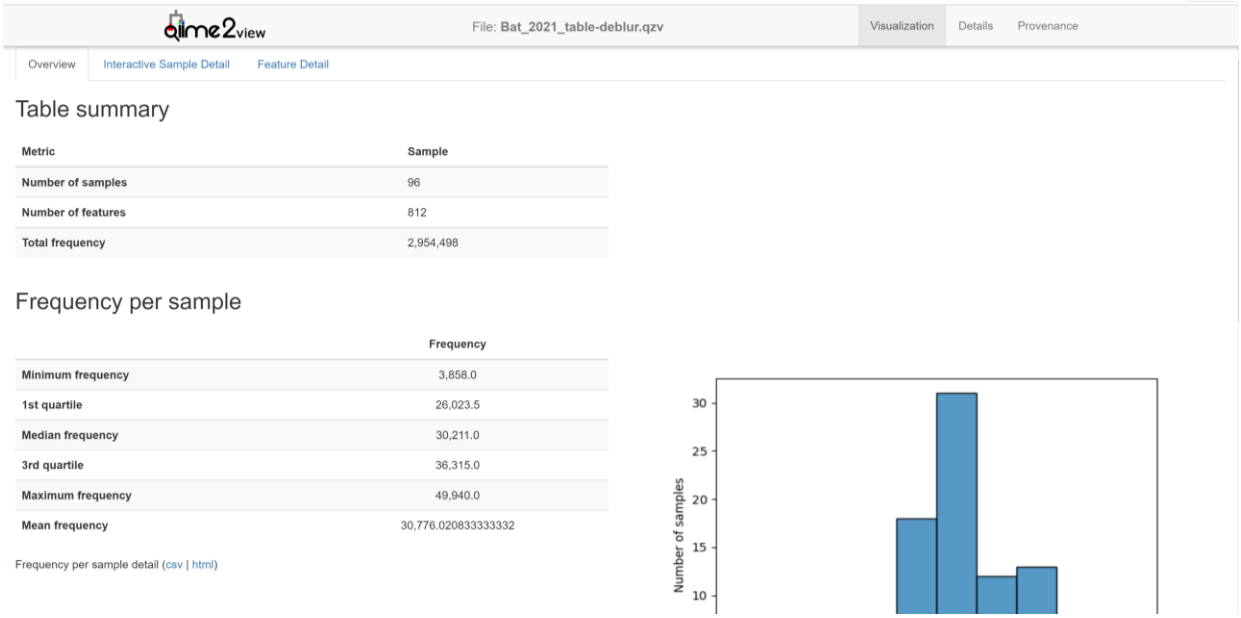
There was an issue with loading the metadata file:
Detected empty metadata ID. IDs must consist of at least one character.
There may be more errors present in the metadata file. To get a full report, sample/feature metadata files can be validated with Keemei: https://keemei.qiime2.org
Find details on QIIME 2 metadata requirements here:
https://docs.qiime2.org/2021.11/tutorials/metadata/

Update: I fixed the problem with my mapping file format (in your face QIIME!):

- Open a new Google Sheets spreadsheet
- Copy-paste all content from your master mapping file (your original excel sheet) into the Google Sheet
- Check it again with the Keemei extension
- File → download → as .tsv (tab separated values) → save to your QIIME folder

The code that actually worked for the visualize/summarize deblur table step:
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--o-visualization Bat_2021_table-deblur.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'

Summarized deblur-table qzv in qiime2 view:



- How many samples are in the dataset? → **96 samples**
- How many features (bacterial "species") are in the dataset? → **812**
- What is the total frequency (total number of DNA sequences in the dataset)? → **2,954,498 total sequences in the dataset**
- What is the frequency per sample? → **[mean] 30,776 per individual bat**

- What is the mean frequency per feature [scroll down!] (the mean number of sequences assigned to a feature, a.k.a., bacterial "species")? → **3,638.54 per bacterial "species"**

Part II – Analyses on your sequence data

3/4/2022: I (tried to) assign taxonomy to my samples using the training feature classifier (qza file) made by Dr. Walke and Shelby Fettig.

a.) Preparing your sequences and assigning taxonomy to your sequence data

Note: the training feature classifier step tells QIIME where to search within the Silva database when assigning taxonomy to your samples, instead of trying to search throughout the whole database. In our case, the training feature classifier step had to be custom-made by Dr. Walke and Shelby Fettig because of the primers we are using (515f, 926r), as QIIME only has the 806r primer pre-installed.

Luckily, we were able to skip the training feature classifier steps since the qiime classifier qza was already made the **Shelby Fettig - silva_99_138.1_qiime_classifer.qza** file. I've included the training feature classifier code anyways in orange below for future reference, but I did not run through this code first-hand.

Update: just kidding, we weren't lucky. We ended up having to run through the training feature classifier steps because the old classifier file was out of date.

The reference database that we will compare our sequences to (the classifier file we uploaded into the taxonomy step) has been trimmed for our primers (515f, 926r). The Silva database has the full 16S rRNA gene (with a lot of sequences and assigned identities), but we only need to look at the V4-V5 region, which is the region specified by out primers (515f, 926r). The database is pared down in the classifier step so QIIME isn't trying to search the entire database.

1.) Training feature classifier:

```
qiime tools import \
--type 'FeatureData[Sequence]' \
--input-path silva_132_99_16S.fna \
--output-path silva_132_99_16S.qza

qiime tools import \
--type 'FeatureData[Taxonomy]' \
--input-format HeaderlessTSVTaxonomyFormat \
--input-path taxonomy_7_levels.txt \
--output-path <file-name_ref-taxonomy.qza>

qiime feature-classifier extract-reads \
--i-sequences silva_132_99_16S.qza \
--p-f-primer GTGYCAGCMGCCGCGGTAA \  → 515f primer sequence: GTGYCAGCMGCCGCGGTAA
--p-r-primer CCGYCAATTYMTTTRAGTTT \  → 926r primer sequence: CCGYCAATTYMTTTRAGTTT
--p-trunc-len 250 \
```

--p-min-length 250 \
--p-max-length 500 \
--o-reads <file-name_ref-seqs.qza>

qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads <drag and drop the ref-seqs.qza from the previous step> \
--i-reference-taxonomy <drag and drop the ref-taxonomy.qza> \
--o-classifier **silva-132-515-926-nb-classifier-PC.qza → Shelby's qza file**

2.) Assign taxonomy to your sequences with Silva database:
qiime feature-classifier classify-sklearn \
--i-classifier <silva-138-515-926-nb-classifier-PC.qza from training feature classifier steps> \
--i-reads <file-name_rep-seqs-deblur.qza from deblur step> \
--o-classification <file-name_taxonomy.qza>

The code I (tried) to use to assign taxonomy to my samples:
qiime feature-classifier classify-sklearn \
--i-classifier '/media/sf_QIIME/QIIME/Shelby Fettig - silva_99_138.1_qiime_classifer.qza' \
--i-reads '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--o-classification Bat_2021_taxonomy.qza

Error code: the pre-made classifier (Shelby's qza) appears to be outdated and therefore incompatible with my most current version of QIIME2. Looking around the QIIME2 forum, people either downloaded the most recent version of a ready-made classifier to use, or they had to re-classify to generate a most recent classifier file.
Plugin error from feature-classifier:

The scikit-learn version (0.23.1) used to generate this artifact does not match the current version of scikit-learn installed (0.24.1). Please retrain your classifier for your current deployment to prevent data-corruption errors.

Debug info has been saved to /tmp/qiime2-q2cli-err-otdrlgyz.log

3/11/2022: I reclassified the training feature to generate an updated qza file classifier in the Silva database, using the code chunks previously in orange, starting from the extract reads step. I had to download the silva sequences qza file (**silva-138-99-seqs.qza**) and the reference taxonomy file (**silva-138-99-tax.qza**) from the QIIME2 website (https://docs.qiime2.org/2021.11/data-resources/).
Note: the extract reads step takes a long time. This step took just under 5 hours to run.

References for downloaded files from QIIME2:
For the sequence reference database:
Michael S Robeson II, Devon R O'Rourke, Benjamin D Kaehler, Michal Ziemski, Matthew R Dillon, Jeffrey T Foster, Nicholas A Bokulich. RESCRIPt: Reproducible sequence taxonomy reference database management for the masses. bioRxiv 2020.10.05.326504;
doi: https://doi.org/10.1101/2020.10.05.326504

For SILVA in general:

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl. Acids Res. 41: D590 – D560

For the taxonomic framework:

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

Training feature classifier: We were able to skip the first few chunks of code since QIIME already had the input files we needed. There were a few tricky parts regarding where to trim the sequences in the qiime feature-classifier extract-reads step. Visit the training feature classifier pages on the QIIME2 website (https://docs.qiime2.org/2022.2/tutorials/feature-classifier/) for more details if you will be training your own feature classifier in the future. The main takeaways though:

- Be very careful with this step!
- The most conservative option was to cut out the --p-trunc-len parameter and set --p-min-length and --p-max-length at 0 to turn them off. This is what I did, to avoid cutting out any good sequences.
- The --p-min-length and --p-max-length exclude amplicons that are far outside of the length you expect with the primers you used, because those are likely non-target sequences and should be excluded. When altering these codes, make sure to select settings that are appropriate for your marker genes!
- The --p-trunc-len parameter should only be used to trim the reference sequences if the search sequences are trimmed to the same length or shorter. If you don't know enough about this, do not mess with it!
    - It's tricky because single end reads may be variable in length.
- For untrimmed single-end reads, QIME2 recommends training a classifier on sequences that have been extracted at the corresponding primer sites (in our case, 515f and 926r) but not trimmed.

qiime feature-classifier extract-reads \
--i-sequences <drag and drop the silva_132_99_16S.qza from Silva website> \
--p-f-primer GTGYCAGCMGCCGCGGTAA \ **→ 515f primer sequence**
--p-r-primer CCGYCAATTYMTTTRAGTTT \ **→ 926r primer sequence**
--p-min-length 0 \ **→ 0 turns this off, --p-trunc-len code line removed too**
--p-max-length 0 \ **→ 0 turns this off**
--o-reads <file-name_ref-seqs.qza>

qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads <drag and drop the file-name_ref-seqs.qza from the previous extract reads step> \
--i-reference-taxonomy <drag and drop the ref-taxonomy.qza from Silva website> \ **→ Note: this is the reference taxonomy file. The new taxonomy file you create in the taxonomy assignment step will be used in your later analyses**
--o-classifier <file-name_silva-132-515-926-nb-classifier-PC.qza>

The code I used to train the feature classifier starting from the extract reads step:
qiime feature-classifier extract-reads \
--i-sequences '/media/sf_QIIME/QIIME/silva-138-99-seqs.qza' \

--p-f-primer GTGYCAGCMGCCGCGGTAA \
--p-r-primer CCGYCAATTYMTTTRAGTTT \
--p-min-length 0 \
--p-max-length 0 \
--o-reads Extract_Reads_ref-seqs.qza → **this step was successful**

qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads '/media/sf_QIIME/QIIME/Extract_Reads_ref-seqs.qza' \
--i-reference-taxonomy '/media/sf_QIIME/QIIME/silva-138-99-tax.qza' \
--o-classifier Updated_2022_silva-132-515-926-nb-classifier-PC.qza → **Unsuccessful, not enough RAM on virtual box to run this**

Error message for qiime feature-classifier fit-classifier-naive-bayes step:
Plugin error from feature-classifier:
Unable to allocate 1.00 GiB for an array with shape (134217728,) and data type float64

It looks like there's not enough RAM in my virtual box to complete this step. The QIIME forum suggested using a pre-trained classifier or running the code on another computer (i.e., not a virtual box) with more memory. I don't know how easy it would be to upgrade the virtual box memory, so it seems like having someone else with a Mac run this chunk with the input files (**Extract_Reads_ref-seqs.qza** and **silva-138-99-tax.qza**) would be simplest. Sigh.

3/13/2022: Dr. Walke ran the qiime feature-classifier fit-classifier-naive-bayes on her Mac using the **Extract_Reads_ref-seqs.qza** that I made earlier and the **silva-138-99-tax.qza** from the QIIME website. The process took 5 hours total to run. Using the updated classifier file from Dr. Walke's run (**Updated_Mar2022_Silva138_515_926_classifier.qza**), I tried to assign taxonomy to my sequence data.

Note: the classifier must be trained using the exact same version of QIIME that you have installed, otherwise it won't work in your assign taxonomy step (because you are using an older version of the classifier that does not match the latest version of the software). Dr. Walke made the classifier on her 2020 version of QIIME (I have the 2022 version) and it was not compatible with mine (the most recent 2022 version):

Plugin error from feature-classifier:

The scikit-learn version (0.22.1) used to generate this artifact does not match the current version of scikit-learn installed (0.24.1). Please retrain your classifier for your current deployment to prevent data-corruption errors.

Debug info has been saved to /tmp/qiime2-q2cli-err-q1ineznd.log

3/14/2022: Dr. Walke assigned taxonomy to my samples by running Shelby's classifier on her matching version of QIIME (my version was too recent to match). I filtered the mitochondria and chloroplasts from my table and converted it to a qzv.

The code Dr. Walke used to assign taxonomy (see #2 for more detail):
qiime feature-classifier classify-sklearn \
--i-classifier /Users/jwalke/Desktop/Dana\ QIIME\
13Mar22/ShelbyFettig_Silva_99_138.1_qiime_classifer.qza \
--i-reads /Users/jwalke/Desktop/Dana\ QIIME\ 13Mar22/Bat_2021_rep-seqs-deblur.qza \
--o-classification Bat_2021_taxonomy.qza

3.) Visualize the taxonomy.qzv file before you try to filter out the mitochondria and chloroplasts. See #6 for the code and the entry note on 3/15/2022 for details on why you should do this.

3.) Filter mitochondria and chloroplasts out of your table:
qiime taxa filter-table \
--i-table <drag and drop the table-deblur.qza from deblur step>\
--i-taxonomy <drag and drop the file-name_taxonomy.qza from taxonomy assignment step> \
--p-exclude D_4__Mitochondria,D_3__Chloroplasts,D_3__Chloroplast,Unassigned \
--o-filtered-table <file-name_filtered-table.qza> → **make sure that this works! Verify that these were removed**

The code I tried to use to filter mitochondria and chloroplasts from my deblur table:
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude <Whatever needs to be filtered out from your data, i.e., mitochondria, chloroplasts, unassigned, etc.> \
--o-filtered-table Bat_2021_filtered-table.qza

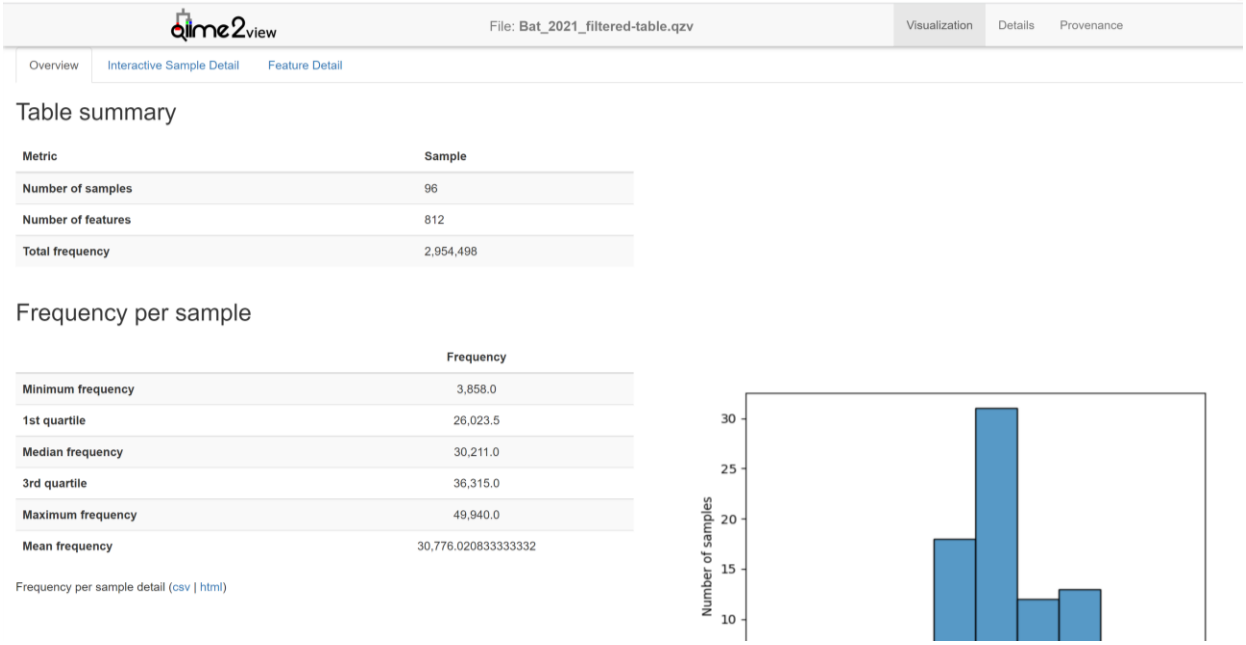
Open output table from that step (filtered table), convert to qzv and open it to check taxonomy
Compare # of features in filtered steps, vs. deblur filtered table → make sure that there are fewer, 1-5 different features (ASV, bacterial "spp") usually matched chloroplasts, mitochondria, unassigned (things that are not bacteria or archaea but made it into the sequence data anyways)

4.) Check that the mitochondria and chloroplasts were filtered out from your table:
qiime feature-table summarize \ → gives qzv file that you can look at, can see your sample list, bacteria, abundance
--i-table <file-name_filtered-table.qza from step #3> \
--o-visualization <matched-file-name_filtered-table.qzv> \
--m-sample-metadata-file <the mapping/metadata file you used earlier.tsv>

The code I tried to use to check that the mitochondria and chloroplasts were filtered from my deblur table:
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_filtered-table.qza' \
--o-visualization Bat_2021_filtered-table.qzv \

--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'

Summarized filtered-table.qzv in QIIME2 View:



qiime2 view — File: Bat_2021_filtered-table.qzv — Visualization | Details | Provenance

Overview | Interactive Sample Detail | Feature Detail

Table summary

Metric	Sample
Number of samples	96
Number of features	812
Total frequency	2,954,498

Frequency per sample

	Frequency
Minimum frequency	3,858.0
1st quartile	26,023.5
Median frequency	30,211.0
3rd quartile	36,315.0
Maximum frequency	49,940.0
Mean frequency	30,776.020833333332

Frequency per sample detail (csv | html)

- How many samples are in the dataset? → **96 samples** → **96 samples**
- How many features (bacterial "species") are in the dataset? → **812** → **812**
- What is the total frequency (total number of DNA sequences in the dataset)? → **2,954,498 total sequences in the dataset** → **2,954,498**
- What is the frequency per sample? → **[mean] 30,776 per individual bat** → **30,776**
- What is the mean frequency per feature [scroll down!] (the mean number of sequences assigned to a feature, a.k.a., bacterial "species")? → **3,638.54 per bacterial "species"** → **3,638.54**

The filtering step did not work this first time around. All values are identical to the unfiltered sequences and tables, so something needs to be adjusted.

Update: the parameters I used in the --p-exclude line were specifically for Phillip's data set since I used his code as a template, so these parameters were not in my dataset. **The --p-exclude parameters are specific to your dataset, so look at your taxonomy.qza file first to see what you will need to filter out.**

Otherwise, if something is not present the default is that QIIME filters nothing, explaining why nothing happened this first time around.

5.) Filter mitochondria and chloroplasts out of your sequences:
qiime taxa filter-seqs \
--i-sequences <file-name_rep-seqs-deblur.qza from deblur step> \ → from sequences too!
--i-taxonomy <file-name_taxonomy.qza from taxonomy step> \
--p-exclude < Whatever needs to be filtered out from your data, i.e., mitochondria, chloroplasts, unassigned, etc.> \
--o-filtered-sequences <file-name_filtered-rep-seqs.qza>

The code I tried to use to filter the mitochondria and chloroplasts from my sequences:
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude D_4__Mitochondria,D_3__Chloroplasts,D_3__Chloroplast,Unassigned \
--o-filtered-sequences Bat_2021_filtered-rep-seqs.qza → **unsuccessful, re-run**

The code I used to visualize the representative sequences:
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--o-visualization Bat_2021_rep-seqs-deblur.qzv

3/15/2022: Dr. Walke and I troubleshooted the mitochondria/chloroplast filtering for the table and sequences, so we got those correctly filtered out. I visualized the taxonomy of my sequences in both the table form (#6) and as taxa bar plots (#7), in addition to making the phylogenetic tree and (finally) getting through the alpha rarefaction step.

When filtering out the mitochondria and chloroplasts, you need to **look in your taxonomy.qza (#6) first to determine what needs to be removed and match the formatting exactly in --p-exclude.** However, when we did this, QIIME didn't like our matched formatting (i.e., g__Chloroplast), probably because of the double underscore. We ended up having to do this as a two-step process for filtering the chloroplasts and mitochondria from the table and sequences, because QIIME didn't like it when we tried to do this all at once (i.e., Mitochondria,Chloroplast), probably because QIIME wanted a semicolon(;) not a comma. I'm not sure if using a semicolon would allow you to do it all in one fell swoop.

Mitochondria are usually from the host (in our case, from the bats) while the chloroplasts are generally from the plants in the environment (i.e., bats roosting in trees). Both are not part of the microbiome and need to be removed before any analyses.

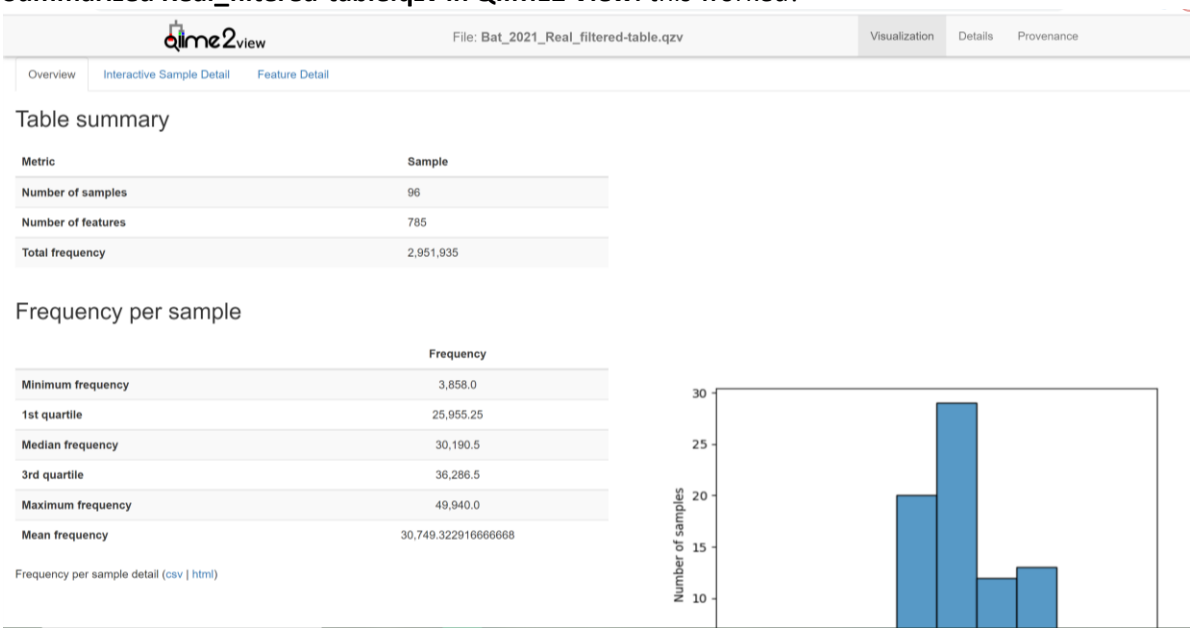
The code I used to [actually] filter mitochondria and chloroplasts from my deblur table:
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Mitochondria \
--o-filtered-table Bat_2021_Real_filtered-table.qza

qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Chloroplast \
--o-filtered-table Bat_2021_Real_filtered-table.qza

→ Use Bat_2021_Real_filtered-table.qza from now on

The code I used to check that the mitochondria and chloroplasts were filtered from my deblur table:
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--o-visualization Bat_2021_Real_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'

Summarized Real_filtered-table.qzv in QIIME2 View: this worked!



- How many samples are in the dataset? **→ 96 samples → 96 samples**
- How many features (bacterial "species") are in the dataset? **→ 812 → 785**
- What is the total frequency (total number of DNA sequences in the dataset)? **→ 2,954,498 total sequences in the dataset → 2,951,935**
- What is the frequency per sample? **→ [mean] 30,776 per individual bat → 30,749.32**
- What is the mean frequency per feature [scroll down!] (the mean number of sequences assigned to a feature, a.k.a., bacterial "species")? **→ 3,638.54 per bacterial "species" → 3,760.42**

The code I used to [actually] filter the mitochondria and chloroplasts from my sequences:
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Mitochondria \

3/17/2022: Looks like I used the wrong input file for this step. I should have used the Real_filtered from the first chunk like I did for the table. See the code on 3/17/2022 for

--o-filtered-sequences Bat_2021_Real_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences **'/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza'** \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Chloroplast \
--o-filtered-sequences Bat_2021_Real_filtered-rep-seqs.qza

The code I used to visualize my Bat_2021_Real_filtered-rep-seqs.qza:
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-rep-seqs.qza' \
--o-visualization Bat_2021_Real_filtered-rep-seqs.qzv

→ Use Bat_2021_Real_filtered-rep-seqs.qza from now on

6.) Visualize the taxonomy of your sequences: ***Do this before the mitochondria/chloroplast filtering step in the future.***
qiime metadata tabulate \
--m-input-file <file-name_taxonomy.qza> \
--o-visualization <matched-file-name_taxonomy.qzv>

The code I used to visualize taxonomy of my sequences:
qiime metadata tabulate \
--m-input-file '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--o-visualization Bat_2021_taxonomy.qzv

Taxonomy visualization in QIIME2 View: very useful file to refer to!

Feature ID #q2:types	Taxon categorical	Confidence categorical
0022c6c55acbfd79fa62454e779c7b05	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Micrococcales; f__Micrococcaceae	0.999711283031692 8
00559626b6cc8e90b491e52ada9c8b4f	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__Brevundimonas	0.999922686265839 1
0091c580b2ea1f8baeed911f2c37a7ae	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae	0.996056082347
00c70932b5b33a1b0f57574566d8d404	d__Bacteria; p__Actinobacteriota; c__Actinobacteria	0.966716790409169 1
00d71527af4c51a79a34f999e9c1fd94	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Beijerinckiaceae; g__1174-901-12	0.999769666376181 9
0116c1f55934b0812fec333f0b92be65	d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Sphingobacteriales; f__Sphingobacteriaceae	0.999999226496163 8
017f8dd33cf8407bf60b7547bfee8473	d__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Stenotrophomonas	0.997805496992278 2
02139f4d0f7e7b5be8177d9d88aa4233	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Micrococcales; f__Microbacteriaceae; g__Leucobacter	0.846866575235863 4
02571eb270b90bfcb92ccf0872b78120	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Corynebacteriales; f__Dietziaceae; g__Dietzia	0.966366842642684 4
025a0ac1904f0f54f3e79797c08a5ba5	d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae	0.837199000784817 1
028701138db649b94390028d42be3dbc	d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Tannerellaceae; g__Macellibacteroides	0.999898116601128 9
02963118ec968a39e9c0137467cce329	d__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae	0.825502058896677 6
02a3c23986d33b79bf077811a244bd9b	d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacterales; f__Acetobacteraceae; g__Endobacter; s__Acetobacteraceae_bacterium	0.728412395346656 4
02c54d0cf51f5ff190be0c045f85fd7e	d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Corynebacteriales; f__Corynebacteriaceae	0.998921632503804 6

- Sift through your data and determine what needs to be filtered out:
    - We had one mitochondrion and ~30 chloroplasts, and no unassigned sequences. The fewer number of features in our Bat_2021_Real_filtered-table.qzv confirms that everything was filtered out correctly.
- Use the search bar at the top of the page to search for anything you need to remove, or fun and wiley bacteria to focus on in your discussion. From a quick glance through this table, we found:

- Family Chitinophagaceae: antifungal bacteria that break down chitin, the primary component of fungal cell walls!
- *Rhodococcus* spp: antifungal bacteria, show up a lot in our lit and widely studied as a naturally-occurring bat skin bacterium being developed as a probiotic for Pd. We appear to have a lot of *Rhodococcus.*
- *Pseudomonas* spp: antifungal bacteria, show up a lot in our lit and widely studied as a naturally-occurring bat skin bacterium being developed as a probiotic for Pd. We appear to have a few *Pseudomonas.*
- *Vibrio* spp: Well-known contaminant of DNA extraction kits. Generally kit contamination is even more pronounced in cases where the amount of host bacteria are low (as in our case with the bat skin bacteria).

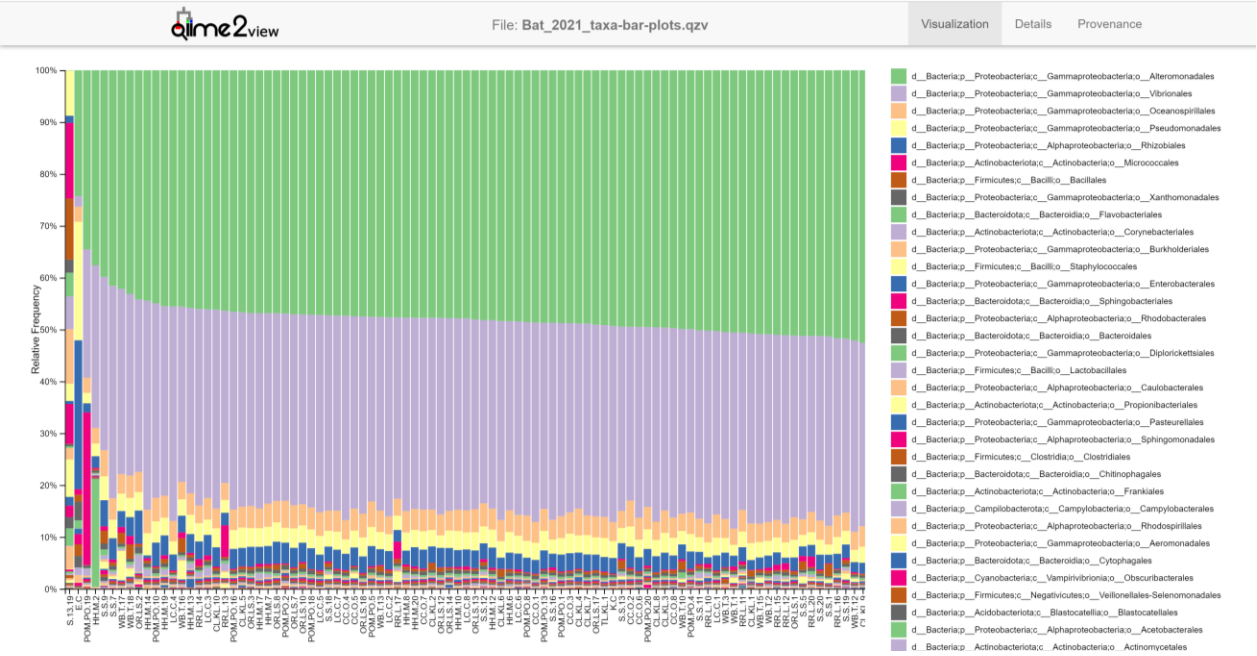
7.) Visualize the microbiome using taxa bar plots:

qiime taxa barplot \
--i-table <file-name_filtered-table.qza from filter table step> \
--i-taxonomy <file-name_taxonomy.qza from taxonomy step> \
--m-metadata-file < the mapping/metadata file you used earlier.tsv > \
--o-visualization <file-name_taxa-bar-plots.qzv>

The code I used to make taxa bar plots:

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv' \
--o-visualization Bat_2021_taxa-bar-plots.qzv

Taxa bar plots in QIIME2 View:



We think the orders Alteromonadales (light green) and Vibrionales (lavender) are attributed to contamination from our Qiagen DNEasy extraction kits. Apparently kit contamination straight from the factory is not uncommon.

b.) Create phylogenetic tree and rarefy sequence data

1.) Generate a tree for phylogenetic diversity analysis

qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences <file-name_filtered-rep-seqs.qza from sequence filter step> \
--o-alignment <file-name_aligned-rep-seqs.qza> \
--o-masked-alignment <file-name_masked-aligned-rep-seqs.qza> \
--o-tree <file-name_unrooted-tree.qza> \
--o-rooted-tree <file-name_rooted-tree.qza>

The code I used to build a phylogenetic tree:

qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_unrooted-tree.qza \
--o-rooted-tree Bat_2021_rooted-tree.qza

2.) Alpha rarefaction plotting → check plots together

qiime diversity alpha-rarefaction \
--i-table <file-name_filtered-table.qza from table filter step> \
--i-phylogeny <file-name_rooted-tree.qza from phylogen. tree step> \
--p-max-depth <The median frequency per sample in your filtered-table.qza > \
--m-metadata-file <the mapping/metadata file you used earlier.tsv > \
--o-visualization <file-name_alpha-rarefaction.qzv>

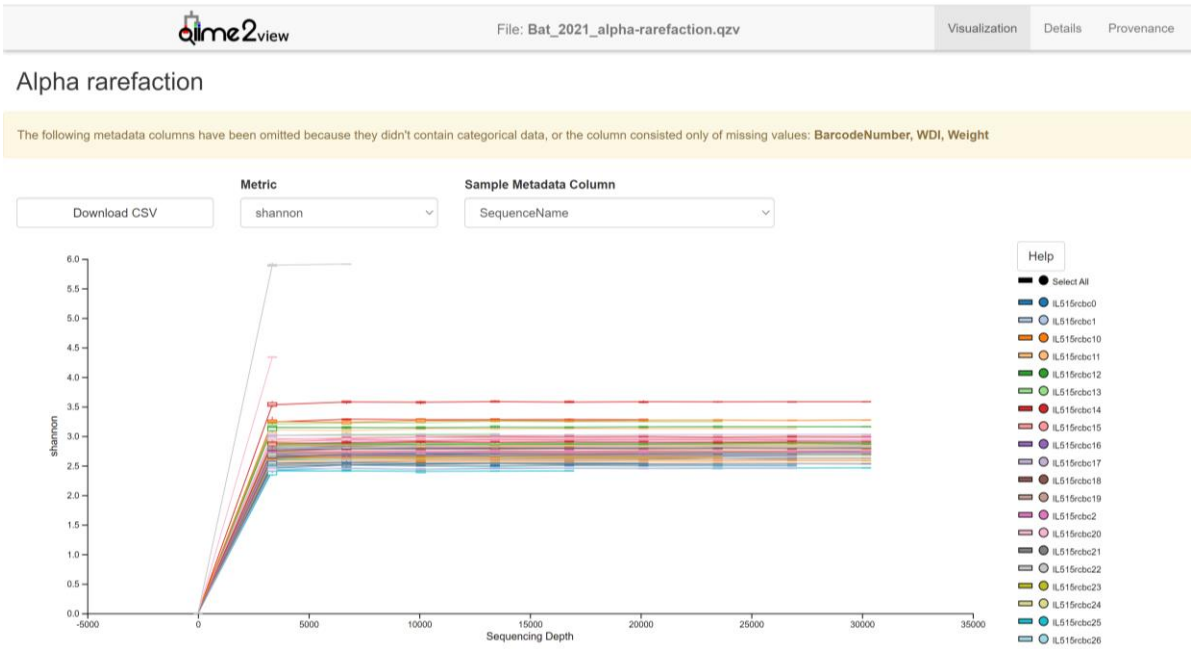
The code I used to plot alpha rarefaction, after much frustration:
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_rooted-tree.qza' \
--p-max-depth 30191 \ → **was 30,190.5, so I rounded to 30191**
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_alpha-rarefaction.qzv

Word of warning on this sucker: this step was extremely fussy. For whatever reason QIIME did not like the forward slashes (/) or periods in my header names in my mapping file, even though using that same exact file was fine up to this point. What I ended up having to do:

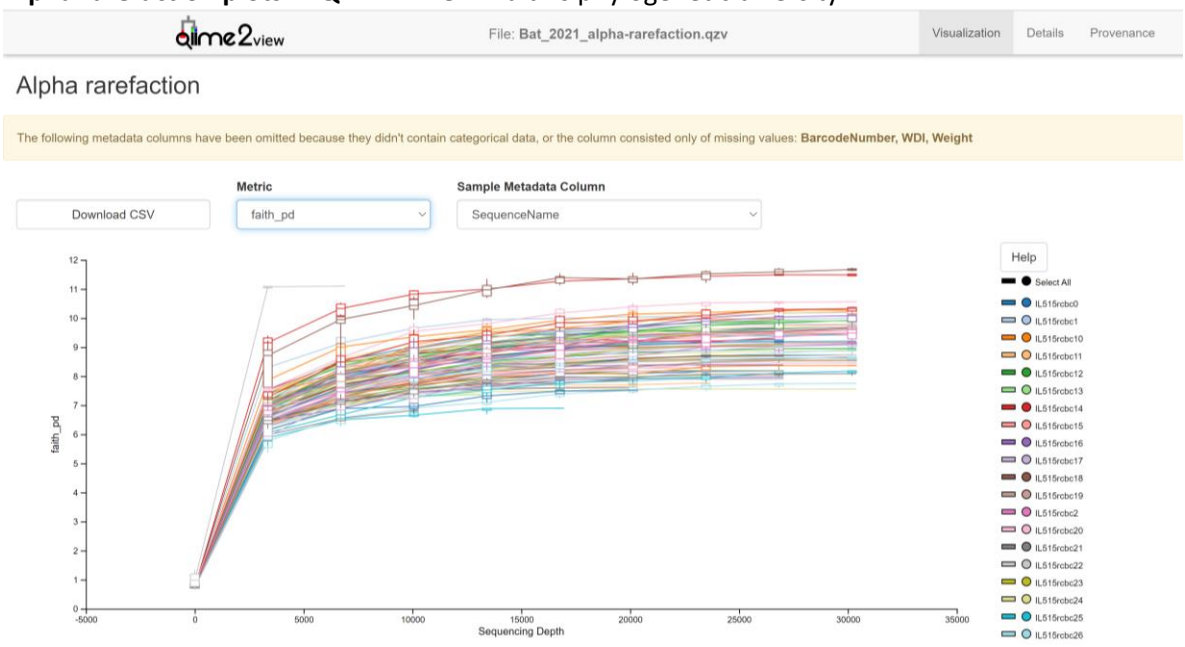
- I made another Google Sheets copy of my original mapping file Google Sheet that I used to make the .tsv (step #7, part I). It was too difficult to edit the .tsv mapping file directly without messing up the entire file formatting.
- In the copy, I removed all forward slashes and periods from my header titles. I also cleaned them up by removing all the information in parentheses, since that would show up on the figures. The information for each column can be looked up on the master excel sheet mapping file.
- I then exported this as a .tsv and dragged the file from my "downloads" folder into my "QIIME" folder so that the .tsv format would stay the same. If you try to open the .tsv after downloading and then "save as" to your directory sometimes it changes it to a text file, rendering it absolutely useless.

→ **Use the Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv mapping file from now on**

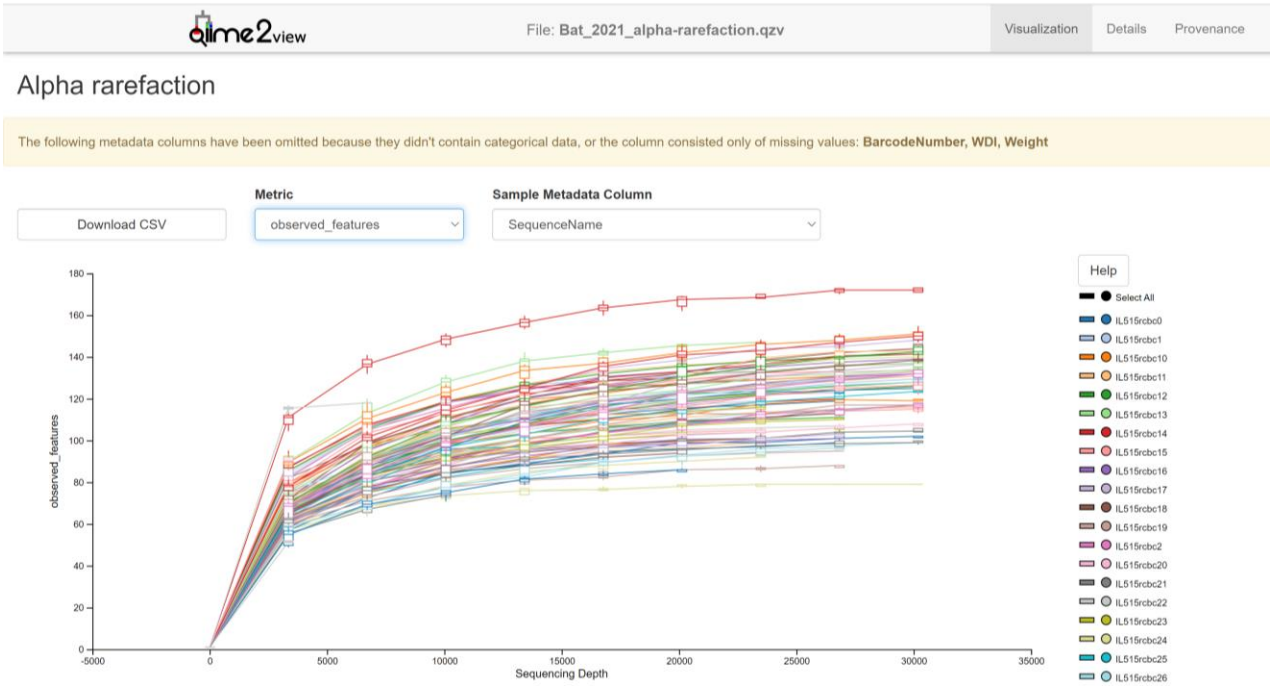
Alpha rarefaction plots in QIIME2 View: Shannon's



Alpha rarefaction plots in QIIME2 View: Faith's phylogenetic diversity



Alpha rarefaction plots in QIIME2 View: Observed features



- From the three alpha rarefaction plots, we decided to rarefy our sequences at a sequencing depth of

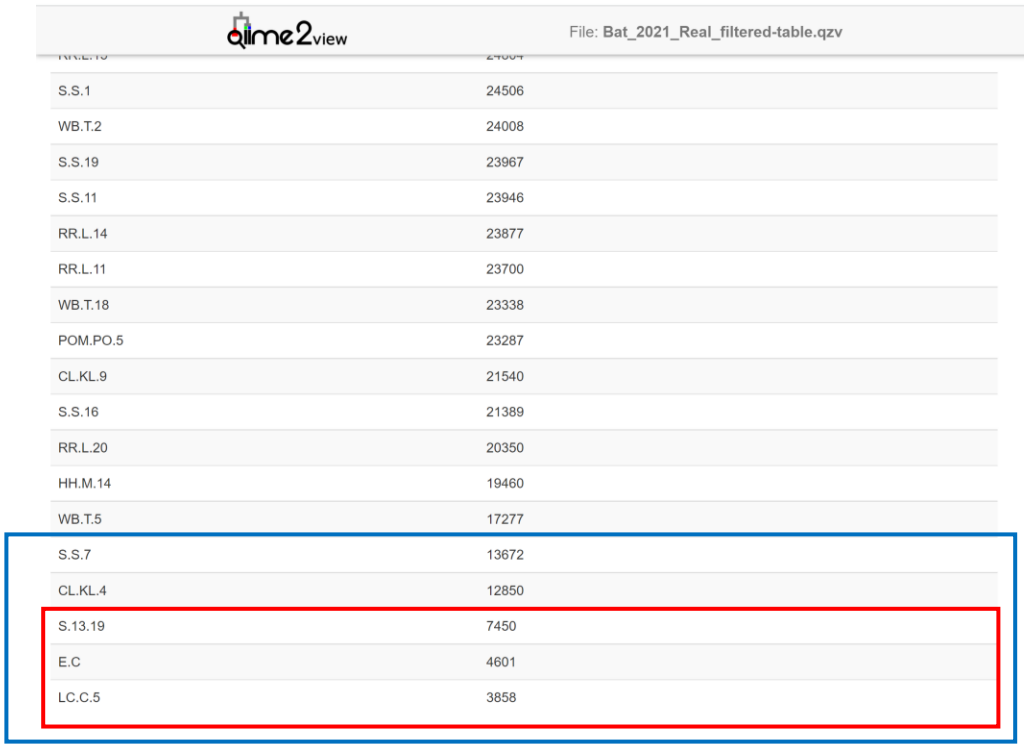
3/17/2022: Dr. Walke and I examined my alpha rarefaction plots, the Real_filtered table list of sequences with the number of features (sequences) and the taxa bar plot (downloaded as a .csv) to determine where to rarefy our sequences. Surprise, there's more complications.

Looking at the alpha rarefaction plots:

- We looked at where the lines began to level off, where there were no more big jumps from sequence depth to sequence depth.
- Ideally, we'd want to rarefy our sequences at the highest depth, but in doing so we'd lose many of our samples. Looking for that sweet spot with the highest sequence depth possible but still have enough sample types represented to answer the study question is the tricky part.
- Across the faith's phylogenetic diversity and observed_features plots, 12,700 and 17,000 appear to be decent areas to clip everything at. The Shannon's is level regardless so it doesn't matter what we do there.

Looking at the Real_filtered-table.qzv:

- Rarefying at 12,700 would remove the LC.C.5, the environmental control (E.C.), and Shelby's control sample (S.13.19).
- Rarefying at 17,000 would remove the three samples above, plus S.S.7 and CL.KL.4.
    - I don't think it would be huge deal to increase our sequence depth to 17,000 and sacrifice two more samples. The Spokane samples were all pretty comparable (all yuma myotis, all females, all no ectoparasites), as were the Klickitat samples (all big brown bats, all females, most had ectoparasites) so removing these samples should not influence our ability to answer our study question.



Looking at the taxa bar plot:

- The three most abundant taxa are likely contaminants:
  - *Pseudoalteromonas, Vibrio, Halomonas*
  - The first two that are listed are well-known contaminants, while the *Halomonas* is a salt-loving bacterium so it makes sense that it would be a contaminant in something salty like our buffer. The large proportions of these taxa are likely not real members of the bat skin microbiome.
  - We will filter out these contaminants, similar to the chloroplast/mitochondria filtering step. After that, we will look at the updated taxa bar plot and re-rarefied plots and make a decision.

The code I used to filter out the contaminants from our table: I had to do this step by step, I couldn't figure out how to do it all at once.

Remember: when you do the filtering one by one, make sure that for the second thing you filter out that you use the output from the previous step, not the original file (otherwise you are only filtering out one thing, not two). Don't do what I did the first time I tried to filter out mitochondria and chloroplasts from my sequences. I redid all my filtering for the sequences just to be sure everything got out.

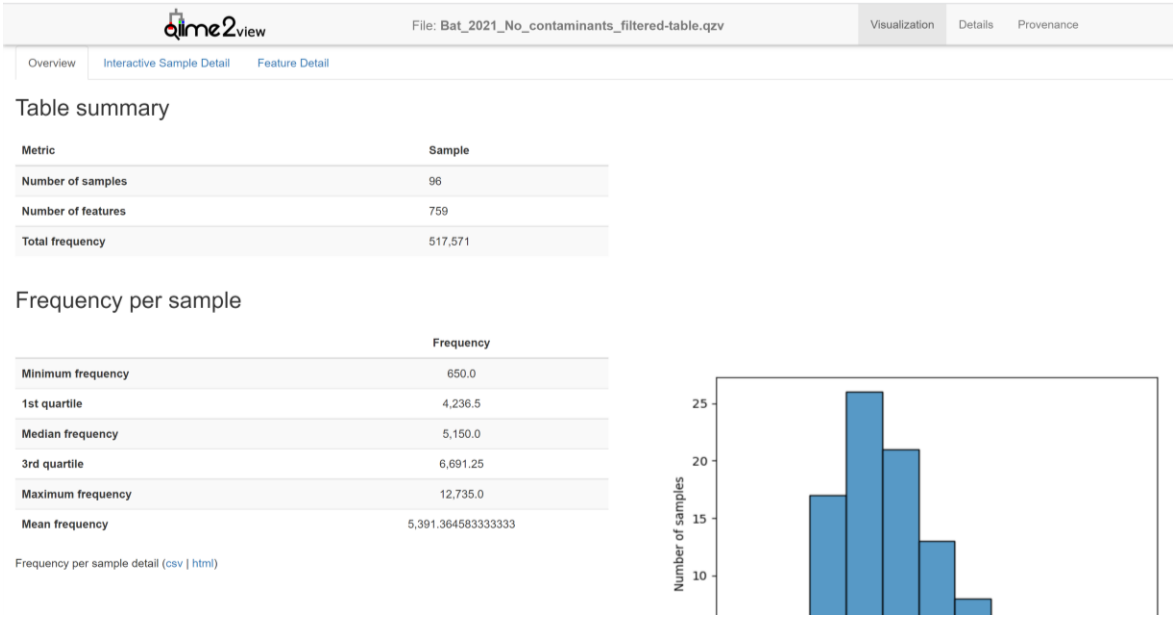
1.) The code I used to filter out the contaminants from our table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Vibrio \
--o-filtered-table Bat_2021_No_contaminants_filtered-table.qza
```

qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \ **← output from previous filter step**
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Halomonas \
--o-filtered-table Bat_2021_No_contaminants_filtered-table.qza

qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Pseudoalteromonas \
--o-filtered-table Bat_2021_No_contaminants_filtered-table.qza

→ Bat_2021_No_contaminants_filtered-table.qza is the most updated table at this point



2.) The code I used to check that the contaminants were filtered from the table:
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--o-visualization Bat_2021_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'

3.) The code I used to filter out the contaminants from our sequences (I re-did mitochondria and chloroplasts too just to be sure they were removed):
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Mitochondria \

--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \ **← output from previous filter step**
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Chloroplast \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

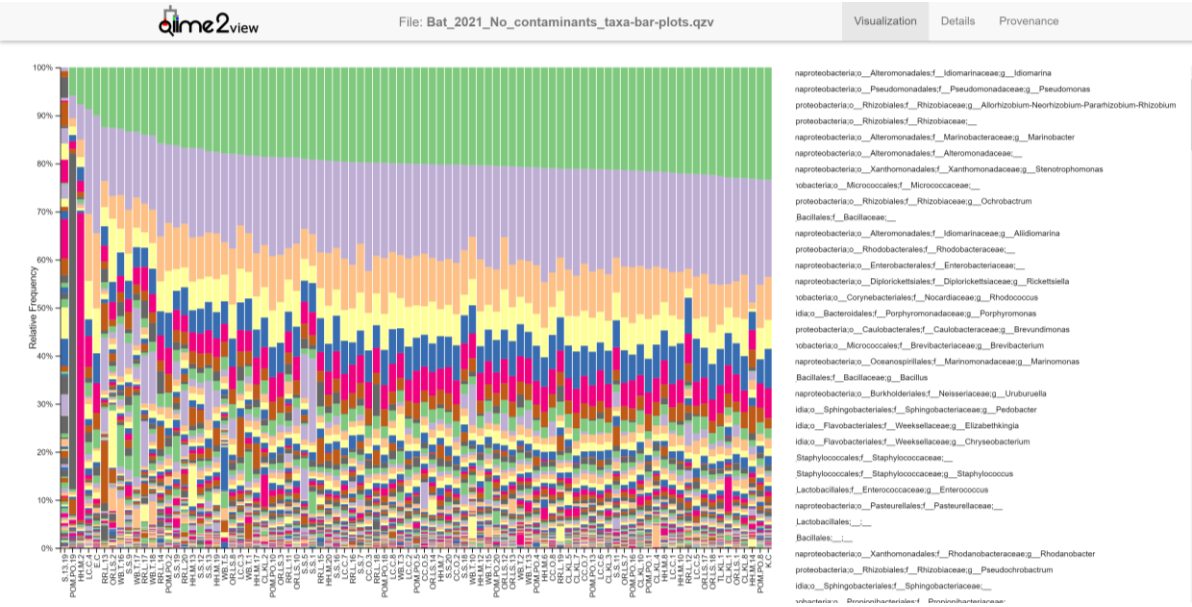
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Vibrio \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Halomonas \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Pseudoalteromonas \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

4.) The code I used to visualize my Bat_2021_No_contaminants_filtered-rep-seqs.qza:
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--o-visualization Bat_2021_No_contaminats_filtered-rep-seqs.qzv

5.) The code I used to visualize the no contaminants microbiome using taxa bar plots:
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_contaminants_taxa-bar-plots.qzv



Good news, the genera *Pseudoalteromonas, Vibrio,* and *Halomonas* were filtered out from the table.

3/18/2022: I made a new phylogenetic tree using my no contaminants filtered sequences and replotted my alpha rarefaction plots using my no contaminants table. From this, Dr. Walke and I were able to determine where to rarefy the sequences for the data analyses.

1.) The code I used to build a phylogenetic tree using my no contaminants sequences:

```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_No_Contaminants_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_No_contaminants_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_No_contaminants_unrooted-tree.qza \
--o-rooted-tree Bat_2021_No_contaminants_rooted-tree.qza
```

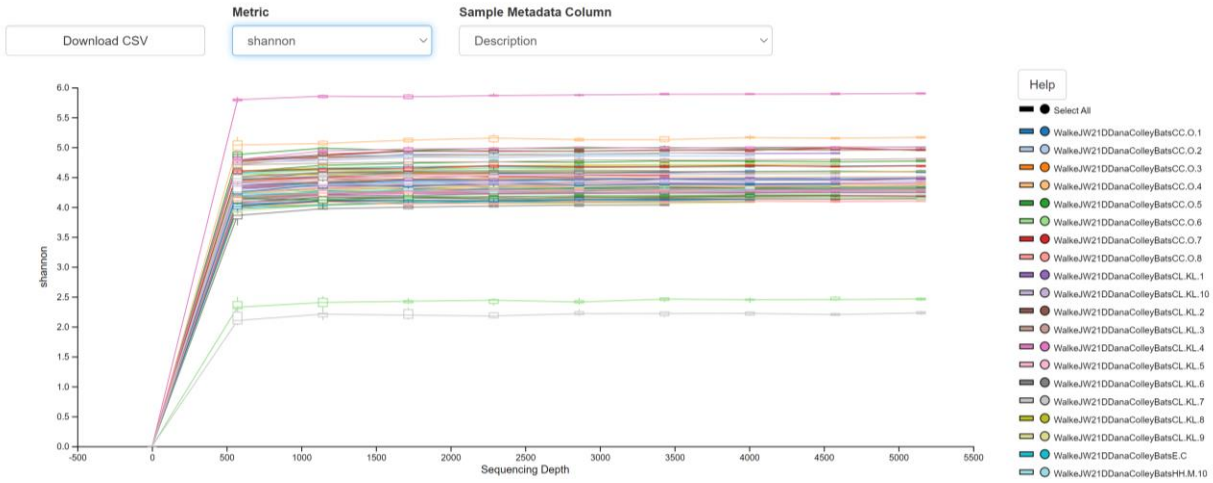
2.) The code I used to replot my alpha rarefaction using my no contaminants table

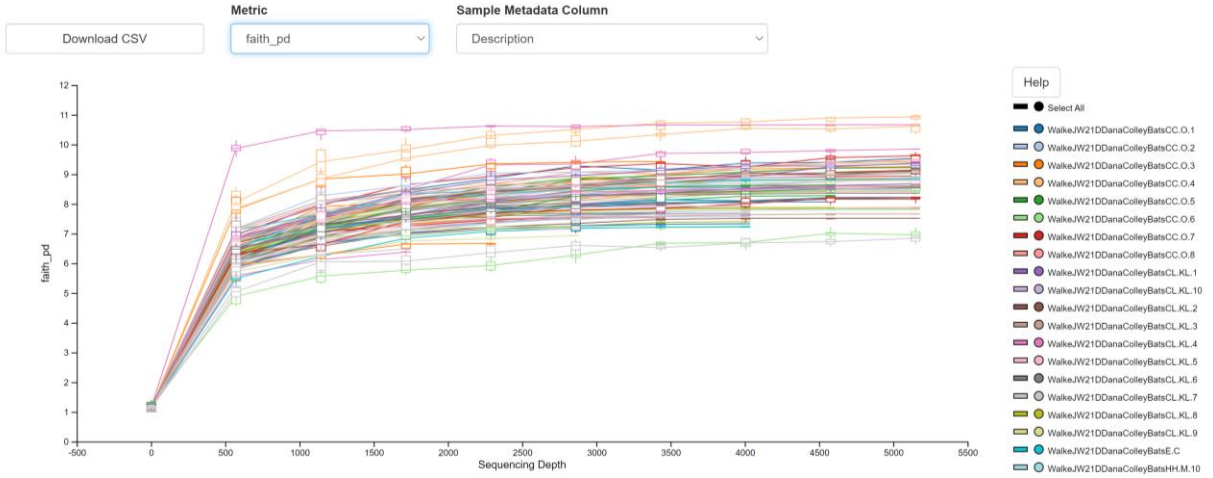
```
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_rooted-tree.qza' \
--p-max-depth 5150 \ ← new value from No_contaminants table
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_contaminants_alpha-rarefaction.qzv
```

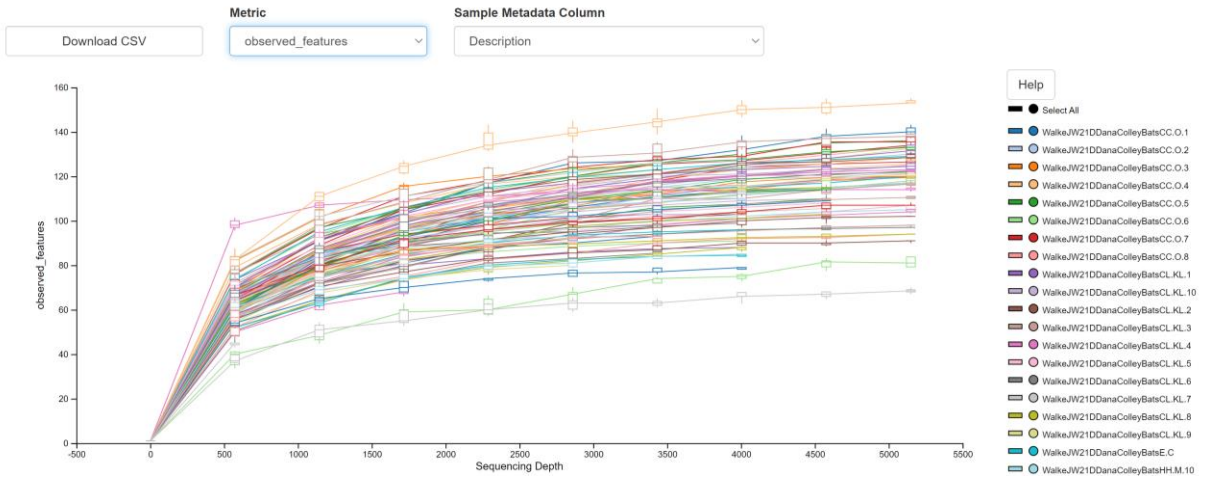
Where should we rarefy our sequences?

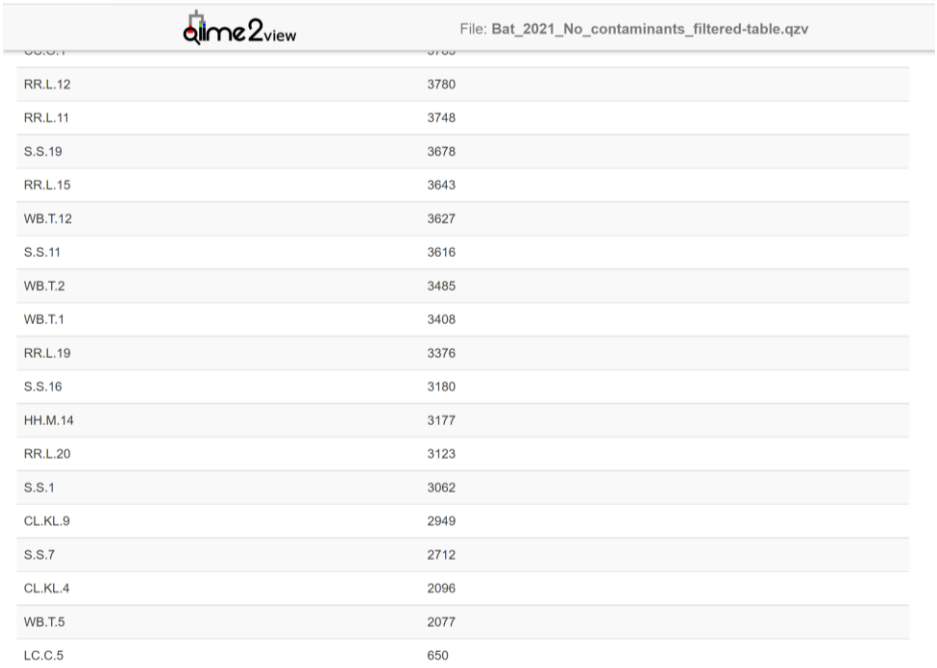
The new no contamination alpha rarefaction plots appear to be more leveled out for the Faith's phylogenetic diversity metric. The Shannon metric is still the same (level) and so is the observed_features (increasing).

The no contaminants table with the list of samples with their number of sequences appears to be more leveled out as well. There are fewer large jumps between the lowest sequence samples.


Metric
Sample Metadata Column
Download CSV
shannon
Description
Help
Select All
WalkeJW21DDanaColleyBatsCC.O.1
WalkeJW21DDanaColleyBatsCC.O.2
WalkeJW21DDanaColleyBatsCC.O.3
WalkeJW21DDanaColleyBatsCC.O.4
WalkeJW21DDanaColleyBatsCC.O.5
WalkeJW21DDanaColleyBatsCC.O.6
WalkeJW21DDanaColleyBatsCC.O.7
WalkeJW21DDanaColleyBatsCC.O.8
WalkeJW21DDanaColleyBatsCL.KL.1
WalkeJW21DDanaColleyBatsCL.KL.10
WalkeJW21DDanaColleyBatsCL.KL.2
WalkeJW21DDanaColleyBatsCL.KL.3
WalkeJW21DDanaColleyBatsCL.KL.4
WalkeJW21DDanaColleyBatsCL.KL.5
WalkeJW21DDanaColleyBatsCL.KL.6
WalkeJW21DDanaColleyBatsCL.KL.7
WalkeJW21DDanaColleyBatsCL.KL.8
WalkeJW21DDanaColleyBatsCL.KL.9
WalkeJW21DDanaColleyBatsE.C
WalkeJW21DDanaColleyBatsHH.M.10
shannon
Sequencing Depth


Metric
Sample Metadata Column
Download CSV
faith_pd
Description
Help
Select All
WalkeJW21DDanaColleyBatsCC.O.1
WalkeJW21DDanaColleyBatsCC.O.2
WalkeJW21DDanaColleyBatsCC.O.3
WalkeJW21DDanaColleyBatsCC.O.4
WalkeJW21DDanaColleyBatsCC.O.5
WalkeJW21DDanaColleyBatsCC.O.6
WalkeJW21DDanaColleyBatsCC.O.7
WalkeJW21DDanaColleyBatsCC.O.8
WalkeJW21DDanaColleyBatsCL.KL.1
WalkeJW21DDanaColleyBatsCL.KL.10
WalkeJW21DDanaColleyBatsCL.KL.2
WalkeJW21DDanaColleyBatsCL.KL.3
WalkeJW21DDanaColleyBatsCL.KL.4
WalkeJW21DDanaColleyBatsCL.KL.5
WalkeJW21DDanaColleyBatsCL.KL.6
WalkeJW21DDanaColleyBatsCL.KL.7
WalkeJW21DDanaColleyBatsCL.KL.8
WalkeJW21DDanaColleyBatsCL.KL.9
WalkeJW21DDanaColleyBatsE.C
WalkeJW21DDanaColleyBatsHH.M.10
faith_pd
Sequencing Depth


Metric
Sample Metadata Column
Download CSV
observed_features
Description
Help
Select All
WalkeJW21DDanaColleyBatsCC.O.1
WalkeJW21DDanaColleyBatsCC.O.2
WalkeJW21DDanaColleyBatsCC.O.3
WalkeJW21DDanaColleyBatsCC.O.4
WalkeJW21DDanaColleyBatsCC.O.5
WalkeJW21DDanaColleyBatsCC.O.6
WalkeJW21DDanaColleyBatsCC.O.7
WalkeJW21DDanaColleyBatsCC.O.8
WalkeJW21DDanaColleyBatsCL.KL.1
WalkeJW21DDanaColleyBatsCL.KL.10
WalkeJW21DDanaColleyBatsCL.KL.2
WalkeJW21DDanaColleyBatsCL.KL.3
WalkeJW21DDanaColleyBatsCL.KL.4
WalkeJW21DDanaColleyBatsCL.KL.5
WalkeJW21DDanaColleyBatsCL.KL.6
WalkeJW21DDanaColleyBatsCL.KL.7
WalkeJW21DDanaColleyBatsCL.KL.8
WalkeJW21DDanaColleyBatsCL.KL.9
WalkeJW21DDanaColleyBatsE.C
WalkeJW21DDanaColleyBatsHH.M.10
observed_features
Sequencing Depth



- ~2,000 or 2,700 appears to be a good place to rarefy. However, the top hit after filtering (*Idiomarina*) appears to be another contaminant (see no contaminants taxa bar plot). It is a deep-sea halophilic bacterium, and it does not appear in other bat skin microbiome literature. It seems very unlikely that this is indeed a true member of the bat skin microbiome (and that it is the most abundant bacterium), so I will filter it out, and re-do all of the phylogenetic trees and rarefaction plot steps to see where we should trim everything.

3.) The code I used to filter *Idiomarina* from my table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Idiomarina \
--o-filtered-table Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza
```

4.) The code I used to check that the contaminants were filtered from the table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--o-visualization Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

5.) The code I used to filter *Idiomarina* from our sequences

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Idiomarina \
--o-filtered-sequences Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington state bats – part 2

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats to compare their relative abundances sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the most abundant culturable bacterial isolates from western bats will differ from the most abundant bacteria in our sequencing data.

This sequencing data comes from maternity colonies sampled across Washington state from April – June 2021 with the WDFW.

Part II – Analyses on your sequence data – continued from part 1 of my bioinformatics notebook

a.) Preparing your sequences and assigning taxonomy to your sequence data – continued from part 1

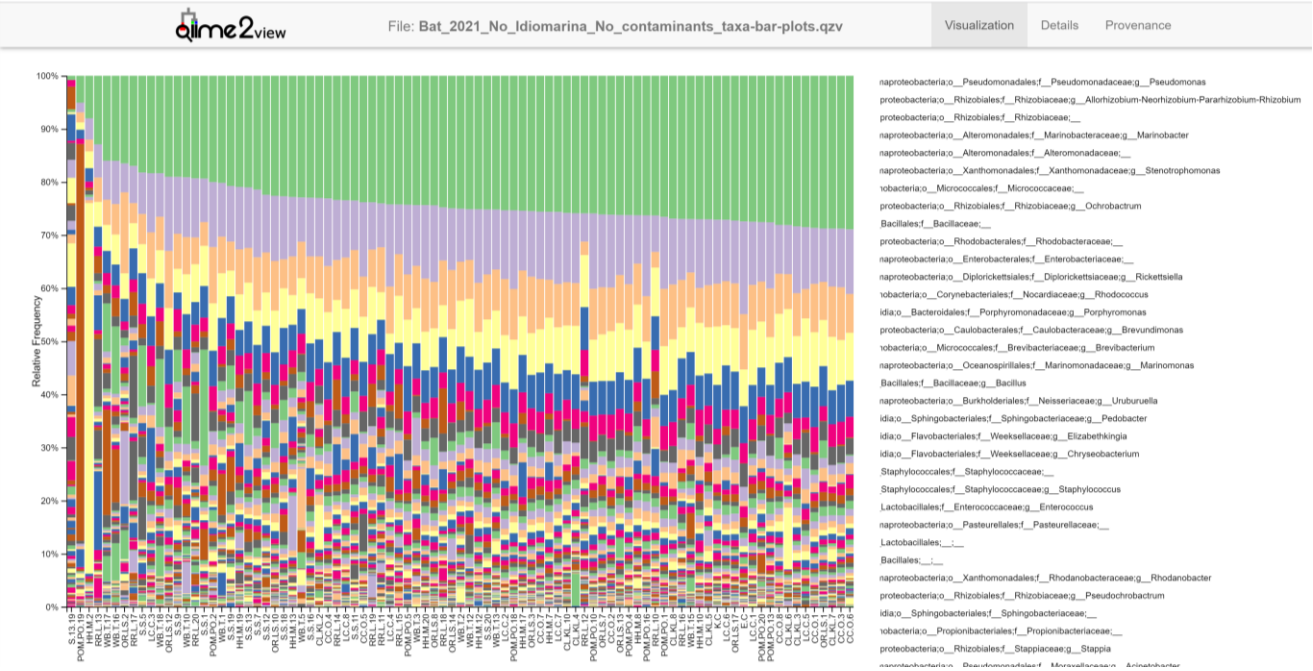
3/19/2022: I visualized the no *Idiomarina* sequences, created an updated no *Idiomarina* taxa bar plot, and re-did the phylogenetic tree and alpha rarefaction plotting step with the no *Idiomarina* sequences and table.

1.) The code I used to visualize my Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza:

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--o-visualization Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qzv
```

2.) The code I used to visualize the Bat_2021_No_Idiomarina_No_contaminants microbiome using taxa bar plots:

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv
```



Hooray! *Idiomarina* was filtered out! *Pseudomonas* is now the most abundant taxa (light green).

b.) Create phylogenetic tree and rarefy sequence data – continued from part 1

1.) The code I used to build a phylogenetic tree using my No_Idiomarina_No_contaminants sequences:

qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_No_Idiomarina_No_Contaminants_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_No_Idiomarina_No_contaminants_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_No_Idiomarina_No_contaminants_unrooted-tree.qza \
--o-rooted-tree Bat_2021_No_Idiomarina_No_contaminants_rooted-tree.qza

2.) The code I used to replot my alpha rarefaction using my No_Idiomarina_No_contaminants table:

qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--p-max-depth 4032 \ **← new value from No_Idiomarina table**
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_Idiomarina_No_contaminants_alpha-rarefaction.qzv

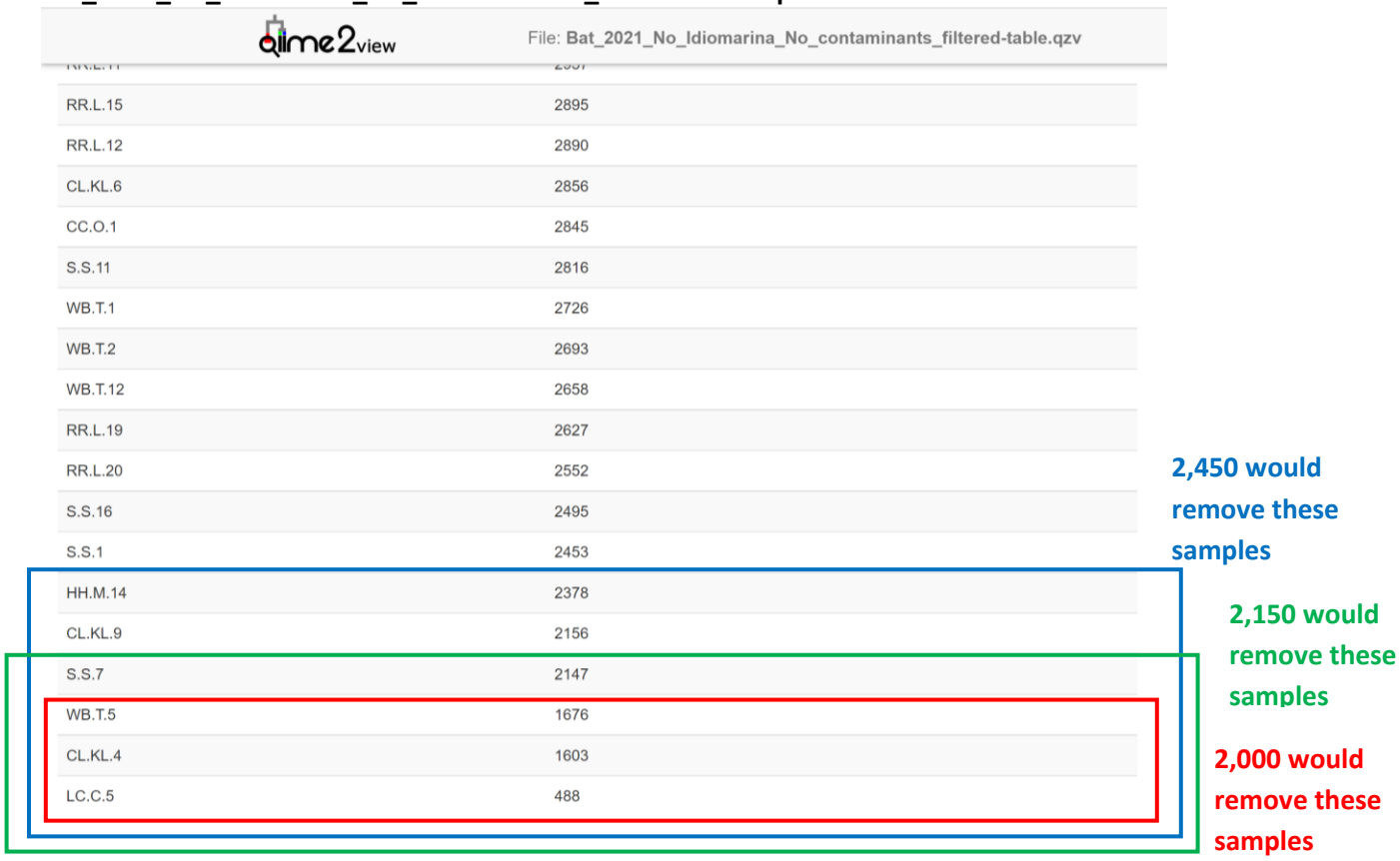
Where should we rarefy our sequences?

- **From No_Idiomarina alpha rarefaction plots:** 2,000 – 2,150 appears to be a decent cutoff point, with no huge jumps in these areas. There doesn't appear to be a huge difference between 2,000 or 2,150, so it would probably be good to go for the higher sequencing depth of **2,150** like we were thinking about for the other versions of the unfiltered alpha rarefaction plots.
    - It looks like the sequences continue to level out more with each filtering step. If we decided to cut off at 2,450 this would remove a sample of importance. As we continue

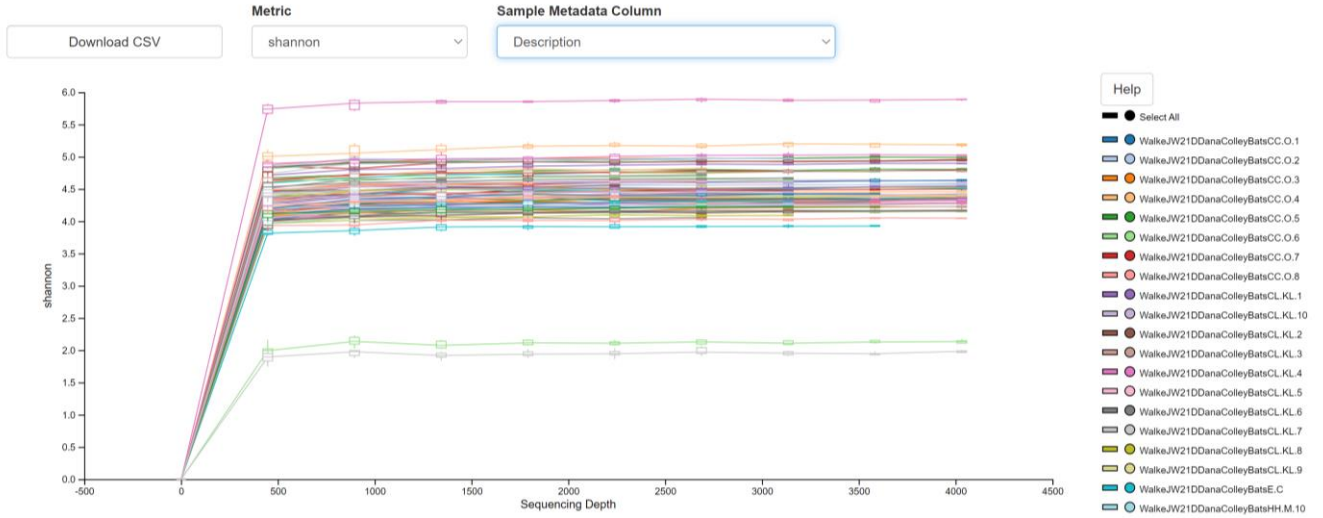
to filter these sequences, more samples are up for cutting as we figure out where to rarefy them.

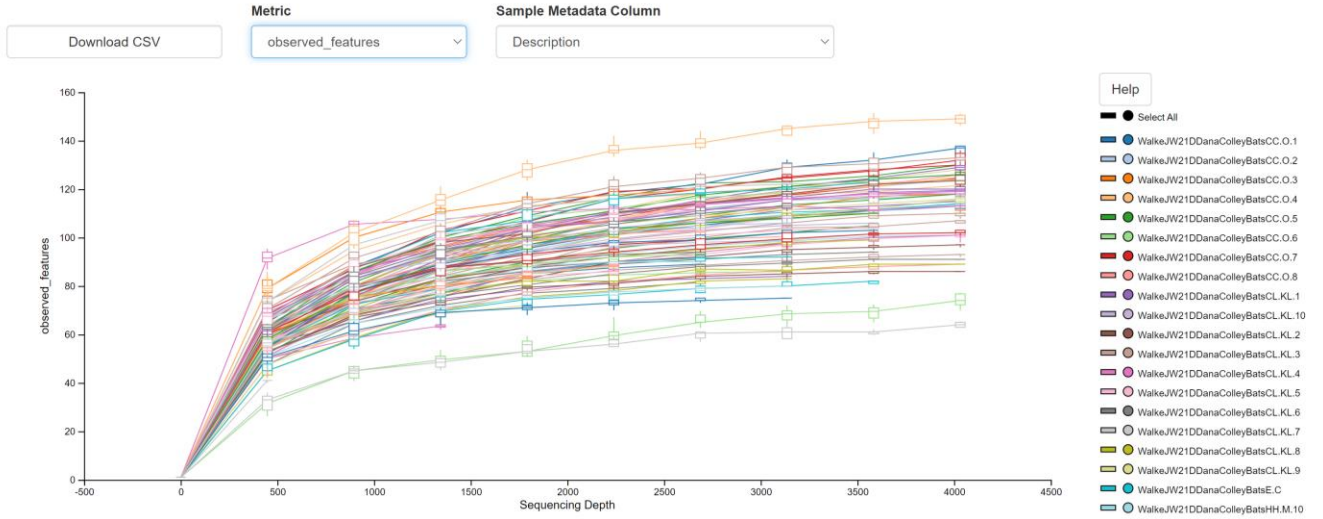
- **Samples up for cutting from No_Idiomarina table:**
  - **HH.M.14:** one of the common species and sex from that site (little brown, female), no ectoparasites. Probably OK to cut.
  - **CL.KL.9**: we need to keep this one, has ectoparasites, ectoparasite presence not as common so we need to keep every ectoparasite bat we can (see note about CL.KL.4).
  - **S.S7:** comparable to other samples from this site (all yuma myotis, all females, all no ectoparasites).
  - **WB.T.5:** comparable to other samples from this site (most yuma myotis females, no ectoparasites).
  - **CL.KL.4:** comparable to other samples from this site (all big brown bats, all females, most had ectoparasites), but ectoparasite bats in general were not as abundant overall so we should try to keep every ectoparasite sample that we can. However, it would probably be OK to cut just one ectoparasite bat to increase our sampling depth to 2,150 (otherwise we would be stuck at 1,500 which would not be desirable), but I would not want to remove any more ectoparasite samples.
  - **LC.C.5:** comparable to other samples from this site (all female little brown bats with no ectoparasites).

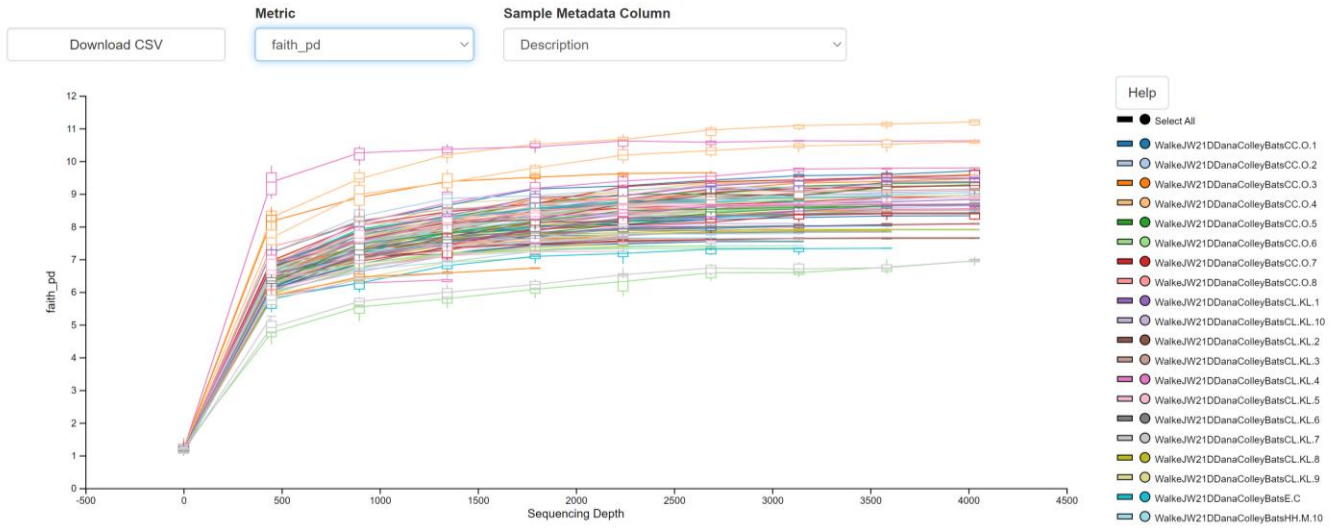
Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qzv:



From Bat_2021_No_Idiomarina_No_contaminants alpha rarefaction plots:







Update: last bout of filtering contaminants. I filtered out *Marinobacter*, *Marinomonas, Salinisphaera,* and *Salinarimonas* from the No_Idiomarina table and sequences. I remade the phylogenetic tree and alpha rarefaction plots using the updated sequences and table.

List of contaminants filtered out of our sequence data:

- *Pseudoalteromonas*
- *Vibrio*
- *Halomonas*
- *Idiomarina*
- *Marinobacter*
- *Marinomonas*
- *Salinisphaera*
- *Salinarimonas*

1.) The code I used to filter out the remaining contaminants from my table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinobacter \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \ ← output from previous step
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinomonas \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinisphaera \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinarimonas \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

2.) The code I used to visualize the final no contaminants table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' -\
-o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv --m-sample-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

3.) The code I used to filter out the remaining contaminants from my sequences:

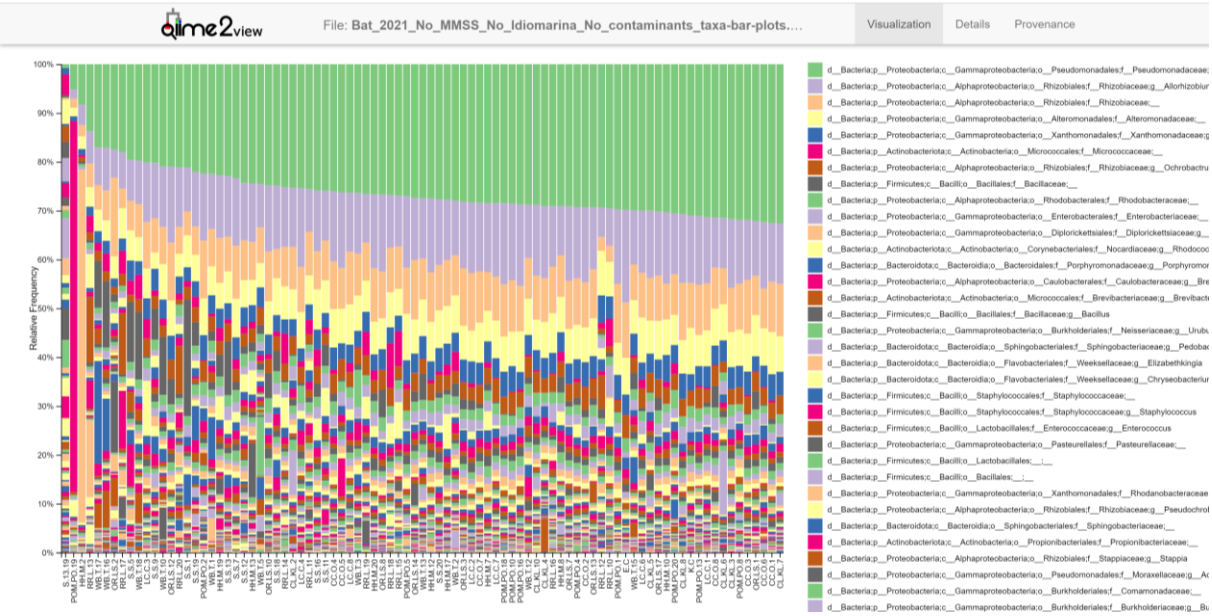
```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinobacter \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \ ← output from previous step
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinomonas \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinisphaera \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinarimonas \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

4.) The code I used to visualize my Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza:

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qzv
```

5.) The code I used to visualize the Bat_2021No_MMSS_No_Idiomarina_No_contaminants microbiome using taxa bar plots:

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv

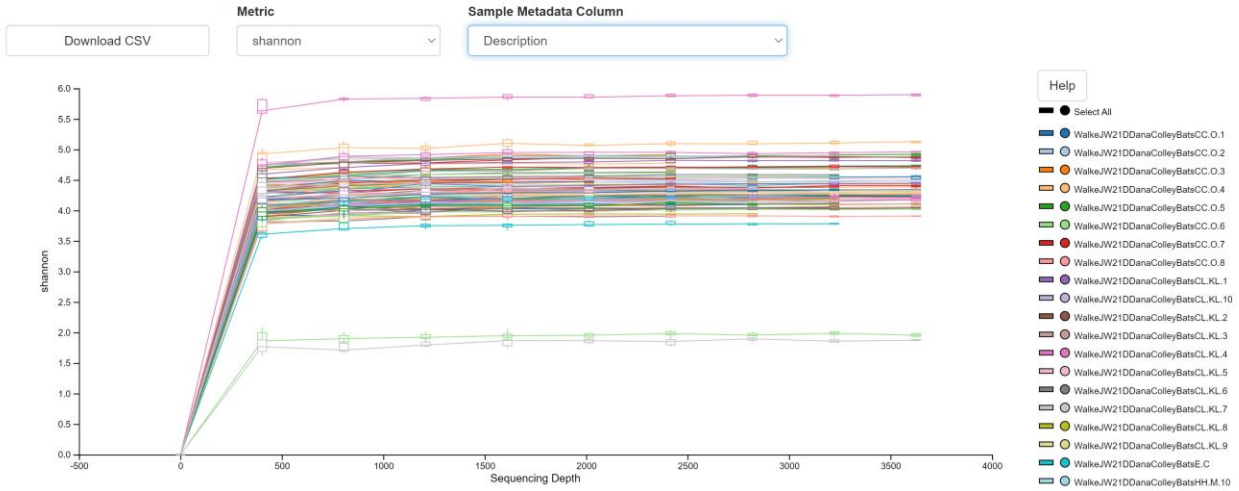


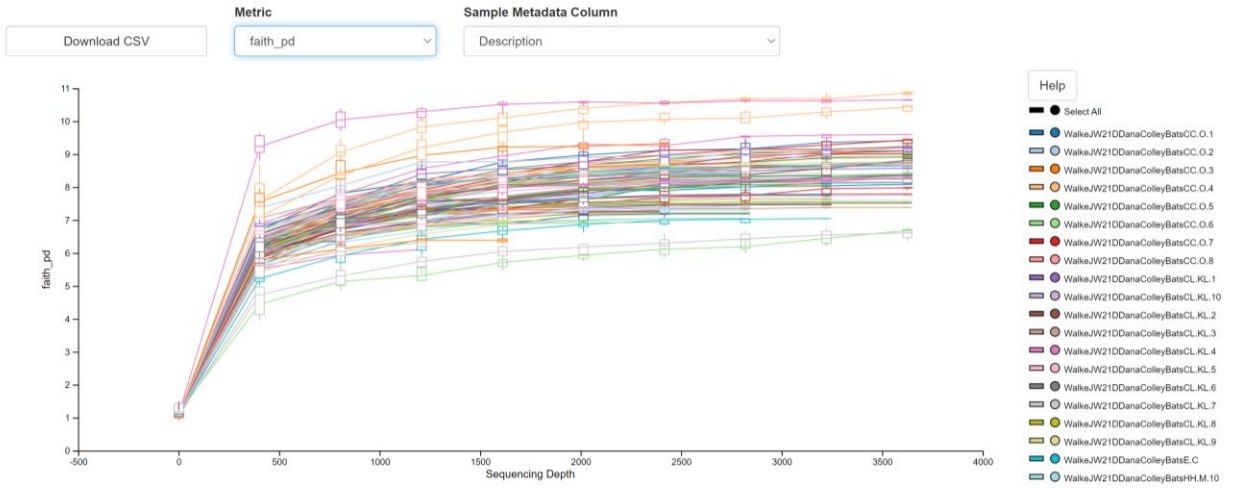
6.) The code I used to build a phylogenetic tree using my No_MMSS sequences:

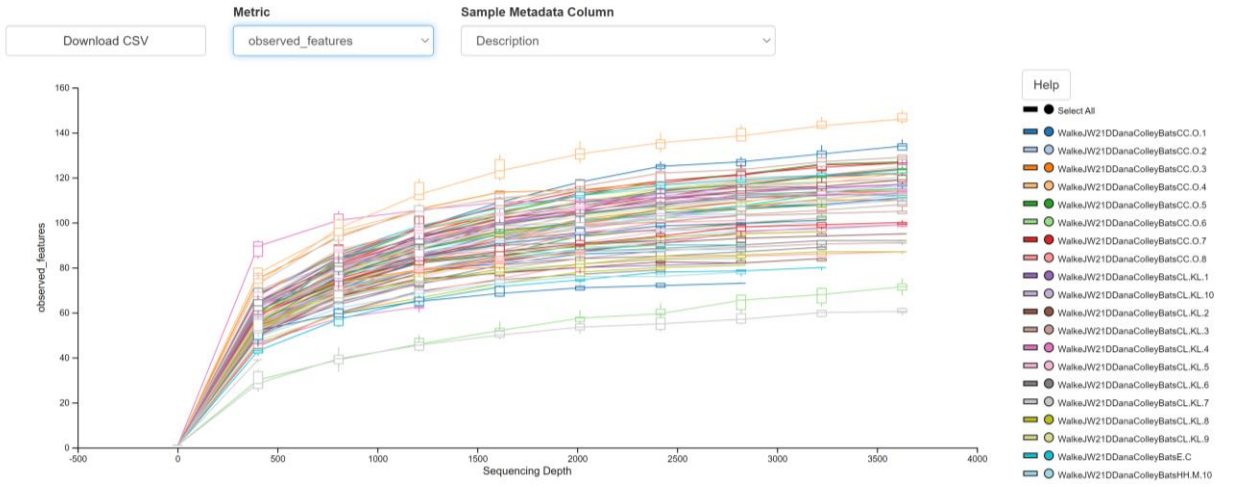
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_No_MMSS_No_Idiomarina_No_Contaminants_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_unrooted-tree.qza \
--o-rooted-tree Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza

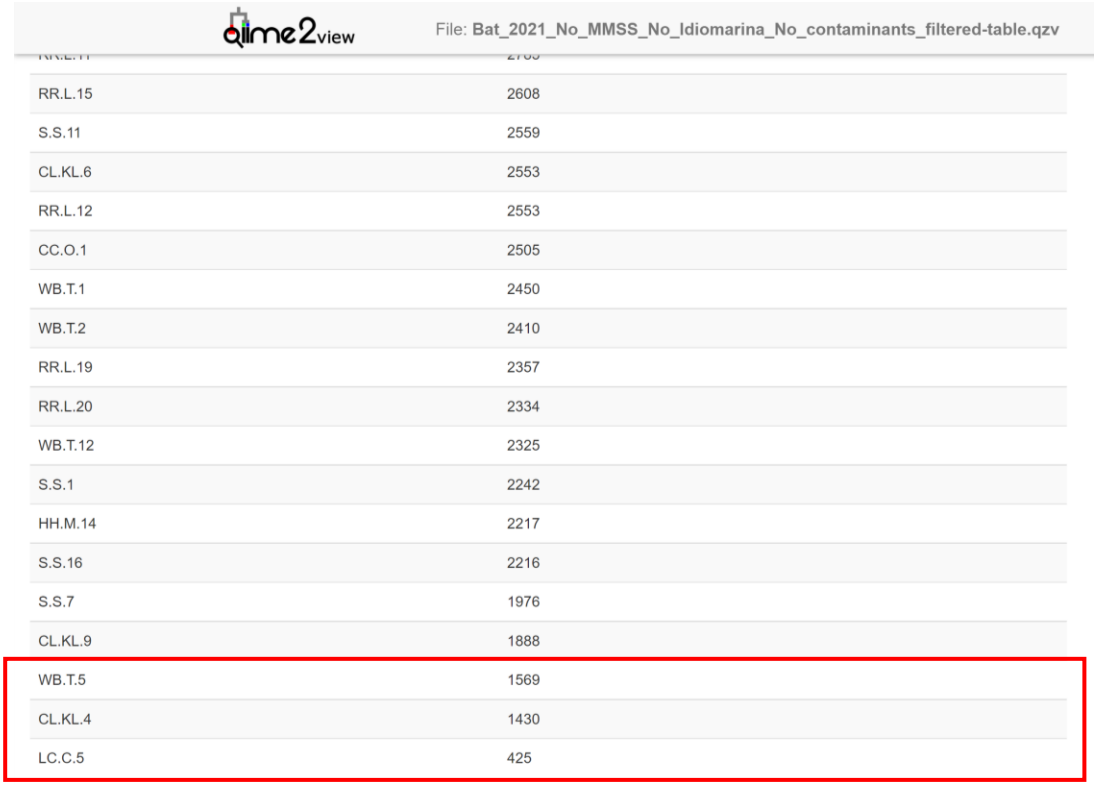
7.) The code I used to replot my alpha rarefaction using my No_MMSS table:

qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--p-max-depth 3626 \ **← new value from No_MMSS table**
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_alpha-rarefaction.qzv


Download CSV
Metric
shannon
Sample Metadata Column
Description
Help
Select All
WalkeJW21DDanaColleyBatsCC.O.1
WalkeJW21DDanaColleyBatsCC.O.2
WalkeJW21DDanaColleyBatsCC.O.3
WalkeJW21DDanaColleyBatsCC.O.4
WalkeJW21DDanaColleyBatsCC.O.5
WalkeJW21DDanaColleyBatsCC.O.6
WalkeJW21DDanaColleyBatsCC.O.7
WalkeJW21DDanaColleyBatsCC.O.8
WalkeJW21DDanaColleyBatsCL.KL.1
WalkeJW21DDanaColleyBatsCL.KL.10
WalkeJW21DDanaColleyBatsCL.KL.2
WalkeJW21DDanaColleyBatsCL.KL.3
WalkeJW21DDanaColleyBatsCL.KL.4
WalkeJW21DDanaColleyBatsCL.KL.5
WalkeJW21DDanaColleyBatsCL.KL.6
WalkeJW21DDanaColleyBatsCL.KL.7
WalkeJW21DDanaColleyBatsCL.KL.8
WalkeJW21DDanaColleyBatsCL.KL.9
WalkeJW21DDanaColleyBatsE.C
WalkeJW21DDanaColleyBatsHH.M.10
shannon
Sequencing Depth


Download CSV
Metric
faith_pd
Sample Metadata Column
Description
Help
Select All
WalkeJW21DDanaColleyBatsCC.O.1
WalkeJW21DDanaColleyBatsCC.O.2
WalkeJW21DDanaColleyBatsCC.O.3
WalkeJW21DDanaColleyBatsCC.O.4
WalkeJW21DDanaColleyBatsCC.O.5
WalkeJW21DDanaColleyBatsCC.O.6
WalkeJW21DDanaColleyBatsCC.O.7
WalkeJW21DDanaColleyBatsCC.O.8
WalkeJW21DDanaColleyBatsCL.KL.1
WalkeJW21DDanaColleyBatsCL.KL.10
WalkeJW21DDanaColleyBatsCL.KL.2
WalkeJW21DDanaColleyBatsCL.KL.3
WalkeJW21DDanaColleyBatsCL.KL.4
WalkeJW21DDanaColleyBatsCL.KL.5
WalkeJW21DDanaColleyBatsCL.KL.6
WalkeJW21DDanaColleyBatsCL.KL.7
WalkeJW21DDanaColleyBatsCL.KL.8
WalkeJW21DDanaColleyBatsCL.KL.9
WalkeJW21DDanaColleyBatsE.C
WalkeJW21DDanaColleyBatsHH.M.10
faith_pd
Sequencing Depth


Download CSV
Metric
observed_features
Sample Metadata Column
Description
Help
Select All
WalkeJW21DDanaColleyBatsCC.O.1
WalkeJW21DDanaColleyBatsCC.O.2
WalkeJW21DDanaColleyBatsCC.O.3
WalkeJW21DDanaColleyBatsCC.O.4
WalkeJW21DDanaColleyBatsCC.O.5
WalkeJW21DDanaColleyBatsCC.O.6
WalkeJW21DDanaColleyBatsCC.O.7
WalkeJW21DDanaColleyBatsCC.O.8
WalkeJW21DDanaColleyBatsCL.KL.1
WalkeJW21DDanaColleyBatsCL.KL.10
WalkeJW21DDanaColleyBatsCL.KL.2
WalkeJW21DDanaColleyBatsCL.KL.3
WalkeJW21DDanaColleyBatsCL.KL.4
WalkeJW21DDanaColleyBatsCL.KL.5
WalkeJW21DDanaColleyBatsCL.KL.6
WalkeJW21DDanaColleyBatsCL.KL.7
WalkeJW21DDanaColleyBatsCL.KL.8
WalkeJW21DDanaColleyBatsCL.KL.9
WalkeJW21DDanaColleyBatsE.C
WalkeJW21DDanaColleyBatsHH.M.10
observed_features
Sequencing Depth



qiime2view File: Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv

RR.L.15	2608
S.S.11	2559
CL.KL.6	2553
RR.L.12	2553
CC.O.1	2505
WB.T.1	2450
WB.T.2	2410
RR.L.19	2357
RR.L.20	2334
WB.T.12	2325
S.S.1	2242
HH.M.14	2217
S.S.16	2216
S.S.7	1976
CL.KL.9	1888
WB.T.5	1569
CL.KL.4	1430
LC.C.5	425

1,750 or 1,800 would remove these samples

3/20/2022: I made a new table, alpha rarefaction, and taxa bar plots using my updated metadata file with the species column included. I forgot to include species so these new plots have species included.

1.) With_spp table:

qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file
'/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'

2.) With_spp taxa bar plots:

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv

3.) With_spp alpha rarefaction:
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--p-max-depth 3626 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_alpha-rarefaction.qzv

4/5/2022: I removed the control samples (E.C, S.13.19, K.C) from the dataset and I ran the alpha and beta diversity analyses. You had to specify the column you wanted to analyze. In our case, all of the controls had "NA" listed for the Species, so we could remove those samples. I visualized all of the alpha diversity stats results.

Removing control samples from the table:
Should everything be lumped together
qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='NA'" \
--p-exclude-ids \
--o-filtered-table Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

Visualizing the table without the controls: check to see that the samples were filtered out
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--o-visualization Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'

→ Use No_control table from now on

c.) Actual data analyses – alpha and beta diversity
1.) Calculating alpha and beta diversity of your samples: → no controls
Filter data table by species, site level differences within a species
qiime diversity core-metrics-phylogenetic \

--i-phylogeny <file-name_rooted-tree.qza from phylog. tree step> \
--i-table <file-name_filtered-table.qza from table filter step> \
--p-sampling-depth <sequence depth number determined from alpha rarefaction step> \
--m-metadata-file <the mapping/metadata file you used earlier> \
--o-rarefied-table <file-name_rarefied-table.qza> \
--output-dir <file-name_core-metrics-results>

The code I used to calculate alpha and beta diversity:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered
-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_rarefied-table.qza  --output-dir Bat_2021_core-metrics-results
```

2.) Calculate alpha diversity statistics with nonparametric Kruskal-Wallis tests:

Make sure to transfer all stat data to a single excel file for each metric.

Shannon diversity:

qiime diversity alpha-group-significance \
--i-alpha-diversity <file-name_core-metrics-results/shannon_vector.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv > \
--o-visualization <file-name_core-metrics-results/shannon_significance.qzv>

The code I used to visualize Shannon's diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/shannon_vector.qza' /
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/shannon_significance.qzv
```

→ Note: the output goes to the core-metrics-results folder

Observed OTU's:

qiime diversity alpha-group-significance
--i-alpha-diversity <file-name_core-metrics-results/observed_features_vector.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_core-metrics-results/observed_features_vector_significance.qzv>

The code I used to visualize Observed OTU's:

```
qiime diversity alpha-group-significance \
```

--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/observed_features_vectors_significance.qzv

Faith's Phylogenetic Diversity:

qiime diversity alpha-group-significance
--i-alpha-diversity <file-name_core-metrics-results/faith_pd_vector.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_core-metrics-results/faith_pd_significance.qzv>

The code I used to visualize Faith's Phylogenetic Diversity:

qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/faith_pd_significance.qzv

Evenness:

qiime diversity alpha-group-significance
--i-alpha-diversity <file-name_core-metrics-results/evenness_vector.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_core-metrics-results/evenness_significance.qzv>

The code I used to visualize Evenness:

qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/evenness_significance.qzv

Discrete: different test

3.) Calculate beta diversity statistics using nonparametric PERMANOVAs:

Bray-Curtis:

qiime diversity beta-group-significance \
--i-distance-matrix <file-name_core-metrics-results/bray_curtis_distance_matrix.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever "treatment" you're comparing diversity values between> \
--o-visualization <file-name_core-metrics-results/bray_curtis_site_significance.qzv>
--p-pairwise

The code I used to visualize Bray-Curtis diversity: all variables included

qiime diversity beta-group-significance \

--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/bray_curtis_site_significance.qzv
--p-pairwise

Jaccard:
qiime diversity beta-group-significance
--i-distance-matrix <file-name_core-metrics-results/jaccard_distance_matrix.qza from diversity metrics folder in step #1>
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., ectoparasites> \
--o-visualization <file-name_core-metrics-results/jaccard_site_significance.qzv>
--p-pairwise

The code I used to visualize Jaccard:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise

Unweighted UniFrac Distances:
qiime diversity beta-group-significance
--i-distance-matrix <file-name_core-metrics-results/unweighted_unifrac_distance_matrix.qza from diversity metrics folder in step #1>
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., ectoparasites> \
--o-visualization <file-name_core-metrics-results/unweighted_unifrac_site_significance.qzv>
--p-pairwise

The code I used to visualize unweighted UniFrac:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
Weighted UniFrac Distances:
qiime diversity beta-group-significance

--i-distance-matrix <file-name_core-metrics-results/weighted_unifrac_distance_matrix.qza from diversity metrics folder in step #1>
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., ectoparasites> \
--o-visualization <file-name_core-metrics-results/weighted_unifrac_site_significance.qzv>
--p-pairwise

The code I used to visualize weighted UniFrac:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise

4/10/2022: I ran alpha and beta diversity analyses on data tables with just one type of species to see if ectoparasite presence does have an influence within the species level. Ectoparasite presence appears currently to not influence the microbiome diversity.

MYYU-only data table
qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='MYYU'" \
--o-filtered-table Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

Visualize the MYYU-only table to check that only MYYU included
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--o-visualization Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'

MYLU-only data table
qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \

--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='MYLU'" \
--o-filtered-table Bat_2021_MYLU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

MYVO-only data table

qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='MYVO'" \
--o-filtered-table Bat_2021_MYVO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

MYYU/LU-only data table

qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='MYYU/LU'" \
--o-filtered-table Bat_2021_MYYU_LU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

EPFU-only data table

qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='EPFU'" \
--o-filtered-table Bat_2021_EPFU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

Calculate alpha and beta diversity for MYYU-only data table:

qiime diversity core-metrics-phylogenetic \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \

```
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contamin
ants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYYU_only_rarefied-table.qza \
--output-dir Bat_2021_MYYU_only_core-metrics-results
```

Visualize MYYU-only alpha diversity results:
Shannon:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/shannon_vector.qza' /
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-
results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-
results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYYU-only beta diversity results:
Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
```

--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise

Jaccard:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise

Unweighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise

Weighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise

Calculate alpha and beta diversity for MYLU-only data table:
qiime diversity core-metrics-phylogenetic \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYLU_only_rarefied-table.qza \

--output-dir Bat_2021_MYLU_only_core-metrics-results

Visualize MYLU-only alpha diversity results:
Shannon:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/shannon_significance.qzv

Observed features:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/observed_features_vectors_significance.qzv

Faith's phylogenetic diversity:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/faith_pd_significance.qzv

Evenness:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/evenness_significance.qzv

Visualize MYLU-only beta diversity results:
Bray-Curtis:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise

Jaccard:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise

Unweighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise

Weighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise

Calculate alpha and beta diversity for MYVO-only data table:
qiime diversity core-metrics-phylogenetic \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYVO_only_rarefied-table.qza \
--output-dir Bat_2021_MYVO_only_core-metrics-results

Visualize MYVO-only alpha diversity results:
Shannon:
qiime diversity alpha-group-significance \

--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/shannon_significance.qzv

Observed features:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/observed_features_vectors_significance.qzv

Faith's phylogenetic diversity:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/faith_pd_significance.qzv

Evenness:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/evenness_significance.qzv

Visualize MYVO-only beta diversity results:
Bray-Curtis:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise

Jaccard:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/jaccard_distance_matrix.qza' \

--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise

Unweighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise

Weighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise

Calculate alpha and beta diversity for MYYU/LU-only data table:
qiime diversity core-metrics-phylogenetic \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYYU_LU_only_rarefied-table.qza \
--output-dir Bat_2021_MYYU_LU_only_core-metrics-results

Visualize MYYU/LU-only alpha diversity results:
Shannon:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/shannon_significance.qzv

Observed features:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/observed_features_vectors_significance.qzv

Faith's phylogenetic diversity:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/faith_pd_significance.qzv

Evenness:
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/evenness_significance.qzv

Visualize MYVO-only beta diversity results:
Bray-Curtis:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise

Jaccard:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

Calculate alpha and beta diversity for EPFU-only data table:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'\
 --o-rarefied-table Bat_2021_EPFU_only_rarefied-table.qza \
--output-dir Bat_2021_EPFU_only_core-metrics-results
```

Visualize EPFU-only alpha diversity results:

Shannon:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
```

```
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYVO-only beta diversity results:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
```

--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise

Weighted UniFrac Distances:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise

4/13/2022: I ran correlation tests on my continuous variables (weight, forearm length, ear length, tragus length) for both alpha diversity (Spearman's rank correlation) and beta diversity (Mantel test). I ran beta diversity analyses for species and site including all variables.

Beta diversity for Species:
Bray-Curtis:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Site \
--o-visualization Bat_2021_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise

4/19/2022: I ran Spearman's rank correlations for WDI and weight for the alpha diversity test.

1.) Alpha diversity for continuous variables: Spearman's rank correlation test
Shannon:
qiime diversity alpha-correlation \
--i-alpha-diversity <file_name_core-metrics-results/shannon_vector.qza> \
--m-metadata-file <the mapping/metadata file you've been using.tsv> \
--o-visualization <Bat_2021_core-metrics-results/shannon_correlation_Spearman.qzv>

qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/shannon_correlation_Spearman.qzv

Faith's Phylogenetic Diversity:
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/faith_pd_correlation_Spearman.qzv

Observed features:
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/observed_features_correlation_Spearman.qzv

Evenness:
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/evenness_correlation_Spearman.qzv

Repeat for species-only comparisons

2.) Calculate beta diversity statistics using Mantel tests:
Bray-Curtis:
qiime diversity beta-correlation \
--i-distance-matrix <file-name_core-metrics-results/bray_curtis_distance_matrix.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., Ectoparasites> \
--p-intersect-ids \
--o-metadata-distance-matrix <file-name_core-metrics-results/bray_curtis _correlation.qza> \
--o-mantel-scatter-visualization <file-name_core-metrics-results/bray_curtis_ correlation.qzv>

Repeat for Jaccard similarity, unweighted UniFrac Distances, weighted UniFrac Distances:

3.) Beta diversity PERMANOVA analyses:
For species:
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Species.qzv \
--p-pairwise

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington state bats – part 3

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats to compare their relative abundances sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the most abundant culturable bacterial isolates from western bats will differ from the most abundant bacteria in our sequencing data.

This sequencing data comes from maternity colonies sampled across Washington state from April – June 2021 with the WDFW.

Part II – Analyses on your sequence data – continued from part 2 of my bioinformatics notebook

c.) Actual data analyses – alpha and beta diversity

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Ectoparasites.qzv \
--p-pairwise

qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/No_slash_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Species.qzv \
--p-pairwise

qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/bray-curtis_significance_Bat_Ectoparasites.qzv \
--p-pairwise
```

4/24/2022: I ran beta diversity analyses on all of the individual species comparisons. I ran the rest of the Spearman's rank correlations for the species-specific tables for alpha diversity.

1.) Calculate beta diversity for site within each species:
Beta diversity among site in MYLU only: repeat for other beta diversity metrics and other species
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Site \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/bray_curtis_site_significance_Site.qzv \
--p-pairwise

2.) Repeat for other alpha diversity metrics and other species
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/shannon_correlation_Spearman.qzv

2.) Calculate beta diversity statistics using Mantel tests:
Bray-Curtis:
qiime diversity beta-correlation \
--i-distance-matrix <file-name_core-metrics-results/bray_curtis_distance_matrix.qza from diversity metrics folder in step #1> \
--m-metadata-file <whatever variable you're interested in> \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/bray_curtis_Ectoparasites_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Weight \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/jaccard_distance_matrix.qza' \

```
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column WDI \
--p-intersect-ids --o-metadata-distance-matrix Bat_2021_core-metrics-results/jaccard_WDI_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/jaccard_WDI_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column WDI \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/unweighted_unifrac_WDI_correlation.qza/
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/unweighted_unifrac_WDI_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column WDI \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/weighted_unifrac_WDI_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/weighted_unifrac_WDI_correlation.qzv
```

I had to remove the control samples from my metadata file, QIIME didn't like those samples because they had NAs for many of the values

Repeat for other continuous variables (WDI), for other beta diversity metrics, for species specific chunks

3.) Beta diversity PERMANOVA analyses:
For species:

For site:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
```

--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Species.qzv \
--p-pairwise

5/4/2022: I ran LEfSe analyses (relative abundance comparisons) on the significant comparisons to see which bacterial taxa were actually driving the significant differences between the different groups (i.e., different sites, different species). I also ran a LEFSe analysis on ectoparasites just to see if any bacterial taxa differ from each other, even though the overall relationships were non-significant.

Update: LEfSe analysis did not work on the Harvard website, so I had to run a similar analysis (indicspecies) in R. See R Notebook for details.

LEfSe Step-by-Step:

1.) Calculate relative frequency for a collapsed table (genus in example at level 6, repeat for other levels like phylum, family, etc.):
qiime taxa collapse \ → **groups bacteria of particular taxonomy (i.e., genus)**
--i-table <file_name_rarefied-table.qza from previous analyses, make sure that it's the **rarefied table** that was produced by the alpha/beta diversity analyses> \
--i-taxonomy <file_name_taxonomy.qza from previous analyses> \
--p-level 6 \ → **genus level here, refer back to the taxonomy bar plots for what the other levels are**
--o-collapsed-table <file_name_filtered-table-l6.qza>

qiime feature-table relative-frequency \
--i-table <file_name_filtered-table-l6.qza from previous code chunk> \
--o-relative-frequency-table <file_name_frequency-table-l6.qza>

Exporting straight ASV table, non-collapsed, rarefied = do not include taxonomy header → total ASV, individual ASVs for the non-collapsed

The code I used to calculate relative frequency for a collapsed table:
qiime taxa collapse \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-level 6 \
--o-collapsed-table Bat_2021_filtered-table-l6.qza

qiime feature-table relative-frequency \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_filtered-table-l6.qza' \
--o-relative-frequency-table Bat_2021_frequency-table-l6.qza

2.) Export biom file from QIIME: variations of this code will work to export other data from QIIME as well for other uses besides LEfSe analyses
qiime tools export \
--input-path <file_name_frequency-table-l6.qza from previous step> \
--output-path <file_name_lefse-files>

qiime tools export \
--input-path '/media/sf_QIIME/QIIME/Bat_2021_frequency-table-l6.qza' \
--output-path Bat_2021_lefse-files

→ This exports a new folder (i.e., here the folder is titled "Bat_2021_lefse-files" in your directory

3.) Convert biom to text file (for LEfSe comparison): biom files from QIIME need to be converted to compatible file types for most other uses as well.
biom convert \ → for feature tables, asv table, species table, genus table → at different levels, genus/phylum/…/individual ASV is non-collapsed one
--input-fp <file_name_lefse-files/feature-table.biom> \
--output-fp <file_name_lefse-files/frequency-table-l6.txt> \
--header-key "taxonomy" / → works for general QIIME format, column with taxonomy indicated
--to-tsv

The code I used to convert the biom file to a text file:
biom convert \
--input-fp Bat_2021_lefse-files/feature-table.biom \
--output-fp Bat_2021_lefse-files/Bat_2021_frequency-table-l6.txt \
--header-key "taxonomy" \
--to-tsv

→ converted text file will be outputted in the folder you just made in your directory (i.e., "Bat_2021_lefse-files"), but the typical green "exported or saved" output after a successful code run does not show up in this step, so don't panic. Just check the new folder you made for the new text file

qiime tools export --input-path --output-path --header-key "xx" --to-tsv

biom convert --input-fp --output-fp

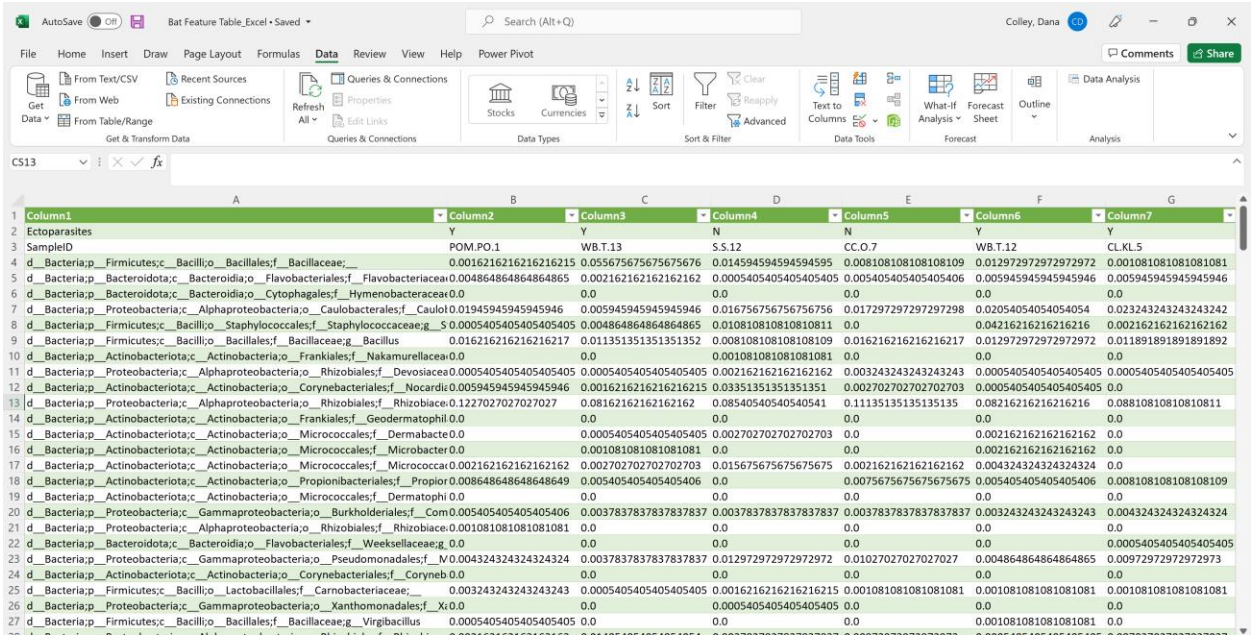
4.) Edit the text file in Excel so you can use it for LEfSe:
How to open a text file in Excel and convert it to an Excel workbook:

a. **Open a new Excel workbook**
b. **Data tab → "get and transform data" section on far left**
c. **Select "from text/CSV" option in the get data section**
d. **Find your text file → import → load**
e. **Save your Excel workbook**

Editing the text file in Excel for LEfSe analyses:

a. Delete **#Constructed from biom file** row
b. Add Ectoparasites as the new 1st row → **whatever variable you are interested, check your mapping file to see what you want to do (i.e., Ectoparasites, Site, Species, Weight, etc.)**
    - Your Ectoparasite row needs to have the data for each individual (i.e., ectoparasite present/absent for each bat), transpose data if necessary (number of rows needs to match the number of columns → paste special → more options → transpose)
c. Replace **#OTU ID with SampleID**

→ This edited Excel file will be used in the next step.



	A	B	C	D	E	F	G
1	Column1	Column2	Column3	Column4	Column5	Column6	Column7
2	Ectoparasites	Y	Y	N	N	Y	Y
3	SampleID	POM.PO.1	WB.T.13	S.S.12	CC.O.7	WB.T.12	CL.KL.5
4	d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;__	0.001621621621621	0.055675675675676	0.014594594594595	0.008108108108109	0.012972972972972	0.001081081081081
5	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriacea	0.004864864864865	0.002162162162162	0.000540540540540	0.005405405405406	0.005945945945946	0.005945945945946
6	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Cytophagales;f__Hymenobacteraceae	0.0	0.0	0.0	0.0	0.0	0.0
7	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Caulob	0.01945945945945946	0.005945945945946	0.016756756756756	0.017297297297298	0.02054054054054	0.023243243243242
8	d__Bacteria;p__Firmicutes;c__Bacilli;o__Staphylococcales;f__Staphylococcaceae;g__S	0.000540540540540	0.004864864864865	0.010810810810811	0.0	0.04216216216216216	0.002162162162162
9	d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus	0.016216216216217	0.011351351351352	0.008108108108109	0.016216216216217	0.012972972972972	0.011891891891892
10	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Frankiales;f__Nakamurellacea	0.0	0.0	0.001081081081081	0.0	0.0	0.0
11	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Devosiacea	0.000540540540540	0.000540540540540	0.002162162162162	0.003243243243243	0.000540540540540	0.000540540540540
12	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Corynebacteriales;f__Nocardia	0.005945945945946	0.001621621621621	0.033513513513513	0.002702702702703	0.000540540540540	0.0
13	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiace	0.1227027027027027	0.08162162162162162	0.085405405405405	0.11135135135135135	0.08216216216216216	0.088108108108108
14	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Frankiales;f__Geodermatophil	0.0	0.0	0.0	0.0	0.0	0.0
15	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Dermabacte	0.0	0.000540540540540	0.002702702702703	0.0	0.002162162162162	0.0
16	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Microbacter	0.0	0.001081081081081	0.0	0.0	0.002162162162162	0.0
17	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Micrococca	0.002162162162162	0.002702702702703	0.015675675675675	0.002162162162162	0.004324324324324	0.0
18	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Propionibacteriales;f__Propior	0.008648648648649	0.005405405405406	0.0	0.007567567675675675	0.005405405405406	0.008108108108109
19	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Micrococcales;f__Dermatophi	0.0	0.0	0.0	0.0	0.0	0.0
20	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Com	0.005405405405406	0.003783783783783837	0.003783783783783837	0.003783783783783837	0.003243243243243	0.004324324324324
21	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Rhizobiace	0.001081081081081	0.0	0.0	0.0	0.0	0.0
22	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Weeksellaceae;g	0.0	0.0	0.0	0.0	0.0	0.000540540540540
23	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__M	0.004324324324324	0.003783783783783837	0.012972972972972	0.01027027027027027	0.004864864864865	0.00972972972972973
24	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Corynebacteriales;f__Coryneb	0.0	0.0	0.0	0.0	0.0	0.0
25	d__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Carnobacteriaceae;__	0.003243243243243	0.000540540540540	0.001621621621621	0.001081081081081	0.001081081081081	0.001081081081081
26	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Xanthomonadales;f__Xa	0.0	0.0	0.000540540540540	0.0	0.0	0.0
27	d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Virgibacillus	0.000540540540540	0.0	0.0	0.0	0.001081081081081	0.0

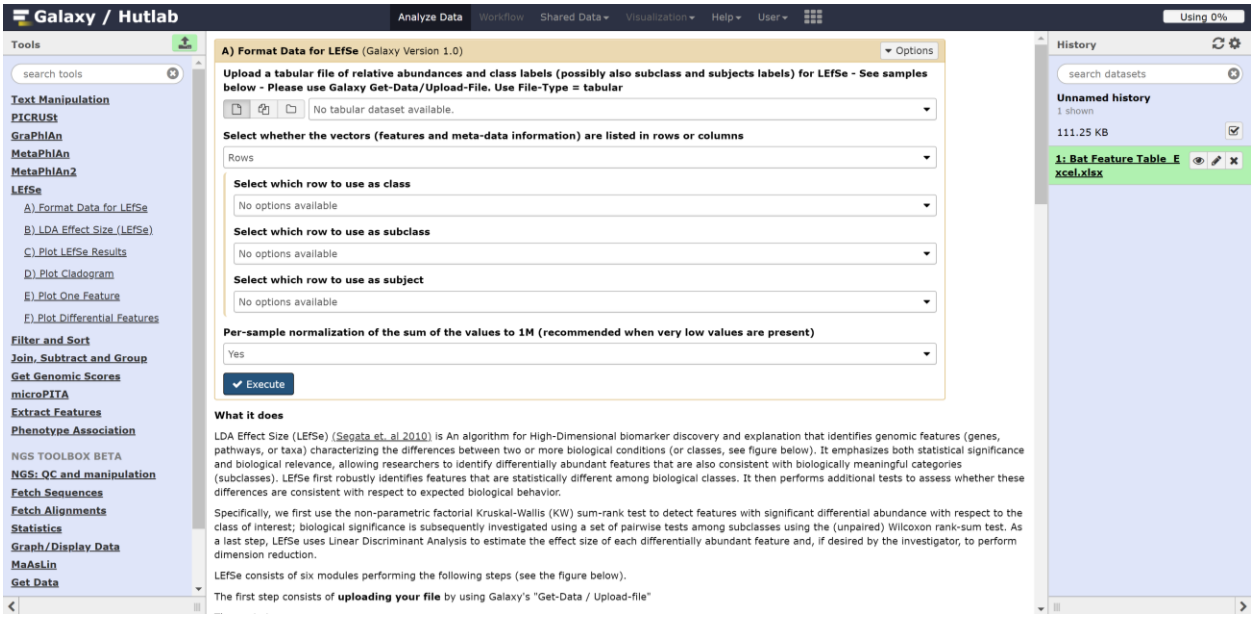
Running the LEfSe analyses:

→ Use the edited Excel file from the previous step:

a. Follow instructions to use LEfSe on Harvard site (http://huttenhower.sph.harvard.edu/galaxy/)
b. Defaults are OK
c. For this project, there is only a class (ChytridResult) and no subclass.

LEfSe:

1. Load in your data to the Galaxy/Hutlab website:
    - Load data in the “history” pane on the right side of the page
2. Format data for LEfSe (part A)
    - Data listed in **columns**
    - All other defaults OK



3. Couldn't get the Lefse analysis to work, so we did an indicspecies analysis in R instead. See R code notebook for details.

5.) Create taxa bar plots for specific comparisons as necessary to go with the LEfSE analysis results
**qiime taxa barplot **

Remember, something like this:

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv

How to export data files in QIIME:

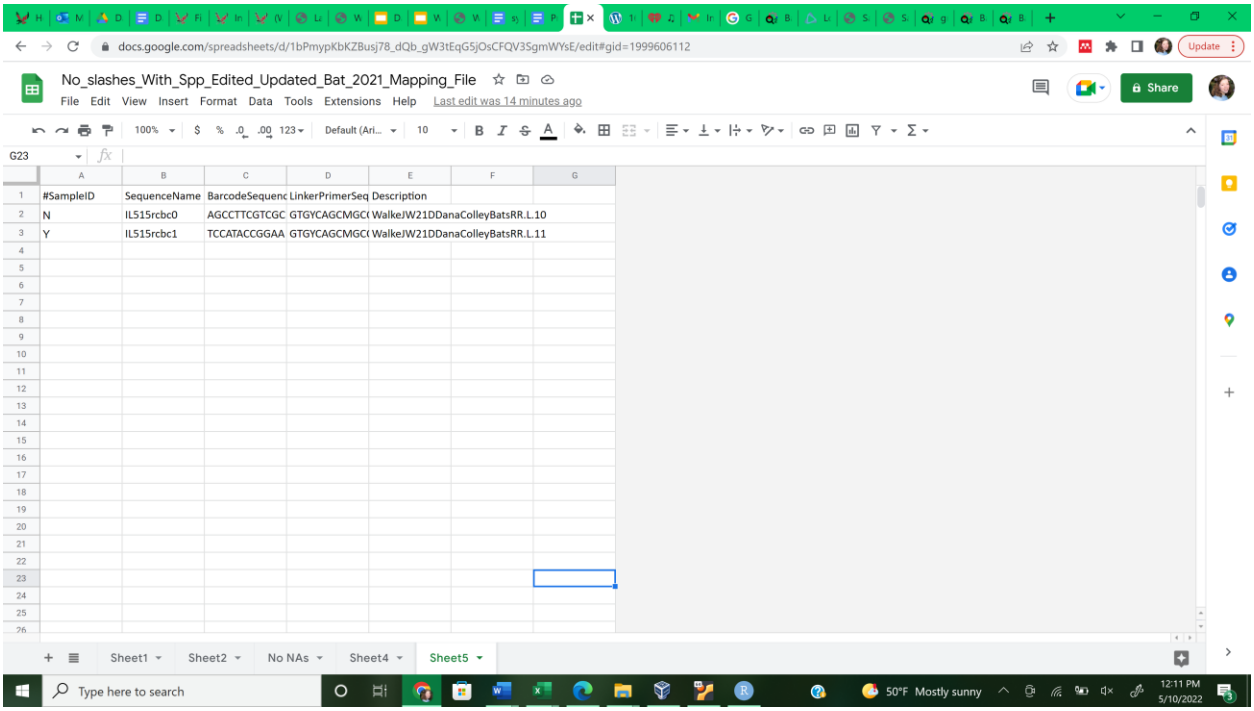
qiime tools export \
--input-path '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--output-path Bat_2021_weighted_unifrac_distance_matrix-files

Making grouped taxa bar plots: we had to make fake metadata files for each grouped bar plot (see photo below). The fake metadata must include all group options (i.e., six different samples for bat species, each sample represents one of the six different bat species). Make sure you include the sequence name, barcode sequence, linker primer sequence, and description. They don't need to correspond to the correct samples, just make sure that all options are represented.

For ectoparasites:

qiime feature-table group \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--p-axis 'sample' \

```
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--p-mode 'mean-ceiling' \
--o-grouped-table Bat_2021_Grouped_EP_Feature_Table

qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_EP_Feature_Table.qza' \
--o-visualization Bat_2021_Grouped_EP_Feature_Table.qzv

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_EP_Feature_Table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/EP_Grouped_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet5.tsv' \
 --o-visualization Bat_2021_Grouped_EP_taxa-bar-plots.qzv
```



For roost location:

```
qiime feature-table group \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--p-axis 'sample' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Site \
--p-mode 'mean-ceiling' \
```

--o-grouped-table Bat_2021_Grouped_Site_Feature_Table.qza

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_Site_Feature_Table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/Site_Grouped_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Site.tsv' \
--o-visualization Bat_2021_Grouped_Site_taxa-bar-plots.qzv

For bat species:
qiime feature-table group \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--p-axis 'sample' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--p-mode 'mean-ceiling' \
--o-grouped-table Bat_2021_Grouped_Species_Feature_Table.qza

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_Species_Feature_Table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/Species_Grouped_Slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Species.tsv' \
--o-visualization Bat_2021_Grouped_Species_taxa-bar-plots.qzv

Filter table by COTO:
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='COTO'" \
--o-filtered-table
Bat_2021_COTO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza

Calculate alpha and beta diversity metrics for COTO:
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_COTO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \

--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-rarefied-table Bat_2021_COTO_only_rarefied-table.qza \
--output-dir Bat_2021_COTO_only_Final_core-metrics-results

qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_COTO_only_Final_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_COTO_only_Final_core-metrics-results/shannon_correlation_Spearman_Weight.qzv

→ Repeat correlations for other alpha diversity metrics

Beta diversity correlation via Mantel tests
qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_COTO_only_Final_core-metrics-results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Weight \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_COTO_only_Final_core-metrics-results/bray_curtis_Weight_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_COTO_only_Final_core-metrics-results/bray_curtis_Weight_correlation.qzv

APPENDIX F

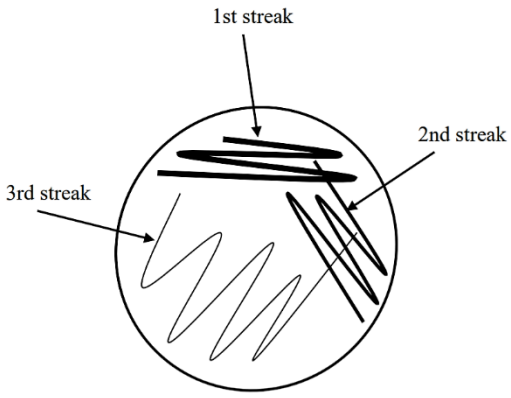
DNA Extraction from Bacterial Colonies via the Freeze-Thaw Method

Part I: isolate preparation and re-streaking

1. Take your samples out of the freezer (the bacteria and TSYE + 20% glycerol in the 2 mL screw-cap cryogenic tubes) and place them in a container of ice.
    - You don't want your samples to thaw completely, just slightly so you can scrape a little off the top.
2. Scrape off the top layer of semi-frozen bacteria and TSYE + 20% glycerol using a sterile toothpick.
3. Gently streak bacteria and TSYE + 20% glycerol on fresh media for isolation (see figure).
4. Incubate plates until distinct colonies appear.

Part II: DNA extraction from isolates (3 – 7 days after re-streaking)

1. Using sterile methods, pipette **200 µL of buffer TE or AE** into a sterile 1.5 mL centrifuge tube **(not low-retention!)**.
2. Add a colony (or loop-full) of bacteria to the tube.
3. Vortex tube.
4. Heat tubes in a **99°C heat block for 1 minute**
5. Cool tubes in a **-80°C freezer for 3 minutes**
6. Heat tubes in a **99°C heat block for 2 minutes**
7. Repeat steps 5-6 **two more times:**
    - Cool tubes in a -80°C freezer for 3 minutes
    - Heat tubes in a 99°C heat block for 2 minutes



8. Centrifuge tubes at **10,000 RPM for 5 minutes**
9. Pipette **100 µL of supernatant** into a new sterile 1.5 mL centrifuge tube
10. Store DNA at -20°C or -80°C until ready to use for PCR.

Reference:

Tsai, Y. and Olson, B.H. (1991) Rapid method for direct extraction of DNA from soil and sediments. Appl. Environ. Microbiol. 57, 1070 – 1074

APPENDIX G

PCR of 16S rRNA gene for Sanger Sequencing Protocol

Edited Nov-April 2018-2019 by: Shelby Fettig, Jeni Walke

Reagents:
UltraClean PCR grade H2O

5 Prime Hot Master Mix

Forward primer 8F (10uM)

Reverse primer 1492R (10uM)

Before beginning:

- Sterilize workspace with RNA Away. If possible, perform in a hood dedicated to PCR set up. UV hood before using; UV hood space 15 minutes and open PCR tubes for additional 15 minutes.
- Sterilize pipettors with bleach and ethanol or with RNA away (**use pipettors dedicated for PCR** reagents and use a separate pipettor for the DNA).
- Clean and sterilize with 5-10% bleach: 1 large centrifuge tube rack and several small PCR tube racks. Rinse and allow to dry.
- Locate samples and reagents. Keep both in fridge until ready to use.

Step 1: Make your PCR reactions

A) For each sample, you will run one PCR reaction.
B) You will run one negative control each PCR run.
C) You will run one positive control too, with a sample you know will work.
D) For samples that might have LOW DNA CONCENTRATIONS, the PCR reactions could be prepared with the same method as below, but with a small change in the volume of the reagents and DNA; additionally, BSA could be added to increase PCR yield.

Per sample		"Cake Batter for N=8"
11 ul UltraClean PCR grade H2O	x N(number of samples incl. cont. +1 extra for pipetting)	88 ul
10 ul 5 Prime Hot Master Mix		80 ul
1 ul Forward primer 8F		8 ul
1 ul Reverse primer 1492R		8 ul

23 ul Total (Before DNA)

+ 2 ul DNA (or water for negative control)

25 ul Total (After DNA)

1. Add all reagents **EXCEPT DNA** into a 1.5 mL centrifuge tube. This is your "cake batter".

2. Pipette 23 ul of "cake batter" into each of your sample PCR tubes.

3. Add DNA (2ul) to each tube EXCEPT the negative control. Add 2ul water (or "cake batter") to negative.

4. Vortex gently and centrifuge each PCR tube, including negative control, briefly.

Step 2: Run reactions in thermocycler

1. Make sure machine is set for 25 ul samples.

2. Thermocycler conditions:

	Temp	Time	
1.	94°C	2 min	
2.	94°C	30 sec	Denaturing
3.	50°C	30 sec	Annealing
4.	65°C	1.5 min	Extension
	Repeat steps 2-4 34x		
5.	65°C	10 min	
6.	4°C	hold	

You can maintain your PCR product in the fridge overnight if you need to wait until the next day to run your gel.

Step 3. Run gels to check amplification and negative controls

1. Make a 1% gel. Combine 1X TBE and agarose in a small Erlenmeyer flask. Microwave until just boiling. Swirl. Continue boiling/swirling until solution is completely clear. Be sure the liquid does not boil over-use appropriate size flask for volume of liquid to prevent this from happening.

a. Mini-gels:

i. 1% 40 mL buffer, 0.4 g agarose

ii. 1.5% 40 mL buffer, 0.6 g agarose

b. Big gels:

i. 1% 140 mL buffer, 1.4 g agarose

ii. 1.5% 140 mL buffer, 2.1 g agarose

2. Once the solution has cooled slightly, add Gel Red stain.

a. Mini-gels: 4 ul Gel Red (or 0.4ul Green Glow)

b. Big gels: 14 ul

Note: Gel red is the dye that stains your DNA for visualization.

Note: Gel red stain is light sensitive--keep away from light as much as possible.

3. Pour gel into mold and allow to cool completely. Don't forget the combs!

4. On a strip of parafilm, combine 4 ul PCR product and 1 ul loading dye. Pipette up and down to combine.

Note: loading dye is the dye that is used to view how far your samples have traveled in the gel during electrophoresis.

5. Reset pipettor to 5 ul. Pipette each sample into gel well.

As the amount of solution decreases (due to evaporation), you may need to reset your pipette ul setting. Avoid air bubbles in the pipette tip as this will cause the DNA to leak out. Gently pipette solution into wells.

6. Load 5 ul of DNA ladder into gel. You can use a broad range 50-10,000 bp ladder.

7. Run gel at a voltage of ~160V for approximately 20 minutes, until dye is about halfway across gel and each of the three colored bands has separated. Longer time for larger DNA fragments, larger gels.

8. Visualize gels using ImageLab software. Do not touch the computer, gel imager, or handle on gel tray with gloved-hands to avoid getting sticky buffer on equipment.

Bands for this primer set will be between 1200 and 1500 bp when comparing to DNA ladder. Sample bands may be a little smeary, but there should not be multiple bands. No bands should be visible for the negative controls.

NOTE: If sample bands are very faint (indicating low or too high DNA content), try the following alternatives (see table):

a) Modify the starting DNA concentration with 1:10 or 1:50 dilutions. Or use ½ of the DNA volume. Dilute in PCR water.

b) Reduce the volume of water and replace with BSA which increases PCR yield (also useful when bands are not amplifying).

c) If the previous troubleshooting methods do not work, is possible that DNA is too low in which case double the volume of DNA (to 4ul) or try to duplicate DNA + BSA

NOTE: If there are bands in the negative control for a sample, redo the PCR

Store PCR products at -20 C until you've accumulated all of the samples that you are going to send for sequencing moving on to Step 4.

APPENDIX H

Protocol: Analyzing Sequences using Geneious and BLAST

Prior to sending isolates to Genewiz, identify the reactions
with the isolate as part of the ID, without the underscore.
Example: G1B_A-8F → G1BA-8F

1. Log into Genewiz online (login info is on a sticky note on computer tower)
2. Select all reactions
3. Select **.ab1 file type** at the top and download all reactions
4. Save files into a folder with appropriate labeling (Ex. "VT bee round X")
5. Open Geneious and create a new folder
6. Drag and drop the sequence files from documents into the Geneious folder
7. Select all files
8. Click "**Align/Assemble**" tab and select "**De Novo Assemble**"
    a. Click "**Assemble by: 1st part of name, separated by – (Hyphen)**"
    b. Select "**Trim sequences: Options**"; default settings:
        - Ensure "Remove new trimmed regions from sequences" is selected
        - Select "Error Probability Limit: 0.05"
        - Ensure "Trim 5' End" and "Trim 3' End" are selected
10. Click OK
11. A new file of the sequence will appear, with "Assembly" in its name
12. Select one "Assembly" sequence at a time
13. Zoom into the contig view and copy the **consensus sequence**
    a. **click on "Consensus" and CTRL+C**
14. Open BLAST through NCBI online; select "Nucleotide BLAST"
15. Paste the consensus sequence into the box
    a. into the "enter query sequence" box
16. Select "rRNA/ITS database", then ensure "16S ribosomal RNA sequences" is selected
17. Click BLAST
18. Results will appear with the highest percent identification of the species at the top of the list
19. Record the species ID, max score, total score, query cover, and percent ID on Excel sheet

CURRICULUM VITAE

Dana Colley

Graduate Student
Department of Biology
Eastern Washington University
Cheney, WA 99004
Phone: (509) 368-0702 | Email: danacolley9@gmail.com

Education

Eastern Washington University	Cheney, WA	Biology	M.S., 2022
Eastern Washington University	Cheney, WA	Biology	B.S., 2020

Honors and Awards

Eastern Washington University Biology Department Spring Minigrant 2022

Northwest Scientific Association Student Grant 2021

Eastern Washington University Biology Department Winter Minigrant 2021

Graduate Assistantship, Eastern Washington University Biology Department 2020 – 2022

Outstanding Graduating Senior in Biology, Eastern Washington University 2020

Eastern Washington University Presidential Scholarship 2016 – 2020

Eastern Washington University Dean's List 2016 – 2020

Professional Experience

Research Assistant, Eastern Washington University, Cheney, WA 2022
Assisted with field sampling of small mammals on the Palouse Prairie from April – May 2022. Field sampling skills included setting up and dismantling Sherman live traps, safe handling of small mammals, identification of species, sex, and reproductive status, DNA sampling and data collection.

White-Nose Syndrome Surveillance Personnel, Washington State 2021 – 2022
Assisted with the Washington Department of Fish and Wildlife's annual White-Nose Syndrome surveillance and monitoring from April – June 2021 and May 2022. Sampled bats from 10 maternity colonies across Washington State, utilizing skills in setting up and decontaminating field equipment for bat surveillance, bat processing including assessing bat health and species, sampling bat skin microbes and recording data for use by the Washington Department of Fish and Wildlife.

Volunteer, Turnbull National Wildlife Refuge, Cheney, WA **2019 – 2020**
Analyzed and recorded field camera data for a total of six volunteer hours in February 2019. Also conducted waterfowl surveys of selected refuge units in May 2019 and analyzed and recorded data. In May – June 2020, authored a species account and video presentation for use by the refuge for a total of 10 volunteer hours.

Research Assistant, Eastern Washington University, Cheney WA **2018 – 2020**
Assisted with research projects related to biomonitoring of riparian wetlands utilizing skills in data collection, field methods, and sample storage. Field methods involved invertebrate sampling, plant sampling, and water quality monitoring using a YSI meter. Lab methods included invertebrate identification and plant identification.

Teaching Experience
Eastern Washington University, Cheney, WA
Biology Department Graduate Student Instructor **2021 – 2022**
Prepared lectures and conducted lab sessions for general biology and microbiology courses, graded labs, and provided feedback for students. Completed a total of seven hours of laboratory safety training and diversity and inclusion training to better assist students.

Biology Department Teaching Assistant **2020 – 2021**
Assisted students during class activities and labs for vertebrate zoology and wildlife management courses, graded labs and provided feedback on classroom activities.

PLUS One-on-One Tutor **2018 – 2020**
Catered to the needs of students in both group and individual tutoring sessions for organic chemistry and general biology courses, prepared lesson plans for tutoring sessions. Completed training to become a certified Level II Advanced Tutor for the College Reading and Learning Association (CRLA).

PLUS Study Group Facilitator **2018**
Created lesson plans for general biology course study group sessions, conducted interactive study group sessions catering to the needs of students.

Presentations
D. E. Colley. 2022. Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington State bats. Eastern Washington University Research and Creative Works Symposium, Cheney, Washington

D. E. Colley. 2022. Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington State bats. Northwest Scientific Association 92nd Annual Meeting, Arcata, California (Virtual)

D. E. Colley. 2020. Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington State bats. Eastern Washington University Biology Department Symposium, Cheney, Washington