MOLECULAR APPROACHES FOR ANALYZING ORGANISMAL AND ENVIRONMENTAL INTERACTIONS

by

Stephanie Fern Hudon

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Biomolecular Sciences

Boise State University

August 2020

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Stephanie Fern Hudon

Dissertation Title: Molecular Approaches for Analyzing Organismal and Environmental Interactions

Date of Final Oral Examination: 16 April 2020

The following individuals read and discussed the dissertation submitted by student Stephanie Fern Hudon, and they evaluated her presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Eric J. Hayden, Ph.D.	Chair, Supervisory Committee
Matthew L. Ferguson, Ph.D.	Member, Supervisory Committee
Jennifer Sorensen Forbey, Ph.D.	Member, Supervisory Committee

The final reading approval of the dissertation was granted by Eric J. Hayden, Ph.D., Chair of the Supervisory Committee. The dissertation was approved by the Graduate College.

DEDICATION

I dedicate this to my husband, Tyler, without whom this would never have happened, and to our children, Amelia, Evangeline, Ivy, Ryan and Ridley, who have sacrificed their time with me to allow me to pursue my dreams. I also dedicate this to my mother, Nancy, who has always believed in my potential and seen me through every chapter of life. I also dedicate this to my in-laws, Linda and Jim, who have edited papers and listened with enthusiasm to all of my ideas, as well as served as surrogate parents to our kids during times when I needed to travel to complete this work.

ACKNOWLEDGEMENTS

Foremost I would like to thank my husband Tyler for his hard work and countless sacrifices that have made this happen. He dealt with my long hours and time away from our family for years to make this possible. He took on both of our roles when needed and supported our family selflessly. I would also like to thank our amazing kids – Amelia, Evangeline, Ivy, Ryan and Ridley. You kept me smiling and grounded and frequently robbed me of much needed sleep. Watching you grow into unique individuals, with so many talents, inspires me every day. I would also like to thank my mom, Nancy Kelly, who single-handedly raised me and who has always been my biggest cheerleader, and my in-laws whose visits gave me a reprieve from the chaos and who selflessly watched kids, took on household chores and repairs and listened to all my ideas.

I would also like to thank my advisor, Dr. Eric Hayden. He is an amazing mentor who has allowed everyone in his lab to find work-life balance and has given me the freedom to pursue research which took us in many new directions and required a lot of extra coursework and training to achieve. He keeps an open door and is always available to talk about life or research over a cup of coffee. We all look forward to seeing his bike outside his door and line up to chat with him. I would also like to thank Dr. Jennifer Sorensen Forbey who has taken an active role as both mentor and committee member. Her unmatched energy and enthusiasm are contagious and I am constantly amazed at the breadth of her influence in creating multidisciplinary collaborations. I am certain she has more hours in the day than the rest of us. I would also like to thank Dr. T. Trevor

Caughlin who was my mentor for the second chapter of this dissertation. His skills in mathematics, modeling, coding and writing are an inspiration, and I have learned so much from our time together. He read, co-authored and edited countless drafts of our manuscript and I found his quiet, steady demeanor to always be a calming influence in my day.

This research was supported by Boise State University (Biomolecular Sciences Graduate Programs), the National Science Foundation (Grant Nos. OIA-1738865, OIA-1826801, OIA-1757324, 1807809, IOS-1258217, DEB-1146194, DEB-1146368) and a National Aeronautics and Space Administration (Grant No. 80NSSC17K0738), an Idaho State Board of Education grant IGEM19-002, Pittman-Robertson 683 funds from the Idaho Department of Fish and Game, Sigma Xi Grants-in-Aid, a Bureau of Land Management Grant #L09AC16253, and an INBRE Bioinformatics Core Technology Access program grant.

ABSTRACT

Our planet is undergoing rapid change due to the expanding human population and climate change, which leads to extreme weather events and habitat loss. It is more important than ever to develop methods which can monitor the impact we are having on the biodiversity of our planet. To influence policy changes in wildlife and resource management practices we need to provide measurable evidence of how we are affecting animal health and fitness and the ecosystems needed for their survival. We also need to pool our resources and work in interdisciplinary teams to find common threads which can help preserve biodiversity and vital habitats. This dissertation showcases how improved molecular biology assays and data analysis approaches can help monitor the fitness of animal populations within changing ecosystems.

Chapter 1 details the development of a universal telomere assay for vertebrates. Recent work has shown the utility of telomere assays in tracking animal health. Telomere lengths can predict extinction events in animal populations, life span, and fitness consequences of anthropogenic activity. Telomere length assays are an improvement over other methods of measuring animal stress, such as cortisol levels, since they are stable during capture and sampling of animals. This dissertation provides a telomere length assay which can be used for any vertebrate. The assay was developed using a quantitative polymerase chain reaction platform which requires low DNA input and is rapid. This dissertation also demonstrates how this assay improves on current telomere assays developed for mice and can be used in a vertebrate not previously

assayed for telomere lengths, the American kestrel. This work has the potential to propel research in vertebrate systems forward as it alleviates the need to develop new reference primers for each species of interest. This improved assay has shown promise in studies in mouse cell line studies, American kestrels, golden eagles, five species of passerine birds, osprey, northern goshawks and bighorn sheep.

Chapter 2 presents a machine learning analysis, using a topic model approach, to integrate big data from remote sensing, leaf area index surveys, metabolomics and metagenomics to analyze community composition in cross-disciplinary datasets. Topic models were applied to understand community organization across a range of distinct, but connected, biological scales within the sagebrush steppe. The sagebrush steppe is home to several threatened species, including the pygmy rabbit (*Brachylagus idahoensis*) and sage-grouse (*Centrocercus urophasianus*). It covers vast swaths of the western United States and is subject to habitat fragmentation and land use conversion for both farming and rangeland use. It is also threatened by increases in fire events which can dramatically alter the landscape. Restoration efforts have been hampered by a lack of resources and often by inadequate collaboration between stakeholders and scientists. This work brought together scientists from four disciplines: remote sensing, field ecology, metabolomics and metagenomics, to provide a framework for how studies can be designed and analyzed that integrate patterns of biodiversity from multiple scales, from the molecular to the landscape scale. A topic model approach was used which groups features (chemicals, bacterial and plant taxa, and light spectrum) into "communities" which in turn can be analyzed for their presence within individual samples and time points. Within the landscape, I found communities which contain encroaching plant species, such as juniper

(*Juniperus* spp.) and cheatgrass (*Bromus tectorum*). Within plants, I found chemicals which are known toxins to herbivores. Within herbivores, I identified differences in bacterial taxonomical communities associated with changes in diet. This work will help to inform restoration efforts and provide a road map for designing interdisciplinary studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DISSERTATION INTRODUCTION

Monitoring organismal and ecosystem health

There is an urgent need to monitor the health of wildlife and the ecosystems of which they are a part. The earth is undergoing significant changes due to anthropogenic effects, which are fragmenting and destroying habitats, polluting ecosystems and causing global warming. It is likely that these activities are leading to the sixth mass extinction, with 322 land vertebrates having gone extinct in the last 500 years (Dirzo *et al.,* 2014; Burgess, 2019). These extinction events do not account for additional population declines in the remaining species. Although conservationists and wildlife managers are trying to preserve and restore vital habitats to slow the rate of species decline and extinction, there is much work to be done. Efforts are in part hampered by a lack of resources and, in many cases, the difficulty of stakeholders and researchers to develop collaborations and shared goals (Sievanen *et al.,* 2012; Pujadas Botey *et al.,* 2014).

There is an opportunity to develop better methods and data analysis tools to track the health of animals and their habitats and to pave the way for interdisciplinary research teams to share a common language and integrate disparate data sets across scales and disciplines. The focus of this dissertation is to improve molecular biology assays and data analysis approaches, which can aid ecologists and conservationists in the preservation and restoration of ecosystems, while also providing a road map for how an interdisciplinary team can co-analyze data sets to improve conservation outcomes. Although many molecular approaches have been developed to monitor human health and

disease, there has been slower progress in other vertebrate organisms due to the difficulty in transferring these assays to genomes and environmental conditions which differ from humans in significant ways.

One emerging molecular method is to measure telomere lengths as an indicator of aging and fitness. However, this assay can be challenging to adapt to new organisms due to a lack of published genomes in many non-model organisms, which makes the development of reference primers challenging. Another important area of research for monitoring the health of an animal is through analysis of the gut microbiome. This type of analysis, when merged with other environmental data, can inform conservationists in their efforts to understand how animals adapt to climate change and other human disturbances and can better inform reintroduction efforts after local extinction events. In this dissertation, I present advancements to two emerging molecular methods for monitoring the health of animals. In Chapter 1, I present an assay to measure fitness and stress responses in *any* vertebrate through telomere length assessment. In Chapter 2, I present an integrated study of the sagebrush steppe, including microbiome analysis of a sagebrush herbivore, plant chemistry data and landscape surveys through remote sensing and leaf area index. I show how topic modeling can identify community structure in each of these data sets and lead to better integration of multi-disciplinary studies.

Telomere background information

Telomeres are critical structures for protecting the ends of linear chromosomes. Telomeres are long, non-coding, sequences of repeating nucleotides (TTAGGG)n which cap the ends of linear chromosomes. These repeats are the same in all eukaryotes, with similar sequences found in organisms across the tree of life (Watson and Riha, 2010;

Gomes *et al.,* 2011). These repeats are bound by a six subunit protein complex called shelterin which protects the single-stranded DNA from DNA damage machinery following strand invasion of the single-stranded DNA to form a lariat-like structure (Greider, 1991; Lange, 2005) (Figure 1).

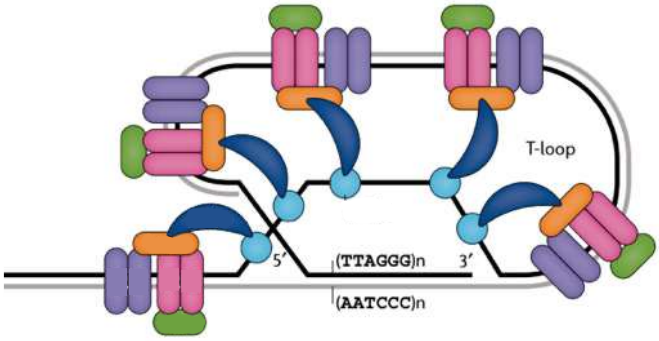


Figure 1. Telomere repeats bound by shelterin protein complex

Despite this structure formation, telomere shortening of the lagging strand still occurs during every cell division due to the end replication problem (Waga and Stillman, 1998). During DNA synthesis, lagging strand synthesis occurs opposite the direction of the replication fork movement, which involves continuous use of RNA primers to generate Okazaki fragments (Read and Brenner, 2001). The DNA template at the end of the lagging strand is unable to be primed, and therefore, the lagging strand is slightly shorter during every round of cell division (Figure 2). This shortening eventually leads to cell senescence (Shay, 2018). Some stems cells, germ cells and most cancer cells express the enzyme telomerase, which is capable of adding new TTAGGG-3' repeats to avoid this shortening.

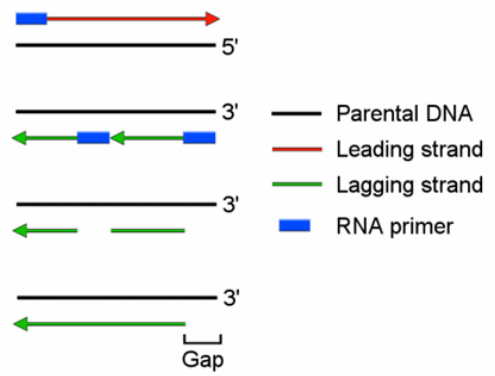


Figure 2. End replication problem.
DNA at the end of eukaryotic chromosome is not able to be fully copied during DNA synthesis, resulting in progressive shortening of telomeric sequences.

In addition to the shortening of telomeres, which occurs during DNA synthesis, increases in the rate of telomere attrition occur from exposure to stress (Sahin *et al.,* 2011; Martens and Nawrot, 2016; Zhang *et al.,* 2016; Barnes *et al.,* 2019). Research into the mechanism of this increased rate of shortening has shown that this likely occurs through oxidative damage from environmental and genetic factors. This has made the measurements of telomere lengths a useful tool for monitoring the impact of stress and disease in humans (Bär and Blasco, 2016; Donati and Valenti, 2016; Factor-Litvak *et al.,* 2016). Despite this potential, and a growing need to assess the response of animals to environmental stressors, less progress has been made using telomere assays in other non-human and non-model vertebrates.

Telomere assays as measurements of animal fitness and health

Despite the slower pace in the use of telomere assays in wildlife studies, there have been several pivotal studies which demonstrate the importance of this approach as a measure of fitness and lifespan. In one recent study by Dupoué *et al.,* common lizards (*Zootoca vivipara*) were sampled along an extinction risk gradient and found to have shorter telomere lengths in populations facing a high risk of extinction. The authors of

this study propose that telomere lengths can be used as a biomarker to predict extinction events (Dupoué *et al.,* 2017). In another study by Whittemore *et al.,* they demonstrate that the rate of telomere shortening can be used to predict the lifespan of a wide variety of birds and mammals and that this is independent of the starting telomere length of individual animals (Whittemore *et al.,* 2019). Other studies have shown a link between anthropogenic causes, such as urbanization, traffic noise and coal exposure, in telomere shortening in vertebrates, indicating that telomere lengths can also be used to measure these stressors (Meillère *et al.,* 2015; Ibáñez-Álamo *et al.,* 2018; Matzenbacher *et al.,* 2019). There have also been several studies which show a direct effect of disease and parasites on telomere shortening in animals (Beirne *et al.,* 2014; Asghar *et al.,* 2015). These studies point to the wide application that telomere assays can have to help understand the health of animal populations.

Developing a universal telomere assay for all vertebrates

Several approaches have been used to measure telomere lengths. Terminal Restriction Fragment (TRF) analysis was the initial approach used to measure the length of telomeres (Mender and Shay, 2015). This approach involves using a labelled probe to the TTAGGGn sequences and performance of a Southern blot to analyze the intensity of telomere smears after digestion by restriction enzymes of all other genomic DNA. This method has long been considered the gold standard of telomere length measurements but is labor intensive and requires large amounts of DNA (approximately 3 µg), which can be particularly difficult to obtain from wildlife. Fluorescent in-situ hybridization (FISH) is another approach where probes specific to the telomere repeats are bound and fluorescence is measured using methods such as flow cytometry (Baerlocher *et al.*, 2006).

This approach allows for telomere measurements in specific cell populations but is also time-consuming, expensive and requires significant technical skills due to the calibrations needed throughout the protocol (Baerlocher *et al.*, 2006). Another approach, single telomere length analysis (STELA), is a ligation-PCR-based method which is capable of measuring the telomere length of a single telomere using primers designed to specific chromosomes (Hemann *et al.*, 2001). This method is unique in its ability to identify critically short telomeres within a population of cells.

Chapter 1 of this dissertation details the development of a quantitative polymerase chain reaction (qPCR) based universal telomere assay for all vertebrates. Many researchers are now measuring relative telomere lengths by qPCR, due to several advantages of this assay (Cawthon, 2002). This approach can be performed with much less DNA (approximately 75 ng) and can use high-throughput platforms, speeding the rate of discovery. In this approach the amount of telomere signal (T) is normalized to reference genes signals (R) and the relative telomere length calculated as a ratio between these signals $T/R_{average}$. The bottleneck to this approach is the development of reference gene primers which are ubiquitous across vertebrates and do not exhibit copy number variants across individuals. This dissertation addresses this key technical issue by developing a qPCR based, universal telomere assay, with five pairs of reference gene primers, validated across the vertebrate tree of life.

Microbiome background information

The microbiome is defined as the collection of microorganisms, in particular bacteria, which live within an environment (Ursell *et al.*, 2012). The study of microbiomes has benefited from advances in whole genome sequencing technology and

from decreased costs. It took nearly 13 years to sequence the human genome, while today we can generate millions of reads in a matter of hours using Next Generation Sequencing. This has led to the ability to rapidly decode the microbiome of individuals or an environment from a single DNA sample. Bacteria inhabit many body systems including the gut, skin and mouth. Changes to health status, diet and the environment can have profound effects on bacterial populations, in both the number and type of bacteria present (Althani *et al.,* 2016; Lynch and Pedersen, 2016). Progress has been rapid in this field, with large projects such as the NIH Human Microbiome Project, established in 2008, contributing to the momentum. Over 200 million dollars is now spent annually on microbiome research (Proctor *et al.,* 2019). However, as with telomere research, the field of generating and understanding microbial communities and their function has grown faster in human health than in other vertebrates.

Analyzing microbiome data sets

Analysis of microbiome data sets has been performed using a variety of computational and modeling approaches. However, many of them do not lead to insights into the structure of bacterial communities or how these communities may interact with other features in the environment, such as chemical profiles. Common approaches to analyzing microbial communities include estimates of alpha and beta diversity and multivariate statistical analysis, such as principal coordinate analysis. These can give us an idea about community composition but do not offer insights into potential interactions across microbial communities or interactions with other communities at other biological scales or scientific disciplines. For example, identifying how changing plant communities influence the type and concentration of plant metabolites consumed by herbivores, which

influence bacteria populations exposed to those metabolites in the guts of herbivores, will help us unravel the functional interactions of these changes. Moving from describing populations of bacteria and chemicals to finding connections between them and to the environment is needed to start to understand functional biodiversity and how it is affected by various stressors (Shoaie *et al.,* 2013; Dorrestein *et al.,* 2014). Several clustering techniques have been used to try to group data points together, but these often fail to find more minor communities (LaMontagne, *et al.*; Widder *et al.,* 2016). This need has led to an increased interest in computational approaches to develop deeper analysis of microbial and metabolic communities. This dissertation (Chapter 2) shows how a topic modeling approach, Latent Dirichlet Allocation (LDA), can help to start to define community structure in datasets including landscape features, metabolomics and metagenomics.

LDA is a method for identifying microbial communities which has some unique advantages over traditional clustering approaches. Normal, unsupervised clustering approaches, which are frequently employed to analyze microbiome datasets, do not unravel the latent community structure in bacterial samples that are biologically relevant (Yan *et al.,* 2017). Importantly, these approaches do not allow bacteria taxa to reside in multiple clusters (mixed community membership). In contrast, LDA allows for mixed membership of bacteria across multiple communities. LDA is a Bayesian based method which applies topic modeling, frequently employed in text mining, to uncover hidden structure in datasets (Blei *et al.,* 2003). The basic concept of LDA is that there are multiple possible bacterial communities and each type of bacteria has some probability of existing in each community. In addition, LDA allows for each microbial sample collected to have multiple communities of bacterial taxa. LDA is used to predict the probability of

each community existing within a given sample (Chen *et al.,* 2012). This represents a middle ground between clustering approaches (well-separated) and ordination approaches (continuous gradients). There was a recent demonstration of the effectiveness of LDA through the analysis of the human gut microbiome (Hosoda *et al.,* 2019). Specifically, LDA identified a fourth classification of bacterial ecosystems in the human gut (enterotype) that unsupervised clustering was not able to identify. LDA's ability to identify these communities of bacteria is an important step towards more accurately identifying microbial interactions that explain their collective effect on host health and well-being. In Chapter 2 of this dissertation I show that LDA can be used to identify communities of bacteria in a sagebrush herbivore that change in response to diet transitions. In addition, I show how LDA can be used to also identify community structure in a range of data types including remote sensing data, metabolomic data and field ecology data.

Developing an approach to integrate and model big data across scientific disciplines

Defining communities of microbes and understanding these communities within the context of environments, at increasing biological scales, will give us new insights into animals and their habitats (Bahrndorff *et al.,* 2016; Xiong, 2018; Allan *et al.,* 2018). Many studies demonstrate a strong association between microbial communities and various environmental stressors, but there is further progress to be made to understand the causality and mechanisms behind the interactions of microbes and their environments. For example, how do the metabolic products produced by one community of microbes influence another microbe and how does host diet affect this interaction? How can scientists, from disparate disciplines, collaborate to better understand how animals and

their microbiomes change in response to the environment? These questions require a new approach to how we analyze large, multi-faceted datasets and a common language to use across disciplines. In this dissertation, I demonstrate the potential for community structure to unite patterns of biodiversity across molecular, organismal and landscape scales.

This dissertation (Chapter 2) demonstrates how a probabilistic, topic modeling approach, Latent Dirichlet Allocation (LDA), can be used to understand community interactions in multi-disciplinary studies (Blei *et al.*, 2003). With a specific focus on how large spatial scales of communities of landscape features from remote sensing and field ecology data can cascade down to explain communities of microbes within herbivores. In this study, we define communities as groups of features that are likely to be found together. These features are discipline specific. For example, in microbiome data these features are bacterial taxa and in field ecology these features are plant species. LDA allows us to find communities of these discipline-specific features and to calculate the probability of the features exiting within a community. Importantly, LDA allows for the existence of multiple communities within a single sampling unit, which prevents the overrepresentation of majority elements (Sankaran and Holmes, 2019). This means that a feature, such as the bacteria genus *Clostridium,* could be placed in multiple communities and in turn several of these communities could be found within a given sampling unit such as an individual herbivore.

In Chapter 2, I apply LDA to a sagebrush steppe environment and look at four distinct data types representing different biological scales and disciplines. Data focuses on the sagebrush steppe ecosystem and includes hyperspectral data from the landscape

acquired by remote sensing, leaf area index of plant plots, metabolomics of sagebrush subspecies and metagenomics of a sagebrush herbivore. The sagebrush steppe covers 165 million acres in the western United States and is threatened by changes in land use as well as an increase in fires, due to climate change (Baker, 2006). Understanding how the environment shapes the lives of the plants and animals living in the sagebrush steppe will allow us to develop conservation and wildlife management approaches with improved outcomes. I demonstrate how LDA can be used to find distinct community structures in landscape ecology, metabolome and microbiome data sets and provide a framework for developing future multidisciplinary studies across scales.

Advancements in molecular approaches to analyze organismal and environmental interactions

This dissertation provides a molecular assay to monitor vertebrate health, fitness and lifespan. The universal telomere assay presented in this dissertation has been used to analyze the effects of parasitism and anthropogenic factors on new, non-model organisms. I also provide a topic modeling approach to analyze big data sets, thus allowing the integration of data from the fields of landscape ecology, metabolomics and metagenomics towards the goal of understanding interactions of animals and their environment. This approach was able to identify invasive species in the landscape, metabolites which can differentiate plant species and predict selection by herbivores and shifts in bacterial communities in response to diet transitions in a sagebrush herbivore. Together these contributions will add to our toolshed in the quest to understand the effects of climate change and human activity on animals and the environment and

provide an approach for the development of interdisciplinary studies across disparate fields of study.

References

Allan N, Knotts T, Pesapane R, *et al.* 2018. Conservation Implications of Shifting Gut Microbiomes in Captive-Reared Endangered Voles Intended for Reintroduction into the Wild. *Microorganisms* **6**: 94–94.

Althani AA, Marei HE, Hamdi WS, *et al.* 2016. Human Microbiome and its Association With Health and Diseases. *J Cell Physiol* **231**: 1688–94.

Asghar M, Hasselquist D, Zehtindjiev P, *et al.* 2015. Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. **347**: 9–12.

Baerlocher GM, Vulto I, Jong G de, and Lansdorp PM. 2006. Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nat Protoc* **1**: 2365–76.

Bahrndorff S, Alemu T, Alemneh T, and Lund Nielsen J. 2016. The Microbiome of Animals: Implications for Conservation Biologyhttps://www.hindawi.com/journals/ijg/2016/5304028/abs/. Viewed 20 Dec 2019.

Baker WL. 2006. Fire and Restoration of Sagebrush Ecosystems. *Wildl Soc Bull* **34**: 177–85.

Bär C and Blasco MA. 2016. Telomeres and telomerase as therapeutic targets to prevent and treat age-related diseases. *F1000Research* **5**.

Barnes RP, Fouquerel E, and Opresko PL. 2019. The impact of oxidative DNA damage and stress on telomere homeostasis. *Mech Ageing Dev* **177**: 37–45.

Beirne C, Delahay R, and Hares M. 2014. Age-Related Declines and Disease-Associated Variation in Immune Cell Telomere Length in a Wild Mammal. *PLOS ONE* **9**: e108964.

Blei DM, Ng AY, and Jordan, Michael I. 2003. Latent Dirichlet Allocation. *J Mach Learn Res* **3**: 993–1022.

Burgess S. 2019. Deciphering mass extinction triggers. *Science* **363**: 815–6.

Cawthon RM. 2002. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**: 47e–47.

Chen X, He T, Hu X, *et al.* 2012. Estimating functional groups in human gut microbiome with probabilistic topic models. *IEEE Trans Nanobioscience* **11**: 203–15.

Dirzo R, Young HS, Galetti M, *et al.* 2014. Defaunation in the Anthropocene. *Science* **345**: 401–6.

Donati B and Valenti L. 2016. Telomeres, NAFLD and Chronic Liver Disease. *Int J Mol Sci* **17**: 383.

Dorrestein PC, Mazmanian SK, and Knight R. 2014. Finding the Missing Links among Metabolites, Microbes, and the Host. *Immunity* **40**: 824–32.

Dupoué A, Rutschmann A, Le Galliard JF, *et al.* 2017. Shorter telomeres precede population extinction in wild lizards. *Sci Rep* **7**: 1–8.

Factor-Litvak P, Susser E, Kezios K, *et al.* 2016. Leukocyte Telomere Length in Newborns: Implications for the Role of Telomeres in Human Disease. *Pediatrics* **137**.

Gomes NMV, Ryder OA, Houck ML, *et al.* 2011. Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell* **10**: 761–8.

Greider CW. 1991. Telomeres. *Curr Opin Cell Biol* **3**: 444–51.

Hemann MT, Strong MA, Hao L-Y, and Greider CW. 2001. The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability. *Cell* **107**: 67–77.

Hosoda S, Nishijima S, Fukunaga T, *et al.* 2019. Revealing microbial assemblage structure in the human gut microbiome using latent Dirichlet allocation. *bioRxiv*: 664219.

Ibáñez-Álamo JD, Pineda-Pampliega J, Thomson RL, *et al.* 2018. Urban blackbirds have shorter telomeres. *Biol Lett* **14**: 20180083.

LaMontagne M, Shetty T, Gajjar T, *et al.* HABase: A Web-Application for the Analysis of Protein Spectra and Identification of Microbial Species. : 2.

Lange T de. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**: 2100–10.

Lynch SV and Pedersen O. 2016. The Human Intestinal Microbiome in Health and Disease. *N Engl J Med* **375**: 2369–79.

Martens DS and Nawrot TS. 2016. Air Pollution Stress and the Aging Phenotype: The Telomere Connection. *Curr Environ Health Rep* **3**: 258–69.

Matzenbacher CA, Da Silva J, Garcia ALH, *et al.* 2019. Anthropogenic Effects on Natural Mammalian Populations: Correlation Between Telomere Length and Coal Exposure. *Sci Rep* **9**: 1–9.

Meillère A, Brischoux F, Ribout C, and Angelier F. 2015. Traffic noise exposure affects telomere length in nestling house sparrows. *Biol Lett* **11**: 20150559.

Mender I and Shay JW. 2015. Telomere Restriction Fragment (TRF) Analysis. *Bio-Protoc* **5**.

Proctor L, LoTempio J, Marquitz A, *et al.* 2019. A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. *Microbiome* **7**: 31.

Pujadas Botey A, Garvin T, and Szostak R. 2014. Interdisciplinary Research for Ecosystem Management. *Ecosystems* **17**: 512–21.

Read J and Brenner S. 2001. DNA Synthesis. In: Brenner S, Miller JH (Eds). Encyclopedia of Genetics. New York: Academic Press.

Sahin E, Colla S, Liesa M, *et al.* 2011. Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature* **470**: 359–65.

Sankaran K and Holmes SP. 2019. Latent variable modeling for the microbiome. *Biostatistics* **20**: 599–614.

Shay JW. 2018. Telomeres and aging. *Curr Opin Cell Biol* **52**: 1–7.

Shoaie S, Karlsson F, Mardinoglu A, *et al.* 2013. Understanding the interactions between bacteria in the human gut through metabolic modeling. *Sci Rep* **3**: 1–10.

Sievanen L, Campbell LM, and Leslie HM. 2012. Challenges to Interdisciplinary Research in Ecosystem-Based Management. *Conserv Biol* **26**: 315–23.

Ursell LK, Metcalf JL, Parfrey LW, and Knight R. 2012. Defining the Human Microbiome. *Nutr Rev* **70**: S38–44.

Waga S and Stillman B. 1998. The DNA replication fork in eukaryotic cells. *Annu Rev Biochem* **67**: 721–51.

Watson JM and Riha K. 2010. Comparative biology of telomeres: Where plants stand. *Febs Lett* **584**: 3752–9.

Whittemore K, Vera E, Martínez-Nevado E, *et al.* 2019. Telomere shortening rate predicts species life span. *Proc Natl Acad Sci* **116**: 15122–15127.

Widder S, Allen RJ, Pfeiffer T, *et al.* 2016. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* **10**: 2557–68.

Xiong W. 2018. Linking land management, the soil microbiome and suppression of Fusarium wilt diseasehttp://dspace.library.uu.nl/handle/1874/363529. Viewed 20 Dec 2019.

Yan J, Chuai G, Qi T, *et al.* 2017. MetaTopics: an integration tool to analyze microbial community profile by topic model. *BMC Genomics* **18**: 962.

Zhang J, Rane G, Dai X, *et al.* 2016. Ageing and the telomere connection: An intimate relationship with inflammation. *Ageing Res Rev* **25**: 55–69.

CHAPTER ONE: UNIVERSAL ASSAY FOR MEASURING VERTEBRATE TELOMERES BY REAL-TIME QUANTITATIVE PCR

Stephanie F. Hudon[1*], Esteban Palencia Hurtado[2], James D. Beck[3], Steven J. Burden[1], Devin P. Bendixsen[1], Kathleen R. Callery[2], Jennifer S Forbey[2], Lisette P. Waits[4], Robert A. Miller[5], Ólafur K. Nielsen[6], Julie A. Heath[2] and Eric J. Hayden[2]

[1] Biomolecular Sciences, Boise State University, Boise, ID 83725
[2] Department of Biological Sciences, Boise State University, Boise, ID 83725
[3] Computational Science and Engineering, Boise State University, Boise, ID 83725
[4] Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID 83844
[5] Intermountain Bird Observatory, Boise, ID 83725
[6] Department of Ecology, Icelandic Institute of Natural History, Garðabær 210, Iceland

Abstract

Telomere length dynamics are an established biomarker of health and aging in animals. The study of telomeres in numerous species has been facilitated by methods to measure telomere length by real-time quantitative PCR (qPCR). In this method, telomere length is determined by quantifying the amount of telomeric DNA repeats in a sample and normalizing this to the total amount of genomic DNA. This normalization requires the development of genomic reference primers suitable for qPCR, which remains challenging in non-model organisms with genomes that have not been sequenced. Here we report reference primers that can be used in qPCR to measure telomere lengths in any vertebrate species. We designed primer pairs to amplify genetic elements that are highly conserved between evolutionarily distant taxa and tested them in species that span the vertebrate tree of life. We report five primer pairs that meet the specificity and

reproducibility standards of qPCR. In addition, we demonstrate how to choose the best primers for a given species by testing the primers on multiple individuals within a species and applying an established computational tool. These reference primers can facilitate the application of qPCR-based telomere length measurements in any vertebrate species of ecological or economic interest.

Introduction

The measurement of telomere lengths is an important approach used to study the health and aging of organisms. Telomeres are structures at the end of Eukaryotic chromosomes that are comprised of proteins bound to repetitive DNA sequences. Telomeres protect the ends of linear chromosomes and provide several important cellular functions (1). The length of telomere DNA shortens at each cell division, and telomere shortening can eventually lead to cellular senescence, which affects tissue function, organismal health and lifespan (2–4). Telomere lengths are considered an indicator of phenotypic quality (5, 6) and telomere length dynamics have predictive power in the future success of organisms. For example, telomere shortening has been shown to predict both lifespan and reproductive success (7, 8), which are proxies for organismal fitness. Short telomere lengths have even been shown to precede extinction events (9). In addition, telomere shortening has been shown to be accelerated by various forms of stress (10–12) and telomere lengths have been used to evaluate environmental quality (13–15). The ability to monitor the health of organisms and to predict their future success has important applications in both wild and captive environments. In addition, understanding telomere dynamics across the vertebrate tree of life could also lead to a better understanding of the mechanisms and evolution of aging in general.

Despite the motivation for measuring telomere lengths, research remains limited to relatively few species, with the majority of studies done in avian species (16–18). Methodological challenges are one reason for the limited use of telomere length measurements. A widely used method is to measure relative telomere length by real-time quantitative PCR (qPCR). This method requires only common lab equipment and techniques, can be done in relatively high-throughput and is robust despite low sample quantity and quality (19, 20). Measurement of telomere length by qPCR requires primers that amplify telomere repeats, where the concentration of telomeric DNA in a sample, determined by qPCR, is proportional to telomere length. The method also requires reference primers that amplify a non-telomeric region of the genome of interest to normalize for the total amount of genomic DNA in the sample (21). The repeating DNA sequence of telomeres (TTAGGG) is identical in all vertebrates, so the established telomere-specific primer pairs should work in any vertebrate species (22). In contrast, the genome specific reference primers optimized in one species may not work in other species because of unknown genetic differences, often requiring the design and optimization of new primers for each new species that is to be studied (23, 24). Reference primer development is especially challenging in non-model organisms with genomes that have not been sequenced. Despite the simplicity of measuring telomere lengths by qPCR, reference primer design may limit the adoption of this assay in newly analyzed species.

Here we report PCR primers that can be used as genomic reference primers for qPCR based-telomere length measurements in any vertebrate species. To design a PCR assay that would work universally in vertebrates, we designed primers to amplify ultra-conserved elements (UCEs) which are genetic elements that are highly conserved

between evolutionarily distant taxa (25–27). We identified primer pairs that amplify five different UCEs and also meet qPCR requirements for specificity and efficiency (28) (Figure 1A). We designed these primers to match the annealing temperature of the established telomere primers so that they can be used in the same PCR plate if desired. We found that all five primers efficiently and reproducibly amplify genomic DNA in every vertebrate sample that we tested. In addition, the targeted UCEs do not vary in copy number in the species we investigated. As a proof of concept, we used the primers to verify a significant decrease in telomere length with age in American kestrels, a species for which telomere lengths had not previously been measured. The reported qPCR primers enable the expansion of telomere length studies to new vertebrate systems for basic and applied research questions.

Materials and Methods

Genomic DNA extraction

DNA was extracted from blood or tissue samples using the Zymo Quick-DNA Microprep Plus Kit (#D4074) according to the manufacturer's protocols for the sample type. Blood samples from Northern goshawk and American kestrel were stored in Queen's lysis buffer (0.01 M Tris, 0.01 M NaCl, 0.01 M EDTA, and 1% n-lauroylsarcosine, pH 7.5) prior to extraction. Human DNA was from a buccal swab. DNA from a red-eared slider was extracted from a shell fragment. DNA from sea squirt was extracted from whole organisms. All other DNA samples were extracted from muscle or liver tissue. All animals were treated in accordance with Boise State University animal care and use policies where applicable. DNA purity was assessed by 260/280 nm

absorbance ratio (NanoDrop). DNA was quantified by absorbance at 260 nm and the concentration was normalized to 2ng/µl in 10 mM Tris-HCl pH 8.5 and 0.1 mM EDTA.

Quantitative PCR

Quantitative PCR (qPCR) was performed using a Roche LC96. A standard curve for each primer pair was included in triplicate on each qPCR plate. Genomic DNA from the organism of interest which had the highest concentration, measured by UV absorbance, was used for the standard curve. The standard curve was prepared by seven serial dilutions (1:5). Reactions were carried out in 20 µL volumes containing approximately 8 ng DNA for unknown samples (or 4 µl of varying concentrations of DNA for serial dilutions), 10 µl of 2x Biotium Fast Plus EvaGreen® qPCR Master Mix, 10 pmol each of forward and reverse primers (500 nM final primer concentration) and water up to 20 µl. For UCE and telomere primers, the two-step thermal cycling profile was 95°C for 2 min, followed by 40 cycles of 95°C for 5 s and 55°C for 30 s, with signal acquisition at the end of the 55°C step and melt curves generated by increasing temperatures from 72 to 95°C, in 0.5°C steps, with a 30 s dwell period per step at the end of the thermal cycling. Reactions with previously reported primers for 36B4 in mice contained a 250 nM final concentration of forward and reverse primers with all other reagents as described above. PCR reactions with the 36B4 primers were thermal cycled for 95°C for 2 min, followed by 35 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 30 s with a final extension for 5 min at 72°C and signal acquisition at the end of the 72°C step and melt curves generated by increasing temperatures from 72 to 95°C, in 0.5°C steps, with a 30 s dwell period per step at the end of the thermal cycling.

All samples were PCR amplified in triplicate along with three no template controls per primer pair. Triplicates were averaged and any sample with a quantification cycle (Cq) standard deviation of greater than 0.5 between triplicates or with Cq values outside of the standard curve were repeated. Concentrations (ng) of telomere repeats and UCEs were determined using an external standard curve approach and Roche LightCycler software (release 4.0) for Absolute Quantification auto-analysis with the second derivative maximum method (proprietary). The efficiencies of each primer pair were determined from the slope of the standard curves using Roche LightCycler software where $E=10^{[-1/slope]}$. Optimal efficiency (100%) is defined as a slope of 3.32. The correlation coefficients (R^2) were determined from the replicates of the dilution series.

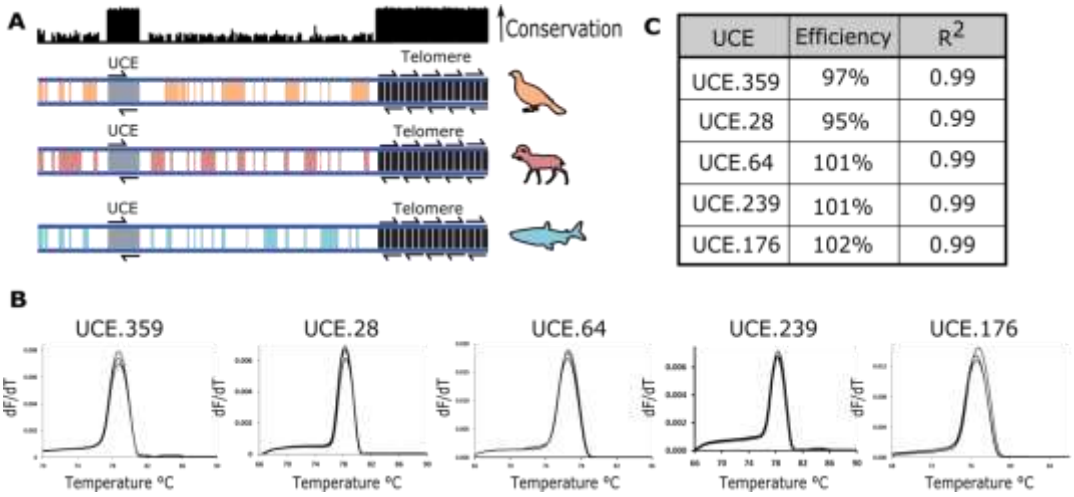
Data analysis

Primer pairs were assessed for their effect on stability values computationally (geNorm) as previously described (29). The same genomic DNA samples from multiple individuals were amplified with 3-5 UCE primer pairs, and the Cq values were used to sequentially eliminate the least stable primer pair. The two most stable UCE primer pairs were then used to generate DNA concentrations (ng) from a standard curve and these concentrations were graphed on a scatter plot with Pearson correlations and trend lines reported. To calculate relative telomere lengths in adult and nestling kestrels, the telomere concentration (T) was divided by the average concentration of two UCE genes (UCE_{ave}) to yield T/UCE_{ave}.

Results

Assay development

We set out to develop primer pairs to several UCEs to enable cross validation. We chose several UCEs from UCbase 2.0 and designed multiple primer pairs for each using a primer design tool (OligoArchitect, Sigma-Aldrich). During design, we limited the product length to 250 bps and designed primers to have a Tm of 60˚C to match the Tm of established telomere-specific primers (30). We ranked the primer pairs computationally (Beacon Designer) and experimentally tested the top four pairs for each of the UCEs by amplifying mouse DNA in a qPCR reaction. Using melt-curve analysis, we identified primer pairs for five of the UCEs that produced a single-peaked melt curve (Figure 1B). We determined the amplification efficiency for each of these five UCE primer pairs by amplifying seven serial dilutions of the mouse DNA in triplicate (Figure 1C). A primer pair for all five UCEs was found that had qPCR efficiencies within the best-practice range of 90-110% and correlations (R^2) of the Cq values for replicates greater than or equal to 0.980 (31) (Figure 1C). The sequences of the chosen UCE primer pairs is reported in Table 1.



C

UCE	Efficiency	R^2
UCE.359	97%	0.99
UCE.28	95%	0.99
UCE.64	101%	0.99
UCE.239	101%	0.99
UCE.176	102%	0.99

Figure 1.1 UCE primer performance in qPCR.

(A) Conceptual diagram of the universal vertebrate telomere assay design. Black bars represent the level of conservation of genomic regions determined by the alignment of multiple genomes such as a bird, a mammal and a fish. Telomeric repeats are conserved in all chordates and are amplified by primers specific to this sequence. The same reference primers can be used in all species because they amplify ultra-conserved elements (UCE) that have high conservation among distant taxa. (B) Melt curves of the amplification product of the qPCR of genomic mouse DNA using the best primer pair for each UCE. (C) Efficiencies and R^2 values of the best primer pair identified for five different UCEs, named by their access number in the UCE database.

Table 1.1 UCE and Telomere Primer Sequences

Name	Sequence
tel 1b	CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT
tel 2b	CAGCCGAAAGGCCCTTGGCAGGAGGGCTGCTGGTGGTCTACCCTT
UCE.359-F	ATCTGAGACTTGTGACAT
UCE.359-R	GTGTTAATTGGTAATGACTATT
UCE.28-F	AAATACCACCCAACAGTTT
UCE.28-R	AAGCCCTATACAGATGGAT
UCE.64-F	GAGTCTCCAATATCATCAGAAGC
UCE.64-R	ACACATGCCACGATCAATG
UCE.239-F	TCAGATGTTCAGCCTATT
UCE.239-R	AATACCATGTTAATTATCCTCAA
UCE.176-F	TTTCTACAGTTCTGATTTAGTTGA
UCE.176-R	TGTTCCCTGTCGCATTAG

Validation across the vertebrate tree of life

We next tested our primers across the vertebrate tree of life (Figure 2A). We collected various tissue types from 19 species for DNA extraction. To test our primers as broadly as possible, we also included a sea squirt, which is amenable to telomere length measurements because basal chordates have the same telomeric repeat sequences as vertebrates. We performed qPCR on each sample in triplicate with all five primer pairs and evaluated the melt-curves. We found that all five primer pairs amplified DNA from every species with a single melt peak. All of the UCE primer pairs also have efficiencies in the best practice range of 90-110% when used to generate a standard curve through serial dilution (32). These results indicate that all five primer pairs are suitable primers for qPCR in all vertebrates that we tested.

An additional requirement for reference genes used for normalizing telomere length is that they must not vary in copy number among individuals in the population. For example, a duplication of a reference gene in one individual would appear as a halving of their telomere lengths relative to a non-duplicated individual. The amplification of closely related homolog variants could also introduce normalization differences that would have the same effect as copy number variation. The challenge of identifying reference primers with minimal variation between individuals for telomere length measurements is similar to choosing primer pairs for normalizing real-time PCR data for gene expression. We therefore evaluated our primer pairs using an algorithm, geNorm, designed for this purpose (29). The geNorm algorithm evaluates the cumulative variation among multiple primer pairs and iteratively eliminates primer pairs that contribute the most variation until final primer pairs are chosen. Two or more primer pairs are preferred

over a single primer pair to reduce random experimental variation (33). Using the NormqPCR package in R which uses the geNorm algorithm, we identified the best combination of UCE primer pairs from DNA extracted from 20 mice (*Mus musculus*), 17 rock ptarmigan (*Lagopus muta*), and 20 northern goshawks (*Accipiter gentilis*), representing both sexes and a range of ages. Samples were amplified by qPCR with each of the five reference primer sets using the same DNA concentration (8 ng) for each reaction. We found that the use of two primer pairs for normalization led to very low stability values in each of the species tested (Figure 2B). Next, we extracted DNA from another mammal (bighorn sheep, *Ovis canadensis*) and tested the top three primer pairs which had been identified in mice. Bighorn sheep were limited to analyzing three primer pairs due to low DNA concentration provided for these samples. Using NormqPCR we identified the two primer pairs with the lowest stability values (Figure 2B). The genomic DNA concentration quantified by these two primer pairs showed high correlation (Pearson), suggesting that neither pair exhibit copy number polymorphisms within the individuals tested (Figure 2C and Supplementary Figure S1).

We next set out to determine if the UCE-based assay yields expected relative telomere length measurements in a species where telomere lengths have not been previously measured. We extracted DNA from the blood of 15 adult and 29 nestling American kestrels. We amplified each DNA sample separately with telomere specific primers UCE.28 and UCE.239 which were chosen because of low stability values and high correlations. Relative telomere length was calculated as the ratio of the telomere amplification product to the average of the UCE amplification products. The results showed that adult American kestrels had a significantly shorter telomere length than

nestlings, which was expected since telomeres tend to shorten with aging (Figure 2D). We note that, prior to developing the UCE primers, we had attempted to use published GAPDH primers to normalize telomere lengths in the American kestrel. We found that both the published (34) and extensively redesigned and optimized GAPDH primers exhibited abnormal PCR variations among individuals, making this commonly used reference gene unsuitable in this organism (Supplementary Figure S2). These results suggest that our primers may even be an improvement for telomere assays in avian species.

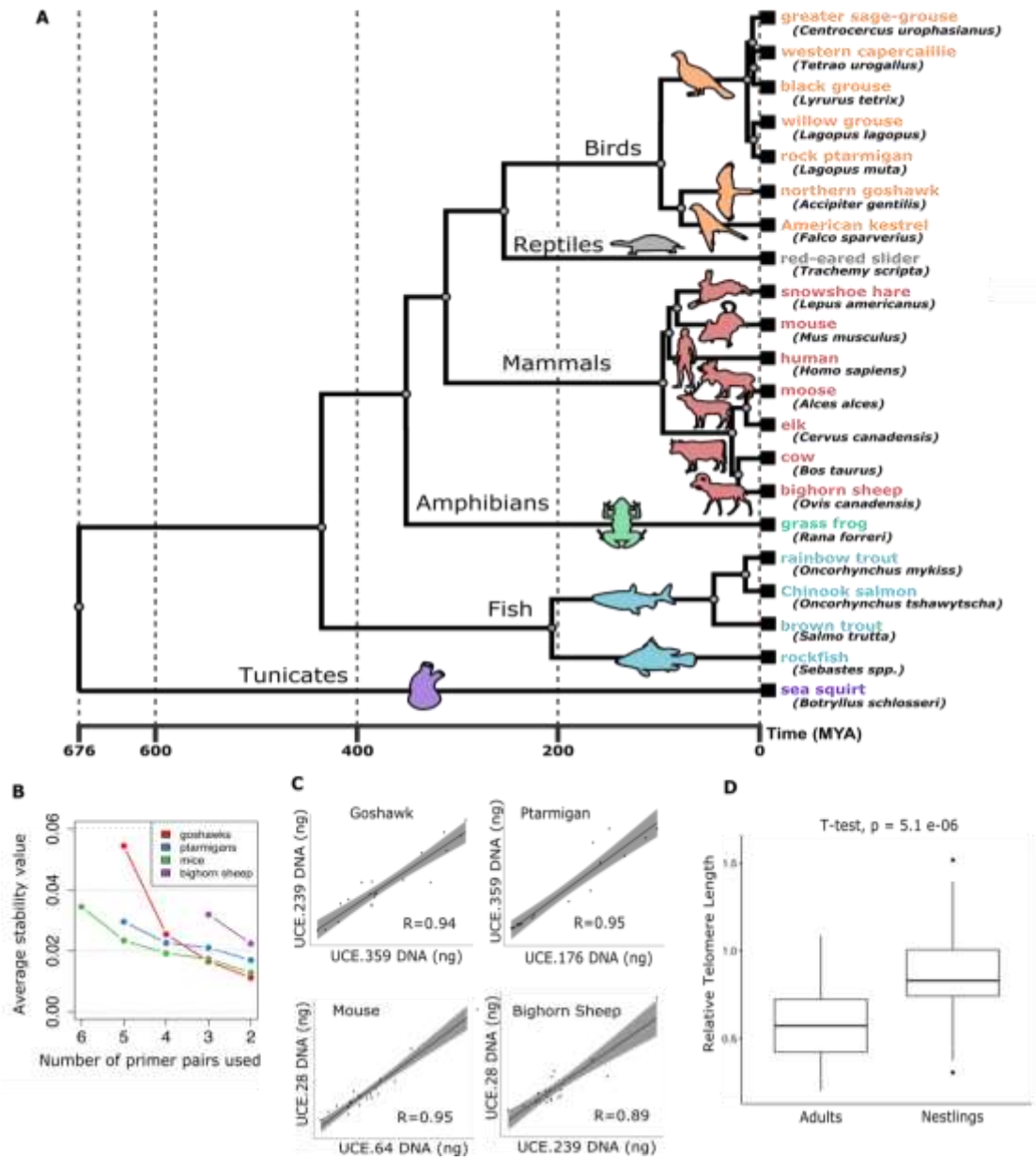


Figure 1.2 Validation of UCE reference primers across the vertebrate tree of life. (A) Phylogenetic tree of organisms in which reference primers were validated. (B) Average stability values generated in geNorm from the utilization of six to two reference gene primers. (C) Pearson correlation between the genomic DNA concentrations (ng) of the two UCE primer pairs that resulted in the lowest average stability values (geNorm) in northern goshawk (*Accipiter gentillis*), rock ptarmigan (*Lagopus muta*), mouse (*Mus musculus*) and bighorn sheep (*Ovis Canadensis*). (D) Relative telomere length for adult and nestling American kestrels (*Falco sparverius*) using UCE primers.

UCE primers are an improvement over previously reported telomere assay reference primers for mice

We next analyzed DNA from 20 mouse ear punches with a reference primer pair previously reported in the literature that was designed for qPCR-based telomere length measurement in mice (35). The previously reported primers target the acidic ribosomal phosphoprotein PO (36B4) gene and have been used in multiple publications (35). We used the published thermal cycling profile of the 36B4 primers to determine the PCR efficiency and evaluate the melt peak. The 36B4 primers had a good efficiency (102%) but showed a broad, multi-peaked melt-curve indicating PCR artifacts or non-specific amplification occurred (Supplementary Figure S3). When we included the 36B4 primer pair in the NormqPCR stability analysis for mouse, this primer pair was the second eliminated by the geNorm algorithm, indicating that our top UCE primer pairs are an improvement with respect to genomic reference stability values in mice. These results further suggest that our UCE reference primers may be useful even in organisms with established reference primers.

Discussion

Real-time qPCR-based telomere length assays provide a reliable and high-throughput method for studying the length of telomeres. The primers reported here make this assay possible in any vertebrate species. For investigators that wish to study telomeres in a new organism, we recommend first testing all five UCE primer pairs on at least ten individuals and performing a stability analysis to find the best two primer pairs (29). Compared to a single reference primer pair, using the average quantity from two reference primer pairs for relative telomere length measurement has the advantage of

averaging out other sources of qPCR noise (36). While three or more reference primers could be used if warranted, this would require more DNA and reagents. We were able to identify primer pairs with low stability values in bighorn sheep from comparing the best three primer pairs identified from mouse DNA samples (UCE.28, UCE.64 and UCE.239). This suggests that these three primers could be a good starting point for any mammalian species. However, the slightly lower correlation between the concentrations measured for the bighorn sheep samples, compared to the other organisms, suggests that a better primer pair might be found if all five primers were tested (Figure 2C). The bighorn sheep samples also had relatively low DNA concentration, which may also have contributed to the lower correlation value. We also recommend that in telomere assays in which all samples do not fit on one plate that the raw T/ UCE_{ave} ratio would be divided by the average T/ UCE_{ave} ratio of internal calibrator samples. In addition to primer choice, it is also important to establish appropriate tissue collection protocols for each new species because telomere length dynamics can vary among different types of cells and tissue (37). The biological relevance of telomere lengths must be determined in each organism which will require additional information such as diseases state and life history.

The ability to determine relative telomere length in any vertebrate species creates new opportunities in basic and applied research. For example, telomere length measurements can be used for monitoring the health and aging of organisms. Telomere lengths have even been considered a proxy for fitness because telomere lengths have been shown to predict lifespan and reproductive success in some species (5, 7). For organisms of conservation concern, telomere lengths have been shown to be an early indicator of extinction risk (9). Telomere lengths in indicator species in different

environments may allow identification of important ecosystems disturbances, including those caused by humans. The ability to study multiple species simultaneously opens up questions about how telomere lengths change during long-term predator-prey cycles or other ecological interactions. The ability to study telomere lengths in diverse taxa will allow for a better understanding of aging across the tree of life as well as from an evolutionary perspective (38).

Availability

geNorm is open source and available in the R software package NormqPCR (https://github.com/jimrperkins/NormqPCR).

Acknowledgements

We would like to thank Boise State University's Raptor Research Center, the Fulbright program, the Department of Biological Sciences, College of Arts and Sciences and Division of Research, American kestrel nest box adopters and landowners who have nest boxes on their property, the Minidoka Ranger District of Sawtooth National Forest and Idaho Fish and Game. We also would like to acknowledge Jack Connelly (for sage-grouse), Örjan Johansson and Gier Rune Rauset (for capercaillie and black grouse), Rolf Brittas and Tomas Willebrand (for willow grouse), Knut Keiland (for snowshoe hares), Henrik Andrén and the Grimsö Wildlife Research Station (for moose).

Funding

The authors acknowledge support from the National Science Foundation [1826801 to J.B., J.F. and E.H.], the National Science foundation [1807809 to E.H. and S.B.], the Semiconductor Research Corporation [2018-SB-2842 to E.H. and S.B.] and the National Aeronautics and Space Administration [80NSSC17K0738 to E.H.].

Additional funding was provided through two National Science Foundation Research Experience for Undergraduates awards, through the Division for Biological Infrastructure [1263167 to K.C. and J.H.] and the Division of Environmental Biology [1145552 to K.C. and J.H.].

Conflict of Interest

None declared.

Author's Contributions

SFH carried out the molecular lab work, participated in data analysis, participated in the design of the study and drafted the manuscript; EPH carried out molecular lab work and participated in design of the study; JDB participated in statistical analysis, data analysis and coding; SJB participated in study design and data analysis; DPB participated in study design, editing the manuscript and creating figures; KRC carried out molecular lab work and participated in data analysis; JSF critically revised the manuscript, collected field samples and provided funding; LPW critically revised the manuscript, and collected field samples; RAM critically revised the manuscript, and collected field samples; OKN critically revised the manuscript, and collected field samples; JAH critically revised the manuscript, and collected field samples, critically revised the manuscript and provided funding; EJH critically revised, conceived of the study, coordinated the study, helped draft the manuscript and provided funding. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

References

1. Gomes,N.M.V., Shay,J.W. and Wright,W.E. (2010) Telomere Biology in Metazoa. *FEBS Lett.*, **584**, 3741–3751.

2. O'Sullivan,R.J. and Karlseder,J. (2010) Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.*, **11**, 171–181.

3. McLennan,D., Armstrong,J.D., Stewart,D.C., Mckelvey,S., Boner,W., Monaghan,P. and Metcalfe,N.B. (2016) Interactions between parental traits, environmental harshness and growth rate in determining telomere length in wild juvenile salmon. *Mol. Ecol.*, **25**, 5425–5438.

4. Wilbourn,R.V., Moatt,J.P., Froy,H., Walling,C.A., Nussey,D.H. and Boonekamp,J.J. (2018) The relationship between telomere length and mortality risk in non-model vertebrate systems: a meta-analysis. *Philos. Trans. R. Soc. B Biol. Sci.*, **373**.

5. Bauch,C., Becker,P.H. and Verhulst,S. (2013) Telomere length reflects phenotypic quality and costs of reproduction in a long-lived seabird. *Proc R Soc B*, **280**, 20122540.

6. Saino,N., Romano,M., Ambrosini,R., Rubolini,D., Boncoraglio,G., Caprioli,M. and Romano,A. (2012) Longevity and lifetime reproductive success of barn swallow offspring are predicted by their hatching date and phenotypic quality. *J. Anim. Ecol.*, **81**, 1004–1012.

7. Eastwood,J.R., Hall,M.L., Teunissen,N., Kingma,S.A., Hidalgo Aranzamendi,N., Fan,M., Roast,M., Verhulst,S. and Peters,A. (2019) Early-life telomere length predicts lifespan and lifetime reproductive success in a wild bird. *Mol. Ecol.*, **28**, 1127–1137.

8. Heidinger,B.J., Blount,J.D., Boner,W., Griffiths,K., Metcalfe,N.B. and Monaghan,P. (2012) Telomere length in early life predicts lifespan. *Proc. Natl. Acad. Sci.*, **109**, 1743–1748.

9. Dupoué,A., Rutschmann,A., Le Galliard,J.F., Clobert,J., Angelier,F., Marciau,C., Ruault,S., Miles,D. and Meylan,S. (2017) Shorter telomeres precede population extinction in wild lizards. *Sci. Rep.*, **7**, 16976.

10. Bauch,C., Boonekamp,J.J., Korsten,P., Mulder,E. and Verhulst,S. (2019) Epigenetic inheritance of telomere length in wild birds. *PLOS Genet.*, **15**, e1007827.

11. Tricola,G.M., Simons,M.J.P., Atema,E., Boughton,R.K., Brown,J.L., Dearborn,D.C., Divoky,G., Eimes,J.A., Huntington,C.E., Kitaysky,A.S., *et al.* (2018) The rate of telomere loss is related to maximum lifespan in birds. *Philos. Trans. R. Soc. B Biol. Sci.*, **373**.

12. Barnes,R.P., Fouquerel,E. and Opresko,P.L. (2019) The impact of oxidative DNA damage and stress on telomere homeostasis. *Mech. Ageing Dev.*, **177**, 37–45.

13. Martens,D.S. and Nawrot,T.S. (2016) Air Pollution Stress and the Aging Phenotype: The Telomere Connection. *Curr. Environ. Health Rep.*, **3**, 258–269.

14. McLennan,D., Armstrong,J.D., Stewart,D.C., Mckelvey,S., Boner,W., Monaghan,P. and Metcalfe,N.B. (2016) Interactions between parental traits, environmental harshness and growth rate in determining telomere length in wild juvenile salmon. *Mol. Ecol.*, **25**, 5425–5438.

15. Meillère,A., Brischoux,F., Ribout,C. and Angelier,F. (2015) Traffic noise exposure affects telomere length in nestling house sparrows. *Biol. Lett.*, **11**, 20150559.

16. Bauch,C., Becker,P.H. and Verhulst,S. (2014) Within the genome, long telomeres are more informative than short telomeres with respect to fitness components in a long-lived seabird. *Mol. Ecol.*, **23**, 300–310.

17. Beaulieu,M., Reichert,S., Le Maho,Y., Ancel,A. and Criscuolo,F. (2011) Oxidative status and telomere length in a long-lived bird facing a costly reproductive event. *Funct. Ecol.*, **25**, 577–585.

18. Quirici,V., Guerrero,C.J., Krause,J.S., Wingfield,J.C. and Vásquez,R. a. (2016) The relationship of telomere length to baseline corticosterone levels in nestlings of an altricial passerine bird in natural populations. *Front. Zool.*, 10.1186/s12983-016-0133-5.

19. Martin-ruiz,C.M., Baird,D., Roger,L., Boukamp,P., Krunic,D., Cawthon,R., Dokter,M.M., Harst,P.V.D., Bekaert,S., Meyer,T.D., *et al.* (2015) Reproducibility of telomere length assessment : an international collaborative study. 10.1093/ije/dyu191.

20. Lin,J., Smith,D.L., Esteves,K. and Drury,S. (2019) Psychoneuroendocrinology Telomere length measurement by qPCR – Summary of critical factors and recommendations for assay design. *Psychoneuroendocrinology*, **99**, 271–278.

21. Cawthon,R.M. (2002) Telomere measurement by quantitative PCR. *Nucleic Acids Res.*, **30**, e47.

22. Meyne,J., Ratliff,R.L. and Moyzis,R.K. (1989) Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci.*, **86**, 7049–7053.

23. Smith,S., Turbill,C. and Penn,D.J. (2011) Chasing telomeres, not red herrings, in evolutionary ecology. *Heredity*, **107**, 372–373.

24. Horn,T., Robertson,B.C. and Gemmell,N.J. (2010) The use of telomere length in ecology and evolutionary biology. *Heredity*, **105**, 497–506.

25. Bejerano,G. (2004) Ultraconserved Elements in the Human Genome. *Science*, **304**, 1321–1325.

26. McCormack,J.E., Faircloth,B.C., Crawford,N.G., Gowaty,P.A., Brumfield,R.T. and Glenn,T.C. (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.*, **22**, 746–754.

27. Asghar,M., Hasselquist,D. and Bensch,S. (2011) Are chronic avian haemosporidian infections costly in wild birds? *J. Avian Biol.*, **42**, 530–537.

28. Bustin,S.A., Benes,V., Garson,J.A., Hellemans,J., Huggett,J., Kubista,M., Mueller,R., Nolan,T., Pfaffl,M.W., Shipley,G.L., *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.

29. Vandesompele,J., De Preter,K., Pattyn,F., Poppe,B., Van Roy,N., De Paepe,A. and Speleman,F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, research0034.1.

30. Roos,G., Kröber,A., Grabowski,P., Kienle,D., Bühler,A., Döhner,H., Rosenquist,R. and Stilgenbauer,S. (2008) Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. *Blood*, **111**, 2246–2252.

31. Kuang,J., Yan,X., Genders,A.J., Granata,C. and Bishop,D.J. (2018) An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. *PLOS ONE*, **13**, e0196438.

32. Rogers-Broadway,K.-R. and Karteris,E. (2015) Amplification efficiency and thermal stability of qPCR instrumentation: Current landscape and future perspectives. *Exp. Ther. Med.*, **10**, 1261–1264.

33. D'haene,B., Vandesompele,J. and Hellemans,J. (2010) Accurate and objective copy number profiling using real-time quantitative PCR. *Methods*, **50**, 262–270.

34. Criscuolo,F., Bize,P., Nasir,L., Metcalfe,N.B., Foote,C.G., Griffiths,K., Gault,E.A. and Monaghan,P. (2009) Real-time quantitative PCR assay for measurement of avian telomeres. *J. Avian Biol.*, **40**, 342–347.

35. Callicott,R.J. and Womack,J.E. (2006) Real-time PCR assay for measurement of mouse telomeres. *Comp. Med.*, **56**, 17–22.

36. Kozera,B. and Rapacz,M. (2013) Reference genes in real-time PCR. *J. Appl. Genet.*, **54**, 391–406.

37. Lin,J., Smith,D.L., Esteves,K. and Drury,S. (2019) Telomere length measurement by qPCR – Summary of critical factors and recommendations for assay design. *Psychoneuroendocrinology*, **99**, 271–278.

38. Gomes,N.M.V., Ryder,O.A., Houck,M.L., Charter,S.J., Walker,W., Forsyth,N.R., Austad,S.N., Venditt,C., Pagel,M., Shay,J.W., *et al.* (2011) Comparative biology

of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*, **10**, 761–768.

Supporting Information

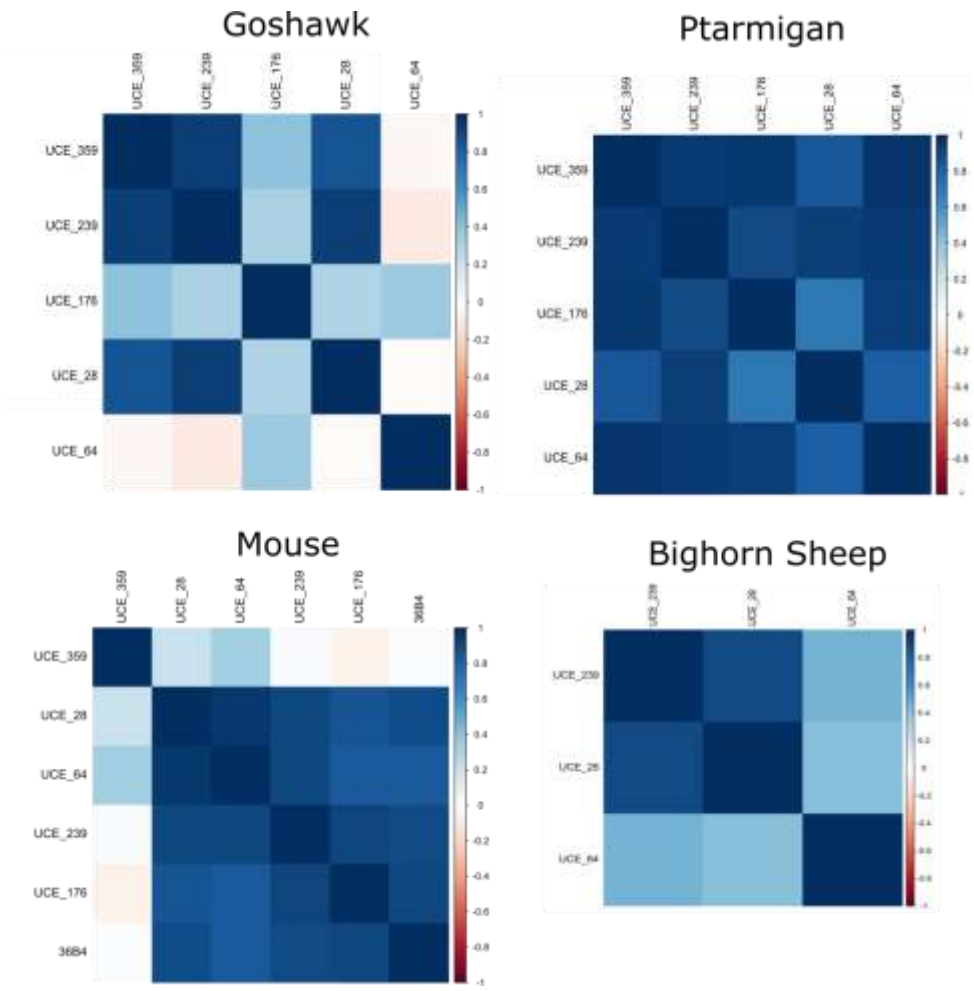


Fig. S1.1 Pearson Correlation matrixes for UCE primers in northern goshawk (Accipiter gentillis), rock ptarmigan (Lagopus muta), mouse (Mus musculus) and bighorn sheep (Ovis canadensis).

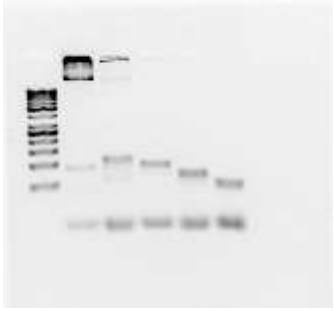


Fig. S1.2 GAPDH PCR variants in American kestrel. PCR products of multiple sizes are formed when utilizing GAPDH primers in American kestrels (Falco sparverius).

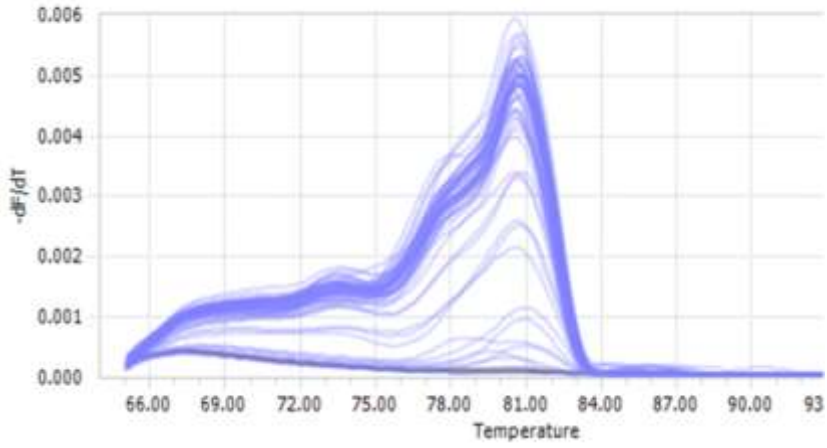


Fig. S1.3 Melt curve of frequently utilized 36B4 reference primers used in qPCR-based telomere assays in mice (Mus musculus).

CHAPTER TWO: UNIFYING COMMUNITY DETECTION ACROSS SCALES FROM GENOMES TO LANDSCAPES

Stephanie F. Hudon[†*1], Andrii Zaiats[†1], Anna Roser[1], Anand Roopsind[1], Cristina Barber[1], Brecken C. Robb[1], Britt A. Pendleton[1], Meghan J. Camp[2], Patrick E. Clark[4], Merry M. Davidson[1], Jonas Frankel-Bricker[1], Marcella Fremgen-Tarantino[1], Jennifer Sorensen Forbey[1], Eric J. Hayden[1], Lora A. Richards[3], Olivia K. Rodriguez[1], T. Trevor Caughlin[1]

[1] Boise State University, Boise, ID 83725, USA
[2] Washington State University, Pullman, WA 99164, USA
[3] University of Nevada, Reno, Reno, NV 89557, USA
[4] Northwest Watershed Research Center, USDA Agricultural Research Service, Boise, ID 83712, USA

* To whom correspondence should be addressed. Tel: +1 208 995 3765; Email: **stephaniehudon@boisestate.edu**
Correspondence may also be addressed to Tel: +1 352 234 3185; Email: **trevorcaughlin@boisestate.edu**
† These authors contributed equally to this work

In a nutshell:

- Understanding how biodiversity responds to global change requires connecting research across many spatial scales, from molecules to landscapes.
- Given modern challenges in conservation, we demonstrate how an interdisciplinary collaboration can further a holistic understanding of biodiversity, efficient conservation research and management planning.

- Using modeling to find common features within datasets, we present a framework to analyze data about landscape vegetation patterns, plant chemicals, and bacteria in the digestive tracts of sagebrush herbivores.
- We demonstrate how our interdisciplinary approach could aid conservation strategies and how models for detecting communities could provide a common language across many types of ecological data.

Abstract

Biodiversity science increasingly encompasses multiple disciplines and biological scales from molecules to landscapes. Each scale has potential to inform conservation strategies and nested interactions between scales are common. Nevertheless, biodiversity data are often analyzed separately with discipline-specific methodologies and resulting inferences may be constrained across scales. To overcome this, we present a topic modeling framework to analyze community composition in cross-disciplinary datasets, including those generated from metagenomics, metabolomics, field ecology and remote sensing. Using topic models, we demonstrate how biodiversity inference from disparate datasets can inform the conservation of interacting plants and vertebrate herbivores. We show how topic models can identify members of molecular, organismal, and landscape-level communities that explain the health and population dynamics of threatened herbivores. We conclude with a future vision for how topic modeling could be used to design cross-scale studies that promote a holistic approach to detect, monitor, and manage both threatened species and biodiversity.

Introduction

Understanding biodiversity will require crossing disciplinary boundaries to link biological organization across scales. While early efforts to quantify biodiversity focused on the organismal scale of plants and animals (Simpson, 1949), modern biodiversity research encompasses molecular scales, including metabolomic, genomic, and microbial diversity, as well as scales beyond individual organisms, including variability of biotic and abiotic features within landscapes, regions, and continents. Studying biodiversity at microscopic and macrosystem scales has led to emerging insights with broad relevance for human health (Mohajeri *et al.,* 2018), global sustainability (Bennett *et al.*, 2015) and wildlife conservation (Trevelline *et al.*, 2019). As recognition of the importance of biodiversity has increased, so have methods for analyzing biodiversity, from molecular approaches such as Next Generation sequencing for genomic data to airborne sensors that can measure large-scale landscape features. These methods are necessarily discipline-specific, thus limiting analysis of biodiversity patterns that may be nested within or interact among scales. The lack of interdisciplinary cohesion in biodiversity studies, due to different terminology and varying scales of interest, is a critical gap that limits our understanding of biological processes vital to sustaining our global ecosystem. One step toward overcoming this lack of cohesion is to identify unifying patterns in data across disciplines that can then be analyzed and discussed with a common organization and language (Mosher *et al.,* 2020).

Community organization is a unifying feature of biodiversity across scales. Metabolites, microbial taxa, plant and animal species, and spectral bands from land surface reflectance are the features that comprise communities. These are often

indicators of function that link disciplines. For example, metabolite features within plants influence the microbial features of individual herbivores and reflectance features of green plants can predict herbivore dynamic features of populations across landscapes. One challenge of detecting communities is mixed membership, including membership of both single features and single samples in multiple communities. The degree of mixed membership in communities depends on whether features arrange themselves as discrete members of disparate communities (Clements, 1916) or as fluid entities with membership in multiple communities (Gleason, 1926). Within cellular units, biomolecular processes such as mutation and differential gene expression can promote mixing of metabolic and genetic features. Within landscape units, processes such as dispersal and anthropogenic disturbances lead to mixing of organismal features (*i.e.*, species) and obscure boundaries between communities (Lortie *et al.*, 2004). Additionally, tradeoffs between sampling extent and resolution impact community detection and represent a common methodological challenge, spanning DNA sequencing to satellite remote sensing. For example, in metagenomics, the benefit of deep sequencing must be weighed against the cost of generating more reads and whether there will be adequate data generated to identify bacterial genes and taxa with low abundance but great importance. Similarly, in remote sensing, there are tradeoffs between quantifying abundance of specific plant taxa at high resolution and the spatial extent by which the broad plant classification are mapped across regions.

Altogether, mixed membership of features within sampling units and communities is common. Nevertheless, many analytical methods, such as clustering and ordination techniques, lack a probabilistic interpretation of community membership, limiting our

understanding of diverse mixtures of coupled molecular and ecological data (McCune and Grace, 2002). One solution is topic modeling of community membership, which has revolutionized multivariate analysis by enabling a single feature or sampling unit to belong to multiple communities. Latent Dirichlet Allocation (LDA) is a topic modeling approach that can identify communities of features, while allowing for mixed membership of features across communities as well as mixtures of communities within individual sampling units (Valle *et al.,* 2014). LDA was first developed in population genetics, motivated by the need to use genotypes as features that could group individuals into populations, while allowing for admixture (Pritchard *et al.,* 2000). Several years later, LDA was independently developed as a tool for text mining and broadly adopted by the machine learning community (Blei *et al.,* 2003). Since then, LDA has resulted in transformative biological insights across disciplines including annotating unknown chemicals in fermented beverages (van Der Hooft *et al.,* 2016), characterizing functional roles of gene regions (Chen *et al.,* 2010) and identifying communities of bird species in citizen science data (Valle *et al.,* 2018). Beyond single discipline applications, we contend that topic modeling has unrealized potential to unify biodiversity science across scales.

Here we demonstrate how to apply LDA across multiple scales to inform conservation of vertebrate herbivores. We focus on the sagebrush steppe ecosystem that once covered ~ 1 million km^2 of land in the western United States but is increasingly threatened by wildfires and invasive species (Requena-Mullor *et al.,* 2019). Sagebrush (*Artemisia* spp.) are the dominant plant species in these ecosystems and are critical for two sagebrush obligate species: the pygmy rabbit (*Brachylagus idahoensis*) and the

Greater sage-grouse (*Centrocercus urophasianus*, hereafter sage-grouse). Both herbivores are considered species of conservation concern across the Intermountain West. However, efforts to conserve and reintroduce populations of pygmy rabbits and sage-grouse have had mixed success due to problems that range from lack of consideration of local molecular adaptations (Oh *et al.,* 2019) to ecosystem fragmentation (Cross *et al.,* 2018).

Management of threatened species, including pygmy rabbits and sage-grouse, will benefit from a deeper and more functional understanding of the biological communities that promote or undermine individual health and population dynamics. We use four case studies co-occurring with the sagebrush steppe ecosystem to demonstrate how LDA can assess community mixtures of metabolites from leaf material of individual plants, microbial species from fecal pellets of herbivores, plant species from field plots within ecosystem patches and spectra from pixels across a landscape (Figure 2.1). Organismal sampling units, ecosystem patches, and the landscape interact in this study system. At the micro-scale, microbial features in herbivores (Kohl *et al.,* 2016) interact with metabolite concentrations in the gut after herbivores consume sagebrush (Kohl *et al.,* 2016) . At the macro-scale, features of herbivores and sagebrush are dependent on metabolite concentrations of plant taxa within habitat patches (Ulappa *et al.,* 2014, Frye et al., 2013) and those plant taxa can be detected with aerial remote sensing platforms (Olsoy *et al.,* 2020). Ultimately, the community patterns that emerge from analyzing features across scales could deepen our understanding of plant-herbivore interactions and identify molecular, organismal, and landscape targets within complex communities for management in changing landscapes.

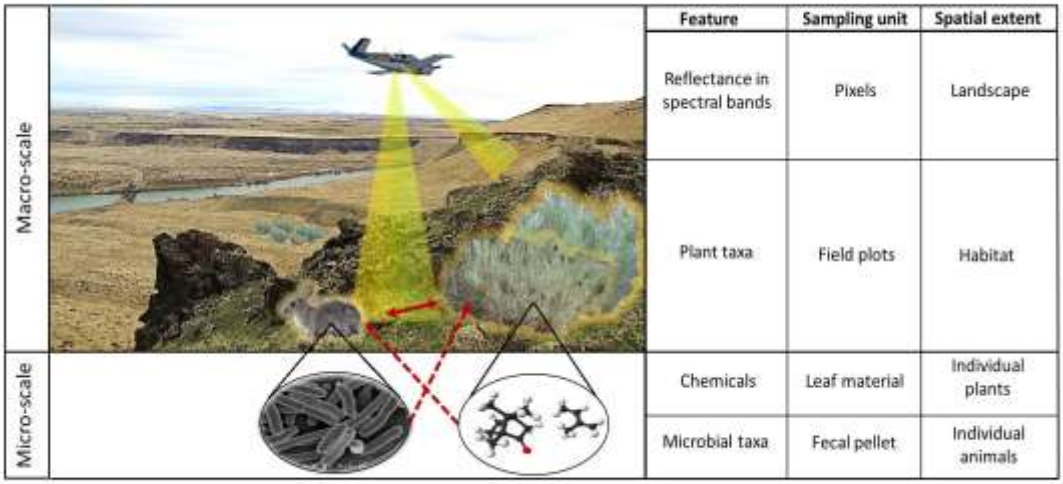


Figure 2.1 Illustration of how communities of features related to biodiversity are measured in sampling units that span micro- and macro-scales.
In the sagebrush steppe ecosystem, these communities are linked across scales. Microbial taxa in fecal pellets from individual herbivores interact with chemical features in leaf material when herbivores consume individual plants. Metabolite features in leaf material consumed by herbivores are dependent on the abundance of individual plant taxa detected within field plots. The distribution of plant taxa can be detected with spectral bands in pixels of aerial imagery obtained remotely within landscapes.

Results

Applying Latent Dirichlet Allocation analysis across data sets

We analyzed each of our datasets with LDA models in RStudio (v. 3.4.4) to determine community membership of features within sampling units. Our models applied a Bayesian framework using the Gibbs sampler from the 'Rlda' package (Albuquerque *et al.,* 2019). We applied the elbow method to identify and focus on communities with biological relevance. We used a binomial variation of LDA to detect metabolite and spectral communities (Valle *et al.,* 2018) and the multinomial parametrization for the analysis of microbial taxa and leaf area index (LAI) (Blei *et al.,* 2003; Valle *et al.,* 2014). We provide an overview of the statistical model underlying LDA and detailed methods for each case study in the supplementary material. For an overview of the statistics that underlie LDA we refer readers to Albuquerque et al. (2019) and Blei et al. (2003).

Case study 1. Reflectance of spectral bands at the landscape scale

Our first case study uses LDA to assess patterns in spectral data obtained from remotely sensed images across the landscape. Understanding impacts of global change on sagebrush ecosystems will require measurements over spatial extents much greater than that provided by ecological field plots alone. We investigated how this challenge may be overcome using aerial remote sensing to detect spatial patterns in vegetation cover. We used a binomial version of LDA to detect patterns in electromagnetic reflectance from aerial imagery of a sagebrush steppe landscape (Figure 2.2a).

Using LDA, we were able to detect ecological patterns related to changing composition of plant taxa. We identified two communities of spectral features characteristic of vegetation, including low reflectance in photosynthetically-active wavelengths (Figure 2.2a). Based on visual interpretation of concurrently collected Red-Green-Blue (RGB) imagery, the first community (Community 1) represents juniper trees (*Juniperus* spp.) while the second community is more associated with low-growing shrubs (Figure 2.2b). Conifer tree encroachment, including juniper range expansion, threatens wildlife species (Severson *et al.,* 2017). The patchiness of Community 1 suggests fine-scale variation in juniper cover during the early stages of woody encroachment (Figure 2.2b). The more uniform representation of Community 2 (Figure 2.2b, right panel) is attributable to a dominant but sparse canopy of shrubs documented in ground observations (National Ecological Observatory Network, NEON.DOM.SITE.DP1.10058.001). Our results demonstrate how high-resolution hyperspectral data can detect and map juniper encroachment in sagebrush steppe. Ultimately, patterns of remotely-sensed, spectral features could be used to monitor

ecological change in patches within and across the landscape where herbivores forage (Frye *et al.*, 2013; Ulappa *et al.*, 2014).

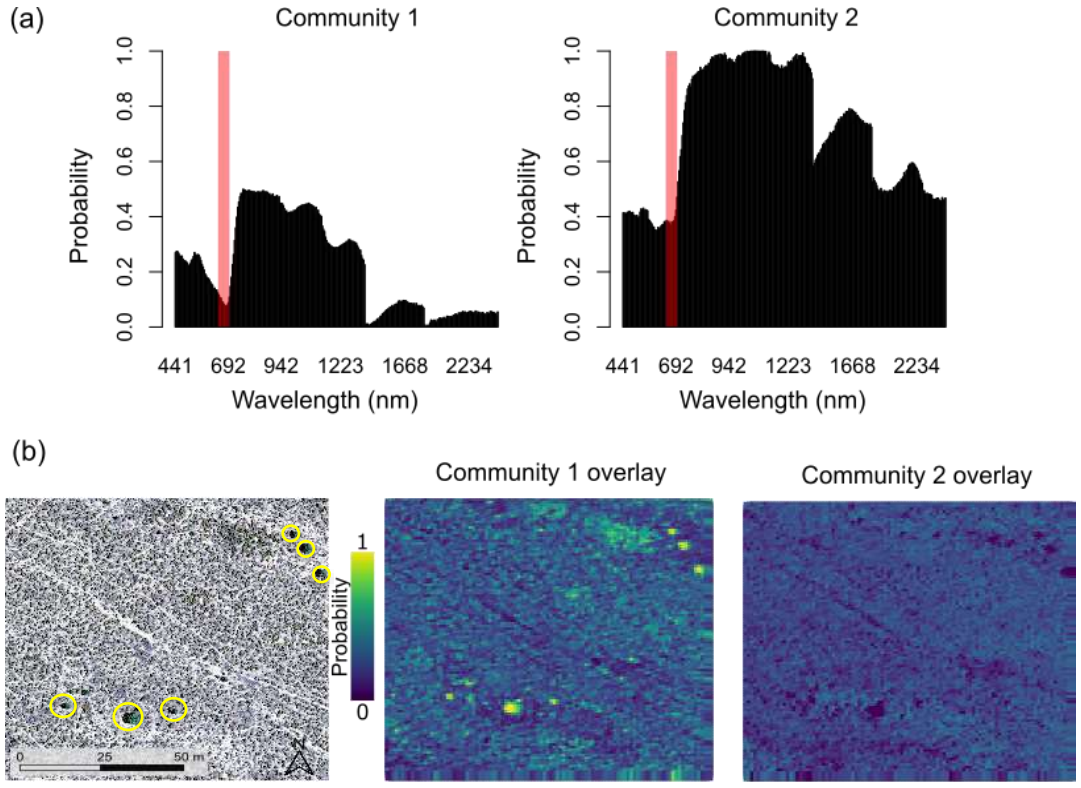


Figure 2.2 LDA applied to a subset of hyperspectral (1 m2 resolution) orthomosaic from sagebrush steppe (Onaqui, Utah, USA).
(a) The probability of each wavelength of reflected light belonging to two communities. The rapid change in reflectance between 690 nm and 750 nm (the "red edge") is representative of changes in plant photosynthetic activity. (b) Red, Green, Blue (RGB) image of the area with encroaching juniper trees circled in yellow (left image) with community 1 overlay (middle) and community 2 overlay (right) in the same area outlining a high probability that junipers belong to Community 1. Colors approaching yellow indicate higher probability of pixel membership from a particular spectral feature.

Case study 2. Plant taxa within plots at the ecosystem scale

Our second case study uses LDA to identify patterns in leaf-area-index (LAI) classified to species in field plots. Ecosystem processes are driven in part through LAI, the relative size of one leaf over a unit of ground surface and a common proxy for photosynthetic biomass (Ewert, 2004). Measurements of LAI in drylands also relate to food availability for herbivores (Olsoy *et al.,* 2015). We quantified LAI in vegetation plots within a habitat dominated by Wyoming big sagebrush (*Artemisia tridentata ssp. wyomingensis*) (Figure 2.3).

Results from applying LDA suggest that plots within this habitat type are characterized by six LAI communities (Figure 2.3). We report on the composition of three of these communities due to their ecological significance. Community 1 and 3 were dominated by the presence of Wyoming big sagebrush and Sandberg bluegrass (*Poa secunda*), respectively. Wyoming big sagebrush and Sandberg bluegrass are of particular importance because their presence indicates habitats favorable for herbivores (Beck *et al.*, 2009). Community 6 was dominated by cheatgrass (*Bromus tectorum*), an invasive annual (Figure 2.3a) that indicates degraded ecosystems less suitable for herbivores (Steenvoorden *et al.,* 2019).

LDA applied to all LAI plots indicate overall dominance by Wyoming big sagebrush and Sandberg bluegrass with generally low probability of the invasive cheatgrass community (Figure 2.3b). These results are also seen at the level of a single sampling unit ($1m^2$, Figure 2.3c). Our leaf-level analysis could be used to quantify fine-scale suitability for particular wildlife species. For example, plots monitored after fires with high probability of Wyoming big sagebrush and low probability of cheatgrass might

indicate more successful post-fire restoration (Baker, 2006), including the regeneration of suitable forage (Beck *et al.*, 2012) for wild and domestic herbivores.

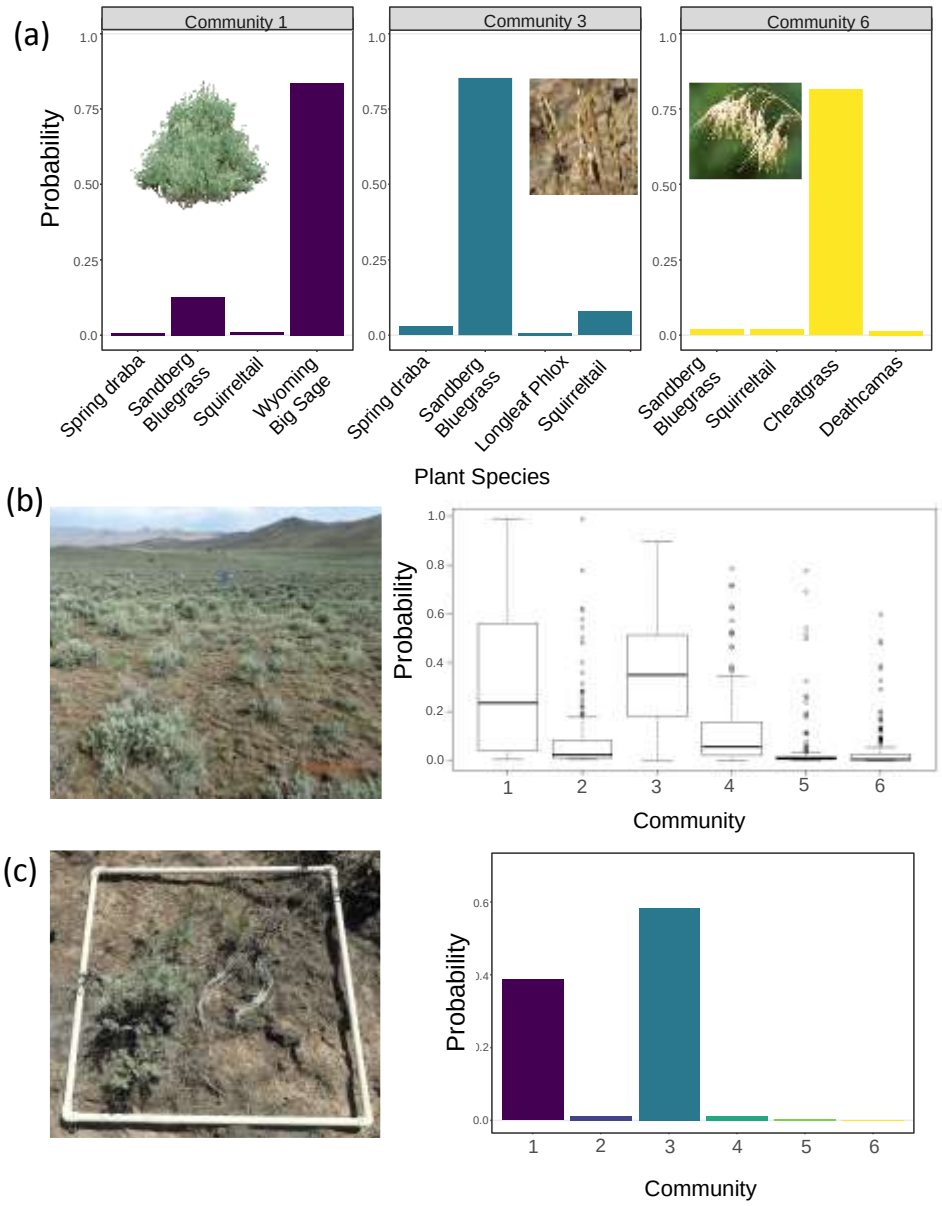


Figure 2.3 The results of LDA analysis on leaf area index (LAI) in a Wyoming big sagebrush habitat (*Artemisia tridentata ssp.wyomingensis*).
(a) The probability of plant species occurring within three communities with an image of the dominant species in inset. (b) A landscape level photo (left) and the probability of the presence of the six most common communities within the habitat sampling units (right). (c) A representative photo of a single 1m^2 field plot sampling unit (left) and the probability of the presence of each community within a single plot (right).

Case study 3. Metabolites within leaves at the plant scale

Our third case study uses LDA to identify patterns in metabolite features between two sagebrush taxa. While several herbivores rely on sagebrush as forage year-round, the volatile monoterpene features of this plant influence selection by herbivores at the species, patch and plant scale (Frye *et al.,* 2013). Although there are known concentration-dependent consequences of individual monoterpenes, the unique mixtures of metabolites in plants may better explain intake by herbivores (Nobler *et al.,* 2019). Moreover, foraging herbivores consume mixtures of metabolites, not individual metabolites. Approaches that focus on the presence or concentration of a specific metabolite likely miss important changes in the relative ratios of compounds that better determine diet selection by herbivores and predict interactions with the microbial features (*e.g.,* case study 4 below) in the gut of herbivores.

We found that LDA can detect communities of monoterpenes that have relevance to herbivore diet selection in two different sagebrush taxa (Figure 2.4). We focused on three communities that contained compounds that predict foraging by herbivores. Community 4 is characterized by a high abundance of an unknown monoterpene (Unk 21.0), while Community 3 is dominated by high abundance of a different unknown monoterpene (Unk 21.5) and Community 1 is co-dominated by yet another unknown monoterpene (Unk 20.5) and β-pinene (Figure 2.4a). There was considerable variation between sagebrush species, with highly dynamic differences among samples (Figure 2.4b). At the individual sampling unit (plant) level, three-tip sagebrush (*Artemisia tripartita*) had a high probability of Community 4, whereas Wyoming big sagebrush was dominated by Communities 1 and 3 (Figure 2.4c). Concentrations of Unk 21.0 (in

Community 4) and Unk 21.5 (in Community 3) predict diet selection by free-ranging sage-grouse (Fremgen-Tarantino *et al.* 2020) and β-pinene was avoided by captive mountain cottontails (*Sylvilagus nuttallii*) (Nobler *et al.,* 2019). Our results demonstrate how LDA can reveal communities of metabolite features that predict foraging decisions by herbivores. A potential application of LDA could be to improve post-fire restoration by reseeding with plants that have similar chemical community profiles to those of plants consumed and preferred by threatened herbivores.

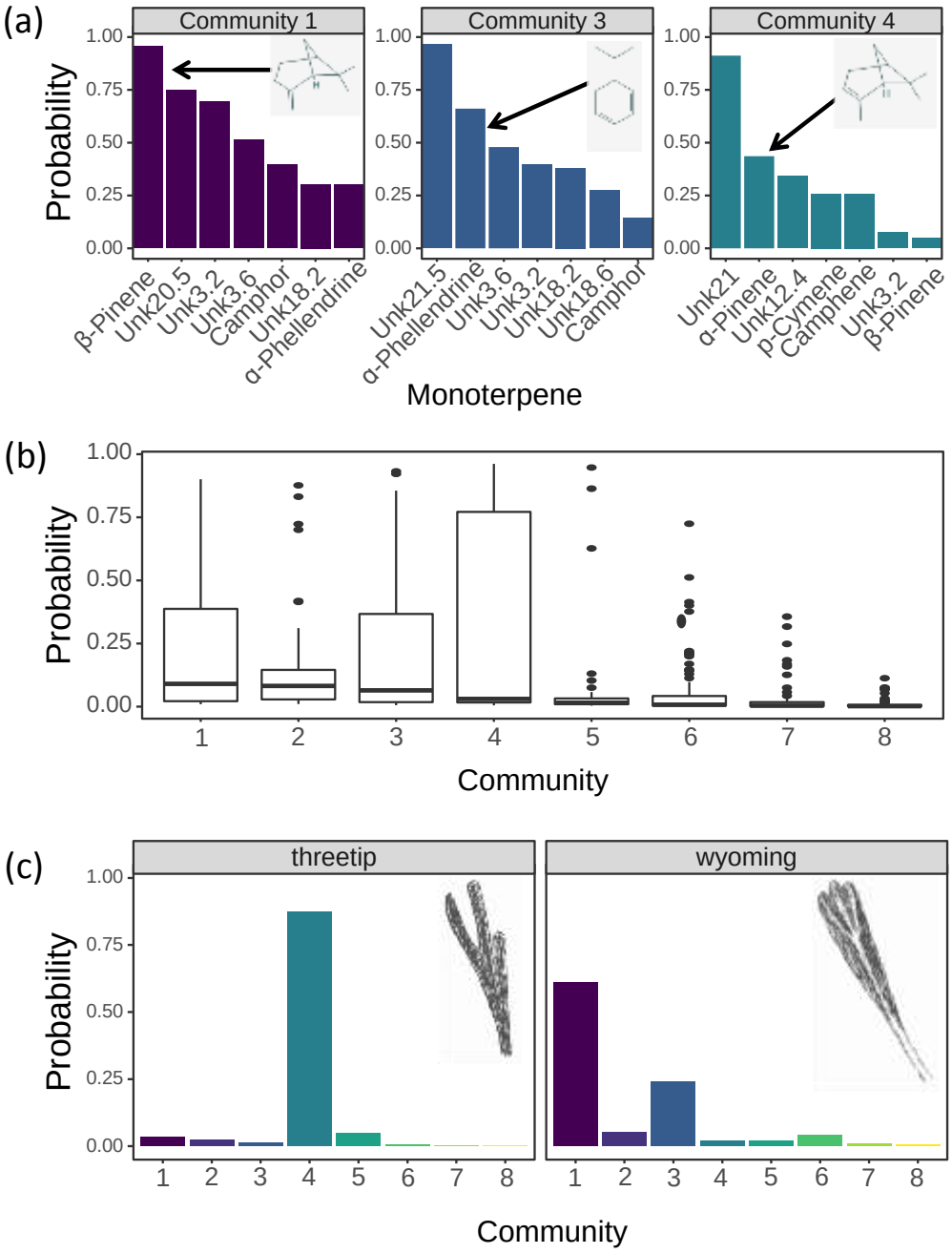


Figure 2.4 Results of LDA analysis of monoterpenes from leaves of sagebrush plants consumed by herbivores with an image of the molecular structure of the dominate known monoterpene in inset.
(a) Probability of monoterpenes occurring within three metabolite communities. (b) Probability of the eight most common metabolite communities across all sagebrush samples. (c) Probability of the eight metabolite communities occurring within an individual three-tip (*Artemisia tripartita*) and a Wyoming big sagebrush (*Artemisia tridentata ssp. wyomingensis*) sampling unit with an image of the leaf morphology of each species in inset. Illustrations by James Hudon.

Case study 4. Microbial taxa within fecal pellets at the herbivore scale

Chemical communities in herbivore forage, including plants in the wild and artificial pellets in captivity, can modify microbial species composition within animal guts (Sandifer *et al.*, 2015; Mohajeri *et al.*, 2018; Kohl et al., 2014). Our fourth case study uses LDA to identify patterns in microbial taxonomic features detected in fecal pellets of pygmy rabbits over time, as they transitioned from a natural field-based diet to captivity. Specifically, we analyzed how the taxa of the fecal microbiome from this obligate sagebrush herbivore would change as they were transitioned from a natural diet containing Wyoming big sagebrush to a captive diet, containing commercial rabbit food, over a seven-day period. Fecal samples from the rabbits on day 1 (sagebrush diet) and day 10 (captive diet) were collected and analyzed using shotgun metagenomics. We used LDA to identify communities of bacteria at the genus level (Figure 2.5). The anaerobes, C*lostridium* and *Bacteroides*, were common features of these bacterial communities (Figure 2.5a). Communities 3 and 8 show the highest probability of being found within all fecal samples (Figure 2.5b). Community 3 was dominated by *Bacteroides* and had a higher probability of being present when the rabbits were on a natural diet, whereas Community 8, which was dominated by *Clostridium* species, was more prevalent after a week of transitioning to a captive diet (Figure 2.5c). Some *Clostridium* species are associated with enteritis and increased mortality in wild and captive animals (Paul and Friend 2019) whereas other *Clostridium* species may improve animal health (Liu *et al.*, 2019). These preliminary results suggest that LDA can be used to monitor changes in bacterial communities associated with dietary shifts, and potential health, in sagebrush-dependent herbivores. Because microbial function is largely driven by communities,

rather than individual species, community-level analyses (*e.g.*, LDA) are crucial for identifying physiologically-relevant changes in herbivore metagenomes.

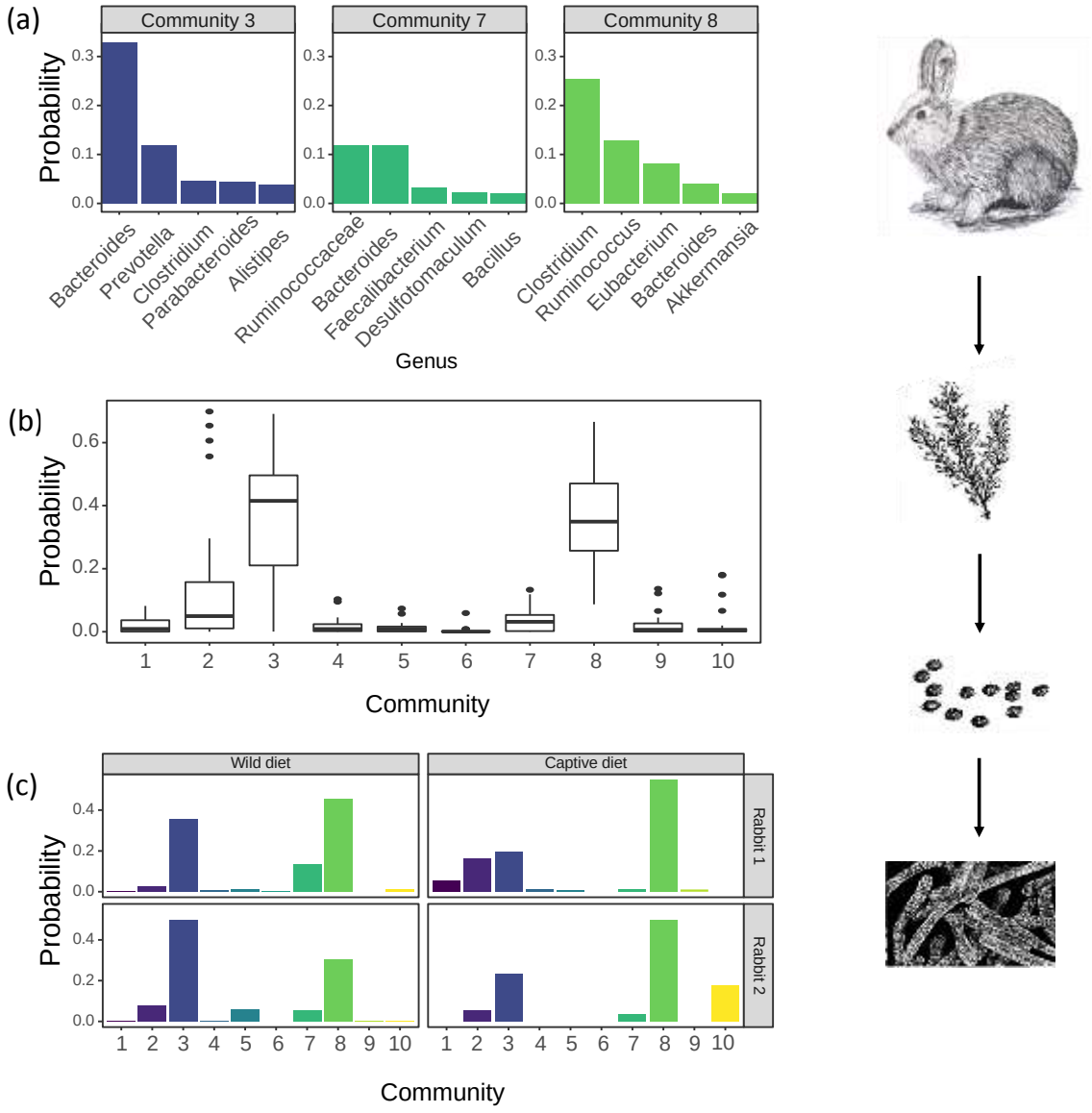


Figure 2.5 LDA analysis using genus level taxonomy counts from metagenomics of fecal samples collected from pygmy rabbits (*Brachylagus idahoensis*, shown top left). (a) Probability of microbial features within the three most prevalent communities detected in fecal samples, each dominated by different microbial taxa. (b) Probability of the ten identified microbial communities within fecal samples from the pygmy rabbit (n=22). (c) Probability of the ten microbial communities in fecal samples from the pygmy rabbit sampling units consuming a natural diet (primarily Wyoming big sagebrush (*Artemisia tridentata ssp. wyomingensis*) and after ten days on an artificial pellet diet in captivity. Illustrations by James Hudon.

Discussion

As biodiversity science grows to encompass scales from molecular to continental, the need for integrative approaches has increased as well. We have demonstrated the potential for community structure to unite patterns of biodiversity across disciplines. We found that Latent Dirichlet Allocation, a topic model that can represent mixed membership of features, enabled us to quantify biological communities across molecular, organismal and landscape scales. Our results have potential relevance for conservation of threatened herbivores in the imperiled sagebrush steppe ecosystem, including the pygmy rabbit. At the landscape scale, LDA detected juniper encroachment, a driver of habitat degradation in sagebrush steppe, from aerial remote sensing data. At the plant scale, LDA enabled discrimination between plant species assemblages, with relevance for habitat structure, including the availability of quality forage for herbivores and the presence of invasive species. At the molecular scale, LDA identified mixtures of secondary metabolites that can differentiate plant species and predict diet selection by herbivores. At the microbial scale, LDA quantified shifts in bacterial communities in response to diet transitions of herbivores, that are predictive of disease and survival. Across all of these scales, LDA enabled our interdisciplinary team to develop a holistic view of plant-herbivore ecology and understand the relevance and community-level connections of research across disciplines. Common models for disparate datasets, including LDA, will enable collaborative studies that can inform cross-scale strategies for conservation.

One realization that emerged from co-analyzing our data is the overarching importance of herbivore gut microbiomes for uniting scales. We argue that studying gut microbiomes has great potential to develop a more complete understanding of herbivore

ecology, particularly if multiple scales are incorporated into analyses. Herbivores, such as the pygmy rabbits in our study, make foraging decisions at individual metabolite, leaf, plant, and landscape scales (Ulappa et al., 2014; Nobler et al., 2019). In turn, foraging herbivores can influence patterns of habitat structure and plant species composition (Eldridge et al., 2016). Over long periods of time, we would expect that gut microbes mediate feedback loops between plants and herbivores, with ecological and evolutionary implications (Ley et al., 2008; Kohl and Dearing, 2016). In a practical sense, the gut microbiome links these disparate scales and represents the net sum of forage availability and quality across landscapes (Figure 2.2), habitats (Figure 2.3) and within plants (Figure 2.4). Considering that herbivore foraging has wide-ranging consequences for above-ground (Frye et al., 2013; Ulappa et al., 2014; Fremgen-Tarantino et al., 2020) and below-ground (Chomel et al., 2016) ecological processes, a more holistic understanding of co-occurring plant, metabolite and microbial communities in the guts of herbivores will have broad relevance. While topic models, such as LDA, present an opportunity to describe microbial community structure (Chen et al., 2012), development of analytical tools that integrate hierarchies of scale and complex network structure will further enable researchers to uncover how microbial communities might interact with communities at other scales, from the molecular to the landscape scale.

We envision designing future studies where data are collected from multiple biological units at the same time and place with a focus around fecal collections. Data collection focused around herbivore fecal pellets could involve collecting feces from herbivores for metagenomic and metabolite analysis while simultaneously collecting leaf tissue from plants browsed by herbivores for metabolite content (parent and

detoxification products), and mapping the GPS location where pellets and plant samples are collected. Subsequently, research teams could assess how communities of microbes, plant-derived metabolites, and plant communities detected in feces are influenced by variation in plant species availability at plot and habitat scales to address questions of resource selection. Remote sensing data, such as hyperspectral aerial images, could then be applied to detect temporal and spatial variation in the composition of plant species and foliar chemistry across the landscape and to relate habitat features to microbial, metabolite, and plant communities. This type of data collection will require extensive interdisciplinary coordination, but will lead to a more connected understanding of coupled biodiversity among scales. Long-term ecological research sites, such as the NEON network, provide a valuable starting point for this type of study where collection and analysis of herbivore metagenomics and metabolites from plants could add substantial value to existing data on plant diversity and soil microbial communities. In the context of planning field studies, LDA could be applied as a generative model to simulate data and estimate appropriate sample sizes for statistical estimation.

Biodiversity data commonly includes features and communities that change over time and space, in response to experimental treatments, environmental covariates, or endogenous dynamics. A next step for the development of LDA will be to incorporate predictor variables into topic models for community detection, with a goal of statistical inference. Dynamic topic models are currently used in text mining to account for changing community membership (Blei and Lafferty, 2006), while dynamic mixture models enable realized proportions of communities to change over time (Wei *et al.*, 2007). In ecology, LDA, in conjunction with breakpoint models, has recently been

applied to interpret temporal shifts in rodent communities (Christensen *et al.*, 2018). As a statistical approach, conceptually related to regression models for proportional data (Douma and Weedon, 2019), LDA could be extended to enable robust statistical inference on community patterns.

Altogether, coordinated studies of community structure across scales will enable researchers to address fundamental questions in ecology and evolution. One such question relates to the long-standing debate over whether biological features, from genes to species assemblages, are organized by neutral processes or deterministic ecological and evolutionary forces (Kreitman, 1996; Lynch, 2007; Lowe and McPeek, 2014). For example, convergent communities of microbes in the soil and guts of herbivores exposed to similar plant metabolite communities across broad biogeographical scales would provide powerful evidence for the role of non-neutral processes. Alternately, random associations between overlain communities could suggest neutral theory as an overarching explanation for observed assemblages. Common models for community structure will provide detailed and cohesive insight into the complex interactions among plants, animals and microbes co-occurring across landscapes. Altogether, we anticipate that interdisciplinary collaboration, facilitated by the common modeling language of LDA, will have payoffs for biodiversity studies that must address complex problems that cross scales.

Acknowledgements

We would like to thank Karenina Sanders, Greg Humphrey, Jon Sanders, Rob Knight and The Center for Microbiome Innovation at UC San Diego for metagenomic data acquisition. We also thank J. Connelly, D. D. Musil, L. Cross, L.A. Shipley and the

GUTT seminar working group in addition to financial support from a NASA grant 80NSCCC17K0738, a Idaho State Board of Education grant IGEM19-002, a Semiconductor Research Corporation grant SRC 2018-SB-2842, Pittman-Robertson 683 funds from the Idaho Department of Fish and Game, Sigma Xi Grants-In-Aid, Bureau of Land Management grant #L09AC16253 and National Science Foundation grants IOS-1258217, DEB-1146194, DEB-1146368, OIA-1826801, OIA-1757324, OIA-1738865, ECCS-1807809 and the USDA Agricultural Research Service.

References

Albuquerque PH, Valle DR do, and Li D. 2019. Bayesian LDA for mixed-membership clustering analysis: The Rlda package. *Knowl-Based Syst* **163**: 988–995.

Baker WL. 2006. Fire and Restoration of Sagebrush Ecosystems. *Wildl Soc Bull* **34**: 177–85.

Beck JL, Connelly JW, and Reese KP. 2009. Recovery of Greater Sage-Grouse Habitat Features in Wyoming Big Sagebrush following Prescribed Fire. *Restor Ecol* **17**: 393–403.

Beck JL, Connelly JW, and Wambolt CL. 2012. Consequences of Treating Wyoming Big Sagebrush to Enhance Wildlife Habitats. *Rangel Ecol Manag* **65**: 444–55.

Bennett EM, Cramer W, Begossi A, *et al.* 2015. Linking biodiversity, ecosystem services, and human well-being: three challenges for designing research for sustainability. *Curr Opin Environ Sustain* **14**: 76–85.

Blei DM and Lafferty JD. 2006. Dynamic topic models. In: Proceedings of the 23rd international conference on Machine learning. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery.

Blei DM, Ng AY, and Jordan MI. 2003. Latent Dirichlet Allocation. *J Mach Learn Res* **3**: 993–1022.

Chen X, Hu X, Lim TY, *et al.* 2012. Exploiting the functional and taxonomic structure of genomic data by probabilistic topic modeling. *IEEE/ACM Trans Comput Biol Bioinform* **9**: 980–991.

Chen X, Hu X, Shen X, and Rosen G. 2010. Probabilistic topic modeling for genomic data interpretation. In: 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE.

Chomel M, Guittonny-Larchevêque M, Fernandez C, *et al.* 2016. Plant secondary metabolites: a key driver of litter decomposition and soil nutrient cycling. *J Ecol* **104**: 1527–41.

Christensen EM, Harris DJ, and Ernest SKM. 2018. Long-term community change through multiple rapid transitions in a desert rodent community. *Ecology* **99**: 1523–9.

Clements FE. 1916. Plant succession; an analysis of the development of vegetation,. Washington,: Carnegie Institution of Washington,.

Cross TB, Schwartz MK, Naugle DE, *et al.* 2018. The genetic network of greater sage-grouse: Range-wide identification of keystone hubs of connectivity. *Ecol Evol 2018 1-19*: 1–19.

Der Hooft JJJ van, Wandy J, Barrett MP, *et al.* 2016. Topic modeling for untargeted substructure exploration in metabolomics. *Proc Natl Acad Sci* **113**: 13738–13743.

Douma JC and Weedon JT. 2019. Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods Ecol Evol* **10**: 1412–30.

Eldridge DJ, Poore AGB, Ruiz-Colmenero M, *et al.* 2016. Ecosystem structure, function, and composition in rangelands are negatively affected by livestock grazing. *Ecol Appl* **26**: 1273–83.

Ewert F. 2004. Modelling Plant Responses to Elevated CO2: How Important is Leaf Area Index? *Ann Bot* **93**: 619–27.

Fremgen-Tarantino MR, Peña JJ, Connelly JW, and Forbey JS. 2020. Winter foraging ecology of Greater Sage-Grouse in a post-fire landscape. *J Arid Environ* **178**: 104154.

Frye GG, Connelly JW, Musil DD, and Forbey JS. 2013. Phytochemistry predicts habitat selection by an avian herbivore at multiple spatial scales. *Ecology* **94**: 308–14.

Gleason HA. 1926. The Individualistic Concept of the Plant Association. *Bull Torrey Bot Club* **53**: 7–26.

Kohl KD, Connelly JW, Dearing MD, and Forbey JS. 2016. Microbial detoxification in the gut of a specialist avian herbivore, the Greater Sage-Grouse (A Endo, Ed). *FEMS Microbiol Lett* **363**: fnw144.

Kohl KD and Dearing MD. 2016. The Woodrat Gut Microbiota as an Experimental System for Understanding Microbial Metabolism of Dietary Toxins. *Front Microbiol* **7**.

Kreitman M. 1996. The neutral theory is dead. Long live the neutral theory. *Bioessays* **18**: 678–683.

Ley RE, Hamady M, Lozupone C, *et al.* 2008. Evolution of Mammals and Their Gut Microbes. *Science* **320**: 1647–51.

Liu L, Zeng D, Yang M, *et al.* 2019. Probiotic Clostridium butyricum Improves the Growth Performance, Immune Function, and Gut Microbiota of Weaning Rex Rabbits. *Probiotics Antimicrob Proteins* **11**: 1278–92.

Lortie CJ, Brooker RW, Choler P, *et al.* 2004. Rethinking plant community theory. *Oikos* **107**: 433–8.

Lowe WH and McPeek MA. 2014. Is dispersal neutral? *Trends Ecol Evol* **29**: 444–450.

Lynch M. 2007. The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* **8**: 803–813.

McCune BP and Grace J. 2002. Analysis of Ecological Communities.

Mohajeri MH, Brummer RJM, Rastall RA, *et al.* 2018. The role of the microbiome for human health: from basic science to clinical applications. *Eur J Nutr* **57**: 1–14.

Mosher BA, Bernard RF, Lorch JM, *et al.* 2020. Successful molecular detection studies require clear communication among diverse research partners. *Front Ecol Environ* **18**: 43–51.

Nobler JD, Camp MJ, Crowell MM, *et al.* 2019. Preferences of Specialist and Generalist Mammalian Herbivores for Mixtures Versus Individual Plant Secondary Metabolites. *J Chem Ecol* **45**: 74–85.

Oh KP, Aldridge CL, Forbey JS, *et al.* 2019. Conservation Genomics in the Sagebrush Sea: Population Divergence, Demographic History, and Local Adaptation in Sage-Grouse (Centrocercus spp.). *Genome Biol Evol* **11**: 2023–34.

Olsoy PJ, Forbey JS, Rachlow JL, *et al.* 2015. Fearscapes: Mapping Functional Properties of Cover for Prey with Terrestrial LiDAR. *BioScience* **65**: 74–80.

Olsoy PJ, Forbey JS, Shipley LA, *et al.* 2020. Mapping foodscapes and sagebrush morphotypes with unmanned aerial systems for multiple herbivores. *Landsc Ecol*.

Paul GC and Friend DG. 2019. Clostridial Enterotoxemia and Coccidiosis in Weanling Cottontail Rabbits (Sylvilagus audubonii, Sylvilagus floridanus, Sylvilagus nuttallii) from Colorado, USA. *J Wildl Dis* **55**: 189–95.

Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–59.

Requena-Mullor JM, Maguire KC, Shinneman DJ, and Caughlin TT. 2019. Integrating anthropogenic factors into regional-scale species distribution models-A novel application in the imperiled sagebrush biome. *Glob Change Biol*.

Sandifer PA, Sutton-Grier AE, and Ward BP. 2015. Exploring connections among nature, biodiversity, ecosystem services, and human health and well-being: Opportunities to enhance health and biodiversity conservation. *Ecosyst Serv* **12**: 1–15.

Severson JP, Hagen CA, Maestas JD, *et al.* 2017. Effects of conifer expansion on greater sage-grouse nesting habitat selection. *J Wildl Manag* **81**: 86–95.

Simpson EH. 1949. Measurement of Diversity. *Nature* **163**: 688–688.

Steenvoorden J, Meddens AJH, Martinez AJ, *et al.* 2019. The potential importance of unburned islands as refugia for the persistence of wildlife species in fire-prone ecosystems. *Ecol Evol* **9**: 8800–12.

Trevelline BK, Fontaine SS, Hartup BK, and Kohl KD. 2019. Conservation biology needs a microbial renaissance: a call for the consideration of host-associated microbiota in wildlife management practices. *Proc R Soc B Biol Sci* **286**: 20182448.

Ulappa AC, Kelsey RG, Frye GG, *et al.* 2014. Plant protein and secondary metabolites influence diet selection in a mammalian specialist herbivore. *J Mammal* **95**: 834–42.

Valle D, Albuquerque P, Zhao Q, *et al.* 2018. Extending the Latent Dirichlet Allocation model to presence/absence data: A case study on North American breeding birds and biogeographical shifts expected from climate change. *Glob Change Biol* **24**: 5560–5572.

Valle D, Baiser B, Woodall CW, and Chazdon R. 2014. Decomposing biodiversity data using the Latent Dirichlet Allocation model, a probabilistic multivariate statistical method. *Ecol Lett* **17**: 1591–1601.

Wei X, Sun J, and Wang X. 2007. Dynamic Mixture Models for Multiple Time-Series. In: Ijcai.

Supplementary Information

Modeling Overview:

The overarching goal of our Latent Dirichlet Allocation model was to assess community membership in features across our datasets, including spectral reflectance, plant species, metabolites, bacterial taxa, and functional genes. We applied two versions of the LDA model to our case studies: a multinomial parametrization where the abundance of features in each sample unit was proportional (i.e., the abundance of features in each sample unit summed to one, $\sum_{n=1}^{n} Y_m = 1$); and a binomial parametrization where this constraint was relaxed. For more details regarding the model description and application of LDA we refer readers to recent work by Sankaran and Holmes (2017) and Albequerque *et al.* (2019b). Briefly, we organized our datasets into abundance matrices where sampling units (m) were represented in rows and features (n) in columns. The multinomial version of LDA implies the following generative model:

$$Y_{m,n} \sim \text{Multinomial}(\phi_{z[m,n]});$$

$$z_{m,n} \sim \text{Multinomial}(\theta_m);$$

$$\phi_k \sim \text{Dirichlet}(\beta);$$

$$\theta_m = V_m \prod_{k=1}^{k-1}(1-V_m), \text{where } V_m \sim \text{Beta}(1,\gamma).$$

where ($Y_{m,n}$) represents the observed abundance of (n) features in (m) sample units. Each entry in the data matrix is assigned to a community estimated as a latent variable $z_{m,n}$ that depends on the distribution of features across a sample unit (m), θ_m. Parameter θ_m is modeled with a stick-breaking prior that ensures the minimum number of latent states $z \in [1,K]$. In turn, the probability of a feature belonging to a community (n) is modeled as parameter ϕ_k. Finally, the observed abundance is modeled according to the feature

distribution ϕ_k for community $z_{m,n}$. We used uninformed hyperparameters β and γ to initiate the model, which assumed a uniform distribution of communities and features across sample units and communities, respectively.

We applied the binomial parametrization of LDA to hyperspectral and metabolome datasets. Instead of a multinomial data generation process and a Dirichlet prior, the model assumes that observations are drawn from a binomial distribution:

$$Y_{m,n} \sim \text{Binomial}(\zeta_n, \phi_{z[m,n]});$$

$$z_{m,n} \sim \text{Multinomial}(\theta_m);$$

$$\phi_{k,n} \sim \text{Beta}(\beta_0, \beta_1);$$

$$\theta_m = V_m \prod_{k=1}^{k-1}(1-V_m), \text{where } V_m \sim \text{Beta}(1, \gamma).$$

The model closely follows the data generation process described above with a few changes. Specifically, each entry in parameter matrix ϕ is given a flat prior drawn from a Beta distribution, and ζ_n is the number of trials in the Binomial distribution, indicating the observed maximum feature abundance across the sample units.

We fit the LDA models to each of our datasets using the 'Rlda' package in R (R Core Team 2018), which relies on Gibbs sampling for parameter estimation. Following the framework outlined in Albuquerque *et al*. (2019a), we assessed model convergence using a trace-plot of log-likelihood in MCMC sampling and ensured that log-likelihood reached a plateau that indicated convergence. We discard the burn-in iterations and present parameter estimates as posterior means. Specifically, for each case study we present θ and ϕ parameter matrices corresponding to the membership of communities in sample units and features in communities, respectively. We set the number of communities as an input parameter, while the truncated stick-breaking prior ensured that

elements are allocated to the minimum number of communities. Therefore, the superfluous communities had probabilities of occurrence approaching zero and we used the "elbow method" to identify a set of biologically significant communities with non-zero probability of occurrence. This method is commonly applied in machine learning analyses and proposes use of a cut-off value to identify the optimal number of clusters or objects in a dataset (Kodinariya and Makwana, 2013; Brieuc *et al.,* 2018). All datasets presented in the study are available from Zenodo Digital Repository. Additionally, we provide fully annotated R scripts that can be used to reproduce the models as well as the results and figures presented in the main text.

Spectral Features:

We used publicly available data collected and maintained under National Ecological Observatory Network (https://www.neonscience.org) program (NEON 2019a). The data analyzed in this study represents a 100 m^2 airborne image mosaic of the Onaqui field site dominated by big sagebrush (*Artemisia tridentata*) with high spatial (1 m^2) and spectral (5 nm) resolution. The instrument used in the survey was a pushbroom spectrometer that includes the electromagnetic range from 381 nm to 2500 nm in 426 bands. For computational efficiency and to avoid noise, we removed spectral bands that are sensitive to atmospheric moisture and carry little ecologically relevant information, resulting in 306 features in the dataset (the list of bands included in the analysis is available with the dataset online). We clipped the hyperspectral tile (extent: 373751, 373850, 4448201, 4448300) in R using the 'raster' package (Hijmans 2019), and trained the model based on a random subset of 200 pixels from the clipped study area. We then forecasted the probabilities of communities over the study area, based on the spectral

reflectance of each pixel. When presenting the model output, we focused on two communities of bands with an absorbance in the red-edge of the spectrum (680-750 nm) and compared our inference with ground vegetation surveys (NEON 2019b). Changes in the red-edge electromagnetic reflectance are sensitive to changes in chlorophyll content (Filella and Penuelas, 1994; Schuster *et al.*, 2012), which is relevant to the biological context of this study.

Leaf Area Index features:

Our study site for this case study is the Reynolds Creek Experimental Watershed (RCEW). RCEW is in Owyhee County in southwestern Idaho, USA. The 239km^2 watershed has served as a natural laboratory to study semi-arid rangeland hydrology since 1960 (W. Slaughter *et al.*, 2001). We used data collected during the summer of 2018 at a site located within the watershed characterized by a dominance of *Artemisia tridentata ssp. wyomingensis*. At this site, there are five one-hectare monitoring sites, and each site has 30 plots from which we collected leaf-area index (LAI) measurements. The LAI plots are 1m^2 and randomly dispersed throughout the hectare. Data collected at the LAI plots were recorded using a metal sampling pin, that was lowered through 20 notches along five transects (100 notches per plot). Every contact between the sampling pin and vegetation was recorded to species, resulting in relative abundance of each species in a plot. In total, this field sampling resulted in 150 LAI plots.

Metabolic features:

Leaves collected in the field from two sagebrush species, including *A. tridentata spp. wyomingensis* and *A. tripartita*, were ground into fine, homogenized powder in liquid nitrogen. Monoterpenes from ground leaf tissue from each plant (100 mg wet

weight) were separated and quantified using headspace gas chromatography. All samples were analyzed using an Agilent 6890 Network Gas Chromatograph (Agilent Technologies, Inc., Santa Clara, CA, USA) ,coupled with a Hewlett-Packard HP 7694 Geadspace AutoSampler (Model: G1289A, Hewlett-Packard, Palo Alto, CA, USA). For more details on the methodology, including the headspace and autosampler sequence parameters, please refer to Fremgen 2015 (Fremgen 2015). Compounds were identified using co-chromatography with a mixture of monoterpene standards. As not all compounds were able to be identified using the co-chromatography, unidentified compound names are based on retention times. Retention times and peak areas (area under the curve, AUC as the metric of concentration) were calculated with HP ChemStation version B.01.00.

Microbial features:

Pygmy rabbits were captured in the field while feeding predominantly on a natural sagebrush diet (*A. tridentata ssp. wyomingensis*) in Blaine County, Idaho (Idaho Department of Fish and Game collection permits 100310). The rabbits were then transported within 48 hours and housed indoors at the Small Mammal Research Facility at Washington State University (Boise State University Institutional Animal Care and Use Committee Protocol # 006-AC12–009, Washington State University Institutional Animal Care and Use Committee Protocol # 04513–001) for the duration of the study. They were given fresh water and pelleted commercial rabbit chow (Purina Professional Rabbit Chow, Purina Mills LLC, St. Louise, MO), void of sagebrush. Fecal pellets were collected on day one (day of capture on natural diet) and day ten (captive diet). Metagenomics of fecal samples were processed by the Knight lab at the University of

California, San Diego, using a magnetic bead based KingFisher Flex Purification System (Marotz *et al.* 2017). Shotgun metagenomic sequencing was performed as 150-base-pair paired-end reads using an Illumina HiSeq 2500. Samples were then sent through the MG-RAST pipeline and were analyzed for taxonomical hits, using the NCBI RefSeq database to the genus level (www.ncbi.nlm.nih.gov/refseq/) (e-value:5, %-ident: 60, length: 15, min.abundance: 1).

References

Albuquerque PH, Valle DR do, and Li D. 2019a. Bayesian LDA for mixed-membership clustering analysis: The Rlda package. *Knowl-Based Syst* **163**: 988–995.

Albuquerque PHM, Valle DR do, and Li D. 2019b. Bayesian LDA for mixed-membership clustering analysis: The Rlda package. *Knowl-Based Syst* **163**: 988–95.

Brieuc MSO, Waters CD, Drinan DP, and Naish KA. 2018. A practical introduction to Random Forest for genetic association studies in ecology and evolution. *Mol Ecol Resour* **18**: 755–66.

Filella I and Penuelas J. 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *Int J Remote Sens* **15**: 1459–70.

Fremgen M. 2015. Plant Toxins Influence Diet Selection and Intestinal Parasites in a Specialist Herbivore. *Boise State Univ Theses Diss*.

Hijmans RJ. 2019. raster: Geographic Data Analysis and Modeling.

Kodinariya TM and Makwana PR. 2013. Review on determining number of Cluster in K-Means Clustering. *Int J* **1**.

Marotz C, Amir A, Humphrey G, *et al.* 2017. DNA extraction for streamlined metagenomics of diverse environmental samples. *BioTechniques* **62**: 290–3.

National Ecological Observatory Network. 2019a. Data Product DP3.30006.001, Spectrometer orthorectified surface directional reflectance - mosaic. Provisional data downloaded from http://data.neonscience.org February 2019.

National Ecological Observatory Network. 2019b. Data Product D15.ONAQ.DP1.10058.001, Plant presence and percent cover. Provisional data downloaded from http://data.neonscience.org February 2019.

R Core Team. 2018. R: A language and environment for statistical computing; 2015.

Sankaran K and Holmes SP. 2017. Latent variable modeling for the microbiome. *ArXiv Prepr ArXiv170604969*.

Schuster C, Förster M, and Kleinschmit B. 2012. Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *Int J Remote Sens* **33**: 5583–99.

W. Slaughter C, Marks D, Flerchinger G, *et al.* 2001. Thirty-Five Years of Research Data Collection at the Reynolds Creek Experimental Watershed, Idaho, United States. *Water Resour Res* **37**.

CONCLUSION

Overall, the work presented here represents a significant contribution to our ability to monitor non-model vertebrates and their ecosystems. To influence policy changes in wildlife and resource management practices, we need to provide measurable evidence of how humans are affecting animal health and the ecosystems needed for their survival. The universal assay for measuring telomere lengths will provide a new tool for assessing how human activity and climate events affect the fitness of vertebrates. For example, it has the potential to assess whether blood parasites effect the fitness of the northern goshawk and how climate change effects the fitness of American kestrels (Appendix A).

This dissertation also provides a framework for how machine learning and topic modeling can yield new insights on community structure in big data sets and unite disparate scientific disciplines to better understand biodiversity. It demonstrates how studies can be designed to span scales from the molecular to the landscape. Understanding how the microbiome of animals is tied to leaf chemistry and how plants are disbursed throughout the landscape will help us to understand the components needed to restore habitats and preserve ecosystems for threatened or endangered species. For example, I found that when a sagebrush herbivore is transitioned off its natural diet their microbiome is dominated by a different community of bacteria then when it is foraging on sagebrush. This could have implications for reintroduction of this threatened herbivore into habitats with suitable foraging material. Overall, this work represents a

significant contribution to the fields of molecular ecology and conservation biology, both in terms of new tools in monitoring animal fitness and new analytical tools to design and understand interdisciplinary studies on biodiversity.

APPENDIX A

Additional Contributions

Manuscripts in Progress

Physiological effects of hematophagous ectoparasites on golden eagle nestlings

Benjamin M. Dudek, Michael T. Henderson, Stephanie F. Hudon, Eric J. Hayden, Julie A. Heath

Preparing for submission to *Conservation Physiology* also found in thesis work: Dudek, Benjamin. (2017). The role of disease and ectoparasites in the ecology of nestling golden eagles.

This work looks at the effects of a hematophagous ectoparasite, the Mexican chicken bug (*Haematosiphon inodorus*), on golden eagle (*Aquila chrysaetos*) nestlings. Hematophagous ectoparasites can have direct effects on young birds by depleting blood volume and reducing energetic resources available for growth and development. Less is known about the effects of ectoparasitism on stress physiology (i.e., glucocorticoid hormones) or changes in behavior. Mexican chicken bugs are blood-sucking ectoparasites that live in bird nests and feed on developing nestlings. Over the past 50 years, the range of *H. inodorus* has expanded, suggesting that new hosts or populations may be vulnerable. We studied the physiological effects of *H. inodorus* on golden eagle nestlings in southwestern Idaho to better understand the physiological and behavioral effects of ectoparasitism. We estimated the level of *H. inodorus* infestation at each nest and measured nestling mass, hematocrit, corticosterone concentrations, telomere lengths, and mortality. At nests with the highest levels of infestation, nestlings had significantly lower mass and hematocrit (Figures A1.1 and A1.2). In addition, heavily parasitized nestlings had corticosterone concentrations twice as high (42.96 ng/mL) as non-parasitized

nestlings (20.2 ng/mL) (Figure A1.3). Finally, in nests with higher infestation levels, eagle nestlings were twenty times more likely to die, often because they left the nest before they could fly (Figure A1.4) and heavily parasitized females had significantly shorter telomeres than non-parasitized females (Figure A1.5). For eagles that survived infestation, chronic elevation of glucocorticoids or shortened telomeres may adversely affect cognitive function or survival in this otherwise long-lived species. These results suggest *H. inodorus* may limit local golden eagle populations by decreasing productivity up to 23%. Emerging threats from ectoparasites should be an important management consideration for raptor species facing range-wide population declines like golden eagles.

My contribution to this study was to measure the relative telomere lengths using the methods described in Chapter 1 of this dissertation. This was done on 50 blood samples from eagle nestlings. Preliminary results suggest that there is a sex-specific affect, with female nestlings exhibiting increased telomere shortening due to high infestation rates of *H. inodorus* infection (Figure A1.5).

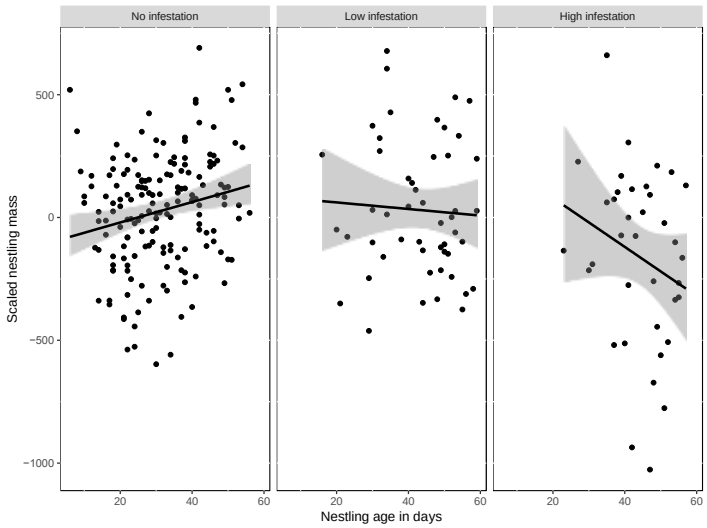


Figure A1.1 Observed golden eagle nestling mass (black circles, scaled to remove the effect of nestling age and sex), predicted mass (dark line), and associated 95% confidence intervals (solid gray area) measured from nestlings experiencing different levels of *H. inodorus* infestation in nests in southwestern Idaho, USA in 2015 and 2016.

Nestlings that experienced high levels of infestation had lower mass than nestlings in nests with no or low levels of infestation ($\chi^2 = 23.86$, $p < 0.01$).

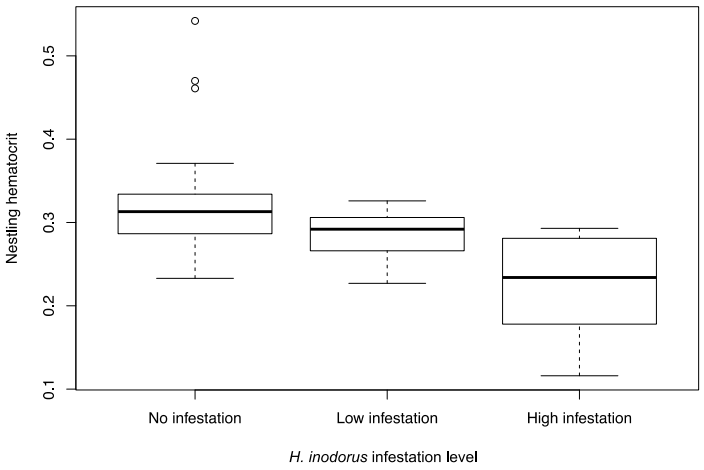


Figure A1.2 Hematocrit measured from golden eagle nestlings experiencing different levels of *H. inodorus* infestation in nests in southwestern Idaho, USA in 2015 and 2016.

Bold lines within boxes represent the median, upper and lower limits of the box are the first and third quartiles, whiskers contain 1.5 times the interquartile. range, and open circles are outliers. Nestling hematocrit decreased as cimicid infestation increased (χ^2 = 27.85, $p < 0.01$).

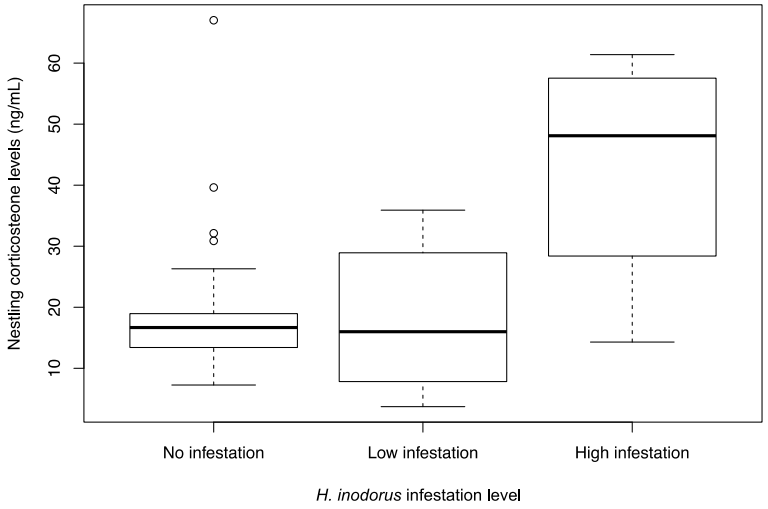


Figure A1.3 Corticosterone levels measured from golden eagle nestlings experiencing different levels of *H. inodorus* infestation in nests in southwestern Idaho, USA in 2015.

Bold lines within boxes represent the median, upper and lower limits of the box are the first and third quartiles, whiskers contain 1.5 times the interquartile range, and open circles are outliers. Nestling corticosterone levels (ng/mL) increased as *H. inodorus* infestation increased (χ^2 = 21.1, $p < 0.01$).

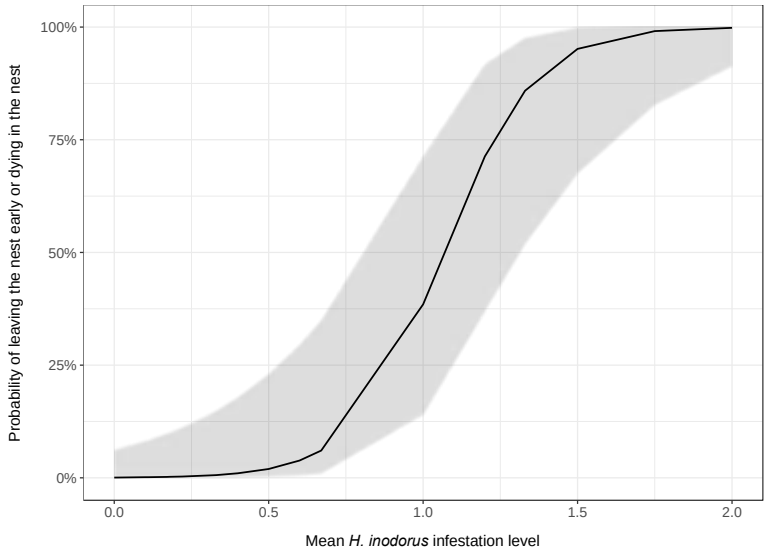


Figure A1.4 Predicted probability (solid dark line) and associated 95% confidence intervals (solid gray area) of golden eagle nestlings leaving the nest early, or dying in the nest, based on the mean infestation level at nests throughout the breeding season in southwestern Idaho, USA in 2015 and 2016.

The probability of leaving the nest early or dying in the nest increased as infestation increased ($\chi^2 = 10.58$, $p < 0.01$).

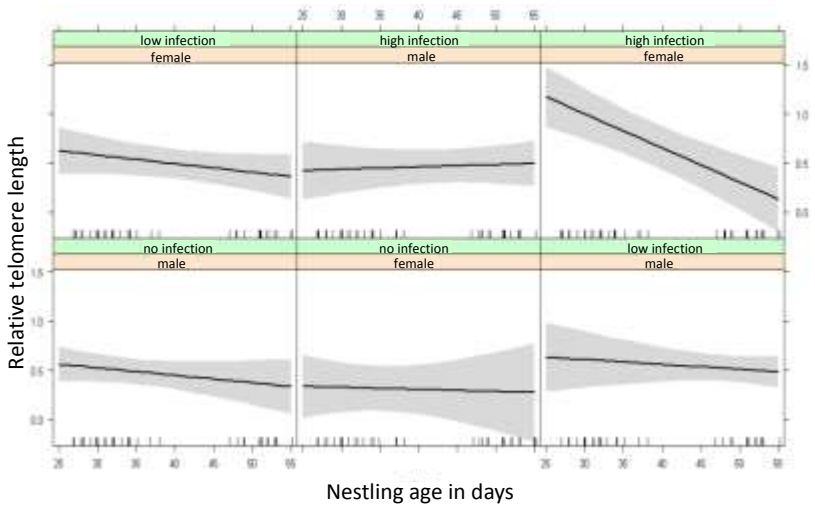


Figure A1.5 Golden eagle nestling relative telomere lengths and associated 95% confidence intervals (solid gray area) measured from nestlings experiencing different levels of H. inodorus infestation in nests in southwestern Idaho, USA in 2015 and 2016.

Female nestlings that experienced high levels of infestation had shorter relative telomere lengths than male nestlings or nestlings in nests with no or low levels of infestation.

Comparing lengths of telomeres extracted from different tissues of American Kestrels (*Falco sparverius*) and Long-billed Curlews (*Numenius americanus*)

Sadie C. Ranck, Stephanie F. Hudon, Ben Wright, Jay D. Carlisle, Eric J. Hayden, Julie A. Heath

Telomere length is used as a metric of fitness, stress and longevity in basic and applied wildlife research. In avian species, most telomere research is based on the collection of blood samples. However, blood collection may be considered invasive and require specialized training for sample collection. Alternatively, recent research suggests that tissue attached to the tips of feathers may provide a source of genomic DNA suitable for telomere analyses. It remains unclear whether this approach is consistent across species or time due to patterns in feather molt and different rates of turnover between tissue types.

We compared relative telomere lengths of DNA extracted from paired feather and blood samples from 12 adult American Kestrels (*Falco sparverius*) wintering in Boise, Idaho, and from six adult Long-billed Curlews (*Numenius americanus*) breeding in Wyoming. Relative telomere lengths were determined by quantitative polymerase chain reaction (qPCR). We found that telomeres in feathers are longer and more variable in kestrels (mean=3.5, σ^2=3.75) and curlews (mean=0.79, σ^2=0.28) when compared to telomere length estimates derived from blood in kestrels (mean=1.31, σ^2=0.11) and curlews (mean=0.75, σ^2=0.27). In kestrels, measured telomere ratios (T/R) in feathers ranged from 1.44 to 7.49 and ranged from 0.74 to 1.90 in blood. In curlews, measured

telomere ratios in feathers ranged from 0.13 to 1.27 and ranged from 0.14 to 1.57 in blood.

We found consistent differences between blood versus feather derived telomere lengths that were species dependent ($p=0.043$), indicating that the relationship between telomere lengths from these two tissue types is significantly different between kestrels and curlews (Fig. A2.1). Although the relationship between blood-derived and feather-derived telomere estimates tended to be different between species, neither species has a significant association between blood-derived and feather-derived relative telomere length estimates (kestrel 95% CI: -5.57 – 0.18, curlew 95% CI: -0.19 to 1.22).

The differences in relative telomere lengths measured in feathers and blood of kestrels and curlew are likely caused by the presence of different cell types in each sample. We expect that DNA extracted from whole blood was primarily sourced from nucleated red blood cells, with some contribution from other leukocytes. As for feathers, the extracted DNA was likely sourced primarily from keratinized pulp caps and feather follicle cells attached to the calamus. Feather follicles contain several types of stem cells that are needed to regenerate feathers after molting, yet the length of telomeres in the stem cells of avian feather follicles relative to differentiated skin cells or blood cells is unknown. It is possible that these stem cells have longer telomeres due to fewer cell divisions than differentiated cells, or due to differential telomere maintenance. Additionally, muscle cells and blood cells could also be present in the feather samples since feathers are integrated into the motor system and contain capillaries.

Results indicate that there was no significant correlation between relative telomere length estimates sourced from blood and feather samples for either species,

although the relationship between blood and feather telomere estimates was significantly different between kestrels and curlews. The lack of correlation within species could be the result of the different cell types present in each tissue type sampled, and different patterns between species suggests that there may be a species effect, or an effect from how long it has been since feather molt on the reliability of telomere length estimates derived from feather samples. These results suggest that additional work is needed before DNA sourced from feather tissue is used for deriving telomere length estimates, and to further understand why telomere lengths may differ between tissues.

My contribution to this work was to train and supervise Sadie and Ben in DNA isolation and perform qPCR-based telomere length assays on both feathers and blood in two species of birds.

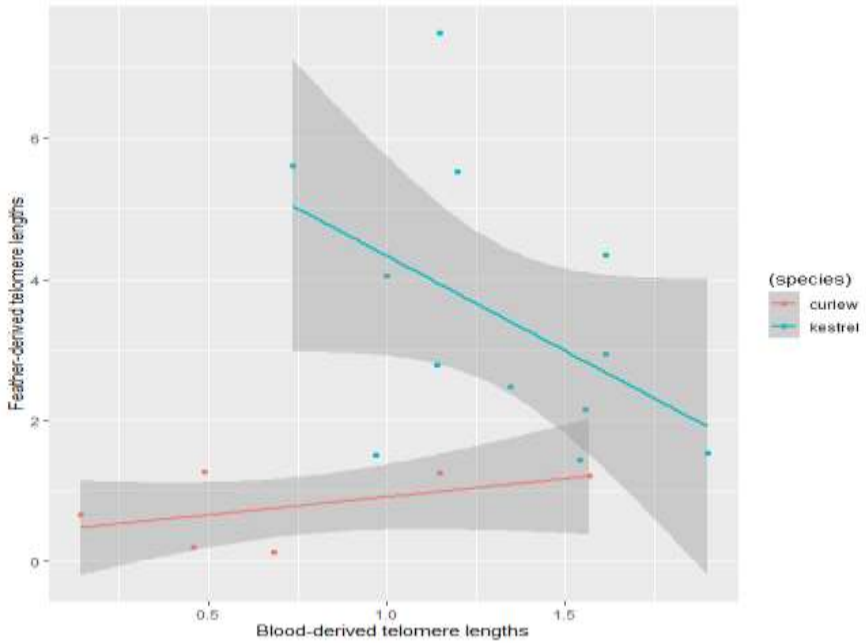


Figure A2.1 The association between telomere length estimates derived from blood and feather samples in 12 adult American Kestrels and 6 adult Long-billed Curlews.

There is no significant effect of blood-derived estimates in explaining feather-derived estimates in either kestrels (95% CI: -5.57 – 0.18) or curlews (95% CI: -0.19 to 1.22) and there is high variability in feather-derived telomere estimates. There is a significant interaction between blood telomere lengths and species (p=0.043), indicating that the relationship between feather and blood telomere lengths is significantly different between kestrels and curlews.

Rapid changes in the gut microbiome of a captive mule deer upon introduction to a natural diet

Stephanie F. Hudon, Alicia N. Wilkening, Eric J. Hayden, Jennifer Sorensen Forbey, Lisa A. Shipley

Submitted to *Microorganisms*, waiting for additional sample sequencing prior to resubmission

Herbivores rely on their gut microbial community for numerous functions. The importance of these functions to wildlife ecology is rapidly emerging. Although it is known that captivity can alter herbivore gut microbial communities, the effects are often species-specific and the dynamics and magnitude of effects for short-term captivity remain poorly studied. Here we report changes in the microbiome from fecal samples of a mule deer (*Odocoileus hemionus*) before and after its release into a natural habitat. Using shotgun metagenomics, we observed rapid shifts in the abundance of Bacteroidia populations in the fecal microbiome of this mule deer after transitions between a pellet diet and a diet with significant browse added (Figure A3.1). Functional annotation of sequence data showed changes in the abundance of genes belonging to carbohydrate and protein metabolism. Deer on a pellet diet have a higher abundance of functional genes for amino acid and protein metabolism and a lower abundance of carbohydrate metabolism genes relative to deer on a wild diet suggesting a link between dietary changes in the deer and metabolic capacity of the associated microbiome (Figure A3.2). In addition, the taxonomic diversity of samples and relatively long read lengths reported here provide an important baseline for future studies. Our results confirm the feasibility

of using shotgun metagenomics to assess the effects of short-term captivity on the function of the microbiome in mule deer.

Additionally, our results indicate that mule deer microbiomes can shift rapidly in response to diet changes associated with short term captivity and release. Most strikingly, we observed a dynamic response from the microbial class Clostridia (Figure A3.1). We find the highest relative abundance of Clostridia at the *acclimation* time point, when the deer had been eating browse from their natural habitat. Importantly, we find that the relative abundance of Clostridia mostly returns to *pre-release* levels after only a one week return to captivity (*post-release*) where the animal again received a pellet diet (Figure A3.1). This dynamic response suggests that the change in Clostridia is not simply an age effect. Instead, we conclude that the relative abundance of Clostridia is dependent on the diet of mule deer.

Another important dietary change associated with captivity is the high levels of protein found in the captive pellet feed. In fact, we found that the functional annotations of our data mirror this diet shift. Specifically, the decrease in relative abundance of genes associated with protein metabolism also decreases in the *acclimation* sample, when protein content is lowest (Figure A3.3). This further highlights the importance of the *acclimation* diet in preparing the microbiome for release conditions. These results suggest that the effect of high-protein on the mule deer microbiome should be investigated further, and future studies should be aimed at isolating the effect of protein alone, without the confounding effect of cellulose introduced by the natural browse.

We performed an in-depth analysis of the mule deer gut microbiota with the aim of understanding how diet within a captive and wild environment influences gut

microbes. Future studies will apply this study format while integrating metabolomics and when possible sampling of multiple gut compartments. Ideally, several gut compartments would be sampled for bacterial content including the stomach, small intestine, ceca and large intestine to get the full spectrum of bacterial colonization. However, this is not always possible when animal survival is necessary for reintroduction, in which case, obtaining as much information as possible from a stool sample becomes of great importance. Plant samples collected from each site can also help identify sources of phytochemicals identified in gut compartments and provide a fuller picture of the pathways uncovered during metabolomics studies.

My contribution to this work was to prepare the DNA libraries for Next Generation sequencing from the mule deer fecal samples. I also performed the data analysis for this study and wrote the manuscript. We have submitted additional samples for sequencing to increase our sample size for the study.

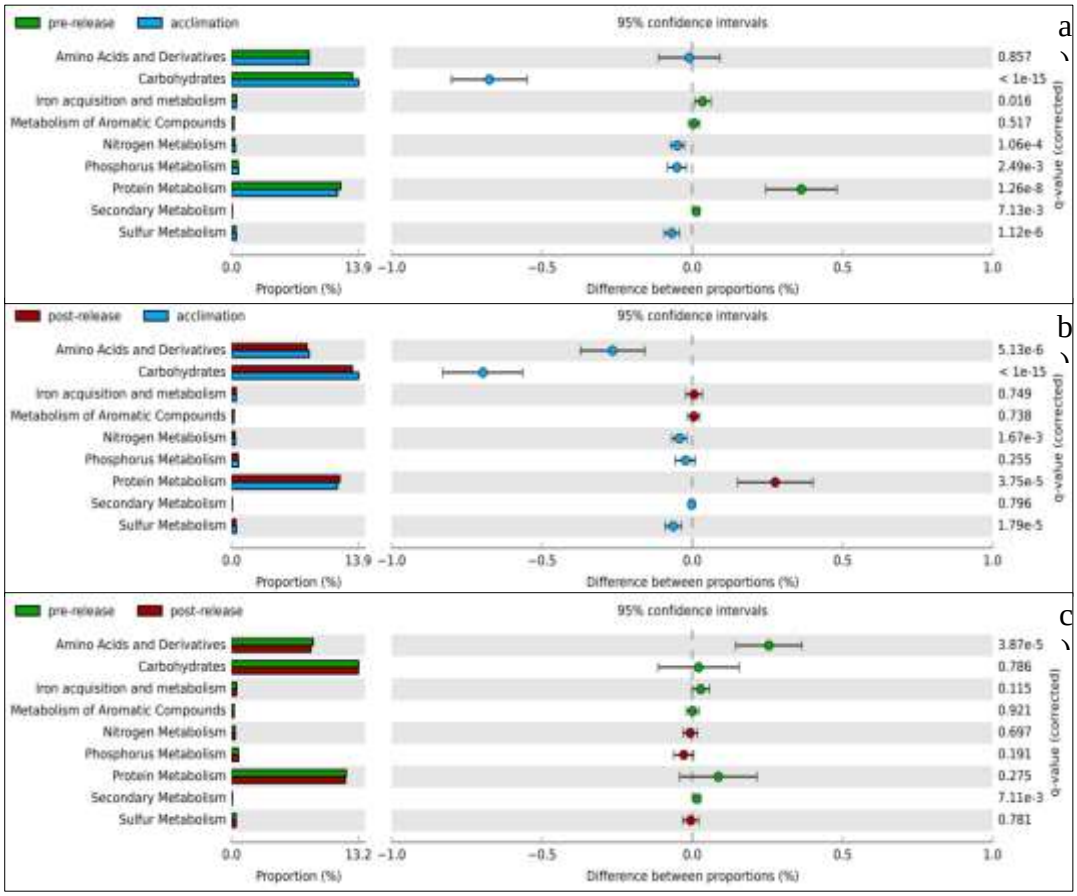


Figure A3.1 Taxonomical differences between a mule deer (*Odocoileus hemionus*) fed a pelleted diet with substantial browse prior to the transition into the wild (*acclimation*, blue) and a deer fed pellets in captivity before and after release into wild conditions (*pre-release*, green, *post-release*, red). Functional categories were compared with STAMP.

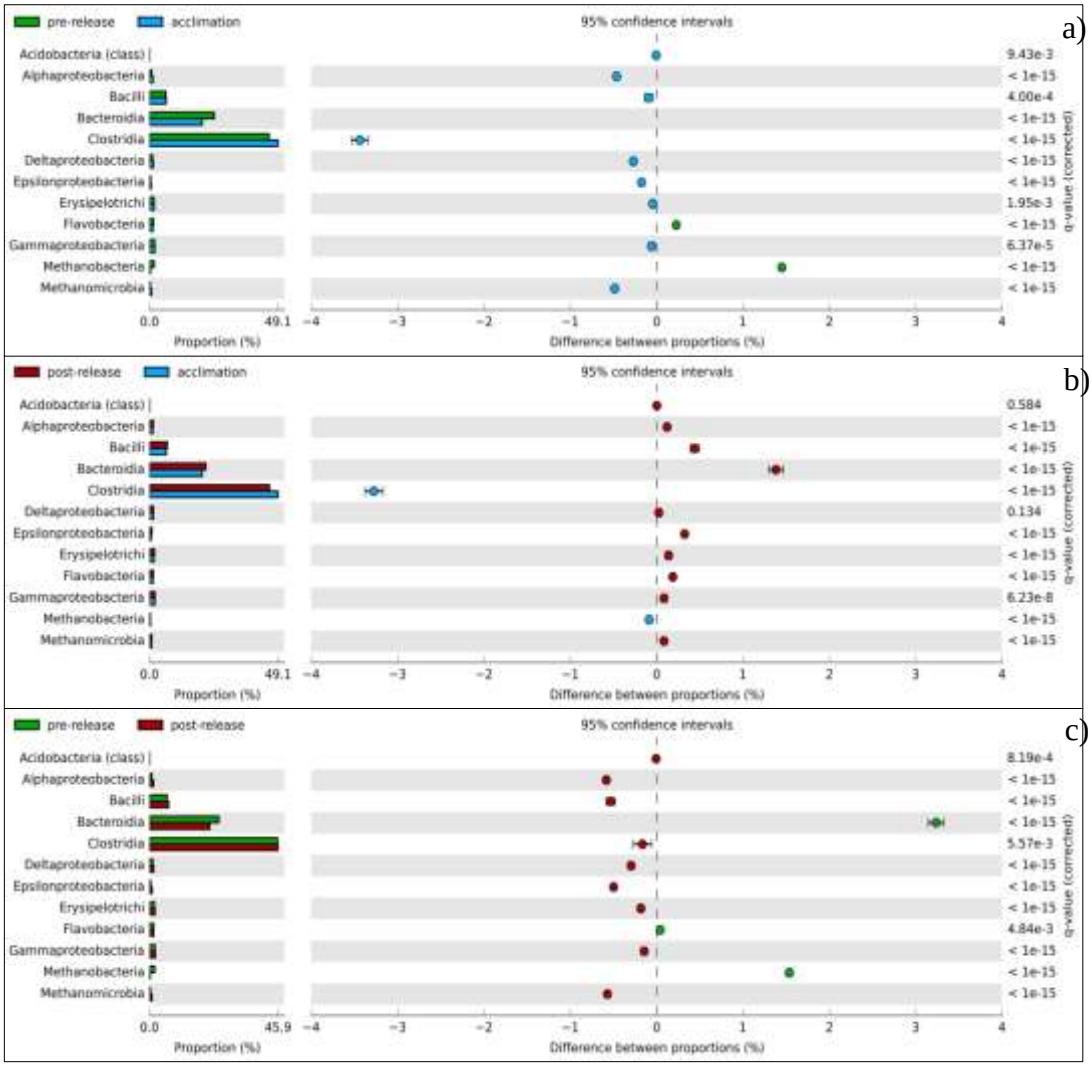


Figure A3.2 Functional categories (KEGG Level 2) between a mule deer (*Odocoileus hemionus*) fed a pelleted diet with substantial browse during the transition into the wild (*acclimation*, blue) and a deer fed pellets in captivity before and after release into wild conditions (*pre-release*, green, *post-release*, red). Functional categories were compared with STAMP.

Additional Projects and Collaborations

Telomere dynamics in a long-term, population study of the American kestrel (*Falco sparverius*)

Stephanie F. Hudon, Sadie C. Ranck, Kathleen R. Callery, Eric J. Hayden, Julie A. Heath

This project looked at how telomere lengths can be used as an index of stress and fitness in a population of American kestrels (*Falco sparverius*), which have been monitored for 30 years. During this time period the population has been affected by climate change, which affects the timing of nesting, as well as by urban sprawl, which has dramatically changed the kestrel habitat. American kestrels are considered a generalist predatory bird and tend to live in human-dominated environments. This makes the kestrel particularly susceptible to anthropogenic interference, which may result in decreased survival or reproduction. In southwestern Idaho the breeding season is becoming earlier since climate change has resulted in farmers planting earlier in fields where kestrels forage. This can result in a mismatch between peak prey levels and nestling life cycles. The Heath lab has also analyzed habitat characteristics such as the amount of native shrub-steppe and human disturbances such as land development and traffic conditions. These types of disturbances can lead to nest abandonment and therefore decreased reproductive success.

Kestrels are cavity-nesting birds which have made them easy to monitor through a system of 89 nest boxes stationed throughout Idaho. Every March the nest boxes are visited every 7-10 days to determine when clutches are produced. Adults are captured for

blood sample collection, banding and measurements and then the nests are surveyed to determine the number of successful nestlings produced. After hatching, the nestlings are also sampled and measured. Following the field collections, I performed a telomere assay on 78 birds which included both nestlings and adults. This represented a significant amount of work since a common reference gene, GAPDH, exhibited variance in the birds and therefore we had to find suitable reference genes for the kestrel before the assay could be performed. We are now in the data analysis stage of this work. Preliminary results suggest that brood size may play significant role in nestling fitness (Figure A.4.1). The average telomere length of broods with three birds appears to be significantly longer than those of larger brood sizes. This may be due to stress produced as siblings compete for limited food resources. Additional analysis will soon be performed to prepare a manuscript on this work.

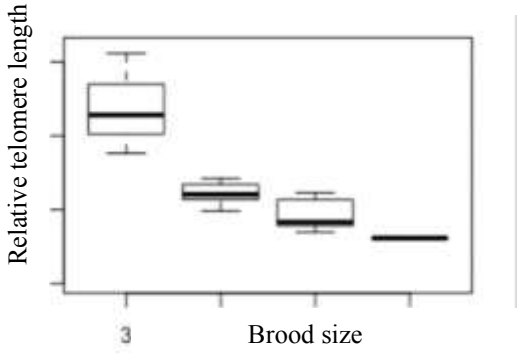


Figure A.4.1 Telomere lengths of American kestrel (*Falco sparverius*) nestlings from broods of different sizes.
The average telomere length of a brood size of 3 is significantly different from a brood size of 4 ($p=0.003$), 5 ($p<0.001$) or 6 ($p<0.001$).

The effects of haemosporidia parasite infections on telomere lengths of Northern Goshawks (*Accipiter gentilis*) in the Minidoka Ranger District of the Sawtooth National Forest

Stephanie F. Hudon, Julianna Ramirez, Robert A. Miller and Eric J. Hayden

The Northern Goshawk (*Accipiter gentilis*) is considered a management indicator species for the Sawtooth National Forest. As such, management plans within the Sawtooth National Forest must consider the effects on the habitat and ecological needs of the Northern Goshawk and the status of the goshawks serves as indicator of how other non-monitored species are doing. Top predators, such as raptors, are indicators of prey abundance within the forest food chain. Within the Sawtooth National Forest, over 50,000 acres are surveyed each year, including the monitoring of nearly 200 Northern Goshawk nests. In addition to surveying landscape features such as canopy closure and stand density, it also is important to establish the health of the goshawk population.

Blood parasites (haemosporidia) have been shown by several studies to play a role in the health and life span of avian species. Previous work using blood smear analysis has shown that the Haemosporidia of the genera, Leucocytozoon, which uses blackflies to infect birds, is present in the Northern Goshawk population. To analyze the effect of these blood parasite infections on the birds we utilized a quantitative polymerase chain reaction (qPCR) approach developed by Tkach et al 2015 to determine the infection level within the goshawk population of three genera of blood parasites, Plasmodium, Haemoproteus and Leucocytozoon. We determined that the qPCR-based approach

correlated well with blood smear counts but with a higher sensitivity level for low-level infections. This assay has now been applied to 56 birds. We are also performing telomere length assays on the goshawks to determine the effects of parasitism on the health and fitness of the bird population and to find the areas of highest infection within the Sawtooth National Forest. This work is ongoing. We have isolated DNA from 124 samples and will continue to process them for both parasite levels and relative telomere lengths. My part in this project has been to perform DNA isolation from blood samples and run the telomere and parasite assays by qPCR. Recently, Julie Ramirez joined this project and plans to run a more specific assay for just the Leucocytozoon parasite as well as to run additional telomere assays on newly acquired samples.

Students Mentored

Julianna Ramirez: trained in single and multichannel pipetting and performing and analyzing telomere and parasite load assays in Northern Goshawks. Julie will continue work on parasites within these samples for the next couple of years. Julie presented a poster at the Boise State Biology Undergraduate Research Showcase titled "Factors Influencing the Genetic Health of Northern Goshawks Across the Great Basin."

Alicia Wilkening: I trained Ali in pipetting and library preparation of samples for Next Generation Sequencing, as well as performing parasite assays on golden eagles using qPCR. She also helped to develop an assay to fragment and purify library samples to similar lengths for improved next generation sequencing results. Ali is an author on a paper in progress on the microbiome analysis of captive and wild populations of mule deer and she presented her poster "An Assay for Blood Parasite Detection in Golden

Eagle (*Aquila chrysaetos*)" at the Midstate Undergraduate Research Symposium as well as the Idaho Conference of Undergraduate Research.

Sadie Ranck: I trained Sadie in DNA isolation from blood samples and single and multichannel pipetting to analyze telomere lengths from American kestrel blood samples. Sadie went on to enroll in Boise State University's Master of Science program in Raptor Biology where she continues to work on the American kestrel. Sadie and I will author two manuscripts together and she presented a poster titled "Heritability of Telomere Length in American Kestrels" at the Idaho Conference on Undergraduate Research.

Kathleen Callery: I trained Katie in DNA isolation from blood samples and single and multichannel pipetting to develop the original telomere assay presented in Chapter 1 of this dissertation. Katie went on to enroll in Boise State University's Master of Science program in Raptor Biology where she works on the American kestrel. Katie is a coauthor on the universal telomere assay manuscript presented in Chapter 1 of this dissertation and presented a poster titled "What Correlates with Telomere Length in American Kestrels (*Falco sparverius*)?"

Esteban Palencia Hurtado: I trained Esteban in single and multichannel pipetting and DNA isolation from a variety of tissue types. Esteban and I worked together for over two years developing the universal telomere assay. Esteban is second author on the submitted manuscript on developing a universal telomere assay and presented a poster at the Idaho

Conference on Undergraduate Research at Boise State titled “Telomere Estimates by qPCR in American kestrels.”

Benjamin Wright: I trained Ben in pipetting and how to isolate DNA from blood and feathers from his curlew samples. Ben learned how to do telomere assays on these samples and will be a co-author on a manuscript in progress. He is currently volunteering in the lab and applying to Boise State’s Biomolecular Sciences program.

Stacie Loisate: I trained Stacie in multichannel pipetting and DNA isolation from primary mouse cell lines for telomere analysis. Stacie is now a co-author on a submitted manuscript and presented her poster titled “Effects of Microgravity and Disruption of LINC Complex on Cellular Compartmentalization of YAP and TAZ” at the Idaho Conference on Undergraduate Research.

AnnaGrace Blomquist: I trained AnnaGrace in single and multichannel pipetting and DNA isolation from primary mouse cell lines for telomere analysis. She presented her poster titled “Dystrophin-Glycoprotein Complex and Reactive Oxygen Species” at the Boise State Undergraduate Research and Scholarship Conference.