UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# ANTHROPOGENIC IMPACTS: BRINGING MULTI-OMIC APPROACHES TO NICHE CONSTRUCTION THEORY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

KRISTEN M. RAYFIELD
Norman, Oklahoma
2023

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/337552

ANTHROPOGENIC IMPACTS: BRINGING MULTI-OMIC APPROACHES TO NICHE CONSTRUCTION THEORY

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ANTHROPOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Courtney Hofman, Chair
Dr. Krithivasan Sankaranarayanan
Dr. Jessica Cerezo-Román
Dr. Tassie Hirschfeld
Dr. Hayley Lanier

© Copyright by KRISTEN M. RAYFIELD 2023
All Rights Reserved.

*To my mother, who's positivity and response to life's challenges inspires me*
*and to my best friend and dearest husband who has always provided unconditional support no matter the cost, time, or distance.*

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my Committee Chair, and Advisor, Dr. Courtney Hofman. Thank you for the constant support, guidance, mentoring, as I obtain my Ph.D. I will forever be indebted to you for accepting me as your graduate student so that I could be a part of OU and the Laboratories of Molecular Anthropology and Microbiome Research to do the research that I love. It has always been my dream since I was five when I first found my dad's dissertation to receive a Ph.D., though at the time, I did not know what a dissertation was, I just thought it was cool he wrote a book! Thank you for guiding me through numerous opportunities including Smithsonian Fellowships, research grants and awards, and domestic/international training workshops. The researcher I am today would not have been possible without your encouragement and mentorship.

I would like to thank Dr. Krithi Sankaranarayanan for first training me in my early laboratory skills and coding as well as teaching me the methodology in microbiome research. I would also like to thank my committee who have helped shape this dissertation and provided me with the necessary feedback and guidance to obtain my degree. This would not have been possible without their help. In addition, I would like to the thank the Anthropology Department for first accepting me into the program and for helping me grow into a teaching role.

This final year has also been made easier with the support of Drs. Sabrina Sholts and Alexis Mychajliw who have read numerous versions of my work and have helped strengthen many of the research ideas discussed in this dissertation. Finally, I would like to thank Drs. Sabrina Shirazi and Lilly Parker, as well as Robin Singleton, Nihan Dagtas, and Karissa Hughes for their friendships, guidance, support, and training. Long days at the lab would not have been as much fun without them. Finally, I would like to thank my family for always encouraging me and giving me confidence throughout this journey.

## ABSTRACT

Anthropogenic effects on the environment, ecology, and biodiversity of animals are not exclusive to the Anthropocene but extend deep into antiquity. In fact, many archaeologists and historical ecologists argue that the complex interplay between humans and their environment for the past millennia have shaped the Anthropocene. Archaeologists, in particular, have long recognized humans as the ultimate niche constructors given their ability to significantly alter the environment to meet their needs, which in turn creates and influences other natural selective pressures. Such effects have been observed within early human-animal interactions and the resulting transition to animal domestication. In addition, it has been argued that these early human alterations to both our environment and animals fostered the rise in zoonotic diseases. New biomolecular tools and advances in the "-omics" coupled with the archaeological record can provide meaningful snapshots in time to address how humans have facilitated changes to domestication patterns and consequentially animal health and the spread of zoonotic diseases. As such, this dissertation research takes a theoretical, methodological, and data driven approach to address human-animal interactions and the long dynamic nature of anthropogenic disruptions. The impact of this dissertation is threefold: increasing our understanding of human-made environmental impacts, increasing the recognition of natural history collections and biorepositories as valuable resources to address historical shifts in biodiversity and ecologies due to human activity, and the incorporation of microbiome data in broadening our understanding of human-animal interactions. As a result, this dissertation research offers a transdisciplinary approach that will fill the gap in studies of past human-animal interactions.

# TABLE OF CONTENTS

**DEDICATION** ..... **IV**

**ACKNOWLEDGEMENTS** ..... **V**

**ABSTRACT** ..... **VI**

**TABLE OF CONTENTS** ..... **VII**

**LIST OF TABLES** ..... **X**

**LIST OF FIGURES** ..... **XI**

**CHAPTER 1** ..... **- 1 -**

**RECONSTRUCTING PAST HUMAN AND ANIMAL ENVIRONMENTS** ..... **- 1 -**

**1.1 INTRODUCTION** ..... **- 1 -**
**1.3 A PROTEOMIC METHOD FOR SEXING ANIMAL REMAINS** ..... **- 4 -**
**1.4 ANTHROPOGENIC EFFECTS ON MAMMALIAN ORAL MICROBIOMES** ..... **- 6 -**
**1.5 CONCLUSION** ..... **- 7 -**

**CHAPTER 2** ..... **- 8 -**

**UNCOVERING THE HOLOCENE ROOTS OF CONTEMPORARY DISEASE-SCAPES: BRINGING ARCHAEOLOGY INTO ONE HEALTH** ..... **- 8 -**

**2.1 INTRODUCTION** ..... **- 8 -**
**2.2 EPIDEMOLOGICAL TRANSITIONS AS MANIFESTATIONS OF HUMAN NICHE CONSTRUCTION** ..... **- 10 -**
2.2.1 FIRST EPIDEMIOLOGICAL TRANSITION ..... - 11 -
2.2.2 SECOND EPIDEMIOLOGICAL TRANSITION ..... - 12 -
2.2.3 THIRD EPIDEMIOLOGICAL TRANSITION ..... - 13 -
2.2.4 OVERLAPPING EPIDEMIOLOGICAL TRANSITIONS ..... - 13 -
**2.3 TOOLS AND DATASETS TO RECONSTRUCT PAST DISEASE-SCAPES** ..... **- 15 -**
2.3.1 RECONSTRUCTING PAST ANIMAL POPULATIONS AND HEALTH ..... - 16 -
2.3.2 RECONSTRUCTING PAST HUMAN POPULATIONS AND HEALTH ..... - 17 -
2.3.3 RECONSTRUCTING PAST ENVIRONMENTS ..... - 19 -
2.3.4 TOOLS TO RECONSTRUCT PAST HUMAN-ANIMAL-ENVIRONMENT INTERACTIONS ..... - 20 -
2.3.5 EVALUATING SYNERGIES: HOW DID HUMAN-ANIMAL-ENVIRONMENT INTERACTIONS CONTRIBUTE TO ZOONOTIC SPREAD? A FOCUS ON RATS. ..... - 21 -

**2.4 NEW AVENUES FOR TRANSDISCIPLINARY RESEARCH AND DATA SOURCES ..... - 22 -**
2.4.1 MUSEUM SPECIMENS AS A SOURCE OF DATA ..... - 23 -
2.4.2 APPLICATIONS TO CONSERVATION PROBLEMS ..... - 23 -
**2.5 ETHICAL CONSIDERATIONS ..... - 23 -**
**2.6 CONCLUSIONS ..... - 24 -**

**CHAPTER 3 ..... - 26 -**

**A PROTEOMIC METHOD TO DETERMINE SEX FROM ZOOARCHAEOLOGICAL DENTAL REMAINS ..... - 26 -**

**3.1 INTRODUCTION ..... - 26 -**
**3.2 BACKGROUND ..... - 27 -**
**3.3 MATERIALS AND METHODS ..... - 29 -**
3.3.1 MASS SPECTROMETER PARAMETERS ..... - 30 -
3.3.2 *IN SILICO* ANALYSIS ..... - 30 -
3.4.1 SEX IDENTIFICATION BY MASS SPECTROMETRY ..... - 36 -
**3.4 RESULTS/DISCUSSION ..... - 38 -**
**3.4.1 CHALLENGES TO DETERMINE SEX IN ANIMALS ..... - 38 -**
**3.6 CONCLUSION AND FUTURE DIRECTIONS ..... - 42 -**

**CHAPTER 4 ..... - 44 -**

**ANTHROPOGENIC EFFECTS ON MARINE MAMMAL ORAL MICROBOMES ..... - 44 -**

**4.1 INTRODUCTION ..... - 44 -**
**4.2 BACKGROUND ..... - 45 -**
**4.3 MATERIALS ..... - 47 -**
**4.4 METHODS ..... - 48 -**
4.4.1 LABORATORY METHODS ..... - 48 -
4.4.2 BIOINFORMATIC WORKFLOW ..... - 50 -
**4.5 RESULTS AND DISCUSSION ..... - 51 -**
4.5.1 PRESERVATION AND AUTHENTICATION ..... - 52 -
4.5.3 MICROBIAL PRESERVATION AND TAXONOMIC PROFILING ..... - 56 -
4.5.4 ECOLOGICAL IMPACTS AND MICROBIAL PRESENCE ..... - 60 -
**4.6 CONCLUSION ..... - 62 -**

**CHAPTER 5 ..... - 64 -**

**CONCLUSION ..... - 64 -**

**5.1 SUMMARY OF WORK** ..... - 64 -
**5.2 FUTURE DIRECTIONS** ..... - 66 -

**REFERENCES** ..... - 68 -

**SUPPLEMENTAL 1** ..... - 93 -

**SUPPLEMENTAL 2** ..... - 94 -

**SUPPLEMENTAL 3** ..... - 97 -

**SUPPLEMENTAL 4** ..... - 98 -

**SUPPLEMENTAL 5** ..... - 102 -

**SUPPLEMENTAL 6** ..... - 107 -

**SUPPLEMENTAL 7** ..... - 112 -

**SUPPLEMENTAL 8** ..... - 116 -

**SUPPLEMENTAL 9** ..... - 120 -

**SUPPLEMENTAL 10** ..... - 124 -

# LIST OF TABLES

Table 1: Summary of specimens in this study ..... - 30 -
Table 2: Generated spectra across taxa samples ..... - 39 -
Table 3: Sex determination overview from archaeological human dental remains ..... - 41 -
Table 4: Sex determination results ..... - 42 -
Table 5: Collection sampling population across three time periods ..... - 49 -
Table 6: Summary of sequencing data ..... - 52 -

# LIST OF FIGURES

Figure 1: Human niche construction of disease-scapes.. ..... - 11 -
Figure 2: Origin and spread of morbilliviruses..... - 15 -
Figure 3: Tools and datasets to reconstruct past disease-scapes.. ..... - 16 -
Figure 4: Protein extraction workflow..... - 30 -
Figure 5: Mammalian AMELX and AMELY alignment ..... - 32 -
Figure 6: *Canis* species and *Cervid* species alignments. ..... - 34 -
Figure 7: Phylogenetic analysis of the amelogenin protein (AMELX and AMELY) across multiple mammalian taxa. ..... - 35 -
Figure 8: N-terminal and C-terminal domain within the amelogenin protein..... - 40 -
Figure 9: Individuals in this study. ..... - 48 -
Figure 10: 5'C toT damage. ..... - 53 -
Figure 11: Average fragment length and endogenous content ..... - 54 -
Figure 12: 1970s NMML and SBMNH damage patterns, fragment length, and endogenous content.. ..... - 55 -
Figure 13: Comparative sources. ..... - 57 -
Figure 14: Mammal calculus. ..... - 57 -
Figure 15: Collection comparison. ..... - 58 -
Figure 16: Heatmap of the top bacteria genera across mammalian oral microbiome samples. - 60 -
Figure 17: Competitive mapping of commensal and pathogenic microbes. ..... - 62 -

# CHAPTER 1

## Reconstructing past human and animal environments

### 1.1 INTRODUCTION

Major shifts in human behavior and their alterations to the environment that we consider as instigators of the Anthropocene can be associated with human behavior deep into antiquity (Braje and Rick, 2011; Smith and Zeder, 2013; Hofman *et al.*, 2015; Boivin *et al.*, 2016; Roberts *et al.*, 2017; Ellis, 2021; Ellis *et al.*, 2021). Such alterations have shaped both human and animal demographics and behavioral responses to ecological niches for the past millennia (Smith and Zeder, 2013; Boivin *et al.*, 2016) – a process known as niche construction (Odling-Smee, Laland and Feldman, 1996; Laland, Odling-Smee and Feldman, 2001). Archaeologists have long recognized humans as the ultimate niche constructors given their ability to significantly alter the environment to meet their needs, which in turn creates and influences other natural selective pressures (Laland, Odling-Smee and Feldman, 2001). Niche Construction Theory (NCT) provides a co-evolutionary framework from which to assess how organisms influence and shape their environment. Within archaeology NCT provides a multilinear approach evaluate the selective ecological pressures on organisms that have allowed for the succession of certain species. This allows archaeologists to incorporate human agency and a behavioral component with NCT to observe genetic inheritance and exchange within a dynamic environment, tying in well with concepts of gene-culture coevolution (Laland and O'Brien, 2010; O'Brien and Laland, 2012). This has been observed within early human-animal interactions and the resulting transition to animal domestication (Zeder, 2016) as well as the consequential rise of zoonotic diseases due to early human alterations to both our environment and animals (Cockburn, 1971; Harper and Armelagos, 2013; Zuckerman *et al.*, 2014).

Niche Construction and the Anthropocene parallel with one another as the Anthropocene is viewed as a transition from "natural" landscapes to cultural landscapes (Crutzen and Stoermer, 2000; Waters *et al.*, 2016; McMahon, Morand and Gray, 2018; Waters and Turner, 2022). The complex interplay between humans and their environment for the past millennia have shaped the Anthropocene (Braje and Rick, 2011; Smith and Zeder, 2013; Hofman *et al.*, 2015; Szabó, 2015; Armstrong *et al.*, 2017; Ellis *et al.*, 2021). As a result, the onset of the Anthropocene has been equated with different periods of time, such as the early 1800s and the onset of the Industrial Revolution (Crutzen and Stoermer 2000; Steffen et al. 2011), the 1610 Orbis Spike and the Euro-colonial expansion into the Americas (Lewis and Maslin, 2015), the Neolithic Demographic Transition and the adoption of agriculture (Smith and Zeder 2013; Ruddiman 2017), to as far back as the presence of controlled fire (~0.5 mya) (Gowlett 2016; Lucock et al. 2013). Consequently, human niche constructions have also led to biodiversity loss through habitat fragmentation, particularly deforestation, and climate change (Laland et al. 2001). These "anthromes or anthropogenic biomes" shape the genetic selection of multiple species around us and have been built upon several successive rounds of ecological inheritance (Fuentes and Baynes-Rock, 2017; Ellis, 2021)

The archaeological record – an archive of the long-term human-environment interactions – can provide insight to the dynamic and contingent nature of how anthropogenic activities alter animal demographics, ecosystems, and disease landscapes over time. Now, new biomolecular tools and advances in the "-omics" coupled with the archaeological record can provide meaningful snapshots in time to address how humans have facilitated changes to domestication patterns and consequentially environmental and animal health and the spread of zoonotic diseases. As such, this dissertation research takes a theoretical (Chapter 2), methodological (Chapter 3), and data-driven (Chapter 4) approach to address human-animal interactions and the long dynamic nature of anthropogenic disruptions.

I first provide a theoretical approach in addressing how NCT coupled with the One Health perspective can be used to reconstruct past human created disease landscapes, from now on referred to as "disease-scapes". Within this review chapter, I discuss how humans, as niche constructors, have driven disease emergence for millennia. I highlight new "-omic" methodologies that can be incorporated to study past human-animal-environment interactions and how contemporary "disease-scapes" are built on successive rounds of anthropogenic disruptions. In addition, I provide alternative datasets that can be incorporated to model how these changes have taken place over time. Within Chapters 3 and 4, I demonstrate how ancient proteomic and genomic methods can be used to address past human-animal-environment interactions. I apply a proteomic methodological approach to determine sex from archaeological animal dental remains and address the application of this technique when studying early animal husbandry practices. In Chapter 4, I tie many of the themes introduced within Chapter 2 (review of One Health and NCT) with a genomic data driven approach to study human alterations to marine mammal oral microbiomes and past ecologies. Overarchingly, this dissertation explores how we can reconstruct past human-animal-environmental interactions coupling theoretical concepts with new biomolecular tools that can fill the gap in these studies.

## 1.2 UNITING ONE HEALTH AND NCT TO STUDY PAST DISEASE-SCAPES

Approximately 60% of human pathogens are associated with non-human animal origins (Woolhouse and Gowtage-Sequeria, 2005; Jones *et al.*, 2013). The recent rise in re-emerging infectious diseases such as malaria, tuberculosis and brucellosis and emerging infectious diseases including HIV, H1N1, SARS, MERS, and now coronaviruses argue for a deeper investigation into human-animal relationships and their associated interactions with their environments. One approach that recognizes and enhances detection of the spread of zoonotic diseases is the adoption of a One Health strategy. A One Health approach that mitigates and prevents future pandemics and enhances the surveillance of human-animal interfaces with the greatest zoonotic disease risk has increasingly gained attraction (Heymann and Dixon, 2013; Plowright *et al.*, 2021). The adoption of this approach stresses how human, animal, and environmental health are codependent of each other and influence health outcomes (Salyer *et al.*, 2017; Bendrey *et al.*, 2020).

Zoonotic transmission is a complex cascade of events and factors that require the alignment of ecological, epidemiological, and behavioral determinants to allow for favorable transmission of a pathogen to a new host species (Plowright *et al.*, 2017). The current One Health model equates this cascade of events to anthropogenically induced ecological changes. Most frequently the

major anthropogenic influences on these ecological changes include landscape alteration, global travel of humans, domesticated animals, wildlife trade, plants and microbes, the introduction of antibiotics and antimicrobial resistance, and the geographic expansion of animals and disease vectors due to climate change (McMahon, Morand and Gray, 2018). Such alterations have been further linked to the rising rate of emerging and re-emerging infectious disease where 75% of these infectious diseases originate from wildlife populations (Woolhouse and Gowtage-Sequeria, 2005; Jones *et al.*, 2013; White and Razgour, 2020). In addition, environmentally introduced antimicrobial resistance through the misuse and overuse of antibiotics in both humans and domesticated animals (Brealey *et al.*, 2021) as well as industrialized food production (Atlas and Maloy, 2014), and the expanding interface between humans, domesticated animals, and wildlife has created new ecological niches for pathogenic microbes to spread to alternative host species (Atlas and Maloy, 2014). These alterations are expected to intensify in the Anthropocene and have been doing so at an alarming rate (Jones *et al.*, 2013; McMahon, Morand and Gray, 2018).

Environmental modifications ultimately change the infections humans and animals are likely to encounter and can therefore provide critical details when addressing coevolutionary disease relationships. Current emerging disease trends have increasingly been linked to anthropogenic land alterations and highlight the impact environmental disruptions, both natural and cultural, play on disease outbreaks. To consider a select few – land conversion and deforestation, habitat fragmentation and climate change – have been significantly linked to both past and present ecological changes and disease dynamics. This has been readily documented with the creation of agricultural land and Nipah virus outbreaks (Epstein *et al.*, 2006; McMahon, Morand and Gray, 2018) and the reversion to woodlands with the expansion of tick borne diseases such as Lyme's Disease (Ostfeld and Keesing, 2000, 2012). Habitat fragmentation and deforestation have fostered the emergence of Ebola, HIV along with numerous vector borne diseases such as Zika (Ali *et al.*, 2017; McMahon, Morand and Gray, 2018). Finally, geographic expansion and seasonal fluctuations of mosquito vectors like Blue Tongue Virus and West Nile Virus are have been linked to anthropogenic climate change (Gibbs and Greiner, 1994; Wilson and Mellor, 2009; Paz, 2015). As a result, changes in the landscape can consequently influence the types of organisms – animals, insect vectors, microbes – humans are now sharing space with, as well as what organisms would benefit from cultural modifications within an anthropogenically constructed niche.

To date, the One Health strategy has most frequently been applied to altered terrestrial ecosystems, focused on either the human-wildlife-interface or industrialized food production with a focus on antibiotic resistance. Archaeology has a long history of recognizing zoonotic disease transmission from animals to humans (Cockburn, 1971; Goodman *et al.*, 1984; Cohen, 1989; Dobson and Carper, 1996; Bocquet-Appel, Naji and Bandy, 2008; Harper and Armelagos, 2013; Harper *et al.*, 2013; Zuckerman *et al.*, 2014; Roberts and Buikstra, 2019). This has been readily documented with the First Epidemiological Transition and the advent of agriculture (Omran, 1971; Harper and Armelagos, 2013; Zuckerman *et al.*, 2014). The resulting altered environments, the transition to a sedentary lifestyle, the domestication of animals, and an increase in population densities consequently created new ecological niches for the transmission of zoonotic and vector borne pathogens (Harper and Armelagos, 2013; Zuckerman *et al.*, 2014). Yet, archaeology has not fully been incorporated in One Health discussions. Within this review chapter I demonstrate that that current disease-scapes are spillovers of ecological inheritance and can be untangled

using archaeological methods. This has been documented with the successive layering of large scale anthropogenic disruptions on the environment as well as the ecology and biogeography of animals that extends to the Late Pleistocene, beginning when ancient peoples first hunted megafauna and burned landscapes (Ellis, 2021; Ellis *et al.*, 2021). Four unique patterns of transitions that are consistent with the Anthropocene include (1) the extinction of megafauna at the end of the Pleistocene (2) the Neolithic Demographic Transition and the adoption of agriculture (3) the colonization of islands (4) and the rise of urbanization and creation of large commercial networks (Boivin *et al.*, 2016). While the start of the Anthropocene is continuously debated, each of these suggested periods have been associated with major environmental disruption, changing demographics, as well as the creation of new ecological niches at the human-animal interface (Atlas and Maloy, 2014; Boivin *et al.*, 2016; Ellis, 2021; Ellis *et al.*, 2021) – the key drivers that One Health recognizes as contributing to new disease-scapes. As a result, the full anthropogenic impact on zoonotic disease ecology and current disease-scapes has not been well addressed with One Health alone (Messenger, Barnes and Gray, 2014; Wolf, 2015). A temporal perspective provided through NCT must be incorporated to understand the disease dynamics within current disease-scapes.

By combining zooarchaeological, bioarchaeological, and palaeoecological data under a One Health framework, I address how humans, as niche constructors, have facilitated new host and reservoir species and disease-scapes from the Late Pleistocene and Holocene directly into the Anthropocene. I also address how the archaeological record has relevance for environmental and conservation concerns, of importance, the archaeological record has informed our understanding of past human-environment interactions and the behaviors adopted under times of environmental stress. These studies have helped in addressing ways to bridge archaeology with current climate and environmental problems we face today within the Anthropocene (climate change, drought, El Niño events, etc.) (Burke *et al.*, 2021). I discuss how museum collections can be used to provide a timestamp of change and a temporal perspective in understanding past disease (DiEuliis *et al.*, 2016; Dunnum *et al.*, 2017). Finally I address how archaeology can be used to understand the impact of past pandemics on health disparities that persist into the present (Dávalos *et al.*, 2020; Wade, 2020) and provide ethical consideration for conducting collaborative and inclusive research with descendant communities.

## 1.3 A PROTEOMIC METHOD FOR SEXING ANIMAL REMAINS

One area we can further explore the scale of human influence, is the study of animal husbandry and domestication. We can do this using proteomic methods as a complementary or alternative method to ancient DNA, as it is generally more reliable, less expensive, and less destructive (Hendy *et al.*, 2018). Proteomics is the study of proteins produced by organism (proteomes) (Hendy *et al.*, 2018; Hendy, 2021) and as a field has shown that proteins degrade slower than DNA and within more temperate climates, thus making them more applicable for analyses in multiple parts of the world (Hendy *et al.*, 2018; Welker *et al.*, 2020). The analysis of proteins and recovery of amino acid sequences has been practiced since the 1950s, however their application to ancient contexts and the innovative development of methods in their analyses has grown exponentially in the past two decades (Cappellini, Collins and Gilbert, 2014; Hendy *et al.*, 2018; Hendy, 2021). The applications of proteomic research can be broadly applied to organic

materials, reconstructing subsistence patterns, diet, and past human diseases (Hendy, 2021; Warinner, Korzow Richter and Collins, 2022).

Residual protein analysis has allowed archaeologists to explore how people were engaging with their environment and they type of resources they were using through the analysis of cultural material. This has been documented in the protein analysis of paint and glue substances used for art as well as pot sherd for dietary reconstructions(Tokarski *et al.*, 2006; Solazzo *et al.*, 2008; Dallongeville *et al.*, 2013). Most notably proteomic research has been used for reconstructing phylogeny and species identification (Cappellini *et al.*, 2012, 2019; Coutu *et al.*, 2021). In particular, the application of Zooarchaeology by Mass Spectrometry (ZooMS) has provided taxonomic resolution and ecological reconstruction where archaeological methods alone cannot (Buckley *et al.*, 2009; Buckley and Collins, 2011; Richter *et al.*, 2022). ZooMS uses the amino acid sequence of collagen peptides recovered from bones to taxonomically identify unidentifiable fragments (Buckley *et al.*, 2009; Buckley and Collins, 2011; Richter *et al.*, 2022).

Stewart and contributors use a different protein, the amelogenin protein, from archaeological human remains for sex determination. This proteomic approach has been applied to determine sex from human dental remains with success (Stewart *et al.*, 2017; Parker *et al.*, 2019; Buonasera *et al.*, 2020). Sex determination of human remains is fundamental in archaeology and forensic cases, especially among juveniles who cannot be sexed by osteology. Within zooarchaeological contexts, the ability to determine sex among animal remains is critical to address animal harvest, husbandry and exploring domestication processes. Doing so sheds light on how people influenced the reproduction of animals. Determining the proportion of males versus females within zooarchaeological sites can identify early herd management practices as one step in the process of domestication (Zeder, 2016). Domestication is often defined as a sustained multigenerational relationship where one organism significantly influences the reproduction and care of another organism (Zeder, 2015). Such alterations can leave both genetic and morphological alterations. However, these alterations cannot be easily identified in early domestication settings. Early markers for domestication are often identified through alterations to the age and sex profiles, such as the selective harvesting of young males, where younger males are killed for consumption purposes, but the number of females and the number of older females remains in the herd. Alternatively, sex specific bones or the size of the bone are relied on to determine sex (Zeder, 2015).

Within this dissertation chapter I provide an innovative proteomic methodological approach to determine sex from zooarchaeological remains using teeth. Teeth preserve well and their enamel mineralization is among the hardest components in the human body, surviving well after decomposition of other tissue. As a result, the dental enamel provides a stable source for protein analysis. The protein amelogenin makes up 90% of the tooth enamel and is coded by the X and Y chromosomes in humans. The protein products of the X and Y gene have been recovered in human tooth enamel and used to determine sex with the absence of the Y protein. This proteomic method has been shown to be less destructive than DNA extraction from teeth, with a potential for less allelic dropout. I apply the same protein extraction method performed by Stewart and contributors to non-human mammals, where sex determination remains challenging due to the preservation and taphonomy of faunal assemblages as well as the lack of sexual dimorphism that can be observed between some species.

I conducted a blind experiment of acid etching tooth enamel and top-down mass spectrometry to identify the amelogenin protein in four mammalian species commonly found in North America, *Canis latrans* (coyote), *Canis lupus familiaris* (dog), *Odocoileus virginianus* (white-tailed deer), and *Castor canadensis* (beaver) from the Sam Noble Oklahoma Museum of Natural History. Our study evaluates this methodological approach in determining sex from non-human animal remains. I demonstrate the possibilities and challenges with the proteomic characterization of amelogenin and provide steps in how to recover sex determining peptides from the amelogenin protein for future analyses.

## 1.4 ANTHROPOGENIC EFFECTS ON MAMMALIAN ORAL MICROBIOMES

In addition to human alterations to domesticated animals, human alterations to the environment contribute significantly to microbial diversity and zoonotic disease transmission. In this final study, I take a novel approach to assess the scale of human impacts by examining oral microbiomes of California sea lions. For this data driven research, I examine the microbial diversity of oral microbiomes in California sea lions to use as a sentinel for documenting transitioning environments and California sea lion health within the Channel Islands.

The California Channel Islands have fostered a long, dynamic, and continuous relationship between humans and marine mammals. Located off the southern coast of California, archaeological and cultural resources extend through more than 12,000 years (Rick and Erlandson, 2008). While anthropogenic disruptions have increased with the onset of the Anthropocene, marine biome modifications and ecosystem alterations due to human activity can extend to the Pleistocene when ancient peoples hunted marine mammals (Braje and Rick, 2011). Historical exploitation of marine mammal fur, meat and oil during the 18$^{th}$ and 19$^{th}$ century consequently fostered a modern population heavily shaped by centuries of anthropogenic disruptions (Braje and Rick, 2011). We can explore the extent and scale of human impacts by using a historical ecological framework that integrates archaeology with biological datasets.

One method of understanding the long and continuous impact of anthropogenic disruptions is through studying the ancient microbiomes of marine mammals. Human and animal microbiome studies have shown environmental and social adaptations that influence the microbial diversity and abundance of phyla within their microbiomes (Trinh *et al.*, 2018). The oral microbiome is a reservoir for both primary and opportunistic pathogens (Warinner, 2016). Investigations of the mammalian oral microbiomes are underexplored within the zooarchaeological record. Mammalian microbiome studies have focused on the gut in conservation efforts or to observe the influence of antibiotics used as growth promoters (Allen and Stanton, 2014; Bahrndorff *et al.*, 2016). These studies have shown that captivity and the loss of wildtype food sources can strongly impact the prevalence of certain microbes (McKenzie *et al.*, 2017). Mammalian oral microbiome studies have observed human influence on wild vs captive and ancient microbiomes of primates and bears, to name a select few (Ottoni *et al.*, 2019; Brealey *et al.*, 2021; Fellows Yates *et al.*, 2021).

I studied ancient marine mammal oral microbiomes from museum collections using dental calculus. Dental calculus – mineralized plaque – is a layered fossilized record of preserved

microbial cells or other debris calcified in situ during the individual's lifespan (Warinner, 2016). The abundance of dental calculus can vary between human populations due to host genetics or dental hygiene practices (Austin *et al.*, 2019). The presence of dental calculus has been identified in marine mammal populations but currently it is unclear what factors contribute to its formation (Wellman *et al.*, 2020). Dental calculus held in museum collections provide a temporal perspective to address shifting microbial abundance and diversity.

Understanding the historical ecology of sea lions can help shed light on past human and marine mammal relations (Rick and Erlandson, 2008; Braje and Rick, 2011; Thurstan *et al.*, 2015; Hofman *et al.*, 2018; Radde, 2021; Scarborough *et al.*, 2022). To date, archaeologists have investigated and assessed these early human impacts on the environment within the Channel Islands by observing the absence or presence of pinnipeds in their abundance, biogeography and breeding behavior as a result of ancient peoples from southern California (Braje and Rick, 2011; Hofman *et al.*, 2018). With the addition of metagenomic analysis of dental calculus, we can recognize patterns of marine biome microbial change associated with human alterations of the demographics, distribution, microbiome, and ecology of marine mammals through past ~130 years.

## 1.5 CONCLUSION

This dissertation research offers a theoretical and multi -omic approach, uniting One Health with archaeology to study past human-animal-environmental interaction. Understanding the past human-animal-environmental interactions and the creation of disease-scapes has can address critical global questions of interest to a broad audience that stimulates new discussion and transdisciplinary research. I demonstrate the application of proteomic and genomic methods that are minimally destructive approaches and can be extrapolated and applied to other collections and specimens to address similar archaeological and zooarchaeological questions. Furthermore, the comprehensive genomic data generated using small samples in our approach will protect and conserve historic and archaeological materials that are important to for future study. The impact of this dissertation is threefold: increasing our understanding of human-made environmental impacts, increasing the recognition of natural history collections and biorepositories as valuable resources to address historical shifts in biodiversity and ecologies due to human activity, and the incorporation of microbiome data in broadening our understanding of human-animal interactions. Incorporating such analyses can open the door in developing datasets to identify unrecognized impacts.

# Chapter 2

## Uncovering the Holocene roots of contemporary disease-scapes: bringing archaeology into One Health

**Authors and affiliations:**
Kristen M. Rayfield[1,2,3], Alexis M. Mychajliw[1,2,3,4], Robin R. Singleton[1,2], Sabrina B. Sholts[3], Courtney A. Hofman[1,2,3]

1. Laboratories of Molecular Anthropology & Microbiome Research, University of Oklahoma
2. Department of Anthropology, University of Oklahoma
3. Department of Anthropology, National Museum of Natural History, Smithsonian Institution
4. Department of Biology & Environmental Studies, Middlebury College

**Abstract:**

The accelerating pace of global zoonotic disease events in the 21st century has motivated academic researchers, health professionals, and policy makers to craft integrative One Health approaches from disease ecology, environmental science, and epidemiology to detect and mitigate future outbreaks. Zoonotic disease events are frequently attributed to human-animal interactions through landscape alterations, industrialized food production, and wildlife trade. However, we demonstrate that anthropogenic effects on the environment, ecology, and biogeography of animals extends to the Late Pleistocene, beginning when ancient peoples first hunted megafauna and burned landscapes. The archaeological record – a long-term archive of human-animal-environment interactions – has largely been untapped in One Health research, thus limiting our understanding of how anthropogenic activities alter ecosystems and the disease dynamics within "disease-scapes" over time. Here, we highlight how new biomolecular tools and advances in the "-omics" can be coupled with paleoecological reconstructions to holistically evaluate past human impacts on environments and animal populations in the service of studying zoonotic disease emergence and reemergence. By combining zooarchaeological, bioarchaeological, and paleoecological data under a One Health perspective, we address how humans, as niche constructors, have facilitated new host and reservoir species and disease-scapes from the Late Pleistocene and Holocene directly into the Anthropocene.

## 2.1 INTRODUCTION

The profound physical and biological alteration of the global environment characteristic of the Great Acceleration, now marking the Anthropocene has effectively introduced new variables into zoonotic pathogen selection (Steffen *et al.*, 2015; McMahon, Morand and Gray, 2018). Over the past 40 years, the rise in emerging pathogens including HIV, H1N1, Ebola Virus and Coronaviruses has highlighted the need to better understand the emergence of zoonotic diseases – infectious diseases that spread between humans and non-human animals – as well as the

reemergence of familiar ones using a One Health approach (Destoumieux-Garzón *et al.*, 2018). As ~60% of human infectious diseases have a non-human origin, significant effort is now directed at predicting which species and populations are prime for emerging zoonotic diseases based on their phylogeny, ecology, environmental conditions, and interactions with humans (Carlson *et al.*, 2021). One Health provides an all-encompassing perspective of health determinants involving wildlife, domesticated animals, human practices, climate, ecological, and environmental factors (Destoumieux-Garzón *et al.*, 2018). While this perspective has been tractable and successful in addressing syndemics – where multiple diseases cluster together and influence public health – a temporal dimension is missing (Singer *et al.*, 2017). Archaeology provides this temporal dimension by recognizing the deep history of human influence in shaping diseases of the present and future but has not been employed in the service of predicting zoonotic disease emergence (Bendrey *et al.*, 2020). With the addition of archaeology as an integrative field of study and practice within One Health, we can highlight how disease-scapes are spillovers of ecological inheritance (Zeder, 2016) and that One Health is really a manifestation of niche construction where all three components (human, animal, and environmental health) and their synergistic activities reflect how humans have been iteratively shaping the syndemics of a population.

The present is a palimpsest of overwritten landscapes and species interactions that form the roots of present-day dynamics. These overwritten constructions can make it challenging to pinpoint how influential long-term human-animal interactions were in shaping the present-day disease ecology – a concept known in archaeology as niche construction theory (NCT). Drivers of disease emergence we see today – deforestation, climate change, species translocation, wildlife trade, as well as urban-adapted mammals harboring more pathogens (Shivaprakash *et al.*, 2021; Albery *et al.*, 2022) – are rooted in antiquity (Hofman *et al.*, 2015; Armstrong *et al.*, 2017; Ellis *et al.*, 2021). Continuous anthropogenic processes facilitate the passage of pathogenic microbes between humans and non-human animals within a shared space, whether through indirect (environmental – water, air, feces) or direct routes of transmission (butchering, consumption) (Barrett *et al.*, 1998; Harper and Armelagos, 2013). Therefore, zoonotic transmission is not a simple event but rather a complex cascade of events and factors that must come together at the same time and place to allow for favorable transmission of a pathogen to a new host species (Plowright *et al.*, 2017). Such a cascade requires the alignment of ecological, epidemiological, and behavioral determinants which foster increased pathogen pressure, the likelihood that the recipient host has adequate exposure to the pathogen, and the receptiveness of the new host (Plowright *et al.*, 2017). Thus, the resulting disease dynamics are not novel but are instead contingently built on successive rounds of anthropogenic disruptions.

Developing predictive models, anticipating outcomes, making them relevant across cultural and ecological contexts requires understanding history, place, and context. Models can assist in determining if spillovers of today recapitulate the initial spillovers of many diseases in the past and if there are parallels between the past and present for drivers in disease emergence. The archaeological record is an archive of the long-term human-animal-environment interactions that provides a temporal context for understanding disease dynamics in a particular place, culture, or ecosystem. Anthropogenically created disease landscapes, or "disease-scapes", are the result of the types of animals, plants, and microbes that are culturally and behaviorally selected within a constructed niche. Here we demonstrate how applying the lens of NCT to the One Health

framework provides a temporal perspective to the complexity of disease dynamics. We outline a technical toolset to extend One Health beyond the present to deeper time data by uniting zooarchaeological, bioarchaeological, and paleoecological data in addressing disease dynamics from the Late Pleistocene to the Anthropocene. In doing so, we address how archaeological/temporal datasets can contribute to documenting changes in environmental, animal, and human health and bring these together to show connectivity between these datasets.

We also address the relevance of the archaeological lens in viewing health disparities, as past pandemics disproportionately affect marginalized communities and create inequities that persist into the present (Dávalos *et al.*, 2020; Wade, 2020). We provide steps forward for interdisciplinary research and highlight ethical considerations for collaborative and inclusive research. Incorporating the archaeological record with One Health highlights the scale of human influence on shifting historical ecologies and overall fosters the development of datasets to identify unrecognized zoonotic pathways.

## 2.2 EPIDEMOLOGICAL TRANSITIONS AS MANIFESTATIONS OF HUMAN NICHE CONSTRUCTION

NCT provides a co-evolutionary framework and recognizes that organisms do not solely adapt to their environment, but rather reconstruct the environment around them which in turn creates or influences other natural selective pressures (Odling-Smee, 1988; Laland, Matthews and Feldman, 2016). It recognizes that the selective pressures created during niche construction have long-lasting effects on multiple taxa and result in greater evolutionary and ecological consequences (Odling-Smee, Laland and Feldman, 1996; McNally and Brown, 2015). Consequently, not only are selected genes passed from generation to generation, but also an altered ecological inheritance (Odling-Smee, 1988; Odling-Smee, Laland and Feldman, 1996). Through ecological inheritance, organisms create repetitive niche constructions that best fit the inherited genes from their ancestors, therefore the environment is re-imposed on future generations and becomes a source of selection (Odling-Smee, Laland and Feldman, 1996). Humans are considered the ultimate niche constructors given that they have greatly influenced their own natural selection and biological evolution by altering the environment for their exclusive benefit, which in turn has created and influenced other natural selective pressures (Smith, 2015; Boivin *et al.*, 2016). We can consider epidemiological transitions as manifestations of human niche construction through time where humans have driven disease dynamics and pathogen selection within their shared environment through niche creation, niche modification, and niche reduction (Figure 1). Three epidemiological transitions are recognized: 1) the rise of zoonotic diseases and souvenir species – coinciding with the adoption of agriculture and transition to a more sedentary lifestyle, 2) a shift from acute infectious diseases to chronic diseases – coinciding with industrialization and colonialism, 3) the emergence and re-emergence of infectious diseases due to antibiotic resistance and global travel (Omran, 1971; Harper and Armelagos, 2013).

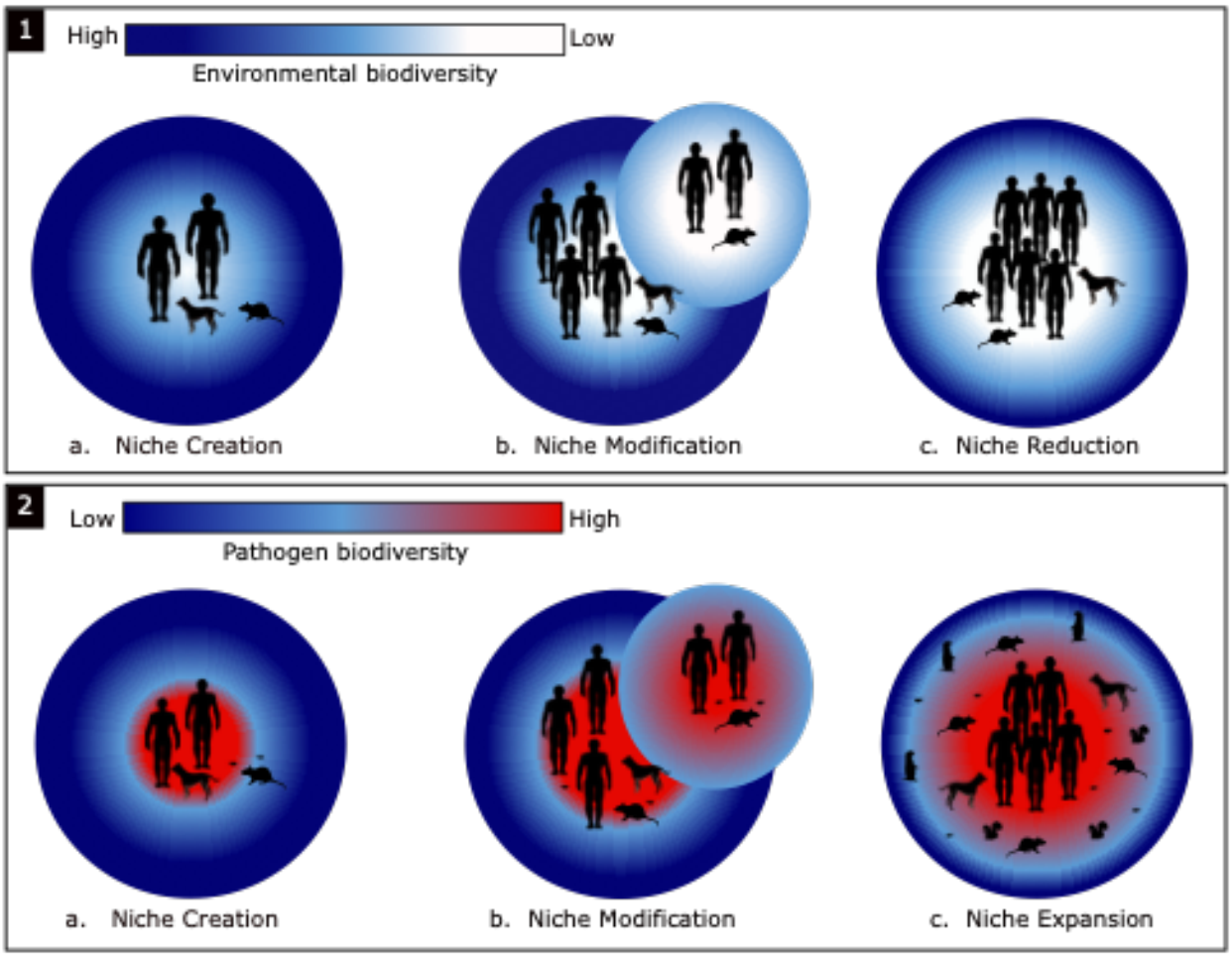



Figure 1: Human niche construction of disease-scapes. We can address anthropogenic impacts on disease dynamics through niche creation, niche modification and niche reduction (1). While ecological niches are further reduced, pathogen niches expand (2). During niche creation a subset of an area is altered (1a); consequently, introducing naïve host populations to new pathogens (2a). Through niche modification, specific plants and animal species are selected, resulting from either domestication, translocation, or extinction events (1b), thus resulting in an influx in zoonotic spillovers (2b). Within niche reduction, we continue to limit the biodiversity within the constructed niche (1c), yet pathogen biodiversity expands as new reservoir hosts are created and diseases become endemic (2c).

### 2.2.1 First epidemiological transition

The transition from the Pleistocene to the Holocene, was accompanied by major climatic shifts, the extinction of the majority of the world's mammalian megafauna, and human population size increase. Together, these factors promoted new developments in lithic technologies (new stone tools such as ground stone tools and pottery for food surplus and storage) and in subsistence strategies including – a transition from hunting and gathering to the adoption of agriculture and sedentism (Smith, 2015). While the transition to agriculture occurred in different regions independently and during different times (the earliest evidence in the Levant ~12,000 years ago

and the most recent in North America ~5,000 years ago), during this period humans start to intentionally interfere with the selective breeding and reproduction of certain traits within plants and animals – a process known as domestication – in addition to adopting monocrop strategies and a less diverse diet (Smith, 2015). Consequently, the resulting altered environments for agriculture and the domestication of animals created new ecological niches for the transmission of zoonotic and vector borne pathogens.

Anthropogenically created landscapes and cultivated land introduced humans to both vector and non-vector parasites through irrigation systems and feces used as fertilizers (Harper and Armelagos, 2013). New parasitic infections, such as blood born parasites, like malaria, take hold as souvenirs of these new behaviors in human-animal-environment interactions (Loy *et al.*, 2017). Agricultural burning, land clearing, terracing, and irrigation resulted in habitat fragmentation while also redefining human, domestic animal, and wildlife relationships. Landscape alterations fostered the opportunity for domesticated animals such as cattle, goats, sheep, pigs and fowl to become reservoirs for zoonotic diseases such as flu, tuberculosis and brucellosis (Pearce-Duvet, 2006; Fournié, Pfeiffer and Bendrey, 2017). Growing human and animal populations clustered in a shared space fostered crowd diseases and the frequent passing of some pathogens between animal and humans to evolve and become more virulent and specific to their new human reservoirs, such as measles and smallpox (Barrett *et al.*, 1998). As a result, the transition to agriculture and sedentism sparked the First Epidemiological Transition. This has been readily documented within the archaeological record by comparing the rates of nutritional deficiencies, oral pathologies, infectious disease, and an increase in co-morbidities and early mortality rates in early agrarian societies and Paleolithic hunter gatherer groups (Goodman *et al.*, 1984; Cohen, 1989; Armelagos, Barnes and Lin, 1996; Barrett *et al.*, 1998; Pearce-Duvet, 2006; Bocquet-Appel, Naji and Bandy, 2008; Armelagos, 2009; Harper and Armelagos, 2010, 2013; Zuckerman *et al.*, 2014). Ultimately, the adoption of agriculture changed disease ecology as humans created new routes and sources of infection associated with urbanization where landscape alterations, human waste, storage of food, and animal husbandry became prominent within densely populated communities.

### 2.2.2 Second epidemiological transition

Colonialism and industrialization brought new challenges as once created niches have now been modified into larger and more elaborate built environments to support growing and densely packed populations. Along with these challenges, increasing social and health disparities became a reflection of the modified environmental conditions both locally and globally due to colonization (Armelagos, 2009). Niche modification built over the last millennia established the culturally selected plants, animal species, as well as microbes resulting from either domestication, translocation, or extinction events. The consequence of niche modification can be defined by the Second Epidemiological Transition in a shift from infectious diseases to chronic, non-degenerative, non-infectious diseases (Armelagos, 2009; Zuckerman and Armelagos, 2014). While infectious diseases are still prevalent, many of them have become established with reoccurring outbreaks of smallpox, typhoid, diphtheria, measles, typhus, yellow fever, as well as tuberculosis and other respiratory diseases due to the harsh working environments, crowded living conditions, and Euro-colonial expansion (Armelagos, 2009). Medical intervention and later public health involvement alleviated some of these stresses and allowed for some

individuals to live longer, thus expanding the population size and shifting the age demographics to include a growing number of older individuals (Armelagos, 2009; Zuckerman and Armelagos, 2014). Within these now modified niches and established anthromes, there is a transition from infectious diseases to chronic disease such as cancer, diabetes, obesity, and diseases related to environmental pollution and metabolic diseases (Zuckerman and Armelagos, 2014). These heavily modified niches changed what microbes we allowed ourselves to be exposed to and fostered the rise in allergies and autoimmune diseases people face in industrialized nations, also known as the hygiene hypothesis.

### 2.2.3 Third epidemiological transition

The Third Epidemiological Transition has been associated with major ecological changes due to biodiversity loss as a result of niche reduction and has fostered the emergence of both novel and familiar diseases in different geographic areas as well as the re-emergence of now antibiotic resistant diseases (Omran, 1971; Barrett *et al.*, 1998; Zuckerman *et al.*, 2014). Within reduced niches, pathogens have not only been selected by the environment but also by human behavior and interference for the past millennia. This epidemiological transition is most often linked to changes in disease ecology today and major anthropogenic disruptions we attribute to disease outbreaks (climate change, geographic expansion of vectors, industrialized food production and food borne pathogens, as well as novel infectious disease). Finally, sanitized environments have resulted in microbiome shifts where individuals from industrialized nations have a less diverse microbiome (Obregon-Tito *et al.*, 2015; Segata, 2015). The lack of microbial diversity and microorganism exposure potentially impacts immune system function and may result in hypersensitivity and allergy development (Zuckerman and Armelagos, 2014). As the Third Epidemiological Transition continues to unfold, now is the time recognize that these transitions are a consequence of continued alterations to modified disease-scapes shaped by successions of reduced biodiversity and pathogen expansion.

### 2.2.4 Overlapping epidemiological transitions

Epidemiological models can help focus on anthropogenic disease ecologies within established anthromes, yet we are still left with large temporal gaps in highlighting region-specific disease ecology baselines. The spark of these epidemiological transitions is compared to archaeologically described “Paleolithic baselines” where population sizes and technologies are problematically viewed as roughly uniform globally and infectious diseases do not start to take a foothold until alterations to the environment and transitions within population demographics change. Consequently, these epidemiological transitions are often viewed as unilinear and uniform across the globe and do not incorporate how these transitions overlap nor incorporate socioeconomic situations that can contribute to a population simultaneously undergoing the first and third epidemiological transition or second and third epidemiological transition (Inhorn and Brown, 1990; Barrett *et al.*, 1998). For instance, the impacts of land appropriation and colonization translocated new diseases to non-immune populations with devastating consequences, as seen with the decimation of Indigenous people in the Americas from measles, smallpox and other diseases introduced via European colonization beginning in the 15$^{th}$ century (Larsen, 1994; Martin and Goodman, 2002; Uhl *et al.*, 2019) (Box 1). New global supply chain demands, mass mining, and the reliance on monoculture has led to global landscape changes that

are still felt today with simplified ecosystems, loss of predator species and the propagation of disease-carrying reservoirs and cosmopolitan species (i.e. mice, rats, bats) (Keesing *et al.*, 2010), as well as the rapid progression from outbreak to pandemics due to globalization (Chin *et al.*, 2020). As a result, the three epidemiological models, particularly the second and third transitions, have centered on westernized nation demographics, economic, and sociological determinants and their associated consequences that influenced health and disease (Omran, 1971). They have also fallen short in addressing the significance of race, gender, and class health disparities within the nation as a whole as well as globally where pandemics have disproportionately impacted marginalized communities (Inhorn and Brown, 1990; Armelagos, Barnes and Lin, 1996; Barrett *et al.*, 1998). We can start to tease apart these overlapping epidemiological transitions and disease dynamics (including exposure, transmission, and host switching) with the inclusion of zooarchaeological, bioarchaeological, and palaeoecological datasets under a One Health framework.

**Box 1: The origin and spread of Morbilliviruses**

Humans have driven disease dynamics and pathogen selection with the origin and spread of crowd diseases like morbilliviruses, which includes rinderpest (RPV), measles (MeV), and canine distemper virus (CDV) (Figure 2). RPV is a ruminant morbillivirus that was eradicated in 2011 (*FAO - News Article: Rinderpest eradicated - what next?*, no date). The origin of MeV has been associated with the early transmission of RPV from domesticated cattle in the Eurasia Steppe given that MeV and RPV are genetically similar (Uhl *et al.*, 2019). The spread of MeV from Europe to the Americas decimated Indigenous populations. This MeV outbreak may have resulted in the eventual spillover into local dogs, as the first recorded case of CDV is noted in Ecuador during this time (Uhl *et al.*, 2019). Since its transmission into domestic canines, the transmission of CDV into alternative hosts has rapidly and globally expanded. Outbreaks have been reported within marine mammals, thus resulting in Phocid distemper virus (PDV) and Cetacean morbillivirus (CeMV), as well as, in ferrets, tigers and lions, non-human primates, pandas, and badgers. Therefore, these ecological niches are a reflection of spillover disease-scapes (Zeder, 2016).

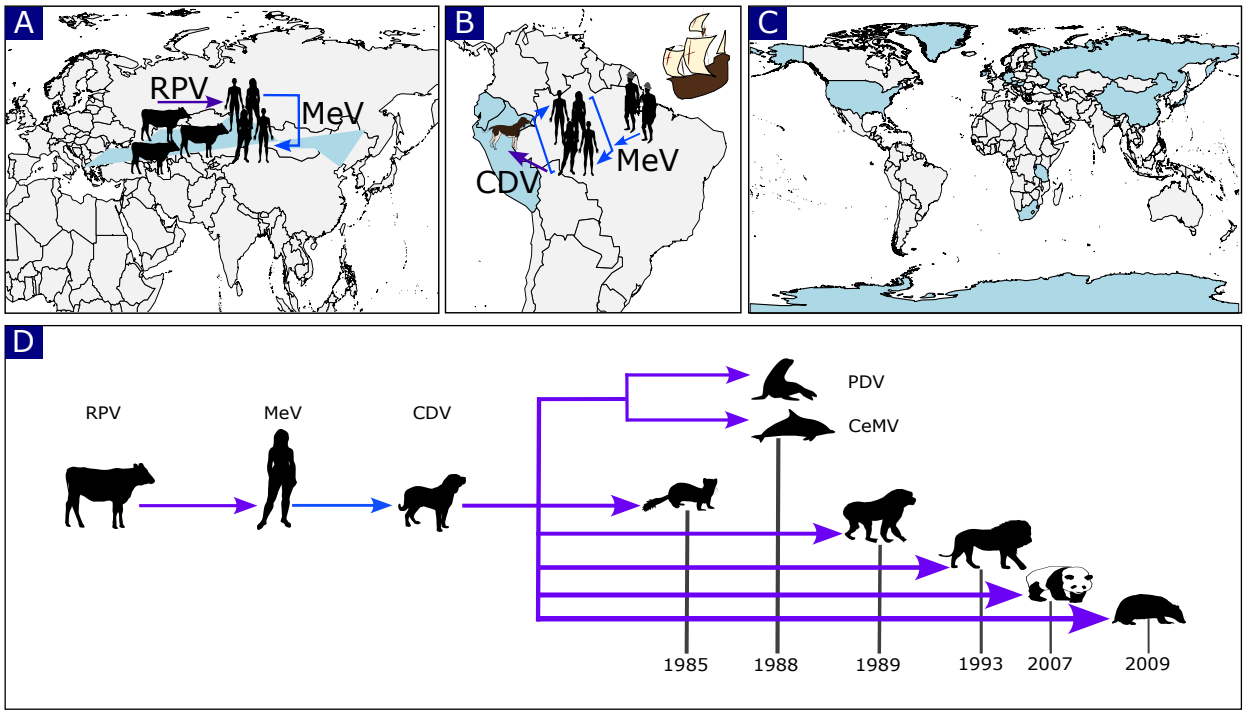



Figure 2: Origin and spread of morbilliviruses. A) transmission of rinderpest virus (RPV) to humans and the origin of measles virus (MeV). B) Transmission of MeV into the Americas and the first documentation of canine distemper virus (CDV). C) Current reported countries with CDV. D) Spread of morbilliviruses into recent alternative host species.

## 2.3 TOOLS AND DATASETS TO RECONSTRUCT PAST DISEASE-SCAPES

Examination of the archaeological record allows for the reconstruction of disease-scapes resulting from niche creation, modification, and reduction. Below, we provide a toolkit to explore how past diseases could have shaped local human history (Figure 3). We separately show how archaeological tools and datasets can contribute to documenting changes in environmental, animal, and human health in the past.

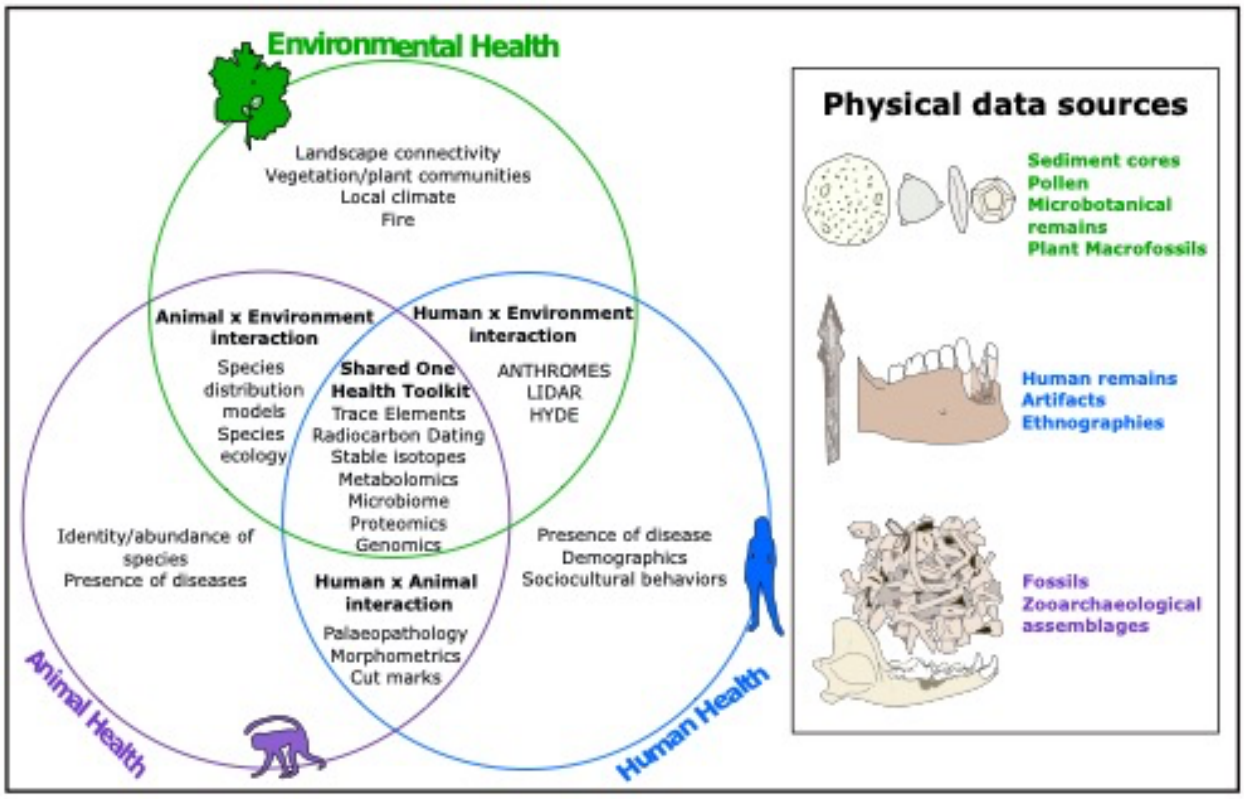



Figure 3: Tools and datasets to reconstruct past disease-scapes. The Venn diagram presents tools that can be implemented into a One Health framework to reconstruct past disease-scapes. The box contains physical data sources for each component – human, animal, environment – that can be studied.

### 2.3.1 Reconstructing past animal populations and health

Zooarchaeology is the study of animal remains (bones, dental remains, hair, feathers, eggshells, hides, crabs, mollusks, and shells) within archaeological contexts (Reitz and Shackley, 2012). Most often this involves the identification of faunal remains when looking at how humans interacted with animals through harvest, consumption, domestication, translocation, or extinction events. These research areas have been explored using tools such as morphometrics, Zooarchaeology by Mass Spectrometry (ZooMS), stable isotopes, as well as ancient genomics research (discussed below). Analysis of the zooarchaeological record permits us to reconstruct past species richness and abundance, which in turn can be used to evaluate hypotheses exploring how humans were interacting with and shaping the abundance of different animal species, and the purpose for interacting with these animals (i.e., food, dairying, ritual, clothing, shelter, tools, cultural material). Such research can provide insight into types of disease transmission routes (vector, butchery, consumption) that may have been present. We focus on specific tools that have been implemented within zooarchaeology to reconstruct past faunal assemblages; analysis of animal remains, morphometrics, paleopathology, ZooMS, as well as Species by Proteome Investigation (SPIN).

The majority of zooarchaeological faunal assemblages consist of fragmented skeletal remains. This can be challenging when trying to determine which species were present and how the animals were used. Usually, this is achieved by analyzing the minimum number of individuals (MNI) within the assemblage. Yet, it is an incomplete record of species present as both cultural and taphonomic filters, such as diagenetic processes and human alterations (butchering, modelling for tools or other objects that have be produced and used by humans) make taxa identification challenging (Buckley, Collins and Thomas-Oates, 2008). Taxonomic resolution is achieved by analyzing the differences in mass of the collagen peptide, which forms a unique fingerprint for different taxa (Buckley, Collins and Thomas-Oates, 2008). ZooMS has been useful in reconstructing past ecologies by looking at species that are infrequently observed in archaeological faunal assemblages (Richter *et al.*, 2022). It has provided taxonomic resolution for animal remains that have been culturally modified (tools, jewelry) (McGrath *et al.*, 2019; Trolle Jensen *et al.*, 2019), as well for animals that are difficult to distinguish through skeletal remains alone (sheep and goat) (Buckley, Collins and Thomas-Oates, 2008; Buckley and Collins, 2011). While these are major achievements in reconstructing past assemblages, the application of ZooMS is limited to taxa where there is enough divergence within the collagen protein. Therefore, canid species (dog, wolf, coyote) or bovids (cattle and bison), as well as equids (horse and donkey) would be difficult to identify at the species level using ZooMS alone. Alternative peptide biomarkers may resolve some of these taxonomic resolution issues using similar methods in equid species (Paladugu *et al.*, 2023). Archaeological context can provide further insight and help with classification where species identification cannot be achieved. A new proteomic method known as SPIN, which uses shotgun proteomics and Liquid Chromatography through Mass Spectrometry (LC-MS/MS) can provide further taxonomic resolution where ZooMS cannot (Rüther *et al.*, 2022).

The analysis of infectious disease within animal remains is an underexplored area (Thomas, 2019). Early medical reports of past animal disease include canine distemper, rabies, cowpox, tuberculosis, plague (*Yersinia pestis*), trichinonosis, to name a few (Pearce-Duvet, 2006; Uhl *et al.*, 2019). The study of past animal health has focused on butchering techniques, trauma, and dental pathologies. Nonetheless, incorporating the tools described to reconstruct the types of animals present, both wild and domestic, can reveal the species of animals that were around and may have contributed to alternative enzootic transmission events or early spillovers between humans and animals. This area has a lot of potential that can be incorporated with the addition of biomolecular technologies to provide further insight into past zoonotic diseases.

### 2.3.2 Reconstructing past human populations and health

We can broaden our toolkit to include techniques to study the bioarchaeological record– human skeletal remains within archaeological contexts – thus documenting disease transmission events. The analyses of the bioarchaeological record contribute to our understanding of how people were living in a constructed, modified, or reduced niche and how their behaviors within these niches have impacted past and future health problems. From this, the bioarchaeological record has provided insight into human origins (Hillson, 1996), human migration (Gregoricka, 2021), past human behavior including funerary and ritual practices (Roberts, 2009), identity and gender roles (Agarwal, 2012), diet (Brickley and Mays, 2019), demography (Roberts, 2009), and the impacts of these behavior on past human health (Grauer and Buikstra, 2019). This has mostly been done

through the analysis of skeletal remains, yet new insights and deeper timescales have been addressed with the incorporation of biomolecular techniques.

The majority of palaeopathology techniques have relied on identifying skeletal lesions or alterations to the bone that signify a bone growth or weakening due to stress or an immune response. This includes trauma (fractures, breaks) and nutritional stress which can lead to metabolic diseases such as rickets, scurvy osteoporosis, osteomalacia, and fluorosis and are often attributed to vitamin deficiencies along with poor diet (Brickley and Mays, 2019). These types of metabolic diseases are identified through demineralization of bone and bone formation alterations and are most often attributed to vitamin deficiencies along with poor diet. Bioarchaeological research on nutritional stress have focused on cribra orbitalia, porotic hyperostosis, stature, and oral pathologies as indicators of poor health and increased susceptibility (Bocquet-Appel, Naji and Bandy, 2008; Brickley and Mays, 2019; Grauer and Buikstra, 2019). For example, some studies focusing on nutritional deficiencies have shown that health did decline after the adoption of agriculture (Goodman *et al.*, 1984; Bocquet-Appel, Naji and Bandy, 2008; Stock and Pinhasi, 2011). Consequently, among the most vulnerable, maternal and infant stress increased, leading to decreased age at death and increased maternal mortality. However, other studies have shown that the frequency of nutritional deficiencies with agriculture are short lived (Stock and Pinhasi, 2011; Wells and Stock, 2020) and therefore health decline is regional specific (Stock and Pinhasi, 2011).

Although most people died of infectious diseases prior to the development of medical interventions, the focus of human skeletal analyses has mostly been on bacterial (i.e., tuberculosis, leprosy, brucellosis, treponemal diseases – particularly syphilis) and parasitic infections. This focus is partly for lack of bony responses to acute, severe infections that are typically observed with viral pathogens. Researchers are often limited in studying viral infections from skeletal remains alone, with the exception of poliomyelitis and variola osteomyelitis (Grauer and Roberts, 2019), especially as many of the bony lesions caused by infectious agents are non-specific indicators of disease or poor health. In addition, they are limited in studying ancient mycoses given that very few fungi species have the ability to alter human bone (Grauer and Roberts, 2019). Within the archaeological record, these are most often limited to studying systemic mycoses infection, such as sporotrichosis and aspergillosis, within a immunologically compromised host (Grauer and Roberts, 2019). CT scanning and radiographs may help resolve some of these issue by locating hidden lesions. Nonetheless, this type of osteological paradox (Wood *et al.*, 1992) can make it a challenge when recording the scale of infectious diseases in the past, both in morbidity and mortality within a population.

The Osteological Paradox questions many conclusions made from health and stress indicators of human skeletal remains to understand the overall health of prehistoric populations (Wood *et al.*, 1992). Of importance, conclusions drawn from skeletal remains can create population bias given that we are observing only a subset of a population, particularly when looking at cemeteries. Furthermore, of that subset, individuals with "sicker" looking bones or those that have a higher frequency of skeletal lesions are determined as representing weaker individuals in the population. Instead, individuals with skeletal lesions present should be viewed as survivors given that they are most likely in the later stage of the disease, and of importance, they have survived long enough for the disease progression to affect the bones, providing a biased sample set.

Therefore, individuals that may have been infected but succumbed to the infection, before skeletal lesion presence, are under-observed and or not accounted for within the bioarchaeological literature. Additionally, while nutritionally deficiencies can be recorded in the skeletal record, we are often limited in recording overt skeletal evidence of identifiable infectious diseases. Where we may begin to resolve the complexity of disease identification in the past is through isotopic and biomolecular techniques.

### 2.3.3 Reconstructing past environments

Environmental modifications change the types of organisms – animals, insect vectors, microbes – humans are now sharing a space with, therefore providing critical details when addressing coevolutionary disease relationships. The archaeological record has informed our understanding of past human-environment interactions and the behaviors adopted under times of environmental stress but can be further implemented in studying disease-scapes. One method that provides incremental snapshots of environmental changes, both natural and anthropogenic, is the analysis of sediment cores. Sediment cores taken from lakes, oceans, and terrestrial environments can reconstruct past vegetation through pollen and microbotanical analysis to identify climate change, landscape alterations and land use practices including the use of fire. Through the remnants of charcoal, some of the earliest evidence of landscape alteration comes from controlled burning and deforestation up to 45,000 years ago in Southeast Asia, Australia, and Papua New Guinea (Roberts *et al.*, 2017) and among Pre-Columbian settlements in the Amazon, known as *terras pretas* (Arroyo-Kalin, 2012; Roberts *et al.*, 2017). Such analyses have highlighted the longevity of human manipulation of tropical forest ecologies, especially when combined with macrobotanical remains (seeds, roots, woods, and fruits) and microbotanical remains (phytoliths, pollen, spores, and starch grains) in archaeological contexts (Reitz and Shackley, 2012). Botanical remains can also provide evidence for plant introduction (native, alien, and invasive species) and human usage in the past. The botanical remains found on tools or ceramics can reveal the types of plants humans were engaging with and how they were being used (i.e. medicinal, diet). Together, these datasets can reconstruct how plant communities changed over time.

Alternative methods used to reconstruct past population sizes, an important component for assessing environmental impact and change, include laser imaging detection and ranging (LiDAR) for archaeological site identification and size estimation, metabolomic datasets (White *et al.*, 2019) to estimate fecal accumulation from stanols, and summed probability distributions (SPDs) which uses the spatial distribution abundance of radiocarbon dated materials from archaeological sites as a proxy for determining past human population sizes (Robinson *et al.*, 2019; Crema and Shoda, 2021). With many open-source databases, archaeological data can now be modeled to explore regional and global population fluctuations and responses to environmental and climatic changes. These include sediment core databases – Neotoma (*NeotomaDB - Home Page*), International Paleofire Network (*Global Paleofire Database*) – radiocarbon databases – AustArch (Williams and Ulm, 2014), Radiocarbon Paleolithic Europe Database (*Radiocarbon Palaeolithic Europe Database v29, Feb 2022*), and p3k14c (Bird *et al.*, 2022) – in addition to environmental models – History Database of the Global Environment (HYDE 3.2) (Goldewijk *et al.*, 2017) and ANTHROMES (Ellis *et al.*, 2021). For instance, paleoenvironmental reconstructions using ANTHROMES data illustrated that more than 95% of

temperate and 90% of tropical woodlands were inhabited or cultivated as far back as 12,000 years ago, thus suggesting that the Anthropocene is advancing the alteration to already modified environments (Ellis *et al.*, 2021).

Overlapping these paleoenvironmental datasets with disease records can link alterations to possible disease emergence in the past. For instance, periods of drought and climatic shift, that can be identified through dendrochronology and sediment cores, are usually followed by famines and high mortality rates (Zuckerman and Dafoe, 2020). The presence of controlled fire and deforestation for agricultural and irrigation purposes signifies large population sizes, pooling water and the propagation of vector borne diseases, as observed with malaria (Harper and Armelagos, 2013). In addition, deforestation changes bat roosting patterns into more favorable habitats for various bat species, thus allowing a greater likelihood for bat-borne viruses to enter human populations (Afelt, Frutos and Devaux, 2018). As a result, paleoecological datasets and environmental reconstructions can elucidate some of the early baselines for current disease-scapes.

### 2.3.4 Tools to reconstruct past human-animal-environment interactions

The incorporation of geochemistry and new "-omic" methodologies has advanced our understanding of past human-animal-environment interactions within the same ecological niche. While there is a plethora of research on how each of these tools have been implemented in different scientific disciplines, we focus on tools that can be incorporated in reconstructing past disease-scapes: stable isotopes, trace elements and genomic, proteomic, and metabolomic methods.

#### *2.3.4.1 Geochemistry*

Chemical signatures representative of the environment to reconstruct past diets, life histories, migration, climates, and husbandry strategies can be studied through the geochemical techniques of stable isotopes and trace elements analysis. Stable isotope analyses of organic remains (plant, animals, and humans) provide quantitative perspectives of the life histories and resources consumed within an ecological niche (Reitz and Shackley, 2012). Most frequently stable isotopes, such as carbon and nitrogen, are observed to access dietary patterns and trophic levels. Strontium isotopes have reconstructed past migration, trade, and diet within humans and animals as they reflect the geochemistry of local water sources (Borić and Price, 2013). Newer methods such as compound specific stable isotopes focus on analysis of essential amino acids within bone collagen to capture information on nutrient sources and physiology, helping to alleviate some of the challenges that come with isotope analysis – such as shifting baselines and overlapping isotope values (Whiteman *et al.*, 2019). In addition to isotopes, trace elements, such as cadmium, lead, manganese, mercury, and zinc, have been studied to reconstruct trophic levels. Trace elements are needed for metabolic processes and can help address past health and disease. The abundance or lack of has been shown to impact the health and immune system of individuals in the present as well as in the past (Nelson, Halling and Buikstra, 2019).

#### *2.3.4.2 "Omics"*

New techniques in the "omics" have opened doors to study past interactions and how animals and plants have been used to make a range of items including, food, creams, medicine, art, etc. This has predominately been led by aDNA and archaeogenomics, where DNA has been recovered from bone, dental calculus, seeds, and paleofeces from animals, plants, and humans. Now environmental DNA (eDNA also known as sedaDNA) methods can reconstruct past ecologies where zooarchaeological and bioarchaeological data may be absent (Kjær *et al.*, 2022). With more standardized methods, targeted enrichment techniques, and high-throughput sequencing, ancient genome analysis has allowed us to explore human and animal biogeographies (Sugiyama *et al.*, 2022). It is also now possible to uncover past epidemiological events, such as the phylogeny and co-evolution of pathogens and humans. With deep sequencing and enrichment techniques, the reconstruction of entire genomes of ancient pathogens is feasible. This has been applied to *Yersinia pestis* (Bos *et al.*, 2011), *Mycobacterium tuberculosis* (Bos *et al.*, 2014), human herpes simplex virus 1 (HSV-1) (Guellil *et al.*, 2022), the 1918 influenza strain (Tumpey *et al.*, 2005), *Salmonella enterica* (Key *et al.*, 2020), as well as *Klebsiella pneumoniae* (Austin *et al.*, 2022), among others. Pathogen aDNA preserves best in the tooth root compared to bone, but dental calculus and paleofeces also provide insight into both host and microbial evolution (Spyrou *et al.*, 2019).

Where aDNA is limited due to preservation, other omic techniques such as paleoproteomics and metabolomics are starting to fill in the gaps, especially since proteins may preserve better than DNA (Welker *et al.*, 2015). Both targeted and metaproteomic techniques have been applied to reconstructing our own evolutionary history as well as determining how past animal byproducts have been used in tool manufacturing, art, and dairying (Welker *et al.*, 2020; Hendy, 2021). The application of metaproteomics allows researchers to investigate ancient diseases through host-pathogen interactions as well as explorations of the host immune response (Warinner *et al.*, 2014). While underexplored in past disease ecology, metabolomics has proven value in highlighting pathogen biomarkers (Tounta *et al.*, 2021). Within the archaeological record, metabolomics has mainly been used for the analysis of organic residues found within ceramics. Yet it can also be applied to recover gut microbiomes from paleofeces (Porru *et al.*, 2021). Further development of methods that target pathogen specific biomarkers within both fields will offer new insights into the pathogen evolution and epidemiology within and between human and animal populations.

### 2.3.5 Evaluating synergies: How did human-animal-environment interactions contribute to zoonotic spread? A focus on rats.

Through evaluating these synergies, we can understand how species introductions can also indirectly alter disease ecology. Humans have been translocating species for at least ~20,000 years (Grayson, 2001; Boivin *et al.*, 2016; Hofman and Rick, 2018). For example, small mammals were translocated through Melanesia as a reliable protein source (Grayson, 2001; Roberts *et al.*, 2017; Hofman and Rick, 2018). Three species of rodents – the black rat (*Rattus rattus*) the house mouse (*Mus musculus*) and the brown rat (*Rattus norvegicus*) are significantly more abundant in anthropogenically modified niches where they serve as commensal organisms. Their remains are often found in food storage pits and linked to agricultural transitions and food surplus (Cucchi, Vigne and Auffray, 2005; Hofman and Rick, 2018). These species of rodents have attained a global distribution due to human migration and have therefore been successful in

inhabiting areas outside their natural bounds with human assistance (Cucchi, Vigne and Auffray, 2005). They have been used as a proxy to study human migration (Matisoo-Smith, 2009; Puckett, Orton and Munshi-South, 2020) but can also serve as a proxy for associated vector borne diseases (Grayson, 2001; Matisoo-Smith, 2009). Rodents are reservoirs for many zoonotic diseases that can infect humans including, hantavirus, hemorrhagic fever, leptospirosis, plague and among others (Young *et al.*, 2017). They are characterized as hyper-reservoirs (Han *et al.*, 2015) and through human facilitated translocation, they have caused extinction events or created new disease reservoirs. This has been documented with the translocation of rodents (and their fleas) to create new plague reservoirs among North American ground squirrels and prairie dogs (Eads and Biggins, 2015) (Box 2) as well as the extinction of Christmas Island rat (*Rattus macleari)* with the introduction of a pathogenic trypanosome (Wyatt *et al.*, 2008). Consequently, the translocation of species, like rodents, has greatly impacted the epizootic transmission of diseases as well as facilitate alternative host species.

**Box 2: What can we learn from past pandemics: Plague Pandemics**

Many of the techniques described above to reconstruct past human-animal-environment interactions have revealed the long-lasting impact and ecological inheritance from Justinian Plague and Black Death pandemics. Through the analysis of documented Plague burial sites and extraction of aDNA from plague victims, Bos and colleagues (Bos *et al.*, 2011) were able to reconstruct the *Yersinia pestis* genome. In addition, research in rat ecology, immunological responses, environmental changes, and vector competence are further unravelling the role of host susceptibility and vector dynamics within a shared environment (Miarinjara *et al.*, 2021). Paleogenomic analysis of black rats populations reflect human migration and trade networks established during the Roman Empire (Yu *et al.*, 2022). The spatial distribution of rat populations may highlight the local vs. regional impact of the Black Death. Palynological records also support spatial heterogeneity in the impact of *Yersinia pestis* outbreaks (Izdebski, no date). The abundance of pollen in sediment cores can be used as a proxy for population size – the greater agricultural turnover, the larger the population – thus suggesting the Black Death was not a uniform pandemic environment as historical accounts suggest.

A key question in ancient pandemic research is what can we learn from past pandemics to predict and respond to future one? The Black Death was a large selective pressure on immunological genes that have lasting effects the on the susceptibility to chronic inflammatory and autoimmune diseases today (Klunk *et al.*, 2022). While a homozygous allele of the *ERAP2 gene* provided better protection against the Black Death (~40% more likely to survive), it has been shown to be a risk factor for Crohn's disease (Klunk *et al.*, 2022). Other factors that have long lasting impacts include social and economic inequalities (Dávalos *et al.*, 2020). Overtime, with fewer plague mortalities among wealthier individuals, plague became seen as a disease associated with the poor (Wade, 2020). As such, the health and social impacts of pandemics are long lasting, resulting in future health implications and further health disparities. Future pandemic responses will need to focus on identifying and mitigating environmental and social pressures that foster the longevity of pandemics at both regional and global fronts.

## 2.4 NEW AVENUES FOR TRANSDISCIPLINARY RESEARCH AND DATA SOURCES

### 2.4.1 Museum specimens as a source of data

Collaborative efforts between zoos, natural history museums, and biorepositories can provide data in genealogy, genetics, physiology, behavior, and life history that can contribute to understanding zoonotic disease host origins and transmission events (DiEuliis *et al.*, 2016; Poo *et al.*, 2022). Museum collections have provided valuable insight into the spread of antimicrobial resistance (Brealey *et al.*, 2021), the H1N1 influenza virus in 1918 (Colella *et al.*, 2021), the epidemiology of chytrid (Schmitt *et al.*, 2019), Lyme disease (Marshall *et al.*, 1994), White Nose Syndrome (Campana *et al.*, 2017), and Sin Nombre (Dunnum *et al.*, 2017), among others. In addition, they offer a snapshot of biodiversity, spanning over important time gaps between archaeological and modern outbreaks. Virtual communities such as Project ECHO's Museums and Emerging Pathogens in the Americas (MEPA) are transforming how collaborative research using biorepositories and specimen vouchers along with field collections can add value in predictive outbreak models (Colella *et al.*, 2021; Thompson *et al.*, 2021). These collaborations and the use of either biorepository for zoonotic diseases remains underutilized but can provide meaningful and temporal insight towards understanding disease dynamics while costing less than reconducting fieldwork for One Health.

### 2.4.2 Applications to conservation problems

Zoonotic disease transmission and epizootic outbreaks remain major threats for conservation efforts. The contemporary spread of chytrid and white-nose syndrome attest to the detrimental impact of infectious diseases within conservation biology. Habitat destruction, globalization, and the expanding human-wildlife interface is a concern for not only the transmission of infectious diseases from animals to humans but also from humans to animals. The archaeological record can fill important gaps to access how human disruptions to past species populations may have impacted their susceptibility to diseases today. While zoonotic disease transmission events are most frequently recognized as being unidirectional from animals to humans, humans can similarly transmit pathogens and alter their environment in ways that can influence disease ecologies and etiologies (Messenger, Barnes and Gray, 2014; Fagre *et al.*, 2022). The transmission of pathogens from humans to multiple species has been referred to as reverse zoonosis or zooanthroponosis (Hubálek, 2003; Messenger, Barnes and Gray, 2014; Fernandes *et al.*, 2018). Some zooanthroponsoses studies have been reported on *Pseudomonas aeruginosa* (Fernandes *et al.*, 2018), *Mycobacterium tuberculosis* (Fagre *et al.*, 2022), influenza strains, *Giardia duodenalis*, and *Crytopsporidium parvum* (Messenger, Barnes and Gray, 2014) and now SARS CoV-2 (Clayton *et al.*, 2022). Overlooking this bidirectional flow complicates our understanding of spillovers and transmission events, leaving major gaps in interpreting disease dynamics for conservation concern. Yet, while the integration of diverse datasets across different timescales can be challenging, especially with taphonomic filters and statistical analyses, developing regional and local specific models along with standardized data collection may help alleviate these challenges (Rick and Lockwood, 2013; Zipkin *et al.*, 2021).

## 2.5 ETHICAL CONSIDERATIONS

While every field of research in this transdisciplinary One Health perspective has their own guidelines and standards for ethical research, we highlight two crosscutting ethical concerns for

marginalized and underrepresented communities: 1) collaborative and inclusive research through community engagement and 2) data sovereignty. Each of these ethical concerns are ongoing discussions (Claw *et al.*, 2018; Carroll *et al.*, 2020; Wagner *et al.*, 2020), however we highlight some ways they have been implemented for a more inclusive and decolonized interpretation of the past.

Collaborative and inclusive research begins with community engagement. Consulting with descendant communities (broadly defined) to co-develop research agendas, identify how the community will benefit and make decisions on research methods, destructive sampling, data storage and availability at the beginning of a project is a powerful approach to address the historical inequities in research (Wagner *et al.*, 2020; Tsosie *et al.*, 2021). Co-interpretation of data can provide accurate cultural and behavioral contexts and avoid inappropriate language use. This has been highlighted within 'omics' research, where indigenous communities are often relied on as a baseline of "non-westernized" vs "westernized" and are frequently described as 'non-modern', 'non-industrialized' or 'traditional' for comparative studies (Mangola *et al.*, 2022). Such terminology relays to the broader audience that the community is somehow stagnant both culturally and genetically while progressing bias and inaccurate narratives (Mangola *et al.*, 2022). The incorporation of ethnographic ethnohistoric records, oral traditions, and languages in conjunction with archaeology and phylogeography have provided valuable insight where archaeological and phylogeographic evidence is lacking and have documented the longevity of zoonotic disease dynamics (Hawkins, O'Connor and Kealy, 2016; Turner, Geralda Armstrong and Lepofsky, 2021). Incorporating community participation into the One Health study design can strengthen our understanding of how humans are niche constructors of emerging and re-emerging infectious diseases.

Finally, data sovereignty is becoming increasingly important, especially for Indigenous communities (Tsosie *et al.*, 2021). Discussions of who has access to data, for long and for what purpose at the onset of research can reduce conflict and harm. In addition, the researcher and community should discuss if future research can be pursued from the generated data as the misuse of data outside the scope of the project can have long lasting and negative effects (Carroll *et al.*, 2020). Indigenous consortiums and open-source platforms such as The Native BioData Consortium ('About Us', no date) and Mukurtu (*Home*, no date) are paving new paths for collaborative and ethical research. While these bioethics are becoming more prominent within archaeology, inclusive research needs to be further implemented in zoonotic outbreaks. This includes the sharing of zoonotic risk technologies and diagnostics with communities in higher risk areas of spillover events to occur.

## 2.6 CONCLUSIONS

While some zoonotic pathogens may not be able to transmit to multiple hosts due to host plasticity, zoonotic diseases that can infect ecologically and taxonomically diverse host range have proven to have devastating impacts. Now, with our accelerating population sizes, rapid global travel, encroaching wildlife habitats, antibiotic resistance, and increasing social inequality, the biggest challenge for human health is preventing devasting overlapping pandemics. Predicting emerging diseases, therefore, cannot simply begin with predicting likely diseases to

infect human populations, but rather focus on diseases that will likely infect secondary hosts and then transmit to humans as well as diseases that humans are likely to transmit to animals.

This review addresses how these complex challenges in studying disease-scapes can be addressed with long-term data. The major shifts in human behavior and alterations to the environment that we consider as instigators under our One Health framework today, have their roots in the past. Disease-scapes, therefore, are spillovers of ecological inheritance where humans have created, modified, and reduced disease-scapes for the past millennia. Both human and animal demographic and behavioral responses to ecological niches have been shaped over centuries of anthropogenic disruptions. By incorporating paleoecological, zooarchaeological, and bioarchaeological tools we can begin to untangle past human-animal-environment interactions and pathogen transmission over the past millennia and start to study past disease-scapes as models for studying future emerging infectious disease events.

**Acknowledgements**

We would like to thank the Smithsonian National Museum of Natural History Associate Director of Research for funding Kristen Rayfield's Graduate Research Fellowship as well as the University of Oklahoma VPRP for their support. We would also like to thank Torben Rick for his guidance.

# Chapter 3

## A Proteomic Method to Determine Sex from Zooarchaeological Dental Remains

**Authors and affiliations:**
K.M. Rayfield[1,2], Lushuang Huang[3], Yanting Guo[3], Si Wu[3], Courtney Hofman[1,2]

[1]Laboratories of Molecular Anthropology and Microbiome Research, The University of Oklahoma
[2]Department of Anthropology, The University of Oklahoma
[3]Department of Chemistry and Biochemistry, The University of Oklahoma

**Keywords:** ancient proteins, amelogenin, tooth enamel, zoology

**Abstract:**

Sex determination from animal skeletal remains challenging, often relying on sex specific bones, osteometrics, or DNA. Determining sex is beneficial in understanding animal husbandry practices, as well as human-animal interactions. Building on previous work with humans, we explore the potential of a proteomic approach for determining sex from tooth enamel in non-human mammals. This proteomic method is less destructive than DNA extraction from teeth and cost-effective. The protein amelogenin, which makes up to 90% of the tooth enamel, is dimorphic and coded by the X and Y chromosomes in humans. To evaluate this approach in non-human mammals, four mammalian species (dogs, coyotes, beavers, and deer) were tested in a blind study from research collections in the Sam Noble Oklahoma Museum of Natural History. Mass Spectrometry-based bottom-up proteomics was applied to identify the amelogenin proteins. Here, we present the findings of this experiment and further characterize species specific amelogenin proteins. We show that the identification of sex may be possible with some taxa, yet the lack of reference peptide sequences for the AMELX and AMELY protein limits our analyses. Our research demonstrates the possibilities and challenges with how the proteomic characterization of the amelogenin can be broadly applied when determining sex from animal remains.

### 3.1 INTRODUCTION

Over the last few decades, the field of palaeoproteomics has grown exponentially. Now with high-sensitivity mass spectrometry, researchers have been able to recover ancient proteomes – the mixture of proteins produced by organisms (Cappellini *et al.*, 2018; Hendy, 2021). Within archaeology, the applications of ancient protein research can be broadly applied to organic materials, reconstructing subsistence patterns, diet, and past human diseases (Hendy, 2021; Warinner, Korzow Richter and Collins, 2022). Recent advances in palaeoproteomic research have documented that proteins generally preserve better than DNA and are impacted less by diagenesis, particularly in warm wet climates which greatly affects the preservation and degradation of DNA (Cappellini, Collins and Gilbert, 2014; Cappellini *et al.*, 2018; Welker *et*

*al.*, 2020). As such, palaeoproteomics has been applied as a complementary and sometimes alternative method to ancient DNA, as it is generally, more reliable, less expensive, and less destructive (Cappellini *et al.*, 2018; Hendy *et al.*, 2018; Hendy, 2021). Most paleoproteomic studies have focused on bones and teeth where the collagen type 1 (COL1) protein is most abundant within bone and dentine proteomes and can be used for species determination and phylogeny (Buckley *et al.*, 2009; Cappellini *et al.*, 2012).

Teeth preserve well in the archaeological record and their enamel mineralization is among the hardest components in the human body, surviving well after decomposition of other tissue (Welker et al., 2020). Unlike bone and dentine proteomes, the enamel proteome is comprised of ten or fewer proteins (Cappellini *et al.*, 2018) and has shown to be stable within both archaeological and human evolutionary time scales (Parker *et al.*, 2019; Buonasera *et al.*, 2020; Welker *et al.*, 2020). As a result, the dental enamel is exceedingly resistant to diagenetic changes (Hillson, 1996) and provides a stable source for protein analysis. Furthermore, the recovery of enamel proteins requires a simple acid etching with HCl and has shown to be minimally destructive compared to other techniques. The dental enamel matrix is comprised of the amelogenin, ameloblastin, and the enamelin proteins (Shaffer *et al.*, 2015) with ~ 90% of dental enamel crystalline matrix composed of the protein amelogenin (Bansal *et al.*, 2012; Stewart *et al.*, 2017). Amelogenin proteins are responsible for the enamel structure by organizing the enamel rods during dental development. During enamel formation, these proteins are cleaved into peptides and left *in situ* (Buonasera *et al.*, 2020). Of importance, in some species, the amelogenin protein is dimorphic and coded by the X and Y chromosomes, allowing for sex determination applications with the presence or absence of the AMELY protein.

Taking advantage of this dimorphism, Stewart and contributors extracted the amelogenin protein from dental enamel in archaeological human remains for sex determination (2017). Similar proteomic approaches have been further applied to determine sex from early hominin and human dental remains with success (Stewart *et al.*, 2017; Lugli *et al.*, 2019, 2020; Parker *et al.*, 2019; Wasinger *et al.*, 2019; Buonasera *et al.*, 2020; Froment *et al.*, 2020; Haas *et al.*, 2020; Welker *et al.*, 2020). In addition, the AMELX protein has been recovered from non-human animals and hominids (Cappellini *et al.*, 2019; Welker *et al.*, 2019; Nogueira *et al.*, 2021).
Sex determination from human skeletal remains is fundamental to forensics and archaeology, and zooarchaeology. While the robustness of this approach has been debated among human remains (Parker *et al.*, 2021; Štamfelj, 2021), the application of this method can easily be extended toward zooarchaeological remains, where other mammalian taxa have differences between the AMELX and AMELY gene sequences. We expand the applications of these proteomic methods to zooarchaeological remains to determine if these methods can be applied to and yield similar results from other mammalian taxa.

## 3.2 BACKGROUND

Sex determination among non-human animals is important for estimating the sex ratio in a wild-living community as well as for population management. Current sex determination methods within zooarchaeology have heavily relied on osteometrics (Davis *et al.*, 2012; Popkin *et al.*, 2012; Gifford-Gonzalez, 2018; Brassard and Callou, 2020; Wolfhagen, 2020), and while these methods have proven value they can be inaccurate. Additionally, osteometric methods often rely

on museum collections for comparative reference specimens, but there are well-documented sex biases in these collections; more males than females (Cooper *et al.*, 2019; Gower *et al.*, 2019). Where osteometric analysis is lacking, genetic sexing analysis has provided insight (Sinding *et al.*, 2016; Brassard and Callou, 2020; Bro-Jørgensen *et al.*, 2021). Within wild populations, the development of sensitive techniques to determine sex from highly degraded samples such as feces and hair is valuable when observing species populations that are difficult to capture (Hrovatin and Kunej, 2018). Molecular sexing methods have started to replace nonmolecular sexing methods as these techniques become faster, less expensive, and more sensitive. Previous molecular techniques have focused on the SRY, AMEL, and Zinc Finger Protein to determine sex among animals (Lu *et al.*, 2007; Svensson, Götherström and Vretemark, 2008; Davis *et al.*, 2012). In particular, the SRY gene has been used to determine sex in both wild and domesticated animals. Of these, the AMEL and ZFX and its homologue ZFY have highly conserved regions across mammalian species. In particular, genetic sexing with the amelogenin gene, which differs in size and nucleotide sequence between males and females in many species, is commonly used to sex mammals (Pfeiffer and Brenig, 2005; Slavkin, 1997; Shaffer et al., 2015). Alternative sexing studies have used high-throughput shotgun sequencing and assigned sex by the proportion of reads mapping to the X and Y chromosomes and autosomes (Nistelberger *et al.*, 2019; Bro-Jørgensen *et al.*, 2021). Yet genetic sexing remains expensive and has its challenges where misidentification is still possible as the absence of the Y signal could be the result of allelic dropout, especially with degraded or low quantity DNA (Faze et al., 2014).

Within zooarchaeological contexts, the ability to determine sex among animal remains is critical for addressing hunting strategies or animal husbandry and domestication processes, such as, how were people influencing the reproduction of animals. The acceleration of domestication began during the early Holocene as animals became more entangled within human constructed niches (Larson and Fuller, 2014; Larson *et al.*, 2014; Zeder, 2015, 2016). However, the when and why this process occurred is continuously debated (Cucchi and Arbuckle, 2021), as humans have selectively hunted, managed, translocated, and altered animal habitats for stable food resources without morphological and genetic domestication traits (Willcox, 2012; Cucchi and Arbuckle, 2021). Domestication is often defined as a sustained multigenerational relationship where one organism significantly influences the reproduction and care of another organism (Zeder 2012). Although this relationship may not always be so heavily and intentionally influenced by humans (Vigne, 2011; Larson and Fuller, 2014; Larson *et al.*, 2014). Zeder offers three alternative pathways, merging biological and social dynamics, that animals followed towards domestication – commensal, prey and directed (2012). Exploring each pathway provides the scale of human influence and intent in the selective process of domestication, noting that some animals may have become domesticated through an axiomatic coevolutionary process as they were attracted to human niches given the accessible food resources (Larson and Fuller, 2014). Other pathways may have been for strategic hunting and increasing the supply of vanishing resources. This is most often explored by measuring the size, sex ratios, and mortality profiles within zooarchaeolgical assemblages. Domestication processes have often been analyzed by observing both genetic and morphological alterations. However, these alterations cannot be easily identified in early domestication instances. Early markers for domestication have therefore relied upon alterations to the age and sex profiles, such as the selective harvesting of young males (Zeder and Hesse, 2000; Larson and Fuller, 2014). Yet, many early animal domestication processes remain archaeologically and genetically poorly documented (Larson and Fuller, 2014).

Using alternative methods for sex identification may help untangle the early herd management, hunting, and domestication strategies. Building on previous work with humans, I present a targeted proteomic approach, following the Stewart *et al*. 2017 method, for determining sex from tooth enamel in non-human mammals. The species chosen span over different orders of mammals with applicability for zooarchaeological questions. Following mass spectrometry analyses, we reveal the possibilities and challenges when applying this method to determine sex on various animal species.

## 3.3 MATERIALS AND METHODS

The samples for this study were specimens housed at the Sam Noble Museum of Natural History Mammals collection. A single blind study was conducted in which the museum curators knew the sex of the animals remains, but the laboratory team of this study did not. The sex of these specimens was known from perimortem morphology during acquisition by the Sam Noble Museum. Four different mammalian species commonly found in North America were used for this study – *Canis lupus familiaris* (dogs), *Canis latrans* (coyotes), *Odocoileus virginianus* (white-tailed deer), and *Castor canadensis* (North American beavers). These species were selected for their varying dental morphology and the presence of both sexes in the collection. Up to five females and five males were chosen from each species. A total of 36 dental samples were used for this study – 10 coyotes, 10 deer, 10 beaver, and 6 dogs. The mandibular teeth were selected for sexing for each animal – the right canine for dogs and coyotes, the right medial incisor for beavers, and the right first premolar for deer (Table 1).

Following the Stewart *et al.,* 2018 protocol, all teeth were washed with 3% hydrogen peroxide and rinsed with ultrapure water. The tooth sample was then lowered into a small volume of 5% Hydrochloric acid (HCl) two times, discarding the first etch and retaining the second etch. This acid etch is used to demineralize the enamel and extract the amelogenin peptides. Only one side of the tooth enamel surface was used during the acid etching process. As an environmental control, we placed 60µl of HCl in the bottom of a low bind protein tube and left the cap open during the entire sampling process to determine if there were any free floating contaminates in the prep laboratory. All dental washing and acid etching were performed in the mammal preparation laboratory at the Sam Noble Museum. Protein extractions were performed in the Laboratories of Molecular Anthropology and Microbiome Research ancient DNA laboratory. This process required the transfer of the 2nd etched solution to a conditioned ZipTip which was then washed six times with 0.1% (vol/vol) formic acid. The resin-bound peptides were then eluted into an elution buffer. The same protocol steps were performed on an extraction negative control, to which no protein was added, along with the previously obtained environmental control, to monitor potential contamination. All samples were then lyophilized for downstream Mass Spectrometry Analysis. The proteins were resuspended in 20 µl of buffer A (0.1% formic acid in HPLC grade water). 8 µl of each sample was then injected onto the Accelaim PepMap C18 column (15 cm × 75 µm) column LC-MS/MS analysis (Figure 4).

Table 1: Summary of Specimens in this study

| Species | Collection Date | Tooth Sampled | Males/Females |
|---|---|---|---|
| *Canis latrans* | 1980-2002 | Right $C_1$ | 5/5 |
| *Canis lupus familiaris* | 1982-1989 | Right $C_1$ | 4/1 |
| *Odocoileus virginianus* | 1982-1989 | Right $PM_1$ | 5/5 |
| *Castor canadensis* | 1987-2010 | Right $I_1$ | 5/5 |

### 3.3.1 Mass spectrometer parameters

All samples were analyzed using the LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Hanover Park, IL, USA) with the following parameters: the heated capillary temperature was 300°C with a spray voltage of 3 kV. MS scans were obtained with a normal scan rate and the m/z range was 350-1350. MS/MS scans were acquired with collisional induced dissociation (CID) at a normalized collision energy setting of 35%. The ten most abundant precursor ions were selected for MS/MS. The AGC for MS scans was 5E6 and 200 ms was the maximum injection time with 2 microscans; while the AGC for MS/MS scans was 1E6 and the maximum ion injection time was 50 ms with 3 microscans. Importantly, a buffer blank was injected and run with the same LC method to wash the column between sample runs. 60 s was used as the dynamic exclusion.

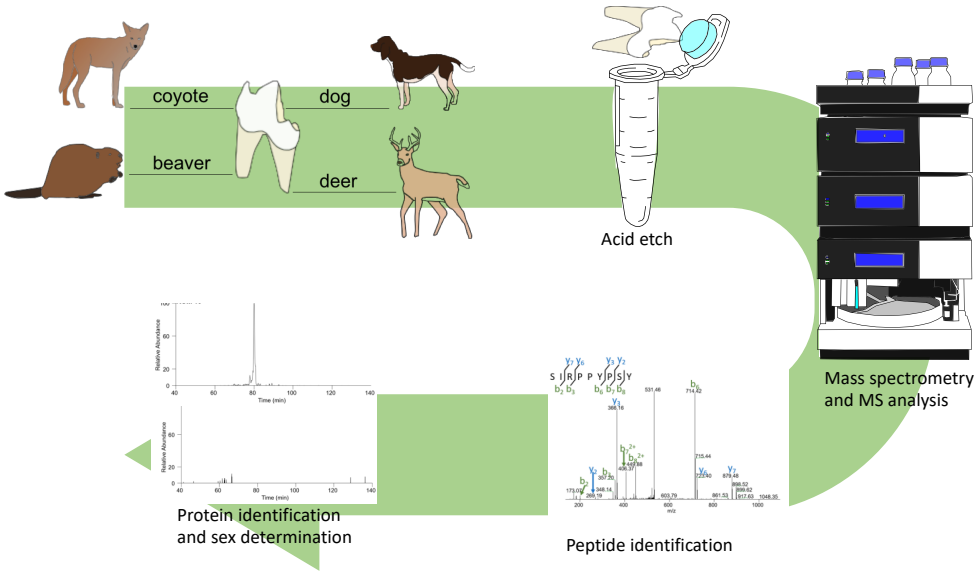



Figure 4: Protein extraction workflow. Acid etching of the tooth enamel and protein extraction are followed by bottom-up LC-MS/MS Mass spectrometry. AMELX and AMELY are determined by the difference in mass and peptide sequence.

### 3.3.2 *In silico* analysis

Within humans, slight variations in peptide length and dimorphic differences have been observed with the presence of an additional methionine in the AMELY peptide sequence. To determine if unique peptides could be identified between the AMELX and AMELY protein across different

taxa, AMELX and AMELY protein sequences were obtained from UniProt for *Canis lupus familiaris* (F6PLF2 – AMELX; A0A0F6T235 – AMELY), *Odocoileus virginianus* (AMELX only – A0A6J0Z0E7), and *Castor canadensis* (AMELX only – A0A2Z2ECS5). Unfortunately, there were no AMELY protein sequences for *Odocoileus virginianus* and *Castor canadensis* available on UniProt nor NCBI GenBank at the time of this analyses. In addition, there were no available AMELX and AMELY protein sequences nor a reference proteome for *Canis latrans*. To resolve this issue, we used the *Canis lupus familiaris* AMELX and AMELY as well as the *Canis lupus* (wolf) AMELX (Q8SPH3) as a comparative reference. For *Odocoileus virginianus,* we downloaded the AMELX and AMELY protein sequences for *Cervus elaphus* (European red deer) (A0A059V6J8 – AMELX; A0A059V6J9 – AMELY) as a comparative reference sequence. We did not find an available AMELY sequence for closely related taxa to *Castor canadensis*. Using the available amelogenin protein sequences, we performed a protein-to-protein BLAST search on NCBI (September 07, 2020) using the non-redundant protein sequences database to identify additional mammalian amelogenin protein sequences. The downloaded amelogenin protein sequences were aligned in Geneious Prime (v. 2022.2.1) using Blossom 62 (Figure 5).

Within this BLAST search only AMELX protein sequences were recovered for rodent species. Further investigation in the literature revealed that the amelogenin protein is only found on the X chromosome within mice and perhaps other rodents (Lau *et al.*, 1989). Therefore, it may not be possible to determine sex for beavers using the amleogenin protein as the AMELY may have been lost within rodent species (Iwase *et al.*, 2007). Our prior BLAST search for alternative amelogenin proteins also did not return AMELY protein sequences belonging to alternative rodent species, therefore making our database to identify AMELY for beavers biased toward AMELX identification. This could be a result of the lack of male rodent species within global databases such as NCBI or UniProt, or that the amelogenin protein is only located on the X chromosome within in rodents.

We aligned the AMELX and AMELY *Cervus elaphus* and *Odocoileus virginianus* amelogenin protein sequences in Geneious Prime (v. 2022.2.1) using Blossom 62 and did the same for *Canis lupus familiaris* AMELX and AMELY and *Canis lupus* (AMELX) protein sequences. We observed several differences between the AMELX and AMELY peptide sequences for *Canis lupus familiaris* with the addition of a Methionine among the AMELX peptides sequences and several substitutions (Figure 6A). There were no differences in the peptide sequences between the *Cervus elaphus* and *Odocoileus virginianus* AMELX protein, besides the length of the peptide sequence and one substitution from Threonine to Proline (Figure 6B). The available AMELX sequences for *Cervus elaphus* are fragmented. However, we did observe a difference between the *Cervus elaphus* AMELX and AMELY and *Odocoileus virginianus* AMELX peptide sequences with an indel at the 139-152 region. Since we did not observe overt differences among the AMELX for canids and cervids across 210 amino acids, we believed that the available protein sequences for AMELX and AMELY could be used to determine sex for *Odocoileus virginianus* and *Canis latrans.*

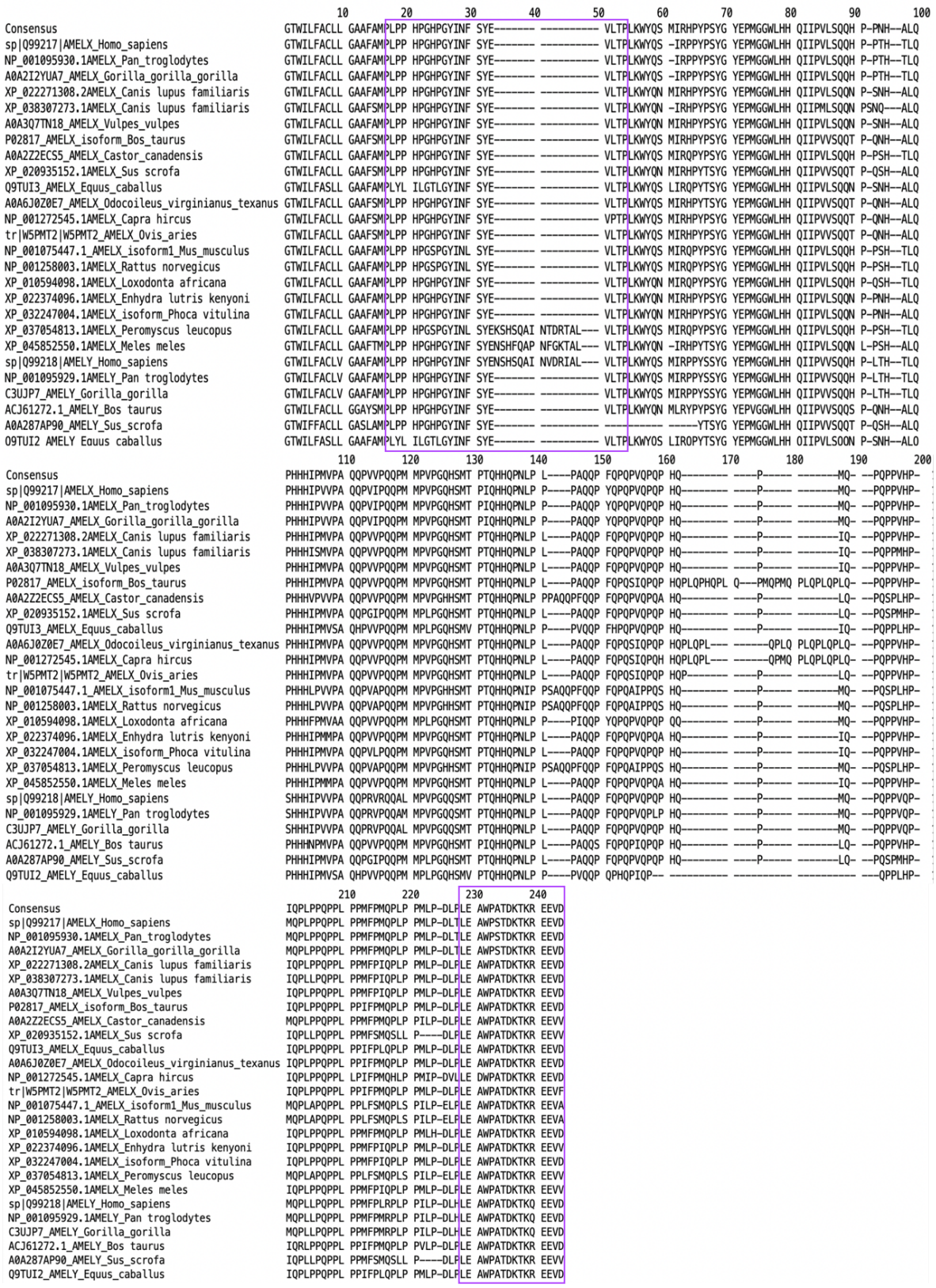



Figure 5: Mammalian AMELX and AMELY. Mammalian AMELX and AMELY sequences from six mammalian orders (Carnivora, Artiodactyla, Perissodactyla, Prosboscidea, Primate, and

Rodentia). Protein sequences have been trimmed and modified in Geneious to correct for incorrect placements of amino acids or gaps. Numbers represent location in amino acid sequence. Purple box represents peptides routinely recovered.

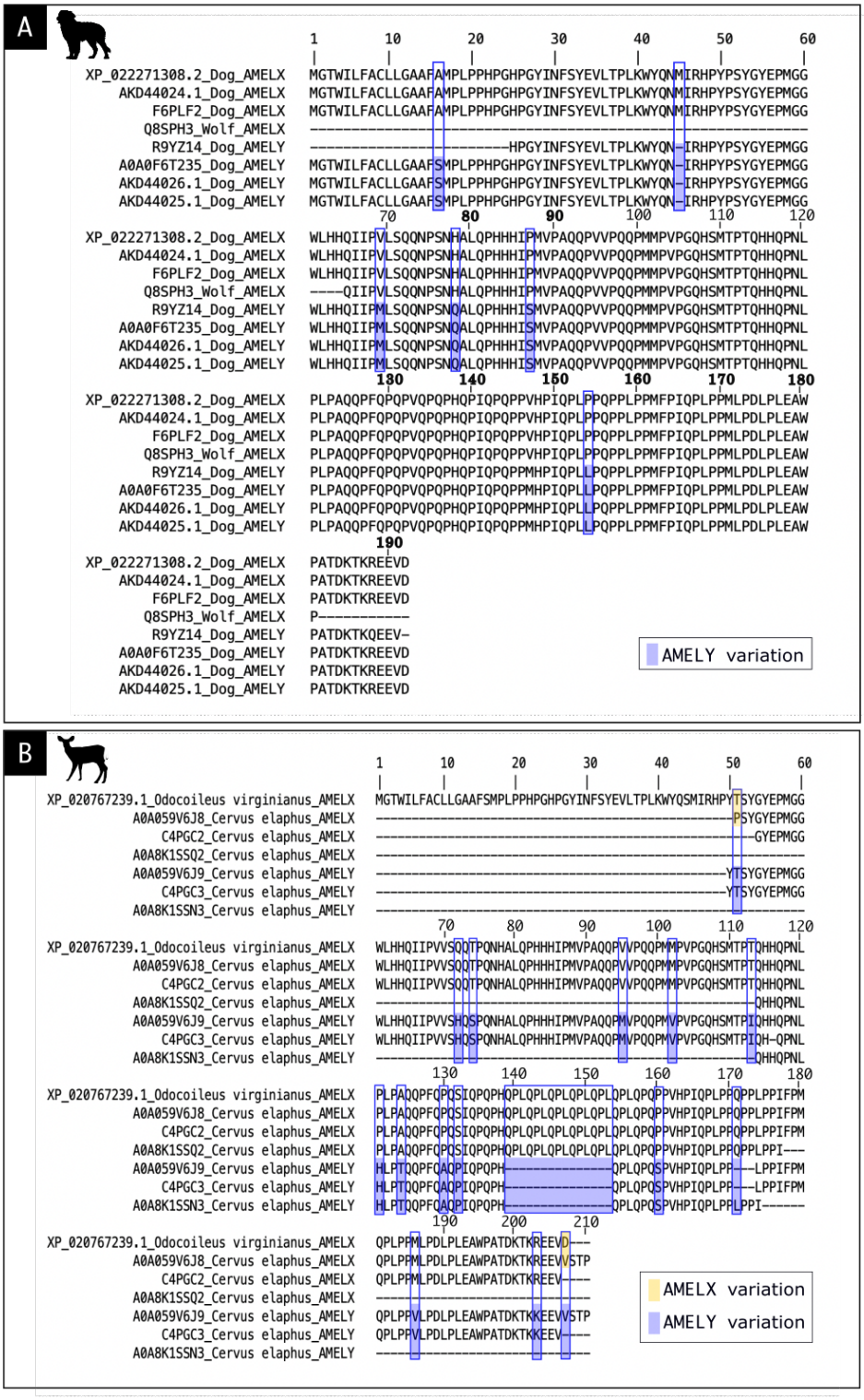


Figure 6: Canis species and Cervid alignments. A) Canis lupus familiaris AMELX and AMELY were aligned with the Canis lupus AMELX protein sequence. Differences between the two sequences are outlined in the blue box with the AMELY amino acids highlighted in blue. B) Odocoileus virginianus AMELX was aligned with the Cervus elaphus AMELX and AMELY protein sequences. Differences between the two sequences are outlined in the blue box with the AMELY amino acids highlighted in blue. Slight amino acid variation was also observed between the Odocoileus virginianus AMELX and the Cervus elaphus AMELX sequences highlighted in yellow.

Given the difference between AMELX and AMELY from the studied species, we wanted to explore the phylogeny of the amelognin protein for sex determination. Using the protein sequences downloaded from the NCBI protein to protein blast search, we performed a maximum likelihood phylogenetic analysis using IQ-tree (version 2.2.2.3). Protein sequences were aligned using Blossom 62 model and curated in Geneious Prime (v. 2022.2.1). We performed ModelFinder (-MFP) to determine the best-fit substitution model for the protein alignment (Figure 7). The Maximum likelihood tree was determined using Q.mammal + R4 for the best-fit model. Branch support was estimated of the best tree with 100 bootstrap replicates. There is not a strong phylogenetic signal to suggest deep divergence between mammalian AMELX and AMELY. Instead, AMELY and AMELX cluster within orders of mammals, with Primates in one clade, Carnivora, Artiodactyla, Perissodactyla, Prosboscidea, and Rodentia in separate clade. However, this is a poorly resolved tree where low bootstrap values do not support these clades due to the highly conserved regions of the amelogenin protein shared across mammalian taxa. Previous evolutionary studies of the amelogenin gene have revealed that the initial autosome encoding *AMEL* became homologous on the short arms of the sex chromosomes 100-180 million years ago (MYA) within eutherian mammals, before rodents diverged (Iwase *et al.*, 2007). Since this translocation, recombination inhibition stabilized AMELX and AMELY and allowed for independent differentiation potentially under different functional constraints. While the AMELX gene is further expressed and functionally active compared to the lowered expression of the AMELY, the continual expression of the AMELY indicates its conservation at the functional level. These early studies using limited mammalian amelogenin sequences suggest that it is possible that AMELX and AMELY in different mammalian lineages underwent positive selection followed by negative selection to retain these functions, where a positive selection established and stabilized the translocation, however subsequent negative selection prevented significant alteration of the gene product. This has been observed among humans and non-human primates, Ruminantia (cattle and goats) and horses (Iwase *et al.*, 2007).

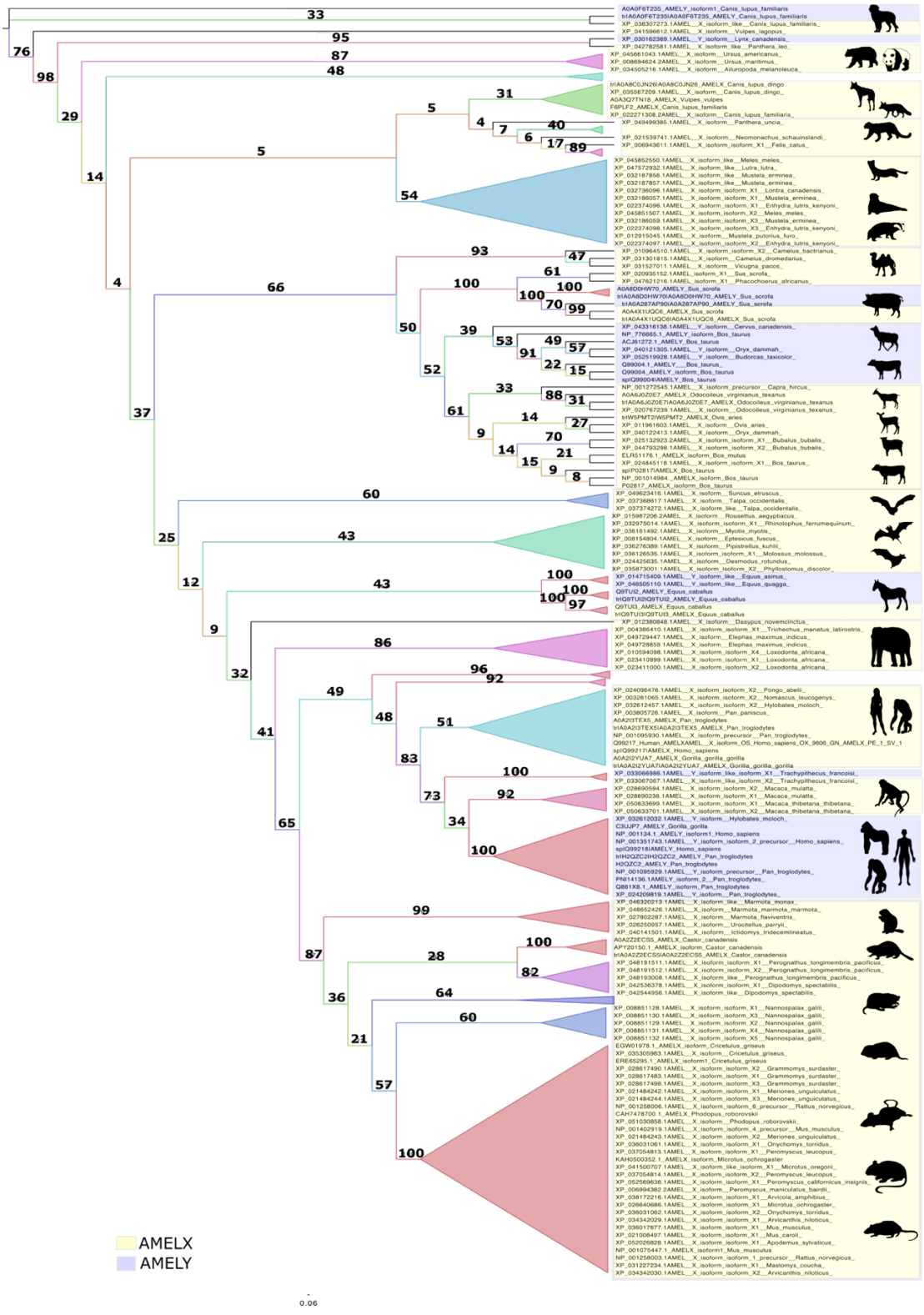



Figure 7: Phylogenetic analysis of the amelogenin protein (AMELX and AMELY) across multiple mammalian taxa. AMELX and AMELY sequences were downloaded from NCBI

following a protein-to-protein blast search. Maximum Likelihood phylogenetic analysis using IQtree and Q.mammal + R4 for the best-fit model. Bootstrap values are shown on the branch.

### 3.4.1 Sex identification by mass spectrometry

Using a targeted proteomic approach, we investigated the recovery of AMELY or AMELX peptides from known sex mammal specimens. We created five custom databases for these analyses: 1) an expanded mammals database containing the available amelogenin protein sequences from our target taxa and BLAST search in addition to the Human Proteome (UP000005640); 2) a restricted mammal's database with amelogenin protein sequences for common domesticated species and their wild counterpart available on Genbank and Uniprot; 3) *Canis lupus familiaris* Proteome (UP000002254) including *Canis lupus* AMELX; 4) *Odocoileus virginianus* Proteome (UP000504613) including *Cervus elaphus* AMELX and AMELY; 5) *Castor canadensis* Proteome (UP000694852). To create the expanded mammal database, we combined our protein-to-protein BLAST search results from NCBI (September 07, 2020) with the human proteome reference database (UP000005640). The human proteome reference database was included to characterize any unique peptides and potential human contamination. The restricted mammal's database included amelogenin protein sequences for human and non-human primates, common domesticated species, and wild canid, cervid, and bovid species available on Genbank and Uniprot on January 05, 2023. The restricted mammal's amelogenin database, *Canis lupus familiaris* proteome, *Odocoileus virginianus* proteome, and *Castor canadensis* proteome databases were all downloaded from UniProt on January 05, 2023.

All mass spectrometry raw files assessed using Xcalibur Quality Browser and were converted into mzXML files using MSConvert, and analyzed against the various reference databases using MS-GF+ (version 2018.10.15)(Kim, Gupta and Pevzner, 2008; Kim and Pevzner, 2014) (Table 2). MS-GF+ scores the MS/MS spectra against peptides within a protein database for peptide identification. This approach has shown to be more sensitive in peptide identification and universal in working with diverse spectra, MS configurations, and experimental protocols (Kim and Pevzner, 2014). Peptide identification is achieved by computing the E-values of the peptide spectrum matches (PSM), which is referred to as a *Spectral E-value*. A decoy database is created to calculate the probability of matched spectra to the database by chance and provides a global confidence assessment, also known as false discovery rates (FDR), to control the false positives among the peptide spectrum matches. Factoring in the *Spectral E-value* and FDR, MS-GF+ provides an E-value for the probability of identifying the peptides by chance. In this study, recovered peptides were filtered (E-Value $\leq 0.01$). We continued the blind study throughout the data analysis process to compare the mass spectrometry results with the known sex identification records. Filtered AMELX and AMELY peptides were considered for sex determination. Sex was determined by the absence or presence of the AMELY protein.

All environmental and extraction controls were run on the mass spectrometer and analyzed with MS-GF+. We recovered no significant peptide spectrum matches from the environmental and extraction controls. With the addition of blanks in between samples, we can confirm there is no carry over between samples, nor is there any evidence of cross contamination at earlier steps.

*3.4.2.1 Expanded mammals amelogenin and Human Proteome database*

For the first analysis, all mzXML files were analyzed against the expanded mammals amelogenin protein sequence and Human Proteome with MS-GF+ default parameters – Carbamidomethyl as a fixed modification, and Oxidations (M), Deamidation (NQ), Pyro-glu from E, Pyro-glu from Q, Acetylation Protein N-term, Acetylation (K), Methylation (K), and Phosphorylation (STY) as variable modifications. We incorporated the Human Proteome as a quality control measure to determine if we were recovering amelogenin protein and not human contaminants. This first analysis proved to be unsuccessful for sex identification as we routinely recovered "INFSYEVLTPLK" peptide at the 28-40 region, with phosphorylation of the Serine. This region of the amelogenin protein is a conserved across AMELX and AMELY and can therefore not distinguish sex (Figure 5). While we did have unique peptides matching to AMELX or AMELY within some of our samples, we received high e-values for the peptides sequences and therefore could not trust these matches to be statistically significant. We believed this could not determine sex due to the limited AMELY protein sequences within the dataset that was representative for the taxa in this study as well as the selected MS-GF+ parameters.

*3.4.2.2 Restricted mammal's amelogenin database*

To improve resolution, all mzXML files were re-analyzed against the restricted mammal's database which included only amelogenin protein sequences for common domesticated species and their wild counterpart that had both AMELX and AMELY protein sequences available, in addition to the available proteins sequences for the specimens used in this study. Since the samples were undigested during protein extraction, we set the cleavage to unspecific and used the same modification parameters from the first analysis. This analysis recovered longer peptides for both AMELY and AMELX proteins, however again they were conserved peptides and not useful for sex determination. In addition, confident peptide matches for amelogenin were not recovered in eight samples (1 dog, 2 beavers, and 5 coyotes) due to high E-values and limited peptide spectrum matches.

Since the specimens within this study are historic (1988-2010) and have been curated in a museum, we redefined our parameters for ancient protein analysis. Ancient proteins are susceptible to high deamidation rates. To factor these modifications that reflect deamidation, all raw files were analyzed against the second database with unspecific digest and variable modifications for Oxidations (M), Deamidation (NQ), Pyro-glu from E, Pyro-glu from Q, Phosphorylation (S), and hydroxyproline. No fixed modifications were specified. We recovered alternative longer peptides of the originally matched AMELX and AMELY peptides within the canid specimens and AMELX within the deer specimens, however we could still not determine sex for each specimen given that the peptides were conserved between the AMELX and AMELY protein. In addition, we received limited results back for the same eight samples discussed earlier and could therefore not determine sex.

*3.4.2.3 Mammal's proteome databases*

In addition to the expanded and restricted mammal's database, we analyzed the samples against their species' proteome to confirm that most of the peptide-spectrum matches (PSM) were

unique to dental remains (enamelin, ameloblastin, and amelogenin). At the time of this analysis, there was no available proteome for coyote (*Canis* latrans), therefore all coyote samples were compared against the dog proteome. We believed that this was an appropriate database for comparison given that dogs and coyotes are closely related taxa. All canid samples were analyzed against the dog proteome reference database using the same ancient proteins parameters discussed earlier. We yielded similar results from the proteome database search compared to the restricted mammals amelogenin database for some of the deer and canid specimens and recovered the same conserved region of AMELX and AMELY in most specimens as previously recovered in the other database searches. However, we could not confidently recover AMELX or AMELY peptides for *Castor canadensis*. In addition, we still could not provide better resolution for sex determining peptides for the eight specimens that previously failed, along with an additional coyote specimen (n=9 undetermined sex specimens). We did, however, recover the enamelin protein unique to dental remains, therefore providing a secondary quality control measure for our specimens. From these analyses we recovered with confidence AMEL proteins in many of the specimens (coyote, dog, and deer) however the recovered peptides from each of these analyses were not unique to the AMELX or AMELY protein for sex identification.

## 3.4 RESULTS/DISCUSSION

### 3.4.1 Challenges to determine sex in animals

We routinely recovered the “INFSYEVLTPLK” peptide with phosphorylation at the serine (S) amino acid (Figure 5). With the restricted mammal’s database and proteome databases, we were able to routinely recover larger peptides such as the “MPLPPHPGHPGYINFSYEVLTPLK” peptide with oxidation at the methionine (M) and phosphorylation at serine (S) within deer, coyote, and dog and occasionally the “LEAWPATDKTKREEVD” peptide within some dog samples. Together these results suggest that we successfully recovered amelogenin from these samples. However, between species and across taxa this is a conserved region of the amelogenin protein where sex cannot be distinguished. The “INFSYEVLTPLK” and “MPLPPHPGHPGYINFSYEVLTPLK” peptides are found within the N-terminal domain of the amelogenin peptide, while the “LEAWPATDKTKREEVD” peptide is found on the C-terminal domain. The amelogenin protein is made up of three domains, a hydrophilic N-terminal domain with the only post-translational modification phosphorylated site, a middle hydrophobic region, and a hydrophilic highly charged C-terminal domain (Norcy *et al.*, 2011).The greatest variation between the AMELX and AMELY across taxa is observed within the central hydrophobic region. Unlike, the central region, the N-terminal and C-terminal domains are highly conserved across mammalian taxa (Figure 8) and are necessary for enamel formation (Norcy *et al.*, 2011; Shin *et al.*, 2020). A single substitution in the amino acid sequence of the N-terminal domain can result in amelogenesis imperfecta (Iwase *et al.*, 2007), thus signifying the N-terminal domain’s importance in the stability of the enamel matrix. In particular, the phosphorylation at the serine site in the N-terminal domain significantly influences the enamel mineralization because it stabilizes the amorphous calcium phosphate and inhibits apatite crystal growth and therefore controls enamel formation (Norcy *et al.*, 2011; Green *et al.*, 2019; Shin *et al.*, 2020). Given the stability of this peptide, it may preserve better than the more variable hydrophobic region. Enamel proteome phosphorylation and proteolytic digestion have been recovered from *Homo*

*antecessor* and *Homo erectus* fossils (Welker *et al.*, 2020), further signifying the stability of this peptide.

While we successfully recovered amelogenin peptides that contribute to the formation and mineralization of the dental enamel, we did not recover peptides from variable regions; therefore, we were unable to determine sex from the specimens between the five database searches (Table 4). In addition, we were unable to confidently recover amelogenin peptides for eight specimens as the recovered peptides had high E values. Sex identification was most challenging for coyote and beaver specimens within this study. As previously mentioned, beaver likely do not have an AMELY protein to recover. Most of the coyote specimens returned low confidence peptide matches, which may suggest that there may be greater protein differences between the two species than initially assumed. To test this, we aligned the *Vulpes vulpes,* AMELX protein sequence with our canid protein sequences and found no differences in the amino acid sequencies between the canid taxa. As *Vulpes* amelogenin is expected to be more divergent than from any *Canis* amelogenin, this does not explain the decreased recovery of amelogenin. Therefore, the low confidence peptide matches may be due to our methodological approach in omitting a digestion stage at outlined in the Stewart *et al.* (2017) protocol. Proteomic methods often use trypsin, a serine protease that is used to breakdown proteins into peptides for mass spectrometry, yet Stewart *et al.* (2016) suggested that this step can be eliminated as it did not improve sequence coverage in their study. The N-terminal and C-terminal domain are shorter in length compared to the variable hydrophobic region, yet as mentioned previously the N-terminal and C-terminal domains are preserved well and may not have fragmented for mass spectrometry analysis. Given the lack of trypsin digestion and the young age of the samples compared to those in Stewart *et al.* (2016), we may have not recovered more informative peptides from the mass spectrometer column given their peptide size.

Table 2: Generated Spectra across taxa samples

| **Species** | **Range of the total no. of generated spectra** |
|---|---|
| *Canis latrans* | 3,993 – 18,664 |
| *Canis lupus familiaris* | 8,481 – 18,218 |
| *Castor canadensis* | 1,223 – 16,615 |
| *Odocoileus virginianus* | 17,236 – 18,953 |

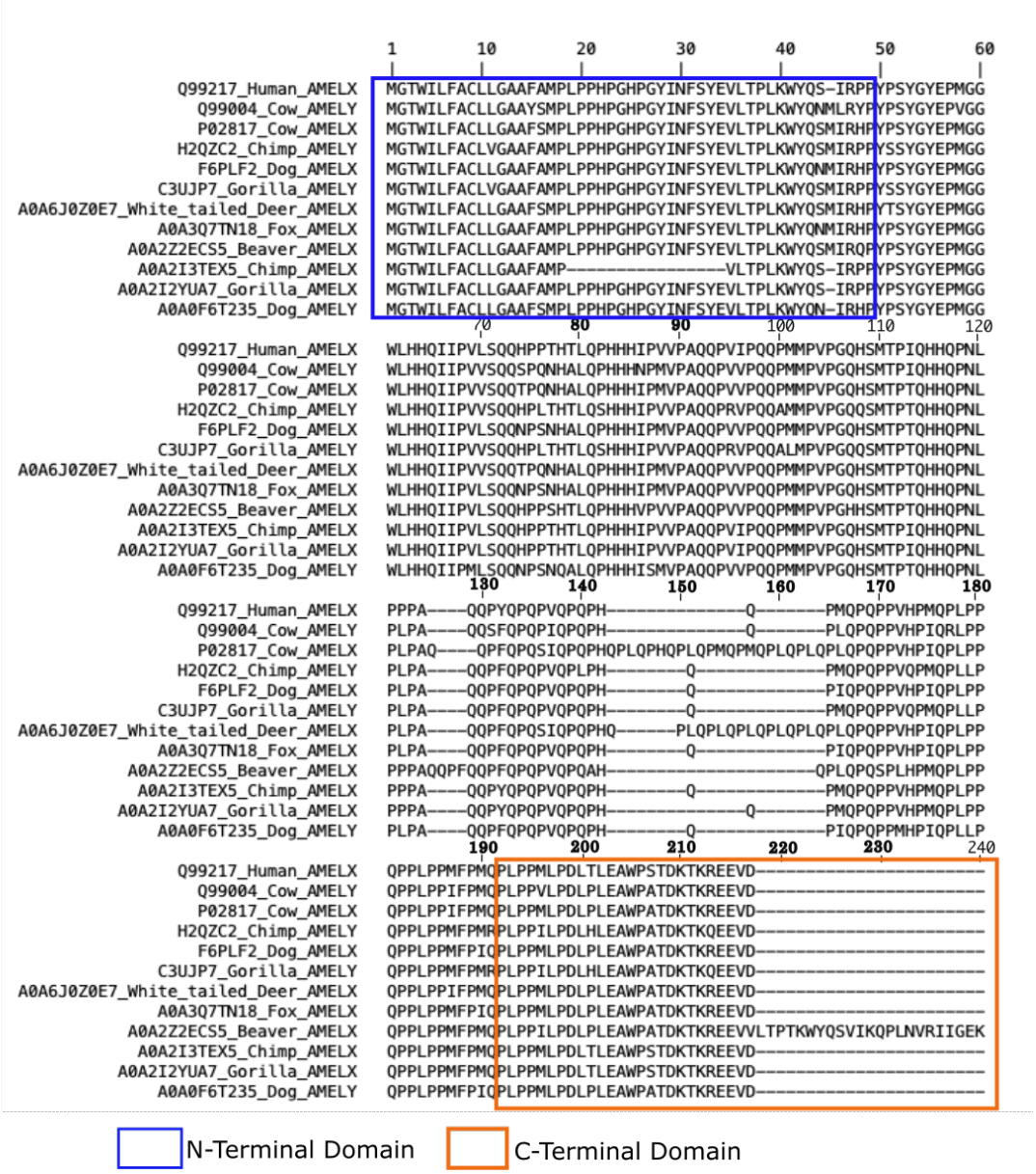



Figure 8: N-terminal and C-terminal Domain within the amelogenin protein. N-terminal domain outlined and blue and C-terminal domain outline in orange. Geneious Prime (v. 2022.2.1) Blossom 62 protein alignment of human, cow, chimp, dog, gorilla, fox and deer amelogenin to show the conserved N-terminal and C-terminal domain across taxa.

Additionally, we followed measures for ancient proteins, assuming that the proteins in this study would be fragmented due to time (~40 years) and curatorial methods. We expected to see PTMs such as deamidation (NQ) where amino acids asparagine and glutamine form glutamate and aspartate, an authenticity measure for ancient proteins (Cappellini, Collins and Gilbert, 2014). However, it is possible that recovered proteins did not have much deamidation, commonly found in ancient proteins, and therefore we did not recover all the peptides given their size. Future attempts might use C4 liquid chromatography columns instead of C18 columns to increase the

length of the recovered peptides in a top-down proteomics approach. Alternatively, the use of trypsin, could digest larger peptides for mass spectrometry analysis. Stewart et al (2017) eliminated this step for their analysis on archaeological specimens, but it may be necessary for historic specimens such as those in this study. Alternatively a trypsin digestion step may increase the peptide signal to recover AMELY specific peptides given the already low signal of AMELY compared to the AMELX (Parker *et al.*, 2019). This signal is further diluted when there are high levels of post-translational modifications like methionine oxidation or deamidation (Parker *et al.*, 2019). Parker *et al* (2019) indicate that these PTMs can further complicate the peptide matches as the number of potential amelogenin peptide masses are increased and that a trypsin digest may return more sexually dimorphic peptides. Within this study, many of the samples experienced PTM methionine oxidation which could potentially have contributed to the low signal for sexually dimorphic peptides. Further investigation in mass differences a peak calling may be necessary to uncover sexually dimorphic peptides that we cannot uncover using MS-GF+ alone.

### 3.4.2 Determining sex from archaeological human dental remains as a comparison

Given the challenges in determining sex from historic non-human animal remains, we performed the same methodological approaches outlined earlier on three archaeological dental remains (two adults and one juvenile) dating to the Roman period ($2^{nd}$-$6^{th}$ centuries CE), as a comparison and positive control (Supplemental 1). This involved the amelogenin protein extraction from two adult and one juvenile individual dental remains. The two adult individuals had been previously sexed using osteological methodologies. Unfortunately, sex determining osteological approaches are only suitable for adult individuals, therefore sex could not be determined for the juvenile individual. Sex identification was also determined through genetic analyses on one of the adult individuals (Table 3). We followed the same protein extraction steps and mass spectrometry parameters as outlined previously for the animal samples. All mass spectrometry raw files were converted into mzXML files using MSConvert, and analyzed against the human proteome reference database using MS-GF+ (version 2018.10.15)(Kim, Gupta and Pevzner, 2008; Kim and Pevzner, 2014). Sex could be determined for the two adult samples, but not for the juvenile sample, due to the absence of the unique AMELY peptides. In addition, we recovered the "SIRPPY" peptide at the 45-50 amino acid location in the peptide sequence. As discussed previously the "SIRPPY" peptide has been used to determine sex as female in human remains in previous studies due to the absence of the Methionine amino acid. We were not able to determine sex with confidence from the PSM for the juvenile individual as we recovered limited non-unique amelogenin peptides. However, the recovery of sex specific peptides using this method shows that it is possible to recover sex specific peptides from tooth enamel and further supports sample age as a contributor to the poor performance of the non-human mammals in this study.

Table 3: Sex determination overview from archaeological human dental remains

| Sample ID | Age | Total Generated Spectra | Osteological Sex determination | Genetic Sex determination | Peptide Sex determination |
|---|---|---|---|---|---|
| ROM3 | Adult | 22,910 | Female | NA | Female |
| ROM10 | Adult | 24,849 | Male | Female | Female |
| TT3 | Juvenile | 16,762 | CND | NA | CND |

*Cannot determine (CND)

## 3.6 CONCLUSION AND FUTURE DIRECTIONS

We presented an approach to determine sex from zooarchaeological remains. While we did not successfully recover sex determining peptides of the AMELX and AMELY protein for sex determination of the specimens used in this study, we did routinely recover amelogenin peptides. The specimens used in this study were collected over the last 40 years and given the stability of the amelogenin protein, may have required a trypsin digest or a different liquid chromatography column for mass spectrometry analysis. The bottom-up approach, without digestion, therefore recovered limited unique peptides in the historic animal remains. However, it was successful for recovering sex specific peptides in two archaeological human samples (up to ~1900 years old), thus highlighting that the methods taken can recover sex specific peptides. This may be due to the age of the proteins and possible fragmentation of the peptide compared to the more peptides recovered from the animal dental remains in this study. While the described approach could not determine sex from historic animal remains, future data analysis calling for the specific mass peaks may provide further sex identification resolution. We anticipate that this approach may still be viable for exploring questions about animal harvest and husbandry practices, but additional methodological refinement is required. We suggest that researchers should consider the preservation of the samples in selecting a protocol and mass spectrometry (top-down vs bottom-up proteomics) approach as well as identify *in silco* amino acid differences to determine sex.

Table 4: Sex Determination Results

| **Sample ID** | **Species** | **Date collected** | **OMNH Sex** | **Expanded Mammals** | **Restricted Mammals** | **Species Proteome** |
|---|---|---|---|---|---|---|
| 56496 | *Canis latrans* | ND | M | U, Peptide | CD | CD |
| 43780 | *Canis latrans* | 1981 | F | U, Peptide | CD | CD |
| 43791 | *Canis latrans* | 1980 | M | U, Peptide | U, Peptide | CD |
| 43833 | *Canis latrans* | 1980 | M | U, Peptide | U, Peptide | U, Peptide |
| 43820 | *Canis latrans* | 1981 | M | U, Peptide | U, Peptide | U, Peptide |
| 43821 | *Canis latrans* | 1982 | F | U, Peptide | U, Peptide | U, Peptide |
| 10924 | *Canis latrans* | 1982 | M | CD | U, Peptide | U, Peptide |
| 39190 | *Canis latrans* | 2002 | F | U, Peptide | U, Peptide | U, Peptide |
| 43381 | *Canis latrans* | 1981 | F | CD | CD | CD |
| 43801 | *Canis latrans* | 1980 | F | CD | CD | CD |
| 54024 | *Canis lupus familiaris* | 1989 | M | U, Peptide | U, Peptide | U, Peptide |
| 44081 | *Canis lupus familiaris* | 1987 | M | U, Peptide | U, Peptide | U, Peptide |
| 43999 | *Canis lupus familiaris* | 1982 | M | CD | CD | CD |
| 11603 | *Canis lupus familiaris* | 1984 | M | U, Peptide | U, Peptide | U, Peptide |
| 48492 | *Canis lupus familiaris* | 1985 | M | U, Peptide | U, Peptide | U, Peptide |
| 44044 | *Canis lupus familiaris* | 1985 | F | U, Peptide | U, Peptide | U, Peptide |
| 60065 | *Odocoileus virginianus* | 1984 | M | U, Peptide | U, Peptide | U, Peptide |
| 58600 | *Odocoileus virginianus* | 1984 | F | U, Peptide | U, Peptide | U, Peptide |
| 58282 | *Odocoileus virginianus* | 1984 | F | U, Peptide | U, Peptide | U, Peptide |
| 45653 | *Odocoileus virginianus* | 1982 | M | U, Peptide | U, Peptide | U, Peptide |
| 60056 | *Odocoileus virginianus* | 1984 | F | U, Peptide | U, Peptide | U, Peptide |
| 45645 | *Odocoileus virginianus* | ND | F | U, Peptide | CD | CD |

| | | | | | | |
|---|---|---|---|---|---|---|
| 45603 | *Odocoileus virginianus* | 1980 | M | U, Peptide | U, Peptide | U, Peptide |
| 45602 | *Odocoileus virginianus* | 1980 | M | U, Peptide | U, Peptide | U, Peptide |
| 45860 | *Odocoileus virginianus* | 1989 | M | U, Peptide | U, Peptide | U, Peptide |
| 45654 | *Odocoileus virginianus* | ND | F | U, Peptide | U, Peptide | U, Peptide |
| 39460 | *Castor canadensis* | 2010 | M | U, Peptide | U, Peptide | U, Peptide |
| 11536 | *Castor canadensis* | ND | F | U, Peptide | U, Peptide | U, Peptide |
| 48596 | *Castor canadensis* | 2009 | F | U, Peptide | U, Peptide | U, Peptide |
| 53129 | *Castor canadensis* | 1998 | M | U, Peptide | CD | CD |
| 53130 | *Castor canadensis* | 1998 | F | U, Peptide | U, Peptide | U, Peptide |
| 53131 | *Castor canadensis* | 1999 | M | U, Peptide | U, Peptide | U, Peptide |
| 53145 | *Castor canadensis* | 1988 | M | U, Peptide | U, Peptide | U, Peptide |
| 53147 | *Castor canadensis* | 1988 | M | U, Peptide | U, Peptide | U, Peptide |
| 15702 | *Castor canadensis* | 1987 | F | U, Peptide | CD | CD |
| 15704 | *Castor canadensis* | 1987 | F | U, Peptide | U, Peptide | U, Peptide |

*ND (no date); U, peptide (unknown sex estimation based on available peptides); CD (cannot determine due to high e-values and few PSMs)

# Chapter 4

## ANTHROPOGENIC EFFECTS ON MARINE MAMMAL ORAL MICROBOMES

**Authors and affiliations:**
Kristen M. Rayfield[1,2,3], Karissa S. Hughes[1,2], Torben C. Rick[3], Krithivasan Sankaranarayanan[1,4], Sabrina B. Sholts[3], Courtney A. Hofman[1,2,3]

1. Laboratories of Molecular Anthropology & Microbiome Research, University of Oklahoma
2. Department of Anthropology, University of Oklahoma
3. Department of Anthropology, National Museum of Natural History, Smithsonian Institution
4. Department of Microbiology and Plant Biology, University of Oklahoma

### 4.1 INTRODUCTION

Marine biome modifications and ecosystem alterations due to significant human activity is not exclusive to the Anthropocene but extends deep into antiquity. Historical exploitation of marine mammal fur, meat and oil during the 18th and 19th century consequently fostered a modern population heavily shaped by millennia of anthropogenic disruptions. The extent and scale of human impacts can be explored using a historical ecological framework that integrates archaeology with biological datasets spanning human history. In this chapter, we explore a novel approach to assess the scale of human impacts by examining oral microbiomes of California sea lions (*Zalophus californianus*). Human microbiome studies have shown environmental and social adaptations influence microbial diversity and abundance. However, these studies have yet to be widely applied in studying animal oral microbiomes. This chapter aims to link microbial diversity within the oral microbiome to anthropogenic alterations to California sea lions. We present the first historic study of California sea lion oral microbiomes using dental calculus. We use dental calculus as a sentinel for documenting transitioning environmental health and microbial dynamics among humans and California sea lions within a shared microbe-scape. We subsampled dental calculus from individuals (n=75) held at the Santa Barbara Museum of Natural History (SBMNH), as well as two collections held at the Smithsonian National Museum of Natural History, the National Marine Mammals Laboratory (NMML) collection and historic Smithsonian collections (USNM). DNA from dental calculus was extracted and prepared for metagenomic sequencing. We address how curatorial methods impact the quality of DNA preservation and find that California sea lions have a unique microbiome compared to other mammalian taxa along with pathogens present such as *Klebsiella pneumoniae* microbial. This approach is designed to be applicable to other mammal populations around the world and will provide valuable data about the human-made environmental changes to the marine biome where humans have altered the demographics, distribution, microbiome and ecology of marine mammals for millennia.

## 4.2 BACKGROUND

The California Channel Islands have fostered a long, dynamic relationship between humans and California sea lions. Located off the southern coast of California, archaeological and cultural resources extend to more than 13,000 years of human habitation (Erlandson *et al.*, 2011). While anthropogenic disruptions have rapidly increased with the onset of the Anthropocene, marine biome modifications and ecosystem alterations due to significant human activity can extend to the Pleistocene when ancient peoples hunted marine mammals (Braje and Rick, 2011; Hofman *et al.*, 2018). This has been observed among the Chumash, Tongva and their Ancestors who lived on the northern and southern Channel Islands, respectively. Both Indigenous groups had strong maritime traditions that allowed them to harvest shellfish and hunt marine mammals while maintaining trading networks between the Channel Islands and the mainland (Braje and Erlandson, 2008; Braje and Rick, 2011). Within the Channel Islands, archaeologists and historical ecologists have developed a long-term record of marine mammal hunting as evidenced by both zooarchaeological and technological artifacts in addition to the absence or presence of pinnipeds in their abundance, biogeography and breeding behavior (Rick and Erlandson, 2008; Braje and Rick, 2011; Thurstan *et al.*, 2015; Hofman *et al.*, 2018; Radde, 2021; Scarborough *et al.*, 2022). Rookeries were most often exploited, with subadults being most sought after given their size for ease of transport and age so that future generations were not dramatically impacted (Radde, 2021). Of historical importance, during the 18th and 19th century, pinniped populations were heavily hunted, almost to extinction, for fur, meat, and oil (Braje and Rick, 2011). Subsequent sheep and cattle ranching stripped the vegetation on the Channel Islands, in particular San Miguel (Roberts and Consultants, 2016). By the 20th century the Channel Islands served as a dumping ground for DDT as well as a strategic US military site for testing long range missiles (Levin, 1989; Roberts and Consultants, 2016; Xia, 2020). The passage of the Marine Mammals Protection Act (MMPA) in 1972, now protects marine mammals from overexploitation, thus allowing their population to grow. However, loss of genetic diversity within marine mammal populations has been linked to increased disease susceptibility(Acevedo-Whitehouse *et al.*, 2003). Such genetic diversity loss may have a greater impact on the population as the introduction of toxic pollutants and waste-water runoff as well as antibiotic resistance due to agricultural use has further increased after the passage of the MMPA (Navon-Venezia, Kondratyeva and Carattoli, 2017).

One way we can study the scale and longevity of these anthropogenic disruptions to California sea lions is through reconstructing their microbiome over time. The study of the human microbiome and its relation to environmental, diet, and health factors that impact microbial diversity is an area of active research (Dewhirst *et al.*, 2010; Hooper, Littman and Macpherson, 2012; Warinner *et al.*, 2014; Schnorr *et al.*, 2016; Fellows Yates *et al.*, 2021). Human and animal microbiome studies have shown environmental and social adaptations that influence the microbial diversity and abundance of phyla within their microbiomes (Trinh *et al.*, 2018). For example, chimpanzees have similar microbiomes to other chimpanzees within their troop as opposed to chimpanzees within different troops (Ezenwa *et al.*, 2012; Moeller *et al.*, 2014), while humans also share similar microbiomes with those living in their same household (Mosites *et al.*, 2017). Consequently, anthropogenic activities have shown to alter the microbial diversity and abundance of phyla within animals and the environment (Trinh *et al.*, 2018). In fact, loss of microbial diversity in mammals due to anthropogenic factors have resulted in similar gut

microbiome profiles to that of humans – increased abundance of *Prevotella* and *Bacteroides* genera and antimicrobial resistance genes (Clayton *et al.*, 2016; McKenzie *et al.*, 2017; Baron, Diene and Rolain, 2018). This loss in microbial diversity has been observed in non-human primates within captivity when compared to wild populations (Ley *et al.*, 2008; Clayton *et al.*, 2016; McKenzie *et al.*, 2017). Microbial diversity loss within non-human primates has been largely attributed to the use of antibiotics, loss of environmental wild microbes in their anthropogenically controlled captive habitat, and a shift to a processed diet (Clayton *et al.*, 2016; Brealey *et al.*, 2021), all of which are also commonly practiced in human agricultural and animal husbandry practices (Rice *et al.*, 2012). Microbiomes and microbe selection are therefore not only environmentally selected but also culturally selected by human interference. However, less is known about the potential transmission of human pathogens to non-human animals during this humanization process.

The oral cavity can be viewed as an entryway for microorganism introduction, colonization, and formation of the commensal relationship as well as an interaction zone with the innate host defenses within the oral microbiome. As a result, it is a reservoir for both primary and opportunistic pathogens, second compared to the gut (Warinner *et al.*, 2014). These microbial communities that make up the oral microbiome are comprised of bacteria, protozoa, fungi, viruses and archaea (Benn *et al.*, 2018). Across mammalian taxa, as well as in domestic fowl, the oral microbiome is influential in both training the immune system and maintaining the microbiome ecology (Allen and Stanton, 2014; Bahrndorff *et al.*, 2016). In a symbiotic relationship between microbes and host epithelial cells, the host will provide nutrients to the microbes and in return the microbes will provide protection against exogenous species through colonization resistance (Marsh, 2000). Similarly, horizontal gene transfer from exogenous species to endogenous oral species can contribute to their adaptive potential and increased fitness within the oral microbiome (Roberts and Kreth, 2014). As a result, the early colonizers of the oral microbiome shape the host physiology as well as the innate host's defenses (Marsh, 2000; Kolenbrander *et al.*, 2002; Fellows Yates *et al.*, 2021). By documenting and reconstructing mammalian oral microbiomes, we can start to look at how anthropogenic disruptions selected for increased pathogenesis, virulence, or antibiotic resistance among microbes and the coevolution of some diseases. Understanding colonization history due to host environmental factors plays a large role in the evolutionary and behavioral components that selected for their abundance and pathogenesis within the mammalian oral microbiome.

The oral microbiome can be reconstructed from both the bioarchaeological and zooarchaeological record using dental calculus. Dental calculus is a biofilm attached to the teeth surfaces and is a fossilized record of the microbial communities for past human and animal populations. Within humans, the abundance of dental calculus can vary between populations due to host genetics or dental hygiene practices (Warinner, 2016). However, the formation of dental calculus is the same, as the calcium phosphate ions found in the host saliva and gingival crevicular fluid (GCF) mineralizes the dental plaque (Warinner, 2016). DNA binds to calcium phosphate minerals, thus allowing for bacterial nucleic acid to survive. This mineralization period occurs in stages throughout an individual's lifespan and can leave a layered record of the preserved microbial cells or other debris calcified in situ (Warinner *et al.*, 2015; Warinner, 2016). It has also been observed that dental calculus may contain up to 1000 times more DNA compared to bone or dentine (Warinner, Speller and Collins, 2015; Weyrich, Dobney and

Cooper, 2015; Warinner, 2016; Mann *et al.*, 2018). The presence of dental calculus has been identified in other marine mammal populations (sea otters and killer whales) but currently it is unclear what factors contribute to dental calculus formation in marine mammals (Hofman, 2018; Wellman *et al.*, 2020). Host mitochondrial and nuclear DNA, antibiotic resistance genes, as well as entire microbial communities have been recovered from dental calculus which allows for researchers to observe host-pathogen interactions, genome rearrangement, horizontal gene transfer, and genetic mutations (Peterson *et al.*, 2009; Ozga *et al.*, 2016; Warinner, 2016; Mann *et al.*, 2020; Brealey *et al.*, 2021; Fellows Yates *et al.*, 2021; Austin *et al.*, 2022). High-throughput sequencing has also allowed metagenomic and metaproteomics methods to be used that incorporates the genes, genomes, and the protein functions of bacteria within the microbiome (Peterson *et al.*, 2009). Given the rapid advancement of high throughput sequencing and the similar development of microbiomes across mammalian species (Brealey *et al.*, 2021), these methods can be easily extrapolated and applied towards studying marine mammal oral microbiomes.

Our research uses historical collections which provides a temporal perspective to document ecological shifts, if any, within our California sea lion population. As such, our objectives are to document and reconstruct the oral microbiome of California sea lions over time and to explore changes in microbial diversity within California sea lion oral microbiomes as a result of human activities. By reconstructing the oral microbiomes of recent and past California sea lions, we can provide a temporal assessment from which we can test if anthropogenic impacts over the last ~130 years have fostered microbial diversity loss within California sea lion oral microbiomes and the potential transmission of human pathogens to California sea lions during this humanization process. Given that our specimens were collected from three different collections, we also explore and compare how different curatorial methods may impact the preservation and quality within our specimens for future research.

This study offers a novel approach in assessing the scale of human influence on shifting historical ecologies through the analysis of dental calculus from historic collections. By linking microbial diversity within the oral microbiome to recognized anthropogenic alterations, we can highlight past human and marine mammal interactions and provide meaningful snapshots in time for assessing the scale of human impacts on the marine biomes. Incorporating such analyses can open the door in developing datasets to reconstruct past ecologies and the scale of human impacts in the past as well as provide an alternative towards studying human behavior without the analysis of human remains.

## 4.3 MATERIALS

Our study population includes California sea lions that were collected from 1884-2014. We collected n=30 specimens from the Santa Barbara Museum of Natural History (SBMNH), n=30 from the National Marine Mammals Laboratory (NMML) collection held at the National Museum of Natural History (NMNH) and n=15 from the US National Museum (USNM) collection held at NMNH (Figure 9). The specimens within these collections are historic (~130 years of age from the earliest collection date) and the majority were collected from deceased, stranded sea lions on the coastal beaches. Sample curation methods vary between collections and over time, but typically involve first flensing the larger tissue, removing any remaining tissue via

boiling or beetles and then maceration in cold water, followed by a drying period. The SBMNH California sea lion specimens were collected between 1920-2014. The NMML specimens were collected between 1950-1990 and the USNM specimens were collected between 1884-2006. All individuals sampled for dental calculus were limited to the primary breeding range extending from the Channel Islands to Central Mexico, with the majority sampled from specimens collected from the Channel Islands. Individuals were also grouped into three ecological time bins: Colonial Development period (1880-1940), the Early Global Market period (1940-1980), and the Late Global Market period (1980-present) reflecting the rapid anthropogenic disruptions introduced over the last ~130 years.

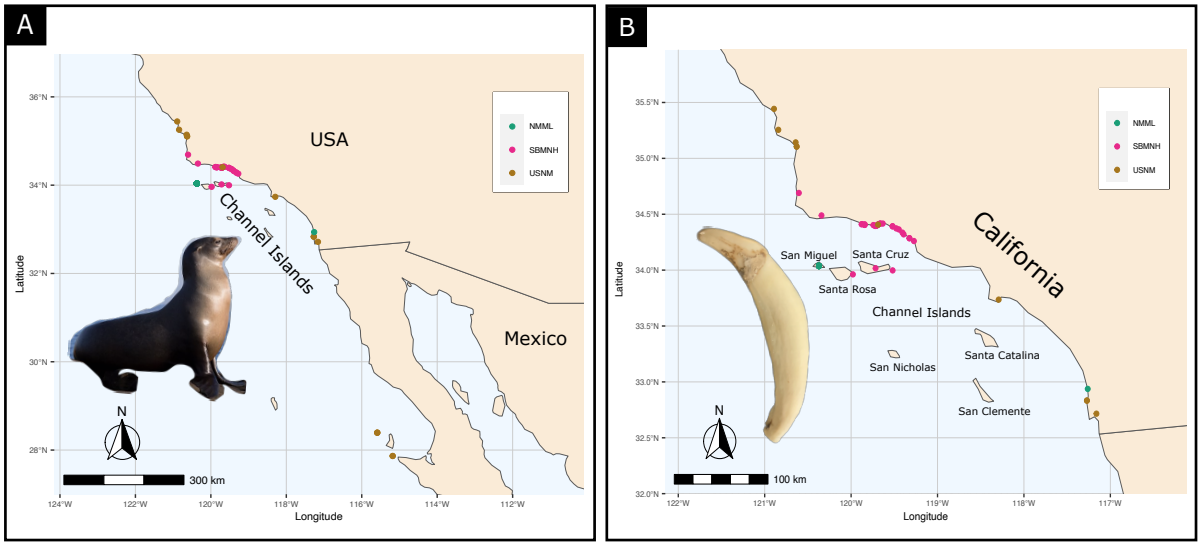



Figure 9: Individuals in this study. A) Map of individuals in this study labeled by collection. B) Map of individuals from California labeled by collection

## 4.4 METHODS

### 4.4.1 Laboratory Methods

#### *4.4.1.2 Subsampling*

Dental calculus was subsampled with permission of the SBMNH Vertebrate Zoology and NMNH Division of Marine Mammals. Subsampling of (n=75) specimens took place at the SBMNH and at the Smithsonian Museum Support Center which houses collections from the USNM and NMML (Table 5). Dental calculus was removed from either the labial surface of incisors or labial surface of canines of California sea lions with dental scalars, making sure to leave enough behind for future analyses. All laboratory procedures were completed in the ancient DNA (aDNA) lab at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR) at the University of Oklahoma.

Table 5: Collection sampling population across three time periods

| Collection | Total no. of individuals | Colonial Development No. of Individuals | Early Global Market No. of Individuals | Late Global Market No. of Individuals |
|---|---|---|---|---|
| NMML | 30 | 0 | 26 | 4 |
| SBMNH | 30 | 2 | 15 | 13 |
| USNM | 15 | 4 | 6 | 5 |

*4.4.1.3 aDNA extractions*

In brief, each extraction used between 1-15mg of dental calculus. An extraction negative control, consisting of buffer, was also included to monitor potential contamination. All dental calculus subsamples were decontaminated using a UV Crosslinker. All dental calculus subsamples were then predigested in 0.5 M of EDTA for 15 minutes. The samples were then incubated for an additional 24 hours in 1ml of 0.5 M of EDTA. 100µl of Qiagen Proteinase K was added to the samples for further digestion and decalcification over ~5 days. DNA was extracted following a modified silico-extraction protocol (Wellman *et al.*, 2020) the samples were purified on a Qiagen MinElute column to remove salts/impurities and to concentrate the DNA extract. The DNA concentrations of the extracts were measured using the double stranded DNA High Sensitivity (dsDNA HS) kit for Qubit 3 Fluorometer.

*4.4.1.4 Illumina library construction*

Dual-indexed Illumina double stranded metagenomic libraries were constructed on (n=75) samples and well as extraction and library controls (n=15). The library preparation steps were conducted using previously published protocols (Carøe *et al.*, 2018; Wellman *et al.*, 2020). All samples underwent partial UDG treatment to repair internal cytosine deamination while still allowing for authentication using terminal end DNA damage (Rohland *et al.*, 2015). The concentration of DNA within the libraries was then measured through quantitative polymerase chain reaction (qPCR) and to determine the number of cycles needed for indexing. For indexing, a unique DNA barcode was added to each sample while the samples were amplified via PCR. Libraries were amplified in triplicate using Kapa HiFi HotStart Uracil, which tolerates uracil residues and low concentration of AT-rich DNA. All qPCR and PCR setup were conducted in the aDNA lab while amplification and further work was conducted in the modern laboratory at the Stephenson Research and Technology Center.

*4.4.1.5 Metagenomic sequencing*

Following three indexing PCR replicates, all library replicates were then pooled and quantified for sequencing. The concentration of each sample was again determined using the dsDNA HS kit for Qubit 3 Fluorometer. 1ng/ul dilutions were then made to run on the Fragment Analyzer (model 5200, FA Version #: 1.2.0.11) following the High Sensitivity protocol. This procedure provides more accurate concentrations, average fragment sizes, and whether dimers were present. As there were dimers in the samples, they were pooled in 20 nM dilutions and run through a 2% Pippin Prep gel to select for fragment between 150 and 500 base pairs in size. We

sequenced the libraries (2 x 150bp chemistry) on an Illumina NovaSeq S4 instrument at the Oklahoma Medical Research Foundation.

### 4.4.2 Bioinformatic Workflow

#### *4.4.2.1 Quality Filtering*

Quality filtering was completed using a pipeline that integrates several different open source bioinformatic tools. The pipeline begins with quality filtering the data by removing adapter sequences, trimming low quality bases, and merging read 1 and read 2 with AdapterRemoval 2 (version 2.1.7; Schubert, Lindgreen and Orlando, 2016). Low quality reads and reads shorter than 30 base pairs were also filtered out at this point.

#### *4.4.2.2 Host DNA Preservation*

Sequencing reads were aligned and mapped to the sea lion reference genome (GCF_009762305.2) using Burrows Wheeler Aligner (BWA) with ancient DNA parameters (Schubert *et al.*, 2012). SAMtools (version 1.5), was then used to filter out unmapped or poor quality (-q set to 37) reads and sort the alignments by coordinates. DeDup (version 0.12.7) was used to filter out duplicate reads. DNA degradation and damage patterns were assessed using MapDamage2 (Jónsson *et al.*, 2013).

#### *4.4.2.3 Oral Microbiome Profiles*

We used three approaches to reconstruct the oral microbiome profiles of the California sea lion specimens 1) MetaPhlAN4, which profiles the abundance and compositions of the microbial communities from microbial gene database (Blanco-Miguez *et al.*, 2022), 2) KRAKEN2 (Wood and Salzberg, 2014) with a custom database containing genomes for a diverse and deep breadth of mammalian species, which will allow for the identification of dietary, microbial and host DNA as well, and 3) QIIME (v.1.9.1) which analyzes microbial communities sampled through 16S marker genes and clusters them at a 97% phylogenetic level into OTUs (Caporaso *et al.*, 2010). With MetaPhlAN4, individual microbiome profile abundance tables and total reads counts for each microbial clade tables, and unknown estimation values were created for each sample (n=75) and negatives controls (n=15) using quality filtered, trimmed R1 and merged reads set for ancient input (minimum read length of 30). For all Kraken analyses, we used a minimum hit group of four to allow for overlapping kmers and a confidence threshold of 0.10. We then mapped the sample data to microbial taxa with the largest assigned read counts across all samples. Analysis ready reads (trimmed and merged reads) were mapped to the Greengenes (GG97) database using Bowtie2. Duplicates were removed using DeDup (version 0.12.7) and a FASTA file containing the sample names and the 16S rRNA reads mapping to the Greengenes database was created for OTU picking.

Given that the post-mortem preparation and curatorial methods of these specimens may contribute to contamination, we also used SourceTracker (v1.0.1) (Knights *et al.*, 2011) to predict the source of microbial communities within the samples and to determine if the samples are of an oral microbial community. Raw sequencing reads were downloaded from SRA for

human skin (n=10), human supragingival plaque (n=10), human subgingival plaque (n=5), soil (n=10), non-human mammal calculus (bear, n=10; orca, n=15; and chimp, n=10), human urban gut (n=10), human rural gut (n=10), and marine water (Pacific Ocean, n=10) (Supplemental 2). Raw-sequencing reads were then trimmed, merged, and remnant adapters were removed using AdapterRemoval2 (v.2.1.7; (Schubert, Lindgreen and Orlando, 2016). Microbiome profiles were again created with MetaPhlAN4 using the trimmed R1 and merged reads and processed with the total number of reads from each clade for each comparative source sample. All source and sample (sink) microbiome profile tables were merged and converted into a single genus level OTU table with MetaPhlAN4. A genus level OTU table was rarefied to 10,000 with the phyloseq R package (McMurdie and Holmes, 2013) and analyzed using SourceTracker.

We used MetaPhlAN4's beta diversity unweighted unifrac distance measures to quantify the dissimilarities between the samples. Given the expected damage and fragmentation in ancient samples, alpha diversity is not an appropriate to measure (Ziesemer *et al.*, 2015; Mann *et al.*, 2018). We performed an unweighted unifrac diversity measure across all calculus samples used for the SourceTracker analysis (bear, chimp, orca, human, and sea lion calculus), between collections (NMML, SBMNH, USNM), and across other sources (soil, marine water, human skin, bear calculus, chimp calculus, orca calculus, human plaque, and human gut). In addition, 16S rRNA reads were extracted from the shotgun sequencing data of all source calculus samples (bear, chimp, human, and orca) for further taxonomic mapping to the GG97 database in QIIME. A genus level OTU table was created and rarefied to 5000 in QIIME.

#### *4.4.2.4 Microbial DNA Preservation*

Following the Kraken and MetaPhlAN4 analyses, we mapped the merged sequencing reads to oral taxa found across all samples and some unique pathogens. Using the same pipeline for mapping to the host genome, we mapped all sample sequencing reads to *Neisseria zalophi* (GCF_008807015.1), *Streptococcus marimammalium* (GCA_000380045.1), *Desulfomicrobium orale* (GCA_001553625.1), *Tannerella forsythia* (GCF_000238215.1), *Enterobacteriacae* species – *Citrobacter werkmanii* (GCF_008693645.1), *Enterobacter cloacae* (NZ_JAMQCM010000001.1), *Klebsiella pneumoniae* (NC_016845.1), and *Franconibacter helveticus* (NZ_QISR01000001.1) – and *Toxoplasma gondii* strains – TGCOUG (GCA_000338675.2), TG-CTG (GCA_016808245.1), TG-Me49 (GCA_000006565.2), and TG-RH88 (GCA_019455545.1). Sequencing reads were aligned and mapped to each reference genome using BWA -q20 and -n0.1 for strict mapping quality. All damage patterns were again assessed using MapDamage2.

## 4.5 RESULTS AND DISCUSSION

The aim of our study was to provide a temporal approach in documenting microbial diversity loss due to the longevity of anthropogenic impacts on the marine ecology. In addition to microbial diversity loss, we hoped to document any humanization of the California sea lion oral microbiomes. The measure of anthropogenic influence on California sea lion oral microbiomes was to be determined by assessing loss of microbial diversity, the presence of human associated pathogen or commensals, and similarity to human oral microbiome profiles from individuals in the US (Human Microbiome Project). We also wanted to assess how museum curation methods

could impact the DNA preservation among our specimens and therefore complicate further downstream analyses (Table 6).

Table 6: Summary of sequencing data

| Collection | Range of Total Reads | Range of Proportion of reads retained | Range of Average Read Pairs after QC |
|---|---|---|---|
| NMML | 30,761,890 – 92,270,890 | 0.83 – 0.94 | 27,044,123 – 80,178,869 |
| SBMNH | 1,415,533 – 520,428,216 | 0.49 – 0.94 | 832,837 – 484,289,635 |
| USNM | 15,785,332 – 45,006,669 | 0.86 – 0.93 | 14,418,558 – 28,152,551 |
| Negatives | 1,153,001 – 10,924,554 | 0.0001 – 0.43 | 434 – 2,919,191 |

### 4.5.1 Preservation and Authentication

We obtained robust sequencing data across all three collections as well as the negatives. We assessed data quality by mapping to the *Zalophus californianus* host genome and evaluating the damage and fragmentation patterns of each sample (Supplemental 4). Archaeological and museum specimens are prone to DNA damage which involves C-T transformations at the ends of the DNA strand and fragmentation, thus resulting in shorter fragment lengths compared to modern DNA (Ginolhac *et al.*, 2011). For some of the SBMNH samples, we observed low levels of read merging during the quality filtering step, which may influence these estimates. We then remapped to the host genome using the R1, R2 and merged reads under the assumption that low merging (<80 of reads merging), might be an effect of longer fragment lengths. However, the addition of R1 and R2 trimmed reads did not substantially change fragment length estimates and further observations showed that quality filtering did not sufficiently remove adapter sequences from the R1 and R2 reads. We, therefore, proceeded with the higher quality merged reads for further downstream analyses.

All 75 samples showed damage patterns associated with partial UDG treatment (Figure 10) (Rohland *et al.*, 2015). To determine if the frequency of damage varied by collection, we compared the first position C-T transformation for each sample according to their collection. We observed a significant difference for damage between the three collections with a Kruskal-Wallis test ($p$= 0.02061). A Pairwise comparisons using Wilcoxon rank sum was performed to assess any statically significant difference between the collection levels. The differences in damage patterns between NMML and SBMNH and USNM were also significant (Pairwise Wilcox test (NMML = 30, SBMNH = 30), $p$ = 0.042; (NMML = 30, USNM =15), $p$= 0.042), thus suggesting that curatorial methods may contribute to damage patterns between the NMML and SBMNH and USNM collections.

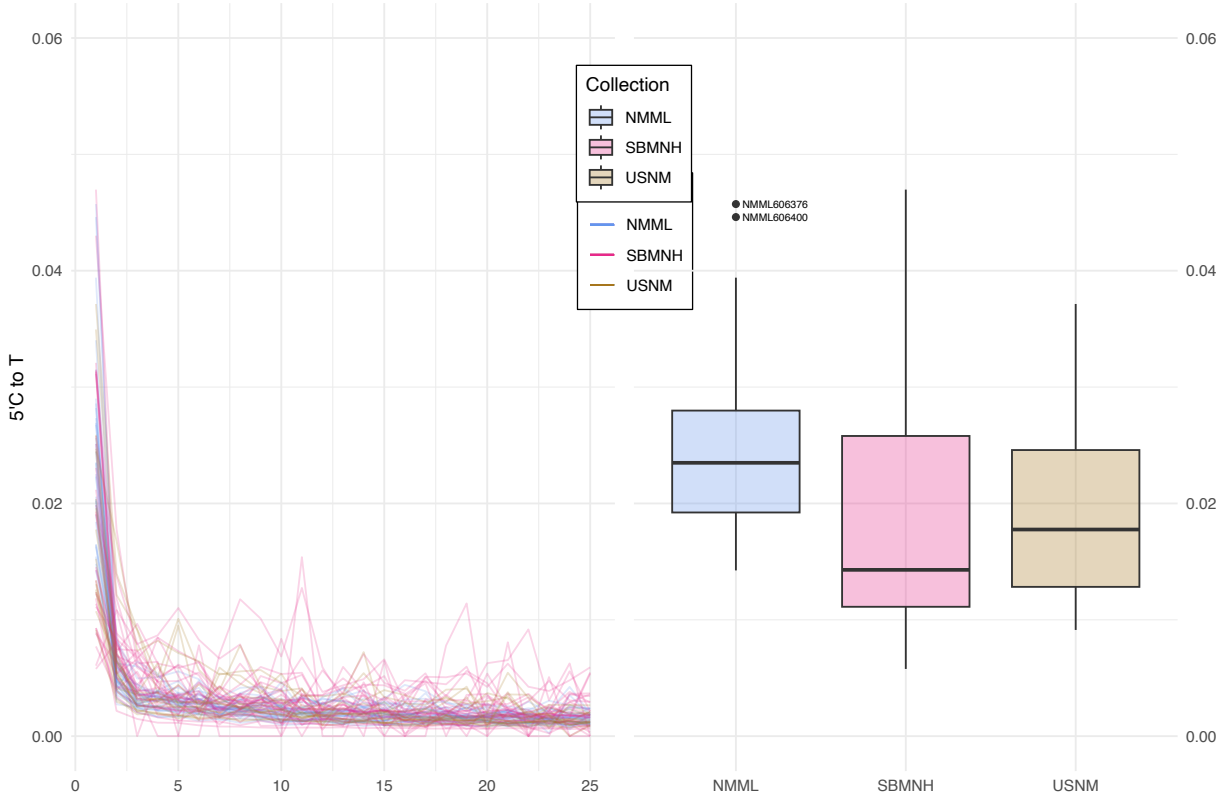



Figure 10: 5'C toT damage. Damage patterns for all samples with read counts >1000 reads mapped. Box plot of the average first position damage.

We compared the average fragment length of mapped reads for each sample across the collections (Figure 11A). A Kruskal-Wallis test was performed on the average fragment length of the three collection. The differences between the collection average fragment length of 69.1 bp (NMML), 63.7bp (SBMNH), and 60.6bp (USNM) were significant, $p = 0.0004$. Post hoc comparisons were conducted with a Pairwise comparisons using Wilcoxon rank sum exact test. The differences between NMML and SBMNH and USNM were statistically significant (Pairwise Wilcox test (NMML = 30, SBMNH = 30), $p = 0.00649$; (NMML = 30, USNM =15), $p = 0.00075$). We also divided the sample dataset by different time periods associated with human disruptions to the Channel Island ecology (Figure 11A). Time does not appear to contribute to fragment length in this study. The higher damage patterns and larger fragment size observed among samples in NMML compared to the SBMNH and USNM is interesting, as it may suggest that while curatorial methods could contribute to the observable damage patterns, fragment length does not seem to be as influenced, as anticipated. This may be due to the preservation of endogenous host DNA compared to microbial DNA between the collections Alternatively, curatorial methods between the collections have impacted DNA preservation within the SBMNH and USNM collections, thus contributing to shorter fragment lengths.

Dental calculus can be used as an alternative source for host endogenous DNA yet compared to other tissue sources (bone vs tooth) contains lower host endogenous DNA (Mann *et al.*, 2018). Within marine mammals, host endogenous DNA has been recovered from sea otter and orca

dental calculus (Hofman, 2018; Wellman *et al.*, 2020). Host endogenous is calculated from the proportion of quality filtered unique mapped reads to the total number of analysis ready reads. For this study we recovered typical calculus host endogenous content (<1 – 4%) except for one outlier, SBMNH 8981, which had 27% of host endogenous content (Figure 11B). Previous studies have highlighted that NETosis, a process associated with an immune response from neutrophils can contribute to fragmented host DNA in calculus (Mann *et al.*, 2018). This is most often attributed to signs of oral infection such as gingivitis or periodontal disease. During sample collection, we also documented the oral pathologies among the specimens in this study. Sample SBMNH 8981 did not exhibit oral pathologies, however the individual was a juvenile and exhibited dental eruption of the canines, which may contribute to the high endogenous recovery. Inflammation of the tissues near the erupting tooth could draw neutrophils to the site inducing NETosis resulting in high endogenous DNA content. Similarly, to fragment size, we divided the sample dataset by the three ecological time periods (Figure 11B). Time does not appear to contribute to recovered endogenous content in this study. However, unlike the observed statistical differences in damage patterns and fragment size between the three collections, we did not observe any differences in endogenous content between the three collections using a Kruskal-Wallis rank sum test ($p = 0.06011$).

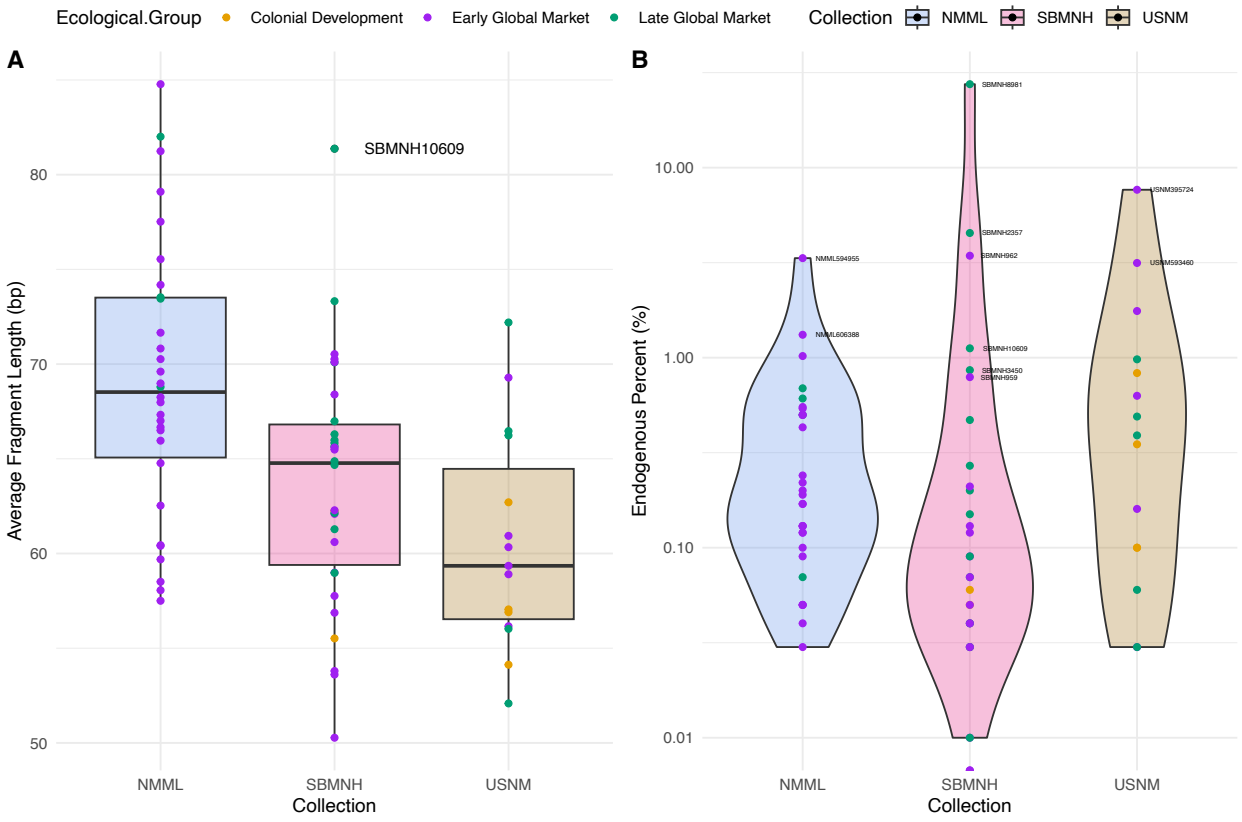



Figure 11: Average fragment length and endogenous content. A) Average fragment length of each sample across the three collections, ecological groups are shown as points and do not show any observable difference in time contributing to average fragment length. B) Endogenous content of all 75 samples across the three collections. Violin plots have been scaled by log 10.

Ecological groups are again shown as points and do not show any observable difference in the time contributing to the preservation of endogenous content.

To further explore if curatorial methods were contributing to the observable differences between the collections, we performed the same preservation and quality measures for a subset of specimens collected from the 1970s to 1980s from the NMML and SBMNH collections. A Kruskal Wallis rank sum test was performed on the damage patterns, fragment size, and endogenous content. The differences between the collection damage patterns were significant ($p = 0.0009374$) (Figure 12A). This may further support that curation processes may impact the quality and preservation of DNA from the specimens. We did not find a significant difference in average fragment length between this subset of the two collections ($p = 0.05835$) (Figure 12B). However, a significant difference in endogenous content 0.427% NMML and 0.388% SBMNH was observed ($p = 0.03692$) (Figure 12C). Many of the SBMNH samples that have inconsistent damage plots in Figure 10 also had lower a lower proportion of reads kept after trimming and merging (post-AdapterRemoval). In addition, there are fewer reads within the SBMNH samples mapping to the host genome compared to the NMML specimen. Given these factors, the difference in endogenous content between the two collections may be that exogenous rather than endogenous is DNA is more abundant in the SBMNH specimens. Further investigation of the curatorial methods used for the SBMNH collections revealed that many of the sea lion skulls were boiled and macerated in cold water until bacteria digested the remaining tissues, followed by soaking in hydrogen peroxide to whiten the bone (personal communication with SBMNH curator, Paul Collins).

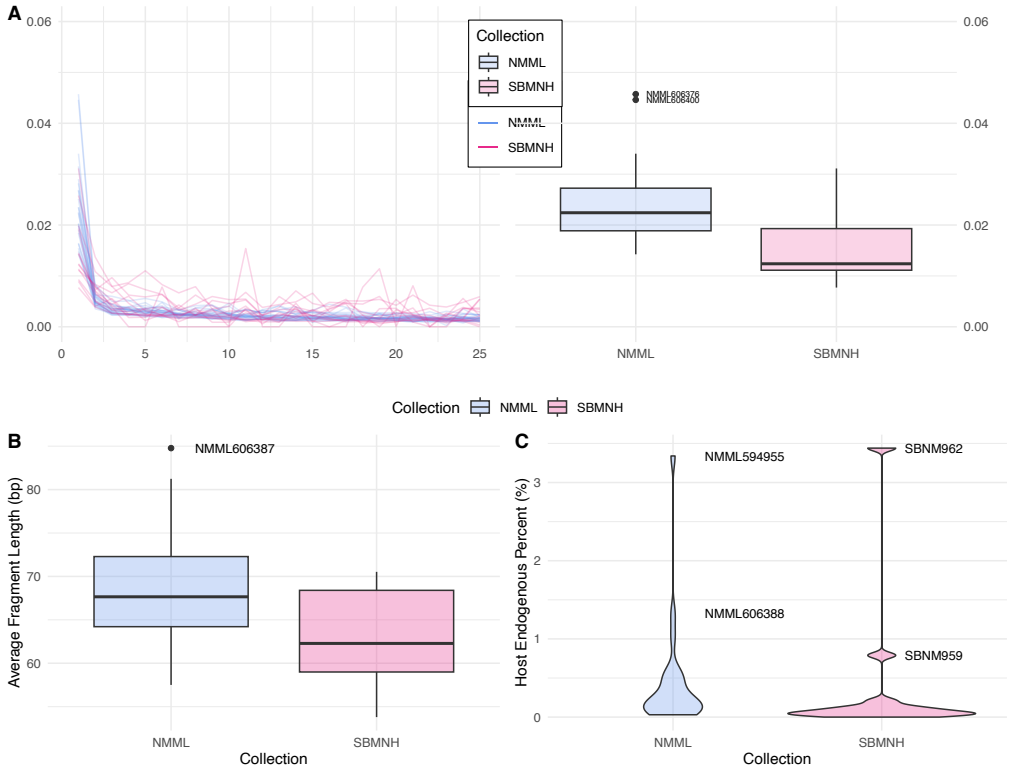



Figure 12: 1970s NMML and SBMNH damage patterns, fragment length, and endogenous content. A) Damage patterns of 37 samples from collected and curated between the 1970s and

1980s from the two collections. B) Average fragment size between the two collections. C) Endogenous content between the two collections.

### 4.5.3 Microbial Preservation and Taxonomic Profiling

To investigate the microbial communities within sea lion dental calculus, we compared sea lions to several sources including other mammal calculus, and possible environmental contaminates using MetaPhlAN4 taxonomic classifications. The majority of samples displayed oral microbiome signatures as estimated by the Bayesian Sourcetracker tool (Supplemental 3). However, eight samples did not pass authentication due to the limited presence of either soil, skin, or gut contamination. Of these samples, gut contamination was the largest contamination source. Interestingly, SBMNH 8981 and SBMNH 3822 had unique microbial profiles worth further investigation. For SBMNH 8981, Sourcetracker predicted high levels of human gut taxa and according to the Kraken RefSeq analysis, of the 50% reads assigned, 95% of the reads were assigned to the *Zalophus californianus* while relatively few mapped to microbial DNA. This is unique given that the vast majority of dental calculus is comprised of microbial DNA. This individual also had extraordinarily high endogenous content when mapping just to the *Zalophus californianus* genome as well. The predominant microbes that were assigned with MetaPhlAN4 included *Desulfovibrio intestinalis*, a mesophilic bacterium found in the hindgut of termites *Mastotermes darwiniensis,* and *Clostridium algidicarnis* a psychrotrophic bacterium originally isolated from refrigerated pork. As a comparative method, the custom refseq Kraken assigned the most reads to *Desulfovibrio sp. 86* and *Clostridium botulinum.* Given that termites are not used as a curation method, we believe that these reads may be mismapping to an unclassified *Desulfovibrio* species. Sourcetracker predicted a large proportion of unknown sources in addition to human gut and mammalian oral taxa. Comparing MetaPhlAN4 and KRAKEN2, most of the classified mapped reads were assigned to Clostridium species - *Clostridium ihumii* and *Hathewaya massiliensis* (MetaPhlAN4 output) and *Clostridium botulinum* (KRAKEN2 output). *Clostridium ihumii* and *Hathewaya massiliensis* are found in the human gut. Given the difference in database assignments, the assigned reads may be mapping to conserved regions shared between these *Clostridium* species. Unique microbial species observed in SBMNH3822 included *Enterobacteriacae species* that have shown to have antibiotic resistance genes – including *Leclercia pneumoniae*, *Citrobacter werkmanii*, and *Enterobacter cloacae*, and *Morganella morganii.*

We proceeded with further downstream analyses and microbiome profiling with the remaining sixty-seven samples that passed authentication. A PCoA was performed with unweighted unifrac beta diversity measures between comparative sources and the samples (Figure 13), between comparative dental calculus samples (Figure 14), and between each collection (Figure 15). California sea lions clustered together in the comparative sources PCoA and showed distinct differences between the environmental controls (soil and water), thus suggesting the samples represent intact oral communities given that they were collected in many ways at different time points.

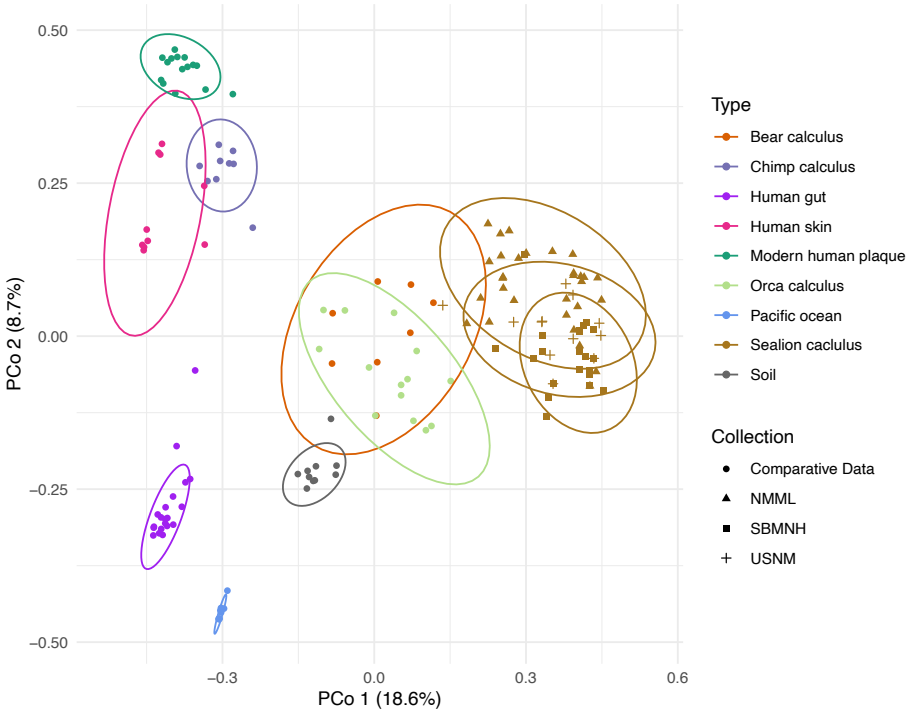



Figure 13: Comparative sources. Unweighted unifrac beta diversity PCoA of sources used for the SourceTracker and the samples within this study. Comparative sources are differentiated by color and a 95% confidence ellipse. Collections are differentiated by shape. California sea lion calculus separates well from the other environmental and comparative dental calculus sources.

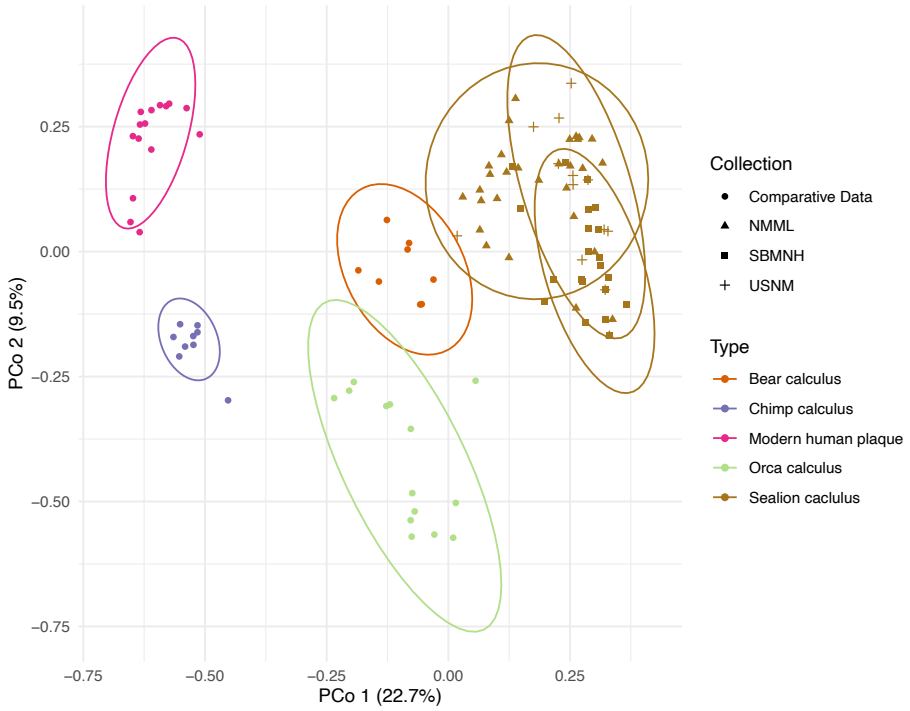



Figure 14: Mammal calculus. Unweighted unifrac beta diversity PCoA of mammalian calculus sources and the samples within this study. Mammal calculus sources are differentiated by color

and a 95% confidence ellipse. Collection is differentiated by shape. California sea lion calculus separates well from the other mammalian dental calculus sources.

We did not identify any observable dissimilarities between collections as well. In addition, we did not observe any beta diversity similarities between humans and non-human primates with the California sea lion specimens. California sea lions appear to have a distinct oral microbiome signature, except for a few individuals clustering with bear calculus and orca calculus. While we did not observe any overt collection dissimilarities (Figure 15A), significant difference between collection microbial communities were tested with PERMANOVA with 999 permutations (*p=0.001*) using the vegan package (v.2.6-4) (Oksanen *et al.* 2019) in R. A pairwise comparison revealed significant differences in the microbial community between all three collections using a Holm-Bonferroni corrected *p*-value (NMML and SBMNH $p=7.9e^{-07}$, NMML and USNM *p=0.0028*, SBMNH and USNM *p=0.0053*) (Figure 15C). We also explored the relationships between individuals' sea lions by collections, time, and location. While there is some clustering of Early Global Market individuals, this could be due the large number of individuals from that time frame within the dataset (Figure 15B).

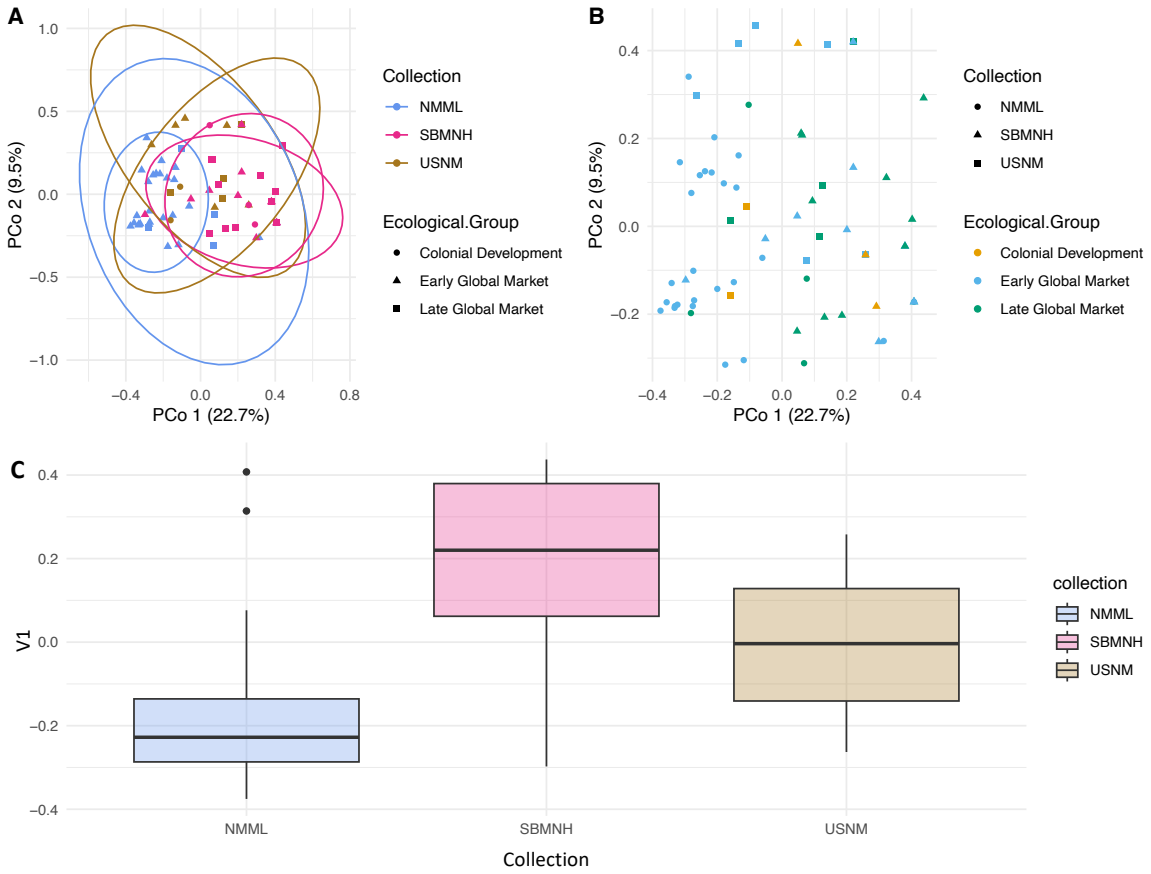



Figure 15: Collection comparison. A) Unweighted unifrac beta diversity PCoA of samples from the three collections. Collections are differentiated by color and a 95% confidence ellipse. Ecological groups are differentiated by shape. B) Unweighted unifrac beta diversity PCoA of every sample from the three collections. Collections are differentiated by shape and ecological

group by color. Significant difference between collection microbial communities were tested using PERMANOVA. C) Boxplot of PCoA variances to distinguish significant difference in collection.

For most samples, the reads assigned to the custom Kraken RefSeq database were assigned to *Zalophus californianus* and *Neisseria zalophi*. *Neisseria zalophi* is a novel species found only in the oral microbiome of California sea lions (Volokhov *et al.*, 2018). From the MetaPhlAN4 and Kraken analyses, we observed high abundance of the expected oral taxa *Neisseria zalophi* and *Streptococcus marimmalianis* and unexpected oral taxa *Tannerella forsythia and Desulfomicrobium orale*, which are associated with periodontal disease in humans(Dewhirst *et al.*, 2010). In addition, we discovered some unexpected taxa not common to the oral microbiome including *Enterobacteriacae* species and low abundance of *Toxoplasma gondii*.

Given that MetaPhlAN4 and Kraken are biased for human associated microbes, we had a large proportion of the microbial data unclassified (~90%). As there are more published 16S rRNA gene sequences from less common taxa, we also conducted 16S rRNA gene analysis. For comparative mapping we extracted reads mapping to the 16S rRNA gene from all the samples within this study and dental calculus sources used for the Sourcetracker analyses by mapping to the GG97 database using QIIME. A genus level OTU table was created and rarefied to 5000, as discussed previously. OTU read counts below 20% were filtered out from the OTU table to retain the 148 genus taxa found in more than 20% of each sample. We created a heatmap to show the abundance of taxa found within each species (sea lion, human, bear, chimp, and orca) calculus sample (Figure 16). Overall, between mammalian species, sea lions have a unique oral microbiome that clusters separately from orca, human, bear, and chimp oral microbiomes. These include *Burkholderia, Marinobacter, Lampropedia, Fusibacter, genera* and unknown genera from *Cardiobacteriacea* and *Peptostreptococcaceae,* to name a select few. Interestingly, sea lions cluster away from the orca microbiome, which clusters with humans and chimps, thus possibly suggesting that the environment may not be as influential on the microbial diversity in the oral microbiome. Compared to other mammalian species, we also did not observe a high unique abundance of a select number of genera within the sea lion oral microbiome. For instance, orcas have a high abundance of *Desulfococcus, Desulfovibrio, Methanobrevibacter, Dethiosulfovibrio* genera and unknown genera of *Desulfobulbaceae.* We did not observe any unique dissimilarities between collection and ecological periods suggesting that sea lions across time and collection curations have a similar oral microbiome signature.

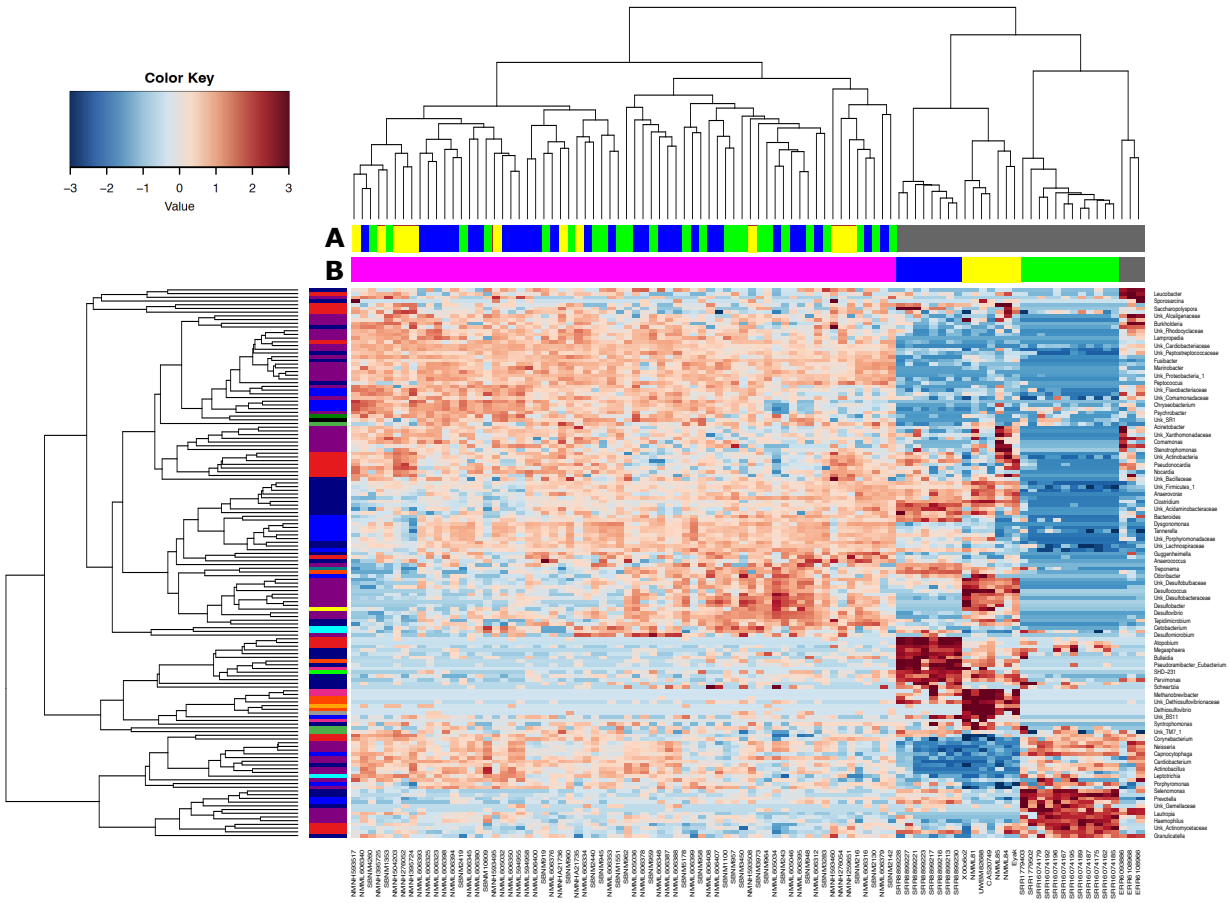



Figure 16: Heatmap of the top bacteria genera across mammalian oral microbiome samples. A) Heatmap of mammalian mapping (top dendrogram and color bar) – Pink represents sea lions, blue represents chimps, yellow is orca, green is human, and grey is bear. A log10 transformation with +1 constant to deal with zeros has been incorporated. Left side dendrogram and color bar is of the bacteria counts at the phylum level. B) Heatmap of collection mapping (top dendrogram and color bar), NMML is blue, SBMNH is green and USNM is yellow, comparative data from unknown collections are grey. A log10 transformation with +1 constant to deal with zeros has been incorporated. Left side dendrogram and color bar is of the bacteria counts at the phylum level.

### 4.5.4 Ecological impacts and microbial presence

To explore how human associated pathogens might have been introduced to the sea lion populations, we mapped to two commensals (*Neisseria zalophi* and *Streptococcus marimmaliani*) and four pathogens (*Tannerella forsythia and Desulfomicrobium orale, Enterobacteriacae* species and *Toxoplasma gondii*) based on identification from the MetaPhlAN4 and KRAKEN2 analyses, or the potential presence of these pathogens documented in the literature (Figure 17; Supplementals 5-10). *Neisseria zalophi* and *Streptococcus marimmaliani* were recovered in all samples, thus suggesting the two taxa may be associated with California sea lion core oral microbiome. Pathogenic oral taxa, such as *Tannerella forsythia* and *Desulfomicrobium orale* were also recovered in many of the samples,

however in lower read counts compared to *Neisseria zalophi* and *Streptococcus marimmaliani*. With each oral taxa, we noticed unique damage patterns not characteristic with typical ancient DNA patterns, where the damage is highest at the first position of the fragment sequence. In this case, damage is highest at the second and third position. This may be due to the limited number of reads available for mapping, where more than 10,000 reads may be necessary to consistently visualize damage patterns from microbial genomes (Mann *et al.*, 2020). This mismapping pattern was particularly observable among *T. forsythia, K. pneumoniae, and T. gondii.* We do not believe that these unique damage patterns are a result of contamination given that our negatives are clean and we have recovered the taxa across the three collections. However, it is unclear why the damage patterns are higher on the second and third position and will require further investigation. If it is connected to strain differences in pathogen reference genomes, de novo assembly and remapping to the new assembly might help this problem.

Of the non-oral pathogenic taxa, we mapped to *T. gondii* and *Enterobacteriacae* species, *C. werkmanii*, *E. cloacae*, *K. pneumoniae*, and *F. helveticus.* These microbial taxa were selected based on identification from the MetaPhlAN4 and KRAKEN2 analyses, and their prevalence in the literature. *T. gondii* has been recorded among California sea lions, but it has particularly been linked to mortality events observed in sea otters (Miller *et al.*, 2023). While we did have some samples containing numerous reads potentially indicative of its presence, the genome coverage was limited, with multiple reads mapping to the same region. In addition, while read counts can help signify the potential presence of a microbe, it is heavily biased by sequence depth and it not an accurate measure of pathogen presence. In mapping to the *Enterobacteriacae* species we found n=8 samples indicative of *K.pneumoniae* presence. In reviewing the unique rescaled bam files, reads were found to map across the *K.pneumoniae* genome. *K .pneumoniae* has been particularly observed in mortality events among California sea lions along San Miguel Island. Further investigation, particularly capture sequencing might be useful in identifying the prevalent strain.




Figure 17: Competitive mapping of commensal and pathogenic microbes. Read counts were limited to >5000 unique mapped reads for the 5' C to T damage lines for *N. zalophi, S. marimammalim, D. orale,* and *T. forsythia*. Read counts were limited to >1000 unique mapped reads for the 5' C to T damage lines for *K. pneumoniae* and *T. gondii*. Scatterplots for each microbe were plotted using the unique mapped read counts. While this is not a definitive measure for microbial presence due to differences in sequencing depth, we chose this measure for easier visualization of potential microbe presence over time.

## 4.6 CONCLUSION

We presented the first California sea lion dental calculus study. The aim of this study was to explore the impact of anthropogenic disruptions on California sea lion populations over the last ~130 years through the analysis of dental calculus. We had anticipated that given the scale of anthropogenic disruptions including, overhunting, wastewater runoff and sewage and consequential introduction of antibiotics into the marine water, along with the marine pollution would have resulted in measurable microbiome shifts among the study population. We therefore separated the samples into ecological bins each associated with different anthropogenic impacts

and changes to the marine ecology. In addition, we anticipated human pathogens to potentially be present because of wastewater runoff. However, we presented some challenges and future directions in alleviating these challenges so that these methods can be applied towards studying other mammalian oral microbiomes.

While we did not observe overt indicators of these transitions, we believe that some of these observations may have been limited due to collection curation methods and available reference database bias for non-human animal studies. From the quality assessment measures we noticed a significant difference between the average fragment length and damage patterns with NMML specimens containing higher damage at the first position and longer average fragment lengths. To further test that curation methods may be impacting these observations, we compared specimens collected and curated during the same time period (1970-1980) between SBMNH and NMML. Again, we observed a significant difference in damage and endogenous content but not in fragment length, thus suggesting that curation methods are contributing in some way to the quality of DNA preservation.

In addition to curation impacts, we were limited in characterizing the oral microbiome of sea lions due to the available reference databases, which are heavily influenced by human associated microbial reference data, and where ~90% of each sample's reads are unclassified in this study using MetaPhlAN4 and Kraken. Therefore, we were limited to observing microbiome changes within <10% of the assigned reads which may be influencing mapping quality and different taxonomic assignments. This is problematic given that MetaPhlAN4 uses a microbial gene database and KRAKEN2 Reqseq which does not have a deep breath of microbial genomes, as discussed previously. To provide better taxonomic resolution we extracted 16S reads and mapped them to the Greengenes database for OTU picking.

Of the definable microbiome, we observed that California sea lions have a unique microbiome signature that is different from other mammalian oral, in particular human and non-human primates. This signifies that although there have been anthropogenic changes to the environment over the last ~130 years, these changes have not been observed within the presented data analyses. In addition, we did not see any changes over time and location of collection origin that suggest there has been a loss in microbial diversity due anthropogenic disruptions. The current evidence presented in this study suggest that the oral microbiome may be more stable and less prone to microbial diversity alterations as observed with the gut microbiome due to diet and environmental impacts. However further investigation into alternative anthropogenic impacts, such as the introduction of antibiotic resistance genes will provide a better assessment and insight into how, if any, anthropogenic disruptions may be impacting the oral microbial diversity and potentially overall health of California sea lions. Future directions will require identifying the causation for unique damage patterns among microbial DNA and how these patterns compare to other marine mammal calculus data using de novo assembly, SNP filtering for phylogenetic tree building, and strain evolution. In addition, we will explore the presence of *K. pneumoniae* within the eight samples and look for alternative signatures of anthropogenic impacts, such as the presence of antibiotic resistance genes in the population.

# Chapter 5

# CONCLUSION

## 5.1 SUMMARY OF WORK

This dissertation took a multi-omic approach to study past human-animal-environment interactions through the lens of NCT. I provided a review of new tools that can be incorporated in One Health and argued the relevance of incorporating archaeology into One Health discussions. Viewing humans as niche constructors of past and future disease-scapes, I provided a theoretical approach on how disease-scapes are spillovers of ecological inheritances. Emerging infectious diseases we face today and, in the future, therefore have been built on successive rounds of anthropogenic disruptions for the past millennia. These have been exemplified through colonialism and globalization. I provide tools that can be combined to untangle the complex layering of disease-scapes in the past and highlight how researchers from different fields can easily work together to uncover unrecognized disease interactions. In addition, I provide a roadmap for tools that can be implemented in this type of research and how researchers can incorporate ethical and inclusive research with descendant communities. I address how new approaches in the "omics" are progressing our interpretation of the past and can be combined for a more holistic interpretation. Finally, I discuss the relevance of museum collections when archaeological excavations may not be possible. Museum collections have been used before to document the origin and phylogeny of past diseases. Though they remain underutilized, their application in disease research has grown given that they offer a series of snapshots in time, documenting past ecologies and are generally associated with metadata.

The integration of diverse datasets across different timescales can be challenging, especially when testing for statistical significance. This can include, but is not limited to, data scale mismatches, unbalanced data, and sampling bias (Zipkin *et al.*, 2021). For instance large scale datasets covering diverse spatiotemporal scales often have spatiotemporal gaps along with unbalanced data as a result of collection bias and can therefore lead to artificial inflations of samples sizes and biased inferences (Zipkin *et al.*, 2021). Researchers still face this problem today as observational datasets (photos, videos, GPS coordinates) outnumber primary data from physical and voucher specimens (Salvador and Cunha, 2020; MacCormick, 2023). Furthermore, data and sample collection bias has been observed in favoring regions (mainly North America and Europe) and organisms that may be more interesting for collection (MacCormick, 2023). Building transdisciplinary research and integrating diverse datasets from different timescales therefore requires standardized methods (Rick and Lockwood, 2013) and a well-constructed model to study the ecological process of interest. When working with archaeological, paleoecological, and zooarchaeological datasets we are limited by the type of observable data that can be analyzed due to taphonomy and preservation. Well-constructed models covering large spaciotemporal scales must therefore be local and context specific. I use many of the techniques framed in chapter 2 for chapters 3 and 4 – proteomic and genomic research with museum collections.

In chapter 3, I offered a proteomic approach that can be incorporated in sexing animals. This technique is of relevance to zooarchaeological assemblages where animal remains are either highly fragmented due to consumption and butchering techniques, taphonomic changes, or mixed with other faunal assemblages where neither species nor sex can be determined. In addition, for species that exhibit little sexual dimorphism, alternative methods are needed. Proteins have shown to preserve better than DNA and have been used as an alternative method for species identification using some protein methods likes ZooMS on the alpha collagen 1 protein. Therefore, proteomic methods have been used as a complementary method to aDNA applications. The ability to determine sex from zooarchaeological assemblages can be instrumental when analyzing early pastoralism or domestication patterns. This is normally used focusing on osteometric bones or sex ratio profiles. Yet these methods have revealed some sex biases in museum collections, misidentifying female specimens as males. The proteomic approach I took is less expensive and less destructive than ancient DNA analyses and has shown to accurately determine sex in archaeological human remains. I highlight the complications with this technique in determining sex from museum specimens with an age range of 10-40 years. This is most likely due to the approach taken and the preservation of the sample. While I practice ancient standards for the process, assuming these samples would have some damage due to curatorial methods, I suggest that the protein probably had very little damage. I routinely recovered peptides from the N-terminal and C-terminal domains that are more stable and compared to the variable region of the AMELX and AMELY protein. This area is generally shorter than the variable region and given the methodological approach taken – without digestion – may have not been fully recovered from the mass spectrometer column. However, I suggest that a digestion stage may be necessary for this approach to recover other variable peptides that can be used for sex determination.

Finally, I took a metagenomic approach incorporating animal remains from museum collections to address anthropogenic disruptions to the environment and marine ecology. In this study I use California sea lion dental calculus as a tool to uncover shifting historical baselines surrounding the Channel Islands. Archaeologists and historical ecologists have long documented the deep relationship and anthropogenic impacts among humans and marine mammals in this region. The Channel Islands in particular have been inhabited for the last 13,000 years and there has been evidence for strong maritime technologies within the region. This study in particular focuses on the last ~130 years where massive changes from colonization to globalization events have greatly impacted the marine and terrestrial ecology. Of importance, coastal runoff has been shown to increase the introduction of antibiotics into the environment, leading to environmental selection of antibiotic resistance within the marine biome (Hatosy and Martiny, 2015; Almakki *et al.*, 2019). Historical alterations to the marine biome linked to the conversion to upland habitat and pollutant runoff as well as the historical exploitation of marine species, specifically California sea lions (*Zalophus californianus*), has inevitably fostered modern marine mammal populations heavily shaped by centuries of anthropogenic disruptions (Crain *et al.*, 2009; Braje and Rick, 2011; Hatosy and Martiny, 2015). The further introduction of antibiotic resistant human pathogens into this environment coupled with the similarities in human and marine mammal biology, consequently, increases the potential for associated epizootic transmission of human diseases to marine mammals (Hatosy and Martiny, 2015).

In this study of 75 California sea lion specimens from three different collections – the Santa Barbara Museum of Natural History (SBMNH), the National Marine Mammals Laboratory (NMML) collection and the historic Smithsonian collection (USNM) – I built oral microbiome profiles from metagenomic sequencing data to address any anthropogenic alterations to the California sea lion microbial diversity. While there were no overt signatures in humanization events or loss of microbial diversity that could be linked to ecological changes over the last ~130 years, I attribute some of the challenges in documenting any microbiome changes to curatorial methods and human biased reference databases. Alternatively, I highlight the unique oral signature among California sea lions and discuss the lack of observable alterations to the oral microbiome over the last ~130 years. I also address the recovery of both core oral microbes and oral pathogens associated with periodontal disease, such as *Tannerella forsythia*. Further investigation of the recovered microbes revealed the potential presence of *Klebsiella pneumoniae* within eight samples. This will require further investigation and capture sequencing as *Klebsiella pneumoniae* has been linked to mortality events among California sea lions within the Channel Islands, particularly along San Miguel Island.

## 5.2 FUTURE DIRECTIONS

Throughout this dissertation, I have emphasized the importance of museum collections and multi-omic approaches to study human-animal-environment relations. While the recovery of sex specific peptides from the proteomic approach taken in Chapter 3 prove to be insufficient to sex some animal species, incorporating a digestive step or a top-down proteomic approach will alleviate this challenge and help with the recovery of the variable region for sex determination for future research. Within this study, further investigation of the mass peaks that have been used to identify sex among human males and females may also prove valuable. This would require the deconvolution of the m/z spectrum. In addition, it will be useful to document any deamination patterns to see if we do or do not see curatorial methods impacting protein preservation as observed with DNA within Chapter 4.

Numerous avenues can be taken to proceed with observing any anthropogenic impacts on the California sea lion oral microbiome. While the oral microbiome of California sea lions appears to have remained intact over time, it may be beneficial to further explore any similarities between human and marine mammal oral microbiomes as a comparative measure for studying mammalian oral microbiomes. This would require further documenting both historic and modern human oral microbiomes to record which shared oral taxa are commensal versus pathogenic. This can be further explored to documenting similarities and differences among other marine mammal oral microbiomes to determine if any ecological factors influence the colonization of marine mammal oral microbiomes and the development of dental calculus. Given the unusual damage patterns observed among the microbial DNA, further investigation to determine if the same patterns can be observed among other marine mammals is of interest. In addition, it would be interesting to see if we notice any difference in microbial DNA damage among marine mammals compared to terrestrial mammals. For instance, does the marine environment impact DNA preservation in any way? Finally, while loss of microbial diversity was not demonstrated over the recent ~130 years within this study and could therefore not be linked to ecological transitions, alternative methods looking for the presence of certain ecological impacts may shed light on this particular question. This would include documenting the presence of antibiotic

resistance genes in the collection that can be associated with environmental disruptions. Furthermore, the characterization of the *Klebsiella pneumoniae* strain observed in the eight samples may shed light on the epidemiology of *Klebsiella pneumoniae* within the California sea lion population. Incorporating these multi-omic approaches with animal remains, particularly within museum collections that have associated meta data, can provide alternative pathways in addressing dynamic human-animal-environment relations for One Health research.

# REFERENCES


'About Us' (no date) *Native BioData Consortium*. Available at: https://nativebio.org/about-us/ (Accessed: 1 March 2023).

Acevedo-Whitehouse, K. *et al.* (2003) 'Disease susceptibility in California sea lions', *Nature*, 422(6927), pp. 35–35. Available at: https://doi.org/10.1038/422035a.

Afelt, A., Frutos, R. and Devaux, C. (2018) 'Bats, Coronaviruses, and Deforestation: Toward the Emergence of Novel Infectious Diseases?', *Frontiers in Microbiology*, 9, p. 702. Available at: https://doi.org/10.3389/fmicb.2018.00702.

Agarwal, S.C. (2012) 'The Past of Sex, Gender, and Health: Bioarchaeology of the Aging Skeleton', *American Anthropologist*, 114(2), pp. 322–335. Available at: https://doi.org/10.1111/j.1548-1433.2012.01428.x.

Albery, G.F. *et al.* (2022) 'Urban-adapted mammal species have more known pathogens', *Nature Ecology & Evolution*, 6(6), pp. 794–801. Available at: https://doi.org/10.1038/s41559-022-01723-0.

Ali, S. *et al.* (2017) 'Environmental and Social Change Drive the Explosive Emergence of Zika Virus in the Americas', *PLoS Neglected Tropical Diseases*, 11(2), p. e0005135. Available at: https://doi.org/10.1371/journal.pntd.0005135.

Allen, H.K. and Stanton, T.B. (2014) 'Altered egos: antibiotic effects on food animal microbiomes', *Annual Review of Microbiology*, 68, pp. 297–315. Available at: https://doi.org/10.1146/annurev-micro-091213-113052.

Almakki, A. *et al.* (2019) 'Antibiotic resistance in urban runoff', *The Science of the Total Environment*, 667, pp. 64–76. Available at: https://doi.org/10.1016/j.scitotenv.2019.02.183.

Armelagos, G.J. (2009) 'The paleolithic disease-scape, the hygiene hypothesis, and the second epidemiological transition', in G.A.W. Rook (ed.) *The Hygiene Hypothesis and Darwinian Medicine*. Basel: Birkhäuser (Progress in Inflammation Research), pp. 29–43. Available at: https://doi.org/10.1007/978-3-7643-8903-1_2.

Armelagos, G.J., Barnes, K.C. and Lin, J. (1996) 'Disease in Human Evolution: The Reemergence of Infectious Disease in the Third Epidemiological Transition'. Available at: http://repository.si.edu/xmlui/handle/10088/22354 (Accessed: 30 November 2022).

Armstrong, C.G. *et al.* (2017) 'Anthropological contributions to historical ecology: 50 questions, infinite prospects', *PLOS ONE*. Edited by A. Parmakelis, 12(2), p. e0171883. Available at: https://doi.org/10.1371/journal.pone.0171883.

Arroyo-Kalin, M. (2012) 'Slash-burn-and-churn: Landscape history and crop cultivation in pre-Columbian Amazonia', *Quaternary International*, 249, pp. 4–18. Available at: https://doi.org/10.1016/j.quaint.2011.08.004.

Atlas, R.M. and Maloy, S. (2014) 'The Future of One Health', *Microbiology Spectrum*, 2(1). Available at: https://doi.org/10.1128/microbiolspec.OH-0018-2012.

Austin, R.M. *et al.* (2019) 'To curate the molecular past, museums need a carefully considered set of best practices', *Proceedings of the National Academy of Sciences*, 116(5), pp. 1471–1474. Available at: https://doi.org/10.1073/pnas.1822038116.

Austin, R.M. *et al.* (2022) 'Remembering St. Louis Individual—structural violence and acute bacterial infections in a historical anatomical collection', *Communications Biology*, 5(1), pp. 1–10. Available at: https://doi.org/10.1038/s42003-022-03890-z.

Bahrndorff, S. *et al.* (2016) 'The Microbiome of Animals: Implications for Conservation Biology', *International Journal of Genomics*, 2016, p. 5304028. Available at: https://doi.org/10.1155/2016/5304028.

Bansal, A.K. *et al.* (2012) 'Amelogenin: A novel protein with diverse applications in genetic and molecular profiling', *Journal of Oral and Maxillofacial Pathology : JOMFP*, 16(3), pp. 395–399. Available at: https://doi.org/10.4103/0973-029X.102495.

Baron, S.A., Diene, S.M. and Rolain, J.-M. (2018) 'Human microbiomes and antibiotic resistance', *Human Microbiome Journal*, 10, pp. 43–52. Available at: https://doi.org/10.1016/j.humic.2018.08.005.

Barrett, R. *et al.* (1998) 'Emerging and Re-Emerging Infectious Diseases: The Third Epidemiologic Transition', *Annual Review of Anthropology*, 27(1), pp. 247–271.

Bendrey, R. *et al.* (2020) 'Approaching ancient disease from a ONE HEALTH perspective: Interdisciplinary review for the investigation of zoonotic brucellosis', *International Journal of Osteoarchaeology*, 30(1), pp. 99–108. Available at: https://doi.org/10.1002/oa.2837.

Benn, A. *et al.* (2018) 'Studying the human oral microbiome: challenges and the evolution of solutions', *Australian Dental Journal*, 63(1), pp. 14–24. Available at: https://doi.org/10.1111/adj.12565.

Bird, D. *et al.* (2022) 'p3k14c, a synthetic global database of archaeological radiocarbon dates', *Scientific Data*, 9(1), p. 27. Available at: https://doi.org/10.1038/s41597-022-01118-7.

Blanco-Miguez, A. *et al.* (2022) 'Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4'. bioRxiv, p. 2022.08.22.504593. Available at: https://doi.org/10.1101/2022.08.22.504593.

Bocquet-Appel, J.-P., Naji, S. and Bandy, M. (2008) 'Demographic and Health Changes During the Transition to Agriculture in North America', in J.-P. Bocquet-Appel (ed.) *Recent Advances in Palaeodemography: Data, Techniques, Patterns*. Dordrecht: Springer Netherlands, pp. 277–292. Available at: https://doi.org/10.1007/978-1-4020-6424-1_10.

Boivin, N.L. *et al.* (2016) 'Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions', *Proceedings of the National*

*Academy of Sciences*, 113(23), pp. 6388–6396. Available at: https://doi.org/10.1073/pnas.1525200113.

Borić, D. and Price, T.D. (2013) 'Strontium isotopes document greater human mobility at the start of the Balkan Neolithic', *Proceedings of the National Academy of Sciences*, 110(9), pp. 3298–3303. Available at: https://doi.org/10.1073/pnas.1211474110.

Bos, K.I. *et al.* (2011) 'A draft genome of Yersinia pestis from victims of the Black Death', *Nature*, 478(7370), pp. 506–510. Available at: https://doi.org/10.1038/nature10549.

Bos, K.I. *et al.* (2014) 'Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis', *Nature*, 514(7523), pp. 494–497. Available at: https://doi.org/10.1038/nature13591.

Braje, T.J. and Erlandson, J.M. (2008) 'Early Maritime Technology from Western San Miguel Island, California'.

Braje, T.J. and Rick, T.C. (2011) *Human impacts on seals, sea lions, and sea otters: integrating archaeology and ecology in the Northeast Pacific*. Univ of California Press.

Brassard, C. and Callou, C. (2020) 'Sex determination of archaeological dogs using the skull: evaluation of morphological and metric traits on various modern breeds', *Journal of Archaeological Science: Reports*, 31, p. 102294. Available at: https://doi.org/10.1016/j.jasrep.2020.102294.

Brealey, J.C. *et al.* (2021) 'The oral microbiota of wild bears in Sweden reflects the history of antibiotic use by humans', *Current Biology*, 31(20), pp. 4650-4658.e6. Available at: https://doi.org/10.1016/j.cub.2021.08.010.

Brickley, M.B. and Mays, S. (2019) 'Chapter 15 - Metabolic Disease', in J.E. Buikstra (ed.) *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)*. San Diego: Academic Press, pp. 531–566. Available at: https://doi.org/10.1016/B978-0-12-809738-0.00015-6.

Bro-Jørgensen, M.H. *et al.* (2021) 'Genomic sex identification of ancient pinnipeds using the dog genome', *Journal of Archaeological Science*, 127, p. 105321. Available at: https://doi.org/10.1016/j.jas.2020.105321.

Buckley, M. *et al.* (2009) 'Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry', *Rapid Communications in Mass Spectrometry*, 23(23), pp. 3843–3854. Available at: https://doi.org/10.1002/rcm.4316.

Buckley, M., Collins, M. and Thomas-Oates, J. (2008) 'A method of isolating the collagen (I) alpha2 chain carboxytelopeptide for species identification in bone fragments', *Analytical Biochemistry*, 374(2), pp. 325–334. Available at: https://doi.org/10.1016/j.ab.2007.12.002.

Buckley, M. and Collins, M.J. (2011) 'Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites', *Antiqua*, 1(1), pp. e1–e1. Available at: https://doi.org/10.4081/antiqua.2011.e1.

Buonasera, T. *et al.* (2020) 'A comparison of proteomic, genomic, and osteological methods of archaeological sex estimation', *Scientific Reports*, 10, p. 11897. Available at: https://doi.org/10.1038/s41598-020-68550-w.

Burke, A. *et al.* (2021) 'The archaeology of climate change: The case for cultural diversity', *Proceedings of the National Academy of Sciences*, 118(30), p. e2108537118. Available at: https://doi.org/10.1073/pnas.2108537118.

Campana, M.G. *et al.* (2017) 'White-Nose Syndrome Fungus in a 1918 Bat Specimen from France', *Emerging Infectious Diseases*, 23(9), pp. 1611–1612. Available at: https://doi.org/10.3201/eid2309.170875.

Caporaso, J.G. *et al.* (2010) 'QIIME allows analysis of high-throughput community sequencing data', *Nature Methods*, 7(5), pp. 335–336. Available at: https://doi.org/10.1038/nmeth.f.303.

Cappellini, E. *et al.* (2012) 'Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient Bone Proteins', *Journal of Proteome Research*, 11(2), pp. 917–926. Available at: https://doi.org/10.1021/pr200721u.

Cappellini, E. *et al.* (2018) 'Ancient Biomolecules and Evolutionary Inference', *Annual Review of Biochemistry*, 87(1), pp. 1029–1060. Available at: https://doi.org/10.1146/annurev-biochem-062917-012002.

Cappellini, E. *et al.* (2019) 'Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny', *Nature*, 574(7776), pp. 103–107. Available at: https://doi.org/10.1038/s41586-019-1555-y.

Cappellini, E., Collins, M.J. and Gilbert, M.T.P. (2014) 'Unlocking Ancient Protein Palimpsests', *Science*, 343(6177), pp. 1320–1322. Available at: https://doi.org/10.1126/science.1249274.

Carlson, C.J. *et al.* (2021) 'The future of zoonotic risk prediction', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1837), p. 20200358. Available at: https://doi.org/10.1098/rstb.2020.0358.

Carøe, C. *et al.* (2018) 'Single-tube library preparation for degraded DNA', *Methods in Ecology and Evolution*, 9(2), pp. 410–419. Available at: https://doi.org/10.1111/2041-210X.12871.

Carroll, S.R. *et al.* (2020) 'The CARE Principles for Indigenous Data Governance', *Data Science Journal*, 19(1), p. 43. Available at: https://doi.org/10.5334/dsj-2020-043.

Chin, A. *et al.* (2020) 'Pandemics and the future of human-landscape interactions', *Anthropocene*, 31, p. 100256. Available at: https://doi.org/10.1016/j.ancene.2020.100256.

Claw, K.G. *et al.* (2018) 'A framework for enhancing ethical genomic research with Indigenous communities', *Nature Communications*, 9(1), p. 2957. Available at: https://doi.org/10.1038/s41467-018-05188-3.

Clayton, E. *et al.* (2022) 'Structural Bases of Zoonotic and Zooanthroponotic Transmission of SARS-CoV-2', *Viruses*, 14(2), p. 418. Available at: https://doi.org/10.3390/v14020418.

Clayton, J.B. *et al.* (2016) 'Captivity humanizes the primate microbiome', *Proceedings of the National Academy of Sciences*, 113(37), pp. 10376–10381. Available at: https://doi.org/10.1073/pnas.1521835113.

Cockburn, T.A. (1971) 'Infectious Diseases in Ancient Populations', *Current Anthropology*, 12(1), pp. 45–62. Available at: https://doi.org/10.1086/201168.

Cohen, M.N. (1989) *Health and the Rise of Civilization*. Yale University Press.

Colella, J.P. *et al.* (2021) 'Leveraging natural history biorepositories as a global, decentralized, pathogen surveillance network', *PLOS Pathogens*, 17(6), p. e1009583. Available at: https://doi.org/10.1371/journal.ppat.1009583.

Cooper, N. *et al.* (2019) 'Sex biases in bird and mammal natural history collections', *Proceedings of the Royal Society B: Biological Sciences*, 286(1913), p. 20192025. Available at: https://doi.org/10.1098/rspb.2019.2025.

Coutu, A.N. *et al.* (2021) 'Palaeoproteomics confirm earliest domesticated sheep in southern Africa ca. 2000 BP', *Scientific Reports*, 11(1), p. 6631. Available at: https://doi.org/10.1038/s41598-021-85756-8.

Crain, C.M. *et al.* (2009) 'Understanding and managing human threats to the coastal marine environment', *Annals of the New York Academy of Sciences*, 1162, pp. 39–62. Available at: https://doi.org/10.1111/j.1749-6632.2009.04496.x.

Crema, E.R. and Shoda, S. (2021) 'A Bayesian approach for fitting and comparing demographic growth models of radiocarbon dates: A case study on the Jomon-Yayoi transition in Kyushu (Japan)', *PLOS ONE*, 16(5), p. 26.

Crutzen, P.J. and Stoermer, E.F. (2000) 'The "Anthropocene"', *Global Change Newsletter*, 41, pp. 17–18.

Cucchi, T. and Arbuckle, B. (2021) 'Animal domestication: from distant past to current development and issues', *Animal Frontiers*, 11(3), pp. 6–9. Available at: https://doi.org/10.1093/af/vfab013.

Cucchi, T., Vigne, J.-D. and Auffray, J.-C. (2005) 'First occurrence of the house mouse (Mus musculus domesticus Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences: HOUSE MOUSE DIFFUSION IN THE MEDITERRANEAN AREA', *Biological Journal of the Linnean Society*, 84(3), pp. 429–445. Available at: https://doi.org/10.1111/j.1095-8312.2005.00445.x.

Dallongeville, S. *et al.* (2013) 'Proteomics applied to the authentication of fish glue: application to a 17th century artwork sample', *The Analyst*, 138(18), pp. 5357–5364. Available at: https://doi.org/10.1039/c3an00786c.

Dávalos, L.M. *et al.* (2020) 'Pandemics' historical role in creating inequality', *Science*. Edited by J. Sills, 368(6497), pp. 1322–1323. Available at: https://doi.org/10.1126/science.abc8953.

Davis, S.J.M. *et al.* (2012) 'Molecular and osteometric sexing of cattle metacarpals: a case study from 15th century AD Beja, Portugal', *Journal of Archaeological Science*, 39(5), pp. 1445–1454. Available at: https://doi.org/10.1016/j.jas.2011.12.003.

Destoumieux-Garzón, D. *et al.* (2018) 'The One Health Concept: 10 Years Old and a Long Road Ahead', *Frontiers in Veterinary Science*, 5. Available at: https://www.frontiersin.org/articles/10.3389/fvets.2018.00014 (Accessed: 31 January 2023).

Dewhirst, F.E. *et al.* (2010) 'The Human Oral Microbiome', *Journal of Bacteriology*, 192(19), pp. 5002–5017. Available at: https://doi.org/10.1128/JB.00542-10.

DiEuliis, D. *et al.* (2016) 'Specimen collections should have a much bigger role in infectious disease research and response', *Proceedings of the National Academy of Sciences*, 113(1), pp. 4–7. Available at: https://doi.org/10.1073/pnas.1522680112.

Dobson, A.P. and Carper, E.R. (1996) 'Infectious Diseases and Human Population History', *BioScience*, 46(2), pp. 115–126. Available at: https://doi.org/10.2307/1312814.

Dunnum, J.L. *et al.* (2017) 'Biospecimen Repositories and Integrated Databases as Critical Infrastructure for Pathogen Discovery and Pathobiology Research', *PLOS Neglected Tropical Diseases*, 11(1), p. e0005133. Available at: https://doi.org/10.1371/journal.pntd.0005133.

Eads, D.A. and Biggins, D.E. (2015) 'Plague bacterium as a transformer species in prairie dogs and the grasslands of western North America', *Conservation Biology: The Journal of the Society for Conservation Biology*, 29(4), pp. 1086–1093. Available at: https://doi.org/10.1111/cobi.12498.

Ellis, E.C. (2021) 'Land Use and Ecological Change: A 12,000-Year History', *Annual Review of Environment and Resources*, 46(1), pp. 1–33. Available at: https://doi.org/10.1146/annurev-environ-012220-010822.

Ellis, E.C. *et al.* (2021) 'People have shaped most of terrestrial nature for at least 12,000 years', *Proceedings of the National Academy of Sciences*, 118(17), p. e2023483118. Available at: https://doi.org/10.1073/pnas.2023483118.

Epstein, J.H. *et al.* (2006) 'Nipah virus: impact, origins, and causes of emergence', *Current Infectious Disease Reports*, 8(1), pp. 59–65. Available at: https://doi.org/10.1007/s11908-006-0036-2.

Erlandson, J.M. *et al.* (2011) 'Paleoindian Seafaring, Maritime Technologies, and Coastal Foraging on California's Channel Islands', *Science*, 331(6021), pp. 1181–1185. Available at: https://doi.org/10.1126/science.1201477.

Ezenwa, V.O. *et al.* (2012) 'Animal Behavior and the Microbiome', *Science*, 338(6104), pp. 198–199. Available at: https://doi.org/10.1126/science.1227412.

Fagre, A.C. *et al.* (2022) 'Assessing the risk of human-to-wildlife pathogen transmission for conservation and public health', *Ecology Letters*. Edited by R. Ostfeld, 25(6), pp. 1534–1549. Available at: https://doi.org/10.1111/ele.14003.

*FAO - News Article: Rinderpest eradicated - what next?* (no date). Available at: https://www.fao.org/news/story/en/item/80894/icode/ (Accessed: 28 February 2023).

Fellows Yates, J.A. *et al.* (2021) 'The evolution and changing ecology of the African hominid oral microbiome', *Proceedings of the National Academy of Sciences*, 118(20), p. e2021655118. Available at: https://doi.org/10.1073/pnas.2021655118.

Fernandes, M.R. *et al.* (2018) 'Zooanthroponotic Transmission of Drug-Resistant Pseudomonas aeruginosa, Brazil', *Emerging Infectious Diseases*, 24(6), pp. 1160–1162. Available at: https://doi.org/10.3201/eid2406.180335.

Fournié, G., Pfeiffer, D.U. and Bendrey, R. (2017) 'Early animal farming and zoonotic disease dynamics: modelling brucellosis transmission in Neolithic goat populations', *Royal Society Open Science*, 4(2), p. 160943. Available at: https://doi.org/10.1098/rsos.160943.

Froment, C. *et al.* (2020) 'Analysis of 5000 year-old human teeth using optimized large-scale and targeted proteomics approaches for detection of sex-specific peptides', *Journal of Proteomics*, 211, p. 103548. Available at: https://doi.org/10.1016/j.jprot.2019.103548.

Fuentes, A. and Baynes-Rock, M. (2017) 'Anthropogenic Landscapes, Human Action and the Process of Co-Construction with other Species: Making Anthromes in the Anthropocene', *Land*, 6(1), p. 15. Available at: https://doi.org/10.3390/land6010015.

Gibbs, E.P.J. and Greiner, E.C. (1994) 'The epidemiology of bluetongue', *Comparative Immunology, Microbiology and Infectious Diseases*, 17(3–4), pp. 207–220. Available at: https://doi.org/10.1016/0147-9571(94)90044-2.

Gifford-Gonzalez, D. (2018) 'Bone's Intrinsic Traits: Inferring Species, Sex, and Age', in D. Gifford-Gonzalez (ed.) *An Introduction to Zooarchaeology*. Cham: Springer International Publishing, pp. 103–123. Available at: https://doi.org/10.1007/978-3-319-65682-3_6.

Ginolhac, A. *et al.* (2011) 'mapDamage: testing for damage patterns in ancient DNA sequences', *Bioinformatics*, 27(15), pp. 2153–2155. Available at: https://doi.org/10.1093/bioinformatics/btr347.

*Global Paleofire Database* (no date). Available at: https://www.paleofire.org/ (Accessed: 1 February 2023).

Goldewijk, K.K. *et al.* (2017) 'Anthropogenic land use estimates for the Holocene – HYDE 3.2', *Earth System Science Data*, 9(2), pp. 927–953. Available at: https://doi.org/10.5194/essd-9-927-2017.

Goodman, A.H. *et al.* (1984) 'Health Changes at Dickson Mounds, Illinois (950-1300AD)', in M. Cohen and G. Armelagos (eds) *Palaeopathology and the Origins of Agriculture*. N.Y.: Academic Press, pp. 217–305.

Gower, G. *et al.* (2019) 'Widespread male sex bias in mammal fossil and museum collections', *Proceedings of the National Academy of Sciences*, 116(38), pp. 19019–19024. Available at: https://doi.org/10.1073/pnas.1903275116.

Grauer, A.L. and Buikstra, J.E. (2019) 'Chapter 3 - Themes in Paleopathology', in J.E. Buikstra (ed.) *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)*. San Diego: Academic Press, pp. 21–33. Available at: https://doi.org/10.1016/B978-0-12-809738-0.00003-X.

Grauer, A.L. and Roberts, C.A. (2019) 'Chapter 12 - Fungal, Viral, Multicelled Parasitic, and Protozoan Infections', in J.E. Buikstra (ed.) *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)*. San Diego: Academic Press, pp. 441–478. Available at: https://doi.org/10.1016/B978-0-12-809738-0.00012-0.

Grayson, D.K. (2001) 'The Archaeological Record of Human Impacts on Animal Populations', *Journal of World Prehistory*, p. 68.

Green, D.R. *et al.* (2019) 'Mapping the Tooth Enamel Proteome and Amelogenin Phosphorylation Onto Mineralizing Porcine Tooth Crowns', *Frontiers in Physiology*, 10, p. 925. Available at: https://doi.org/10.3389/fphys.2019.00925.

Gregoricka, L.A. (2021) 'Moving Forward: A Bioarchaeology of Mobility and Migration', *Journal of Archaeological Research*, 29(4), pp. 581–635. Available at: https://doi.org/10.1007/s10814-020-09155-9.

Guellil, M. *et al.* (2022) 'Ancient herpes simplex 1 genomes reveal recent viral structure in Eurasia', *Science Advances*, 8(30), p. eabo4435. Available at: https://doi.org/10.1126/sciadv.abo4435.

Haas, R. *et al.* (2020) 'Female hunters of the early Americas', *Science Advances*, 6(45), p. eabd0310. Available at: https://doi.org/10.1126/sciadv.abd0310.

Han, B.A. *et al.* (2015) 'Rodent reservoirs of future zoonotic diseases', *Proceedings of the National Academy of Sciences*, 112(22), pp. 7039–7044. Available at: https://doi.org/10.1073/pnas.1501598112.

Harper, K. and Armelagos, G. (2010) 'The Changing Disease-Scape in the Third Epidemiological Transition', *International Journal of Environmental Research and Public Health*, 7(2), pp. 675–697. Available at: https://doi.org/10.3390/ijerph7020675.

Harper, K.N. *et al.* (2013) 'Primates, Pathogens, and Evolution: A Context for Understanding Emerging Disease', in J.F. Brinkworth and K. Pechenkina (eds) *Primates, Pathogens, and Evolution*. New York, NY: Springer (Developments in Primatology: Progress and Prospects), pp. 389–409. Available at: https://doi.org/10.1007/978-1-4614-7181-3_13.

Harper, K.N. and Armelagos, G.J. (2013) 'Genomics, the origins of agriculture, and our changing microbe-scape: Time to revisit some old tales and tell some new ones: Genomics, Agriculture, and Human Microbes', *American Journal of Physical Anthropology*, 152, pp. 135–152. Available at: https://doi.org/10.1002/ajpa.22396.

Hatosy, S.M. and Martiny, A.C. (2015) 'The ocean as a global reservoir of antibiotic resistance genes', *Applied and Environmental Microbiology*, 81(21), pp. 7593–7599. Available at: https://doi.org/10.1128/AEM.00736-15.

Hawkins, S., O'Connor, S. and Kealy, S. (2016) 'Late Quaternary hominin-bat (Chiroptera) interactions in the Asia-Pacific: Late Quaternary hominin-bat interactions in the Asia-Pacific', *Archaeology in Oceania*, 51(1), pp. 7–17. Available at: https://doi.org/10.1002/arco.5084.

Hendy, J. *et al.* (2018) 'Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers', *Nature Communications*, 9(1), p. 4064. Available at: https://doi.org/10.1038/s41467-018-06335-6.

Hendy, J. (2021) 'Ancient protein analysis in archaeology', *Science Advances*, 7(3), p. eabb9314. Available at: https://doi.org/10.1126/sciadv.abb9314.

Heymann, D.L. and Dixon, M. (2013) 'The Value of the One Health Approach: Shifting from Emergency Response to Prevention of Zoonotic Disease Threats at Their Source', *Microbiology Spectrum*. Edited by R.M. Atlas and S. Maloy, 1(1), p. 1.1.09. Available at: https://doi.org/10.1128/microbiolspec.OH-0011-2012.

Hillson, S. (1996) *Dental Anthropology*. Cambridge University Press.

Hofman, C.A. *et al.* (2015) 'Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene', *Trends in Ecology & Evolution*, 30(9), pp. 540–549. Available at: https://doi.org/10.1016/j.tree.2015.06.008.

Hofman, C.A. *et al.* (2018) 'Collagen Fingerprinting and the Earliest Marine Mammal Hunting in North America', *Scientific Reports*, 8(1), p. 10014. Available at: https://doi.org/10.1038/s41598-018-28224-0.

Hofman, C.A. (2018) 'Reconstruction of the Killer Whale Oral Microbiome'. *SMBE in Tokyo*.

Hofman, C.A. and Rick, T.C. (2018) 'Ancient Biological Invasions and Island Ecosystems: Tracking Translocations of Wild Plants and Animals', *Journal of Archaeological Research*, 26(1), pp. 65–115. Available at: https://doi.org/10.1007/s10814-017-9105-3.

*Home* (no date) *Mukurtu CMS*. Available at: https://mukurtu.org/ (Accessed: 1 March 2023).

Hooper, L.V., Littman, D.R. and Macpherson, A.J. (2012) 'Interactions between the microbiota and the immune system', *Science (New York, N.Y.)*, 336(6086), pp. 1268–1273. Available at: https://doi.org/10.1126/science.1223490.

Hrovatin, K. and Kunej, T. (2018) 'Genetic sex determination assays in 53 mammalian species: Literature analysis and guidelines for reporting standardization', *Ecology and Evolution*, 8(2), pp. 1009–1018. Available at: https://doi.org/10.1002/ece3.3707.

Hubálek, Z. (2003) 'Emerging Human Infectious Diseases: Anthroponoses, Zoonoses, and Sapronoses', *Emerging Infectious Diseases*, 9(3), pp. 403–404. Available at: https://doi.org/10.3201/eid0903.020208.

Inhorn, M.C. and Brown, P.J. (1990) 'The Anthropology of Infectious Disease', *Annual Review of Anthropology*, 19, pp. 89–117.

Iwase, M. *et al.* (2007) 'Evolutionary History of Sex-Linked Mammalian Amelogenin Genes', *Cells Tissues Organs*, 186(1), pp. 49–59. Available at: https://doi.org/10.1159/000102680.

Izdebski, A. (no date) 'Palaeoecological data indicates land-use changes across Europe linked to spatial heterogeneity in mortality during the Black Death pandemic', p. 17.

Jones, B.A. *et al.* (2013) 'Zoonosis emergence linked to agricultural intensification and environmental change', *Proceedings of the National Academy of Sciences*, 110(21), pp. 8399–8404. Available at: https://doi.org/10.1073/pnas.1208059110.

Jónsson, H. *et al.* (2013) 'mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters', *Bioinformatics (Oxford, England)*, 29(13), pp. 1682–1684. Available at: https://doi.org/10.1093/bioinformatics/btt193.

Keesing, F. *et al.* (2010) 'Impacts of biodiversity on the emergence and transmission of infectious diseases', *Nature*, 468(7324), pp. 647–652. Available at: https://doi.org/10.1038/nature09575.

Key, F.M. *et al.* (2020) 'Emergence of human-adapted Salmonella enterica is linked to the Neolithization process', *Nature Ecology & Evolution*, 4(3), pp. 324–333. Available at: https://doi.org/10.1038/s41559-020-1106-9.

Kim, S., Gupta, N. and Pevzner, P.A. (2008) 'Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases', *Journal of Proteome Research*, 7(8), pp. 3354–3363. Available at: https://doi.org/10.1021/pr8001244.

Kim, S. and Pevzner, P.A. (2014) 'MS-GF+ makes progress towards a universal database search tool for proteomics', *Nature Communications*, 5(1), p. 5277. Available at: https://doi.org/10.1038/ncomms6277.

Kjær, K.H. *et al.* (2022) 'A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA', *Nature*, 612(7939), pp. 283–291. Available at: https://doi.org/10.1038/s41586-022-05453-y.

Klunk, J. *et al.* (2022) 'Evolution of immune genes is associated with the Black Death', *Nature*, 611(7935), pp. 312–319. Available at: https://doi.org/10.1038/s41586-022-05349-x.

Knights, D. *et al.* (2011) 'Bayesian community-wide culture-independent microbial source tracking', *Nature Methods*, 8(9), pp. 761–763. Available at: https://doi.org/10.1038/nmeth.1650.

Kolenbrander, P.E. *et al.* (2002) 'Communication among oral bacteria', *Microbiology and molecular biology reviews: MMBR*, 66(3), pp. 486–505, table of contents. Available at: https://doi.org/10.1128/MMBR.66.3.486-505.2002.

Laland, K., Matthews, B. and Feldman, M.W. (2016) 'An introduction to niche construction theory', *Evolutionary Ecology*, 30(2), pp. 191–202. Available at: https://doi.org/10.1007/s10682-016-9821-z.

Laland, K.N. and O'Brien, M.J. (2010) 'Niche Construction Theory and Archaeology', *Journal of Archaeological Method and Theory*, 17(4), pp. 303–322. Available at: https://doi.org/10.1007/s10816-010-9096-6.

Laland, K.N., Odling-Smee, J. and Feldman, M.W. (2001) 'Cultural niche construction and human evolution: Niche construction and human evolution', *Journal of Evolutionary Biology*, 14(1), pp. 22–33. Available at: https://doi.org/10.1046/j.1420-9101.2001.00262.x.

Larsen, C.S. (1994) 'In the wake of Columbus: Native population biology in the postcontact Americas', *American Journal of Physical Anthropology*, 37(S19), pp. 109–154. Available at: https://doi.org/10.1002/ajpa.1330370606.

Larson, G. *et al.* (2014) 'Current perspectives and the future of domestication studies', *Proceedings of the National Academy of Sciences*, 111(17), pp. 6139–6146. Available at: https://doi.org/10.1073/pnas.1323964111.

Larson, G. and Fuller, D.Q. (2014) 'The Evolution of Animal Domestication', *Annual Review of Ecology, Evolution, and Systematics*, 45(1), pp. 115–136. Available at: https://doi.org/10.1146/annurev-ecolsys-110512-135813.

Lau, E.C. *et al.* (1989) 'Human and mouse amelogenin gene loci are on the sex chromosomes', *Genomics*, 4(2), pp. 162–168. Available at: https://doi.org/10.1016/0888-7543(89)90295-4.

Levin, M. (1989) *Old Nuclear Dump Off Coast Is a Relic of Era of Naivete in Toxic Disposal - Los Angeles Times*. Available at: https://www.latimes.com/archives/la-xpm-1989-07-31-me-449-story.html (Accessed: 31 March 2023).

Ley, R.E. *et al.* (2008) 'Evolution of mammals and their gut microbes', *Science (New York, N.Y.)*, 320(5883), pp. 1647–1651. Available at: https://doi.org/10.1126/science.1155725.

Loy, D.E. *et al.* (2017) 'Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax', *International Journal for Parasitology*, 47(2–3), pp. 87–97. Available at: https://doi.org/10.1016/j.ijpara.2016.05.008.

Lu, W. *et al.* (2007) 'Amplification and application of the HMG box of bovine SRY gene for sex determination', *Animal Reproduction Science*, 100(1), pp. 186–191. Available at: https://doi.org/10.1016/j.anireprosci.2006.08.023.

Lugli, F. *et al.* (2019) 'Enamel peptides reveal the sex of the Late Antique "Lovers of Modena"', *Scientific Reports*, 9(1), p. 13130. Available at: https://doi.org/10.1038/s41598-019-49562-7.

Lugli, F. *et al.* (2020) 'Sex-related morbidity and mortality in non-adult individuals from the Early Medieval site of Valdaro (Italy): the contribution of dental enamel peptide analysis', *Journal of Archaeological Science: Reports*, 34, p. 102625. Available at: https://doi.org/10.1016/j.jasrep.2020.102625.

MacCormick, H.A. (2023) *Comparison of specimens and field observations reveals biases in biodiversity data*. Available at: https://phys.org/news/2023-05-comparison-specimens-field-reveals-biases.html (Accessed: 3 May 2023).

Mangola, S.M. *et al.* (2022) 'Ethical microbiome research with Indigenous communities', *Nature Microbiology*, 7(6), pp. 749–756. Available at: https://doi.org/10.1038/s41564-022-01116-w.

Mann, A.E. *et al.* (2018) 'Differential preservation of endogenous human and microbial DNA in dental calculus and dentin', *Scientific Reports*, 8(1), p. 9822. Available at: https://doi.org/10.1038/s41598-018-28091-9.

Mann, A.E. *et al.* (2020) 'Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus', *Quaternary International*, 653–654, pp. 33–46. Available at: https://doi.org/10.1016/j.quaint.2020.11.019.

Marsh, P.D. (2000) 'Role of the Oral Microflora in Health', *Microbial Ecology in Health and Disease*, 12(3), pp. 130–137. Available at: https://doi.org/10.1080/089106000750051800.

Marshall, W.F. *et al.* (1994) 'Detection of Borrelia burgdorferi DNA in museum specimens of Peromyscus leucopus', *The Journal of Infectious Diseases*, 170(4), pp. 1027–1032. Available at: https://doi.org/10.1093/infdis/170.4.1027.

Martin, D.L. and Goodman, A.H. (2002) 'Health conditions before Columbus: paleopathology of native North Americans', *Western Journal of Medicine*, 176(1), pp. 65–68.

Matisoo-Smith, E. (2009) 'The Commensal Model for Human Settlement of the Pacific 10 Years on—What Can We Say and Where to Now?', *The Journal of Island and Coastal Archaeology*, 4(2), pp. 151–163. Available at: https://doi.org/10.1080/15564890903155273.

McGrath, K. *et al.* (2019) 'Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points', *Scientific Reports*, 9(1), p. 11027. Available at: https://doi.org/10.1038/s41598-019-47299-x.

McKenzie, V.J. *et al.* (2017) 'The Effects of Captivity on the Mammalian Gut Microbiome', *Integrative and Comparative Biology*, 57(4), pp. 690–704. Available at: https://doi.org/10.1093/icb/icx090.

McMahon, B.J., Morand, S. and Gray, J.S. (2018) 'Ecosystem change and zoonoses in the Anthropocene', *Zoonoses and Public Health*, 65(7), pp. 755–765. Available at: https://doi.org/10.1111/zph.12489.

McMurdie, P.J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLOS ONE*, 8(4), p. e61217. Available at: https://doi.org/10.1371/journal.pone.0061217.

McNally, L. and Brown, S.P. (2015) 'Building the microbiome in health and disease: niche construction and social conflict in bacteria', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1675), p. 20140298. Available at: https://doi.org/10.1098/rstb.2014.0298.

Messenger, A.M., Barnes, A.N. and Gray, G.C. (2014) 'Reverse Zoonotic Disease Transmission (Zooanthroponosis): A Systematic Review of Seldom-Documented Human Biological Threats to Animals', *PLoS ONE*. Edited by B.S. Schneider, 9(2), p. e89055. Available at: https://doi.org/10.1371/journal.pone.0089055.

Miarinjara, A. *et al.* (2021) 'Poor vector competence of the human flea, Pulex irritans, to transmit Yersinia pestis', *Parasites & Vectors*, 14(1), p. 317. Available at: https://doi.org/10.1186/s13071-021-04805-3.

Miller, M.A. *et al.* (2023) 'Newly detected, virulent Toxoplasma gondii COUG strain causing fatal steatitis and toxoplasmosis in southern sea otters (Enhydra lutris nereis)', *Frontiers in Marine Science*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1116899 (Accessed: 17 April 2023).

Moeller, A.H. *et al.* (2014) 'Rapid changes in the gut microbiome during human evolution', *Proceedings of the National Academy of Sciences*, 111(46), pp. 16431–16435. Available at: https://doi.org/10.1073/pnas.1419136111.

Mosites, E. *et al.* (2017) 'Microbiome sharing between children, livestock and household surfaces in western Kenya', *PLOS ONE*. Edited by P.E. Larsen, 12(2), p. e0171017. Available at: https://doi.org/10.1371/journal.pone.0171017.

Navon-Venezia, S., Kondratyeva, K. and Carattoli, A. (2017) 'Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance', *FEMS Microbiology Reviews*, 41(3), pp. 252–275. Available at: https://doi.org/10.1093/femsre/fux013.

Nelson, E.A., Halling, C.L. and Buikstra, J.E. (2019) 'Evidence of Skeletal Fluorosis at the Ray Site, Illinois, USA: a pathological assessment and discussion of environmental factors', *International Journal of Paleopathology*, 26, pp. 48–60. Available at: https://doi.org/10.1016/j.ijpp.2019.05.003.

*NeotomaDB - Home Page* (no date). Available at: https://www.neotomadb.org/ (Accessed: 1 February 2023).

Nistelberger, H.M. *et al.* (2019) 'Sexing Viking Age horses from burial and non-burial sites in Iceland using ancient DNA', *Journal of Archaeological Science*, 101, pp. 115–122. Available at: https://doi.org/10.1016/j.jas.2018.11.007.

Nogueira, F.C.S. *et al.* (2021) 'Ancient enamel peptides recovered from the South American Pleistocene species Notiomastodon platensis and Myocastor cf. coypus', *Journal of Proteomics*, 240, p. 104187. Available at: https://doi.org/10.1016/j.jprot.2021.104187.

Norcy, E.L. *et al.* (2011) 'Potential Role of the Amelogenin N-Terminus in the Regulation of Calcium Phosphate Formation in vitro', *Cells Tissues Organs*, 194(2–4), pp. 188–193. Available at: https://doi.org/10.1159/000324827.

Obregon-Tito, A.J. *et al.* (2015) 'Subsistence strategies in traditional societies distinguish gut microbiomes', *Nature Communications*, 6(1), p. 6505. Available at: https://doi.org/10.1038/ncomms7505.

O'Brien, M.J. and Laland, K.N. (2012) 'Genes, Culture, and Agriculture: An Example of Human Niche Construction', *Current Anthropology*, 53(4), pp. 434–470. Available at: https://doi.org/10.1086/666585.

Odling-Smee, F.J. (1988) 'Niche constructing phenotypes', in H.C. Plotkin (ed.) *The role of behavior in evolution*. The MIT Press, pp. 73–132.

Odling-Smee, F.J., Laland, K.N. and Feldman, M.W. (1996) 'Niche Construction', *The American Naturalist*, 147(4), pp. 641–648. Available at: https://doi.org/10.1086/285870.

Oksanen J, Blanchet FG, Friendly M et al. vegan: community ecology packag e. 2019. https://CRAN.R-project.org/package=vegan

Omran, A.R. (1971) 'The Epidemiologic Transition: A Theory of the Epidemiology of Population Change', *The Milbank Memorial Fund Quarterly*, 49(4), pp. 509–538. Available at: https://doi.org/10.2307/3349375.

Ostfeld, R.S. and Keesing, F. (2000) 'Biodiversity and Disease Risk: the Case of Lyme Disease', *Conservation Biology*, 14(3), p. 7.

Ostfeld, R.S. and Keesing, F. (2012) 'Effects of Host Diversity on Infectious Disease', p. 28.

Ottoni, C. *et al.* (2019) 'Metagenomic analysis of dental calculus in ancient Egyptian baboons', *Scientific Reports*, 9(1), p. 19637. Available at: https://doi.org/10.1038/s41598-019-56074-x.

Ozga, A.T. *et al.* (2016) 'Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus', *American Journal of Physical Anthropology*, 160(2), pp. 220–228. Available at: https://doi.org/10.1002/ajpa.22960.

Paladugu, R. *et al.* (2023) 'Your horse is a donkey! Identifying domesticated equids from Western Iberia using collagen fingerprinting', *Journal of Archaeological Science*, 149, p. 105696. Available at: https://doi.org/10.1016/j.jas.2022.105696.

Parker, G.J. *et al.* (2019) 'Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel', *Journal of Archaeological Science*, 101, pp. 169–180. Available at: https://doi.org/10.1016/j.jas.2018.08.011.

Parker, G.J. *et al.* (2021) 'AMELY deletion is not detected in systematically sampled reference populations: A Reply to Štamfelj', *Journal of Archaeological Science*, 130, p. 105354. Available at: https://doi.org/10.1016/j.jas.2021.105354.

Paz, S. (2015) 'Climate change impacts on West Nile virus transmission in a global context', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1665), p. 20130561. Available at: https://doi.org/10.1098/rstb.2013.0561.

Pearce-Duvet, J.M.C. (2006) 'The origin of human pathogens: evaluating the role of agriculture and domestic animals in the evolution of human disease', *Biological Reviews*, 81(3), pp. 369–382. Available at: https://doi.org/10.1017/S1464793106007020.

Peterson, J. *et al.* (2009) 'The NIH Human Microbiome Project', *Genome Research*, 19(12), pp. 2317–2323. Available at: https://doi.org/10.1101/gr.096651.109.

Plowright, R.K. *et al.* (2017) 'Pathways to zoonotic spillover', *Nature Reviews Microbiology*, 15(8), pp. 502–510. Available at: https://doi.org/10.1038/nrmicro.2017.45.

Plowright, R.K. *et al.* (2021) 'Land use-induced spillover: a call to action to safeguard environmental, animal, and human health', *The Lancet Planetary Health*, 5(4), pp. e237–e245. Available at: https://doi.org/10.1016/S2542-5196(21)00031-0.

Poo, S. *et al.* (2022) 'Bridging the Research Gap between Live Collections in Zoos and Preserved Collections in Natural History Museums', *BioScience*, 72(5), pp. 449–460. Available at: https://doi.org/10.1093/biosci/biac022.

Popkin, P.R.W. *et al.* (2012) 'The Sheep Project (1): determining skeletal growth, timing of epiphyseal fusion and morphometric variation in unimproved Shetland sheep of known age, sex, castration status and nutrition', *Journal of Archaeological Science*, 39(6), pp. 1775–1792. Available at: https://doi.org/10.1016/j.jas.2012.01.018.

Porru, E. *et al.* (2021) 'Bile acids and oxo-metabolites as markers of human faecal input in the ancient Pompeii ruins', *Scientific Reports*, 11(1), p. 3650. Available at: https://doi.org/10.1038/s41598-021-82831-y.

Puckett, E.E., Orton, D. and Munshi-South, J. (2020) 'Commensal Rats and Humans: Integrating Rodent Phylogeography and Zooarchaeology to Highlight Connections between Human Societies', *BioEssays*, 42(5), p. 1900160. Available at: https://doi.org/10.1002/bies.201900160.

Radde, H. (2021) 'Sea lion hunting strategies among late Holocene hunter-gatherer-fishers on Santa Catalina Island, California, USA', *Quaternary International*, 597, pp. 24–36. Available at: https://doi.org/10.1016/j.quaint.2020.12.011.

*Radiocarbon Palaeolithic Europe Database v29, Feb 2022* (no date). Available at: https://ees.kuleuven.be/en/geography/projects/14c-palaeolithic/index.html (Accessed: 1 February 2023).

Reitz, E.J. and Shackley, M. (2012) *Environmental Archaeology*. Boston, MA: Springer US (Manuals in Archaeological Method, Theory and Technique). Available at: https://doi.org/10.1007/978-1-4614-3339-2.

Rice, W.C. *et al.* (2012) 'Influence of wet distillers grains diets on beef cattle fecal bacterial community structure', *BMC Microbiology*, 12(1), p. 25. Available at: https://doi.org/10.1186/1471-2180-12-25.

Richter, K.K. *et al.* (2022) 'A primer for ZooMS applications in archaeology', *Proceedings of the National Academy of Sciences*, 119(20), p. e2109323119. Available at: https://doi.org/10.1073/pnas.2109323119.

Rick, T.C. and Erlandson, J.M. (2008) *Human Impacts on Ancient Marine Ecosystems: A Global Perspective*. 1st edn. University of California Press. Available at: https://www.jstor.org/stable/10.1525/j.ctt1pphh3 (Accessed: 11 March 2023).

Rick, T.C. and Lockwood, R. (2013) 'Integrating Paleobiology, Archeology, and History to Inform Biological Conservation: Paleobiology, Archeology, and History', *Conservation Biology*, 27(1), pp. 45–54. Available at: https://doi.org/10.1111/j.1523-1739.2012.01920.x.

Roberts, A.P. and Kreth, J. (2014) 'The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome', *Frontiers in Cellular and Infection Microbiology*, 4. Available at: https://www.frontiersin.org/articles/10.3389/fcimb.2014.00124 (Accessed: 31 March 2023).

Roberts, C.A. (2009) *Human Remains in Archaeology: a Handbook*. 1st edition. York: Council for British Archaeology.

Roberts, C.A. and Buikstra, J.E. (2019) 'Chapter 11 - Bacterial Infections', in J.E. Buikstra (ed.) *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)*. San Diego: Academic Press, pp. 321–439. Available at: https://doi.org/10.1016/B978-0-12-809738-0.00011-9.

Roberts, L.W. and Consultants, C. (2016) 'with portions adapted and updated from Historic Resource Study Channel Islands National Monument and San Miguel Island, California'.

Roberts, P. *et al.* (2017) 'The deep human prehistory of global tropical forests and its relevance for modern conservation', *Nature Plants*, 3(8), p. 17093. Available at: https://doi.org/10.1038/nplants.2017.93.

Robinson, E. *et al.* (2019) 'Spatiotemporal dynamics of prehistoric human population growth_ Radiocarbon "dates as data" and population ecology models', *Journal of Archaeological Science*, 101, pp. 63–71.

Rohland, N. *et al.* (2015) 'Partial uracil–DNA–glycosylase treatment for screening of ancient DNA', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660), p. 20130624. Available at: https://doi.org/10.1098/rstb.2013.0624.

Rüther, P.L. *et al.* (2022) 'SPIN enables high throughput species identification of archaeological bone by proteomics', *Nature Communications*, 13(1), p. 2458. Available at: https://doi.org/10.1038/s41467-022-30097-x.

Salvador, R.B. and Cunha, C.M. (2020) 'Natural history collections and the future legacy of ecological research', *Oecologia*, 192(3), pp. 641–646. Available at: https://doi.org/10.1007/s00442-020-04620-0.

Salyer, S.J. *et al.* (2017) 'Prioritizing Zoonoses for Global Health Capacity Building—Themes from One Health Zoonotic Disease Workshops in 7 Countries, 2014–2016', *Emerging Infectious Diseases*, 23(13). Available at: https://doi.org/10.3201/eid2313.170418.

Scarborough, C. *et al.* (2022) 'The historical ecology of coastal California', *Ocean & Coastal Management*, 230, p. 106352. Available at: https://doi.org/10.1016/j.ocecoaman.2022.106352.

Schmitt, C.J. *et al.* (2019) 'Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1763), p. 20170387. Available at: https://doi.org/10.1098/rstb.2017.0387.

Schnorr, S.L. *et al.* (2016) 'Insights into human evolution from ancient and contemporary microbiome studies', *Current Opinion in Genetics & Development*, 41, pp. 14–26. Available at: https://doi.org/10.1016/j.gde.2016.07.003.

Schubert, M. *et al.* (2012) 'Improving ancient DNA read mapping against modern reference genomes', *BMC genomics*, 13, p. 178. Available at: https://doi.org/10.1186/1471-2164-13-178.

Schubert, M., Lindgreen, S. and Orlando, L. (2016) 'AdapterRemoval v2: rapid adapter trimming, identification, and read merging', *BMC Research Notes*, 9(1), p. 88. Available at: https://doi.org/10.1186/s13104-016-1900-2.

Segata, N. (2015) 'Gut Microbiome: Westernization and the Disappearance of Intestinal Diversity', *Current Biology*, 25(14), pp. R611–R613. Available at: https://doi.org/10.1016/j.cub.2015.05.040.

Shaffer, J.R. *et al.* (2015) 'Genetic susceptibility to dental caries differs between the sexes: a family-based study', *Caries Research*, 49(2), pp. 133–140. Available at: https://doi.org/10.1159/000369103.

Shin, N.-Y. *et al.* (2020) 'Amelogenin phosphorylation regulates tooth enamel formation by stabilizing a transient amorphous mineral precursor', *Journal of Biological Chemistry*, 295(7), pp. 1943–1959. Available at: https://doi.org/10.1074/jbc.RA119.010506.

Shivaprakash, K.N. *et al.* (2021) 'Mammals, wildlife trade, and the next global pandemic', *Current Biology*, 31(16), pp. 3671-3677.e3. Available at: https://doi.org/10.1016/j.cub.2021.06.006.

Sinding, M.-H.S. *et al.* (2016) 'Sex determination of baleen whale artefacts: Implications for ancient DNA use in zooarchaeology', *Journal of Archaeological Science: Reports*, 10, pp. 345–349. Available at: https://doi.org/10.1016/j.jasrep.2016.11.001.

Singer, M. *et al.* (2017) 'Syndemics and the biosocial conception of health', *The Lancet*, 389(10072), pp. 941–950. Available at: https://doi.org/10.1016/S0140-6736(17)30003-X.

Smith, B.D. (2015) 'A Comparison of Niche Construction Theory and Diet Breadth Models as Explanatory Frameworks for the Initial Domestication of Plants and Animals', *Journal of Archaeological Research*, 23(3), pp. 215–262. Available at: https://doi.org/10.1007/s10814-015-9081-4.

Smith, B.D. and Zeder, M.A. (2013) 'The onset of the Anthropocene', *Anthropocene*, 4, pp. 8–13. Available at: https://doi.org/10.1016/j.ancene.2013.05.001.

Solazzo, C. *et al.* (2008) 'Identification of protein remains in archaeological potsherds by proteomics', *Analytical Chemistry*, 80(12), pp. 4590–4597. Available at: https://doi.org/10.1021/ac800515v.

Spyrou, M.A. *et al.* (2019) 'Ancient pathogen genomics as an emerging tool for infectious disease research', *Nature Reviews Genetics*, 20(6), pp. 323–340. Available at: https://doi.org/10.1038/s41576-019-0119-1.

Štamfelj, I. (2021) 'Sex estimation based on the analysis of enamel peptides: False assignments due to AMELY deletion', *Journal of Archaeological Science*, 130, p. 105345. Available at: https://doi.org/10.1016/j.jas.2021.105345.

Steffen, W. *et al.* (2015) 'The trajectory of the Anthropocene: The Great Acceleration', *The Anthropocene Review*, 2(1), pp. 81–98. Available at: https://doi.org/10.1177/2053019614564785.

Stewart, N.A. *et al.* (2016) 'The identification of peptides by nanoLC-MS/MS from human surface tooth enamel following a simple acid etch extraction', *RSC Advances*, 6(66), pp. 61673–61679. Available at: https://doi.org/10.1039/C6RA05120K.

Stewart, N.A. *et al.* (2017) 'Sex determination of human remains from peptides in tooth enamel', *Proceedings of the National Academy of Sciences*, 114(52), pp. 13649–13654. Available at: https://doi.org/10.1073/pnas.1714926115.

Stock, J.T. and Pinhasi, R. (2011) 'Introduction: Changing Paradigms in Our Understanding of the Transition to Agriculture: Human Bioarchaeology, Behaviour and Adaptaion', in R. Pinhasi and J.T. Stock (eds) *Human Bioarchaeology of the Transition to Agriculture*. 1st edn. Wiley, pp. 1–13. Available at: https://doi.org/10.1002/9780470670170.ch1.

Sugiyama, N. *et al.* (2022) 'Earliest evidence of primate captivity and translocation supports gift diplomacy between Teotihuacan and the Maya', *Proceedings of the National Academy of Sciences*, 119(47), p. e2212431119. Available at: https://doi.org/10.1073/pnas.2212431119.

Svensson, E.M., Götherström, A. and Vretemark, M. (2008) 'A DNA test for sex identification in cattle confirms osteometric results', *Journal of Archaeological Science*, 35(4), pp. 942–946. Available at: https://doi.org/10.1016/j.jas.2007.06.021.

Szabó, P. (2015) 'Historical ecology: past, present and future', *Biological Reviews*, 90(4), pp. 997–1014. Available at: https://doi.org/10.1111/brv.12141.

Thomas, R. (2019) 'Chapter 23 - Nonhuman Animal Paleopathology—Are We so Different?', in J.E. Buikstra (ed.) *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)*. San Diego: Academic Press, pp. 809–822. Available at: https://doi.org/10.1016/B978-0-12-809738-0.00023-5.

Thompson, C.W. *et al.* (2021) 'Preserve a Voucher Specimen! The Critical Need for Integrating Natural History Collections in Infectious Disease Studies', *mBio*, 12(1), pp. e02698-20. Available at: https://doi.org/10.1128/mBio.02698-20.

Thurstan, R.H. *et al.* (2015) 'Filling historical data gaps to foster solutions in marine conservation', *Ocean & Coastal Management*, 115, pp. 31–40. Available at: https://doi.org/10.1016/j.ocecoaman.2015.04.019.

Tokarski, C. *et al.* (2006) 'Identification of Proteins in Renaissance Paintings by Proteomics', *Analytical Chemistry*, 78(5), pp. 1494–1502. Available at: https://doi.org/10.1021/ac051181w.

Tounta, V. *et al.* (2021) 'Metabolomics in infectious diseases and drug discovery', *Molecular Omics*, 17(3), pp. 376–393. Available at: https://doi.org/10.1039/d1mo00017a.

Trinh, P. *et al.* (2018) 'One Health Relationships Between Human, Animal, and Environmental Microbiomes: A Mini-Review', *Frontiers in Public Health*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fpubh.2018.00235 (Accessed: 11 March 2023).

Trolle Jensen, T.Z. *et al.* (2019) 'The biomolecular characterization of a finger ring contextually dated to the emergence of the Early Neolithic from Syltholm, Denmark', *Royal Society Open Science*, 7(1), p. 191172. Available at: https://doi.org/10.1098/rsos.191172.

Tsosie, K.S. *et al.* (2021) 'We Have "Gifted" Enough: Indigenous Genomic Data Sovereignty in Precision Medicine', *The American Journal of Bioethics*, 21(4), pp. 72–75. Available at: https://doi.org/10.1080/15265161.2021.1891347.

Tumpey, T.M. *et al.* (2005) 'Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus', *Science*, 310(5745), pp. 77–80. Available at: https://doi.org/10.1126/science.1119392.

Turner, N.J., Geralda Armstrong, C. and Lepofsky, D. (2021) 'Adopting a Root: Documenting Ecological and Cultural Signatures of Plant Translocations in Northwestern North America',

*American Anthropologist*, 123(4), pp. 879–897. Available at: https://doi.org/10.1111/aman.13658.

Uhl, E.W. *et al.* (2019) 'New world origin of canine distemper: Interdisciplinary insights', *International Journal of Paleopathology*, 24, pp. 266–278. Available at: https://doi.org/10.1016/j.ijpp.2018.12.007.

Vigne, J.-D. (2011) 'The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere', *Comptes Rendus Biologies*, 334(3), pp. 171–181. Available at: https://doi.org/10.1016/j.crvi.2010.12.009.

Volokhov, D.V. *et al.* (2018) 'Neisseria zalophi sp. nov., isolated from oral cavity of California sea lions (Zalophus californianus)', *Archives of Microbiology*, 200(5), pp. 819–828. Available at: https://doi.org/10.1007/s00203-018-1499-x.

Wade, L. (2020) 'An unequal blow', *Science*, 368(6492), pp. 700–703. Available at: https://doi.org/10.1126/science.368.6492.700.

Wagner, J.K. *et al.* (2020) 'Fostering Responsible Research on Ancient DNA', *American Journal of Human Genetics*, 107(2), pp. 183–195. Available at: https://doi.org/10.1016/j.ajhg.2020.06.017.

Warinner, C. *et al.* (2014) 'Pathogens and host immunity in the ancient human oral cavity', *Nature Genetics*, 46(4), pp. 336–344. Available at: https://doi.org/10.1038/ng.2906.

Warinner, C. *et al.* (2015) 'Ancient human microbiomes', *Journal of Human Evolution*, 79, pp. 125–136. Available at: https://doi.org/10.1016/j.jhevol.2014.10.016.

Warinner, C. (2016) 'Dental Calculus and the Evolution of the Human Oral Microbiome', *Journal of the California Dental Association*, 44(7), pp. 411–420.

Warinner, C., Korzow Richter, K. and Collins, M.J. (2022) 'Paleoproteomics', *Chemical Reviews*, 122(16), pp. 13401–13446. Available at: https://doi.org/10.1021/acs.chemrev.1c00703.

Warinner, C., Speller, C. and Collins, M.J. (2015) 'A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660), p. 20130376. Available at: https://doi.org/10.1098/rstb.2013.0376.

Wasinger, V.C. *et al.* (2019) 'Analysis of the Preserved Amino Acid Bias in Peptide Profiles of Iron Age Teeth from a Tropical Environment Enable Sexing of Individuals Using Amelogenin MRM', *Proteomics*, 19(5), p. e1800341. Available at: https://doi.org/10.1002/pmic.201800341.

Waters, C.N. *et al.* (2016) 'The Anthropocene is functionally and stratigraphically distinct from the Holocene', *Science*, 351(6269), p. aad2622. Available at: https://doi.org/10.1126/science.aad2622.

Waters, C.N. and Turner, S.D. (2022) 'Defining the onset of the Anthropocene', *Science*, 378(6621), pp. 706–708. Available at: https://doi.org/10.1126/science.ade2310.

Welker, F. *et al.* (2015) 'Ancient proteins resolve the evolutionary history of Darwin's South American ungulates', *Nature*, 522(7554), pp. 81–84. Available at: https://doi.org/10.1038/nature14249.

Welker, F. *et al.* (2019) 'Enamel proteome shows that Gigantopithecus was an early diverging pongine', *Nature*, 576(7786), pp. 262–265. Available at: https://doi.org/10.1038/s41586-019-1728-8.

Welker, F. *et al.* (2020) 'The dental proteome of Homo antecessor', *Nature*, 580(7802), pp. 235–238. Available at: https://doi.org/10.1038/s41586-020-2153-8.

Wellman, H.P. *et al.* (2020) 'Archaeological mitogenomes illuminate the historical ecology of sea otters (Enhydra lutris) and the viability of reintroduction', *Proceedings of the Royal Society B: Biological Sciences*, 287(1940), p. 20202343. Available at: https://doi.org/10.1098/rspb.2020.2343.

Wells, J.C.K. and Stock, J.T. (2020) 'Life History Transitions at the Origins of Agriculture: A Model for Understanding How Niche Construction Impacts Human Growth, Demography and Health', *Frontiers in Endocrinology*, 11, p. 325. Available at: https://doi.org/10.3389/fendo.2020.00325.

Weyrich, L.S., Dobney, K. and Cooper, A. (2015) 'Ancient DNA analysis of dental calculus', *Journal of Human Evolution*, 79, pp. 119–124. Available at: https://doi.org/10.1016/j.jhevol.2014.06.018.

White, A.J. *et al.* (2019) 'Fecal stanols show simultaneous flooding and seasonal precipitation change correlate with Cahokia's population decline', *Proceedings of the National Academy of Sciences*, 116(12), pp. 5461–5466. Available at: https://doi.org/10.1073/pnas.1809400116.

White, R.J. and Razgour, O. (2020) 'Emerging zoonotic diseases originating in mammals: a systematic review of effects of anthropogenic land-use change', *Mammal Review*, 50(4), pp. 336–352. Available at: https://doi.org/10.1111/mam.12201.

Whiteman, J.P. *et al.* (2019) 'A Guide to Using Compound-Specific Stable Isotope Analysis to Study the Fates of Molecules in Organisms and Ecosystems', *Diversity*, 11(1), p. 8. Available at: https://doi.org/10.3390/d11010008.

Willcox, G. (2012) 'Pre-Domestic Cultivation during the Late Pleistocene and Early Holocene in the Northern Levant', in A.B. Damania et al. (eds) *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*. Cambridge: Cambridge University Press, pp. 92–109. Available at: https://doi.org/10.1017/CBO9781139019514.007.

Williams, A. and Ulm, S. (2014) *AustArch: A Database of 14C and Luminescence Ages from Archaeological Sites in Australia*, *Archaeology Data Service*. Available at:

https://archaeologydataservice.ac.uk/archives/view/austarch_na_2014/ (Accessed: 6 December 2022).

Wilson, A.J. and Mellor, P.S. (2009) 'Bluetongue in Europe: past, present and future', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530), pp. 2669–2681. Available at: https://doi.org/10.1098/rstb.2009.0091.

Wolf, M. (2015) 'Is there really such a thing as "one health"? Thinking about a more than human world from the perspective of cultural anthropology', *Social Science & Medicine*, 129, pp. 5–11. Available at: https://doi.org/10.1016/j.socscimed.2014.06.018.

Wolfhagen, J. (2020) 'Re-examining the use of the LSI technique in zooarchaeology', *Journal of Archaeological Science*, 123, p. 105254. Available at: https://doi.org/10.1016/j.jas.2020.105254.

Wood, D.E. and Salzberg, S.L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biology*, 15(3), p. R46. Available at: https://doi.org/10.1186/gb-2014-15-3-r46.

Wood, J.W. *et al.* (1992) 'The Osteological Paradox: Problems of Inferring Prehistoric Health from Skeletal Samples [and Comments and Reply]', *Current Anthropology*, 33(4), pp. 343–370.

Woolhouse, M.E.J. and Gowtage-Sequeria, S. (2005) 'Host Range and Emerging and Reemerging Pathogens', *Emerging Infectious Diseases*, 11(12), pp. 1842–1847. Available at: https://doi.org/10.3201/eid1112.050997.

Wyatt, K.B. *et al.* (2008) 'Historical Mammal Extinction on Christmas Island (Indian Ocean) Correlates with Introduced Infectious Disease', *PLOS ONE*, 3(11), p. e3602. Available at: https://doi.org/10.1371/journal.pone.0003602.

Xia, R. (2020) *How the waters off Catalina became a DDT dumping ground*, *Los Angeles Times*. Available at: https://www.latimes.com/projects/la-coast-ddt-dumping-ground/ (Accessed: 31 March 2023).

Young, H.S. *et al.* (2017) 'Interacting effects of land use and climate on rodent-borne pathogens in central Kenya', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1722), p. 20160116. Available at: https://doi.org/10.1098/rstb.2016.0116.

Yu, H. *et al.* (2022) 'Palaeogenomic analysis of black rat (Rattus rattus) reveals multiple European introductions associated with human economic history', *Nature Communications*, 13(1), p. 2399. Available at: https://doi.org/10.1038/s41467-022-30009-z.

Zeder, M.A. (2015) 'Core questions in domestication research', *Proceedings of the National Academy of Sciences*, 112(11), pp. 3191–3198. Available at: https://doi.org/10.1073/pnas.1501711112.

Zeder, M.A. (2016) 'Domestication as a model system for niche construction theory', *Evolutionary Ecology*, 30(2), pp. 325–348. Available at: https://doi.org/10.1007/s10682-015-9801-8.

Zeder, M.A. and Hesse, B. (2000) 'The Initial Domestication of Goats ( *Capra hircus* ) in the Zagros Mountains 10,000 Years Ago', *Science*, 287(5461), pp. 2254–2257. Available at: https://doi.org/10.1126/science.287.5461.2254.

Ziesemer, K.A. *et al.* (2015) 'Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification', *Scientific Reports*, 5, p. 16498. Available at: https://doi.org/10.1038/srep16498.

Zipkin, E.F. *et al.* (2021) 'Addressing data integration challenges to link ecological processes across scales', *Frontiers in Ecology and the Environment*, 19(1), pp. 30–38. Available at: https://doi.org/10.1002/fee.2290.

Zuckerman, M. *et al.* (2014) 'The evolution of disease: anthropological perspectives on epidemiologic transitions', *Global Health Action*, 7(1), p. 23303. Available at: https://doi.org/10.3402/gha.v7.23303.

Zuckerman, M.K. and Armelagos, G.J. (2014) 'The Hygiene Hypothesis and the Second Epidemiologic Transition', in *Modern Environments and Human Health*. John Wiley & Sons, Ltd, pp. 301–320. Available at: https://doi.org/10.1002/9781118504338.ch16.

Zuckerman, M.K. and Dafoe, A.C. (2020) 'Disease in the context of environmental change', in G.R. Schug (ed.) *The Routledge Handbook of the Bioarchaeology of Climate and Environmental Change*. 1st edn. Routledge, pp. 43–59. Available at: https://doi.org/10.4324/9781351030465-4.

**Image References**

California sea lion (*Zalophus californianus*) and a Western gull (*Larus occidentalis*) in Scripps Park in La Jolla (San Diego, California). Rhododendrites. Available at: https://creativecommons.org/licenses/by-sa/4.0/deed.en

*Homo sapiens.* T. Michael Keesey, 2010 Jan 10. Available at: http://creativecommons.org/publicdomain/mark/1.0.

*Rousettus aegyptiacus.* 2020 Mar 26. Melissa Ingala. Available at: https://creativecommons.org/licenses/by/3.0/

*Potamops mascahehenes.* 2015 Dec 20. Zimices (Julián Bayona). Available at: https://creativecommons.org/licenses/by-nc/3.0/

*Eptesicus fuscus.* 2021Apr 3 by Margot Michaud. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Ailuropoda melanoleuca* 2020 Jan 26. Margot Michaud. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Ursus americanus*. 2013 Jun 29 by Tracy Heath. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Canis familiaris dingo.* 2013 Jul 15 by T. Michael Keesey. Available at: https://creativecommons.org/licenses/by/3.0/

*Vulpes vulpes*. 2018 Oct 10 by Anthony Caravaggi. Available at: https://creativecommons.org/licenses/by-nc-sa/3.0/

*Meles meles*. 2018 Oct 10. Anthony Caravaggi. Available at: https://creativecommons.org/licenses/by-nc-sa/3.0/

*Canis familiaris familiaris* 2023 Feb 5. Margot Michaud. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Panthera uncia*. 2021 Mar 16. Gabriela Palomo-Munoz. Available at: https://creativecommons.org/licenses/by-nc/3.0/

*Mustela erminea*. 2022 Sep 8. Margot Michaud. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Enhydra lutris*. 2020 Jun 1. Margot Michaud. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Camelus ferus*. 2019 Mar 15. Jody Taylor. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Sus scrofa domestica*. 2012 Feb 10. Steven Traver. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Cervus canadensis* 2013 Jul 30. Tracy Heath. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Bos primigenius taurus* 2023 Feb 15. Katy Lawler. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Capra aegagrus hircus*. 2012 Feb 12. Steven Traver. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Odocoileus hemionus* 2021 Nov 8. Gabriela Palomo-Munoz. Available at: https://creativecommons.org/licenses/by-nc/3.0/

*Ovis aries dolichura*. 2021 Nov 8. Gabriela Palomo-Munoz. Available at: https://creativecommons.org/licenses/by-nc/3.0/

*Macaca mulatta*. 2013 Sep 21. T. Michael Keesey. Available at: https://creativecommons.org/publicdomain/mark/1.0/

*Gorilla gorilla*. 2013 May 6. T. Michael Keesey. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Pan troglodytes*. 2020 Mar 30. Jonathan Lawley. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Castor canadensis* 2019 Sep 25. Margot Michaud. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Mus musculus domesticus*. 2013 Jul 20. T. Michael Keesey. Available at: https://creativecommons.org/licenses/by-sa/3.0/

*Mus musculus*. 2020 Jan 21. Soledad Miranda-Rottmann. Available at: https://creativecommons.org/licenses/by/3.0/

*Georychus capensis*. 2020 Oct 21. Kai Caspar. Available at: https://creativecommons.org/licenses/by/3.0/

*Rattus rattus*. 2021 Sep 10. Ferran Sayol. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Geomys bursarius*. 2020 Jan 21. Xavier Jenkins. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Equus ferus caballus*. 2011 Feb 25. David Orr. Available at: https://creativecommons.org/publicdomain/zero/1.0/

*Elephas maximus*. 2019 Mar 15. Jody Taylor. Available at: https://creativecommons.org/publicdomain/zero/1.0/

# Supplemental 1

The archaeological human remains used in Chapter 3 for peptide sexing were collected under the supervision of Dr. Adam Rabinowitz at the University of Texas at Austin and with permission of Andrei Soficaru at the "Francisc J. Rainer" Institute of Anthropology in Romania. Mr. Sterling Wright collected dental remains for genomic, proteomic, and microbiome analysis in 2018. All laboratory analyses were conducted at the University of Oklahoma Laboratories of Molecular Anthropology and Microbiome Research ancient DNA laboratory.

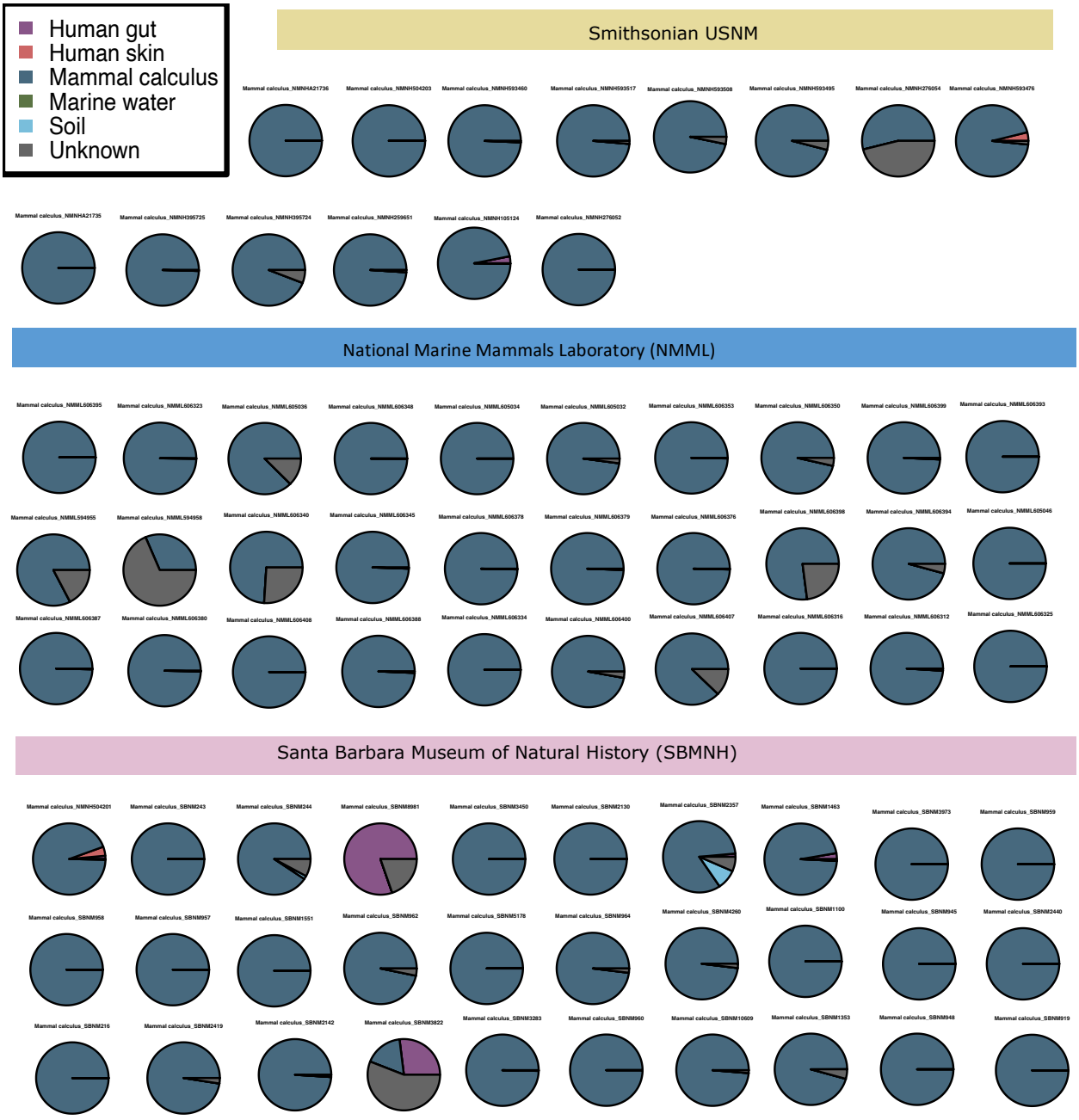
# Supplemental 2

| #SampleID | Study | Description | Notes |
|---|---|---|---|
| ERR6093813 | Brealy etal 2021 | Bear calculus | PRJEB42014 |
| ERR6093818 | Brealy etal 2022 | Bear calculus | PRJEB42014 |
| ERR6093852 | Brealy etal 2023 | Bear calculus | PRJEB42014 |
| ERR6093886 | Brealy etal 2024 | Bear calculus | PRJEB42014 |
| ERR6096923 | Brealy etal 2025 | Bear calculus | PRJEB42014 |
| ERR6108966 | Brealy etal 2026 | Bear calculus | PRJEB42014 |
| ERR6108968 | Brealy etal 2027 | Bear calculus | PRJEB42014 |
| ERR6108972 | Brealy etal 2028 | Bear calculus | PRJEB42014 |
| ERR6108976 | Brealy etal 2029 | Bear calculus | PRJEB42014 |
| ERR6109023 | Brealy etal 2030 | Bear calculus | PRJEB42014 |
| SRR9163889 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163890 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163895 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163898 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163900 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163906 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163907 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163908 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163911 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR9163912 | Suttner2020 | California_sediment | PRJNA545542 |
| SRR1631060 | Oh et al., 2014 | HMP | PRJNA4633 |
| SRR1631061 | Oh et al., 2014 | HMP | PRJNA4634 |
| SRR1631063 | Oh et al., 2014 | HMP | PRJNA4635 |
| SRR1631064 | Oh et al., 2014 | HMP | PRJNA4636 |
| SRR1633008 | Oh et al., 2014 | HMP | PRJNA4637 |
| SRR3184100 | Oh et al., 2014 | HMP | PRJNA4638 |
| SRR3184876 | Oh et al., 2014 | HMP | PRJNA4639 |
| SRR3189411 | Oh et al., 2014 | HMP | PRJNA4640 |
| SRR3189416 | Oh et al., 2014 | HMP | PRJNA4641 |
| SRR3189418 | Oh et al., 2014 | HMP | PRJNA4642 |
| SRR1779378 | Shi et al, 2015 | human Subgingival plaque microbiome | PRJNA255922 |
| SRR1779506 | Shi et al, 2015 | human Subgingival plaque microbiome | PRJNA255922 |
| SRR1779502 | Shi et al, 2015 | human Subgingival plaque microbiome | PRJNA255922 |
| SRR1779403 | Shi et al, 2015 | human Subgingival plaque microbiome | PRJNA255922 |

| | | | |
|---|---|---|---|
| SRR177949 | Shi et al, 2015 | human Subgingival plaque microbiome | PRJNA255922 |
| SRR16074162 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074167 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074175 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074179 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074185 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074187 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074189 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074192 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074195 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR16074196 | Pang et al., 2021 | human adolescent Dental Plaque on Pit and Fissure Site | PRJNA766357 |
| SRR5788152 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788154 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788156 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788158 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788162 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788324 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788393 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788415 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| SRR5788418 | Biller et al., 2018 | Pacific ocean | PRJNA385854 |
| 00o6o2 | Hofman | Orca calculus | NA |
| CAS20749 | Hofman | Orca calculus | NA |
| CAS29219 | Hofman | Orca calculus | NA |
| CAS30729 | Hofman | Orca calculus | NA |
| Eyak | Hofman | Orca calculus | NA |
| NMML78 | Hofman | Orca calculus | NA |
| NMML79 | Hofman | Orca calculus | NA |
| NMML81 | Hofman | Orca calculus | NA |
| NMML82 | Hofman | Orca calculus | NA |
| NMML84 | Hofman | Orca calculus | NA |
| NMML85 | Hofman | Orca calculus | NA |
| UWBM82332 | Hofman | Orca calculus | NA |
| UWBM82333 | Hofman | Orca calculus | NA |
| UWBM82688 | Hofman | Orca calculus | NA |

| | | | |
|---|---|---|---|
| Yakat | Hofman | Orca calculus | NA |
| SRR8899213 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899216 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899217 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899219 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899221 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899223 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899225 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899227 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899228 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SRR8899230 | Ozga et al 2019 | Chimp calculus | PRJNA531027 |
| SM01 | ObregonTito2015 | Matses rural gut | PRJNA268964 |
| SM05 | ObregonTito2015 | Matses rural gut | PRJNA268964 |
| SM18 | ObregonTito2015 | Matses rural gut | PRJNA268964 |
| SM31 | ObregonTito2015 | Matses rural gut | PRJNA268964 |
| SM43 | ObregonTito2015 | Matses rural gut | PRJNA268964 |
| SRR5763450 | Schnorr2017 | Hadza rural gut | PRJNA392180 |
| SRR5763460 | Schnorr2017 | Hadza rural gut | PRJNA392180 |
| SRR5763463 | Schnorr2017 | Hadza rural gut | PRJNA392180 |
| SRR5763476 | Schnorr2017 | Hadza rural gut | PRJNA392180 |
| SRR5763484 | Schnorr2017 | Hadza rural gut | PRJNA392180 |
| HMP_700016610 | HMP2013 | HMP urban gut | NA |
| HMP_700021824 | HMP2013 | HMP urban gut | NA |
| HMP_700023337 | HMP2013 | HMP urban gut | NA |
| HMP_700024318 | HMP2013 | HMP urban gut | NA |
| HMP_700033153 | HMP2013 | HMP urban gut | NA |
| HMP_700037501 | HMP2013 | HMP urban gut | NA |
| HMP_700114653 | HMP2013 | HMP urban gut | NA |
| HMP_700116201 | HMP2013 | HMP urban gut | NA |
| HMP_700161856 | HMP2013 | HMP urban gut | NA |
| HMP_700164870 | HMP2013 | HMP urban gut | NA |

# Supplemental 3




Samples NMNH593476, NMNH105124, NMNH504201, SBMNH8981, SBMNH2357, SBMNH1463, SBMNH244, SBMNH3822 had presence of skin, gut, and soil contamination.

# Supplemental 4

*Zalophus californianus* host genome mapping summary

| Sample | Total Reads | Analysis-ready reads (trimmed and merged) | Proportion of reads kept (post-Adapter Removal) | Q37 Mapped reads | Proportion of reads mapped | Unique Q37 mapped reads | Avg. length of mapped reads | Res-caled or Not | Mean cover-age | Std dev of mean coverage | % ref. covered >= 1X | % ref. covered >= 5X | Cluster factor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | 2 | 0.0003 | 1 | 31 | No | 0X | 0.0001X | 0% | 0% | 2 |
| dsENB | 1,982,046 | 12,067 | 0.006 | 34 | 0.0028 | 2 | 76 | No | 0X | 0.0003X | 0% | 0% | 17 |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 161 | 0.0026 | 1 | 44 | No | 0X | 0.0001X | 0% | 0% | 161 |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| dsLNB | 3,362,917 | 434 | 0.0001 | 3 | 0.0069 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 5,840 | 0.002 | 46 | 68.913 | Yes | 0X | 0.0012X | 0% | 0% | 126.9565 |
| dsExtNegE | 2,068,072 | 194,559 | 0.094 | 350 | 0.0017 | 50 | 50.7 | Yes | 0X | 0.0011X | 0% | 0% | 7 |
| dsExtNegF | 5,075,412 | 243,767 | 0.048 | 286 | 0.0011 | 18 | 52.7778 | No | 0X | 0.0007X | 0% | 0% | 15.8888 |
| dsLibNegE | 2,352,887 | 1,510 | 0.0006 | 1 | 0.0006 | 1 | 31 | No | 0X | 0.0001X | 0% | 0% | 1 |
| dsLibNegF | 1,673,188 | 698 | 0.0004 | 5 | 0.0071 | 1 | 140 | No | 0X | 0.0002X | 0% | 0% | 5 |
| dsLND | 4,612,074 | 17,550 | 0.0038 | 116 | 0.0066 | 3 | 50.6667 | No | 0X | 0.0003X | 0% | 0% | 38.6666 |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 2,617,816 | 0.0439 | 1,992,940 | 77.5172 | Yes | 0.0639X | 0.2703X | 6.08% | 0% | 1.3135 |
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 44,951 | 0.0014 | 32,599 | 58.5075 | Yes | 0.0008X | 0.0282X | 0.08% | 0% | 1.3789 |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 178,216 | 0.0065 | 137,680 | 68.7931 | Yes | 0.0039X | 0.0635X | 0.39% | 0% | 1.2944 |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 31,793 | 0.0009 | 24,516 | 73.5288 | Yes | 0.0007X | 0.0275X | 0.07% | 0% | 1.2968 |
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 267,667 | 0.0079 | 205,422 | 82.0077 | Yes | 0.0069X | 0.0856X | 0.67% | 0% | 1.303 |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 367,138 | 0.0087 | 289,908 | 73.4534 | Yes | 0.0088X | 0.0953X | 0.87% | 0% | 1.2663 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 137,531 | 0.0025 | 71,425 | 70.8216 | Yes | 0.0021X | 0.0469X | 0.20% | 0% | 1.9255 |
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 61,019 | 0.0011 | 27,716 | 59.6919 | Yes | 0.0007X | 0.027X | 0.07% | 0% | 2.2015 |
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 39,440 | 0.0007 | 20,568 | 62.5291 | Yes | 0.0005X | 0.0236X | 0.05% | 0% | 1.9175 |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 133,125 | 0.0028 | 82,510 | 60.4387 | Yes | 0.0021X | 0.0459X | 0.21% | 0% | 1.6134 |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 147,845 | 0.003 | 110,280 | 58.0594 | Yes | 0.0027X | 0.0519X | 0.26% | 0% | 1.3406 |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 696,601 | 0.0117 | 328,249 | 66.502 | Yes | 0.009X | 0.0975X | 0.89% | 0% | 2.1221 |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 95,625 | 0.0022 | 58,280 | 67.0066 | Yes | 0.0016X | 0.0407X | 0.16% | 0% | 1.6407 |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 552,463 | 0.0088 | 318,949 | 75.5405 | Yes | 0.01X | 0.1049X | 0.97% | 0% | 1.7321 |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 562,904 | 0.0078 | 361,796 | 81.241 | Yes | 0.0121X | 0.1157X | 1.16% | 0% | 1.5558 |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 30,416 | 0.0007 | 22,402 | 60.405 | Yes | 0.0006X | 0.0239X | 0.06% | 0% | 1.3577 |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 63,323 | 0.0016 | 47,774 | 57.5032 | Yes | 0.0011X | 0.0343X | 0.11% | 0% | 1.3254 |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 212,580 | 0.0059 | 155,915 | 67.3253 | Yes | 0.0043X | 0.0671X | 0.43% | 0% | 1.3634 |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 25,586 | 0.0005 | 18,652 | 70.263 | Yes | 0.0005X | 0.0235X | 0.05% | 0% | 1.3717 |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 52,440 | 0.0013 | 38,466 | 68.9811 | Yes | 0.0011X | 0.0333X | 0.11% | 0% | 1.3632 |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 1,022,188 | 0.0141 | 742,910 | 84.7773 | Yes | 0.026X | 0.1735X | 2.51% | 0% | 1.3759 |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 1,421,337 | 0.0181 | 1,037,173 | 69.6023 | Yes | 0.0298X | 0.1898X | 2.88% | 0% | 1.3703 |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 199,421 | 0.0024 | 141,927 | 65.9568 | Yes | 0.0039X | 0.0628X | 0.38% | 0% | 1.405 |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 159,732 | 0.0027 | 119,287 | 64.7707 | Yes | 0.0032X | 0.057X | 0.32% | 0% | 1.339 |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 85,651 | 0.0026 | 64,557 | 68.2509 | Yes | 0.0018X | 0.0432X | 0.18% | 0% | 1.3267 |
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 42,343 | 0.0007 | 30,497 | 74.1822 | Yes | 0.0009X | 0.0308X | 0.09% | 0% | 1.3884 |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 106,775 | 0.002 | 71,654 | 66.6672 | Yes | 0.002X | 0.0451X | 0.20% | 0% | 1.4901 |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 144,589 | 0.0036 | 98,137 | 67.9814 | Yes | 0.0028X | 0.0534X | 0.27% | 0% | 1.4733 |
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 316,955 | 0.007 | 244,315 | 79.0987 | Yes | 0.008X | 0.0907X | 0.79% | 0% | 1.2973 |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 67,845 | 0.0016 | 50,355 | 71.657 | Yes | 0.0015X | 0.0391X | 0.15% | 0% | 1.3473 |
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 277,955 | 0.0157 | 198,039 | 81.3716 | Yes | 0.0067X | 0.0842X | 0.66% | 0% | 1.4035 |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 18,814 | 0.0012 | 14,456 | 62.0917 | Yes | 0.0004X | 0.0193X | 0.04% | 0% | 1.3014 |
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 9,419 | 0.0006 | 6,310 | 58.9919 | Yes | 0.0002X | 0.0125X | 0.02% | 0% | 1.4927 |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 189 | 0.0002 | 139 | 50.2734 | Yes | 0X | 0.0017X | 0% | 0% | 1.3597 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 2,995 | 0.0001 | 1,154 | 68.3951 | Yes | 0X | 0.0058X | 0% | 0% | 2.5953 |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 9,416 | 0.0005 | 6,298 | 56.8712 | Yes | 0.0001X | 0.0123X | 0.01% | 0% | 1.495 |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 16,556 | 0.0002 | 9,625 | 61.2825 | Yes | 0.0002X | 0.0158X | 0.02% | 0% | 1.7201 |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 60,117 | 0.0031 | 39,072 | 66.2952 | Yes | 0.0011X | 0.0331X | 0.11% | 0% | 1.5386 |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 8,767 | 0.0009 | 6,028 | 55.5174 | Yes | 0.0001X | 0.0118X | 0.01% | 0% | 1.4543 |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 988,346 | 0.0647 | 691,936 | 65.8346 | Yes | 0.0189X | 0.1438X | 1.85% | 0% | 1.4283 |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 5,580 | 0.0004 | 3,886 | 58.9743 | Yes | 0.0001X | 0.0098X | 0.01% | 0% | 1.4359 |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 6,107 | 0.0006 | 4,640 | 62.244 | Yes | 0.0001X | 0.011X | 0.01% | 0% | 1.3161 |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 2,086 | 0.0001 | 1,372 | 53.6071 | Yes | 0X | 0.0056X | 0% | 0% | 1.5204 |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 25,300 | 0.0013 | 18,366 | 65.665 | Yes | 0.0005X | 0.0224X | 0.05% | 0% | 1.3775 |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 787,870 | 0.0016 | 346,602 | 70.0891 | Yes | 0.01X | 0.1066X | 0.95% | 0% | 2.2731 |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 175,851 | 0.0125 | 120,699 | 66.9847 | Yes | 0.0033X | 0.0584X | 0.33% | 0% | 1.4569 |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 116,849 | 0.0093 | 59,512 | 64.6737 | Yes | 0.0016X | 0.0517X | 0.15% | 0% | 1.9634 |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 35,657 | 0.0021 | 26,660 | 62.1374 | Yes | 0.0007X | 0.0263X | 0.07% | 0% | 1.3374 |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 57,912 | 0.0037 | 42,673 | 64.8722 | Yes | 0.0011X | 0.0341X | 0.11% | 0% | 1.3571 |
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 46,030 | 0.001 | 21,503 | 65.987 | Yes | 0.0006X | 0.0247X | 0.06% | 0% | 2.1406 |
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 8,024,991 | 0.5664 | 3,895,371 | 73.3216 | Yes | 0.1165X | 19.1584X | 6.26% | 0.01% | 2.0601 |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 41,756 | 0.0028 | 31,395 | 62.2829 | Yes | 0.0008X | 0.0286X | 0.08% | 0% | 1.33 |
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 10,971 | 0.0018 | 8,201 | 65.4862 | Yes | 0.0002X | 0.0149X | 0.02% | 0% | 1.3377 |
| SBNM948 | 14,347,850 | 10,598,99 | 0.7387 | 17,268 | 0.0016 | 13,489 | 70.5263 | Yes | 0.0004X | 0.0199X | 0.04% | 0% | 1.2801 |
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 5,377 | 0.0006 | 4,009 | 60.6054 | Yes | 0.0001X | 0.0101X | 0.01% | 0% | 1.3412 |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 7,135 | 0.001 | 5,272 | 65.5971 | Yes | 0.0001X | 0.012X | 0.01% | 0% | 1.3533 |
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 163,631 | 0.0112 | 115,732 | 70.263 | Yes | 0.0034X | 0.0587X | 0.33% | 0% | 1.4138 |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 9,917 | 0.0005 | 6,600 | 53.7961 | Yes | 0.0001X | 0.0123X | 0.01% | 0% | 1.5025 |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 523,190 | 0.0444 | 405,182 | 70.1045 | Yes | 0.0117X | 0.1127X | 1.15% | 0% | 1.2912 |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 13,067 | 0.0007 | 9,233 | 57.7596 | Yes | 0.0002X | 0.015X | 0.02% | 0% | 1.4152 |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 121,549 | 0.006 | 72,329 | 54.1279 | Yes | 0.0016X | 0.041X | 0.16% | 0% | 1.6805 |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 61,753 | 0.0014 | 42,931 | 62.7034 | Yes | 0.0011X | 0.034X | 0.11% | 0% | 1.4384 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 29,462 | 0.0012 | 23,564 | 57.0496 | Yes | 0.0006X | 0.0243X | 0.06% | 0% | 1.2502 |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 174,200 | 0.0103 | 140,296 | 56.9024 | Yes | 0.0033X | 0.0579X | 0.33% | 0% | 1.2416 |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 2,219,414 | 0.0946 | 1,793,838 | 60.9329 | Yes | 0.0453X | 0.2474X | 4.36% | 0% | 1.2372 |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 58,737 | 0.002 | 47,133 | 59.3494 | Yes | 0.0012X | 0.0341X | 0.12% | 0% | 1.2461 |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 12,517 | 0.0005 | 7,006 | 60.3341 | Yes | 0.0002X | 0.0143X | 0.02% | 0% | 1.7866 |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 160,636 | 0.008 | 126,706 | 58.904 | Yes | 0.0031X | 0.056X | 0.31% | 0% | 1.2677 |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 793,641 | 0.04 | 626,683 | 56.1602 | Yes | 0.0146X | 0.1256X | 1.44% | 0% | 1.2664 |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 530,107 | 0.0214 | 436,826 | 69.2846 | Yes | 0.0125X | 0.1143X | 1.24% | 0% | 1.2135 |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 102,807 | 0.006 | 84,862 | 66.2331 | Yes | 0.0023X | 0.0484X | 0.23% | 0% | 1.2114 |
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 104,398 | 0.0048 | 84,433 | 66.4636 | Yes | 0.0023X | 0.0484X | 0.23% | 0% | 1.2364 |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 270,013 | 0.0122 | 215,605 | 72.1975 | Yes | 0.0064X | 0.0813X | 0.64% | 0% | 1.2523 |
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 12,311 | 0.0008 | 9,914 | 56.0207 | Yes | 0.0002X | 0.0153X | 0.02% | 0% | 1.2417 |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 6,592 | 0.0003 | 5,220 | 52.0843 | Yes | 0.0001X | 0.0107X | 0.01% | 0% | 1.2628 |
| NMNHextNegA | 4,113,971 | 580,441 | 0.141 | 61 | 0.0001 | 20 | 46.35 | Yes | 0X | 0.0007X | 0% | 0% | 3.05 |
| NMNHextNegB | 3,868,171 | 1,273,359 | 0.3291 | 62 | 0 | 24 | 53.8333 | Yes | 0X | 0.0007X | 0% | 0% | 2.5833 |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | 6 | 0.0001 | 1 | 94 | No | 0X | 0.0002X | 0% | 0% | 6 |

# Supplemental 5

*Neisseria zalophi* mapping summary

| Sample | Total Reads | Analysis-ready reads | Proportion of reads kept (post-Adapter-Removal) | Mapped reads | Proportion of reads mapped | Unique mapped reads | Avg. length of mapped reads | Res-caled or Not | Mean coverage | Std dev of mean coverage | % ref. covered >= 1X | % ref. covered >= 5X | Cluster factor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | - | 0 | - | no unique mapped reads | Yes | N/A | N/A | N/A | N/A | |
| dsENB | 1,982,046 | 12,067 | 0.006 | 4 | 0.0003 | - | no unique mapped reads | Yes | N/A | N/A | N/A | N/A | |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 40 | 0.0006 | 4 | 67.5 | Yes | 0.0001X | 0.0106X | 0.01% | 0% | 10 |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 545 | 0.0001 | 2 | 57.5 | Yes | 0X | 0.0069X | 0% | 0% | 272.5 |
| dsExt NegE | 2,068,072 | 194,559 | 0.094 | 83 | 0.0004 | 24 | 66.7917 | Yes | 0.0007X | 0.0257X | 0.07% | 0% | 3.4583 |
| dsExt NegF | 5,075,412 | 243,767 | 0.048 | 97 | 0.0003 | 4 | 71.25 | Yes | 0.0001X | 0.0108X | 0.01% | 0% | 24.25 |
| dsLib NegE | 2,352,887 | 1,510 | 0.0006 | 2 | 0.0013 | 1 | 35 | Yes | 0X | 0.0038X | 0% | 0% | 2 |
| dsLib NegF | 1,673,188 | 698 | 0.0004 | 2 | 0.0028 | 1 | 124 | Yes | 0.0001X | 0.0072X | 0.01% | 0% | 2 |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | no unique mapped reads | Yes | N/A | N/A | N/A | N/A | |
| dsLNB | 3,362,917 | 434 | 0.0001 | 1 | 0.0023 | - | no unique mapped reads | Yes | N/A | N/A | N/A | N/A | |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | no unique mapped reads | Yes | N/A | N/A | N/A | N/A | |
| dsLND | 4,612,074 | 17,550 | 0.0038 | - | 0 | - | no unique mapped reads | Yes | N/A | N/A | N/A | N/A | |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 1,440,095 | 0.0242 | 1,132,909 | 82.3747 | Yes | 38.5051X | 28.9159X | 92.64% | 84.33% | 1.2711 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 150,642 | 0.0047 | 115,824 | 70.9185 | Yes | 3.3891X | 3.1756X | 78.10% | 32.33% | 1.3006 |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 269,199 | 0.0099 | 219,282 | 80.4582 | Yes | 7.2796X | 6.0086X | 85.28% | 62.16% | 1.2276 |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 304,312 | 0.0091 | 245,648 | 77.5079 | Yes | 7.856X | 12.0694X | 62.74% | 34.85% | 1.2388 |
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 188,943 | 0.0056 | 146,548 | 87.6043 | Yes | 5.2972X | 5.2044X | 80.45% | 47.48% | 1.2892 |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 203,625 | 0.0048 | 164,653 | 85.8984 | Yes | 5.8357X | 4.4946X | 88.04% | 58.47% | 1.2366 |
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 662,267 | 0.0122 | 341,198 | 82.6952 | Yes | 11.6419X | 10.9948X | 85.81% | 67.89% | 1.941 |
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 135,212 | 0.0025 | 58,922 | 70.18 | Yes | 1.7062X | 1.8417X | 69.66% | 6.78% | 2.2947 |
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 542,256 | 0.0098 | 299,875 | 69.7704 | Yes | 8.6327X | 6.9476X | 87.42% | 69.44% | 1.8082 |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 408,128 | 0.0088 | 257,301 | 69.4669 | Yes | 7.3749X | 6.1537X | 87.13% | 64.96% | 1.5861 |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 507,112 | 0.0105 | 389,913 | 68.1 | Yes | 10.9558X | 8.8274X | 88.39% | 73.22% | 1.3005 |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 1,515,750 | 0.0254 | 758,815 | 73.6358 | Yes | 23.0543X | 16.9219X | 86.28% | 78.20% | 1.9975 |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 1,482,244 | 0.0343 | 861,864 | 77.5252 | Yes | 27.5683X | 19.4803X | 91.42% | 86.53% | 1.7198 |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 956,634 | 0.0152 | 540,204 | 95.3558 | Yes | 21.2538X | 15.9468X | 90.93% | 77.66% | 1.7708 |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 2,259,982 | 0.0316 | 1,357,308 | 101.68 | Yes | 56.9452X | 41.6147X | 90.98% | 82.79% | 1.665 |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 271,427 | 0.0063 | 214,225 | 73.3915 | Yes | 6.4871X | 5.1564X | 90.23% | 62.07% | 1.267 |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 240,608 | 0.0063 | 191,357 | 63.1564 | Yes | 4.9866X | 4.3891X | 83.26% | 49.09% | 1.2573 |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 296,226 | 0.0082 | 246,459 | 68.5964 | Yes | 6.9757X | 6.2544X | 81.77% | 57.29% | 1.2019 |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 401,522 | 0.0089 | 317,297 | 68.686 | Yes | 8.9924X | 8.9269X | 86.56% | 61.03% | 1.2654 |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 588,834 | 0.0151 | 451,099 | 91.2971 | Yes | 16.9928X | 13.5434X | 92.60% | 78.06% | 1.3053 |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 2,242,156 | 0.031 | 1,660,094 | 97.3776 | Yes | 66.7004X | 48.964X | 94.49% | 87.09% | 1.3506 |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 1,336,926 | 0.017 | 1,057,830 | 76.3691 | Yes | 33.3328X | 27.998X | 90.15% | 78.77% | 1.2638 |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 2,027,009 | 0.0252 | 1,427,319 | 85.4703 | Yes | 50.3349X | 33.7896X | 91.15% | 86.05% | 1.4201 |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 1,049,827 | 0.0183 | 840,518 | 83.0246 | Yes | 28.7928X | 17.4414X | 92.31% | 87.77% | 1.249 |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 137,444 | 0.0042 | 107,267 | 84.5322 | Yes | 3.7414X | 3.5563X | 78.60% | 35.51% | 1.2813 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 247,291 | 0.0044 | 179,733 | 90.1903 | Yes | 6.6884X | 6.3879X | 89.32% | 66.75% | 1.3758 |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 265,707 | 0.005 | 202,912 | 88.9842 | Yes | 7.4499X | 5.784X | 86.68% | 65.15% | 1.3094 |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 140,148 | 0.0035 | 103,723 | 76.6779 | Yes | 3.2816X | 4.0661X | 78.31% | 28.65% | 1.3511 |
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 189,114 | 0.0041 | 146,808 | 87.5896 | Yes | 5.3057X | 4.9762X | 76.69% | 50.19% | 1.2881 |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 191,008 | 0.0046 | 149,347 | 83.4652 | Yes | 5.1432X | 4.3003X | 80.30% | 52.54% | 1.2789 |
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 908,288 | 0.0516 | 598,289 | 98.8317 | Yes | 24.3971X | 20.761X | 92.96% | 81.37% | 1.5181 |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 15,477 | 0.001 | 11,199 | 77.6061 | Yes | 0.3586X | 0.8587X | 25.97% | 0.23% | 1.3819 |
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 268,482 | 0.0191 | 199,421 | 66.779 | Yes | 5.4948X | 4.4042X | 83.83% | 55.21% | 1.3463 |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 10,025 | 0.012 | 8,281 | 62.2653 | Yes | 0.2127X | 0.4754X | 18.65% | 0% | 1.2106 |
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 78,402 | 0.0047 | 53,532 | 80.9547 | Yes | 1.7881X | 2.1246X | 67.09% | 8.25% | 1.4645 |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 147,085 | 0.0082 | 114,662 | 73.7718 | Yes | 3.4902X | 2.8767X | 82.34% | 33.32% | 1.2827 |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 78,421 | 0.0011 | 41,284 | 81.8051 | Yes | 1.3935X | 2.5675X | 65.03% | 3.03% | 1.8995 |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 60,816 | 0.0032 | 41,409 | 72.0042 | Yes | 1.2302X | 1.6862X | 58.56% | 2.85% | 1.4686 |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 25,769 | 0.0026 | 18,082 | 63.6013 | Yes | 0.4745X | 0.9147X | 30.14% | 0.50% | 1.4251 |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 852,221 | 0.0558 | 681,771 | 63.9026 | Yes | 17.9757X | 12.3532X | 90.58% | 82.73% | 1.25 |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 215,974 | 0.0173 | 156,229 | 79.686 | Yes | 5.1366X | 4.0492X | 85.89% | 52.31% | 1.3824 |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 81,587 | 0.0092 | 67,061 | 80.2249 | Yes | 2.2198X | 3.2411X | 57.93% | 17.95% | 1.2166 |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 105,356 | 0.0075 | 78,822 | 64.529 | Yes | 2.0986X | 2.3093X | 67.61% | 13.59% | 1.3366 |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 355,161 | 0.0189 | 273,771 | 85.6066 | Yes | 9.67X | 6.7027X | 91.84% | 72.89% | 1.2972 |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 4,121,908 | 0.0085 | 1,873,185 | 74.4031 | Yes | 57.5049X | 42.4845X | 95.43% | 90.32% | 2.2004 |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 111,584 | 0.0079 | 79,974 | 68.2709 | Yes | 2.2528X | 2.0681X | 73.78% | 14.64% | 1.3952 |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 260,418 | 0.0208 | 200,198 | 59.663 | Yes | 4.9283X | 3.8478X | 85.63% | 50.45% | 1.3008 |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 299,340 | 0.0176 | 241,456 | 70.4373 | Yes | 7.0174X | 4.6159X | 94.07% | 67.98% | 1.2397 |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 268,420 | 0.0171 | 209,463 | 76.3589 | Yes | 6.5993X | 5.4669X | 83.69% | 59.73% | 1.2814 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 161,761 | 0.0036 | 74,726 | 83.6359 | Yes | 2.5787X | 2.445X | 77.58% | 18.82% | 2.1647 |
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 1,343 | 0 | 141 | 66.9078 | Yes | 0.0039X | 0.0684X | 0.38% | 0% | 9.5248 |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 58,072 | 0.0039 | 44,312 | 82.5462 | Yes | 1.5092X | 1.7999X | 65.94% | 4.67% | 1.3105 |
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 28,439 | 0.0048 | 22,441 | 85.2752 | Yes | 0.7896X | 1.0761X | 47.46% | 0.73% | 1.2672 |
| SBNM948 | 14,347,850 | 10,598,993 | 0.7387 | 34,843 | 0.0032 | 27,480 | 69.6589 | Yes | 0.7898X | 1.1064X | 46.76% | 1.03% | 1.2679 |
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 28,591 | 0.0036 | 22,840 | 77.5275 | Yes | 0.7306X | 1.0037X | 45.70% | 0.52% | 1.2517 |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 13,930 | 0.0019 | 10,568 | 78.6556 | Yes | 0.343X | 0.7161X | 26.17% | 0.16% | 1.3181 |
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 197,229 | 0.0136 | 147,853 | 72.3926 | Yes | 4.4164X | 3.757X | 79.43% | 45.96% | 1.3339 |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 73,391 | 0.004 | 49,894 | 68.4695 | Yes | 1.4096X | 2.1733X | 59.92% | 5.16% | 1.4709 |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 88,561 | 0.0075 | 71,950 | 81.0307 | Yes | 2.4056X | 2.6978X | 70.98% | 18.18% | 1.2308 |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 96,042 | 0.0051 | 74,002 | 72.4357 | Yes | 2.2118X | 2.2624X | 70.77% | 14.87% | 1.2978 |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 645,581 | 0.0321 | 556,124 | 51.3063 | Yes | 11.7725X | 8.5215X | 87.97% | 73.73% | 1.1608 |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 321,499 | 0.0076 | 255,026 | 66.6688 | Yes | 7.0151X | 5.5564X | 83.76% | 63.17% | 1.2606 |
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 299,118 | 0.0131 | 251,844 | 59.6526 | Yes | 6.1985X | 5.1276X | 81.71% | 60.89% | 1.1877 |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 29,393 | 0.0017 | 21,240 | 60.7582 | Yes | 0.5325X | 1.05X | 36.19% | 0.39% | 1.3838 |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 650,302 | 0.0277 | 569,014 | 64.6354 | Yes | 15.1748X | 10.8304X | 86.62% | 76.25% | 1.1428 |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 178,088 | 0.0063 | 147,173 | 67.6394 | Yes | 4.1073X | 3.4853X | 85.48% | 41.27% | 1.21 |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 167,919 | 0.0073 | 137,275 | 57.3048 | Yes | 3.2457X | 3.1621X | 77.46% | 29.38% | 1.2232 |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 457,924 | 0.023 | 386,530 | 60.7498 | Yes | 9.6886X | 6.7695X | 87.60% | 73.84% | 1.1847 |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 209,963 | 0.0105 | 175,808 | 67.4629 | Yes | 4.8937X | 3.7839X | 86.98% | 49.94% | 1.1942 |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 243,579 | 0.0098 | 205,286 | 69.9797 | Yes | 5.9273X | 4.787X | 86.66% | 57.56% | 1.1865 |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 134,131 | 0.0078 | 113,877 | 78.6207 | Yes | 3.694X | 3.1846X | 82.41% | 35.57% | 1.1778 |
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 199,440 | 0.0092 | 171,301 | 71.0018 | Yes | 5.0183X | 3.6703X | 88.11% | 53.13% | 1.1642 |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 285,220 | 0.0129 | 239,400 | 81.052 | Yes | 8.0061X | 5.6524X | 88.15% | 70.72% | 1.1913 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 104,615 | 0.0072 | 84,933 | 59.9573 | Yes | 2.1011X | 2.188X | 72.59% | 12.06% | 1.2317 |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 165,490 | 0.0099 | 140,866 | 56.7402 | Yes | 3.2978X | 3.005X | 77.90% | 31.21% | 1.1748 |
| NMNHextNegA | 4,113,971 | 580,441 | 0.141 | 910 | 0.0015 | 24 | 47.6667 | Yes | 0.0005X | 0.0314X | 0.03% | 0% | 37.9166 |
| NMNHextNegB | 3,868,171 | 1,273,359 | 0.3291 | 1,894 | 0.0014 | 36 | 46.7778 | Yes | 0.0007X | 0.0536X | 0.03% | 0% | 52.6111 |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |

# Supplemental 6

*Streptococcus marimammalium* mapping summary

| Sample | Total Reads | Analysis-ready reads | Proportion of reads kept post-Adapter-Removal | Mapped reads | Proportion of reads mapped | Unique mapped reads | Avg. length of mapped reads | Rescaled or Not | Mean coverage | Std dev of mean coverage | % ref. covered >= 1X | % ref. covered >= 5X | Cluster factor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsENB | 1,982,046 | 12,067 | 0.006 | 2 | 0.0001 | 1 | 52 | No | N/A | N/A | N/A | N/A | N/A |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 2 | 0 | 1 | 39 | No | N/A | N/A | N/A | N/A | N/A |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 51 | 0 | 3 | 36 | Yes | 0.0001X | 0.0085X | 0.01% | 0% | 17 |
| dsExtNegE | 2,068,072 | 194,559 | 0.094 | 17 | 0 | 11 | 52.8182 | Yes | 0.0004X | 0.0199X | 0.04% | 0% | 1.5454 |
| dsExtNegF | 5,075,412 | 243,767 | 0.048 | 19 | 0 | 7 | 48.2857 | No | N/A | N/A | N/A | N/A | N/A |
| dsLibNegE | 2,352,887 | 1,510 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLibNegF | 1,673,188 | 698 | 0.0004 | 1 | 0.0014 | 1 | 40 | No | N/A | N/A | N/A | N/A | N/A |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLNB | 3,362,917 | 434 | 0.0001 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLND | 4,612,074 | 17,550 | 0.0038 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 134,288 | 0.0022 | 110,518 | 78.7985 | Yes | 5.7844X | 19.965X | 40.49% | 23.26% | 1.215 |
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 63,839 | 0.002 | 53,550 | 66.7255 | Yes | 2.3733X | 11.6527X | 34.07% | 13.44% | 1.1921 |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 81,095 | 0.0029 | 69,663 | 70.4345 | Yes | 3.2591X | 12.6147X | 38.05% | 17.26% | 1.1641 |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 8,621 | 0.0002 | 7,230 | 77.4357 | Yes | 0.3719X | 4.4913X | 10.12% | 1.16% | 1.1923 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 64,386 | 0.0019 | 54,034 | 81.8007 | Yes | 2.9359X | 13.0198X | 37.11% | 15.42% | 1.1915 |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 30,505 | 0.0007 | 26,142 | 74.2157 | Yes | 1.2887X | 8.8264X | 26.86% | 7.06% | 1.1668 |
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 199,651 | 0.0036 | 108,593 | 78.1734 | Yes | 5.6385X | 17.3707X | 42.62% | 23.64% | 1.8385 |
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 58,271 | 0.0011 | 29,336 | 63.9582 | Yes | 1.2462X | 7.7211X | 27.98% | 7.32% | 1.9863 |
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 88,610 | 0.0016 | 53,416 | 64.9581 | Yes | 2.3047X | 9.9865X | 36.46% | 14.11% | 1.6588 |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 100,557 | 0.0021 | 69,594 | 65.4407 | Yes | 3.025X | 12.2487X | 37% | 16.60% | 1.4449 |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 50,523 | 0.001 | 41,181 | 59.565 | Yes | 1.6293X | 10.1729X | 32.39% | 9.27% | 1.2268 |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 293,184 | 0.0049 | 160,335 | 69.6147 | Yes | 7.4136X | 22.1086X | 43.51% | 26.61% | 1.8285 |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 119,823 | 0.0027 | 79,964 | 69.7976 | Yes | 3.7071X | 12.1643X | 39.19% | 19.62% | 1.4984 |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 110,572 | 0.0017 | 66,830 | 84.7431 | Yes | 3.7617X | 14.7506X | 36.34% | 17.26% | 1.6545 |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 115,628 | 0.0016 | 77,875 | 86.2757 | Yes | 4.4627X | 16.6311X | 37.57% | 18.37% | 1.4847 |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 87,538 | 0.002 | 74,491 | 68.3369 | Yes | 3.3811X | 14.5069X | 38.33% | 17.80% | 1.1751 |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 119,396 | 0.0031 | 101,676 | 64.2489 | Yes | 4.339X | 16.6966X | 42.16% | 21.55% | 1.1742 |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 243,390 | 0.0067 | 206,343 | 73.912 | Yes | 10.1295X | 30.5737X | 47.49% | 30.09% | 1.1795 |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 68,957 | 0.0015 | 56,842 | 64.4468 | Yes | 2.4332X | 14.5462X | 34.75% | 13.22% | 1.2131 |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 76,862 | 0.0019 | 62,716 | 82.0458 | Yes | 3.4178X | 13.6488X | 36.66% | 16.74% | 1.2255 |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 234,977 | 0.0032 | 184,284 | 88.565 | Yes | 10.8405X | 37.9408X | 46.13% | 28.75% | 1.275 |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 359,266 | 0.0045 | 290,139 | 74.9538 | Yes | 14.4447X | 50.6675X | 48.43% | 32.26% | 1.2382 |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 96,962 | 0.0012 | 77,276 | 73.943 | Yes | 3.7953X | 17.4751X | 39.44% | 18.85% | 1.2547 |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 68,383 | 0.0011 | 57,513 | 71.5612 | Yes | 2.7337X | 13.3988X | 36.35% | 14.88% | 1.189 |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 38,926 | 0.0011 | 32,651 | 81.1983 | Yes | 1.7609X | 8.9171X | 30.07% | 10.32% | 1.1921 |
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 57,297 | 0.001 | 45,947 | 81.0904 | Yes | 2.4748X | 15.3112X | 31.19% | 11.60% | 1.247 |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 42,477 | 0.0008 | 35,106 | 75.3064 | Yes | 1.756X | 8.6549X | 30.48% | 10.39% | 1.2099 |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 129,117 | 0.0032 | 106,093 | 71.933 | Yes | 5.0688X | 19.0683X | 41.78% | 21.70% | 1.217 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 41,622 | 0.0009 | 34,630 | 80.2329 | Yes | 1.8455X | 11.6278X | 28.85% | 9.56% | 1.2019 |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 56,806 | 0.0013 | 46,540 | 77.2269 | Yes | 2.3873X | 10.9544X | 32.93% | 12.62% | 1.2205 |
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 69,233 | 0.0039 | 54,254 | 77.9336 | Yes | 2.8084X | 11.5669X | 35.70% | 15.26% | 1.276 |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 9,237 | 0.0006 | 7,999 | 71.4943 | Yes | 0.3799X | 4.0934X | 11.93% | 1.13% | 1.1547 |
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 32,417 | 0.0023 | 25,669 | 58.173 | Yes | 0.9919X | 4.1462X | 29.39% | 5.89% | 1.2628 |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 2,518 | 0.003 | 2,214 | 56.0208 | Yes | 0.0824X | 0.4069X | 6.44% | 0.06% | 1.1373 |
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 58,163 | 0.0034 | 43,861 | 70.5743 | Yes | 2.0561X | 7.7306X | 30.65% | 11.81% | 1.326 |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 26,171 | 0.0014 | 21,897 | 64.1711 | Yes | 0.9333X | 5.25X | 25.59% | 5.23% | 1.1951 |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 70,811 | 0.001 | 44,034 | 73.8776 | Yes | 2.1608X | 14.8193X | 30.38% | 11.22% | 1.608 |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 24,184 | 0.0012 | 18,499 | 60.9264 | Yes | 0.7487X | 4.4694X | 23.13% | 3.43% | 1.3073 |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 8,173 | 0.0008 | 6,452 | 62.2111 | Yes | 0.2666X | 3.2288X | 10.78% | 0.64% | 1.2667 |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 76,253 | 0.0049 | 63,949 | 62.3129 | Yes | 2.6467X | 8.9922X | 38.06% | 16.34% | 1.1924 |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 30,289 | 0.0024 | 24,482 | 73.7126 | Yes | 1.1987X | 5.6733X | 26.88% | 6.58% | 1.2371 |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 17,889 | 0.002 | 15,820 | 77.8025 | Yes | 0.8175X | 3.5415X | 21.19% | 5.39% | 1.1307 |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 25,719 | 0.0018 | 21,317 | 61.7928 | Yes | 0.875X | 5.1826X | 24.14% | 4.44% | 1.2065 |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 24,845 | 0.0013 | 20,270 | 74.0179 | Yes | 0.9966X | 5.4448X | 24.42% | 5.75% | 1.2257 |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 958,989 | 0.0019 | 369,926 | 74.2436 | Yes | 18.2419X | 68.2457X | 48.75% | 31.47% | 2.5923 |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 30,720 | 0.0021 | 23,593 | 65.9037 | Yes | 1.0328X | 3.9294X | 28.81% | 6.66% | 1.302 |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 52,228 | 0.0041 | 42,937 | 57.6599 | Yes | 1.6445X | 8.6864X | 33.13% | 10.65% | 1.2163 |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 13,375 | 0.0007 | 11,018 | 59.8433 | Yes | 0.438X | 4.3581X | 16.74% | 1.14% | 1.2139 |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 107,329 | 0.0068 | 88,877 | 69.0416 | Yes | 4.0757X | 12.3948X | 39.21% | 19.95% | 1.2076 |
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 34,726 | 0.0007 | 18,294 | 67.3149 | Yes | 0.818X | 5.9918X | 22.75% | 4.12% | 1.8982 |
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 3,743 | 0.0002 | 2,513 | 59.2654 | Yes | 0.0989X | 4.865X | 0.16% | 0.12% | 1.4894 |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 30,535 | 0.002 | 26,033 | 71.0743 | Yes | 1.229X | 6.1039X | 27.25% | 7.63% | 1.1729 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 8,948 | 0.0015 | 7,751 | 72.3258 | Yes | 0.3723X | 1.8306X | 16.60% | 1.21% | 1.1544 |
| SBNM948 | 14,347,850 | 10,598,993 | 0.7387 | 10,644 | 0.001 | 9,277 | 64.9496 | Yes | 0.4002X | 2.93X | 16.39% | 1.14% | 1.1473 |
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 8,629 | 0.001 | 7,545 | 68.1406 | Yes | 0.3415X | 2.3335X | 14.90% | 1.01% | 1.1436 |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 8,553 | 0.0012 | 7,499 | 71.0724 | Yes | 0.354X | 2.0288X | 15.02% | 1.21% | 1.1405 |
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 31,878 | 0.0021 | 25,495 | 67.0239 | Yes | 1.135X | 4.8664X | 27.29% | 7.81% | 1.2503 |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 29,199 | 0.0016 | 22,904 | 64.5003 | Yes | 0.9813X | 5.7815X | 25.96% | 5.16% | 1.2748 |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 46,702 | 0.0039 | 41,044 | 73.1017 | Yes | 1.9929X | 8.6593X | 33.69% | 11.62% | 1.1378 |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 11,262 | 0.0006 | 9,153 | 64.6461 | Yes | 0.393X | 4.6032X | 14.04% | 0.95% | 1.2304 |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 92,459 | 0.0045 | 81,666 | 53.3038 | Yes | 2.8914X | 10.4905X | 40.10% | 17.64% | 1.1321 |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 132,584 | 0.0031 | 110,463 | 64.5034 | Yes | 4.7326X | 17.4998X | 41.62% | 22.07% | 1.2002 |
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 85,171 | 0.0037 | 76,417 | 58.7762 | Yes | 2.9833X | 11.1545X | 40.71% | 17.10% | 1.1145 |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 26,874 | 0.0015 | 24,052 | 60.4946 | Yes | 0.9665X | 5.4904X | 26.94% | 5.15% | 1.1173 |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 229,597 | 0.0097 | 205,667 | 64.972 | Yes | 8.8756X | 25.7255X | 46.75% | 29.54% | 1.1163 |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 136,698 | 0.0048 | 121,574 | 64.7627 | Yes | 5.2297X | 16.6675X | 45.41% | 24.69% | 1.1244 |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 37,700 | 0.0016 | 33,717 | 60.7215 | Yes | 1.3598X | 5.8794X | 31.07% | 8.97% | 1.1181 |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 54,850 | 0.0027 | 48,026 | 59.6667 | Yes | 1.9034X | 8.6363X | 31.18% | 11.84% | 1.142 |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 33,017 | 0.0016 | 27,457 | 58.228 | Yes | 1.062X | 11.081X | 23.98% | 4.27% | 1.2024 |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 80,397 | 0.0032 | 72,820 | 68.9112 | Yes | 3.3331X | 13.0109X | 38.91% | 17.89% | 1.104 |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 15,066 | 0.0008 | 13,817 | 70.3251 | Yes | 0.6454X | 3.9607X | 21.32% | 2.91% | 1.0903 |
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 26,365 | 0.0012 | 23,966 | 66.7358 | Yes | 1.0623X | 6.4071X | 27.32% | 5.85% | 1.1001 |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 53,616 | 0.0024 | 47,217 | 72.2154 | Yes | 2.2648X | 9.1593X | 33.83% | 13.83% | 1.1355 |
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 24,572 | 0.0017 | 22,166 | 57.2851 | Yes | 0.8434X | 4.494X | 25.91% | 4.41% | 1.1085 |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 36,785 | 0.0022 | 33,683 | 57.4259 | Yes | 1.2848X | 5.8885X | 29.02% | 8.38% | 1.092 |
| NMNHextNegA | 4,113,971 | 580,441 | 0.141 | 99 | 0.0001 | 86 | 47.0465 | Yes | 0.0027X | 0.176X | 0.04% | 0.02% | 1.1511 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMNHextNegB | 3,868,171 | 1,273,359 | 0.3291 | 172 | 0.0001 | 121 | 48.0992 | Yes | 0.0039X | 0.246X | 0.04% | 0.02% | 1.4214 |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |

# Supplemental 7

*Desulfomicrobium orale* mapping summary

| Sample | Total Reads | Analysis-ready reads | Proportion of reads kept post-Adapter-Removal | Mapped reads | Proportion of reads mapped | Unique mapped reads | Avg. length of mapped reads | Rescaled or Not | Mean coverage | Std dev of mean coverage | % ref covered >= 1X | % ref covered >= 5X | Cluster factor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsENB | 1,982,046 | 12,067 | 0.006 | - | 0 | - | no unique mapped reads | BAM file was empty, | N/A | N/A | N/A | N/A | N/A |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 2 | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 17 | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsExtNegE | 2,068,072 | 194,559 | 0.094 | 38 | 0.0001 | 16 | 57.0625 | Yes | 0.0003X | 0.0181X | 0.03% | 0% | 2.375 |
| dsExtNegF | 5,075,412 | 243,767 | 0.048 | 19 | 0 | 3 | 52.3333 | Yes | 0.0001X | 0.0075X | 0.01% | 0% | 6.3333 |
| dsLibNegE | 2,352,887 | 1,510 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLibNegF | 1,673,188 | 698 | 0.0004 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLNB | 3,362,917 | 434 | 0.0001 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |
| dsLND | 4,612,074 | 17,550 | 0.0038 | 39 | 0.0022 | 1 | 39 | No | N/A | N/A | N/A | N/A | N/A |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 100,679 | 0.0016 | 70,909 | 76.2997 | Yes | 1.9438X | 2.7194X | 58.81% | 14.08% | 1.4198 |
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 23,823 | 0.0007 | 4,531 | 58.3454 | Yes | 0.095X | 1.8894X | 3.96% | 0.34% | 5.2577 |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 4,182 | 0.0001 | 228 | 46.8553 | Yes | 0.0038X | 0.1373X | 0.17% | 0.02% | 18.3421 |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 13,370 | 0.0003 | 3,612 | 72.3314 | Yes | 0.0939X | 0.5788X | 7.03% | 0.17% | 3.7015 |
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 142,347 | 0.0042 | 111,579 | 81.3223 | Yes | 3.2602X | 4.1284X | 65.01% | 32.06% | 1.2757 |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 210,407 | 0.005 | 162,348 | 78.7824 | Yes | 4.5953X | 5.2331X | 67.72% | 43.30% | 1.296 |
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 30,719 | 0.0005 | 12,149 | 75.0429 | Yes | 0.3276X | 0.7733X | 24.17% | 0.14% | 2.5285 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 21,527 | 0.0004 | 2,945 | 58.1579 | Yes | 0.0615X | 0.884X | 3.88% | 0.14% | 7.3096 |
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 21,119 | 0.0003 | 7,530 | 64.1918 | Yes | 0.1737X | 0.5266X | 14.31% | 0.08% | 2.8046 |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 13,707 | 0.0002 | 1,504 | 54.6197 | Yes | 0.0295X | 0.6783X | 0.71% | 0.18% | 9.1136 |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 172,277 | 0.0035 | 128,105 | 57.0271 | Yes | 2.6247X | 3.3006X | 64.84% | 22.54% | 1.3448 |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 30,574 | 0.0005 | 13,070 | 59.9272 | Yes | 0.2814X | 0.6268X | 21.97% | 0.05% | 2.3392 |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 34,918 | 0.0008 | 15,812 | 72.1256 | Yes | 0.4097X | 1.0371X | 27.67% | 0.30% | 2.2083 |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 164,301 | 0.0026 | 90,770 | 76.3395 | Yes | 2.4896X | 3.6844X | 62.50% | 21.18% | 1.81 |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 64,430 | 0.0009 | 36,141 | 78.4586 | Yes | 1.0188X | 1.5938X | 47.57% | 2.90% | 1.7827 |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 145,411 | 0.0034 | 100,818 | 66.8352 | Yes | 2.4209X | 3.885X | 63.19% | 20.12% | 1.4423 |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 64,677 | 0.0017 | 46,680 | 52.7533 | Yes | 0.8848X | 1.3317X | 46.65% | 1.58% | 1.3855 |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 194,228 | 0.0054 | 156,764 | 71.8163 | Yes | 4.0449X | 4.5369X | 68.05% | 39.60% | 1.2389 |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 33,568 | 0.0007 | 21,442 | 63.0785 | Yes | 0.486X | 0.9811X | 32.40% | 0.29% | 1.5655 |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 5,397 | 0.0001 | 332 | 54.0904 | Yes | 0.0065X | 0.1589X | 0.33% | 0.03% | 16.256 |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 491,536 | 0.0067 | 366,806 | 84.3409 | Yes | 11.1152 X | 11.8527 X | 72.32% | 59.25% | 1.34 |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 1,068,713 | 0.0136 | 804,189 | 66.2959 | Yes | 19.1551 X | 19.1031 X | 76% | 67.30% | 1.3289 |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 179,113 | 0.0022 | 105,459 | 80.6411 | Yes | 3.0554X | 6.6863X | 63.81% | 27.45% | 1.6984 |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 30,505 | 0.0005 | 4,506 | 67.933 | Yes | 0.11X | 2.6592X | 0.98% | 0.49% | 6.7698 |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 69,327 | 0.0021 | 17,831 | 75.9138 | Yes | 0.4863X | 6.9415X | 18.51% | 0.60% | 3.888 |
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 14,691 | 0.0002 | 2,151 | 66.9916 | Yes | 0.0518X | 1.1579X | 0.88% | 0.33% | 6.8298 |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 31,179 | 0.0005 | 17,433 | 79.7627 | Yes | 0.4996X | 1.2405X | 31.70% | 0.47% | 1.7885 |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 19,987 | 0.0005 | 2,641 | 66.7039 | Yes | 0.0633X | 1.7984X | 0.82% | 0.34% | 7.5679 |
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 12,754 | 0.0002 | 1,254 | 63.067 | Yes | 0.0284X | 0.6752X | 0.73% | 0.14% | 10.1706 |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 16,059 | 0.0003 | 2,295 | 75.2492 | Yes | 0.062X | 1.5583X | 0.83% | 0.34% | 6.9973 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 2,666 | 0.0001 | 205 | 60.0878 | Yes | 0.0044X | 0.088X | 0.38% | 0% | 13.0048 |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 13,009 | 0.0008 | 2,242 | 73.0772 | Yes | 0.0589X | 1.5239X | 0.76% | 0.34% | 5.8024 |
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 3,231 | 0.0002 | 593 | 59.5734 | Yes | 0.0127X | 0.1401X | 1.12% | 0% | 5.4485 |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 1,159 | 0.0013 | 597 | 52.9782 | Yes | 0.0114X | 0.1181X | 1.08% | 0% | 1.9413 |
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 33,050 | 0.0019 | 22,334 | 80.9308 | Yes | 0.6494X | 1.2031X | 38.25% | 0.73% | 1.4798 |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 5,571 | 0.0003 | 585 | 64.7949 | Yes | 0.0136X | 0.3767X | 0.50% | 0.05% | 9.523 |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 32,003 | 0.0004 | 3,227 | 63.6142 | Yes | 0.0737X | 2.0882X | 0.89% | 0.37% | 9.9172 |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 4,487 | 0.0002 | 1,249 | 64.767 | Yes | 0.0291X | 0.2025X | 2.64% | 0.02% | 3.5924 |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 7,247 | 0.0007 | 2,574 | 59.521 | Yes | 0.055X | 0.3348X | 4.71% | 0.05% | 2.8154 |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 7,021 | 0.0004 | 3,021 | 58.6011 | Yes | 0.0636X | 0.2703X | 5.97% | 0% | 2.324 |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 2,351 | 0.0001 | 213 | 62.0704 | Yes | 0.0047X | 0.1422X | 0.24% | 0.02% | 11.0375 |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 5,223 | 0.0005 | 786 | 65.9606 | Yes | 0.0186X | 0.5156X | 0.56% | 0.09% | 6.645 |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 2,419 | 0.0001 | 156 | 49.3205 | Yes | 0.0028X | 0.1032X | 0.14% | 0.01% | 15.5064 |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 56,028 | 0.0029 | 31,943 | 78.3827 | Yes | 0.8995X | 2.5599X | 43.62% | 1.91% | 1.7539 |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 427,101 | 0.0008 | 37,019 | 66.1598 | Yes | 0.8799X | 14.1548X | 19.14% | 0.91% | 11.5373 |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 136,536 | 0.0097 | 98,580 | 62.5842 | Yes | 2.2166X | 2.6431X | 62.77% | 17.75% | 1.385 |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 31,794 | 0.0025 | 22,669 | 53.892 | Yes | 0.4389X | 0.7979X | 31.01% | 0.14% | 1.4025 |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 7,177 | 0.0004 | 929 | 54.3423 | Yes | 0.0181X | 0.4775X | 0.64% | 0.07% | 7.7255 |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 3,076 | 0.0001 | 179 | 54.4078 | Yes | 0.0035X | 0.1216X | 0.19% | 0.01% | 17.1843 |
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 62,697 | 0.0014 | 9,358 | 69.6042 | Yes | 0.234X | 3.3737X | 11.12% | 0.45% | 6.6998 |
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 1,557 | 0.0001 | 111 | 36.5766 | Yes | 0.0015X | 0.0668X | 0.09% | 0% | 14.027 |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 3,640 | 0.0002 | 999 | 63.5285 | Yes | 0.0228X | 0.1993X | 1.95% | 0.02% | 3.6436 |
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 18,600 | 0.0031 | 12,042 | 75.2386 | Yes | 0.3255X | 0.9301X | 23.34% | 0.20% | 1.5445 |
| SBNM948 | 14,347,850 | 10,598,993 | 0.7387 | 17,255 | 0.0016 | 3,010 | 59.0498 | Yes | 0.0639X | 1.5135X | 1.28% | 0.34% | 5.7325 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 7,706 | 0.0009 | 5,134 | 64.2043 | Yes | 0.1184X | 0.3724X | 10.46% | 0.01% | 1.5009 |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 13,946 | 0.002 | 5,337 | 65.202 | Yes | 0.125X | 1.0913X | 8.67% | 0.24% | 2.613 |
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 249,452 | 0.0172 | 186,478 | 71.1307 | Yes | 4.7657X | 5.2705X | 68.99% | 44.69% | 1.3377 |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 3,516 | 0.0001 | 355 | 56.6873 | Yes | 0.0072X | 0.1254X | 0.57% | 0.01% | 9.9042 |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 25,665 | 0.0021 | 20,078 | 70.6514 | Yes | 0.5097X | 0.9332X | 33.50% | 0.28% | 1.2782 |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 21,426 | 0.0011 | 3,530 | 66.9725 | Yes | 0.0849X | 2.1098X | 0.92% | 0.41% | 6.0696 |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 103,691 | 0.0052 | 82,733 | 45.7145 | Yes | 1.3588X | 1.7018X | 58.42% | 4.87% | 1.2533 |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 43,229 | 0.0010 | 9,724 | 63.0661 | Yes | 0.2203X | 3.1771X | 10.90% | 0.45% | 4.4455 |
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 6,488 | 0.0003 | 1,395 | 49.4523 | Yes | 0.0248X | 0.2076X | 2.25% | 0.01% | 4.6508 |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 104,777 | 0.0062 | 88,425 | 56.7412 | Yes | 1.8026X | 2.2651X | 60.55% | 11.21% | 1.1849 |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 7,673 | 0.0003 | 2,816 | 54.8462 | Yes | 0.0555X | 0.2705X | 5.17% | 0.01% | 2.7247 |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 7,997 | 0.0003 | 3,654 | 58.9431 | Yes | 0.0774X | 0.3038X | 7.13% | 0.01% | 2.1885 |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 176,245 | 0.0077 | 144,752 | 56.6456 | Yes | 2.9459X | 3.577X | 65.88% | 27.17% | 1.2175 |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 22,255 | 0.0011 | 15,845 | 52.1673 | Yes | 0.297X | 0.6218X | 23.16% | 0.05% | 1.4045 |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 20,338 | 0.0010 | 12,641 | 59.9956 | Yes | 0.2725X | 0.6212X | 21.48% | 0.07% | 1.6088 |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 21,000 | 0.0008 | 15,343 | 69.1601 | Yes | 0.3812X | 0.7309X | 27.55% | 0.09% | 1.3687 |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 21,287 | 0.0013 | 16,744 | 74.6614 | Yes | 0.4492X | 0.8149X | 30.74% | 0.20% | 1.2713 |
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 22,623 | 0.0011 | 3,651 | 66.5316 | Yes | 0.0873X | 2.0963X | 0.86% | 0.41% | 6.1963 |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 3,703 | 0.0002 | 1,262 | 67.7607 | Yes | 0.0307X | 0.1901X | 2.92% | 0% | 2.9342 |
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 85,035 | 0.0059 | 71,329 | 56.8604 | Yes | 1.4572X | 1.9077X | 56.40% | 6.81% | 1.1921 |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 3,450 | 0.0002 | 203 | 46.2365 | Yes | 0.0034X | 0.1007X | 0.23% | 0.01% | 16.995 |
| NMNHextNeg A | 4,113,971 | 580,441 | 0.141 | 105 | 0.0002 | 2 | 37 | No | N/A | N/A | N/A | N/A | N/A |
| NMNHextNeg B | 3,868,171 | 1,273,359 | 0.3291 | 195 | 0.0002 | 5 | 39 | No | N/A | N/A | N/A | N/A | N/A |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | - | - | - | no unique mapped reads | BAM file was empty | N/A | N/A | N/A | N/A | N/A |

# Supplemental 8

*Tannerella forsythia* mapping summary

| Sample | Total Reads | Analysis-ready reads | Proportion of reads kept post-Adapter-Removal | Mapped reads | Proportion of reads mapped | Unique mapped reads | Average length of mapped reads | Rescaled or Not | Mean coverage | Std dev of mean coverage | % ref covered >= 1X | % ref covered >= 5X |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsENB | 1,982,046 | 12,067 | 0.006 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 2 | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 102 | 0 | 1 | 49 | No | 0X | 0.0038X | 0% | 0% |
| dsExtNegE | 2,068,072 | 194,559 | 0.094 | 18 | 0 | 2 | 82 | No | 0X | 0.0069X | 0% | 0% |
| dsExtNegF | 5,075,412 | 243,767 | 0.048 | 14 | 0 | 1 | 39 | No | 0X | 0.0034X | 0% | 0% |
| dsLibNegE | 2,352,887 | 1,510 | 0.0006 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsLibNegF | 1,673,188 | 698 | 0.0004 | 1 | 0.0014 | 1 | 33 | No | 0X | 0.0031X | 0% | 0% |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsLNB | 3,362,917 | 434 | 0.0001 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| dsLND | 4,612,074 | 17,550 | 0.0038 | - | 0 | - | no unique mapped reads | No | N/A | N/A | N/A | N/A |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 42,171 | 0.0007 | 21,281 | 76.0022 | Yes | 0.475X | 4.6238X | 3.71% | 1.89% |
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 28,371 | 0.0008 | 13,953 | 61.8774 | Yes | 0.2535X | 3.2436X | 2.88% | 0.99% |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 27,962 | 0.001 | 12,638 | 67.3079 | Yes | 0.2498X | 2.824X | 3.10% | 1.05% |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 16,715 | 0.0004 | 7,893 | 73.7305 | Yes | 0.1709X | 1.7368X | 3.27% | 1.06% |
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 13,220 | 0.0003 | 5,544 | 75.6917 | Yes | 0.1232X | 1.6722X | 2.84% | 0.72% |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 35,750 | 0.0008 | 18,753 | 76.4756 | Yes | 0.4212X | 4.0188X | 3.57% | 1.71% |
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 49,824 | 0.0009 | 14,649 | 76.5557 | Yes | 0.3293X | 3.1446X | 3.67% | 1.62% |
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 50,313 | 0.0009 | 15,102 | 63.4507 | Yes | 0.2814X | 2.9409X | 3.75% | 1.34% |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 56,528 | 0.001 | 19,766 | 63.1987 | Yes | 0.3668X | 4.6058X | 3.84% | 1.27% |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 25,946 | 0.0005 | 9,838 | 62.9627 | Yes | 0.1819X | 2.1872X | 2.72% | 0.85% |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 66,120 | 0.0013 | 33,946 | 56.7609 | Yes | 0.5658X | 5.5784X | 4.88% | 1.95% |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 33,253 | 0.0005 | 9,301 | 68.564 | Yes | 0.1872X | 2.2092X | 2.10% | 0.87% |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 43,310 | 0.001 | 16,826 | 65.4215 | Yes | 0.3232X | 4.4117X | 2.83% | 1.03% |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 35,615 | 0.0005 | 11,569 | 79.2415 | Yes | 0.2692X | 2.9424X | 3.40% | 1.31% |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 45,304 | 0.0006 | 15,720 | 82.2404 | Yes | 0.3796X | 4.1025X | 3.45% | 1.45% |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 50,593 | 0.0011 | 25,864 | 64.9341 | Yes | 0.4932X | 5.5012X | 4.05% | 1.74% |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 38,484 | 0.001 | 19,414 | 59.5262 | Yes | 0.3393X | 3.7452X | 3.47% | 1.21% |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 29,392 | 0.0008 | 13,107 | 75.2403 | Yes | 0.2896X | 2.5579X | 3.73% | 1.56% |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 43,984 | 0.0009 | 21,034 | 60.091 | Yes | 0.3712X | 3.5696X | 4.68% | 1.60% |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 17,403 | 0.0004 | 6,965 | 72.7945 | Yes | 0.1489X | 1.859X | 2.31% | 0.71% |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 36,205 | 0.0005 | 15,402 | 79.1173 | Yes | 0.3578X | 3.2919X | 4.14% | 1.80% |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 59,977 | 0.0007 | 26,112 | 62.6494 | Yes | 0.4804X | 3.9865X | 4.98% | 2.17% |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 44,951 | 0.0005 | 17,751 | 70.4293 | Yes | 0.3671X | 4.6106X | 3.01% | 1.19% |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 57,432 | 0.001 | 27,792 | 66.2908 | Yes | 0.541X | 6.3032X | 3.72% | 1.53% |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 27,472 | 0.0008 | 12,322 | 72.4928 | Yes | 0.2623X | 2.9804X | 3.28% | 1.08% |
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 30,371 | 0.0005 | 12,496 | 68.0465 | Yes | 0.2497X | 3.7267X | 2.53% | 0.87% |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 74,020 | 0.0013 | 34,150 | 77.9257 | Yes | 0.7814X | 10.3583X | 4.36% | 1.67% |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 30,025 | 0.0007 | 13,805 | 65.9025 | Yes | 0.2671X | 3.5671X | 2.58% | 0.96% |
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 28,505 | 0.0006 | 12,140 | 70.5887 | Yes | 0.2517X | 2.5788X | 3.64% | 1.23% |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 31,180 | 0.0007 | 14,215 | 68.7515 | Yes | 0.287X | 2.8355X | 3.73% | 1.34% |
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 6,869 | 0.0003 | 2,417 | 70.2826 | Yes | 0.0499X | 0.4702X | 1.92% | 0.31% |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 11,848 | 0.0007 | 5,921 | 62.5768 | Yes | 0.1088X | 1.1209X | 2.56% | 0.66% |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 11,539 | 0.0008 | 5,230 | 57.4979 | Yes | 0.0883X | 1.0621X | 1.90% | 0.54% |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 780 | 0.0009 | 425 | 52.8376 | Yes | 0.0066X | 0.1001X | 0.52% | 0% |
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 9,954 | 0.0005 | 4,107 | 66.8227 | Yes | 0.0806X | 0.9807X | 1.81% | 0.51% |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 24,992 | 0.0014 | 12,336 | 62.8534 | Yes | 0.2277X | 2.6872X | 3.11% | 1.01% |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 47,942 | 0.0006 | 19,726 | 61.7229 | Yes | 0.3575X | 4.5625X | 3.27% | 1.21% |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 22,224 | 0.0011 | 10,342 | 63.1075 | Yes | 0.1917X | 2.1243X | 3.05% | 0.95% |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 8,334 | 0.0008 | 3,575 | 59.8769 | Yes | 0.0629X | 0.7179X | 1.96% | 0.38% |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 14,176 | 0.0009 | 7,521 | 60.6362 | Yes | 0.1339X | 0.6521X | 7.98% | 0.41% |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 14,317 | 0.0011 | 6,849 | 63.0183 | Yes | 0.1267X | 1.712X | 1.97% | 0.62% |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 6,108 | 0.0006 | 3,113 | 66.3951 | Yes | 0.0607X | 0.7711X | 1.99% | 0.35% |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 17,212 | 0.0012 | 7,934 | 60.1012 | Yes | 0.14X | 1.6885X | 2.24% | 0.74% |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 16,521 | 0.0008 | 7,569 | 68.9117 | Yes | 0.1532X | 1.8679X | 2.78% | 0.73% |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 540,427 | 0.0011 | 140,442 | 70.1979 | Yes | 2.8949X | 36.7859X | 6.03% | 2.93% |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 15,483 | 0.0011 | 6,942 | 62.501 | Yes | 0.1274X | 1.103X | 3.24% | 0.83% |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 9,586 | 0.0007 | 3,591 | 49.6906 | Yes | 0.0524X | 0.4923X | 2.21% | 0.29% |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 24,056 | 0.0014 | 12,236 | 60.3987 | Yes | 0.217X | 2.1812X | 3.53% | 1.16% |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 11,110 | 0.0007 | 4,835 | 58.9344 | Yes | 0.0837X | 1.0197X | 1.93% | 0.51% |
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 34,028 | 0.0007 | 10,699 | 66.3377 | Yes | 0.2084X | 2.0076X | 3.56% | 1.22% |
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 1,036 | 0 | 7 | 40.4286 | No | 0.0001X | 0.0102X | 0.01% | 0% |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 9,095 | 0.0006 | 4,208 | 64.7854 | Yes | 0.0801X | 0.8902X | 2.04% | 0.51% |
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 6,009 | 0.001 | 2,948 | 72.711 | Yes | 0.0629X | 0.8208X | 1.58% | 0.38% |
| SBNM948 | 14,347,850 | 10,598,993 | 0.7387 | 11,137 | 0.001 | 6,058 | 56.9777 | Yes | 0.1014X | 1.0444X | 2.74% | 0.61% |
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 7,155 | 0.0009 | 3,221 | 60.6209 | Yes | 0.0573X | 0.562X | 2.24% | 0.34% |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 4,552 | 0.0006 | 2,366 | 61.3931 | Yes | 0.0427X | 0.5634X | 1.38% | 0.28% |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 14,044 | 0.0009 | 5,997 | 69.4444 | Yes | 0.1223X | 1.11X | 3.04% | 0.81% |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 22,745 | 0.0012 | 10,517 | 58.9053 | Yes | 0.1819X | 2.2416X | 2.59% | 0.85% |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 9,206 | 0.0007 | 4,427 | 66.0865 | Yes | 0.0859X | 0.7685X | 2.59% | 0.61% |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 19,357 | 0.001 | 8,954 | 61.8416 | Yes | 0.1626X | 1.6038X | 3.22% | 0.89% |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 156,725 | 0.0077 | 128,020 | 45.428 | Yes | 1.7077X | 3.3374X | 39.33% | 13.66% |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 37,021 | 0.0008 | 19,080 | 58.2771 | Yes | 0.3265X | 3.5289X | 3.25% | 1.29% |
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 9,379 | 0.0004 | 3,769 | 51.875 | Yes | 0.0574X | 0.6968X | 1.81% | 0.36% |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 17,485 | 0.001 | 9,502 | 55.5823 | Yes | 0.1551X | 1.3631X | 3.23% | 1% |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 5,730 | 0.0002 | 573 | 50.0785 | Yes | 0.0084X | 0.2168X | 0.53% | 0.01% |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 14,536 | 0.0005 | 6,778 | 58.3504 | Yes | 0.1161X | 1.4051X | 2.04% | 0.68% |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 33,443 | 0.0014 | 21,006 | 59.9296 | Yes | 0.3697X | 1.9408X | 13.93% | 1.48% |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 12,498 | 0.0006 | 5,793 | 52.1507 | Yes | 0.0887X | 0.8777X | 2.56% | 0.57% |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 11,950 | 0.0006 | 4,661 | 53.3358 | Yes | 0.073X | 0.7469X | 2.34% | 0.45% |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 12,334 | 0.0004 | 5,867 | 61.8173 | Yes | 0.1065X | 1.0174X | 2.71% | 0.68% |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 18,150 | 0.001 | 10,365 | 64.1325 | Yes | 0.1952X | 2.277X | 2.63% | 0.91% |
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 12,229 | 0.0005 | 5,994 | 59.2107 | Yes | 0.1042X | 1.1124X | 2.46% | 0.66% |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 8,138 | 0.0003 | 3,257 | 66.3417 | Yes | 0.0634X | 0.7726X | 1.65% | 0.41% |
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 10,143 | 0.0007 | 5,289 | 55.4997 | Yes | 0.0862X | 0.7679X | 2.76% | 0.53% |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 15,703 | 0.0009 | 9,223 | 54.1751 | Yes | 0.1467X | 1.4209X | 2.77% | 0.92% |
| NMNHextNegA | 4,113,971 | 580,441 | 0.141 | 61 | 0.0001 | 4 | 40.25 | Yes | 0X | 0.0104X | 0% | 0% |
| NMNHextNegB | 3,868,171 | 1,273,359 | 0.3291 | 91 | 0 | 5 | 39.2 | No | 0.0001X | 0.0135X | 0% | 0% |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | 1 | 0 | 1 | 144 | No | 0X | 0.0065X | 0% | 0% |

# Supplemental 9

*Klebsiella pneumoniae* mapping summary

| Sample | Total Reads | Analysis-ready reads (trimmed and merged) | Proportion of reads kept after trimming and merging (post-AdapterRemoval) | Q20 Mapped reads | Proportion of reads mapped (Mapped reads / Analysis-ready reads) | Unique Q20 mapped reads | Average length of mapped reads | Rescaled or Not |
|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | - | 0 | - | There were no unique mapped reads | BAM file was empty |
| dsENB | 1,982,046 | 12,067 | 0.006 | 1 | 0 | - | There were no unique mapped reads | BAM file was empty |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 7 | 0.0001 | - | There were no unique mapped reads | BAM file was empty |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 168 | 0 | 3 | 31.6667 | Yes |
| dsExtNegE | 2,068,072 | 194,559 | 0.094 | 45 | 0.0002 | 5 | 42.2 | Yes |
| dsExtNegF | 5,075,412 | 243,767 | 0.048 | 42 | 0.0001 | 5 | 57.8 | Yes |
| dsLibNegE | 2,352,887 | 1,510 | 0.0006 | - | 0 | - | There were no unique mapped reads | BAM file was empty |
| dsLibNegF | 1,673,188 | 698 | 0.0004 | - | 0 | - | There were no unique mapped reads | BAM file was empty |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | There were no unique mapped reads | BAM file was empty |
| dsLNB | 3,362,917 | 434 | 0.0001 | 2 | 0.0046 | - | There were no unique mapped reads | BAM file was empty |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | There were no unique mapped reads | BAM file was empty |
| dsLND | 4,612,074 | 17,550 | 0.0038 | - | 0 | - | There were no unique mapped reads | BAM file was empty |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 11,211 | 0.0001 | 168 | 46.0119 | Yes |
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 8,488 | 0.0002 | 274 | 38.6277 | Yes |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 6,315 | 0.0002 | 151 | 40.3444 | Yes |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 3,923 | 0.0001 | 94 | 36.734 | Yes |
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 5,310 | 0.0001 | 137 | 40.9854 | Yes |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 5,562 | 0.0001 | 110 | 38.9364 | Yes |
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 6,496 | 0.0001 | 62 | 42.0161 | Yes |
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 7,855 | 0.0001 | 102 | 40.3431 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 15,565 | 0.0002 | 607 | 47.6129 | Yes |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 13,115 | 0.0002 | 277 | 40.7906 | Yes |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 15,261 | 0.0003 | 611 | 37.0442 | Yes |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 15,332 | 0.0002 | 300 | 40.0467 | Yes |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 10,350 | 0.0002 | 336 | 47.8065 | Yes |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 8,245 | 0.0001 | 142 | 37.9085 | Yes |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 12,268 | 0.0001 | 274 | 38.2774 | Yes |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 10,412 | 0.0002 | 332 | 37.7229 | Yes |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 10,332 | 0.0002 | 375 | 35.3227 | Yes |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 6,429 | 0.0001 | 223 | 37.6906 | Yes |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 11,854 | 0.0002 | 367 | 42.6839 | Yes |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 6,640 | 0.0001 | 174 | 36.9943 | Yes |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 9,299 | 0.0001 | 118 | 43.8475 | Yes |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 18,225 | 0.0002 | 631 | 36.8241 | Yes |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 14,917 | 0.0001 | 250 | 41.364 | Yes |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 13,210 | 0.0002 | 387 | 44.7933 | Yes |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 4,454 | 0.0001 | 124 | 38.5645 | Yes |
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 12,744 | 0.0002 | 391 | 37.3504 | Yes |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 6,270 | 0.0001 | 155 | 38.0129 | Yes |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 7,162 | 0.0001 | 151 | 40.5695 | Yes |
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 8,329 | 0.0001 | 219 | 38.2922 | Yes |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 5,497 | 0.0001 | 130 | 36.9769 | Yes |
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 6,796 | 0.0003 | 169 | 41.6391 | Yes |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 2,577 | 0.0001 | 67 | 38.0299 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 4,930 | 0.0003 | 189 | 37.6085 | Yes |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 241 | 0.0002 | 9 | 36.1111 | Yes |
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 3,457 | 0.0002 | 144 | 52.6042 | Yes |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 3,930 | 0.0002 | 165 | 36.503 | Yes |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 10,384 | 0.0001 | 398 | 57.9196 | Yes |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 4,359 | 0.0002 | 144 | 40.1667 | Yes |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 2,273 | 0.0002 | 61 | 38.0656 | Yes |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 7,568 | 0.0004 | 493 | 37.927 | Yes |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 2,935 | 0.0002 | 84 | 36.3095 | Yes |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 1,239 | 0.0001 | 18 | 38.2778 | Yes |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 4,631 | 0.0003 | 134 | 39.5672 | Yes |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 3,329 | 0.0001 | 78 | 36.3462 | Yes |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 70,524 | 0.0001 | 1,469 | 42.501 | Yes |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 2,304 | 0.0001 | 49 | 38.6531 | Yes |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 11,541 | 0.0009 | 3,971 | 62.7477 | Yes |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 3,962 | 0.0002 | 102 | 37.7059 | Yes |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 5,032 | 0.0003 | 147 | 37.9932 | Yes |
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 7,737 | 0.0001 | 161 | 37.2919 | Yes |
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 1,008 | 0 | 13 | 49.8462 | Yes |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 2,811 | 0.0001 | 48 | 41.8333 | Yes |
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 899 | 0.0001 | 31 | 38.4194 | Yes |
| SBNM948 | 14,347,850 | 10,598,993 | 0.7387 | 2,104 | 0.0001 | 91 | 36.2637 | Yes |
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 1,606 | 0.0002 | 47 | 40.2979 | Yes |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 1,408 | 0.0002 | 40 | 37.55 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 2,810 | 0.0001 | 74 | 40.2838 | Yes |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 5,441 | 0.0003 | 160 | 39.1625 | Yes |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 2,506 | 0.0002 | 54 | 39.7593 | Yes |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 3,774 | 0.0002 | 117 | 38.6838 | Yes |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 10,761 | 0.0005 | 680 | 35.7588 | Yes |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 8,114 | 0.0001 | 152 | 39.5263 | Yes |
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 9,741 | 0.0004 | 482 | 36.1141 | Yes |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 3,521 | 0.0002 | 73 | 38.4247 | Yes |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 8,913 | 0.0003 | 325 | 37.6062 | Yes |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 7,314 | 0.0002 | 164 | 42.1951 | Yes |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 6,463 | 0.0002 | 446 | 54.7108 | Yes |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 6,513 | 0.0003 | 287 | 37.2021 | Yes |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 7,548 | 0.0003 | 605 | 42.2116 | Yes |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 5,082 | 0.0002 | 87 | 42.7241 | Yes |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 3,378 | 0.0001 | 105 | 37.7905 | Yes |
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 4,205 | 0.0001 | 77 | 39.9481 | Yes |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 4,602 | 0.0002 | 105 | 43.5429 | Yes |
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 4,643 | 0.0003 | 159 | 36.5786 | Yes |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 5,896 | 0.0003 | 228 | 38.0614 | Yes |
| NMNHextNegA | 4,113,971 | 580,441 | 0.141 | 365 | 0.0006 | 34 | 39.2941 | Yes |
| NMNHextNegB | 3,868,171 | 1,273,359 | 0.3291 | 681 | 0.0005 | 53 | 42.7736 | Yes |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | - | 0 | - | There were no unique mapped reads | BAM file was empty |

# Supplemental 10

*Toxoplasma gondii* mapping summary

| Sample | Total Reads | Analysis-ready reads | Proportion of reads kept post-Adapter-Removal | Mapped reads | Proportion of reads mapped | Unique mapped reads | Avg. length of mapped reads | Rescaled or Not | Mean cove-rage | Std dev of mean coverage | % ref. covered >= 1X | % ref. covered >= 5X | Cluster factor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| dsENA | 1,153,001 | 6,250 | 0.0054 | 4 | 0.00064 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| dsENB | 1,982,046 | 12,067 | 0.006 | 12 | 0.00099445 | 1 | 73 | No | N/A | N/A | N/A | N/A | N/A |
| dsENC | 10,924,554 | 60,701 | 0.0055 | 48 | 0.00079076 | 3 | 46.3333 | No | N/A | N/A | N/A | N/A | N/A |
| dsEND | 6,808,226 | 2,919,191 | 0.4287 | 5,248 | 0.00179776 | 50 | 62.66 | Yes | 0X | 0.0039X | 0% | 0% | 104.96 |
| dsExtNegE | 2,068,072 | 194,559 | 0.094 | 121 | 0.00062192 | 13 | 51 | Yes | 0X | 0.0016X | 0% | 0% | 9.3076 |
| dsExtNegF | 5,075,412 | 243,767 | 0.048 | 228 | 0.00093532 | 17 | 67.7059 | Yes | 0X | 0.0029X | 0% | 0% | 13.4117 |
| dsLibNegE | 2,352,887 | 1,510 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| dsLibNegF | 1,673,188 | 698 | 0.0004 | 5 | 0.00716332 | 1 | 51 | No | N/A | N/A | N/A | N/A | N/A |
| dsLNA | 1,362,415 | 3,086 | 0.0022 | - | 0 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| dsLNB | 3,362,917 | 434 | 0.0001 | - | 0 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| dsLNC | 10,043,828 | 7,021 | 0.0006 | - | 0 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| dsLND | 4,612,074 | 17,550 | 0.0038 | 182 | 0.01037037 | 1 | 84 | No | N/A | N/A | N/A | N/A | N/A |
| NMML594955 | 65,728,408 | 59,498,044 | 0.9052 | 6,170 | 0.0001037 | 1,099 | 66.0446 | Yes | 0.0003X | 0.0634X | 0% | 0% | 5.6141 |
| NMML594958 | 35,152,363 | 31,734,352 | 0.9027 | 2,578 | 8.1237E-05 | 62 | 51.5645 | Yes | 0X | 0.0064X | 0% | 0% | 41.5806 |
| NMML605032 | 30,761,890 | 27,044,123 | 0.8791 | 1,899 | 7.0219E-05 | 125 | 59.776 | Yes | 0X | 0.0086X | 0% | 0% | 15.192 |
| NMML605034 | 37,882,083 | 33,440,234 | 0.8827 | 904 | 2.7033E-05 | 25 | 63.12 | Yes | 0X | 0.0027X | 0% | 0% | 36.16 |
| NMML605036 | 38,670,150 | 33,523,354 | 0.8669 | 1,471 | 4.388E-05 | 123 | 63.2846 | Yes | 0X | 0.009X | 0% | 0% | 11.9593 |
| NMML605046 | 47,143,190 | 41,860,400 | 0.8879 | 1,655 | 3.9536E-05 | 182 | 62.5714 | Yes | 0X | 0.0112X | 0% | 0% | 9.0934 |
| NMML606312 | 59,723,018 | 54,241,046 | 0.9082 | 1,810 | 3.337E-05 | 64 | 53.3906 | Yes | 0X | 0.0053X | 0% | 0% | 28.2812 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606316 | 57,156,414 | 52,273,909 | 0.9145 | 2,313 | 4.4248E-05 | 27 | 46.9259 | Yes | 0X | 0.0025X | 0% | 0% | 85.6666 |
| NMML606323 | 59,527,711 | 54,867,830 | 0.9217 | 2,185 | 3.9823E-05 | 30 | 46.6667 | Yes | 0X | 0.0038X | 0% | 0% | 72.8333 |
| NMML606325 | 50,777,151 | 46,333,884 | 0.9124 | 2,329 | 5.0266E-05 | 59 | 52.7288 | Yes | 0X | 0.0049X | 0% | 0% | 39.4745 |
| NMML606334 | 51,308,495 | 48,264,303 | 0.9406 | 3,450 | 7.1481E-05 | 139 | 49.482 | Yes | 0X | 0.0086X | 0% | 0% | 24.8201 |
| NMML606340 | 65,325,647 | 59,467,102 | 0.9103 | 24,224 | 0.00040735 | 662 | 48.932 | Yes | 0.0001X | 0.0407X | 0% | 0% | 36.5921 |
| NMML606345 | 45,726,026 | 43,144,840 | 0.9435 | 1,794 | 4.1581E-05 | 40 | 56.925 | Yes | 0X | 0.0036X | 0% | 0% | 44.85 |
| NMML606348 | 66,963,615 | 62,726,537 | 0.9367 | 3,420 | 5.4522E-05 | 226 | 55.9027 | Yes | 0X | 0.0131X | 0% | 0% | 15.1327 |
| NMML606350 | 77,428,327 | 71,493,952 | 0.9233 | 3,671 | 5.1347E-05 | 241 | 57.6307 | Yes | 0.0001X | 0.0133X | 0% | 0% | 15.2323 |
| NMML606353 | 49,634,441 | 42,588,122 | 0.858 | 2,160 | 5.0718E-05 | 22 | 49.4545 | Yes | 0X | 0.0032X | 0% | 0% | 98.1818 |
| NMML606376 | 44,408,652 | 37,877,421 | 0.8529 | 2,694 | 7.1124E-05 | 51 | 43.451 | Yes | 0X | 0.0041X | 0% | 0% | 52.8235 |
| NMML606378 | 41,644,691 | 35,801,392 | 0.8596 | 8,731 | 0.00024387 | 373 | 65.193 | Yes | 0.0001X | 0.0365X | 0% | 0% | 23.4075 |
| NMML606379 | 53,598,337 | 44,735,065 | 0.8346 | 3,130 | 6.9967E-05 | 47 | 53.6809 | Yes | 0X | 0.0047X | 0% | 0% | 66.5957 |
| NMML606380 | 42,521,083 | 38,801,722 | 0.9125 | 1,459 | 3.7601E-05 | 44 | 52.6818 | Yes | 0X | 0.0043X | 0% | 0% | 33.159 |
| NMML606387 | 84,460,612 | 72,293,991 | 0.8559 | 4,031 | 5.5758E-05 | 491 | 69.3116 | Yes | 0.0001X | 0.0328X | 0% | 0% | 8.2097 |
| NMML606388 | 88,579,894 | 78,363,858 | 0.8846 | 16,187 | 0.00020656 | 982 | 53.2749 | Yes | 0.0002X | 0.0434X | 0.01% | 0% | 16.4837 |
| NMML606393 | 92,270,890 | 80,178,869 | 0.8689 | 3,628 | 4.5249E-05 | 131 | 55.9695 | Yes | 0X | 0.0088X | 0% | 0% | 27.6946 |
| NMML606394 | 64,883,563 | 57,362,812 | 0.884 | 3,830 | 6.6768E-05 | 147 | 53.3401 | Yes | 0X | 0.011X | 0% | 0% | 26.0544 |
| NMML606395 | 37,605,076 | 32,643,410 | 0.868 | 1,229 | 3.7649E-05 | 57 | 53.4035 | Yes | 0X | 0.0045X | 0% | 0% | 21.5614 |
| NMML606398 | 63,287,248 | 55,201,712 | 0.8722 | 2,591 | 4.6937E-05 | 195 | 92.4769 | Yes | 0.0001X | 0.0436X | 0% | 0% | 13.2871 |
| NMML606399 | 57,636,607 | 52,910,925 | 0.918 | 2,129 | 4.0237E-05 | 107 | 63.2617 | Yes | 0X | 0.0123X | 0% | 0% | 19.8971 |
| NMML606400 | 44,636,965 | 39,285,679 | 0.8801 | 2,462 | 6.2669E-05 | 90 | 55.6778 | Yes | 0X | 0.0064X | 0% | 0% | 27.3555 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMML606407 | 51,136,679 | 45,034,904 | 0.8806 | 2,274 | 5.0494E-05 | 152 | 56.6645 | Yes | 0X | 0.0096X | 0% | 0% | 14.9605 |
| NMML606408 | 45,600,231 | 41,117,218 | 0.9016 | 2,079 | 5.0563E-05 | 263 | 95.7909 | Yes | 0.0001X | 0.0669X | 0% | 0% | 7.9049 |
| SBNM10609 | 19,718,130 | 17,596,084 | 0.8923 | 1,442 | 8.195E-05 | 164 | 62.5366 | Yes | 0X | 0.01X | 0% | 0% | 8.7926 |
| SBNM1100 | 20,213,041 | 14,930,711 | 0.7386 | 732 | 4.9026E-05 | 21 | 53.8571 | Yes | 0X | 0.0035X | 0% | 0% | 34.8571 |
| SBNM1353 | 17,198,232 | 14,005,734 | 0.8143 | 883 | 6.3046E-05 | 21 | 60.6667 | Yes | 0X | 0.0032X | 0% | 0% | 42.0476 |
| SBNM1378 | 1,415,533 | 832,837 | 0.5883 | 63 | 7.5645E-05 | - | no unique mapped reads | BAM file empty | N/A | N/A | N/A | N/A | N/A |
| SBNM1463 | 19,950,831 | 16,649,240 | 0.8345 | 671 | 4.0302E-05 | 25 | 66.04 | Yes | 0X | 0.003X | 0% | 0% | 26.84 |
| SBNM1551 | 19,767,675 | 17,740,890 | 0.8974 | 1,665 | 9.3851E-05 | 58 | 67.8793 | Yes | 0X | 0.0104X | 0% | 0% | 28.7068 |
| SBNM2130 | 79,503,873 | 70,354,799 | 0.8849 | 2,597 | 3.6913E-05 | 68 | 73.2794 | Yes | 0X | 0.0054X | 0% | 0% | 38.1911 |
| SBNM2142 | 22,021,137 | 18,950,565 | 0.8605 | 1,457 | 7.6884E-05 | 46 | 53.8696 | Yes | 0X | 0.0055X | 0% | 0% | 31.6739 |
| SBNM216 | 11,388,418 | 9,598,568 | 0.8428 | 588 | 6.1259E-05 | 8 | 51.875 | Yes | 0X | 0.0014X | 0% | 0% | 73.5 |
| SBNM2357 | 17,789,705 | 15,253,329 | 0.8574 | 26,800 | 0.00175699 | 2,101 | 48.3812 | Yes | 0.0004X | 0.0787X | 0.01% | 0% | 12.7558 |
| SBNM2419 | 13,946,668 | 12,424,680 | 0.8908 | 1,050 | 8.4509E-05 | 160 | 78.9562 | Yes | 0X | 0.0081X | 0% | 0% | 6.5625 |
| SBNM243 | 17,968,771 | 8,779,521 | 0.4885 | 316 | 3.5993E-05 | 11 | 80.9091 | Yes | 0X | 0.0018X | 0% | 0% | 28.7272 |
| SBNM244 | 17,623,351 | 13,964,536 | 0.7923 | 871 | 6.2372E-05 | 10 | 58.9 | Yes | 0X | 0.0017X | 0% | 0% | 87.1 |
| SBNM2440 | 20,015,325 | 18,753,180 | 0.9369 | 960 | 5.1191E-05 | 28 | 73.5 | Yes | 0X | 0.0031X | 0% | 0% | 34.2857 |
| SBNM3283 | 520,428,216 | 484,289,635 | 0.9305 | 20,143 | 4.1593E-05 | 384 | 55.8516 | Yes | 0.0001X | 0.019X | 0% | 0% | 52.4557 |
| SBNM3450 | 16,275,126 | 13,993,223 | 0.8597 | 1,837 | 0.00013128 | 136 | 49.5662 | Yes | 0X | 0.0083X | 0% | 0% | 13.5073 |
| SBNM3822 | 14,560,300 | 12,486,156 | 0.8575 | 16,099 | 0.00128935 | 804 | 61.0672 | Yes | 0.0002X | 0.1053X | 0% | 0% | 20.0236 |
| SBNM3973 | 19,792,725 | 16,953,585 | 0.8565 | 1,346 | 7.9393E-05 | 46 | 50.9565 | Yes | 0X | 0.0035X | 0% | 0% | 29.2608 |
| SBNM4260 | 21,337,098 | 15,620,127 | 0.732 | 1,074 | 6.8757E-05 | 37 | 48.3243 | Yes | 0X | 0.0032X | 0% | 0% | 29.027 |
| SBNM5178 | 50,811,613 | 44,020,788 | 0.8663 | 1,829 | 4.1549E-05 | 42 | 49.5952 | Yes | 0X | 0.0043X | 0% | 0% | 43.5476 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SBNM8981 | 17,380,777 | 14,166,983 | 0.815 | 137,700 | 0.00971978 | 10,832 | 54.4558 | Yes | 0.0023 X | 1.12X | 0.01% | 0% | 12.7123 |
| SBNM919 | 16,712,549 | 14,665,995 | 0.8775 | 1,647 | 0.0001123 | 304 | 72.3553 | Yes | 0.0001 X | 0.0118X | 0.01% | 0% | 5.4177 |
| SBNM945 | 8,651,877 | 5,913,537 | 0.6834 | 358 | 6.0539E-05 | 15 | 64.0667 | Yes | 0X | 0.0025X | 0% | 0% | 23.8666 |
| SBNM948 | 14,347,850 | 10,598,993 | 0.7387 | 634 | 5.9817E-05 | 13 | 55.6923 | Yes | 0X | 0.0017X | 0% | 0% | 48.7692 |
| SBNM957 | 13,817,668 | 7,907,440 | 0.5722 | 522 | 6.6014E-05 | 31 | 57.4516 | Yes | 0X | 0.0034X | 0% | 0% | 16.8387 |
| SBNM958 | 10,229,462 | 6,965,729 | 0.6809 | 430 | 6.1731E-05 | 20 | 51.45 | Yes | 0X | 0.0029X | 0% | 0% | 21.5 |
| SBNM959 | 17,134,482 | 14,498,960 | 0.8461 | 1,712 | 0.00011808 | 118 | 56.7542 | Yes | 0X | 0.0079X | 0% | 0% | 14.5084 |
| SBNM960 | 20,006,820 | 18,102,614 | 0.9048 | 1,040 | 5.745E-05 | 16 | 49.8125 | Yes | 0X | 0.0031X | 0% | 0% | 65 |
| SBNM962 | 17,513,482 | 11,778,529 | 0.6725 | 5,303 | 0.00045023 | 649 | 55.6995 | Yes | 0.0001 X | 0.0383X | 0% | 0% | 8.171 |
| SBNM964 | 22,239,835 | 18,590,491 | 0.8359 | 5,013 | 0.00026965 | 1,259 | 72.1263 | Yes | 0.0003 X | 0.0333X | 0.02% | 0% | 3.9817 |
| NMNH105124 | 21,940,836 | 20,102,329 | 0.9162 | 3,585 | 0.0001 | 193 | 44.456 | Yes | 0X | 0.0096X | 0% | 0% | 18.5751 |
| NMNH259651 | 45,006,669 | 41,985,246 | 0.9328 | 2,344 | 0 | 53 | 55.566 | Yes | 0X | 0.0043X | 0% | 0% | 44.2264 |
| NMNH276052 | 24,611,743 | 22,684,707 | 0.9217 | 2,362 | 0.0001 | 166 | 58.8133 | Yes | 0X | 0.0225X | 0% | 0% | 14.2289 |
| NMNH276054 | 19,380,958 | 16,840,063 | 0.8688 | 1,256 | 0 | 102 | 50.2941 | Yes | 0X | 0.0062X | 0% | 0% | 12.3137 |
| NMNH395724 | 25,486,848 | 23,446,764 | 0.9199 | 4,772 | 0.0002 | 1,100 | 54.97 | Yes | 0.0002 X | 0.0511X | 0% | 0% | 4.3381 |
| NMNH395725 | 31,471,551 | 28,152,551 | 0.8945 | 1,432 | 0 | 50 | 59.72 | Yes | 0X | 0.0039X | 0% | 0% | 28.64 |
| NMNH504201 | 24,699,645 | 22,784,088 | 0.9224 | 1,695 | 0 | 101 | 57.8713 | Yes | 0X | 0.0143X | 0% | 0% | 16.7821 |
| NMNH504203 | 22,849,165 | 19,897,919 | 0.8708 | 1,768 | 0 | 121 | 50.7025 | Yes | 0X | 0.0071X | 0% | 0% | 14.6115 |
| NMNH593460 | 21,934,115 | 19,836,468 | 0.9043 | 4,082 | 0.0002 | 557 | 49.2334 | Yes | 0.0001 X | 0.0243X | 0% | 0% | 7.3285 |
| NMNH593476 | 28,148,793 | 24,717,796 | 0.8781 | 1,986 | 0 | 277 | 62.1047 | Yes | 0.0001 X | 0.0118X | 0% | 0% | 7.1696 |
| NMNH593495 | 18,250,229 | 17,018,983 | 0.9325 | 730 | 0 | 73 | 59.9726 | Yes | 0X | 0.0056X | 0% | 0% | 10 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NMNH593508 | 23,076,638 | 21,470,609 | 0.9304 | 1,218 | 0 | 81 | 55.3827 | Yes | 0X | 0.0063X | 0% | 0% | 15.037 |
| NMNH593517 | 23,645,460 | 21,972,414 | 0.9292 | 1,244 | 0 | 111 | 57.4234 | Yes | 0X | 0.0079X | 0% | 0% | 11.2072 |
| NMNHA21735 | 15,785,332 | 14,418,558 | 0.9134 | 922 | 0 | 29 | 57.7241 | Yes | 0X | 0.0038X | 0% | 0% | 31.7931 |
| NMNHA21736 | 19,291,927 | 16,569,038 | 0.8588 | 1,041 | 0 | 30 | 58.6667 | Yes | 0X | 0.0036X | 0% | 0% | 34.7 |
| NMNHextNegA | 4,113,971 | 580,441 | 0.141 | 70 | 0.0001 | 6 | 67.1667 | Yes | 0X | 0.0012X | 0% | 0% | 11.6666 |
| NMNHextNegB | 3,868,171 | 1,273,359 | 0.3291 | 67 | 0 | 7 | 54 | Yes | 0X | 0.0012X | 0% | 0% | 9.5714 |
| NMNHlibNeg | 2,538,168 | 48,804 | 0.0192 | 11 | 0.0002 | 1 | 58 | No | N/A | N/A | N/A | N/A | N/A |