The Role of Epigenetic Modifications and Microbiome Evolution in Bovid Adaptation to Environmental Changes

Yichen Liu

Supervised by
Dr Bastien Llamas,
Dr Laura S. Weyrich,
and Prof Alan Cooper

THE UNIVERSITY
of ADELAIDE

A thesis presented for the degree of Doctor of Philosophy

SCHOOL OF BIOLOGICAL SCIENCES
THE UNIVERSITY OF ADELAIDE

March 2019

The Role of Epigenetic Modifications and Microbiome Evolution in Bovid Adaptation to Environmental Changes

Yichen Liu

Abstract

Ancient DNA (aDNA) techniques have demonstrated its power to reveal past events and reconstruct the evolutionary histories of animals. However, many questions regarding the rapid adaptation of animals cannot be solely explained by genomic evidence and thus remain to be further investigated.

Gut bacterial communities (microbiota) perform essential functions for their hosts, including nutrient synthesis, dietary toxin degradation, and host immunity development. Epigenetics, including DNA methylation and microRNAs (miRNAs), participates in the regulation of the gene expression as well as many critical cellular processes. These non-genomic mechanisms tend to be highly dynamic and susceptible to internal (*i.e.,* genetics) and external factors (*i.e.,* environmental cues). It is possible that microbiota alterations and epigenetic modifications swiftly transfer external cues to animal phenotypic alterations and exert durative influence on animal fitness and adaptation, whereas the evidence from modern animal models remains scarce and controversial.

The most recently developed aDNA techniques allowed the recovery of microbiomic and epigenomic information from ancient animal remains. In this thesis, I employed advanced ancient DNA (aDNA) techniques to explore the role of non-genomic mechanisms in bovid adaptation to environmental changes over an evolutionary timescale. Two bovid taxa, *Myotragus* and *Bison*, were used for case studies to explore the possible roles of microbiota and epigenetics in animal adaption. In the first case study, I was able to find evidence suggesting the gut microbiota of *Myotragus* facilitate its adaptation to a toxic diet; in the second case study, the potential methylation hotspots responding to mammal-environment interactions were identified. Furthermore, novel protocols and bioinformatics tools were developed and optimised to retrieve ancient epigenetic and microbial information from extremely degraded and contaminated DNA sources. Overall, these studies suggest the microbiota and epigenetic modifications play a role in the adaption to the environment of the past animals. These findings also highlight the potential of aDNA techniques for resolving long-standing evolutionary questions.

Declaration

Yichen Liu

Signed:

Date: 04/03/2019

Acknowledgements

I'd like to thank my supervisors, Bastien, Laura, and Alan. Thank you for training me to do good science — I learnt to be critical and objective to my results as well as others'; I learnt how to design and manage a project; I learnt how to interpret and present my findings, and I learnt to be always curious and not giving up until I find the answer.

Bastien, I enjoyed the time when we had meetings and discussed project design and data interpretation. I like the lab work of the aRNA project the most, as this was some real old-school biology stuff.

Laura, thanks for training me to write scientific papers from scratch. A huge thanks for your kind help when I had trouble with my PhD, I could never make it this far without your help.

Alan, thanks for all the exciting ideas and inspiring discussions during the megafauna meetings, and you always knew the right person to ask when I needed help with my project.

I'd like to thank Pere for the goat poo jokes, the licking-tube-at-the-TB jokes, and all the fun time in the lab! Steve R., thank you for your help with my project! You know everything regarding the wet lab techniques!

Thank you, Maria, for always be there to keep ACAD functioning. You always know the answer to my questions and you saved me from milk-depletion crisis several times! Thank you, Holly, Corrine, Nicole and Steve J., for keeping the lab working. I could never get my experiments done without your hard work! Special thanks to Holly for training me to work at the TB!

I'd like to thank Luis and Graham, for your help with the (huge number of) bugs in my scripts, and teaching me all the bioinformatics stuff. Thank you, Oscar, it was good to have someone also working late and to chat with when I was stuck with my writing.

I'd like to thank the girls and all the wine and fun we had along the Torrens River. Thanks to the duck families and pigeons, too, although it seems that you guys don't like my carrot sticks. Thanks to all the PhD students at ACAD and the fun we had during the board game nights and the paintball battle!

Thanks to my family for being supportive. Thanks to my cat and dogs, for being

so cute, except when you forgot me after not seeing me for half a year. Thank you my friends, Cao Jie, Li Yuan, Mu Yao, Ding Zhen Hao, A Lang, Mia, and Ayla, for your friendship.

Contents

List of Figures

Chapter 1

Introduction

1.1 Ancient DNA research

DNA can be preserved within organic remains for thousands, or even hundreds of thousands of years (Hofreiter et al.; 2001). The ability to extract ancient DNA (aDNA) allowed the retrieval of genetic information from past animals, which provided insights into the interactions between animals and the environment as well as their potential genetic adaptations (Miller et al.; 2008; Shapiro and Hofreiter; 2014; Soubrier et al.; 2016).

Although aDNA is now routinely recovered from sub-fossil (*i.e.* not yet fully fossilised) specimens, aDNA research faces several challenges (Pääbo et al.; 2004). First, the opportunistic nature of sampling, whereby making experimental groups from the same place and time is often impossible—at least for non-human organisms that are not systematically buried in given places. Taphonomy represents another major challenge in aDNA research. After an organism dies, organic tissues immediately start to decompose, and DNA is initially degraded by endogenous and microbial nucleases (Dabney et al.; 2013). Subsequently, biochemical processes further break down DNA at a rate that is dependent on environmental conditions, such as temperature, humidity, and oxygen availability (Pääbo et al.; 2004; Hebsgaard et al.; 2005). However, taphonomic processes do not affect all samples uniformly, resulting in a high experimental failure rate, discrepancies between results even from the same individual, and a decrease in the number of usable samples. Even using approaches tailored for low-biomass samples (*e.g.,* (Gansauge and Meyer; 2013; Gansauge et al.; 2017)), the amount of retrieved aDNA is very low and usually only quantifiable after amplification. The surviving aDNA is highly fragmented, with an average length typically less than 100 bp. The aDNA damage also results in 1) lesions that inhibit *in vitro* DNA amplification, and 2) nucleotide substitutions (*i.e.* C-to-T and G-to-A substitutions) in the sequencing data (Figure 1.1) (Briggs et al.; 2007; Rohland et al.; 2015).

Moreover, the proportion of endogenous aDNA retrieved from a given sample is typically very low due to post-mortem contamination with exogenous DNA (mainly from microbial sources), as the sub-fossil material used in ancient genomic research is generally exposed to the environment for a long period (Hofreiter et al.; 2001). During experimental processing, laboratory staff, reagents, and consumables can also introduce additional contaminant and human DNA directly onto the sample or into the aDNA extracts (Llamas et al.; 2017). Therefore, it is critical to minimise and monitor contamination and authenticate aDNA sequences. Several measures can be taken to reduce such bias in paleogenetic research, including carrying out experiments in low-DNA facilities, using blank controls throughout the whole experimental process, authenticating aDNA sequences based on DNA damage patterns,
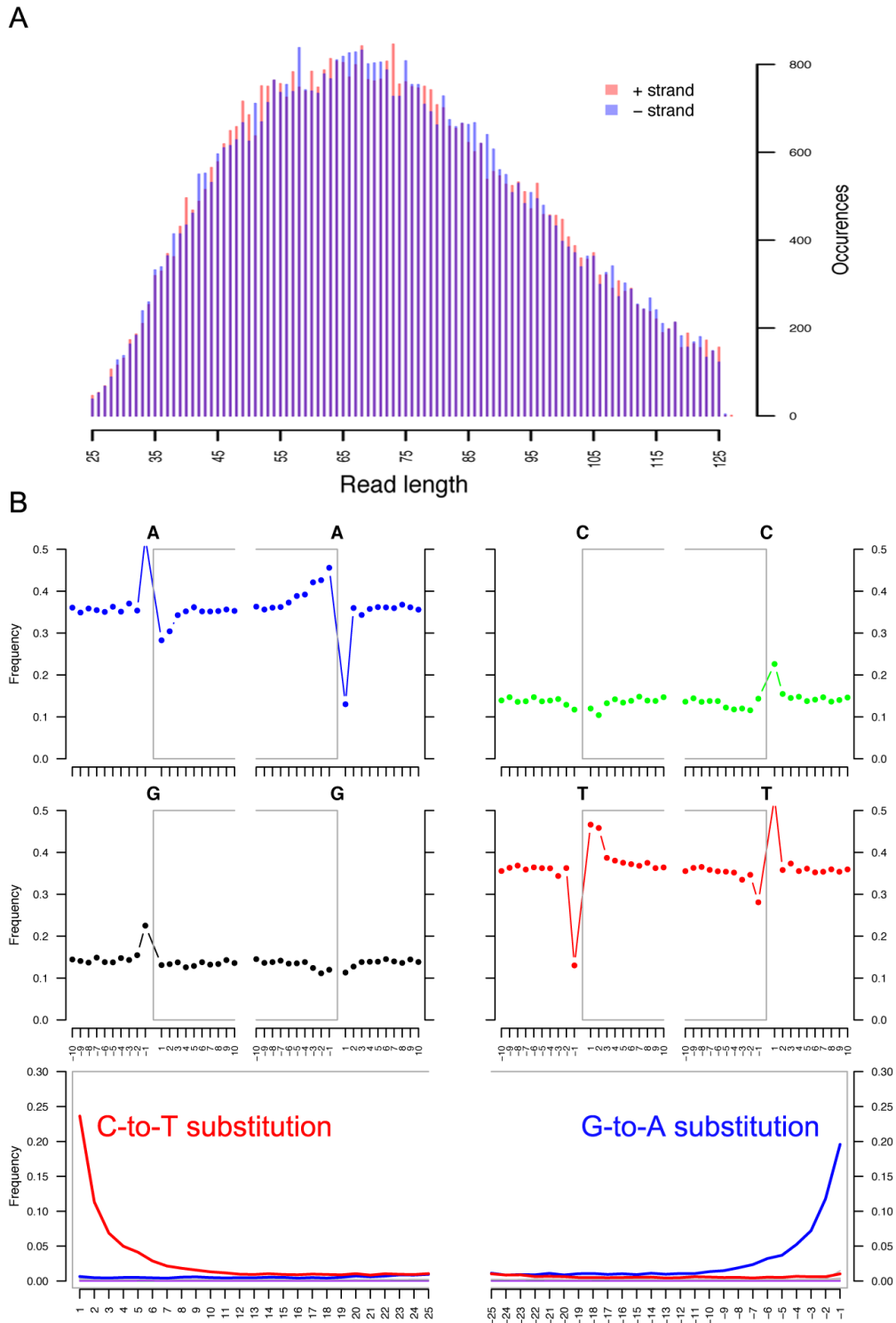
Figure 1.1: An example of aDNA damage. A. the fragment length distribution of aDNA, with an average length of ˜70 bp. B. The nucleotide substitutions in aDNA. C-to-T and G-to-A substitutions are increased at the end of the aDNA molecules.

and assessing the reproducibility of the findings (Llamas et al.; 2017).

Despite the challenges and difficulties of aDNA research, this field experienced an explosion in the past decade following the advent of high throughput sequencing (HTS) techniques (Marciniak and Perry; 2017). Taking human research as an example, only two archaic hominin genomes (and no ancient anatomically modern human, AMH) were recovered at low depth of coverage ($<1\times$) in 2007, while in 2017 over a thousand of ancient AMH and archaic hominin genomes had been reconstructed, and over 35% of the reconstructed genomes had a depth of coverage greater than $1\times$ (Marciniak and Perry; 2017). Advances in HTS techniques not only drastically increased the accessibility to ancient genetic information, but also paved the way for new avenues of research such as paleomicrobiomes and paleoepigenomes.

1.2   Paleomicrobiome and paleoepigenome studies in animals

Paleomicrobiome and paleoepigenome informations have recently been recovered from ancient animals (Table 1.1 and Table 1.2) (Adler et al.; 2013; Warinner et al.; 2014; Weyrich et al.; 2017; Cano et al.; 2014; Gokhman et al.; 2014, 2017; Hanghøj et al.; 2016; Pedersen et al.; 2014; Seguin-Orlando et al.; 2015).

Paleomicrobiome information (the taxonomic and functional profile of ancient bacterial communities) largely originated from human dental calculus (calcified dental plaque) and coprolites (ancient faeces deposits). Dental calculus can entrap and preserve microbial DNA, and it has been routinely used for the recovery of oral microbiome information in paleomicrobiology studies (Lieverse; 1999; Jin and Yip; 2002; Adler et al.; 2013; Weyrich et al.; 2017). As dental calculus is less prevalent in non-human animals, the application of this technique was restricted to ancient human research. Coprolites can potentially preserve diet, pathogen, and gut microbiome information of past animals (Wood et al.; 2008; Tito et al.; 2012; Boast et al.; 2018). However, it is very difficult to recover reliable gut microbiome information from coprolites because once exposed to the environment, animal faeces are very susceptible to contamination (Warinner, Speller, Collins and Lewis Jr; 2015; Eisenhofer et al.; 2017). Thus, a large proportion of DNA in the coprolites can come from environmental sources, and it is hard to distinguish environmental DNA from the gut microbial DNA. Also, because faeces contain organic material, the gut bacterial community, along with bacteria from the environment, can continue to change over time, which can bias the abundance of the endogenous bacterial community (Amir et al.; 2017).

Table 1.1 Paleomicrobiome studies on animals

Study	Animal (genus)	Sample	Method	Data	Age of samples (years BP)
Weyrich et al.; 2017	*Homo*	Dental calculus	Amplicon and shotgun sequencing	Oral microbiome	70 to ~36,000
Adler et al.; 2014	*Homo*	Dental calculus	Amplicon sequencing	Oral microbiome	400 to 7,550
Warinner et al.; 2014	*Homo*	Dental calculus	Amplicon and shotgun sequencing	Oral microbiome	~900 to ~1,000
Welker et al.; 2014	*Myotragus*	Coprolites	DNA barcoding	Two bacterial species	~5,000
Tito et al.; 2012	*Homo*	Coprolites	Amplicon sequencing	Gut microbiome	1,400 to ~8000
Cano et al.; 2014	*Homo*	Coprolites	Amplicon sequencing	Gut microbiome	~250 to 2000
Santiago-Rodriguez et al.; 2013, 2016	*Homo*	Coprolites	Amplicon sequencing	Gut microbial community information (the data have been suggested to be biased by contaminants (Eisenhofer et al;	~ 900

Table 1.2 Paleoepigentics studies on animals

Study	Animal (genus)	Sample	Method	Data	Age of samples (years BP)
Briggs et al.; 2009	*Homo*	Bone	DNA damage based	Genome wide methylation signals	~38,000
Llamas et al.; 2012	*Bison*	Bone	Bisulfite conversion	Methylation status of several locus	~26,000
Smith et al.; 2014	*Homo*	Bone	Bisulfite conversion	Methylation status of a retrotranposon	~200 to 4,500
Pedersen et al.; 2014	*Homo*	Hair	DNA damage-based	Methylome and nucleosome map	~4,000
Gokhman et al.; 2014	*Homo*	Bone	DNA damage-based	Methylome	~38,000 to > 50,000
Gokhman et al.; 2017	*Homo*	Bone	DNA damage-based	Methylome	7,000 to > 50,000
Seguin-Orlando, 2015	*Homo, Mammuthus, Ursus, and Equus*	Bone, hair, skin, skeletal muscle, heart, and liver	Methylated Binding Domains-based enrichment combined with DNA damage-based	Estimated regional methylation levels	~8,000 to 45,000
Hanghøj et al.; 2016	*Homo*	Bone, tooth, and hair	DNA damage-based	Methylome	1,400 to ~50,000
Keller et al.; 2017	*Homo*	Muscle tissue, stomach mucosa, and stomach	RT-qPCR	miRNA profiles	5,300
Smith et al.; 2019	*Canis*	Liver	RNA extraction adapted for ancient samples	Transcriptome (RNA)	12,000

Ancient bacterial communities can be reconstructed using amplicon sequencing and shotgun sequencing methods (Weyrich et al.; 2015; Warinner et al.; 2017). Amplicon sequencing is cost-effective, but it can bias the bacterial abundance of the ancient microbiome (Ziesemer et al.; 2015). This is mainly because aDNA are highly degraded and fragmented, and taxa with short amplicon lengths tends to be over-amplified (Ziesemer et al.; 2015). Shotgun sequencing obtains both the taxonomic and functional information and can be used for the assembly of ancient bacterial genomes, but it is relatively expensive (Adler et al.; 2013; Lloyd-Price et al.; 2017). Two major concerns in the paleomicrobiome data analysis are contamination and database bias (Warinner, Speller and Collins; 2015). Bioinformatically, contamination can be largely removed from animal paleogenomic data through filtering out sequences that failed to map to the reference genome. Additionally, damage patterns of the retained sequences—such as the increase of C-to-T and G-to-A substitutions at the end of mapped sequencing reads—can further help for authentication of the results (*e.g.*, Figure 1.1). However, paleomicrobiome data cannot be mapped to a single bacterial genome, thus removal of unmapped reads will not work. For the same reason, damage patterns of the whole microbiome dataset cannot be generated. An alternative is to check the damage pattern of individual bacterial genomes, but it can only authenticate a small proportion of the dataset. Paleomicrobiome data are also susceptible to database bias (Warinner, Speller and Collins; 2015). For

example, MALT, a metagenomic data analysis pipeline, tends to assign conserved sequences to well-studied organisms (Warinner, Speller and Collins; 2015; Herbig et al.; 2016). Additionally, a large proportion of paleomicrobiome data cannot be identified (Eisenhofer; 2018). This is probably partially because there is still a large number of bacteria that remain unknown, and potentially large phylogenetic distances between modern and ancient bacteria further reduce the power of identification (*e.g.*, using BLAST for the identification). This is an important reason why the assembly of ancient bacteria genomes is crucial, as the newly assembled genomes can be incorporated into mapping databases and can assist future identification of closely related ancient bacteria.

Methylation patterns in DNA sequences, an epigenomic feature, participate in the regulation of gene expression and many critical cellular processes (Jones; 2012). Paleomethylomes have been reconstructed from the damage patterns of ancient genomes (Gokhman et al.; 2014; Hanghøj et al.; 2016; Pedersen et al.; 2014). This method uses the differential damage patterns of cytosines and methylated cytosines to distinguish between the two. It performs well on ancient genomes with high coverage but damage patterns are typically estimated over sliding windows, thus the resolution of this approach is limited (Hanghøj et al.; 2016). Methylation patterns at single-base resolution can be recovered using bisulfite conversion (Llamas et al.; 2012; Smith et al.; 2014a), but this experimental method performs poorly on degraded samples. Therefore, the application of bisulfite was restricted to the profiling of the methylation of a few loci from well-preserved ancient samples.

Unlike genomic information that is similar across cells within an individual, different types of tissues have different methylation patterns (Ziller et al.; 2013; Lokk et al.; 2014). Methylome profiles also vary in regard to sex, age, and physiology of the individual (Ziller et al.; 2013; Lokk et al.; 2014). In modern methylome studies, such tissue- and individual-specificity can be accounted for by utilising replicates (Bock 2012). However, using replicates is not applicable to paleomethylome studies due to the opportunistic nature of sampling. Statistical methods can compensate this issue to some degree (Wu et al.; 2015; Teschendorff and Relton; 2018), but paleomethylome studies are still susceptible to bias caused by random variances. Better databasing of tissue-, sex-, age-, and disease-specific methylation may help to reduce such bias. Also, the function and phenotypic consequences of unique features identified from paleomethylomes can be only inferred and need to be validated using modern animal models.

RNA has been recovered from several ancient specimens (Fordyce et al.; 2013; Keller et al.; 2017; Smith et al.; 2017). The majority of ancient RNA (aRNA) was recovered from plant remains (Smith et al.; 2014b, 2017), but the recovery of aRNA from ancient animal tissues has been recently reported (Keller et al.; 2017;

Smith et al.; 2019). However, our current understanding of aRNA is very limited. Nucleotide substitutions similar to aDNA have been observed in aRNA, but it is not clear if it is sufficient for the authentication of the results (Smith et al.; 2019). Analysis of ancient microRNA (miRNA; a miRNA is a small non-coding RNA that regulates target transcripts), is very challenging, as the majority of aRNA have a length <50 nt (*e.g.*, (Smith et al.; 2019)), and the mappability of sequence with such short length is very limited. Therefore, it is very difficult to determine if the obtained aRNA is from ancient organisms or from the environment. Moreover, the abundance of miRNA estimated using aRNA data is highly biased due to degradation and *in vitro* amplification, and it might bias the inference of related gene expression levels.

1.3   Studying past animal adaptation from epigenetic and microbiomic perspectives

Darwinian natural selection affects traits that are heritable and variable. Mendelian inheritance and genetic variation addressed the theoretical and molecular bases for the natural selection theory, resulting in the proposal of the Modern Synthesis (Provine; 2001). However, even with the unprecedented ability to obtain genetic information from various organisms, it is still difficult to underpin the genetic basis of many adaptive traits (Ellegren and Sheldon; 2008; Mackay et al.; 2009). With the increasing awareness that non-genetic information can have lasting and profound influences on animal phenotypes (reviewed in Chapter 2), it is reasonable to hypothesise that mechanisms beyond genetics play a role in animal adaption (Danchin et al.; 2011; Laland et al.; 2014).

Adaptive evolutionary research faces several challenges (Merilä and Hendry; 2014). Primarily, it is very difficult, if not impossible, to accurately infer the past animal-environmental interactions using data from contemporary populations (Merilä and Hendry; 2014). Directly generating data from experimental studies also has limitations, as animal model analyses usually require long-term studies on a small number of populations or are limited to small organisms with short generation time (Charmantier et al.; 2008; Mackay et al.; 2009). One possible alternative is to use ancient DNA (aDNA) techniques to obtain desired information from animals over evolutionary timescales (Hofreiter et al.; 2001). Using advanced aDNA techniques, this thesis examined how epigenetic mechanisms and the microbiome of bovid taxa has changed and potentially provided a non-genetic way to adapt to changing environments over millennia.

1.4 Bovid taxa used for case studies in the thesis

Two bovidae taxa, the *Myotragus* and bison (extant American bison, *Bison bison*, and extinct steppe bison, *Bison priscus*), were used for case studies in this thesis. *Myotragus* is an extinct dwarf fossil caprine that lived in the Balearic Islands (Mediterranean Sea) from the Late Pleistocene to the Holocene (Figure 2) (Martínez, Suriñach et al. 1997). The ancestors of *Myotragus* are estimated to have arrived to the islands *circa* five million years ago (Ramis and Bover; 2001). Since then, *Myotragus* has evolved in this isolated environment until its human-caused extinction around 4,300 years ago (Ramis and Bover; 2001). The fossil records (skeletal fossils and coprolites) suggest many evolutionary features that were shaped by the insular environment: morphologically, *Myotragus* had a small body size (dwarfism) with strong limb bones; behaviourally, they seemed to be able to exploit all available plant food on the islands, including *Buxus balearica*, a plant that is toxic to ruminants (Alcover et al.; 1999; Köhler and Moyà-Solà; 2004; Ata and Andersh; 2008; Bover et al.; 2014). As the gut microbiome play important roles in food digestion and resilience to harmful dietary components (Huttenhower et al.; 2012; Kohl et al.; 2014), we hypothesised that the gut microbiome assisted *Myotragus* to adapt to its toxic diet.

The Late Pleistocene (110 to 11.65 thousand years ago—kya) was characterised by a series of climate fluctuations (Cooper et al.; 2015) and is associated to the mass extinction of many megafauna species. However, steppe bison were part of the few survivors who made it into the Holocene (last 11,000 years) and they are now identified as American bison (Cooper et al.; 2015). The morphological classification of the extinct steppe bison into tens of species based on their fossil record is poorly-supported by genetic evidence (Shapiro and Hofreiter; 2014; Soubrier et al.; 2016). Non-genetic mechanisms like epigenetics possibly played a role in the past bison morphological diversity. Therefore, we reconstructed a methylome history of bison and explored the possible role of methylome in their adaptation to climate changes. We also explored the possibility to obtain oral microbiome information from ancient bison teeth.

1.5 Thesis overview

The thesis consists of one review chapter (Chapter 2) and four research chapters (Chapter 3–6). Chapter 3 and 5 investigated how the microbiome and methylome assisted the adaptation of the two bovid taxa. Chapter 4 and 6 explored the possibility to extend the scopes of current paleomicrobiology and paleoepigenetics research using laboratory methods optimised for highly degraded samples.

Figure 1.2: Bovid taxa used for case studies in the thesis (left: the reconstructed *Myotragus* (Lalueza-Fox et al.; 2002); right: bison (Dykinga; 2018))

Chapter 2: More arrows in the ancient DNA quiver: use of paleoepigenomes and paleomicrobiomes to investigate animal adaptation to environment

This chapter reviewed the existing evidence indicating that non-genetic mechanisms can play a role in animal adaptation. Epigenetics and microbiome are dynamically shaped by the environment and can lead to phenotypic variation, thereby affecting animal fitness. A proportion of the epigenetics and microbiome alterations can be maintained over multiple generations. In this chapter, we underscored the feasibility of using aDNA techniques to reconstruct a history of epigenetics and microbiome of various animal species over evolutionary timescales, thus providing new insights into the animal adaption.

Chapter 3: Ancient DNA analysis of coprolites reveals the role of gut microbiota in mammal adaptation to environment

This chapter is a case study that investigated whether and how the gut microbiome assisted an extinct cave goat (*M. balearicus*) to adapt to a diet containing the toxic plant *Buxus balearica*. Robust ancient gut microbiome results were retrieved from coprolites, and taxonomical and functional analyses revealed a detoxification role of the *Myotragus* gut microbiome. These results suggest that *Myotragus* harboured a unique gut microbial community, which likely played a key role in their adaptation to a toxic diet.

Chapter 4: Recovery of oral microbiome signal from ancient bison teeth

This chapter explored the possibility to recover oral microbiome information by utilising cementum and a dental "plaque-like" structure from bison teeth. Although a large proportion of the obtained DNA originated from the environment, DNA of oral, mucosal, and ruminal microorganisms was successfully obtained from ancient and modern bison specimens. This chapter demonstrates that ancient teeth can serve as a promising proxy to recover oral microorganisms and host information

from ancient animals.

Chapter 5: Reconstruction of a bison methylome history over the past 50,000 years

This chapter optimised a DNA sequencing library method (the hairpin method) for the retrieval of high-quality methylome data from highly degraded DNA samples. Using the hairpin method, 11 ancient bison and 14 modern bison methylomes with single-base resolution were reconstructed. The robustness of this method was verified, and tissue-specificity was observed in ancient methylomes. Through reconstruction of the bison methylome history over 50,000 years, we were able to identify potential methylation hotspots responding to mammal-environment interactions. This chapter shows the strengths and the great potential of the hairpin method in revealing novel dynamics in mammal methylome history.

Chapter 6: Recovery of ancient RNA from a 30, 000-year-old bison bone

This chapter applied an optimised RNA extraction protocol to a 30,000-year-old bison bone and aRNA was successfully obtained. We evaluated potential contamination by co-extracted DNA and the necessity of an RNA molecule end repair step for sequencing library construction. This study highlights the feasibility of recovering aRNA from sub-fossil bones—which represent most of the animal sub-fossil record—and the necessity for a better understanding of aRNA properties.

Bibliography

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A., Alt, K. W. et al. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the neolithic and industrial revolutions, *Nature genetics* **45**(4): 450.

Alcover, J. A., Perez-Obiol, R., YLL, E.-I. and Bover, P. (1999). The diet of myotragus balearicus bate 1909 (artiodactyla: Caprinae), an extinct bovid from the balearic islands: evidence from coprolites, *Biological Journal of the Linnean Society* **66**(1): 57–74.

Amir, A., McDonald, D., Navas-Molina, J. A., Debelius, J., Morton, J. T., Hyde, E., Robbins-Pianka, A. and Knight, R. (2017). Correcting for microbial blooms in fecal samples during room-temperature shipping, *MSystems* **2**(2): e00199–16.

Ata, A. and Andersh, B. J. (2008). Buxus steroidal alkaloids: chemistry and biology, *The Alkaloids: Chemistry and Biology* **66**: 191–213.

Boast, A. P., Weyrich, L. S., Wood, J. R., Metcalf, J. L., Knight, R. and Cooper, A. (2018). Coprolites reveal ecological interactions lost with the extinction of new zealand birds, *Proceedings of the National Academy of Sciences* **115**(7): 1546–1551.

Bover, P., Valenzuela, A., Guerra, C., Rofes, J., Alcover, J. A., Ginés, J., Fornós, J. J., Cuenca-Bescós, G. and Merino, A. (2014). The cova des pas de vallgornera (llucmajor, mallorca): a singular deposit bearing an exceptional well preserved early pleistocene vertebrate fauna, *International Journal of Speleology* **43**(2): 6.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. et al. (2007). Patterns of damage in genomic dna sequences from a neandertal, *Proceedings of the National Academy of Sciences* **104**(37): 14616–14621.

Cano, R. J., Rivera-Perez, J., Toranzos, G. A., Santiago-Rodriguez, T. M., Narganes-Storde, Y. M., Chanlatte-Baik, L., García-Roldán, E., Bunkley-Williams, L. and Massey, S. E. (2014). Paleomicrobiology: revealing fecal microbiomes of ancient indigenous cultures, *PloS one* **9**(9): e106833.

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. and Sheldon, B. C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population, *science* **320**(5877): 800–803.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. and Bradshaw, C. J. (2015). Abrupt warming events drove late pleistocene holarctic megafaunal turnover, *Science* **349**(6248): 602–606.

Dabney, J., Meyer, M. and Pääbo, S. (2013). Ancient dna damage, *Cold Spring Harbor perspectives in biology* **5**(7): a012567.

Danchin, É., Charmantier, A., Champagne, F. A., Mesoudi, A., Pujol, B. and Blanchet, S. (2011). Beyond dna: integrating inclusive inheritance into an extended theory of evolution, *Nature Reviews Genetics* **12**(7): 475.

Dykinga, J. (2018). *Agricultural Research Service, United States Department of Agriculture; https://www.ars.usda.gov/oc/images/photos/k5680-1/* .

Eisenhofer, R. A. (2018). *New and Refined Tools and Guidelines to Expand the Scope and Improve the Reproducibility of Palaeomicrobiological Research*, PhD thesis, School of Biological Sciences, The University of Adelaide.

Eisenhofer, R., Cooper, A. and Weyrich, L. S. (2017). Reply to santiago-rodriguez et al.: Proper authentication of ancient dna is essential, *FEMS microbiology ecology* **93**(5).

Ellegren, H. and Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations, *Nature* **452**(7184): 169.

Fordyce, S. L., Avila-Arcos, M. C., Rasmussen, M., Cappellini, E., Romero-Navarro, J. A., Wales, N., Alquezar-Planas, D. E., Penfield, S., Brown, T. A., Vielle-Calzada, J.-P. et al. (2013). Deep sequencing of rna from ancient maize kernels, *PLoS One* **8**(1): e50961.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A. and Meyer, M. (2017). Single-stranded dna library preparation from highly degraded dna using t4 dna ligase, *Nucleic acids research* **45**(10): e79–e79.

Gansauge, M.-T. and Meyer, M. (2013). Single-stranded dna library preparation for the sequencing of ancient or damaged dna, *Nature protocols* **8**(4): 737.

Gokhman, D., Lavi, E., Prüfer, K., Fraga, M. F., Riancho, J. A., Kelso, J., Pääbo, S., Meshorer, E. and Carmel, L. (2014). Reconstructing the dna methylation maps of the neandertal and the denisovan, *Science* **344**(6183): 523–527.

Gokhman, D., Tamir, L. A., Housman, G., Rafinia, M. N., Colón, M. N., Gu, H., Ferrando, M., Gelabert, P., Lipende, I., Quillen, E. E. et al. (2017). Recent regulatory changes shaped the human facial and vocal anatomy, *bioRxiv* p. 106955.

Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J. S., Willerslev, E. and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epipaleomix, *Molecular biology and evolution* **33**(12): 3284–3298.

Hebsgaard, M. B., Phillips, M. J. and Willerslev, E. (2005). Geologically ancient dna: fact or artefact?, *Trends in microbiology* **13**(5): 212–220.

Herbig, A., Maixner, F., Bos, K. I., Zink, A., Krause, J. and Huson, D. H. (2016). Malt: Fast alignment and analysis of metagenomic dna sequence data applied to the tyrolean iceman, *BioRxiv* p. 050559.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. and Pääbo, S. (2001). ancient dna, *Nature Reviews Genetics* **2**(5): 353.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S. et al. (2012). Structure, function and diversity of the healthy human microbiome, *nature* **486**(7402): 207.

Jin, Y. and Yip, H.-K. (2002). Supragingival calculus: formation and control, *Critical reviews in oral biology & medicine* **13**(5): 426–441.

Jones, P. A. (2012). Functions of dna methylation: islands, start sites, gene bodies and beyond, *Nature Reviews Genetics* **13**(7): 484.

Keller, A., Kreis, S., Leidinger, P., Maixner, F., Ludwig, N., Backes, C., Galata, V., Guerriero, G., Fehlmann, T., Franke, A. et al. (2017). mirnas in ancient tissue specimens of the tyrolean iceman, *Molecular biology and evolution* **34**(4): 793–801.

Kohl, K. D., Weiss, R. B., Cox, J., Dale, C. and Denise Dearing, M. (2014). Gut microbes of mammalian herbivores facilitate intake of plant toxins, *Ecology letters* **17**(10): 1238–1246.

Köhler, M. and Moyà-Solà, S. (2004). Reduction of brain and sense organs in the fossil insular bovid myotragus, *Brain, Behavior and Evolution* **63**(3): 125–140.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E. et al. (2014). Does evolutionary theory need a rethink?, *Nature News* **514**(7521): 161.

Lalueza-Fox, C., Shapiro, B., Bover, P., Alcover, J. A. and Bertranpetit, J. (2002). Molecular phylogeny and evolution of the extinct bovid myotragus balearicus, *Molecular phylogenetics and evolution* **25**(3): 501–510.

Lieverse, A. R. (1999). Diet and the aetiology of dental calculus, *International Journal of osteoarchaeology* **9**(4): 219–232.

Llamas, B., Holland, M. L., Chen, K., Cropley, J. E., Cooper, A. and Suter, C. M. (2012). High-resolution analysis of cytosine methylation in ancient dna, *PloS one* **7**(1): e30226.

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A. and Haak, W. (2017). From the field to the laboratory: Controlling dna contamination in human ancient dna research in the high-throughput sequencing era, *STAR: Science & Technology of Archaeological Research* **3**(1): 1–14.

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G. et al. (2017). Strains, functions and dynamics in the expanded human microbiome project, *Nature* **550**(7674): 61.

Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson, T. K., Vilo, J., Salumets, A. et al. (2014). Dna methylome profiling of human tissues identifies global and tissue-specific methylation patterns, *Genome biology* **15**(4): 3248.

Mackay, T. F., Stone, E. A. and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects, *Nature Reviews Genetics* **10**(8): 565.

Marciniak, S. and Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation, *Nature Reviews Genetics* **18**(11): 659.

Merilä, J. and Hendry, A. P. (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence, *Evolutionary applications* **7**(1): 1–14.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A. et al. (2008). Sequencing the nuclear genome of the extinct woolly mammoth, *Nature* **456**(7220): 387.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. and Hofreiter, M. (2004). Genetic analyses from ancient dna, *Annu. Rev. Genet.* **38**: 645–679.

Pedersen, J. S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S. et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome, *Genome research* **24**(3): 454–466.

Provine, W. B. (2001). *The origins of theoretical population genetics: With a new afterword*, University of Chicago Press.

Ramis, D. and Bover, P. (2001). A review of the evidence for domestication of myotragus balearicus bate 1909 (artiodactyla, caprinae) in the balearic islands, *Journal of Archaeological Science* **28**(3): 265–282.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. and Reich, D. (2015). Partial uracil–dna–glycosylase treatment for screening of ancient dna, *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**(1660): 20130624.

Seguin-Orlando, A., Gamba, C., Der Sarkissian, C., Ermini, L., Louvel, G., Boulygina, E., Sokolov, A., Nedoluzhko, A., Lorenzen, E. D., Lopez, P. et al. (2015). Pros and cons of methylation-based enrichment methods for ancient dna, *Scientific reports* **5**: 11826.

Shapiro, á. and Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient dna, *Science* **343**(6169): 1236573.

Smith, O., Clapham, A. J., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2014a). Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process, *Scientific reports* **4**: 5559.

Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2014b). A complete ancient rna genome: identification, reconstruction and evolutionary history of archaeological barley stripe mosaic virus, *Scientific reports* **4**: 4003.

Smith, O., Dunshea, G., Sinding, M.-H. S., Fedorov, S., Germonpre, M., Gilbert, M. et al. (2019). Ancient rna from late pleistocene permafrost and historical canids shows tissue-specific transcriptome survival, *BioRxiv* p. 546820.

Smith, O., Palmer, S. A., Clapham, A. J., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2017). Small rna activity in archeological barley shows novel germination inhibition in response to environment, *Molecular biology and evolution* **34**(10): 2555–2562.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y., Kosintsev, P., Lee, M. S., Baryshnikov, G. et al. (2016). Early cave art and ancient dna record the origin of european bison, *Nature communications* **7**: 13158.

Teschendorff, A. E. and Relton, C. L. (2018). Statistical and integrative system-level analysis of dna methylation data, *Nature Reviews Genetics* **19**(3): 129.

Tito, R. Y., Knights, D., Metcalf, J., Obregon-Tito, A. J., Cleeland, L., Najar, F., Roe, B., Reinhard, K., Sobolik, K., Belknap, S. et al. (2012). Insights from characterizing extinct human gut microbiomes, *PloS one* **7**(12): e51146.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., Orlando, L. and Krause, J. (2017). A robust framework for microbial archaeology, *Annual review of genomics and human genetics* **18**: 321–356.

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S. et al. (2014). Pathogens and host immunity in the ancient human oral cavity, *Nature genetics* **46**(4): 336.

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): 20130376.

Warinner, C., Speller, C., Collins, M. J. and Lewis Jr, C. M. (2015). Ancient human microbiomes, *Journal of human evolution* **79**: 125–136.

Weyrich, L. S., Dobney, K. and Cooper, A. (2015). Ancient dna analysis of dental calculus, *Journal of Human Evolution* **79**: 119–124.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V. et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus, *Nature* **544**(7650): 357.

Wood, J. R., Rawlence, N. J., Rogers, G. M., Austin, J. J., Worthy, T. H. and Cooper, A. (2008). Coprolite deposits reveal the diet and ecology of the extinct new zealand megaherbivore moa (aves, dinornithiformes), *Quaternary Science Reviews* **27**(27-28): 2593–2602.

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P. and Conneely, K. N. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates, *Nucleic acids research* **43**(21): e141–e141.

Ziesemer, K. A., Mann, A. E., Sankaranarayanan, K., Schroeder, H., Ozga, A. T., Brandt, B. W., Zaura, E., Waters-Rist, A., Hoogland, M., Salazar-Garcia, D. C. et al. (2015). Intrinsic challenges in ancient microbiome reconstruction using 16s rrna gene amplification, *Scientific Reports* **5**: 16498.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E. et al. (2013). Charting a dynamic dna methylation landscape of the human genome, *Nature* **500**(7463): 477.

Statement of Authorship

Title of Paper	More arrows in the ancient DNA quiver: use of paleoepigenomes and paleomicrobiomes to investigate animal adaptation to environment
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	In preparation for submission to Molecular Biology and Evolution

Principal Author

Name of Principal Author (Candidate)	Yichen Liu		
Contribution to the Paper	Wrote the manuscript		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	04/03/19

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    i.   the candidate's stated contribution to the publication is accurate (as detailed above);

    ii.   permission is granted for the candidate in include the publication in the thesis; and

    iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Reviewed manuscript		
Signature		Date	

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	Reviewed manuscript		
Signature		Date	5/3/19

Please cut and paste additional co-author panels here as required.

Chapter 2

More arrows in the ancient DNA quiver: use of paleoepigenomes and paleomicrobiomes to investigate animal adaptation to environment

Abstract

Whether and how epigenome mechanisms and the microbiome play a role in mammalian adaptation raised considerable attention and controversy, mainly because they have the potential to add new insights into the Modern Synthesis. Recent attempts to reconcile neo-Darwinism and neo-Lamarckism in a unified theory of molecular evolution give epigenetic mechanisms and microbiome a prominent role. However, supporting empirical data is still largely missing. Because experimental studies using modern animals can hardly be done over evolutionary timescales, we propose that advances in ancient DNA techniques provide a valid alternative. In this piece, we evaluate: (1) the possible roles of epigenomes and microbiomes in animal adaptation; (2) advances in the retrieval of paleoepigenome and paleomicrobiome data using ancient DNA techniques; and (3) the plasticity of either and interactions between the epigenome and the microbiome, while emphasising that it is essential to take both into account, as well as the underlying genetic factors that may confound the findings. We propose that advanced ancient DNA techniques should be applied to a wide range of past animals so novel dynamics in animal evolution and adaption can be revealed.

Key words: Ancient DNA, paleomicrobiome, paleoepigenome, animal evolution

2.1 Introduction

How animals adapt to a changing environment is a fundamental question in evolutionary biology. Efforts have focused on the characterisation of genetic processes, such as mutation, drift, and selection, that underlie animal adaptation (Gillespie; 2004). However, increasing evidence suggests that epigenetic modifications play a significant role in shaping animal phenotypes and response to environmental stimuli (Jirtle and Skinner; 2007). In addition, the recently proposed concept of "hologenome" — which considers the genome of an organism as an aggregation of the genomes of both the host and its resident microorganisms (microbiota) — challenges the traditional genetic paradigms that focus on the sole host genome (Zilber-Rosenberg and Rosenberg; 2008). Hence, both epigenetic modifications and microbiome alterations need to be considered when discussing the evolutionary history of a particular species.

Most evolutionary studies largely rely on the comparison and interpretation of modern DNA sequences, which provide only indirect evidence of past events and usually require temporal calibration using known fossil or biogeographical evidence (Telford et al.; 2015). Alternatively, genetic information can be directly obtained from archaeological and paleontological remains using ancient DNA (aDNA) techniques (Rohland and Hofreiter; 2007). In this case, dating methods can be applied to the sample, providing a reliable timestamp for the genetic data (Lorenzen et al.; 2011; Cooper et al.; 2015). Furthermore, methylomes (the cytosine methylation profile of a genome) and microbiomes (the totality of microbial genomes within a niche) have been recently recovered from ancient human remains (Adler et al.; 2013; Gokhman et al.; 2014; Pedersen et al.; 2014; Hanghøj et al.; 2016), which paves the way for the study of past interactions between animals and environments using epigenetics and microbiome.

This review aims to discuss the possible roles of the epigenome and microbiome in animal adaptation to rapidly changing environments. We further propose that the implementation of aDNA techniques to retrieve paleoepigenomes and paleomicrobiomes from a wide range of ancient animals has a great potential to track novel dynamic processes underlying animal adaptation over evolutionary timescales.

2.2 Challenges and opportunities of using aDNA in studying animal evolution

Ancient DNA is DNA retrieved from sub-fossil biological specimens (*i.e.*, not yet fossilised). Ancient DNA sequencing is a powerful way to recover ancient genomic information. However, aDNA is subject to post-mortem decay and contamination,

which can lead to biases and misinterpretation of paleogenetic data (Cooper and Poinar; 2000; Hofreiter et al.; 2001). Even under favourable circumstances (*e.g.*, permafrost), DNA continually degrades with time (Hofreiter et al.; 2001; Dabney et al.; 2013). Consequently, the proportion of endogenous aDNA retrieved from a given sample is typically very low; the DNA molecules are also highly fragmented, with an average fragment length typically less than 100bp (Sawyer et al.; 2012). DNA damage is also characterised by miscoding lesions (Hofreiter et al.; 2001). In particular, cytosines (especially those in single-stranded overhangs that form the sticky ends of aDNA fragments) are susceptible to conversion to uracil via hydrolytic deamination (Briggs et al.; 2009). After experimental amplification of DNA, cytosine deamination leads to cytosine-to-thymine (C-to-T) substitutions on the damaged DNA strand and guanine-to-adenine (G-to-A) substitutions on the complementary strand. Although such damage causes sequence errors in the final DNA sequence, the increase of C-to-T and G-to-A substitutions at fragment termini now serve as the gold standard to authenticate aDNA (Llamas, Valverde, Fehren-Schmitz, Weyrich, Cooper and Haak; 2017).

Despite the technical challenges of aDNA research, the advent of high–throughput sequencing (HTS) techniques enabled genome-wide analyses of ancient specimens, providing critical insights into the evolutionary history of several species (Soubrier et al.; 2016; Miller et al.; 2008; Shapiro and Hofreiter; 2014). However, many questions regarding mammals evolution remain unsolved. For example, one of the most intriguing areas of research is the mass megafauna extinctions that took place during the last glacial period (Late Pleistocene; 110 to 11.65 thousand years ago-—kya) (Cooper et al.; 2015). During this period, many megafaunal species went extinct, including the iconic woolly rhinoceros, mammoth, short-faced bear, short-faced kangaroo, and ground sloth. Whether humans, climate change, or both caused the mass extinctions remains highly controversial (Lorenzen et al.; 2011; Cooper et al.; 2015; Sandom et al.; 2014). Such controversy partially stems from missing data and would be drastically improved with additional information, *e.g.*, data regarding the behaviour, diet, and physiology of ancient animals. Paleoepigenome and paleomicrobiome analyses offer alternative access to such information (Gokhman, Malul and Carmel; 2017; Warinner, Speller, Collins and Lewis Jr; 2015). For instance, changes in the oral microbiota might indicate a change in diet, while methylation level in genomic regulatory regions can relate to specific phenotypes; alternatively, a diseased state might be linked to pathogens identified within the microbiome or abnormal methylation profiles (De Filippo et al.; 2010; Jones; 2012; Cui et al.; 2013).

More importantly, epigenomes and microbiomes respond to environmental cues and thus have the potential to capture fine-scale dynamics between animals and paleoenvironment, which might be obscured in genomic data (Figure 1). For ex-

ample, extinct steppe bison experienced and survived the last ice age. They have been morphologically classified into over 50 species based on the past fossil record (McDonald; 1981), while genetic data suggest that steppe bison are a single morphologically plastic species (Shapiro et al.; 2004). Animal morphology can be affected by epigenetic modifications without changing the genomic sequences (Cropley et al.; 2012); it is thus possible that the past bison morphological diversity is driven by epigenetic changes triggered by environmental factors. *Myotragus balearicus*, an extinct insular cave goat endemic to the Gymnesic Islands in the Mediterranean sea, appears to have adapted to feed on a plant (*Buxus balearica*) that is toxic to ruminants (Welker et al.; 2014). Many mammalian herbivores employ their gut microbes to facilitate the degradation of the harmful components in the diet (*e.g.*, the desert woodrat (Kohl et al.; 2014), koala (Shiffman et al.; 2017), and the Japanese large wood mouse (Sasaki et al.; 2005). The gut microbiome of this goat might play a similar role by enabling tolerance to the toxic plant.

In these examples, aside from offering additional information about phenotypic alterations (*e.g.*, modified body size and the ability to gain nutrition from various diets) and environmental cues (*e.g.*, ice age and toxic vegetation), the epigenome and microbiome might serve as a mechanism that facilitated animal adaptation to the environment. However, this posit has not been fully explored. To investigate the evolutionary role of epigenetic modifications and microbiome variations, three major questions need to be addressed. Firstly, how do epigenome and microbiome respond to environmental stimuli, and what are the phenotypic consequences? Secondly, can those epigenetic responses and microbiome changes be maintained over multiple generations and thus influence animal adaptation in the long term? Finally, how can research verify this hypothesis on an evolutionary timescale?

2.3 DNA methylation patterns as a proxy to infer animal-environment interactions

Epigenetics refers to mechanisms that alter the expression of genes without modifying the underlying genetic sequence. This include DNA methylation, histone modifications, nucleosome positioning, and non-coding RNAs (Holliday; 2006). In this review, the discussion of epigenetic modification will focus on DNA methylation, as it is likely the most accessible epigenetic signal that can be recovered from aDNA (Llamas et al.; 2012; Smith et al.; 2015).

In mammalian genomes, DNA methylation is found almost exclusively in the context of CpG dinucleotides and typically occurs at the fifth carbon position of cytosines (Jones and Takai; 2001). Taking humans as an example, the DNA methyla-
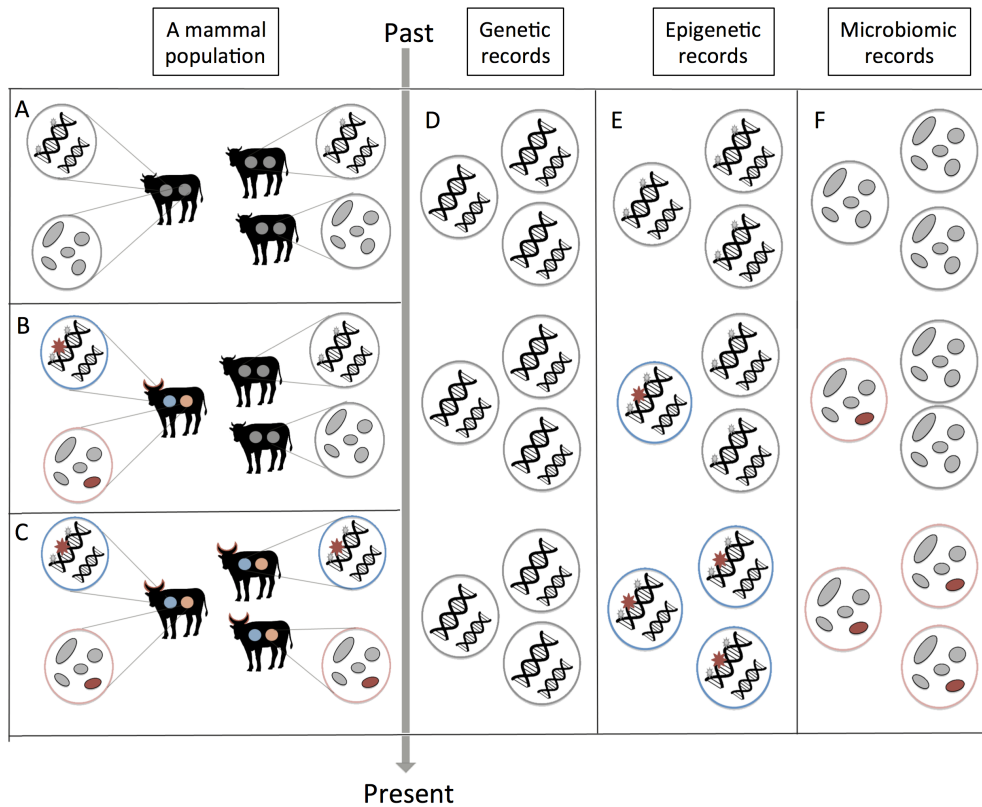
Figure 2.1: Schematic figure of how environmentally-induced epigenetic and microbiome changes accumulate in mammal populations. A-C: Environmentally-induced epigenetic and microbiome alterations steadily increase within a mammal population. D-F: Epigenetic and microbiome alterations are preserved within epigenetic and microbiome records, in the absence of genetic variation.

tion levels are relatively stable across differentiated cells, with 4-8% of total cytosines and 70-80% of CpG islands (loci rich in CpGs) methylated (Ziller et al.; 2013). However, locus-specific methylation can vary across differentiated cells and throughout the development of an organism (Smith and Meissner; 2013). DNA methylation in promoter regions typically blocks the initiation of transcription, while methylation occurring within the gene body can stimulate transcript elongation (Jones; 2012). Due to their regulatory role, DNA methylation patterns are associated with a wide range of biological traits (Jones; 2012).

2.3.1 Environmentally-induced epigenetic modifications

Animals are constantly exposed to the surrounding environment, which encompasses a wide range of beneficial or adverse factors that can exert physiological and psychological changes and stimulate a series of adaptive responses (Mellor; 2015; Hay et al.; 2016; Kent et al.; 2014; Koolhaas et al.; 1999). DNA methylation can respond to various environmental cues, including early life nutrition, temperature, and many other stressors (*e.g.*, chemical compounds, hypoxia, and mental stress) (Shen et al.; 2002; Cao-Lei et al.; 2014; Bollati and Baccarelli; 2010).

Early life nutrition is critical to foetal epigenetic programming and could give rise to persistent and systematic epigenetic alterations (Cao-Lei et al.; 2014; Heijmans et al.; 2008; Tobi et al.; 2009). One possible explanation is that maternal nutrients are associated with the levels of methyl donors available as a biochemical substrate for DNA methylation (Anderson et al.; 2012; Niculescu and Zeisel; 2002). DNA methylation is enzymatically catalysed by DNA methyltransferases, which transfer methyl groups from S-adenosylmethionine (SAM) to the fifth carbon position of cytosines (Fuso et al.; 2005; Chiang et al.; 1996). The synthesis of SAM requires the presence of methyl donors, such as folate, B6, B12, and some other dietary B vitamins (Fuso et al.; 2005). In animals, the availability of methyl donors in the maternal diet can affect foetal epigenetic states, and consequently, affect the offsrping's phenotype (Cooney et al.; 2002; Waterland et al.; 2006). For example, methyl donor supplementation of female mice before and during pregnancy increases the DNA methylation at a metastable (*i.e.* the epigenetic state is mitotically inherited once established) epiallele, axin fused ($Axin^{Fu}$), which results in a decreased incidence of tail abnormality in their offspring (Waterland et al.; 2006). Many epigenetically-regulated genes play important roles in embryogenesis and development, and altered DNA methylation patterns caused by maternal nutrition can exert a life-long influence in mammals (Okano et al.; 1999; Li et al.; 1993).

Temperature fluctuations can pose challenges to animal adaptation to the environment and have the potential to trigger alterations of DNA methylation in animals

(Cooper et al.; 2015; Cahill et al.; 2013; Marsh and Pasqualone; 2014; Parrott et al.; 2014; Navarro-Martín et al.; 2011). This temperature-related methylation has been best characterised in temperature-dependent sex determination in poikilothermic animals (Parrott et al.; 2014; Navarro-Martín et al.; 2011). Temperature-related DNA methylation and demethylation is less understood in mammals, but evidence suggests that temperature can prompt alterations in the mammalian methylome as well. For example, exposure to altered ambient temperature during adulthood is associated with the change of DNA methylation of multiple genes in blood cells (Bind et al.; 2014, 2016). However, it is unclear if such alterations can happen to germ cells and be transmitted to offspring.

In nature, the alteration of nutrition availability and temperature is likely coupled with physiological and psychological stress, and prenatal exposure to such stress can trigger the alteration of DNA methylation of genes with crucial functions. Such environmentally-induced methylation alterations have been observed in humans who experienced natural disasters (*e.g.*, Hunger Winter in 1944-–45 and 1998 Quebec ice storm) (Cao-Lei et al.; 2014; Veenendaal et al.; 2013). It is highly possible that epigenetic changes also occur in animals when the environment changes drastically (*e.g.*, the Late Pleistocene, which was characterised by a series of dramatic cooling and warming events that altered the extent of ice sheets across the globe (Cooper et al.; 2015)). However, these environmentally-induced epigenetic marks must be passed down over multiple generations to ultimately be a substrate for natural selection.

2.3.2 Trans-generational effects of epigenetic modifications on animal adaptation

The plasticity and regulatory role of epigenetic modifications enable short-term exposure to environmental cues to be translated into phenotypic traits, and such epigenetic alterations (epimutations) might be passed down to the next generation (Anway and Skinner; 2006). Such potential makes epimutation-mediated natural selection theoretically possible: the environment triggers epimutations, which lead to phenotypic alterations that can be generationally transmitted and subjected to natural selection. This process is best modelled in isogenic (*i.e.* all individuals are genetically identical) viable yellow agouti (A^{vy}) mice (Cropley et al.; 2012). The A^{vy} allele is epigenetically regulated in mice and its expression impacts coat colour. In this model, the environmentally-induced epimutation was simulated by manipulating methyl donor supplementation in the mice diet, and natural selection was mimicked using selective breeding. Interestingly, the prevalence of epimutation-associated phenotypes was steadily increased in a population for five generations (Cropley et al.; 2012). As these mice are otherwise genetically identical, the change

of coat colour is only driven by an epigenetic response to an external factor and selective forces.

Heritable epigenetic modifications have been observed in pigs, rodents, and humans (Cropley et al.; 2012; Wang et al.; 2014; Weyrich et al.; 2018; Skvortsova et al.; 2018), where many genes (*e.g.*, *IGF2R*, *Snrpn*, *Peg3*, *mest*, and *H19*) are epigenetically imprinted in gametes——*i.e.* genes are expressed in a monoallelic and parent-of-origin manner (Dean et al.; 1998). In the case of imprinting, and contrary to the A^{vy} allele-related phenotype that needs to be maintained by dietary supplements (Cropley et al.; 2012), epialleles can persistently affect animal traits throughout their life even in the absence of obvious external stimuli. Environmental factors (including maternal nutrition) can influence imprinting during foetal development (Kappil et al.; 2015). Such adaptation would be invisible in the genetic record, as it does not entail genetic change. Nevertheless, the epigenetic modifications that occur in gametes or during early developmental stages can be mitotically passed down to different types of cells, and thus can be preserved in sub-fossil records (*e.g.*, bones, teeth, and hair).

2.3.3 Methods and progress in ancient epigenetic research

Several methods are available to retrieve methylomes from modern samples, including bisulfite-based approaches and antibody-based enrichment (Harris et al.; 2010). However, the characteristic fragmentation and damage of aDNA molecules pose serious limitations to these methods when analysing ancient samples. In aDNA research, bisulfite conversion followed by targeted amplification is at best limited to a few methylated loci even using well-preserved samples, while antibody-based enrichment is biased towards large fragments and CpG-rich regions (Llamas et al.; 2012; Smith et al.; 2015; Seguin-Orlando et al.; 2015). The inefficacy and bias of these methods has hindered the retrieval of methylome-wide data from ancient samples. However, the extensive damage that occurs in aDNA offers a proxy to evaluate cytosine methylation levels from ancient samples (Gokhman et al.; 2014; Pedersen et al.; 2014). The deamination of cytosines and methylated cytosines is the most frequently observed damage in aDNA (Hofreiter et al.; 2001). This process converts cytosines to uracils (C-to-U) and methylated cytosines to thymines (5mC-to-T). After enzymatically removing the uracils using uracil—DNA—glycosylase (UDG), the remaining C-to-T substitutions in the sequencing data reflect the methylation level in a given region (Briggs et al.; 2009). This method has helped to reconstruct some ancient and archaic human methylomes (Gokhman et al.; 2014; Pedersen et al.; 2014; Hanghøj et al.; 2016).

However, the resolution of this method is coarse and limited to a regional charac-

terisation of methylation. It seems experimental bisulfite conversion is still the gold standard for obtaining base-resolution methylomes (Leontiou et al.; 2015). However, the biggest obstacle to apply bisulfite sequencing to ancient samples is the combination of short fragment length and reduced sequence complexity. Bisulfite converts cytosines to uracils and leaves methylated cytosines unchanged, making it possible to differentiate cytosines from methylated cytosines (Leontiou et al.; 2015). Consequently, in most of the genome (which is typically hypomethylated), cytosines are not methylated and will be displayed as thymine, which means that the original ATGC-coded DNA becomes an ATG-coded sequence. In aDNA, this means small fragments can no longer be mapped to the reference genome unambiguously. To illustrate this, we estimate that accurate mapping of a bison genome requires at least 25-bp-long DNA fragments, while a minimum of 35bp is necessary after bisulfite conversion of cytosines (Figure 2). Given the fact that most aDNA fragments are very small, 10bp can make a significant difference in terms of the amount of data that can be obtained from an individual sample.
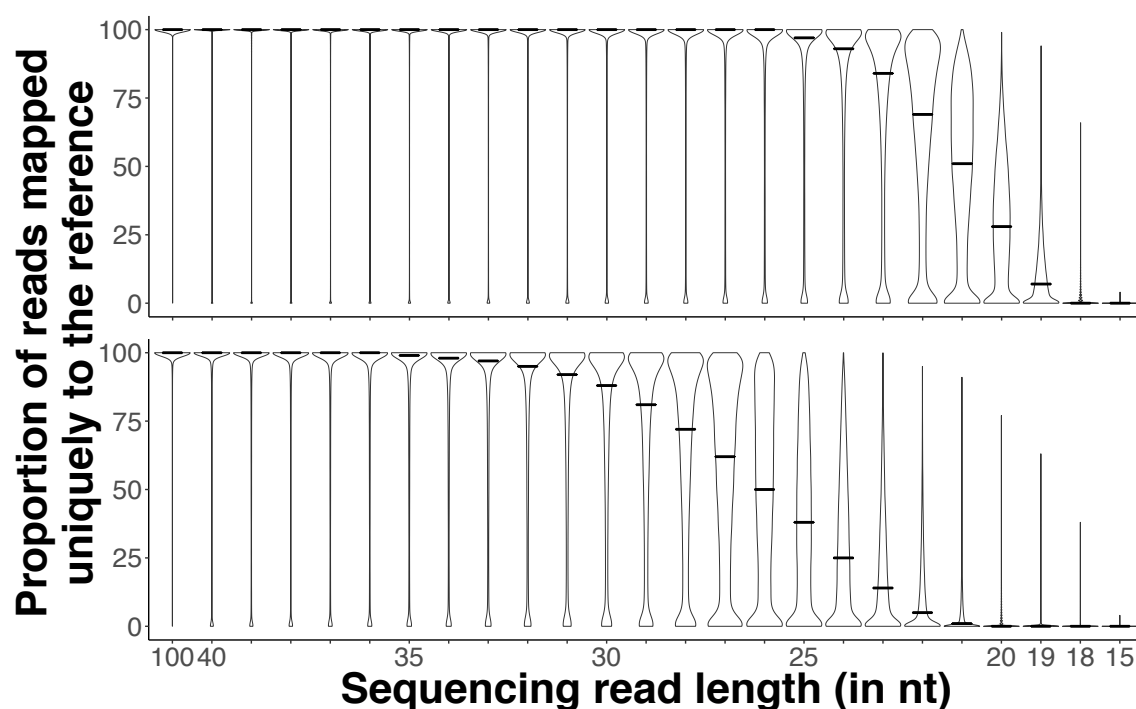
Figure 2.2: The proportion of reads mapped uniquely to the reference genome. Top: the proportion of uniquely mapped ATCG-coded reads drops when the sequencing read length less than 25bp. Bottom: the proportion of uniquely mapped ATG-coded reads drops when the sequencing read length less than 35bp

This issue has been recently addressed by a bisulfite sequencing method optimised for aDNA (Llamas, Heiniger, Gower, Liu and Cooper; 2017) based on (Laird et al.; 2004; Zhao et al.; 2014). Specifically, a hairpin adaptor is ligated to one end of the DNA molecules before bisulfite conversion. The bisulfite treatment de-

natures the hairpin molecules but the plus and minus strands remain attached via the hairpin adaptor. Both strands are thus sequenced together using paired-end sequencing and, after folding the sequencing reads bioinformatically, it is possible to reconstruct the original DNA sequence and identify methylation. Because the original ATGC-coded DNA can be recovered, the method allows accurate mapping of short fragments and detection of methylation at a single-base resolution. Thus, this method can be applied to retrieve highly resolved ancient methylomes.

Ancient epigenetics research is still in its infancy and only a small number of ancient methylomes have been reconstructed. However, this small dataset revealed intriguing findings. Altered DNA methylation profiles were detected from hominids such as Neanderthals and Denisovans (Gokhman et al.; 2014; Gokhman, Tamir, Housman, Rafinia, Colón, Gu, Ferrando, Gelabert, Lipende, Quillen, Meissner, Stone, Pusey, Mjungu, Kandel, Liebergall, Prada, Vidal, Krause, Yakir, Pääbo, Reich, Lalueza-Fox, Marques-Bonet, Meshorer and Carmel, unpublished data). The identified alterations have been related to phenotypic variation, including limb, facial, and vocal tract morphology. These phenotypes are very likely to play critical roles in hominid evolution. For instance, the facial and vocal morphology can affect speech, which is an adaptive feature considered unique to anatomically modern humans (Rauschecker and Scott; 2009). Different hominin lineages also seem to have a distinctive set of methylation signatures (Gokhman, Tamir, Housman, Rafinia, Colón, Gu, Ferrando, Gelabert, Lipende, Quillen, Meissner, Stone, Pusey, Mjungu, Kandel, Liebergall, Prada, Vidal, Krause, Yakir, Pääbo, Reich, Lalueza-Fox, Marques-Bonet, Meshorer and Carmel, unpublished data), which might be a projection of corresponding epigenetic response to both the environment and underlying genetics. With the new methods tailored for aDNA, both the quality and quantity of paleomethylome data are likely to rapidly increase.

2.4 Host-microbiome interactions and their implications in environmental adaptation

The microbiome is another indispensable part of animal biology that may play vital roles in animal adaptation. Trillions of microorganisms (microbiota) inhabit various body sites within animals, but the importance of these microbial communities and their genomic diversity (microbiome) was underestimated until recently (Huttenhower et al.; 2012). It has been recently revealed that microbiota outnumber their host's cells (Sender et al.; 2016). The vast numbers of microbes dynamically interact with their host in a complex and often beneficial manner: they can influence the maturation of host immunity, provide key nutrients, and affect host metabolism

(Huttenhower et al.; 2012; Lloyd-Price et al.; 2016; Kinross et al.; 2011). The disruption of microbiota and their functions is often associated with the development of many diseases (*e.g.* obesity, diabetes, mental health, skin disorders, and cancer) (Schwabe and Jobin; 2013; Grice and Segre; 2011; Hartstra et al.; 2015; Devaraj et al.; 2013). Because of their known relationships to animal health, gut and oral microbiomes are among the most extensively studied areas of microbiome research.

Different body sites are colonised by distinct microbiota, and the oral cavity harbours an especially diverse microbiome that is unique from other body sites (Turnbaugh et al.; 2007; Gill et al.; 2006; He et al.; 2015). The mouth of a healthy animal is typically home to over 200 microbial species (Dewhirst et al.; 2010, 2012), which belong to Firmicutes, Proteobacteria, Actinobacteria, Bacteroidetes and Fusobacteria phyla (He et al.; 2015). While many microbes in the mouth are planktonic, other species readily form a biofilm on the surface of teeth and soft epithelial tissues. Several *Streptococcus* species are key, 'primary' players in dental plaque formation, as they can adhere to tooth enamel and allow the secondary binding of other bacteria, which can dictate oral health outcomes (Dewhirst et al.; 2010). Furthermore, the oral microbiota plays roles in systemic diseases, including cardiovascular disease, diabetes, and cancer (Hartstra et al.; 2015; Ahn et al.; 2012; Shoemark and Allen; 2015; Nakano et al.; 2009).

Like the mouth, the gastrointestinal tract harbours an extensively studied microbiota that is responsible for food digestion, nutrition absorption, and intestine functions (Arumugam et al.; 2011; Turnbaugh et al.; 2009; Kau et al.; 2011). The gut microbiota can transform indigestible molecules into smaller and digestible nutrients, thereby increasing the nutrient bioaccessibility for the hosts (Kau et al.; 2011). For example, host enzymes within the intestine cannot digest dietary fibres, while Bacteroidetes bacteria, part of the core gut microbiota, can transform fibres into physiologically active metabolites, such as short-chain fatty acids (Trompette et al.; 2014). Gut microbiota can also synthesise essential vitamins (including B-group vitamins and vitamin K) and thus act as micronutrient suppliers (LeBlanc et al.; 2013). Besides facilitating intestine function, gut microbiota are also involved in microbial detoxification and play a role in modulating brain development and behaviour (Kohl et al.; 2014, 2016; Heijtz et al.; 2011).

2.4.1 Factors that contribute to microbiota variation

The gut microbiota exhibits great flexibility and plasticity and is dynamically shaped by the host and environment (Kinross et al.; 2011; Spor et al.; 2011). While gut microbial communities vary according to host species, the predominant factor that influences the animal gut microbiota is diet. Herbivorous, carnivorous and omniv-

orous animal gut microbiotas cluster into distinct groups (Ley et al.; 2008), and different microbiota compositions can be found in same animal species with different dietary habits. Bacteriodetes dominate the gut microbiota of children whose diet has a high content of carbohydrates, fibre, and non-animal protein, while Firmicutes dominate the gut microbiota of children whose diet is rich in animal proteins, sugar, and starch (De Filippo et al.; 2010). These dietary differences can also be sex-based. In the Hadza hunter-gatherers, *Treponema* species capable of fibre digestion are increased in Hadza women, which likely results from an adaptation to the higher amount of tubers in their diet compared to men (Schnorr et al.; 2014). In addition to diet, numerous other factors, including seasonality, habitat, and altitude, can also impact the microbiota in animals (Smits et al.; 2017; Guan et al.; 2017; Adak et al.; 2013).

While multiple factors can affect animal gut microbiota, these changes can potentially influence host metabolism, disease susceptibility, behaviour, and consequently impact animal adaptation to environment. The ability to synthesise various micronutrients and influence the host digestive efficiency might be vital for animals' survival during rapid environmental changes and during times when food intake is insufficient and nutrient deficiencies occur (Allen et al.; 2010). Dietary- or environmentally-induced adaptive changes in microbiota have been observed in mammals. For instance, some desert woodrats (*Neotoma lepida*) have a specialised gut microbiota for detoxification, allowing an adaptation to a toxic diet (Kohl et al.; 2014), while Giant pandas (*Ailuropoda melanoleuca*) harbour a gut microbiota with increased cellulose and lignin degradation activities for their adaptation to the bamboo-dominated diet (Zhu et al.; 2011). In addition, during times of infectious disease stress, unique microbes may provide protection, or past exposure to related microorganisms may even provide immunity to some diseases (Kinross et al.; 2011).

2.4.2 Trans-generational effects of microbiota on animal adaptation

One of the most important foundations of co-evolution and co-adaptation between animals and their microbiota is that parental (mostly maternal) microbiota can be passed down through vertical transmission to the offspring (Li et al.; 2005; Cho and Blaser; 2012). Direct contact between infants and vaginal microbes during birth, as well as subsequent mother-infant interactions (*e.g.*, breastfeeding, direct food sharing) can mediate the direct transmission of microbiota (Hyman et al.; 2014). The transmission of microorganisms in later stages of life is less clear, although there is evidence suggesting it continues through life (Song et al.; 2013).

In the long term, vertically-transmitted microbes can contribute to animal adaptation independently of the genome. The giant panda (*Ailuropoda melanoleuca*) is such an example. Genomic and morphological evidence strongly supports that the giant panda belonging to the Ursidae family, which is primarily composed of carnivore species. However, despite the fact that the genome of the giant panda contains genes encoding enzymes for meat digestion, the giant panda's diet is primarily herbivorous and consists almost exclusively of bamboo (Li et al.; 2010; Dierenfeld et al.; 1982). The discordance between phylogeny and feeding habit may be explained by its gut microbiota. Although the overall gut microbiome profile of giant panda is similar to that of a carnivore (Xue et al.; 2015), functional analysis of its gut microbiome revealed the atypical presence of microbial genes encoding enzymes that participate in cellulose metabolism, which is likely a key element that enables them to adapt to a bamboo-centric diet (Zhu et al.; 2011). Similarly, koalas rely on tailored gut microorganisms to aid in the detoxification of the plant material that they consume and mediate the transfer of these crucial microbes through coprophagy (Osawa et al.; 1993). In either case, the direct transfer of microorganisms is essential for specific adaptations to a given environment or diet.

In some cases, such co-evolution signal is so strong that the evolution path of a bacterium mirrors that of its host (*i.e.*, phylosymbiosis (Brooks et al.; 2016); *e.g.*, (Linz et al.; 2007)). Such phylosymbiosis provides an alternative to reconstruct the animal evolutionary history, and the ancient microbes can sometimes serve as a timestamp for the calibration. For example, the history of human migration recovered from microbes that colonize the human body and human mitochondrial genomes are strikingly concordant (Comas et al.; 2013; Linz et al.; 2007). While modern *Helicobacter pylori* population supports the "out-of-Africa" theory, the 5300-year-old *H. pylori* genome pinpoints the timing of the African arrival in Europe (Linz et al.; 2007; Maixner et al.; 2016). Nevertheless, some microbes preserve stronger co-evolutionary signal than others. In hominid gut microbiota, the phylogeny of Bacteroidaceae and Bifidobacteriaceae parallels their hosts, while those of other gut species do not (Groussin et al.; 2017). Several bacterial strains from Bacteroidaceae and Bifidobacteriaceae also show strong evidence of vertical transmission (Jost et al.; 2014; Milani et al.; 2015). A proportion of vertically-transmitted bacteria are beneficial to the host (*e.g. Lactobacillus, Bifidobacterium*) and likely aid the adaptation of the host for the prolonged periods (Milani et al.; 2015; Matsumiya et al.; 2002).

2.4.3 Methods and progress in ancient microbiome research

Although the evolutionary history of a microbiome can be reconstructed by examining similarities and differences in the microbiome from related species (Moeller

et al.; 2014), a temporal record is essential to draw an accurate picture of how microbiota contribute to animal adaptation to the environment. Recent advances in paleomicrobiology provide powerful tools for examining the co-evolutionary history of animals and their microbiota from alternative sources (Darling and Donoghue; 2014; Warinner, Speller and Collins; 2015). Currently, paleomicrobiome information comes from two main sources: fossilised faeces (coprolites) and dental calculus, which reflect gut and oral microbiomes, respectively (Warinner, Speller, Collins and Lewis Jr; 2015; Warinner, Speller and Collins; 2015).

Animal dung that is quickly desiccated or covered by clay sediment can be preserved as a coprolite (McAllister; 1985). Coprolites contain food remnants (such as plant debris, pollen, or prey skeletal elements), host DNA, gut microbial DNA, and DNA originated from the environment (Rawlence et al.; 2016; Wood and Wilmshurst; 2013; Tito et al.; 2012). Several studies have recovered bacterial, fungal, and archaeal information from human and animal coprolites (Wood et al.; 2012; Santiago-Rodriguez et al.; 2013; van Geel et al.; 2011), alongside digested and undigested food used to infer the diet of the host. However, this method is not without limitations. Because faeces are rich in organic material that can be used by microbes after deposition outside of the host, the microbial community continues to change after defecation. For example, the microbial community within a coprolite from a Latin American mummy resembled that in a modern compost pile, rather than that of the human gut (Tito et al.; 2012). Additionally, coprolites are susceptible to contamination from the surrounding environment. Thus, the ancient gut microbiome information obtained from coprolites is usually biased and heavily subject to environmental contamination (Warinner, Speller, Collins and Lewis Jr; 2015). Furthermore, faeces typically decay rapidly and very few of them can be fossilised and preserved over a long time period, which makes the coprolite record more broken and incomplete than skeletal records (Warinner, Speller, Collins and Lewis Jr; 2015). Although host genetic information can be obtained from coprolite in some cases (some gut epithelial cells might be present in the faeces and provide host DNA), the lack of skeletal evidence of a specific host makes it difficult to identify a coprolite's origins.

In contrast to coprolites, dental calculus (calcified matrix formed from biofilm on the teeth surface) is frequently found on the surface of ancient human teeth and is a more accessible source material than coprolites for recovering the evolutionary history of microbiomes (Warinner, Speller and Collins; 2015; Jin and Yip; 2002; Lieverse; 1999). Calcified and non-calcified bacteria have been observed in dental calculus using transmission electron microscopy and gold-labelled antibodies (Warinner, Speller and Collins; 2015). In addition, dental calculus provides an environment suitable for the preservation of ancient microbial DNA (Warinner, Speller

and Collins; 2015). Several studies have shown that microbial DNA can be successfully extracted and amplified from human dental calculus (Adler et al.; 2013; Warinner, Speller and Collins; 2015; Weyrich et al.; 2017). Adler *et al.* (2013) introduced HTS into a paleomicrobiome study and collected the first detailed genetic data from oral microbiome of 34 ancient European human remains (Adler et al.; 2013). They obtained ancient oral microbiome profiles using 16S ribosomal RNA (rRNA) gene amplicon sequencing techniques and observed a shift in the microbiome linked to dietary alterations. Several studies have since identified biases in using 16S rRNA gene amplicon sequencing, which is heavily subject to taphonomic and amplification biases (Weyrich et al.; 2017; Ziesemer et al.; 2015). Shotgun sequencing is now accepted as the gold-standard to reconstruct ancient microbiomes (Ziesemer et al.; 2015). Shotgun libraries include a subsample of all the DNA fragments instead of just a prokaryotic genetic marker. Thus, it can provide additional information on the host, the diet and environment, lifestyle, and functional profile of the microbiome. Such information can be used to infer ancient animal diet, behaviour, and disease, as well as the interaction between ancient microbiome and their host (Baker et al.; 2017; Weyrich et al.; 2017). Nevertheless, amplicon sequencing is more cost-effective than shotgun sequencing and can still provide important microbiome information in terms of the presence and absence of taxa—but not relative or absolute abundance—due to the differential impact of degradation processes on microbial species (Boast et al.; 2018).

These aDNA studies demonstrate the ability to accurately reconstruct animal microbiome records across evolutionary timescales, and thereby to investigate the interactions of microbiome, animal, and environment, revealing the roles that the microbiome may play in animal adaptation. Notably, dental calculus deposits are rare on most non-human mammals, but it is very likely that some oral microbiome information can be obtained from ancient mammal tooth specimens (*e.g.* similar, non-calcified biofilms formed by oral microbiome or food debris preserved in the occlusal surfaces and gaps of mammalian herbivore teeth). In addition to ancient microbial community, specific pathogens (*e.g.*, *Yersinia pestis* and *Mycobacterium tuberculosis*) identified from ancient samples also have the potential to reveal epidemic events in the past that had immense impacts on ancient animals (Scott; 1988; Bos et al.; 2011). As current research is mainly limited to ancient humans, we advocate here that an emphasis should be placed on ancient microbiome research in other animals as well.

2.5 Dynamics among epigenome, microbiome and environment

The external factors that shape epigenetic modifications and the microbiome largely overlap (*e.g.*, diet composition, lifestyle, and exposure to stressors). When animals are exposed to changing environments, it is highly likely that epigenetics and the microbiome respond to the stimuli simultaneously. Some phenotypic consequences of the alteration of epigenetics and the microbiome, such as embryonic development and immunity establishment, are crucial to the ability to survive an adverse environment. Unlike genetic adaptation, which is a long-term process, epigenetics and microbiome can respond rapidly to environmental cues (Yona et al.; 2015; Alberdi et al.; 2016). In particular, modifications that occur during the prenatal period or early stages of life can have life-long or even a trans-generational influence on animals (Anderson et al.; 2012; Cooney et al.; 2002; Li et al.; 2005; Mueller et al.; 2015).

Furthermore, emerging evidence suggests that the microbiome can directly interact with the host epigenome (Hullar and Fu; 2014; Paul et al.; 2015). Some pathogenic bacteria (*e.g.*, *M. leprae* or *H. pylori*) can induce epigenome modifications of the infected host cells, and sometimes even trigger epigenome-wide alterations (Cizmeci et al.; 2016; De Monerri and Kim; 2014). Global disruption of methylation reprogramming has been detected in germ-free conditions (*i.e.*, in the absence of a microbiome) (Yu et al.; 2015), and the gut microbiome composition shows a clear association to host epigenomic profiles (Kumar et al.; 2014). The crosstalk between epigenome and microbiome is not unexpected, as epigenetics plays an important part in shaping immunity (Amarasekera et al.; 2013), which directly affects the community composition and ecology of indigenous microorganisms(Kau et al.; 2011) ; conversely, the microbiome can release molecules (such as folate and transposases) that are directly or indirectly involved in the modification of the host epigenome (De Monerri and Kim; 2014; LeBlanc et al.; 2013). In this context, it is likely that epigenome, microbiome, and the environment form a complex and dynamic three-way interaction that could be the basis for rapid adaptation of animals to changing environments.

2.6 Epigenome and microbiome interactions with the genome

Fast response of animal epigenomes and microbiomes to changing environments is the key to rapid adaptation discussed in this review. However, the observed changes in epigenomes and microbiomes do not necessarily originate independently of the

host genome. It is also possible that genetic mutations cause subsequent alterations in epigenetics and microbiome (Jones; 2012; Fulde et al.; 2018). Such a case should not be considered as adaptation via epigenetic modification or microbiome alteration, because it fundamentally stems from the genome and will be much slower than the within-generation alterations of microbiome and epigenetics. Thus, in paleomicrobiome and paleoepigenetics studies, it is critical to differentiate three possible scenarios. First, environmental changes directly triggered animal microbiome and/or epigenome alternations, in which no associated genetic alterations should be detected. Second, genetic mutations led to subsequent epigenetic and microbiome modifications. In this case, genetic changes with a causal link to the epigenetic and/or microbiome signatures should be detected. Third, the genome, epigenome, and microbiome changed in parallel. In this scenario, the effects should be independently observed and verified in epigenome, microbiome and genome. It can be difficult to determine if a genetic mutation can cause epigenome or microbiome modifications. One possible solution is to carry out additional experiments using animal models. Alternatively, genomic variation might be modelled as a confounding factor and then removed using statistical approaches (*e.g.*, surrogate variable analysis or SVA)(Leek and Storey; 2007).

The complex interactions between environment, genome, microbiome, and epigenome make animal adaptation to the environment an extremely complicated yet fascinating process. In order to recover a comprehensive evolutionary history, it is important to obtain epigenetic and microbiome information along with genomic information, as each factor will likely play a role in past animal adaptation.

2.7 Concluding remarks

The role of epigenetics and the microbiome in animal adaptation has attracted increased attention in recent years, including the resurgence of two key theories: neo-Lamarckism and the hologenome theory of evolution (Zilber-Rosenberg and Rosenberg; 2008; Skinner; 2015; Danchin et al.; 2019). It also raised extensive debate (Laland et al.; 2014; van Opstal and Bordenstein; 2015; Horsthemke; 2018), which is mainly due to the limited understanding of both fields and inconsistency of experimental results in different animal models. There are several major challenge in current studies, including (1) the lack of a universally accepted animal model and experimental system that can serve as a gold standard; (2) the plasticity and tissue- and niche-specificity of epigenome and microbiome responses that make results very difficult to reproduce independently; and (3) experimental studies using modern animals can hardly be done over microevolutionary timescales (Rosenfeld; 2010; Tripathi et al.; 2018). Within this context, it is necessary to develop a model

system in which both external- and internal- confounding factors are controlled and monitored, especially for the sake of elucidating the basic molecular mechanisms and dynamics underlying epigenetics and microbiomes.

On the other hand, numerous animal species that have experienced past climate and environmental turnovers provide an extraordinary repertoire for case studies. Although natural environments are more complex and less controlled than laboratory conditions, the availability of paleogenomic and paleoenvironmental data is fast accumulating, and recently developed approaches offer access to high quality paleoepigenome and paleomicrobiome data. Teasing apart the epigenomic, microbiomic, and genomic components in animal adaptation to environment might become increasingly feasible. More importantly, various animal species that have lived in diverse environments and have distinct evolutionary paths offer invaluable resources for future research.

Here, we propose that increasing efforts should be placed in paleoepigenomic and paleomicrobiomic research across the tree of life, including but not limited to (1) experimental and bioinformatics approaches further tailored for recovering epigenomic and microbiomic data from short and damaged DNA molecules; (2) applying cutting-edge approaches to retrieve paleoepigenomic and paleomicrobiomic data from non-human species, especially those have a large number of sub-fossil specimens available; (3) the generated data should be deposited into public repository with detailed metadata; (4) pipelines should be developed to make data generated using different approaches comparable, such as paleoepigenomic data generated using aDNA damage profile and bisulfite sequencing, as well as paleomicrobiomic data generated using shotgun sequencing and amplicon sequencing. In conclusion, we believe that, with the aggregation of advanced aDNA techniques and increasing understanding of epigenome and microbiome, novel insights into animal adaption to rapidly changing environments can be yielded in the near future.

Bibliography

Adak, A., Maity, C., Ghosh, K., Pati, B. R. and Mondal, K. C. (2013). Dynamics of predominant microbiota in the human gastrointestinal tract and change in luminal enzymes and immunoglobulin profile during high-altitude adaptation, *Folia microbiologica* **58**(6): 523–528.

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A., Alt, K. W. et al. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the neolithic and industrial revolutions, *Nature genetics* **45**(4): 450.

Ahn, J., Chen, C. Y. and Hayes, R. B. (2012). Oral microbiome and oral and gastrointestinal cancer risk, *Cancer Causes & Control* **23**(3): 399–404.

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L. and Gilbert, M. T. P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation?, *Trends in ecology & evolution* **31**(9): 689–699.

Allen, J. R., Hickler, T., Singarayer, J. S., Sykes, M. T., Valdes, P. J. and Huntley, B. (2010). Last glacial vegetation of northern eurasia, *Quaternary Science Reviews* **29**(19-20): 2604–2618.

Amarasekera, M., Prescott, S. L. and Palmer, D. J. (2013). Nutrition in early life, immune-programming and allergies: the role of epigenetics, *Asian Pac J Allergy Immunol* **31**(3): 175–82.

Anderson, O. S., Sant, K. E. and Dolinoy, D. C. (2012). Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and dna methylation, *The Journal of nutritional biochemistry* **23**(8): 853–859.

Anway, M. D. and Skinner, M. K. (2006). Epigenetic transgenerational actions of endocrine disruptors, *Endocrinology* **147**(6): s43–s49.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M. et al. (2011). Enterotypes of the human gut microbiome, *nature* **473**(7346): 174.

Baker, J. L., Bor, B., Agnello, M., Shi, W. and He, X. (2017). Ecology of the oral microbiome: beyond bacteria, *Trends in microbiology* **25**(5): 362–374.

Bind, M.-A. C., Coull, B. A., Baccarelli, A., Tarantini, L., Cantone, L., Vokonas, P. and Schwartz, J. (2016). Distributional changes in gene-specific methylation associated with temperature, *Environmental research* **150**: 38–46.

Bind, M.-A., Zanobetti, A., Gasparrini, A., Peters, A., Coull, B., Baccarelli, A., Tarantini, L., Koutrakis, P., Vokonas, P. and Schwartz, J. (2014). Effects of temperature and relative humidity on dna methylation, *Epidemiology (Cambridge, Mass.)* **25**(4): 561.

Boast, A. P., Weyrich, L. S., Wood, J. R., Metcalf, J. L., Knight, R. and Cooper, A. (2018). Coprolites reveal ecological interactions lost with the extinction of new zealand birds, *Proceedings of the National Academy of Sciences* p. 201712337.

Bollati, V. and Baccarelli, A. (2010). Environmental epigenetics, *Heredity* **105**(1): 105.

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S. et al. (2011). A draft genome of yersinia pestis from victims of the black death, *Nature* **478**(7370): 506.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M. and Pääbo, S. (2009). Removal of deaminated cytosines and detection of in vivo methylation in ancient dna, *Nucleic acids research* **38**(6): e87–e87.

Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. and Bordenstein, S. R. (2016). Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history, *PLoS biology* **14**(11): e2000225.

Cahill, A. E., Aiello-Lammens, M. E., Fisher-Reid, M. C., Hua, X., Karanewsky, C. J., Yeong Ryu, H., Sbeglia, G. C., Spagnolo, F., Waldron, J. B., Warsi, O. et al. (2013). How does climate change cause extinction?, *Proceedings of the Royal Society B: Biological Sciences* **280**(1750): 20121890.

Cao-Lei, L., Massart, R., Suderman, M. J., Machnes, Z., Elgbeili, G., Laplante, D. P., Szyf, M. and King, S. (2014). Dna methylation signatures triggered by prenatal maternal stress exposure to a natural disaster: Project ice storm, *PloS one* **9**(9): e107653.

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G., Doctor, B., Pardhasaradhi, K. and McCann, P. P. (1996). S-adenosylmethionine and methylation., *The FASEB journal* **10**(4): 471–480.

Cho, I. and Blaser, M. J. (2012). The human microbiome: at the interface of health and disease, *Nature Reviews Genetics* **13**(4): 260.

Cizmeci, D., Dempster, E. L., Champion, O. L., Wagley, S., Akman, O. E., Prior, J. L., Soyer, O. S., Mill, J. and Titball, R. W. (2016). Mapping epigenetic changes to the host cell genome induced by burkholderia pseudomallei reveals pathogen-specific and pathogen-generic signatures of infection, *Scientific reports* **6**: 30861.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G. et al. (2013). Out-of-africa migration and neolithic coexpansion of mycobacterium tuberculosis with modern humans, *Nature genetics* **45**(10): 1176.

Cooney, C. A., Dave, A. A. and Wolff, G. L. (2002). Maternal methyl supplements in mice affect epigenetic variation and dna methylation of offspring, *The Journal of nutrition* **132**(8): 2393S–2400S.

Cooper, A. and Poinar, H. N. (2000). Ancient dna: do it right or not at all, *Science* **289**(5482): 1139–1139.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. and Bradshaw, C. J. (2015). Abrupt warming events drove late pleistocene holarctic megafaunal turnover, *Science* **349**(6248): 602–606.

Cropley, J. E., Dang, T. H., Martin, D. I. and Suter, C. M. (2012). The penetrance of an epigenetic trait in mice is progressively yet reversibly increased by selection and environment, *Proceedings of the Royal Society of London B: Biological Sciences* **279**(1737): 2347–2353.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L. A., Wang, Z., Guo, Z., Xu, L. et al. (2013). Historical variations in mutation rate in an epidemic pathogen, yersinia pestis, *Proceedings of the National Academy of Sciences* **110**(2): 577–582.

Dabney, J., Meyer, M. and Pääbo, S. (2013). Ancient dna damage, *Cold Spring Harbor perspectives in biology* p. a012567.

Danchin, É., Pocheville, A. and Huneman, P. (2019). Early in life effects and heredity: reconciling neo-darwinism with neo-lamarckism under the banner of the inclusive evolutionary synthesis, *Philosophical Transactions of the Royal Society B* **374**(1770): 20180113.

Darling, M. I. and Donoghue, H. D. (2014). Insights from paleomicrobiology into the indigenous peoples of pre-colonial america-a review, *Memórias do Instituto Oswaldo Cruz* **109**(2): 131–139.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G. and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from europe and rural africa, *Proceedings of the National Academy of Sciences* **107**(33): 14691–14696.

De Monerri, N. C. S. and Kim, K. (2014). Pathogens hijack the epigenome: a new twist on host-pathogen interactions, *The American journal of pathology* **184**(4): 897–911.

Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J. J., Reik, W. and Feil, R. (1998). Altered imprinted gene methylation and expression in completely es cell-derived mouse fetuses: association with aberrant phenotypes, *Development* **125**(12): 2273–2282.

Devaraj, S., Hemarajata, P. and Versalovic, J. (2013). The human gut microbiome and body metabolism: implications for obesity and diabetes, *Clinical chemistry* **59**(4): 617–628.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W.-H., Lakshmanan, A. and Wade, W. G. (2010). The human oral microbiome, *Journal of bacteriology* **192**(19): 5002–5017.

Dewhirst, F. E., Klein, E. A., Thompson, E. C., Blanton, J. M., Chen, T., Milella, L., Buckley, C. M., Davis, I. J., Bennett, M.-L. and Marshall-Jones, Z. V. (2012). The canine oral microbiome, *PloS one* **7**(4): e36067.

Dierenfeld, E., Hintz, H., Robertson, J., Van Soest, P. and Oftedal, O. (1982). Utilization of bamboo by the giant panda, *The Journal of Nutrition* **112**(4): 636–641.

Fulde, M., Sommer, F., Chassaing, B., van Vorst, K., Dupont, A., Hensel, M., Basic, M., Klopfleisch, R., Rosenstiel, P., Bleich, A. et al. (2018). Neonatal selection by toll-like receptor 5 influences long-term gut microbiota composition, *Nature* **560**(7719): 489.

Fuso, A., Seminara, L., Cavallaro, R. A., D'anselmi, F. and Scarpa, S. (2005). S-adenosylmethionine/homocysteine cycle alterations modify dna methylation status with consequent deregulation of ps1 and bace and beta-amyloid production, *Molecular and Cellular Neuroscience* **28**(1): 195–204.

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. and Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome, *science* **312**(5778): 1355–1359.

Gillespie, J. H. (2004). *Population genetics: a concise guide*, JHU Press.

Gokhman, D., Lavi, E., Prüfer, K., Fraga, M. F., Riancho, J. A., Kelso, J., Pääbo, S., Meshorer, E. and Carmel, L. (2014). Reconstructing the dna methylation maps of the neandertal and the denisovan, *Science* p. 1250368.

Gokhman, D., Malul, A. and Carmel, L. (2017). Inferring past environments from ancient epigenomes, *Molecular biology and evolution* **34**(10): 2429–2438.

Gokhman, D., Tamir, L. A., Housman, G., Rafinia, M. N., Colón, M. N., Gu, H., Ferrando, M., Gelabert, P., Lipende, I., Quillen, E. E., Meissner, A., Stone, A. C., Pusey, A. E., Mjungu, D., Kandel, L., Liebergall, M., Prada, M. E., Vidal, J. M., Krause, J., Yakir, B., Pääbo, S., Reich, D., Lalueza-Fox, C., Marques-Bonet, T., Meshorer, E. and Carmel, L. (2017). Recent regulatory changes shaped the human facial and vocal anatomy. unpublished data, https://www.biorxiv.org/node/30708.abstract, last accessed September 25, 2017.

Grice, E. A. and Segre, J. A. (2011). The skin microbiome, *Nature Reviews Microbiology* **9**(4): 244.

Groussin, M., Mazel, F., Sanders, J. G., Smillie, C. S., Lavergne, S., Thuiller, W. and Alm, E. J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time, *Nature communications* **8**: 14319.

Guan, Y., Yang, H., Han, S., Feng, L., Wang, T. and Ge, J. (2017). Comparison of the gut microbiota composition between wild and captive sika deer (cervus nippon hortulorum) from feces by high-throughput sequencing, *AMB Express* **7**(1): 212.

Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J. S., Willerslev, E. and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epipaleomix, *Molecular biology and evolution* **33**(12): 3284–3298.

Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y. et al. (2010). Comparison of sequencing-based methods to profile dna methylation and identification of monoallelic epigenetic modifications, *Nature biotechnology* **28**(10): 1097.

Hartstra, A. V., Bouter, K. E., Bäckhed, F. and Nieuwdorp, M. (2015). Insights into the role of the microbiome in obesity and type 2 diabetes, *Diabetes care* **38**(1): 159–165.

Hay, K., Morton, J., Mahony, T., Clements, A. and Barnes, T. S. (2016). Associations between animal characteristic and environmental risk factors and bovine respiratory disease in australian feedlot cattle, *Preventive veterinary medicine* **125**: 66–74.

He, J., Li, Y., Cao, Y., Xue, J. and Zhou, X. (2015). The oral microbiome diversity and its relation to human diseases, *Folia microbiologica* **60**(1): 69–80.

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E. and Lumey, L. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans, *Proceedings of the National Academy of Sciences* **105**(44): 17046–17049.

Heijtz, R. D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., Samuelsson, A., Hibberd, M. L., Forssberg, H. and Pettersson, S. (2011). Normal gut microbiota modulates brain development and behavior, *Proceedings of the National Academy of Sciences* **108**(7): 3047–3052.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. and Pääbo, S. (2001). ancient dna, *Nature Reviews Genetics* **2**(5): 353.

Holliday, R. (2006). Epigenetics: a historical overview, *Epigenetics* **1**(2): 76–80.

Horsthemke, B. (2018). A critical view on transgenerational epigenetic inheritance in humans, *Nature communications* **9**(1): 2973.

Hullar, M. A. and Fu, B. C. (2014). Diet, the gut microbiome, and epigenetics, *Cancer journal (Sudbury, Mass.)* **20**(3): 170.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S. et al. (2012). Structure, function and diversity of the healthy human microbiome, *Nature* **486**(7402): 207.

Hyman, R. W., Fukushima, M., Jiang, H., Fung, E., Rand, L., Johnson, B., Vo, K. C., Caughey, A. B., Hilton, J. F., Davis, R. W. et al. (2014). Diversity of the vaginal microbiome correlates with preterm birth, *Reproductive sciences* **21**(1): 32–40.

Jin, Y. and Yip, H.-K. (2002). Supragingival calculus: formation and control, *Critical reviews in oral biology & medicine* **13**(5): 426–441.

Jirtle, R. L. and Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility, *Nature reviews genetics* **8**(4): 253.

Jones, P. A. (2012). Functions of dna methylation: islands, start sites, gene bodies and beyond, *Nature Reviews Genetics* **13**(7): 484.

Jones, P. A. and Takai, D. (2001). The role of dna methylation in mammalian epigenetics, *Science* **293**(5532): 1068–1070.

Jost, T., Lacroix, C., Braegger, C. P., Rochat, F. and Chassard, C. (2014). Vertical mother–neonate transfer of maternal gut bacteria via breastfeeding, *Environmental microbiology* **16**(9): 2891–2904.

Kappil, M., Lambertini, L. and Chen, J. (2015). Environmental influences on genomic imprinting, *Current environmental health reports* **2**(2): 155–162.

Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. and Gordon, J. I. (2011). Human nutrition, the gut microbiome and the immune system, *Nature* **474**(7351): 327.

Kent, R., Bar-Massada, A. and Carmel, Y. (2014). Bird and mammal species composition in distinct geographic regions and their relationships with environmental factors across multiple spatial scales, *Ecology and evolution* **4**(10): 1963–1971.

Kinross, J. M., Darzi, A. W. and Nicholson, J. K. (2011). Gut microbiome-host interactions in health and disease, *Genome medicine* **3**(3): 14.

Kohl, K. D., Connelly, J. W., Dearing, M. D. and Forbey, J. S. (2016). Microbial detoxification in the gut of a specialist avian herbivore, the greater sage-grouse, *FEMS microbiology letters* **363**(14): fnw144.

Kohl, K. D., Weiss, R. B., Cox, J., Dale, C. and Denise Dearing, M. (2014). Gut microbes of mammalian herbivores facilitate intake of plant toxins, *Ecology letters* **17**(10): 1238–1246.

Koolhaas, J., Korte, S., De Boer, S., Van Der Vegt, B., Van Reenen, C., Hopster, H., De Jong, I., Ruis, M. and Blokhuis, H. (1999). Coping styles in animals: current status in behavior and stress-physiology, *Neuroscience & Biobehavioral Reviews* **23**(7): 925–935.

Kumar, H., Lund, R., Laiho, A., Lundelin, K., Ley, R. E., Isolauri, E. and Salminen, S. (2014). Gut microbiota as an epigenetic regulator: pilot study based on whole-genome methylation analysis, *MBio* **5**(6): e02113–14.

Laird, C. D., Pleasant, N. D., Clark, A. D., Sneeden, J. L., Hassan, K. A., Manley, N. C., Vary, J. C., Morgan, T., Hansen, R. S. and Stöger, R. (2004). Hairpin-bisulfite pcr: assessing epigenetic methylation patterns on complementary strands of individual dna molecules, *Proceedings of the National Academy of Sciences* **101**(1): 204–209.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E. et al. (2014). Does evolutionary theory need a rethink?, *Nature News* **514**(7521): 161.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D. and Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective, *Current opinion in biotechnology* **24**(2): 160–168.

Leek, J. T. and Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis, *PLoS genetics* **3**(9): e161.

Leontiou, C. A., Hadjidaniel, M. D., Mina, P., Antoniou, P., Ioannides, M. and Patsalis, P. C. (2015). Bisulfite conversion of dna: performance comparison of different kits and methylation quantitation of epigenetic biomarkers that have the potential to be used in non-invasive prenatal testing, *PLoS One* **10**(8): e0135058.

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. and Gordon, J. I. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota, *Nature Reviews Microbiology* **6**(10): 776.

Li, E., Beard, C. and Jaenisch, R. (1993). Role for dna methylation in genomic imprinting, *Nature* **366**(6453): 362.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y. et al. (2010). The sequence and de novo assembly of the giant panda genome, *Nature* **463**(7279): 311.

Li, Y., Caufield, P., Dasanayake, A., Wiener, H. and Vermund, S. (2005). Mode of delivery and other maternal factors influence the acquisition of streptococcus mutans in infants, *Journal of dental research* **84**(9): 806–811.

Lieverse, A. R. (1999). Diet and the aetiology of dental calculus, *International Journal of osteoarchaeology* **9**(4): 219–232.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S. W. et al. (2007). An african origin for the intimate association between humans and helicobacter pylori, *Nature* **445**(7130): 915.

Llamas, B., Heiniger, H., Gower, G., Liu, Y. and Cooper, A. (2017). Extinct bison methylomes using bisulfite sequencing, *The role of climate and environmental change in megafaunal extinctions*, 12th International Mammalogical Congress, Perth, Western Australia, p. 485.

Llamas, B., Holland, M. L., Chen, K., Cropley, J. E., Cooper, A. and Suter, C. M. (2012). High-resolution analysis of cytosine methylation in ancient dna, *PLoS One* **7**(1): e30226.

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A. and Haak, W. (2017). From the field to the laboratory: Controlling dna contamination in human ancient dna research in the high-throughput sequencing era, *STAR: Science & Technology of Archaeological Research* **3**(1): 1–14.

Lloyd-Price, J., Abu-Ali, G. and Huttenhower, C. (2016). The healthy human microbiome, *Genome medicine* **8**(1): 51.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R. et al. (2011). Species-specific responses of late quaternary megafauna to climate and humans, *Nature* **479**(7373): 359.

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M. R., Hallows, J. L., Kusebauch, U., Vigl, E. E., Malfertheiner, P., Megraud, F. et al. (2016). The 5300-year-old helicobacter pylori genome of the iceman, *Science* **351**(6269): 162–165.

Marsh, A. G. and Pasqualone, A. A. (2014). Dna methylation and temperature stress in an antarctic polychaete, spiophanes tcherniai, *Frontiers in physiology* **5**: 173.

Matsumiya, Y., Kato, N., Watanabe, K. and Kato, H. (2002). Molecular epidemiological study of vertical transmission of vaginal lactobacillus species from mothers

to newborn infants in japanese, by arbitrarily primed polymerase chain reaction, *Journal of infection and chemotherapy* **8**(1): 43–49.

McAllister, J. A. (1985). Reevaluation of the formation of spiral coprolites.

McDonald, J. N. (1981). North american bison, *Berkeley, CA: University of* .

Mellor, D. (2015). Positive animal welfare states and encouraging environment-focused and animal-to-animal interactive behaviours, *New Zealand veterinary journal* **63**(1): 9–16.

Milani, C., Mancabelli, L., Lugli, G. A., Duranti, S., Turroni, F., Ferrario, C., Mangifesta, M., Viappiani, A., Ferretti, P., Gorfer, V. et al. (2015). Exploring vertical transmission of bifidobacteria from mother to child, *Applied and environmental microbiology* pp. AEM–02037.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A. et al. (2008). Sequencing the nuclear genome of the extinct woolly mammoth, *Nature* **456**(7220): 387.

Moeller, A. H., Li, Y., Ngole, E. M., Ahuka-Mundeke, S., Lonsdorf, E. V., Pusey, A. E., Peeters, M., Hahn, B. H. and Ochman, H. (2014). Rapid changes in the gut microbiome during human evolution, *Proceedings of the National Academy of Sciences* **111**(46): 16431–16435.

Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z. and Dominguez-Bello, M. G. (2015). The infant microbiome development: mom matters, *Trends in molecular medicine* **21**(2): 109–117.

Nakano, K., Nemoto, H., Nomura, R., Inaba, H., Yoshioka, H., Taniguchi, K., Amano, A. and Ooshima, T. (2009). Detection of oral bacteria in cardiovascular specimens, *Oral microbiology and immunology* **24**(1): 64–68.

Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L. and Piferrer, F. (2011). Dna methylation of the gonadal aromatase (cyp19a) promoter is involved in temperature-dependent sex ratio shifts in the european sea bass, *PLoS genetics* **7**(12): e1002447.

Niculescu, M. D. and Zeisel, S. H. (2002). Diet, methyl donors and dna methylation: interactions between dietary folate, methionine and choline, *The Journal of nutrition* **132**(8): 2333S–2335S.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development, *Cell* **99**(3): 247–257.

Osawa, R., Blanshard, W. and Ocallaghan, P. (1993). Microbiological studies of the intestinal microflora of the koala, phascolarctos-cinereus. 2. pap, a special maternal feces consumed by juvenile koalas, *Australian Journal of Zoology* **41**(6): 611–620.

Parrott, B. B., Kohno, S., Cloy-McCoy, J. A. and Guillette Jr, L. J. (2014). Differential incubation temperatures result in dimorphic dna methylation patterning of the sox9 and aromatase promoters in gonads of alligator (alligator mississippiensis) embryos, *Biology of reproduction* **90**(1): 2–1.

Paul, B., Barnes, S., Demark-Wahnefried, W., Morrow, C., Salvador, C., Skibola, C. and Tollefsbol, T. O. (2015). Influences of diet and the gut microbiome on epigenetic modulation in cancer and other diseases, *Clinical epigenetics* **7**(1): 112.

Pedersen, J. S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S. et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome, *Genome research* **24**(3): 454–466.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing, *Nature neuroscience* **12**(6): 718.

Rawlence, N. J., Wood, J. R., Bocherens, H. and Rogers, K. M. (2016). Dietary interpretations for extinct megafauna using coprolites, intestinal contents and stable isotopes: Complimentary or contradictory?, *Quaternary Science Reviews* **142**: 173–178.

Rohland, N. and Hofreiter, M. (2007). Ancient dna extraction from bones and teeth, *Nature protocols* **2**(7): 1756.

Rosenfeld, C. S. (2010). Animal models to study environmental epigenetics, *Biology of reproduction* **82**(3): 473–488.

Sandom, C., Faurby, S., Sandel, B. and Svenning, J.-C. (2014). Global late quaternary megafauna extinctions linked to humans, not climate change, *Proc. R. Soc. B* **281**(1787): 20133254.

Santiago-Rodriguez, T. M., Narganes-Storde, Y. M., Chanlatte, L., Crespo-Torres, E., Toranzos, G. A., Jimenez-Flores, R., Hamrick, A. and Cano, R. J. (2013). Microbial communities in pre-columbian coprolites, *PloS one* **8**(6): e65191.

Sasaki, E., Shimada, T., Osawa, R., Nishitani, Y., Spring, S. and Lang, E. (2005). Isolation of tannin-degrading bacteria isolated from feces of the japanese large wood mouse, apodemus speciosus, feeding on tannin-rich acorns, *Systematic and applied microbiology* **28**(4): 358–365.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and dna fragmentation in ancient dna, *PloS one* **7**(3): e34131.

Schnorr, S. L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M. et al. (2014). Gut microbiome of the hadza hunter-gatherers, *Nature communications* **5**: 3654.

Schwabe, R. F. and Jobin, C. (2013). The microbiome and cancer, *Nature Reviews Cancer* **13**(11): 800.

Scott, M. E. (1988). The impact of infection and disease on animal populations: implications for conservation biology, *Conservation biology* **2**(1): 40–56.

Seguin-Orlando, A., Gamba, C., Der Sarkissian, C., Ermini, L., Louvel, G., Boulygina, E., Sokolov, A., Nedoluzhko, A., Lorenzen, E. D., Lopez, P. et al. (2015). Pros and cons of methylation-based enrichment methods for ancient dna, *Scientific reports* **5**: 11826.

Sender, R., Fuchs, S. and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body, *PLoS biology* **14**(8): e1002533.

Shapiro, á. and Hofreiter, M. (2014). A paleogenomic perspective on evolution and gene function: new insights from ancient dna, *Science* **343**(6169): 1236573.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J. et al. (2004). Rise and fall of the beringian steppe bison, *Science* **306**(5701): 1561–1565.

Shen, L., Ahuja, N., Shen, Y., Habib, N. A., Toyota, M., Rashid, A. and Issa, J.-P. J. (2002). Dna methylation and environmental exposures in human hepatocellular carcinoma, *Journal of the National Cancer Institute* **94**(10): 755–761.

Shiffman, M. E., Soo, R. M., Dennis, P. G., Morrison, M., Tyson, G. W. and Hugenholtz, P. (2017). Gene and genome-centric analyses of koala and wombat fecal microbiomes point to metabolic specialization for eucalyptus digestion, *PeerJ* **5**: e4075.

Shoemark, D. K. and Allen, S. J. (2015). The microbiome and disease: reviewing the links between the oral microbiome, aging, and alzheimer's disease, *Journal of Alzheimer's Disease* **43**(3): 725–738.

Skinner, M. K. (2015). Environmental epigenetics and a unified theory of the molecular aspects of evolution: a neo-lamarckian concept that facilitates neo-darwinian evolution, *Genome biology and evolution* **7**(5): 1296–1302.

Skvortsova, K., Iovino, N. and Bogdanović, O. (2018). Functions and mechanisms of epigenetic inheritance in animals, *Nature Reviews Molecular Cell Biology* p. 1.

Smith, R. W., Monroe, C. and Bolnick, D. A. (2015). Detection of cytosine methylation in ancient dna from five native american populations using bisulfite sequencing, *PloS one* **10**(5): e0125344.

Smith, Z. D. and Meissner, A. (2013). Dna methylation: roles in mammalian development, *Nature Reviews Genetics* **14**(3): 204.

Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J. E. et al. (2017). Seasonal cycling in the gut microbiome of the hadza hunter-gatherers of tanzania, *Science* **357**(6353): 802–806.

Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., Caporaso, J. G., Knights, D., Clemente, J. C., Nakielny, S. et al. (2013). Cohabiting family members share microbiota with one another and with their dogs, *elife* **2**: e00458.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y., Kosintsev, P., Lee, M. S., Baryshnikov, G. et al. (2016). Early cave art and ancient dna record the origin of european bison, *Nature communications* **7**: 13158.

Spor, A., Koren, O. and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome, *Nature Reviews Microbiology* **9**(4): 279.

Telford, M. J., Budd, G. E. and Philippe, H. (2015). Phylogenomic insights into animal evolution, *Current Biology* **25**(19): R876–R887.

Tito, R. Y., Knights, D., Metcalf, J., Obregon-Tito, A. J., Cleeland, L., Najar, F., Roe, B., Reinhard, K., Sobolik, K., Belknap, S. et al. (2012). Insights from characterizing extinct human gut microbiomes, *PloS one* **7**(12): e51146.

Tobi, E. W., Lumey, L., Talens, R. P., Kremer, D., Putter, H., Stein, A. D., Slagboom, P. E. and Heijmans, B. T. (2009). Dna methylation differences after exposure to prenatal famine are common and timing-and sex-specific, *Human molecular genetics* **18**(21): 4046–4053.

Tripathi, A., Marotz, C., Gonzalez, A., Vázquez-Baeza, Y., Song, S. J., Bouslimani, A., McDonald, D., Zhu, Q., Sanders, J. G., Smarr, L. et al. (2018). Are microbiome studies ready for hypothesis-driven research?, *Current opinion in microbiology* **44**: 61–69.

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blanchard, C., Junt, T., Nicod, L. P., Harris, N. L. et al. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis, *Nature medicine* **20**(2): 159.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P. et al. (2009). A core gut microbiome in obese and lean twins, *nature* **457**(7228): 480.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R. and Gordon, J. I. (2007). The human microbiome project, *Nature* **449**(7164): 804.

van Geel, B., Guthrie, R. D., Altmann, J. G., Broekens, P., Bull, I. D., Gill, F. L., Jansen, B., Nieman, A. M. and Gravendeel, B. (2011). Mycological evidence of coprophagy from the feces of an alaskan late glacial mammoth, *Quaternary Science Reviews* **30**(17-18): 2289–2303.

van Opstal, E. J. and Bordenstein, S. R. (2015). Rethinking heritability of the microbiome, *Science* **349**(6253): 1172–1173.

Veenendaal, M. V., Painter, R. C., de Rooij, S. R., Bossuyt, P. M., van der Post, J. A., Gluckman, P. D., Hanson, M. A. and Roseboom, T. J. (2013). Transgenerational effects of prenatal exposure to the 1944–45 dutch famine, *BJOG: An International Journal of Obstetrics & Gynaecology* **120**(5): 548–554.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W. et al. (2014). Programming and inheritance of parental dna methylomes in mammals, *Cell* **157**(4): 979–991.

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): 20130376.

Warinner, C., Speller, C., Collins, M. J. and Lewis Jr, C. M. (2015). Ancient human microbiomes, *Journal of human evolution* **79**: 125–136.

Waterland, R. A., Dolinoy, D. C., Lin, J.-R., Smith, C. A., Shi, X. and Tahiliani, K. G. (2006). Maternal methyl supplements increase offspring dna methylation at axin fused, *genesis* **44**(9): 401–406.

Welker, F., Duijm, E., van der Gaag, K. J., van Geel, B., de Knijff, P., van Leeuwen, J., Mol, D., van der Plicht, J., Raes, N., Reumer, J. et al. (2014). Analysis of coprolites from the extinct mountain goat myotragus balearicus, *Quaternary Research* **81**(1): 106–116.

Weyrich, A., Jeschek, M., Schrapers, K., Lenz, D., Chung, T., Rübensam, K., Yasar, S., Schneemann, M., Ortmann, S., Jewgenow, K. et al. (2018). Diet changes alter paternally inherited epigenetic pattern in male wild guinea pigs, *Environmental epigenetics* **4**(2): dvy011.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V. et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus, *Nature* **544**(7650): 357.

Wood, J. R. and Wilmshurst, J. M. (2013). Pollen analysis of coprolites reveals dietary details of heavy-footed moa (pachyornis elephantopus) and coastal moa (euryapteryx curtus) from central otago, *New Zealand Journal of Ecology* **37**(1): 151.

Wood, J. R., Wilmshurst, J. M., Wagstaff, S. J., Worthy, T. H., Rawlence, N. J. and Cooper, A. (2012). High-resolution coproecology: using coprolites to reconstruct the habits and habitats of new zealand's extinct upland moa (megalapteryx didinus), *PloS one* **7**(6): e40025.

Xue, Z., Zhang, W., Wang, L., Hou, R., Zhang, M., Fei, L., Zhang, X., Huang, H., Bridgewater, L. C., Jiang, Y. et al. (2015). The bamboo-eating giant panda harbors a carnivore-like gut microbiota, with excessive seasonal variations, *MBio* **6**(3): e00022–15.

Yona, A. H., Frumkin, I. and Pilpel, Y. (2015). A relay race on the evolutionary adaptation spectrum, *Cell* **163**(3): 549–559.

Yu, D.-H., Gadkari, M., Zhou, Q., Yu, S., Gao, N., Guan, Y., Schady, D., Roshan, T. N., Chen, M.-H., Laritsky, E. et al. (2015). Postnatal epigenetic regulation of intestinal stem cells requires dna methylation and is guided by the microbiome, *Genome biology* **16**(1): 211.

Zhao, L., Sun, M.-a., Li, Z., Bai, X., Yu, M., Wang, M., Liang, L., Shao, X., Arnovitz, S., Wang, Q. et al. (2014). The dynamics of dna methylation fidelity during mouse embryonic stem cell self-renewal and differentiation, *Genome research* pp. gr–163147.

Zhu, L., Wu, Q., Dai, J., Zhang, S. and Wei, F. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome, *Proceedings of the National Academy of Sciences* **108**(43): 17714–17719.

Ziesemer, K. A., Mann, A. E., Sankaranarayanan, K., Schroeder, H., Ozga, A. T., Brandt, B. W., Zaura, E., Waters-Rist, A., Hoogland, M., Salazar-García, D. C. et al. (2015). Intrinsic challenges in ancient microbiome reconstruction using 16s rrna gene amplification, *Scientific Reports* **5**: 16498.

Zilber-Rosenberg, I. and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution, *FEMS microbiology reviews* **32**(5): 723–735.

Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. T.-Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E. et al. (2013). Charting a dynamic dna methylation landscape of the human genome, *Nature* **500**(7463): 477.

Statement of Authorship

Title of Paper	Ancient DNA analysis of coprolites reveals the role of gut microbiota in mammal adaptation to environment
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	In preperation for submission to Current Biology

Principal Author

Name of Principal Author (Candidate)	Yichen Liu
Contribution to the Paper	Data analyses, data interpretation, made the figures and drafted the manuscript
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 05/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    i.   the candidate's stated contribution to the publication is accurate (as detailed above);

    ii.   permission is granted for the candidate in include the publication in the thesis; and

    iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Adam Rohrlach
Contribution to the Paper	Helped design statistical analysis with principal author. Helped to draft statistical component of manuscript with principal author.
Signature	Date 01-03-2019
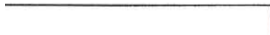
Name of Co-Author	Luis Arriola
Contribution to the Paper	Genome assembly, data interpretation, and reviewed the manuscript
Signature	Date 01-MAR-2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

iv.   the candidate's stated contribution to the publication is accurate (as detailed above);

v.   permission is granted for the candidate in include the publication in the thesis; and

vi.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Josep Antoni Alcover		
Contribution to the Paper	Obtaining Cova Estreta coprolites, discussion on chronologies 14C, incorporation of recent literature on evolution and diet of Myotragus, general discussion.		
Signature		Date	25 February 2019

Name of Co-Author	Jamie Wood		
Contribution to the Paper	Coprolite DNA extractions		
Signature		Date	22 Feb 2018

Name of Co-Author	Pere Bover		
Contribution to the Paper	Experiment and project design, laboratory work, sequencing data filtering and analyses, data interpretation, review of manuscript.		
Signature		Date	26/02/2019

Name of Co-Author	Laura S. Weyrich		
Contribution to the Paper	data interpretation and manuscript editing, project design.		
Signature		Date	5/3/19

Please cut and paste additional co-author panels here as required.

Chapter 3

Ancient DNA analysis of coprolites reveals the role of gut microbiota in mammal adaptation to environment

Abstract

Gut bacterial communities (microbiota) perform essential functions for their hosts, including nutrient synthesis, dietary toxin degradation, and host immunity development. Several key examples have recently identified additional, yet unique, roles that the gut microbiota can play in mammalian adaption. However, the role of gut microbiota in mammalian adaptation to past environments remains limited, especially within extinct animals. Here, we used the extinct cave goat (*Myotragus balearicus*), which maintained a toxic diet, as a model species to investigate how the gut microbiota aided mammal adaptation in the past. DNA was extracted from eight ancient goat coprolites (faecal remains), and metagenomic shotgun sequencing was preformed by creating both double- and single-stranded DNA libraries. Robust ancient gut microbiota were reconstructed from the coprolites by utilizing stringent controls and a novel model to assess DNA damage of unmapped DNA sequences. The probiotic bacterium *Romboutsia ilealis* was also enriched in the coprolites, allowing the reconstruction of three ancient draft *Romboutsia* genomes with coverage up to 80.5% and 12.1× depth. Further, functional analyses of the coprolite microbiomes (genetic and functional information maintained by microbiota) highlighted a detoxification role, suggesting that gut microbiota may have played a key role in the digestion of toxic plants. These results suggest that this extinct goat harboured a unique gut microbiota, which likely provided the ability to detoxify plants. This study demonstrates the strength of paleomicrobiology to explore adaptive processes of the past.

3.1 Introduction

Trillions of microorganisms (microbiota) inhabit the mammalian gut (Sender, Fuchs et al. 2016). These bacteria outnumber the host cells and play essential physiological roles for the hosts, by synthesising nutrients, degrading indigestible and toxic components, shaping the host immune system, and influencing host metabolism (Kinross et al.; 2011; Huttenhower et al.; 2012). Despite the fact that the gut microbiota of herbivores are generally conserved regarding the community structure and function (Muegge et al.; 2011), the gut microbiota of many herbivores possess unique features that are beneficial to their hosts . For instance, some dessert woodrats (*Neotoma lepida*) have gut microbiota specialised for the detoxification of toxic plants in the woodrat diet (Kohl et al.; 2014). Similarly, giant pandas (*Ailuropoda melanoleuca*) harbour a gut microbiota with enhanced cellulose and lignin degradation activities, which are likely essential for their adaptation to a bamboo-dominated diet (Zhu et al.; 2011).

Dynamically shaped by the diet, environment, and host, the animal gut microbiota exhibits great plasticity (Kinross et al.; 2011; Spor et al.; 2011). Parental (mostly maternal) microbiota can be passed down through vertical transmission to the offspring (Li et al.; 2005; Cho and Blaser; 2012). Some of the vertically-transmitted bacteria are beneficial to the host (*e.g. Lactobacillus and Bifidobacterium*) and likely increase the host's fitness for the prolonged periods (Matsumiya et al.; 2002; Milani et al.; 2015). In contrast to genetic adaptation, the gut microbiota can be altered by external stimuli within a short period (*e.g.*, the human gut microbiota can be altered by diet in two weeks (David et al.; 2014)), thus can potentially facilitate animal adaptation to the environment in a rapid manner (Alberdi et al.; 2016). However, this hypothesis has yet been fully explored on an evolutionary timescale, and if true, similar interactions would have been widespread throughout extinct species of the past and remain completely undiscovered.

A key place to examine environmental and dietary adaptation is on island. Islands are considered to be laboratories of evolutionary features that shaped the morphology and the ecological relationships of the wildlife living in them. Although a wide variety of fossil taxa have been recorded in islands around the World, Mediterranean islands hosted several mammals in the past that have been deeply studied because of the singular morphological characteristics (*e.g.*, (van Geel et al.; 2011)). The Eastern Balearic Islands, located at the Western Mediterranean Sea, hosted the bovid *Myotragus balearicus*, a dwarf fossil caprine (Antilopinae, Bovidae) from the Late Pleistocene to Holocene. In Mallorca, a whole lineage with six chronospecies (Bover et al.; 2014; Mas-Peinado et al.; 2018) can be traced back until the Early Pliocene. The ancestor of *M. balearicus* probably arrived through the land bridges

between mainland Europe and the Balearics during the Messinian Salinity Crisis (MSC) (Mas-Peinado et al.; 2018) that occurred 5.97-5.33 Mya (Krijgsman et al.; 1999; Manzi et al.; 2013), until its human-caused extinction around 4,300 years ago (Bover et al.; 2016). During this long evolution in a predator-free insular environment, *Myotragus* acquired numerous morphological changes indicative of adaptation to its environment, including a decrease in body size, increase of limb bone robustness, decrease of bone length (especially in metapodials and stylopodium elements), progressive reduction of number and size of incisiform and premolar teeth, and reduction of brain size and sense organs (*e.g.*, (Alcover; 1981; Köhler and Moyà-Solà; 2004; Bover and Tolosa; 2005)). In addition, a delay in life history schedules has been suggested for *M. balearicus* (Köhler and Moyà-Solà; 2009; Jordana and Köhler; 2011; Jordana et al.; 2012), further indicating a deep history of adaptation to its environment.

The diet of *M. balearicus* was likely also highly adapted to the Eastern Balearic Islands. Several dental irregularities, teeth wear patterns, dental topometry, and enamel surface texture, have suggested that this extinct goat maintained a generalistic feeding strategy and was likely able to exploit basically all available plant food (Alcover et al.; 1999; Bover and Alcover; 1999; Winkler et al.; 2013). This view was additionally supported by the analysis of micro and macroparticles from *M. balearicus* coprolites. Alcover et al. 1999 analyzed the pollen content of 14 well-preserved *M. balearicus* coprolites from the Late Pleistocene-Holocene deposit of Cova Estreta (Pollença, Mallorca). In addition to traces of pollen from common Balearic plants, as Poaceae, Pinus, Corylus, Plantago, *etc.*, these authors also observed a high content (~98%) of Buxus balearica pollen – a well known toxic plant on the Balearic Islands, suggesting that *M. balearicus* could feed on a toxic plant species. Similarly, Bartolomé et al. 2011; 2011 analyzed the plant fiber content of 10 coprolites from the same cave and identified Buxus plant fibers, albeit at lower abundances than observed with pollen (30-82%). Other sites have generated mixed findings, likely due to the poor conditions for biological preservation in the Balearic Islands. For example, Welker et al. 2014 failed to identify Buxus using a DNA metabarcoding approach on a 4,900 year-old coprolite from Cova de Muleta (Sóller, Mallorca), despite the dominant presence of Buxus pollen. Together, pollen and plant fiber evidence suggests *Myotragus* fed on a toxic plant and therefore would have needed an adaptive mechanism to consume this widely distributed toxic plant on the islands (Yll et al.; 1997; Burjachs et al.; 1994); however, it remains unclear how these goats obtained this tolerance. As many mammalian herbivores employ the gut microbiota to detoxify the dietary toxins(Sasaki et al.; 2005; Kohl et al.; 2014; Shiffman et al.; 2017), we hypothesised that *Myotragus* employed a similar strategy to adapt to the toxic diet. Information on the gut microbiota of past

species can be recovered from well-preserved coprolites (faecal remains) using paleomicrobiological techniques (Santiago-Rodriguez et al.; 2013; Welker et al.; 2014). Previously attempts to reconstruct the microbiome from *M. balearicus* coprolites recovered limited information (Welker et al.; 2014), as retrieving microbiome data from coprolites has a number of difficulties (Warinner et al.; 2015). Coprolites are exposed to environmental conditions for long periods of time and are additionally susceptible to high levels of exogenous contamination. Therefore, authentication of genuine biological signal and minimising the biases caused by contamination are of paramount importance, especially within locations with variable ancient DNA preservation. The characteristic damage of DNA molecules (increased C-to-T substitution towards the end of the molecules) has served as a proxy for authentication of ancient DNA (Briggs et al.; 2007; Jónsson et al.; 2013). However, this method requires mapping data to a single reference genome, which can be problematic for paleomicrobiome data. In addition, several methodologies have emerged to account for and mitigate the impacts of exogenous DNA contamination within paleomicrobiome research (Llamas et al.; 2017; Weyrich et al.; 2017).

Here, we applied shotgun sequencing to eight *M. balearicus* coprolites and recovered robust paleomicrobiolome data to assess any adaptation strategies that may have been employed within the gut of this extinct species. We used stringent laboratory and environmental controls to monitor exogenous DNA, and we authenticated the microbiota within the coprolites utilizing a novel model to assess DNA damage of unmapped DNA sequences. To our knowledge, this work is the first to identify symbiotic microbial adaptation within an extinct species and reveals the power of paleomicrobiome research to reveal how host-microbe relationships can have evolutionary impacts.

3.2 Methods

3.2.1 Sample details

Eight coprolites were unequivocally identified — both in terms of morphology (Alcover et al.; 1999) and age — as belonging to *M. balearicus* from two different Mallorcan deposits (Table S3.1). Two of the coprolites (MbCopro7 and 8) come from the Abric de Son Matge deposit (Valldemossa), where a stratum of coprolites was found and radiocarbon dated at 5,820±360 BP [CSIC-176, 5,489-3,965 calBC 2σ, Waldren, 1992, although this radiocarbon date has been rejected because of high standard deviation (Martínez et al.; 1997)], and 6,680±120 BP [QL-29, 5,837-5,380 calBC 2σ, Waldren, 1992]. The other six coprolites come from Cova Estreta (Pollença, (Encinas; 1997), where five were collected from a surface level of excavation grid M4

(MbCopro1-5) and a -40 cm deep level in excavation grid O8 (MbCopro6)(Alcover et al.; 1999). The surface coprolite level has been radiocarbon dated using a coprolite to 4,950±38 BP [Wk-33010, 3,798-3,650 calBC 2σ](Rivera et al.; 2014), as well as bones from endemic extinct mammals found in contact with the coprolite package (Encinas; 1997): a bone from the Gliridae (Rodentia) Hypnomys morheus (Gliridae, Rodentia) mixed with the coprolites, 6,357±44 BP (Utc-5175, 5,469-5,227 calBC 2σ); and a femur of the bovid *M. balearicus* on top of the coprolites, 5,720±60 BP (Utc-5171, 4,716-4,449 calBC 2σ). Coprolites from the same stratigraphical unit and square have been previously examined for diet and parasite analyses (Alcover et al.; 1999; Bartolomé et al.; n.d.; Bartolomé, Retuerto, Martínez, Alcover, Bover, Cassinello and Baraza; 2011) or for the analysis of stomata density in plant fibers (Rivera et al.; 2014) . The ancient molar used as the environmental control (EnvCtrl) comes from a deposit from where all *Myotragus* remains display unequivocally morphological traits with affinities to *M. kopperi*, the *Myotragus* species recorded in the Mallorcan Early Pleistocene (Bover et al.; 2014).

To evaluate the preservation of the gut microbiome information and the presence of potential DNA contaminants from the environment within the coprolites, we also included shotgun sequencing data of five cattle faeces that were sampled from the Australian State of Victoria for a previous study (CattleFaeces 1-5, original ID in (Ross et al.; 2012): 6803, 6838, 6852, 6859, 7920) (Ross et al.; 2012), as well as shotgun sequencing data of five environmental samples from the MG-RAST metagenomics database (MG-RAST ID: air-mgm4516952; brackish water-mgm4536373; fresh groundwater-mgm4536380; grassland soil-mgm4511193; USA soil-mgm4477876) (Wilke et al.; 2015).

3.2.2 Extraction

All the coprolites were extracted in a dedicated ancient DNA laboratory at the Australian Centre for Ancient DNA following the protocol of Wood et al. (2008). Briefly, about 1 mm of the surface of each coprolite was removed using a surgical blade, and the trimmed coprolites were UV-irradiated for 10 min. Each coprolite was then cut in half using a new surgical blade, and material from the inside of the core was obtained using sterile tweezers until half of a 1.5 mL Eppendorf tube was filled. 0.5-0.6 mL of certified DNA-free, ultrapure water was added to each sample, and the coprolite was rehydrated for 24 hours. Between 100-300 mg of rehydrated coprolite was then extracted using the MoBio Power Soil Kit (MoBIO Laboratories, Inc.) following manufacturer's instructions. Two negative controls using only the water (extraction blank controls; EBCs) were extracted alongside the samples. A second upper molar (EnvCtrl) from a 2.4 My old *Myotragus* individual identified as

Myotragus aff. kopperi, from Cova des Pas de Vallgonera (Llucmajor, Mallorca) was extracted as an environmental control to assess microbial DNA from the cave site (Bover et al.; 2014). This bone was preserved inside the cave and was independent of other mammalian bones. This sample was extracted using a previously published silica-based method (Brotherton et al.; 2013). Briefly, approximately 1 mm of the exterior surface was removed using a Dremel rotary tool with a cutting disc, and bone fragments were powdered using a Braun Mikrodismembrator U (B. Braun Biotech International, Germany) with an 8 mm tungsten ball for 5 seconds at 3,000 revolutions per minute. Up to 260 mg was decalcified and digested overnight at 55 °C on a rotary wheel in 4 mL 0.5 M EDTA (pH 8.0), 200 μL of 10% SDS, and 40 μL of 20 mg/mL Proteinase K. A DNA extraction was performed using a modified QG buffer (QIAGEN) and 100 μL suspended silicon dioxide, as previously described (Brotherton et al. 2010). DNA was then purified using 80% ethanol and eluted in TLE buffer (10 mM Tris, 0.1 mM EDTA ph 8.0) to obtain a final volume of 200 μL.

3.2.3 Library construction, sequencing and data filtering

Double-stranded DNA libraries were constructed for each sample using a protocol based on Meyer and Kircher 2010 with modifications (*e.g.*, using heat to deactivate the Bst enzyme following the adapter fill-in step; see (Llamas et al.; 2016)). Briefly, we uniquely labeled our libraries with a 5-mer P5 barcode (in sample Mb-Copro5 and EnvCtrl) or dual 7-mer P5 and P7 barcodes (the remaining samples). We used Platinum Taq Hifi (Invitrogen) for the post-Bst amplification of all libraries. A single-stranded DNA library (SSL) was constructed for extract of sample MbCopro5, following the protocol of Gansauge and Meyer (2013), in order to explore the ability of this approach to retrieve endogenous microbiome DNA (*i.e.*, host DNA and gut microbiome DNA). While direct improvements in recovery of host DNA were obtained (Table S2.1), no significant differences were observed on the microbial community structure obtained using the two methods (Figure S3A, t-test, p=0.7916). Surprisingly, we observed more contaminant bacteria species from SSL dataset (Figure S2.3B), which inflated alpha diversity differences between the two library preparation methods (Shannon-Weaver index: DSL: 8.604; SSL: 8.609; Simpson's reciprocal index: DSL: 115.586; SSL: 137.565). In the SSL dataset, we find a wide diversity of Acinetobacter species and Halobacteriaceae species not previously detected using DSL methods. Thus, SSL protocol was not applied to other samples, and DSL libraries were utilized for all downstream analysis. Both libraries of MbCopro5 were shotgun sequenced in a HiSeq 2500 RapidRun (2×100, paired-end), and the remaining samples in a NextSeq 500 (2×75, paired-end). The quality of sequencing reads was analyzed with FastQC

v.0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). We filtered reads containing the correct P5 barcode (and P7, for the double-barcoded libraries) using Sabre v.1.0 (http://github.com/najoshi/sabre) allowing one mismatch (option m -1). Adapter sequences were trimmed with AdapterRemoval v.2.1.7 (Schubert et al.; 2016) using the following parameters: mismatch rate 0.1, minimum Phred quality 4, quality base 33, trim ambiguous bases (N) and trim bases with qualities equal or smaller than the given minimum quality. Paired reads with at least 11 bp overlapping were collapsed into a single read. After adapter trimming, reads shorter than 25 bp were discarded before downstream analysis of the remaining reads.

3.2.4 Bioinfomatics analyses

Data mapping and analyses

Collapsed reads from all libraries were mapped to a reference using the BWA v.0.7.13 backtrack algorithm (Li and Durbin; 2009) (with options: -l 1024, -n 0.01, -o 2), removing mapped reads with quality lower than a Phred score 25 using SAMtools v.1.3.1 (Li et al.; 2009). Duplicate reads were filtered using FilterUniqueSAM-Cons.py (Kircher 2012), and mapping results were visualized in Geneious v.7.1.7 (Biomatters, http:/www.geneious.com, (Kearse et al.; 2012)). Mitochondrial genomes were obtained of extant and extinct caprine from Mallorca: *Myotragus balearicus* (unpublished; Bover et al. in preparation, 2019), *Capra hircus* (domestic goat, Accession Number KP231536), and *Ovis aries* (sheep, Accession Number KF938320). Additionally, we mapped same reads to the chloroplast genome from the genera of plant species (or available species available at GenBank) recorded in Mallorca during Late Pleistocene-Holocene : *Buxus microphylla* (Accession Number NC_009599, no *B. balearica* sequence available), *Corylus avellana* (Accession Number KX822768), *Pinus halepensis* (Accession Number JN854197, partial genome) and *Ephedra equisetina* (Accession Number AP010819, as no *E. fragilis* sequence available).

Dietary analysis

To identify the eukaryotic DNA in the shotgun data, the collapsed reads were aligned to NCBI plastid database (updated on December 2017) using DIAMOND (v 0.9.13) (Buchfink et al.; 2015). The generated DAA files were meganized using the daa-meganizer script includedon age in MEGAN6 (parameters used: Weighted-LCA=80%, minimum bitscore=42, minimum E-value=0.01, minimum support percent=0.01). The resulting data were subsequently analysed in MEGAN6 (v 6.11.1) (Huson et al.; 2016).

Taxonomic analysis

The collapsed reads were taxonomically identified using MEGAN Alignment Tool (MALT, v 0.3.8) (Herbig et al.; 2016), which compared the reads against an in-house database that including 47,696 archaeal and bacterial genome assemblies from the NCBI Assembly database (A; 2018). The generated alignment based blast-text files were converted to RMA files using blast2rma script included in MEGAN6 (v 6.11.1) (parameters used: Weighted-LCA=80%, minimum bitscore=42, minimum E-value=0.01, minimum support percent=0.01) (Huson et al.; 2016). The RMA files were analysed in MEGAN6. Species identified from the laboratory control and extraction blank controls were removed from coprolites samples. Alpha diversity was calculated from all taxa using Simpson's and Shannon's inverse indexes implemented in MEGAN6. Samples were normalized to the smallest number of reads of any of the selected samples (n=255,989). PCoA was generated using the Bray-Curtis distance of the species identified in each sample.

Functional analysis

The collapsed reads were functionally identified by using DIAMOND (v 0.9.13) to align the reads against the NCBI nr database (updated on December 2017) (Benson et al.; 2005; Burnham and Anderson; 2004) using a minimum open reading frame of 20 (command: -min-orf 20), as the NCBI database was required >1.5TB of memory to assess using MALT. The generated DIAMOND alignment archive (DAA) files were annotated using the daa-meganizer script included in MEGAN6 (parameters used: Weighted-LCA=80%, minimum bitscore=44, minimum E-value=0.01, minimum support percent=0.01; databases: SEED and COG) (Tatusov et al.; 2000; Overbeek et al.; 2005). All samples were then normalized to the smallest number of DNA sequence in any of the selected samples (n=255,989). For the comparison of the amino acid metabolism that typical to herbivore and carnivore gut microbiomes, the functions (level 3) were extracted from SEED-annotated data in a text format (a csv file) using MEGAN6 (v 6.11.1). Then the amino acid metabolism functions related to the herbivore and carnivore diet (Table S5 of (Muegge et al.; 2011)) were extracted from the csv file using an in-house awk script. In order to compare the functional profiles of the coprolites and cow faeces, the abundance of certain functions of coprolites and cow faeces was compared using script group_significance.py that implemented in QIIME (v 1.9.1) (Caporaso et al.; 2010). The threshold for statistical significance was a p=0.05 and was calculated using goodness of fit (g-test). The null hypothesis is the frequency of any given function is equal across all sample groups. To minimise the exogenous signals introduced by laboratory and environmental contamination, the function data were filtered on two levels. First, functions

(level 3) identified from the laboratory control and extraction blank controls were removed from coprolites. Second, the functional profiles (level 3) of coprolites were compared to that of environmental samples (air, brackish water, fresh groundwater, USA soil, and grassland soil), and any function was removed if the function was not significantly ($p > 0.05$; calculated using goodness of fit) different between two groups.

Authentication of the function data

FASTA files of the reads assigned to certain function groups (Level 1) were extracted and the proportion of A, T, C, and G was calculated from the termini of the reads. We used the proportions of A/T towards the ends of the sequence reads as a proxy to evaluate evidence for significant DNA damage using a two-component, piecewise linear mixed-effects model of the form

$$\log(f_i) = \beta_0 + \beta_1 \times [log(i)]x_i + log(p_C)(1 - x_i)],$$

where i is the position of interest, f_i is the proportion of damage at the ith position, and p_c is the changepoint at which the relationship between $\log(f_i)$ and i becomes constant. We treat the position i as a fixed effect, and the function of the protein as a random effect. All mixed effects models are fit using the lme4 (Bates et al.; 2014) package for the R-statistical software (Team et al.; 2013). We find the value of p_c by fitting the mixed effects model for values of i=1,...,20, and calculating the Bayesian information criterion (BIC) for each model (Burnham and Anderson; 2004). We then selected the value of p_c minimised BIC. If the value of p_c selected was greater than the first position, we simulated 5000 independent datasets under the null model (p_c=1), and recorded the repeated the test to find the empirical distribution of the BIC scores. We used this empirical distribution to identify when p_c>1 was significant.

CYP450 contigs assembly

To identify the potential origin of *CYT450* gene in the samples, the FASTA files containing the reads that assigned to IPR001128 Cytochrome P450 was extracted from DAA files. The extracted reads were assembled using Geneious (v 9.1.5; https://www.geneious.com/ (Kearse et al.; 2012)); assembler with high sensitivity. Contigs assembled from less than 10 reads were discarded. The consensus sequences were generated using the most common basses to minimise the ambiguities. The consensus sequences then compared against NCBI nr database using BLASTX with BLOSUM62 matrix.

Genome assembly and analysis

Iterative mapping

In the absence of a closely related reference sequence, we used an iterative mapping assembly approach to reconstruct the genomic sequence of a Candidatus *Romboutsia* species from coprolite metagenomic data, as implemented by MITObim (v1.8) (Hahn et al.; 2013). Briefly, FASTQ files of collapsed reads were initially mapped against the reference sequence of *R. ilealis* CRIB (GCA_900015215.1) using MIRA (v4.0.2) (Chevreux et al.; 1999, 2004) on accurate genome mapping mode, generating a new reference based on the most conserved regions. Later, reads were iteratively baited and assembled using MitoBIM (v.1.8) with default parameters and a mismatch value of 3%, effectively extending the conserved regions from the initial assembly until it reached a stationary state. The assembled genomes from three samples showed a similar pattern of coverage of the *R. ilealis* reference, with the best samples presenting an average depth of coverage of 12× and a breadth of coverage of 80.5% (Figure 3.4).

Quality assessment of the assembled genomes

The quality of the assembled genomes was assessed using two tools: MapDamage (Jónsson et al.; 2013) was used to assess if the assembled genomes have an ancient origin, and CheckM (Parks et al.; 2015) was used for the assessment of the potential cross-mapping in the assembled genomes. First, the nucleotide misincorporation and fragmentation patterns were calculated using MapDamage2.0 with the default parameters (Jónsson et al.; 2013). Seocond, MitoBIM output files from each sample were then transferred to Geneious (v10.2.3) where a consensus sequence was generated (options: Tools- Generate Consensus Sequence) using three different consensus calling approaches: first, a minimum coverage of 1× using a majority call threshold; second, a minimum coverage of 3× using a base call threshold of 85%; and finally, a minimum coverage of 3× and a base call threshold of 75%. The resulting consensus sequences were assessed using CheckM with default parameters (Parks et al.; 2015). The first approach (Majority, 1×) presented the best results across the three samples, with completeness ranging from 63.85% to 96.83% and presenting little contamination or strain heterogeneity, with exception of the reference reconstructed from sample MbCopro3, which presented minimal contamination (0.47%). As the Majority-1× consensus sequences showed the best quality, we used these sequences for the downstream analyses.

Annotation of the assembled genomes

The consensus sequences were annotated using online RAST (http://rast.nmpdr.org) with ClassicRAST scheme and default parameters (Overbeek et al.; 2005).

Estimation of coverage and GC content of assembled genomes

The coverage of the assemble genomes was calculated using samtools (v 1.3.1) bedcov on a window size of 2500 bp (Li et al.; 2009). The GC content was calculated using a python script GCcalc.py (https://github.com/WenchaoLin/GCcalc/blob/master/GCcalc.py) with a window size of 2500 bp. The GC skew was calculated using a perl script gcSkew.pl (https://github.com/Geo-omics/scripts/blob/master/AssemblyTools/gcSkew.pl) with a window size of 2500 bp. The coverage data of the bam files were exported from Geneious and then fed into an in-house awk script to generate a bed file include the range of covered regions. The results were visualized using Circos (v 0.67-7)(Krzywinski et al.; 2009).

Phylogenetic analysis

The reconstructed genomes of the three ancient *R. ilealis* strains, and genome sequences of *R. ilealis* CRIB (GenBank assembly accession: GCA_900015215.1), *R. sp* MT17 (GCA_900074625.1), *Intestinibacter bartlettii* DSM (GCA_000154445.1), *R. maritimum* (GCA_002251085.1), *R. timonensis* Marseille-P326 (GCA_900106845.1), *R. lituseburensis* DSM 797 (GCA_900103615.1), and *R. weinsteinii* CCRI-19649 (GCF_002250835.1) were aligned using mauveAligner algorithm in Geneious. Local collinear blocks (LCBs) were concatenated with a spacer of 100 Ns. The resulting alignment was then used to build a phylogenetic tree using RAxML in Geneious (parameters used: GTR GAMMA model, 1000 bootstrap replicates) (Stamatakis; 2014). The *I. bartlettii* DSM 16795 was used as an outgroup to root the tree.

3.3 Results

3.3.1 Coprolites originate from *M. balearicus*

We first verified the presence of host (*M. balearicus*) DNA preserved within the coprolites (host DNA was identified in 3 out of a total of 8 coprolites), by mapping the metagenomic data to the *M. balearicus* mitochondrial genome. On average, host DNA recovery from the coprolites was low, as only approximately one in every million reads uniquely mapped to the host mitochondrial genome in three, deeply sequenced samples (MbCopro3, MbCopro5, and MbCopro8). No host DNA was obtained from the remaining five samples, likely due to poor preservation or an insufficient sequencing effort (on average 807,391 reads per sample). The number of mapped sequences also varied depending up on library preparation method. For example, seven sequences in the double-stranded library (DSL) preparation in the deepest sequenced sample (MbCopro5), compared to 31 mapped *M. balearicus* sequences in the single-stranded library (SSL) preparation (Table S3.2). While this suggests varied preservation amongst the ancient coprolites, as expected, we

also identified minimal host sequences from comparative modern fecal material; for example, 3 to 9 of every million sequences mapped to the cattle (*Bos taurus*) mitochondrial genome (Table S3.2). Similarly, no mammalian DNA was obtained by Welker et al. 2014 using *M. balearicus* coprolites from two different deposits. It is likely that more host DNA can be identified from coprolites with increased sequencing effort or DNA enrichment techniques. However, we conclude that the coprolites originated from *M. balearicus* due to the presence of host DNA, in addition to robust chronology, and morphological preservation.

3.3.2 Limited dietary signal preserved within coprolites

To recover the dietary signals of the *Myotragus*, we aligned the coprolite metagenome against the NCBI plastid database. Few sequences (0.1% - 0.5%) could be assigned to potential dietary taxa in both the coprolite or modern cattle gut microbiome (Table S3.2). The majority of taxa identified from coprolite and cattle faeces samples fall within three phyla: Chlorophyta (77 species), Streptophyta (73 species) and Bacillariophyta (20 species). Although several potential dietary taxa were identified, such as Eudicotyledon species and Juniperus species, from coprolites, only a few sequences (1-3) were assigned to each taxon. We next explored the presence of Buxus DNA within the coprolites, as previous pollen analysis suggested that *Myotragus* ate the native, toxic Buxus plant (Welker et al.; 2014). A few sequences from well-preserved samples MbCopro5 (DSL) (n=3) and MbCopro3 (n=1) were uniquely assigned to *B. microphylla* – the only Buxus chloroplast genome currently available in the plastid database (GenBank Accession NC_009599). This results was similar to the mapping success of wheat, a dietary staple of modern cattle (Ross et al.; 2012), in our modern cattle fecal samples; only 7 to 25 sequences were assigned to family Poaceae from each cattle faecal metagenome. Unsurprisingly, the majority of identified eukaryotic species identified within the dietary analysis of coprolites were algae and diatoms, which are likely from environmental contamination or alignment misclassification. Nevertheless, this minimal DNA evidence supports previous findings that *Myotragus* ate a generalist diet that included the toxic Buxus plant.

3.3.3 Coprolites produce robust gut signal

To evaluate the preservation of gut microbiome signal and contamination of environmental DNA, we compared the coprolite microbiome to that of the modern cattle gut microbiome (CattleFaeces1-5), as cattle are the closest ruminants with published shotgun data of the gut microbiome (Ross et al.; 2012). We also compared the coprolites to two laboratory negative controls (EBC1 and EBC2) extracted alongside
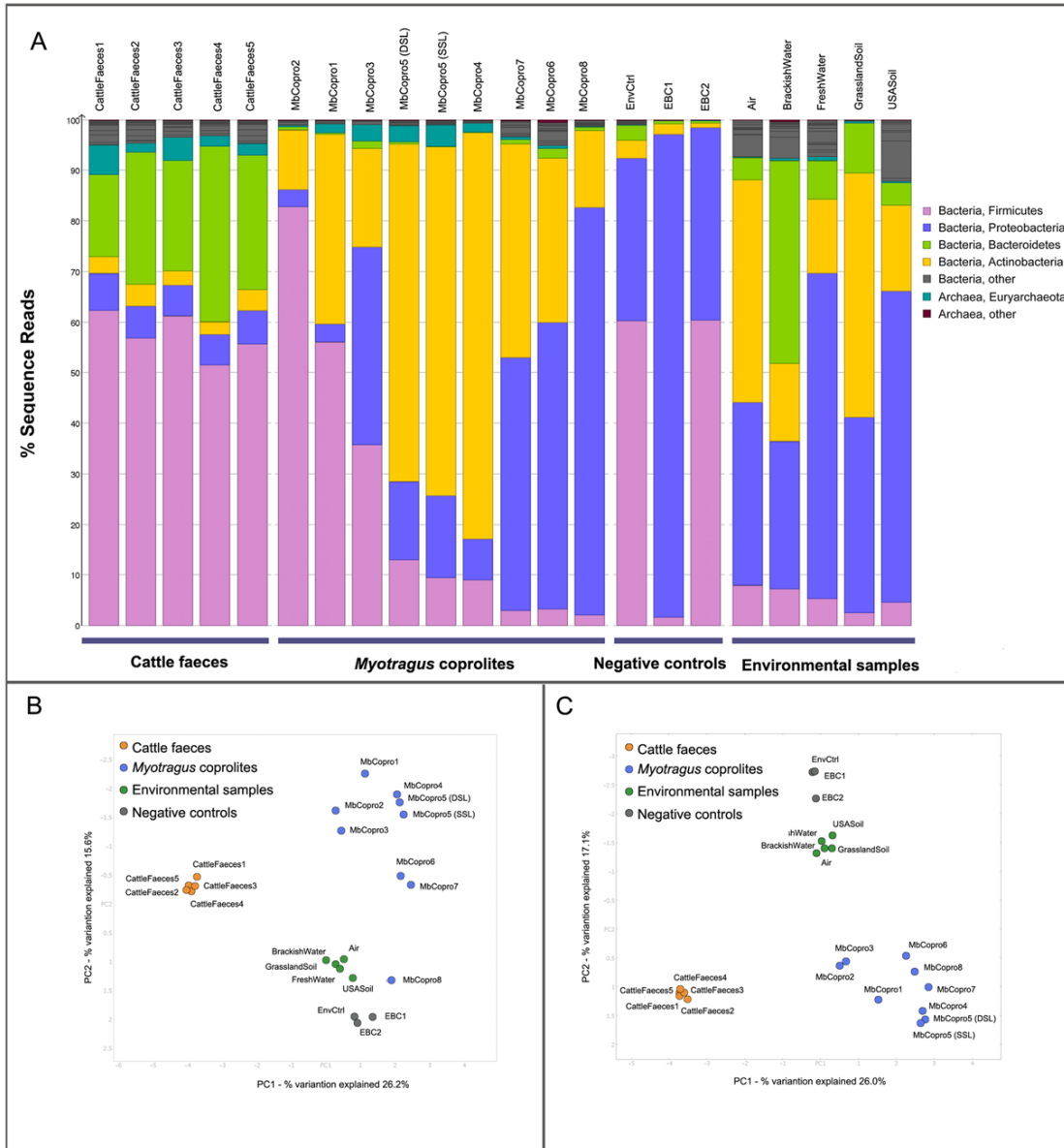
Figure 3.1: Bacteria community composition and the PCoA plot of the filtered and unfiltered data. A. Bacterial community composition at the phyla level of the coprolites, cattle faeces, and controls. B and C. PCoA plot generated using Bray-Curtis distance of the species identified in each sample before (B) and after (C) removing species identified from negative controls.

the coprolites and eight environmental controls, a two million-year-old *Myotragus* molar (EnvCtrl) sampled from Cova des Pas de Vallgornera (Llucmajor, Mallorca; Bover et al., 2014) and five soil and water samples (SoilCtrl) (Table S3.1). As it has been estimated that the upper limit of DNA survival is about one million years (Lindahl et al.; 1993; Willerslev et al.; 2007), the molar sample (EnvCtrl) should not contain any endogenous ancient DNA and thus serves as a control for assessing the environmental signal within the cave system. The microbiome reconstructed from *Myotragus* coprolites contains the three dominant phyla typically observed in extant ruminant gut microbiota: Firmicutes, Proteobacteria, and Actinobacteria (Figure 3.1A) (Ross et al.; 2012; Puniya et al.; 2015). Another typical phyla of modern ruminant microbiota, Bacteroidetes were observed at a limited capacity within coprolites, which can be explained by taphonomic bias that frequently observed from ancient and modern gut microbiome datasets (Hauther et al.; 2015). PCoA analysis of the Bray-Curtis distance shows the *M. balearicus* coprolites clustering away from all environmental samples and laboratory controls, suggesting a microbiome signal that is unique to the *Myotragus* coprolites (Figure 3.1B, Figure 3.1C). Notably, three coprolites (MbCopro 6-8) clustered towards negative controls and environmental samples, indicating poor preservation. To further evaluate the preservation of gut microbiota in the coprolites, we examined the presence of major gut microbiota species shared between modern cattle faeces and *Myotragus* coprolites (Figure 3.2). *Clostridium, Peptostreptococcaceae*, and *Methanobrevibacter* species were dominant within the cattle microbiota and were identified within five coprolite samples (MbCopro1-5). Additional ruminant gut taxa, including Fusobacteriales, Campylobacterales, Enterobacteriaceae, Bifidobacteriaceae, and Bacilli are also detected from these five well-preserved coprolites. In contrast, these ruminant species were not widespread in poorly preserved specimens that clustered near laboratory and environmental controls on a PCoA plot (Figure 3.1C) (MbCopro6-8), highlighting varied preservation amongst *M. balearicus* coprolites. In addition, bacteria that are more likely present in the laboratory or environmental controls (*e.g.*, Rhizobiales, Burkholderiales, Pseudomonadales, and Mycobacterium) were also enriched in these three coprolites, again suggesting minimal preservation of ruminant microbiota. Therefore, these three poorly preserved samples (MbCopro6-8) were only maintained in downstream analysis as an additional control to assess the variation in coprolite preservation. While the preservation of the coprolites was variable, our results show that a strong gut microbiota signal can be preserved within some ancient *M. balearicus* coprolites.
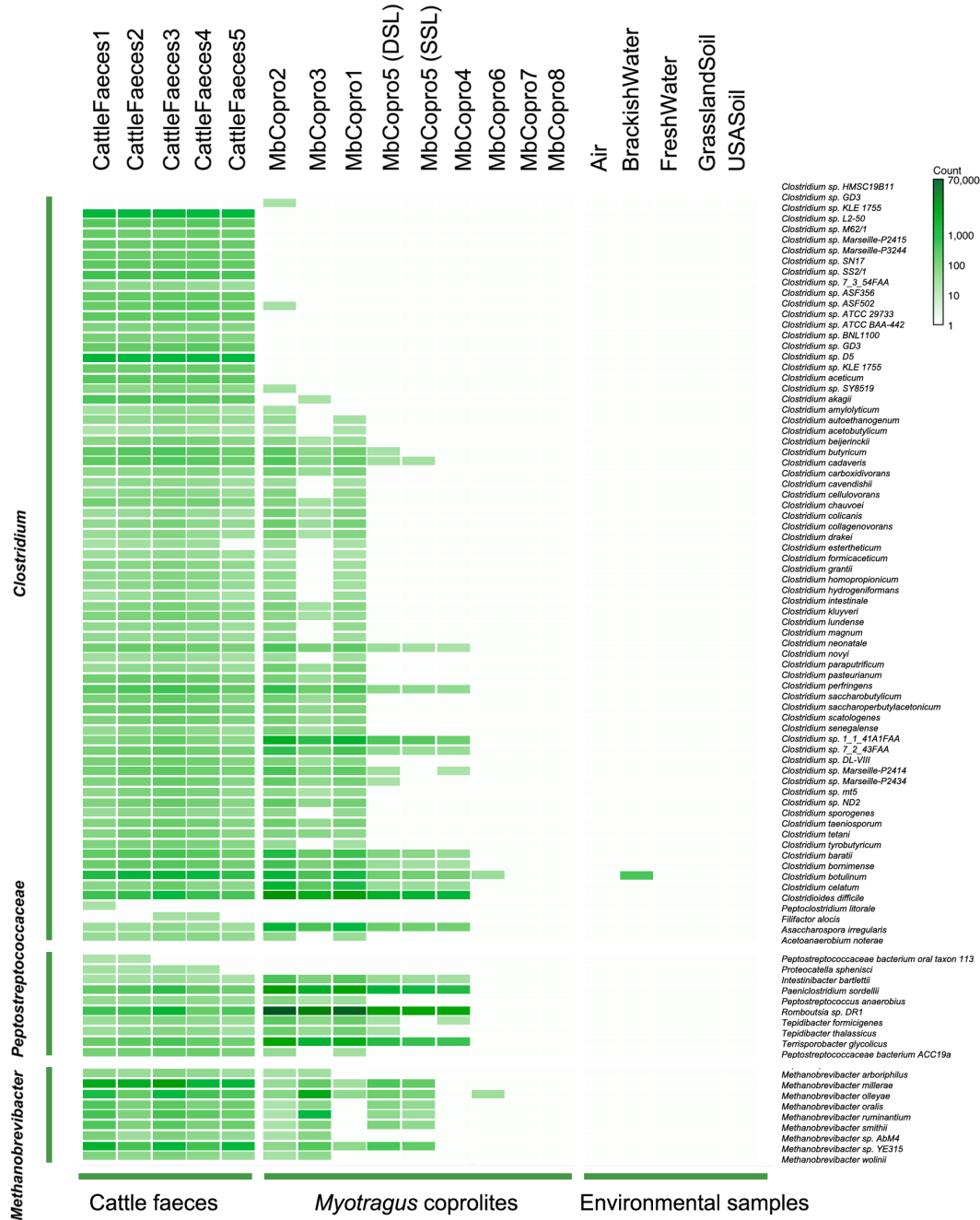
Figure 3.2: Heat map of typical gut bacteria shared between *Myotragus* coprolites and cattle faeces.

3.3.4 Metabolism profiles specific to the *Myotragus* gut microbiome

To investigate the gut microbial metabolism profile of the *Myotragus*, we annotated the gene functions of the coprolite microbiome, cattle faecal microbiome, negative controls and five environmental samples. A total of 6,763 gene functions were identified for all the samples. As coprolites are a mixture of endogenous and exogenous microbes, these gene functions could also originate from laboratory or environmental microbes present within the coprolites. Therefore, we excluded any function identified within laboratory controls (EBCs) and the ancient molar from Mallorca, which should not contain endogenous *Myotragus* gut microbiota. This filtering left 3,537 functions identified in coprolites and modern cattle faeces for further examination. We next compared the functions within coprolites and cattle faeces and identified 151 functions that significantly differed between *Myotragus* coprolites and cow faeces (p<0.05; Figure S3.1). As it is possible that these functions could also be environmental in nature, we additionally removed functions that were similarly abundant (p>0.05) between coprolites and the five environmental soil and water samples, generating a conservative list of 40 gene functions likely endogenous and unique to the *Myotragus* gut microbiome (Figure 3.3A). This includes seven functions involved in amino acids and protein metabolism, six functions involved in carbohydrate metabolism, 10 functions involved in nucleotide and nucleoside metabolism, five functions involved in dormancy and sporulation, four functions involved in regulation and cell signalling, and 13 functions of miscellaneous functions. While most categories can be linked to the unique dietary habits of *Myotragus*, the functional differences in dormancy and sporulation may be due to taphonomic processes that occur during coprolite fossilisation. Thus, we excluded the functions involved in dormancy and sporulation. This analysis identified 45 metabolic activities unique to the *Myotragus* gut microbiome compared to another modern, domesticated ruminant species.

3.3.5 Authentication of function profiling results using ancient DNA damage

Subtractive filtering of functions shared across coprolites and laboratory and environmental controls is a very conservative approach, as many basic microbial functions are conserved across bacterial species, regardless of biological or environmental origin. This approach also leads to a major reduction of data, as only 4.7% of the functions remained for downstream analyses after the subtractive filtering. To tackle this issue, we developed a model based on the damage profiles of ancient DNA, to

authenticate the presence of ancient functional groups independent of their taxonomic origin. We chose six, high-level representative functional groups (level 1) for the authentication, which were present in all samples (coprolites, cattle faeces, and all environmental and laboratory controls). These functional genes include the metabolism of major dietary compounds – proteins, amino acids, fatty acids, and phosphorus – as well as the metabolism of aromatic compounds and stress responses that are likely to be intermixed with environmental signals. First, we assessed damage in the cattle faeces, laboratory controls, and environmental control samples. No significant damage were detected in the metagenomes from cattle faeces, laboratory controls (EBCs), molar tooth, or three environmental controls (air, fresh groundwater, or USA soil sample). However, one and four functional groups examined in two environmental controls (the amino acid metabolism of the brackish water sample, and the metabolism of amino acids, fatty acids, and protein, and stress response of the grassland soil sample) showed signs of DNA damage (Table S3.5), likely resulting from the presence of ancient extracellular DNA. We next examined the damage present in all of the coprolites; we initially examined the poorly preserved *Myotragus* coprolites that were excluded from downstream examination as an addition control for the introduction of exogenous DNA into ancient coprolites. Importantly, all three poorly preserved coprolites (MbCopro6-8) did not show significant signs of ancient DNA damage, consistent with the previous observations (Table S3.5, Figure 3.2). In stark contrast, most functional groups show significant damage in the well-preserved coprolites (MbCopro 1-5) (Table S3.5; $p<=0.001$). There were three minor exceptions; the metabolism of aromatic compounds in MbCopro5 and the metabolism of phosphorus in MbCopro1 and MbCopro5 did now show signs of ancient DNA damage, indicating the presence of some exogenous microbial functions within well preserved coprolites. Overall, this method identified six authenticated ancient microbial functional classes and provides a new way to authenticate functional groups in ancient metagenomic data sets before further downstream analysis.

3.3.6 Amino acid metabolism in *Myotragus* suggests an herbivorous diet

Muegge et al. (2011) identified specific amino acid metabolic functions associated with distinct animal diets: herbivory, carnivory, and omnivory. We examined specific amino acid functions within the authenticated functional groups to describe the dietary habits present in well-preserved *Myotragus* coprolites (*i.e.*, those that contain significant aDNA damage and endogenous gut microbiome; MbCopro 1-5) compared to cattle faeces (Figure 3B). Specific amino acid metabolism associated with herbivore gut microbiomes was detected from coprolites comparable in abundance to that

observed in the cattle gut microbiome. This suggests that the *Myotragus* gut contains genes consistent with a herbivorous diet, as expected. However, there were also several unique observations within *Myotragus* coprolites. Amino acid metabolic gene functions associated with a carnivorous gut microbiome were depleted in coprolites, with two notable exceptions. The abundance proline degradation (EC 1.5.1.12) and a methionine reversible reaction (EC 2.1.1.14) were significantly increased in *Myotragus* coprolites (Bonferroni corrected p<0.001) compared to cattle faeces (Figure 3.3B). This observation potentially indicates of an abundance of proline and methionine in the *Myotragus* diet relative to modern cattle.

3.3.7 Functional profiling highlights a detoxification role in the *Myotragus* gut microbiome

We next examined specific functions within the two authenticated functional groups (stress responses and metabolism of aromatic compounds) of the gut microbiome associated with the tolerance of plant toxins. The gut microbiomes of living herbivores (such as *N. lepida*) that are capable of digesting toxic plants are enriched in genes within these functional groups (Kohl et al.; 2014). To explore if the gut microbiota facilitated the ability of *Myotragus* to consume toxic plants, we examined the presence of these two authenticated functional groups within the well-preserved coprolites (MbCopro 1-5). The mean abundance of genes linked to metabolism of aromatic compounds was increased by 254.37% (t-test; p=0.062) in the *Myotragus* coprolites compared to cattle, while the mean abundance of genes linked to stress responses was increased by 7.2% (t-test, p = 0.106) (Figure 3.3C, Figure 3.3D), although insignificantly. Within these functional groups, 16 specific functions involved in the metabolism of aromatic compounds and 14 functions involved in stress response were significantly increased in *Myotragus* coprolites compared to modern cattle faeces, suggesting that *Myotragus* may have used these metabolic functions to digest plant toxins. Within the detoxification functional group, gene *CYP450* is a likely candidate to contribute to the degradation of toxic compounds in the ruminant microbiome (Plessis-Rosloniec and Zofia 2011). Therefore, we explored the presence of *CYP450* in the well-preserved coprolites compared to cattle faeces using the normalised abundance of sequences assigned to *CYP450* gene and the origins of the assembled *CYP450* contigs. *CYP450* abundance was significantly higher in coprolites compared to modern cattle faeces (Bonferroni corrected p=0.005), suggesting that this function may have been enriched in the *Myotragus* gut. We additionally mapped sequences to the *CYP450* gene from all the samples and assembled contigs to identify the gut microbiota species that maintained this gene. Within the three most deeply sequenced coprolites (MbCopro1-3), two, three and five *CYP450*
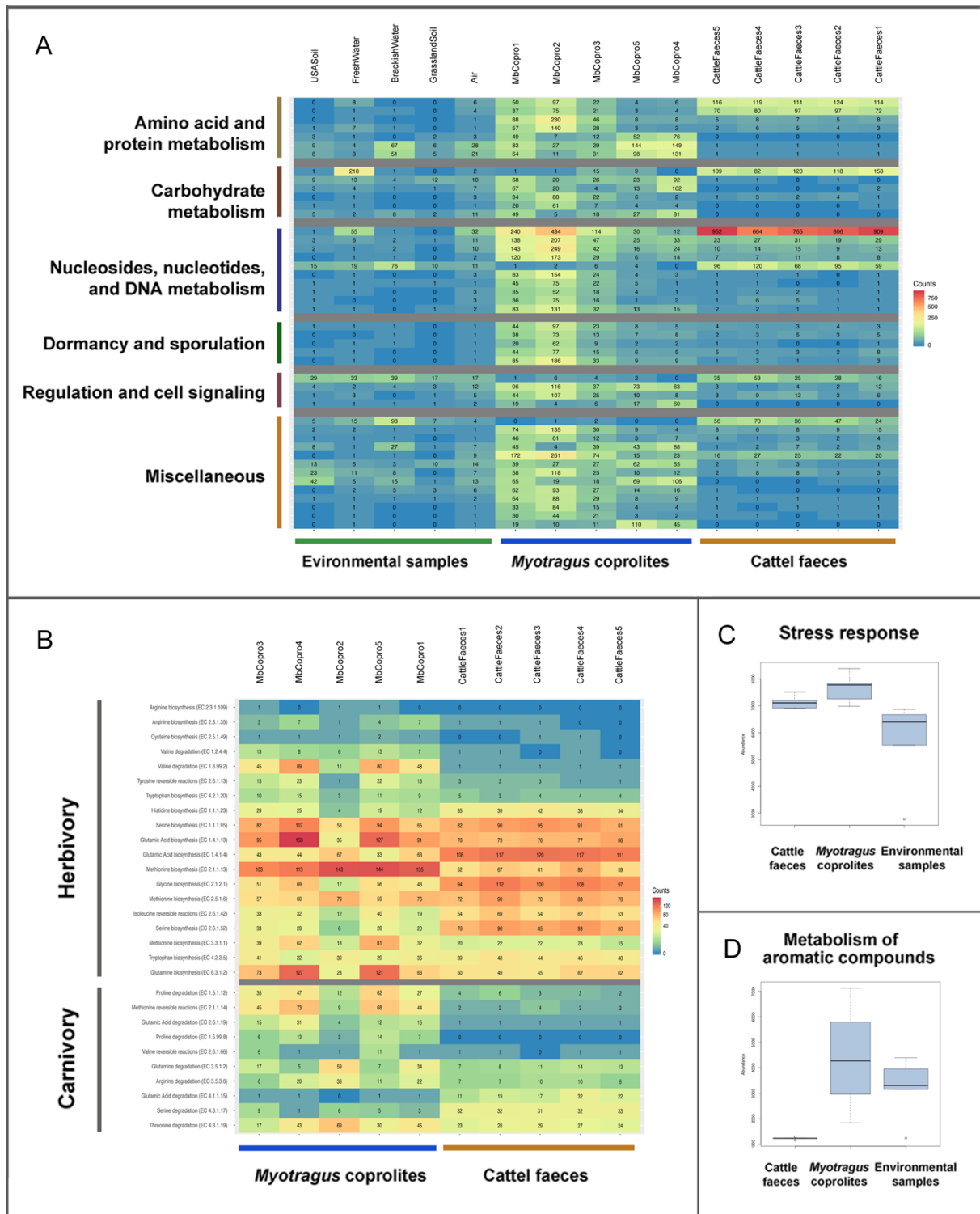
Figure 3.3: Functional profiles of the *Myotragus* coprolite microbiomes. A. Functions specific to the *Myotragus* gut microbiome. B. Amino acids metabolism functions that specific to a carnivores and herbivores. C and D. Comparison of the abundance of the detoxification functional group in the *Myotragus* coprolites and cattle faeces (C: stress response; D: metabolism of aromatic compounds).

contigs with a length from 301 bp to 1272 bp were assembled, respectively (Table S3.5) and identified via BLAST. The majority of contigs originate from species within the typical gut microbiome genera *Clostridium* and *Clostridioides*, as well as *Pseudomonas*, *Streptomyces*, and *Nocardiopsis*. Three and four *CYP450* contigs were assembled from two other well preserved coprolites (MbCopro 4-5) and identified as *Micromonospora, Mycobacterium, Kitasatospora, Lentzea, Amycolatopsis*, and *Kibdelosporangium*, respectively. While most of these taxa were also present in the cattle faeces, *Lentzea* and *Kibdelosporangium* are commonly found in the environment, thus their origin is unclear; a more shallow sequencing depth of these three samples also hindered the ability to assemble long contigs. Lastly, no *CYP450* contigs could be assembled from poorly preserved coprolites (MbCopro6-8), as well as cattle faecal samples, laboratory controls, or the molar tooth specimen. Only a few *CYP450* contigs could be assembled from other environmental samples, and all the contigs assembled from environmental samples were aligned to typical environmental microbes (*e.g., Kocuria, Nevskia*). This finding suggests that the *CYP450* contigs in coprolites originate from the *Myotragus* gut microbiome and therefore may be an important adaptation mechanism to consume Buxus toxins.

3.3.8 Potential co-evolutionary relationship between mammals and gut symbionts

This study is one of the few to examine ancient non-human paleofaeces and represents an excellent opportunity to examine the long-term co-evolutionary relationship of gut microbial species in non-human mammals. To explore this further, we assembled draft genome sequences of the most abundant species identified in *Myotragus* coprolites and cattle faeces – *R. ilealis*. *R. ilealis* was in high abundance (12%-50%) in well-preserved coprolite samples (MbCopro 1-3). Draft genomes of ancient *R. ilealis* strains were reconstructed from these coprolites using *R. ilealis* CRIB genome as reference genome (GeneBank ID: LN555523) (Figure 3.4A, Table S10). Ancient DNA damage signatures (5′-C-to-T and 3′-G-to-A substitution and fragmentation) were consistent with samples of this age (Briggs, Stenzel et al. 2007) in each sample (Table S3.6). No contamination or strain heterogeneity was detected from two genomes (assembled from MbCopro1 and 2) using CheckM, but minimal contamination (0.47%) was detected from one genome (assembled from MbCopro3) (Table S3.7). No reads could be assembled to the reference plasmid genome, which may suggest that these ancient strains did not maintain this plasmid. We then annotated the most complete ancient *R. ilealis* genome using RAST, and a total of 1,757 features were identified. The ancient and modern *R. ilealis* strains show similar metabolic profiles (Figure S3.2) and include functions linked to stress re-

sponse and reduction in the inflammatory response of animal gut (*e.g.*, enolase, NAD-dependent glyceraldehyde-3-phosphate dehydrogenase, elongation factor Tu, heat shock protein GrpE, and chaperone protein DnaK). These functions may be linked to the use of this species as a probiotic to improve gut function (Gerritsen; 2015; Siciliano and Mazzeo; 2012).

A *Romboutsia* phylogeny was constructed from the alignment of the whole genome in RAxML with 1,000 bootstrap replicates (Figure 3.4B). The ancient *R. ilealis* strains clustered together with *Romboutsia* species isolated from the animal gut rather than the environment, suggesting these species originate from the *Myotragus* gut microbiota (Gerritsen; 2015; Gerritsen et al.; 2017; Maheux et al.; 2017). Interestingly, the topology of the *Romboutsia* species phylogenetic tree also mirrors that of their hosts (*e.g.*, ungulates, rodents, and primates). This may reflect an extremely deep co-evolutionary history of *Romboutsia* strains and their mammalian hosts over the past >100 million of years. While available genomes of animal-origin *Romboutsia* species are limited, this provides the first evidence of long-term co-evolutionary history between mammals and this microbe, which likely represents a mutually beneficial relationship.

3.4 Discussion

3.4.1 Paleomicrobiology offers a window into the diet and behaviour of extinct animals

Using *M. balearicus* as a model species, our results demonstrate the power of paleomicrobiology to reveal the diet and behaviour of past mammals. First, we were able to interrogate the dietary behaviours of an extinct species. Although previous studies obtained dietary information from coprolites using targeted amplification of marker genes (*e.g.*, chloroplast rbcL gene and 18s rRNA) (Welker et al.; 2014; Boast et al.; 2018), our results suggest that minimal eukaryotic DNA can also be obtained using shotgun metagenomic approaches. Alternatively, bacterial DNA predominated the endogenous DNA signal from the coprolites and can also be utilized as a generalized signal for extinct species diets. Functional analysis of amino acid metabolism preserved within *Myotragus* coprolites suggested a typical herbivore diet (Muegge et al.; 2011), confirming previous morphological assessments of pollen and plant fibres. While extensive morphological analysis has already been conducted on *Myotragus* coprolites, functional analysis of gut microbial metabolism may provide more insight into newly discovered or lesser studied species. Specific dietary information may also be obtained for assessing microbial functions. For example, the increased gene abundance involved in proline degradation might relate to the
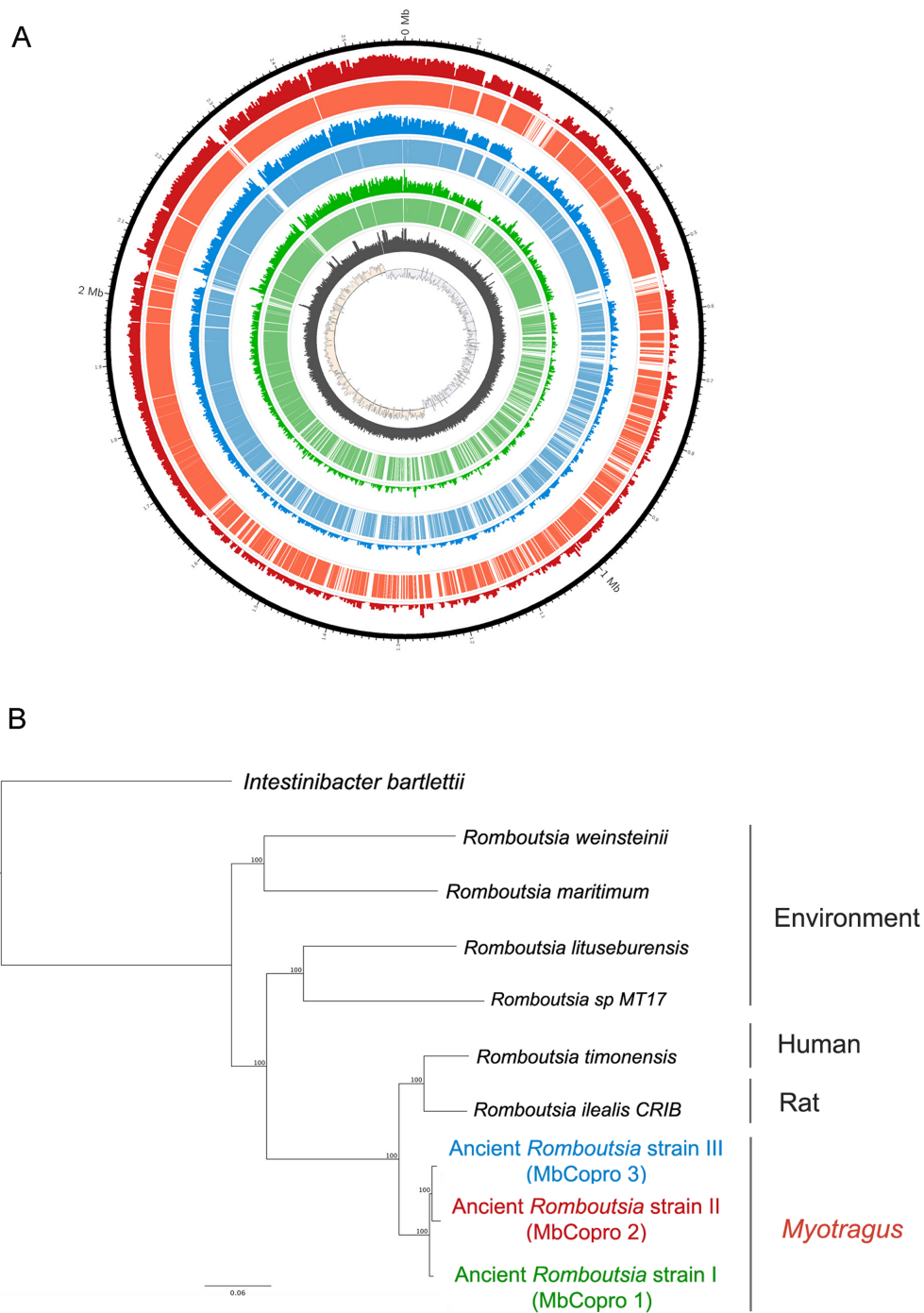
Figure 3.4: Draft genome and phylogeny of three 5000-year-old *R. ilealis* strains. A. The assembled ancient *Romboutsia* genomes. From inside-out: GC skew (light grey<0, light pink>0) of the reference genome *R. ilealis*, GC content of of the reference genome *R. ilealis*, the coverage and depth of the ancient *R. ilealis* strain 1 (green), 2 (red) and 3 (blue). B The *Romboutsia* phylogeny constructed from the alignment of the whole genome in RAxML with 1000 bootstrap replicates.

osteophagic behaviour of the *Myotragus* (Ramis and Bover; 2001), as collagen is one of the most abundant sources of proline and hydroxyproline maintained in bone collagen (90% of the total organic material in bone is collagen) (Young; 2003; Wu et al.; 2011). However, we currently have a limited understanding of the gut microbiomes of osteophagic animals, and further research will need to be done to assess how this behaviour may be linked to gut microbiome function. Additionally, this signal could stem from other proline-rich sources in the *Myotragus* diet, but this is less likely because there is currently little evidence to suggest this may be the case (Alcover et al.; 1999; Welker et al.; 2014). This observation could also be related to the increase in proline-rich proteins present in the saliva of browser animals. Browsers synthesize large quantities of proline-rich saliva proteins to neutralise the toxic components (*e.g.*, tannins) in their diet (McDougall; 1948; Austin et al.; 1989) , which may have an impact on their functions identified within their gut microbiome. Further research is needed in the examining microbial gut functions in extant species to improve our understanding in the past.

3.4.2 Coprolites reveal the role of the gut microbiome in animal adaptation to environment

We found strong evidence from coprolites that the gut microbiota of *M. balearicus* is specialised for detoxification and probiotic function, which might be a key in their adaptation to consume the local vegetation. We found genes involved in stress response and metabolism of aromatic compounds were enriched in *M. balearicus* coprolites. These same functional groups are enriched in other herbivore gut microbiomes that digest plant toxins (Kohl et al.; 2014). We also detected an increased abundance of the *CYP450* gene in coprolite microbiomes. *CYP450* is a highly active enzyme involved in detoxification. It can initiate the degradation of toxic steroidal compounds in Rhodococcus species (Plessis-Rosloniec and Zofia; 2011), which are similar to the toxins identified from the Buxus plant in Mallorca (Ata and Andersh; 2008). Although it is possible that abundance of detoxification functional groups derives from environmental DNA, our novel authentication approach of ancient DNA damage profiles suggests that this function has an ancient biological origin. Additionally, BLAST results of assembled *CYP450* contigs indicate that this function was maintained by microorganisms known to preferentially live in animal guts. It is also possible that the enrichment of these genes in coprolites compared to cattle faeces resulted from the difference between the caprine and bovine gut microbiome. However, this possibility seems unlikely, as the majority of the bacterial community and functional genes were conserved between these two ruminants (Kohl et al.; 2014).

The presence of *R. ilealis* in the coprolites may also suggest that *Myotragus* relied on its gut microbiome to consume toxic plants. We observed high abundances of *R. ilealis* in the coprolite microbiome (MbCopro 1-5; 2% - 50% of the total coprolite microbiome). *R. ilealis* is a probiotic bacterium that can improve the gut function, as the increased the abundance of *R. ilealis* is associated with the reduction of the innate inflammatory response (Gerritsen; 2015). Buxus toxins can cause severe, innate inflammatory responses (Van Soest et al.; 1965), indicating that *Myotragus* would have needed to resist inflammatory responses to the toxin in addition to degrading the toxin. A probiotic bacteria, such as *R. ilealis*, that can suppress inflammatory responses could help reduce the health issues associated with consuming Buxus. However, the abundance of *R. ilealis* required to have a beneficial effect is unknown, and with taphonomic processes at play in coprolites, the abundance and activity of this bacterium remains unknown. A study using mice models showed that the relative abundance of *R. ilealis* can vary drastically in gut microbiota (from 0% to 20% of the total gut microbiota), especially when it plays a probiotic role (Gerritsen; 2015). Nevertheless, it may be possible that *R. ilealis* represents a genuine, anti-inflammatory feature of the *M. balearicus* gut microbiome. Together, our results suggest that *Myotragus* was equipped with a gut microbiota adapted for the consumption of the toxic plant B. balearica. To our knowledge, this study provides the first evidence of how gut microbiota facilitated the adaptation to the diet of an extinct animal.

3.4.3 Ancient bacterial genomes likely reveal a deep co-evolutionary history between the gut microbiota and mammals

Co-evolution events have been identified between the gut microbes and mammals (Ley et al.; 2008; Moeller et al.; 2016). Sometimes the co-evolutionary signal is so strong that the evolutionary history of the gut microorganism is parallel to that of their hosts (Comas et al.; 2013). We speculated that the mirrored phylogeny of *Romboutsia* species and their hosts might represent an extremely deep co-evolutionary relationship (>100 million of years) between this bacterium and mammalian species. However, the current resolution of this relationship is low due to the limited availability of *Romboutsia* genomes. Additional *Romboutsia* genomes from other mammals are required to construct a detailed phylogeny. In addition, direct evidence of vertical transmission of *Romboustia* species is necessary to confirm its ability to maintain a strong and unbiased co-evolutionary signal. Vertical transmission of *Romboutsia* species currently remains unexplored, although it has been shown for other probiotic gut species (*e.g. Lactobacillus* and *Bifidobacterium*) (Matsumiya et al.; 2002;

Milani et al.; 2015). It is also necessary to investigate the potential horizontal transfer events of this species, as it might confound the co-evolutionary history (Moeller et al.; 2016) . Nevertheless, this study illustrates the potential of using ancient bacterial genomes to reveal deep co-evolutionary patterns that may link to adaptation over long time-scales. We also highlight the necessity of advancing the gut microbiome research on modern mammals, as it will pave the way for revealing important insights into the past evolutionary history.

3.4.4 New insights into biological and environmental signals in paleomicrobiome data sets

Gaining biological information from coprolites is extremely difficult (Warinner et al.; 2017). This study took a series of stringent and novel measures to obtain a robust biological signal from coprolites and provided new insights into paleomicrobiological studies. We show that the bias from non-biological sources can be minimised through a number of precautions, such as (1) using negative controls to monitor the contamination that potentially introduced by laboratory work; (2) including environmental samples throughout the analysis to monitor exogenous signals; (3) excluding all the potential contaminant species from the analysis to isolate the biological signal; (4) using damage profiles to authenticate ancient DNA; and (5) assessing the quality and phylogenetic history of assembled ancient genomes. It is critical to take multiple measures to monitor the potential bias, as the contamination can originate from different sources (*e.g.*, sampling location and laboratories).

Authentication of ancient DNA is typically done on assembled or mapped genome sequences. Currently, authentication of paleo-microbiome data largely depends on the ancient DNA damage profiles of a few bacteria species within the dataset (Jónsson et al.; 2013). This method can only be successfully applied to bacteria with nearly complete reference genomes, which is problematic for most paleomicrobiome datasets that may not have many sequences assigned to a specific species. Moreover, over 1,000 sequences from each species are required to obtain robust authentication (Warinner et al.; 2017), and therefore only a few bacteria can be assessed in any given microbiome, representing a small proportion of any given paleomicrobiome data set. To address these issues, we developed a new model for reference-free authentication of paleomicrobiome data. We applied this method to the sequences identified in particular functions that span a range of different microbial species. Significant DNA damage (p<0.001) was detected from coprolites that preserved strong gut microbiome signal, while no significant DNA damage was detected from coprolites with poor gut microbiome signal (MbCopro6-8) or from modern cattle faecal samples. Interestingly, a coprolite from the same deposit as

MbCopro7-8 also did not yield any positive, identifiable ancient DNA in a previous study examining coprolites for fibers, seeds or pollen (Welker et al.; 2014), further supporting the successful use of this authentication approach. The concordant conclusion about the preservation of the biological signal between the taxonomy analysis and the damage profiles, alongside modern comparisons, suggest this method will accurate detect ancient DNA damage.

Significant levels of DNA damage were also detected in three functional groups in two independent environmental samples. This is likely due to the environmental samples containing high proportions of degraded exogenous DNA, as has been obtained in several recent ancient DNA studies (Weiß et al.; 2019). While it is possible that the DNA damage detected in the coprolites is also from the environment, it is highly unlikely given the lack of signal in poorly preserved coprolites. If the DNA decay in well-preserved coprolites resulted from exogenous environmental DNA, consistent damage should be observed from all the functional groups of all the samples from the Balearic Islands (all eight *M. balearicus* coprolites and the molar tooth used as environmental control), regardless of the preservation of biological signals. However, only coprolites containing robust gut microbiome signals show this DNA decay pattern, suggesting this signal is from an ancient biological source rather than the environment. The damage observed for functional groups was also representative of the damage observed using a reference-based approach for *R. ilealis*, further suggesting that reference free predictions of DNA damage will be useful for future ancient microbiome research. Nevertheless, the observation that some environmental samples possess a mixture of damaged and undamaged DNA signatures highlights the necessity of proper controls and in-depth investigations into microbial authentication in ancient samples.

Finally, we tested the ability of a SSL protocol to recover paleomicrobiome information compared to DSL methods. SSL has been shown to increase the yield of DNA fragments, especially the small ones, in paleo-genetic studies (Gansauge and Meyer; 2013). We applied SSL protocol to one coprolite DNA extract and explored its potential in the paleo-microbiological study (Figure S3.1). Surprisingly, we observed more laboratory contamination in the SSL library approach, indicating that the SSL methodologies need further testing to determine their biases and artefacts in paleomicrobiome research. However, the sample size of SSL method (n=1) is limited and further tests are necessary to confirm our observation. Overall, our results represent an exemplary framework for mining authentic paleomicrobiological data from ancient coprolites, while minimising the potential confounding signals.

3.5   Conclusion

In summary, our paleo-microbiome analysis of coprolites offers new insights into the diet and behaviour of an extinct mammal, as well as the co-evolutionary relationships between the gut microbiota and hosts. We identified a potential case in which the gut microbiota directly assisted in the adaptation of its mammal-host to an environment. We demonstrated that robust paleomicrobiome information can be obtained from ancient faecal remains when a combination of multidisciplinary approaches are applied to assess contamination and ancient DNA authentication. Although we tried to be as thorough as possible, the dataset could be further explored and additional insights might be yielded through deeper sequencing, better-curated reference databases and bioinformatics tools further tailored for the identification of short DNA fragments. Nevertheless, this study provides a basic guideline and a transferable toolset for future paleomicrobiome studies of mammalian coprolites, demonstrating the power of paleomicrobiology in reconstructing ancient microbiomes.

Bibliography

A, E. R. (2018). *New and Refined Tools and Guidelines to Expand the Scope and Improve the Reproducibility of Palaeomicrobiological Research*, PhD thesis, School of Biological Sciences, The University of Adelaide. An optional note.

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L. and Gilbert, M. T. P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation?, *Trends in ecology & evolution* **31**(9): 689–699.

Alcover, J. A. (1981). Les quimeres del passat. els vertebrats fossils del plio-quaternari de les balears i pitiuses, *Monografies Cientifiques, Edit. Moll* **1**: 1–260.

Alcover, J. A., Perez-Obiol, R., YLL, E.-I. and Bover, P. (1999). The diet of myotragus balearicus bate 1909 (artiodactyla: Caprinae), an extinct bovid from the balearic islands: evidence from coprolites, *Biological Journal of the Linnean Society* **66**(1): 57–74.

Ata, A. and Andersh, B. J. (2008). Buxus steroidal alkaloids: chemistry and biology, *The Alkaloids: Chemistry and Biology* **66**: 191–213.

Austin, P. J., Suchar, L. A., Robbins, C. T. and Hagerman, A. E. (1989). Tannin-binding proteins in saliva of deer and their absence in saliva of sheep and cattle, *Journal of Chemical Ecology* **15**(4): 1335–1347.

Bartolomé, J., Cassinello, J. and Baraza Ruíz, E. (n.d.). Sobre la dieta de myotragus balearicus, un bóvido sumamente modificado del pleistoceno-holoceno de las baleares.

Bartolomé, J., Plaixats, J., Piedrafita, J., Fina, M., Adrobau, E., Aixàs, A., Bonet, M., Grau, J. and Polo, L. (2011). Foraging behavior of alberes cattle in a mediterranean forest ecosystem, *Rangeland ecology & management* **64**(3): 319–324.

Bartolomé, J., Retuerto, C., Martínez, X., Alcover, J. A., Bover, P., Cassinello, J. and Baraza, E. (2011). Consumo de boj balear (buxus balearica lam. 1785) por el extinto myotragus balearicus bate 1909, *LÓPEZ, C.; RODRÍGUEZ, MP; SAN MIGUEL, A* pp. 491–495.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2014). Fitting linear mixed-effects models using lme4, *arXiv preprint arXiv:1406.5823* .

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Wheeler, D. L. (2005). Genbank, *Nucleic acids research* **33**(suppl_1): D34–D38.

Boast, A. P., Weyrich, L. S., Wood, J. R., Metcalf, J. L., Knight, R. and Cooper, A. (2018). Coprolites reveal ecological interactions lost with the extinction of new zealand birds, *Proceedings of the National Academy of Sciences* **115**(7): 1546–1551.

Bover, P. and Alcover, J. A. (1999). The evolution and ontogeny of the dentition of myotragus balearicus bate, 1909 (artiodactyla, caprinae): evidence from new fossil data, *Biological Journal of the Linnean Society* **68**(3): 401–428.

Bover, P. and Tolosa, F. (2005). The olfactory ability of myotragus balearicus: preliminary notes, *Proceedings of the International Symposium" Insular Vertebrate Evolution: the Palaeontological Approach": September, 16-19 Mallorca*, Societat d'Història Natural de les Balears, pp. 85–94.

Bover, P., Valenzuela, A., Guerra, C., Rofes, J., Alcover, J. A., Ginés, J., Fornós, J. J., Cuenca-Bescós, G. and Merino, A. (2014). The cova des pas de vallgornera (llucmajor, mallorca): a singular deposit bearing an exceptional well preserved early pleistocene vertebrate fauna, *International Journal of Speleology* **43**(2): 6.

Bover, P., Valenzuela, A., Torres, E., Cooper, A., Pons, J. and Alcover, J. A. (2016). Closing the gap: New data on the last documented myotragus and the first human evidence on mallorca (balearic islands, western mediterranean sea), *The Holocene* **26**(11): 1887–1891.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. et al. (2007). Patterns of damage in genomic dna sequences from a neandertal, *Proceedings of the National Academy of Sciences* **104**(37): 14616–14621.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adler, C. J., Richards, S. M., Der Sarkissian, C., Ganslmeier, R., Friederich, S. et al. (2013). Neolithic mitochondrial haplogroup h genomes and the genetic origins of europeans, *Nature communications* **4**: 1764.

Buchfink, B., Xie, C. and Huson, D. H. (2015). Fast and sensitive protein alignment using diamond, *Nature methods* **12**(1): 59.

Burjachs, F., Pérez-Obiol, R., Roure, J. and Julià, R. (1994). Dinámica de la vegetación durante el holoceno en la isla de mallorca, *Trabajos de Palinología básica y aplicada* pp. 199–210.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection, *Sociological methods & research* **33**(2): 261–304.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I. et al. (2010). Qiime allows analysis of high-throughput community sequencing data, *Nature methods* **7**(5): 335.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E., Wetter, T. and Suhai, S. (2004). Using the miraest assembler for reliable and automated mrna transcript assembly and snp detection in sequenced ests, *Genome research* **14**(6): 1147–1159.

Chevreux, B., Wetter, T., Suhai, S. et al. (1999). Genome sequence assembly using trace signals and additional sequence information., *German conference on bioinformatics*, Vol. 99, Citeseer, pp. 45–56.

Cho, I. and Blaser, M. J. (2012). The human microbiome: at the interface of health and disease, *Nature Reviews Genetics* **13**(4): 260.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G. et al. (2013). Out-of-africa migration and neolithic coexpansion of mycobacterium tuberculosis with modern humans, *Nature genetics* **45**(10): 1176.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A. et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome, *Nature* **505**(7484): 559.

Encinas, J. A. (1997). El jaciment fossilífer de la cova estreta (pollença, mallorca), *Endins: publicació d'espeleologia* (21): 83–92.

Gansauge, M.-T. and Meyer, M. (2013). Single-stranded dna library preparation for the sequencing of ancient or damaged dna, *Nature protocols* **8**(4): 737.

Gerritsen, J. (2015). *The genus Romboutsia: genomic and functional characterization of novel bacteria dedicated to life in the intestinal tract*, Wageningen University.

Gerritsen, J., Hornung, B., Renckens, B., van Hijum, S. A., dos Santos, V. A. M., Rijkers, G. T., Schaap, P. J., de Vos, W. M. and Smidt, H. (2017). Genomic and functional analysis of romboutsia ilealis cribt reveals adaptation to the small intestine, *PeerJ* **5**: e3698.

Hahn, C., Bachmann, L. and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach, *Nucleic acids research* **41**(13): e129–e129.

Hauther, K. A., Cobaugh, K. L., Jantz, L. M., Sparer, T. E. and DeBruyn, J. M. (2015). Estimating time since death from postmortem human gut microbial communities, *Journal of forensic sciences* **60**(5): 1234–1240.

Herbig, A., Maixner, F., Bos, K. I., Zink, A., Krause, J. and Huson, D. H. (2016). Malt: Fast alignment and analysis of metagenomic dna sequence data applied to the tyrolean iceman, *BioRxiv* p. 050559.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J. and Tappu, R. (2016). Megan community edition-interactive exploration and analysis of large-scale microbiome sequencing data, *PLoS computational biology* **12**(6): e1004957.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S. et al. (2012). Structure, function and diversity of the healthy human microbiome, *nature* **486**(7402): 207.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. and Orlando, L. (2013). mapdamage2. 0: fast approximate bayesian estimates of ancient dna damage parameters, *Bioinformatics* **29**(13): 1682–1684.

Jordana, X. and Köhler, M. (2011). Enamel microstructure in the fossil bovid myotragus balearicus (majorca, spain): implications for life-history evolution of dwarf mammals in insular ecosystems, *Palaeogeography, Palaeoclimatology, Palaeoecology* **300**(1-4): 59–66.

Jordana, X., Marín-Moratalla, N., DeMiguel, D., Kaiser, T. M. and Köhler, M. (2012). Evidence of correlated evolution of hypsodonty and exceptional longevity in endemic insular mammals, *Proceedings of the Royal Society B: Biological Sciences* **279**(1741): 3339–3346.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics* **28**(12): 1647–1649.

Kinross, J. M., Darzi, A. W. and Nicholson, J. K. (2011). Gut microbiome-host interactions in health and disease, *Genome medicine* **3**(3): 14.

Kohl, K. D., Weiss, R. B., Cox, J., Dale, C. and Denise Dearing, M. (2014). Gut microbes of mammalian herbivores facilitate intake of plant toxins, *Ecology letters* **17**(10): 1238–1246.

Köhler, M. and Moyà-Solà, S. (2004). Reduction of brain and sense organs in the fossil insular bovid myotragus, *Brain, Behavior and Evolution* **63**(3): 125–140.

Köhler, M. and Moyà-Solà, S. (2009). Physiological and life history strategies of a fossil large mammal in a resource-limited environment, *Proceedings of the National Academy of Sciences* **106**(48): 20354–20358.

Krijgsman, W., Hilgen, F., Raffi, I., Sierro, F. J. and Wilson, D. (1999). Chronology, causes and progression of the messinian salinity crisis, *Nature* **400**(6745): 652.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics, *Genome research* **19**(9): 1639–1645.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R. et al. (2008). Evolution of mammals and their gut microbes, *Science* **320**(5883): 1647–1651.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform, *bioinformatics* **25**(14): 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). The sequence alignment/map format and samtools, *Bioinformatics* **25**(16): 2078–2079.

Li, Y., Caufield, P., Dasanayake, A., Wiener, H. and Vermund, S. (2005). Mode of delivery and other maternal factors influence the acquisition of streptococcus mutans in infants, *Journal of dental research* **84**(9): 806–811.

Lindahl, T. et al. (1993). Instability and decay of the primary structure of dna, *nature* **362**(6422): 709–715.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A. et al. (2016). Ancient mitochondrial dna provides high-resolution time scale of the peopling of the americas, *Science advances* **2**(4): e1501385.

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A. and Haak, W. (2017). From the field to the laboratory: Controlling dna contamination in human ancient dna research in the high-throughput sequencing era, *STAR: Science & Technology of Archaeological Research* **3**(1): 1–14.

Maheux, A. F., Boudreau, D. K., Bérubé, È., Boissinot, M., Cantin, P., Raymond, F., Corbeil, J., Omar, R. F. and Bergeron, M. G. (2017). Draft genome sequence of romboutsia weinsteinii sp. nov. strain ccri-19649t isolated from surface water, *Genome Announc.* **5**(40): e00901–17.

Manzi, V., Gennari, R., Hilgen, F., Krijgsman, W., Lugli, S., Roveri, M. and Sierro, F. J. (2013). Age refinement of the messinian salinity crisis onset in the mediterranean, *Terra Nova* **25**(4): 315–322.

Martínez, P. V. C., Suriñach, S. G., Marcén, P. G., Lull, V., Pérez, R. M. and Herrada, C. R. (1997). Radiocarbon dating and the prehistory of the balearic islands, *Proceedings of the Prehistoric Society*, Vol. 63, Cambridge University Press, pp. 55–86.

Mas-Peinado, P., Buckley, D., Ruiz, J. L. and García-París, M. (2018). Recurrent diversification patterns and taxonomic complexity in morphologically conservative ancient lineages of pimelia (coleoptera: Tenebrionidae), *Systematic entomology* **43**(3): 522–548.

Matsumiya, Y., Kato, N., Watanabe, K. and Kato, H. (2002). Molecular epidemiological study of vertical transmission of vaginal lactobacillus species from mothers to newborn infants in japanese, by arbitrarily primed polymerase chain reaction, *Journal of infection and chemotherapy* **8**(1): 43–49.

McDougall, E. (1948). Studies on ruminant saliva. 1. the composition and output of sheep's saliva, *Biochemical journal* **43**(1): 99.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing, *Cold Spring Harbor Protocols* **2010**(6): pdb–prot5448.

Milani, C., Mancabelli, L., Lugli, G. A., Duranti, S., Turroni, F., Ferrario, C., Mangifesta, M., Viappiani, A., Ferretti, P., Gorfer, V. et al. (2015). Exploring vertical transmission of bifidobacteria from mother to child, *Appl. Environ. Microbiol.* **81**(20): 7078–7087.

Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., Pusey, A. E., Peeters, M., Hahn, B. H. and Ochman, H. (2016). Cospeciation of gut microbiota with hominids, *Science* **353**(6297): 380–382.

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., Henrissat, B., Knight, R. and Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans, *Science* **332**(6032): 970–974.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R. et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes, *Nucleic acids research* **33**(17): 5691–5702.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. and Tyson, G. W. (2015). Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, *Genome research* **25**(7): 1043–1055.

Plessis-Rosloniec, D. and Zofia, K. (2011). *Steroid transformation by Rhodococcus strains and bacterial cytochrome P450 enzymes*, Dissertation, University of Groningen.

Puniya, A. K., Singh, R. and Kamra, D. N. (2015). *Rumen microbiology: from evolution to revolution*, Springer.

Ramis, D. and Bover, P. (2001). A review of the evidence for domestication of myotragus balearicus bate 1909 (artiodactyla, caprinae) in the balearic islands, *Journal of Archaeological Science* **28**(3): 265–282.

Rivera, L., Baraza, E., Alcover, J. A., Bover, P., Rovira, C. M. and Bartolomé, J. (2014). Stomatal density and stomatal index of fossil buxus from coprolites of extinct myotragus balearicus bate (artiodactyla, caprinae) as evidence of increased co2 concentration during the late holocene, *The Holocene* **24**(7): 876–880.

Ross, E. M., Moate, P. J., Bath, C. R., Davidson, S. E., Sawbridge, T. I., Guthridge, K. M., Cocks, B. G. and Hayes, B. J. (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing, *BMC genetics* **13**(1): 53.

Santiago-Rodriguez, T. M., Narganes-Storde, Y. M., Chanlatte, L., Crespo-Torres, E., Toranzos, G. A., Jimenez-Flores, R., Hamrick, A. and Cano, R. J. (2013). Microbial communities in pre-columbian coprolites, *PloS one* **8**(6): e65191.

Sasaki, E., Shimada, T., Osawa, R., Nishitani, Y., Spring, S. and Lang, E. (2005). Isolation of tannin-degrading bacteria isolated from feces of the japanese large wood mouse, apodemus speciosus, feeding on tannin-rich acorns, *Systematic and applied microbiology* **28**(4): 358–365.

Schubert, M., Lindgreen, S. and Orlando, L. (2016). Adapterremoval v2: rapid adapter trimming, identification, and read merging, *BMC research notes* **9**(1): 88.

Shiffman, M. E., Soo, R. M., Dennis, P. G., Morrison, M., Tyson, G. W. and Hugenholtz, P. (2017). Gene and genome-centric analyses of koala and wombat fecal microbiomes point to metabolic specialization for eucalyptus digestion, *PeerJ* **5**: e4075.

Siciliano, R. A. and Mazzeo, M. F. (2012). Molecular mechanisms of probiotic action: a proteomic perspective, *Current opinion in microbiology* **15**(3): 390–396.

Spor, A., Koren, O. and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome, *Nature Reviews Microbiology* **9**(4): 279.

Stamatakis, A. (2014). Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics* **30**(9): 1312–1313.

Tatusov, R. L., Galperin, M. Y., Natale, D. A. and Koonin, E. V. (2000). The cog database: a tool for genome-scale analysis of protein functions and evolution, *Nucleic acids research* **28**(1): 33–36.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

van Geel, B., Guthrie, R. D., Altmann, J. G., Broekens, P., Bull, I. D., Gill, F. L., Jansen, B., Nieman, A. M. and Gravendeel, B. (2011). Mycological evidence of coprophagy from the feces of an alaskan late glacial mammoth, *Quaternary Science Reviews* **30**(17-18): 2289–2303.

Van Soest, H., Gotink, W., Vooren, L. et al. (1965). Poisoning of pigs and cattle by buxus., *Tijdschrift voor Diergeneeskunde* **90**: 387–389.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., Orlando, L. and Krause, J. (2017). A robust framework for microbial archaeology, *Annual review of genomics and human genetics* **18**: 321–356.

Warinner, C., Speller, C., Collins, M. J. and Lewis Jr, C. M. (2015). Ancient human microbiomes, *Journal of human evolution* **79**: 125–136.

Weiß, C. L., Gansauge, M.-T., Aximu-Petri, A., Meyer, M. and Burbano, H. A. (2019). Mining ancient microbiomes using selective enrichment of damaged dna molecules, *bioRxiv* p. 397927.

Welker, F., Duijm, E., van der Gaag, K. J., van Geel, B., de Knijff, P., van Leeuwen, J., Mol, D., van der Plicht, J., Raes, N., Reumer, J. et al. (2014). Analysis of coprolites from the extinct mountain goat myotragus balearicus, *Quaternary Research* **81**(1): 106–116.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V. et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus, *Nature* **544**(7650): 357.

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., Paczian, T., Trimble, W. L., Bagchi, S., Grama, A. et al. (2015). The mg-rast metagenomics database and portal in 2015, *Nucleic acids research* **44**(D1): D590–D594.

Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M. B., Brand, T. B., Hofreiter, M., Bunce, M., Poinar, H. N., Dahl-Jensen, D. et al. (2007). Ancient biomolecules from deep ice cores reveal a forested southern greenland, *Science* **317**(5834): 111–114.

Winkler, D. E., Schulz, E., Calandra, I., Gailer, J.-P., Landwehr, C. and Kaiser, T. M. (2013). Indications for a dietary change in the extinct bovid genus myotragus (plio-holocene, mallorca, spain), *Geobios* **46**(1-2): 143–150.

Wood, J. R., Rawlence, N. J., Rogers, G. M., Austin, J. J., Worthy, T. H. and Cooper, A. (2008). Coprolite deposits reveal the diet and ecology of the extinct new zealand megaherbivore moa (aves, dinornithiformes), *Quaternary Science Reviews* **27**(27-28): 2593–2602.

Wu, G., Bazer, F. W., Burghardt, R. C., Johnson, G. A., Kim, S. W., Knabe, D. A., Li, P., Li, X., McKnight, J. R., Satterfield, M. C. et al. (2011). Proline and hydroxyproline metabolism: implications for animal and human nutrition, *Amino acids* **40**(4): 1053–1063.

Yll, E.-I., Perez-Obiol, R., Pantaleon-Cano, J. and Roure, J. M. (1997). Palynological evidence for climatic change and human activity during the holocene on minorca (balearic islands), *Quaternary Research* **48**(3): 339–347.

Young, M. F. (2003). Bone matrix proteins: their function, regulation, and relationship to osteoporosis, *Osteoporosis international* **14**(3): 35–42.

Zhu, L., Wu, Q., Dai, J., Zhang, S. and Wei, F. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome, *Proceedings of the National Academy of Sciences* **108**(43): 17714–17719.

3.6 Supplementary materials

3.6.1 Figure S3.1
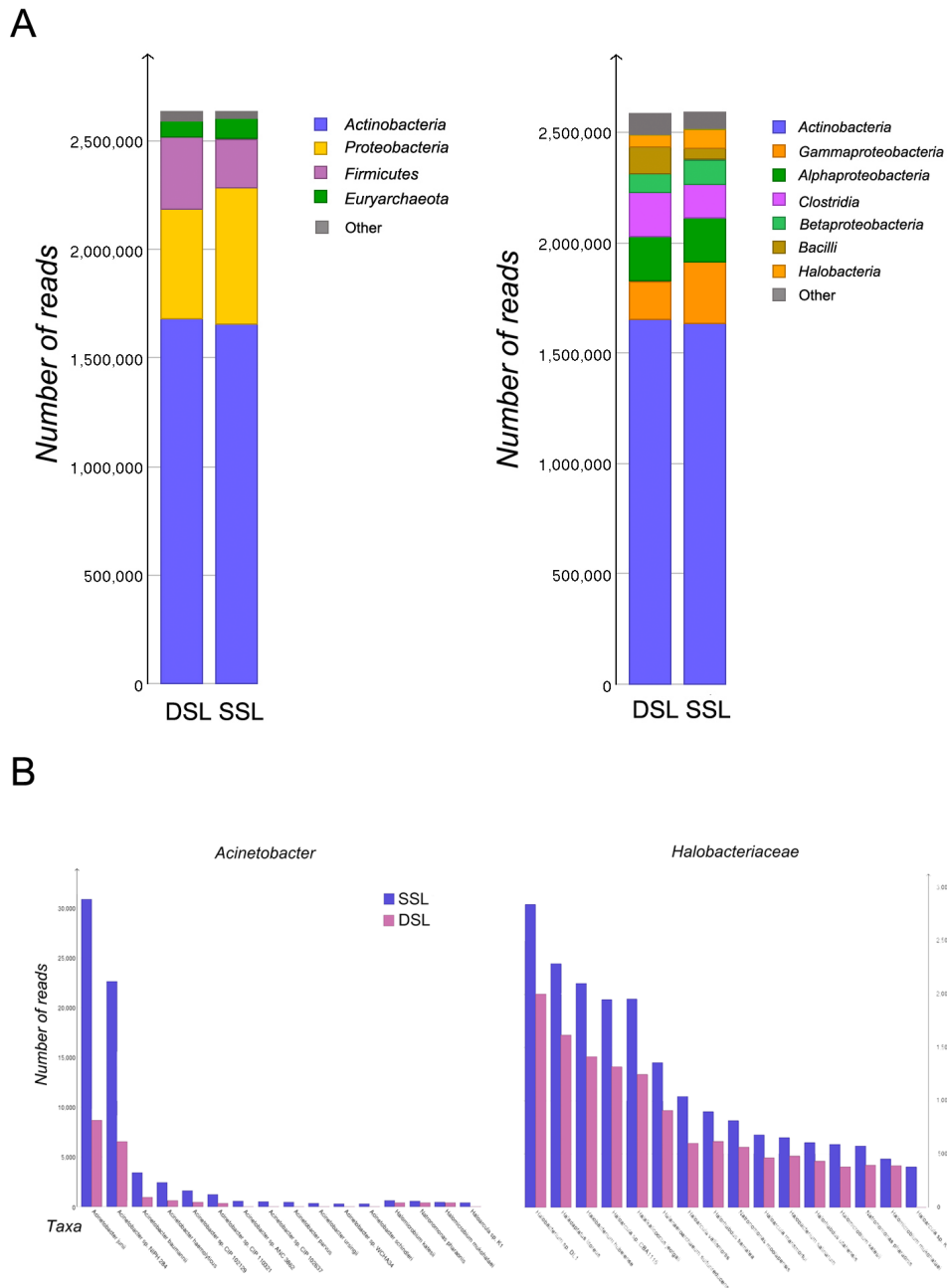
Figure 3.5: A. Bacterial composition at phyla (left) and class (right) level of MbCo-pro5 prepared using single- (SSL) and double-stranded library (DSL) protocols. B. The abundance of typical contaminant species identified from SSL and DSL data.
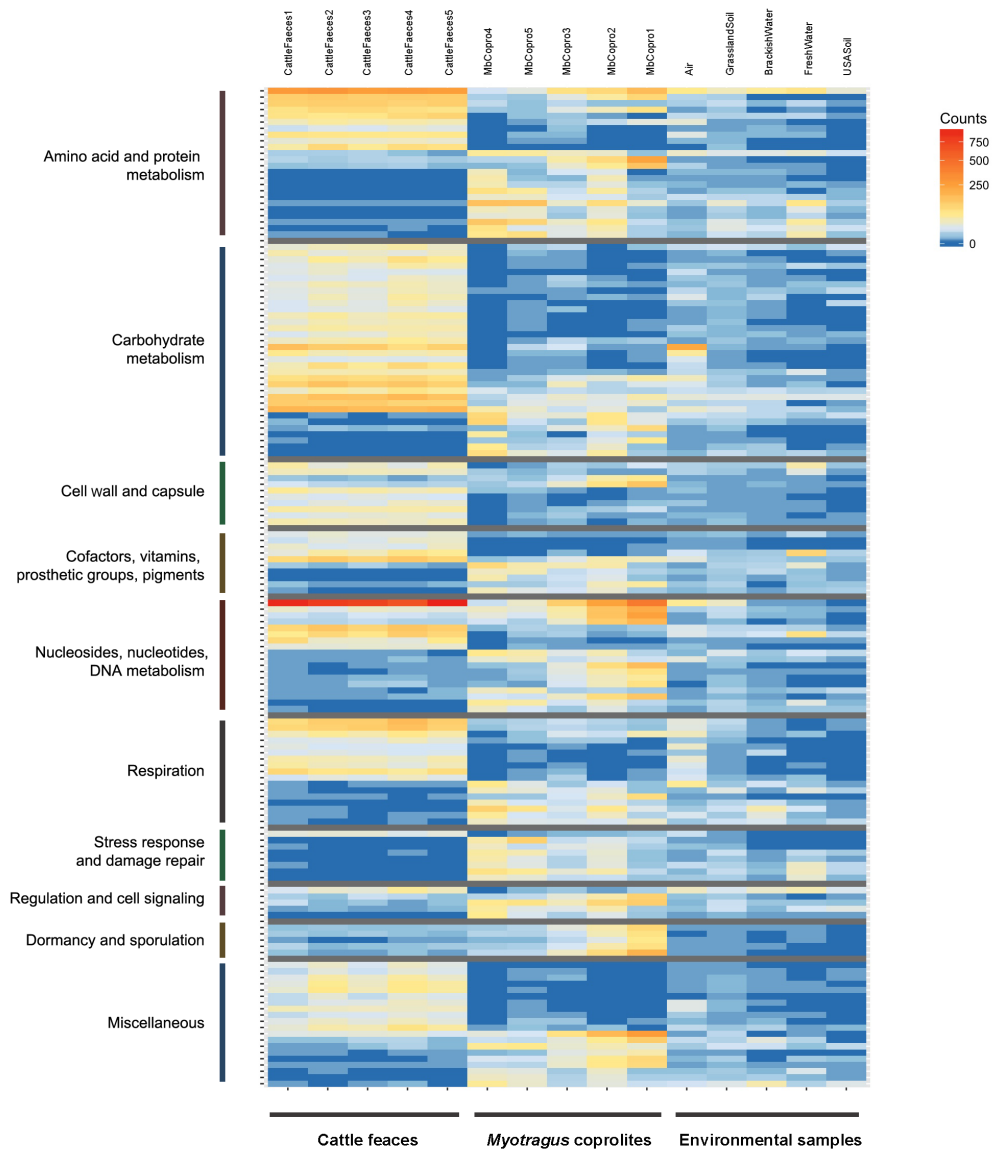
Figure 3.6: Microbiome functions specific to the *Myotragus* coprolites. Any functions identified from laboratory controls were removed.

3.6.3 Figure S3.3
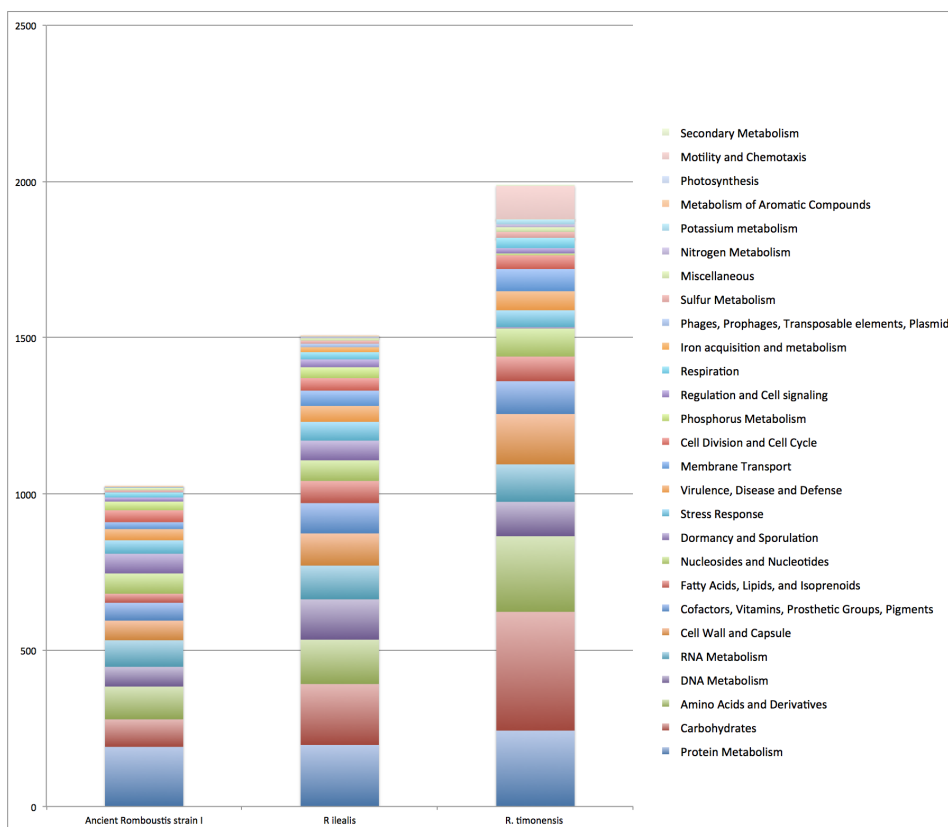
Figure 3.7: Number of annotated features within different functional groups of the ancient *Romboutsia* strain 1, *R. ilealis*, and *R. timonensis*.

3.6.4 Table S3.1

Sample details.

Figure abbreviations	Species	Sample type	Locality	Sequenced read count
MbCopro1	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	1,192,932
MbCopro2	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	1,026,009
MbCopro3	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	1,493,931
MbCopro4	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	1,010,160
MbCopro5 (SSL)	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	25,154,932
MbCopro5 (DSL)	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	7,598,107
MbCopro6	*Myotragus balearicus*	Coprolite	Pollença, Mallorca	481,246
MbCopro7	*Myotragus balearicus*	Coprolite	Valldemossa, Mallorca	463,775
MbCopro8	*Myotragus balearicus*	Coprolite	Valldemossa, Mallorca	843,898
CattleFaeces1	*Bos taurus*	Feces	Ellinbank, Victoria	6,321,628
CattleFaeces2	*Bos taurus*	Feces	Ellinbank, Victoria	5,381,779
CattleFaeces3	*Bos taurus*	Feces	Ellinbank, Victoria	3,851,729
CattleFaeces4	*Bos taurus*	Feces	Ellinbank, Victoria	5,295,240
CattleFaeces5	*Bos taurus*	Feces	Ellinbank, Victoria	4,427,042
EBC1	N/A	Extraction blank conrol	N/A	2,054,975
EBC2	N/A	Extraction blank conrol	N/A	330,633
EnvCtrl	*Myotragus aff. kopperi*	Molar	Llucmajor, Mallorca	921,582
Air	NA	Air	N/A	10,042,512
BrackishWater	NA	Water	N/A	2,830,843
FreshWater	NA	Water	N/A	5,863,944
GrasslandSoil	NA	Soil	N/A	3,962,726
USASoil	NA	Soil	N/A	6,543,903

3.6.5 Table S3.2

Presence of host mitochondrial DNA in coprolite and fecal samples.

Sample	Category	Refence	Input sequence (collapsed reads)	Unique	Reads mapped to host mito genome (per million)
MbCopro1	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	1192582	0	NA
MbCopro2	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	1026009	0	NA
MbCopro3	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	1493931	2	1.34
MbCopro4	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	1010160	0	NA
MbCopro5 (DSL)	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	7598107	7	0.92
MbCopro5 (SSL)	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	25154932	31	1.23
MbCopro6	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	481246	0	NA
MbCopro7	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	463775	0	NA
MbCopro8	Myotragus balearicus coprolite	Myotragus balearicus mitochondrial genome	843898	1	1.18
EBC1	Negative control	Myotragus balearicus mitochondrial genome	2054975	1	NA
EBC2	Negative control	Myotragus balearicus mitochondrial genome	330633	0	NA
EnvCtrl	Negative control	Myotragus balearicus mitochondrial genome	921582	0	NA
CattleFaeces1	Bos taurus feces	Bos taurus mitochondrial genome (GeneBank ID: GU947021.1)	6321628	39	6.17
CattleFaeces2	Bos taurus feces	Bos taurus mitochondrial genome (GeneBank ID: GU947021.1)	5381779	21	3.90
CattleFaeces3	Bos taurus feces	Bos taurus mitochondrial genome (GeneBank ID: GU947021.1)	3815729	12	3.14
CattleFaeces4	Bos taurus feces	Bos taurus mitochondrial genome (GeneBank ID: GU947021.1)	5295240	48	9.06
CattleFaeces5	Bos taurus feces	Bos taurus mitochondrial genome (GeneBank ID: GU947021.1)	4427042	19	4.29

3.6.6 Table S3.3

Proportion of sequences mapped to NCBI plastid database (updated on December 2017).

Specimen	Sequenced Read count	Reads assigned at all levels	Reads been assigned at all levels (%)	Reads assigned to species level (filtered)	Filtered reads been assigned at species levels (%)
MbCopro1	1,192,932	3062	0.26 %	426	0.0357%
MbCopro2	1,026,009	5537	0.54 %	722	0.0704%
MbCopro3	1,493,931	2827	0.19 %	378	0.0253%
MbCopro4	1,010,160	2765	0.27 %	303	0.0300%
MbCopro5 (DSL)	7,598,107	18510	0.24 %	2420	0.0319%
MbCopro5 (SSL)	25,154,932	27813	0.11 %	3748	0.0149%
MbCopro6	481,246	1416	0.29 %	163	0.0339%
MbCopro7	463,775	1825	0.39 %	215	0.0464%
MbCopro8	843,898	1572	0.19 %	196	0.0232%
CattleFaeces1	6,321,628	21835	0.35 %	2938	0.0465%
CattleFaeces2	5,381,779	20845	0.39 %	2757	0.0512%
CattleFaeces3	3,851,729	12898	0.33 %	1735	0.0450%
CattleFaeces4	5,295,240	20078	0.38 %	2778	0.0525%
CattleFaeces5	4,427,042	13539	0.31 %	1954	0.0441%
EBC1	2,054,975	672	0.03 %	NA	NA
EBC2	330,633	973	0.29 %	NA	NA
EnvCtrl	921,582	977	0.11 %	NA	NA

3.6.7 Table S3.4

CYP450 contigs assembled from coprolites microbiome.

Sample	Contigs	# Sequences	Contig Length	% Pairwise Identity	Accession	Organism
		Assembly info			**Annotation**	
MbCopro3	Assembly Contig 1	18	520	81.70 %	WP_044499210	Pseudomonas
	Assembly Contig 2	14	323	75.70 %	WP_054274025	Clostridioides
MbCopro2	Assembly Contig 1	78	1272	74.10 %	WP_044035940	Clostridium
	Assembly Contig 2	15	301	71.40 %	WP_041348209	Clostridium
	Assembly Contig 3	14	530	75.70 %	WP_054274025	Clostridioides
MbCopro1	Assembly Contig 1	17	513	64.10 %	WP_069930300	Streptomyces
	Assembly Contig 2	17	430	81.90 %	WP_078977127	Streptomyces
	Assembly Contig 3	16	509	69.60 %	WP_044035940	Clostridium
	Assembly Contig 4	15	388	70.50 %	WP_054274025	Clostridioides
	Assembly Contig 5	11	421	71.70 %	WP_094932885	Nocardiopsis
MbCopro5 (DSL)	Assembly Contig 1	153	978	56.80 %	WP_089002513	Micromonospora
	Assembly Contig 2	36	310	56.60 %	WP_091242943	Micromonospora
	Assembly Contig 3	10	445	91.90 %	EUA13778	Mycobacterium
MbCopro4	Assembly Contig 1	35	492	67.10 %	WP_051812853	Kitasatospora
	Assembly Contig 2	26	715	72.10 %	WP_086667108	Lentzea
	Assembly Contig 3	23	469	73.70 %	WP_091510136	Amycolatopsis
	Assembly Contig 4	12	332	64.50 %	WP_033381420	Kibdelosporangium
CattleFeace1	No contig assembled	NA	NA	NA	NA	NA
CattleFeace2	No contig assembled	NA	NA	NA	NA	NA
CattleFeace3	No contig assembled	NA	NA	NA	NA	NA
CattleFeace4	No contig assembled	NA	NA	NA	NA	NA
CattleFeace5	No contig assembled	NA	NA	NA	NA	NA
EBC1	No contig assembled	NA	NA	NA	NA	NA
EBC2	No contig assembled	NA	NA	NA	NA	NA
EnvCtrl	No contig assembled	NA	NA	NA	NA	NA
Air	Assembly Contig 1	43	911	99.40 %	WP_017834491	Kocuria
	Assembly Contig 2	15	444	91.90 %	WP_017833470	Kocuria
	Assembly Contig 3	14	224	96.20 %	WP_035928135	Kocuria
	Assembly Contig 4	11	323	92.50 %	WP_017833470	Kocuria
	Assembly Contig 5	11	295	99.30 %	WP_062737362	Kocuria
BrackishWater	Assembly Contig 1	17	897	74.20 %	WP_028659200	Nocardioides
	Assembly Contig 2	11	628	74.90 %	WP_014172850	Streptomyces
FreshWater	Assembly Contig 1	60	1224	96.10 %	WP_013983851	Pseudomonas
	Assembly Contig 2	16	897	94.20 %	WP_003288995	Pseudomonas
	Assembly Contig 3	11	875	83.30 %	WP_022977550	Nevskia
	Assembly Contig 4	11	672	93.50 %	WP_003288878	Pseudomonas
GrasslandSoil	No contig assembled	NA	NA	NA	NA	NA
USASoil	No contig assembled	NA	NA	NA	NA	NA

3.6.8 Table S3.5

Change point results of functional groups.

Sample ID	Functional group	bbg.pc	p
MbCopro1	amino acid metabolism	4	2.00E-04
	metabolism of aromatic compounds	3	2.00E-04
	metabolism of fatty acids	4	2.00E-04
	metabolism of phosphorus	7	0.2577
	protein metabolism	5	2.00E-04
	stress responses	4	2.00E-04
MbCopro2	amino acid metabolism	5	2.00E-04
	metabolism of aromatic compounds	3	2.00E-04
	metabolism of fatty acids	5	2.00E-04
	metabolism of phosphorus	5	2.00E-04
	protein metabolism	5	2.00E-04
	stress responses	5	2.00E-04
MbCopro3	amino acid metabolism	6	2.00E-04
	metabolism of aromatic compounds	6	2.00E-04
	metabolism of fatty acids	6	2.00E-04
	metabolism of phosphorus	6	2.00E-04
	protein metabolism	6	2.00E-04
	stress responses	6	2.00E-04
MbCopro4	amino acid metabolism	6	2.00E-04
	metabolism of aromatic compounds	6	2.00E-04
	metabolism of fatty acids	6	2.00E-04
	metabolism of phosphorus	8	2.00E-04
	protein metabolism	6	2.00E-04
	stress responses	5	2.00E-04
MbCopro5 (DSL)	amino acid metabolism	2	2.00E-04
	metabolism of aromatic compounds	1	NA
	metabolism of fatty acids	2	2.00E-04
	metabolism of phosphorus	2	0.0022
	protein metabolism	2	2.00E-04
	stress responses	2	2.00E-04
MbCopro6	amino acid metabolism	1	NA
	metabolism of aromatic compounds	1	NA
	metabolism of fatty acids	1	NA
	metabolism of phosphorus	12	0.972
	protein metabolism	1	1.9998
	stress responses	1	1.9998
MbCopro7	amino acid metabolism	2	0.0618
	metabolism of aromatic compounds	3	0.1126
	metabolism of fatty acids	1	NA
	metabolism of phosphorus	1	NA
	protein metabolism	3	0.1228
	stress responses	1	NA

MbCopro8	amino acid metabolism	3	0.0002**
	metabolism of aromatic compounds	6	0.0706
	metabolism of fatty acids	3	0.0002**
	metabolism of phosphorus	3	0.0758
	protein metabolism	6	0.0004**
	stress responses	1	NA
EBC1	amino acid metabolism	7	0.5293
	metabolism of aromatic compounds	7	0.4297
	metabolism of fatty acids	10	0.7802
	metabolism of phosphorus	10	0.7802
	protein metabolism	7	0.0818
	stress responses	7	0.5693
EBC2	amino acid metabolism	8	0.6863
	metabolism of aromatic compounds	9	0.838
	metabolism of fatty acids	13	0.9506
	metabolism of phosphorus	1	NA
	protein metabolism	2	0.122
	stress responses	14	0.8578
CattleFaeces1	amino acid metabolism	10	0.4997
	metabolism of aromatic compounds	1	NA*
	metabolism of fatty acids	10	0.7247
	metabolism of phosphorus	12	0.763
	protein metabolism	10	0.6065
	stress responses	12	0.5751
CattleFaeces2	amino acid metabolism	9	0.4621
	metabolism of aromatic compounds	6	0.3777
	metabolism of fatty acids	9	0.3579
	metabolism of phosphorus	10	0.7177
	protein metabolism	10	0.6689
	stress responses	10	0.7708
CattleFaeces3	amino acid metabolism	10	0.5979
	metabolism of aromatic compounds	4	0.044
	metabolism of fatty acids	3	0.0544
	metabolism of phosphorus	5	0.2791
	protein metabolism	3	0.0214
	stress responses	10	0.6451
CattleFaeces4	amino acid metabolism	10	0.2757
	metabolism of aromatic compounds	19	0.8778
	metabolism of fatty acids	12	0.6209
	metabolism of phosphorus	4	0.027
	protein metabolism	10	0.5551
	stress responses	10	0.5341

	amino acid metabolism	10	0.6523
	metabolism of aromatic compounds	9	0.6581
CattleFaeces5	metabolism of fatty acids	9	0.6581
	metabolism of phosphorus	3	0.0284
	protein metabolism	10	0.6757
	stress responses	3	0.0036
	amino acid metabolism	2	0.0424
	metabolism of aromatic compounds	1	NA
EnvCtrl	metabolism of fatty acids	1	NA
	metabolism of phosphorus	1	NA
	protein metabolism	1	NA
	stress responses	1	NA
	amino acid metabolism	2	0.0618
	metabolism of aromatic compounds	2	0.1296
Air	metabolism of fatty acids	2	0.0686
	metabolism of phosphorus	2	0.0844
	protein metabolism	2	0.0452
	stress responses	2	0.0804
	amino acid metabolism	4	2.00E-04
	metabolism of aromatic compounds	2	0.0024
BrackishWater	metabolism of fatty acids	3	0.0022
	metabolism of phosphorus	3	0.0366
	protein metabolism	4	0.0012
	stress responses	4	0.0692
	amino acid metabolism	5	0.0046
	metabolism of aromatic compounds	6	0.0262
FreshWater	metabolism of fatty acids	6	0.013
	metabolism of phosphorus	5	0.013
	protein metabolism	5	0.0068
	stress responses	6	0.0058
	amino acid metabolism	3	2.00E-04
	metabolism of aromatic compounds	3	0.0022
GrasslandSoil	metabolism of fatty acids	3	4.00E-04
	metabolism of phosphorus	3	0.001
	protein metabolism	3	2.00E-04
	stress responses	3	4.00E-04
	amino acid metabolism	1	NA
	metabolism of aromatic compounds	1	NA
USASoil	metabolism of fatty acids	1	NA
	metabolism of phosphorus	1	NA
	protein metabolism	1	NA
	stress responses	1	NA

*A p-value of NA indicates that the first position was selected

** Not matching aDNA damage pattern (decreased TA ratio rather than increased TA ratio towards the end of the read)

3.6.9 Table S3.6

Assembly information of the ancient Romboutsia genomes.

Sample	Reference genome	Number of reads assembled	Mean depth	Coverage	GC content (%)	5' C-to-T	3' G-to-A
MbCopro2	*R. ilealis* CRIB	361,174	12.1	80.50 %	28	0.23	0.19
MbCopro1	*R. ilealis* CRIB	196,163	5.6	70.80 %	28	0.07	0.22
MbCopro3	*R. ilealis* CRIB	76,970	2.2	54.10 %	29.1	0.23	0.077

3.6.10 Table S3.7

CheckM results of the assembled Romboutsia genomes.

Strain	Completeness (%)	Contamination (%)	Strain heterogeneity (%)
R. ilealis (RefSeq)	99.30	0	0
R. ilealis MbCopro1	96.83	0	0
R. ilealis MbCopro2	63.85	0	0
R. ilealis MbCopro3	91.67	0.47	0

Statement of Authorship

Title of Paper	Recovery of oral microbiome signal from ancient bison teeth
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	In preperation for submission

Principal Author

Name of Principal Author (Candidate)	Yichen
Contribution to the Paper	Experiments, data analysis, wrote the manuscript
Overall percentage (%)	75%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 05/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

  i. the candidate's stated contribution to the publication is accurate (as detailed above);
  ii. permission is granted for the candidate in include the publication in the thesis; and
  iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Keith Dobney
Contribution to the Paper	Identification of the cementum samples, data interpretation
Signature	Date 25/02/2019

Name of Co-Author	( )
Contribution to the Paper	Data interpretation, provided samples
Signature	Date

Please cut and paste additional co-author panels here as required.

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    iv.   the candidate's stated contribution to the publication is accurate (as detailed above);

    v.   permission is granted for the candidate in include the publication in the thesis; and

    vi.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Laura Weyrich	
Contribution to the Paper	Data interpretation, manuscript editing	
Signature		Date	5/3/19

Chapter 3 Yichen Liu

Chapter 4

Recovery of oral microbiome signal from ancient bison teeth

Abstract

Robust oral microbiome signals are routinely obtained from calcified dental plaque (calculus) present on the teeth of ancient humans, providing information to infer ancient human behavior, diet, and health, as well as the co-evolutionary relationships between commensal bacteria and their hosts. However, a lack of dental calculus present on non-human animal remains has currently limited these studies to humans, despite the wealth of information that oral microbiome data from extinct mammals could provide about past extinction and adaptive processes. Here, we sequenced ancient DNA preserved in cementum and a dental "plaque-like" structure on bison teeth as a model to explore the oral microbiome signal preserved in past non-human mammals. Although a large proportion of the obtained signal originated from the environment, DNA of oral, mucosal, and ruminal bacteria were successfully obtained from ancient and modern bison specimens. DNA of *Actinomyces* species was notably well-preserved, and the draft genomes of two ancient *A. ruminicola* strains (~33kyr) were reconstructed. Dietary and host mitochondrial DNA was also detected from both ancient cementum and "plaque" samples, offering an alternative non-invasive, taxonomic identification method. Although this is an exploratory study, we show that the ancient non-human teeth can serve as a promising proxy to recover oral microorganisms and host information from ancient animals.

Key words: Ancient DNA, oral microbiome, paleomicrobiology, bison

4.1   Introduction

Diverse microbes, including more than 200 species belong to Firmicutes, Proteobacteria, Actinobacteria, Bacteroidetes and Fusobacteria phyla, inhabit the mammalian oral cavity, and the microbes and their genetic contact is referred to as the oral microbiome (Dewhirst et al.; 2010, 2012). Certain bacteria can form diverse biofilms on the tooth and mucosal surfaces in the oral cavity. Biofilms on teeth (*i.e.* plaque) comprise of *Streptococcus, Actinomyces*, *Propionibacterium, Haemophilus*, and several other species (Kolenbrander et al.; 2010). While microorganisms in dental plaque have been extensively related to oral health, increasing evidence suggests that these bacteria also play key roles in systemic diseases, such as diabetes and cardiovascular diseases (Dewhirst et al.; 2010; Bowen et al.; 2018). Many of these oral bacteria appear to be vertically transmitted and likely co-evolve with their hosts (Mueller et al.; 2015; Dominguez-Bello et al.; 2010; Guerrero et al.; 2013). However, co-evolutionary relationships and the roles that oral microbiome plays in mammalian evolution remain largely unexplored.

Ancient DNA techniques can be utilized to obtain detailed records of past oral microbes, providing a genetic window into the past (Hofreiter et al.; 2001). However, ancient DNA extracted from such specimens needs to be authenticated and is subject to contamination, as the specimens have been exposed to the environment for a long period (Cooper and Poinar; 2000). Therefore, ancient DNA research usually requires an assessment of the ancient origin of the extracted DNA (*i.e.*, authentication) and stringent decontamination measures (Weyrich et al.; 2015, 2018). Authentication of ancient DNA largely relies on the DNA damage patterns, which includes 1) fragmentation: ancient DNA usually was fragmented into small fragments with an average length smaller than 100 bp; 2) depurination, which leads to the overrepresentation of the purines (guanine and adenine) at positions adjacent to the molecule breaks; and 3) deamination, which causes increased cytosine to thymine and guanine to adenine substitution at the termini of DNA after amplification and sequencing (Briggs et al.; 2007). While the deamination rate of ancient DNA increases over time, no obvious pattern was previously observed between the fragment lengths and the age of the sample (Sawyer et al.; 2012). To minimize the biases cause by contamination, ancient samples are also routinely decontaminated using bleach (sodium hypochlorite), UV irradiation, and EDTA presoak (Kemp and Smith; 2005; Farrer; 2016).

These advanced ancient DNA techniques have been recently applied to obtain ancient human oral microbiomes from calcified dental plaque (calculus) (Weyrich et al.; 2015; Adler et al.; 2013; Weyrich et al.; 2017; Warinner et al.; 2014). Dental calculus can preserve bacterial cells, dietary microfossils, and host DNA (Weyrich

et al.; 2015), thus providing important information about ancient human behaviors, diets, diseases, and the responses of the oral bacteria community to major biocultural changes, such as the adoption of agriculture (Adler et al.; 2013; Weyrich et al.; 2015). While dental calculus is widely distributed on the ancient human teeth, it is less prevalent on animal teeth, which poses a major challenge to paleomicrobiology studies on non-human mammals. To date, no published studies have examined the oral microbiome signatures preserved on the teeth of non-human mammals, especially those that have gone extinct.

Steppe bison (*Bison priscus*) are a megafauna that experienced and survived the last ice age and have been used as a model species to investigate the long-term interactions between large mammals and environmental changes (Cooper et al.; 2015; Shapiro et al.; 2004; Soubrier et al.; 2016). Similarly, modern bison (*Bison bison*) have also responded to large scale environmental changes, including several population bottlenecks and a significant loss of genetic diversity in extant populations – a continuing concern with their conservation (Hartl and Pucek; 1994; Gates et al.; 2010). Bison have high-crowned cheek teeth (hypsodont dentition) (Koenigswald; 2011; Wilson; 1988), the structure of which is showed in Figure 4.1A. On the enamel surface of bison teeth, coronal cementum can be widespread as a thin layer. In ancient, buried teeth, the cementum is not strongly bonded to the enamel and is subject to exfoliation or chemical weathering (Wilson; 1988; Lyman; 2018). Dental plaque can form on the surface of bison teeth, but it is likely less resilient in the taphonomic process. Together, it is possible that an oral microbial signal in ancient bison is present.

Here, we used bison as a model species to explore ancient oral microbiome information preserved on the teeth of non-human mammals. We explore 12 ancient individuals and compare these to 17 modern specimens that were buried underground for one year to mimic taphonomic processes. Together, we explore the preservation of oral, mucosal, and ruminal microorganisms within these tissues and how their preservation is impacted by sample type, preservation status, or decontamination method.

4.2 Methods and Materials

4.2.1 Sampling

Oral samples were collected from 12 ancient bison individuals from the Yukon, Canada and Minnesota, USA (Table S4.1). All the ancient samples were stored and examined in a purpose built ancient DNA facility at Australian Centre for Ancient DNA (ACAD), the University of Adelaide. Samples from the single 33 kyr bison

(A3275) were stored at -20°C, as it was collected from permafrost, while the remaining ancient samples were stored at 4°C. We observed that the cementum was peeling away from the surface of some ancient bison teeth (Figure 4.1 F). However, the some ancient bison teeth remained intact and maintained a soft brown layer on the surface of the teeth (*i.e.* individual A3275 (Figure 4.1 E)), which was visually similar to the dental plaque formations present in modern bovid animal teeth. Therefore, we characterized ancient bison oral samples into two sample types: cementum and "plaque". While we were unable to obtain proteomic or chemical analysis of this plaque-like structure, we will refer to this substance as "plaque" throughout this paper for simplification. These samples include 54 cementum samples from 11 individuals range from (2kyr to >50kyr) and 11 plaque specimens from a single 33kyr bison jaw from Beringian permafrost (Table S4.1). We also included 34 cementum samples from 17 modern bison individuals (Table S4.1). The modern bison teeth samples were collected from an organic, grass-fed bison ranch in South Dakota, USA. Bison ranged between 1-2 years of age and are prairie harvested (slaughtered) in the field. After harvest, bones from each individual were then buried in the field approximately 5 feet underground and left for one year to mimic taphonomic and environmental contamination processes, similar to ancient specimen. Samples were then unearthed and whole mandibles were shipped to the University of Adelaide on ice. Upon arrival at the University of Adelaide, the modern bison samples were stored at 4°C. All modern samples were stored and processed within modern DNA laboratories. To evaluate the preservation of the oral microbiome information and potential contamination from the environment within the samples, we additionally included shotgun sequencing data of environmental samples from MG-RAST metagenomics database (MG-RAST ID: air-mgm4516952; brackish water-mgm4536373; fresh groundwater-mgm4536380; grassland soil-mgm4511193) (Wilke et al.; 2015), as well as three ancient bison bone fragments (which should not contain oral microbiome information).

4.2.2 Decontamination

Cementum Samples

Ancient cementum samples were decontaminated using two methods currently applied in the analysis of ancient human dental calculus to initially examine the impacts of different decontamination methods on the identified species (Figure 4.2). In the first group (UV-bleach), cementum samples (n=52 ancient; n=9 modern) were decontaminated using 15 min UV radiation followed by soaking in 3% bleach for 2min. The residual bleach was removed by rinsing samples in 80% ethanol for 3 min and then dried at room temperature for 2min. In the second group (UV-EDTA

group), cementum samples (n=2 ancient; n=17 modern) were decontaminated using 15 min UV radiation followed by 10 min 0.5M EDTA (ethylenediaminetetraacetic acid) wash, instead of 2 min bleach treatment.

Plaque Samples

As bleach is a strong oxidant, we observed several early plaque samples dissolve during the UV-bleach decontamination method. Therefore, the 11 ancient plaque samples from a 33kyr old ancient bison individual (A3275) were only decontaminated using 15 min UV radiation followed by 10 min 0.5M EDTA (ethylenediaminetetraacetic acid) wash, again instead of 2 min bleach treatment (Figure 4.1 E).

4.2.3 DNA extraction and Shotgun library preparation

DNA was extracted using an in-house silica-based method, as previously described (Brotherton et al.; 2013). Briefly, each decontaminated sample was crushed into powder and then decalcified using 1.6ml of 0.5M EDTA, 0.1ml of 10% sodium dodecyl sulphate (SDS), and 20 μL of 20 mg/mL proteinase K and left to rotate at 55 °C for 20 hours. Released DNA was then bound to silica suspensions with modified QG buffer (4 mL modified QG buffer contains 3.7 ml QG buffer (QIAGEN), 20 μL of 5M sodium chloride (NaCl), 222 μL of 3M sodium acetate (NaAc), and 61 μL of water. Bound DNA was then washed twice in 80% ethanol and then eluted using 100 μL of Tris-HCl buffer.

The DNA extracts were then used to construct shotgun libraries, as previously described (Meyer and Kircher; 2010). Briefly, 20 μL of DNA extract was repaired using T4 polynucleotide kinase (New England Biolabs) and T4 DNA polymerase (New England Biolabs) for 15 minutes at 25°C. Two 7bp barcoded Illumina adapters were then ligated to the repaired DNA using T4 DNA ligase (Fermentas) for 60min at 22°C. A MinElute Reaction Cleanup Kit (QIAGEN) was used for the cleanup of the enzymatic reactions. The gaps on the DNA molecules were then filled using Bst DNA polymerase (New England Biolabs) for 30 minutes at 37°C, and then the enzyme was denatured at 80°C for 10 min. The resulting reactions were amplified with 2.5 μL of 10× Gold buffer, 1.25 μL of 50 mM MgSO$_4$, 0.25 μL of 25mM dNTP, 1.25 μL of IS7 primer, and 1.25 μL of IS8 primer under the following conditions: 12 min at 94°C; 13 cycles of 30 sec at 94°C, 30 sec at 60°C, 45 sec at 72°C; and 10 min at 72°C. The resulting reactions were cleaned using Ampure XP and then reamplified using GAII indexed primers for 13 cycles. The reactions were then cleaned using Ampure XP, quantified using Agilent TapeStation, and pooled at equimolar concentration before 150 bp paired-end sequencing on Illumina NextSeq and Ten-X platforms.
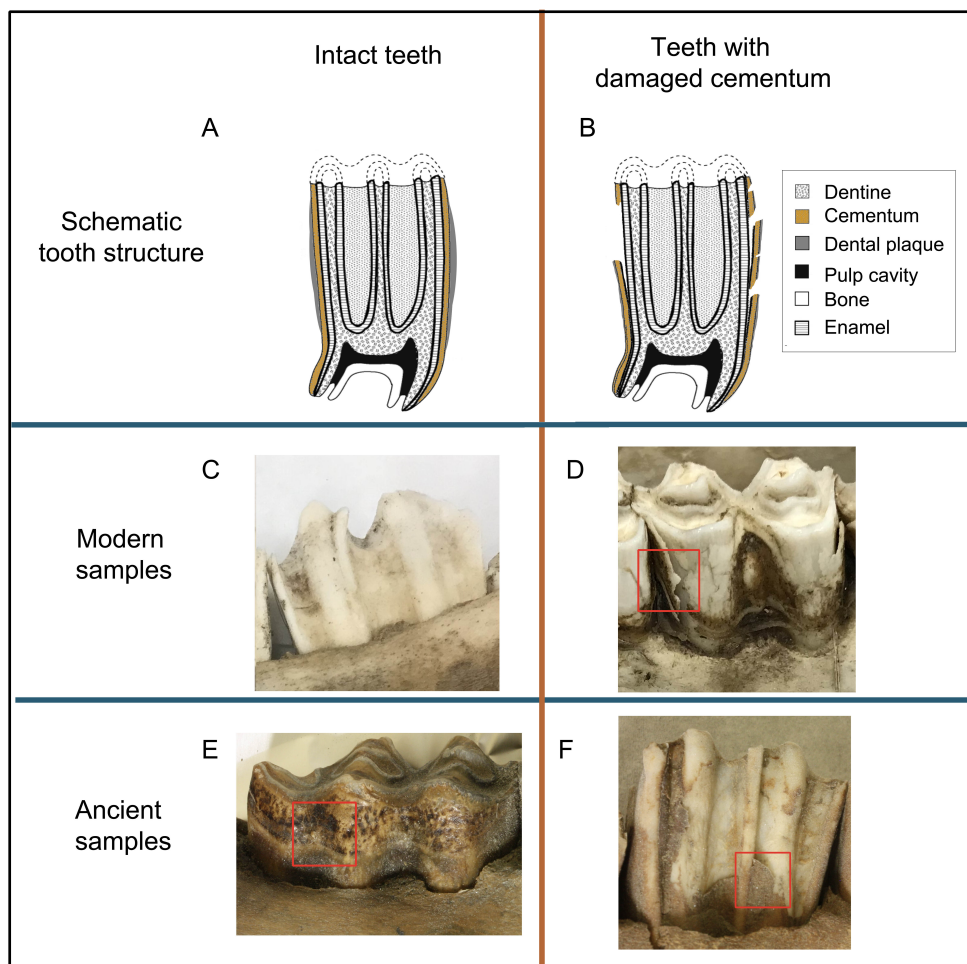
Figure 4.1: Ancient and modern bison teeth samples. A and B. Schematic structure of an intact ruminant tooth (A) and ruminant tooth with cementum exfoliation (B) (adapted from:(Stromberg; 2006)). C. A modern bison tooth sample with intact cementum. D. A modern bison tooth sample with cementum peeling away from enamel surface and was sampled (the red square). E. An exceptionally well-preserved ancient bison tooth sample. The cementum was intact, and a soft layer (likely a bacterial biofilm that similar to dental plaque, the red square) was on the cementum and was sampled. F. An ancient bison tooth. The cementum was peeled off and was sampled (the red square)
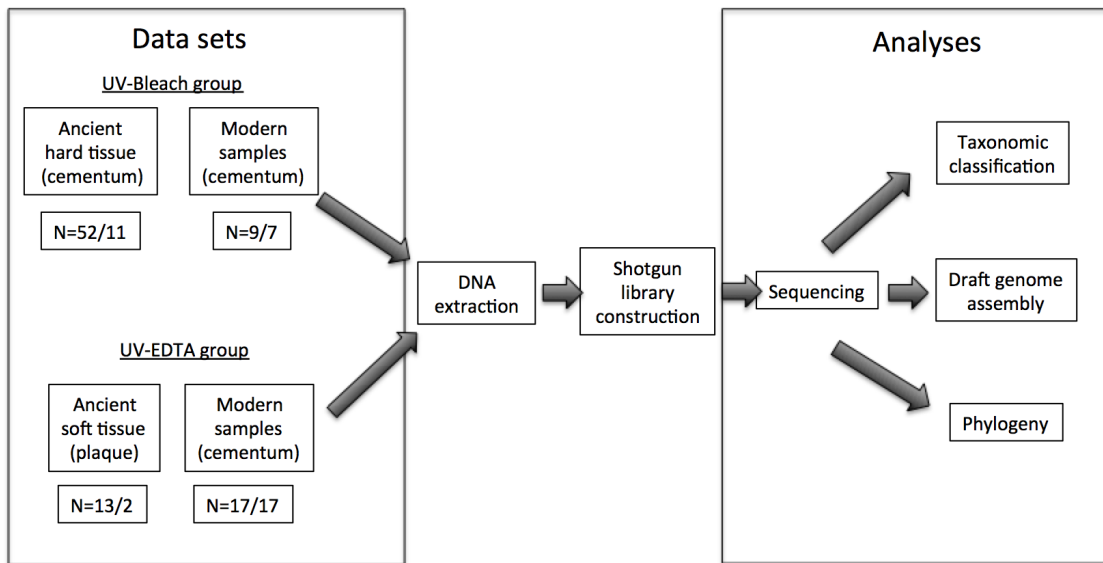
Figure 4.2: A flowchart describing the sample and data processing. Ancient samples were characterized into two sample types: cementum (UV-Bleach group) and plaque (UV-EDTA group). Samples were decontaminated using UV-bleach or UV-EDTA according to the tolerance to bleach. To avoid biases caused by different decontamination methods, ancient samples were only compared to modern samples that were decontaminated using the same approach. DNA was then extracted from decontaminated samples and shotgun libraries were prepared. Processed sequencing data were used for downstream microbiomic analyses.

4.2.4 Data processing and taxonomic analyses

FASTQ data from the Illumina machine was demultiplexed using Sabre v.1.0 (http://github.com/najoshi/sabre) allowing one mismatch (option m -1). The demultiplexed data were then trimmed and collapsed using AdapterRemoval v.2.1.7 with the following parameters: mismatch rate 0.1, minimum Phred quality 4, quality base 33, trim ambiguous bases (N) and bases with qualities equal or smaller than the given minimum quality, and a minimal overlapping length of 11 bp (Schubert et al.; 2016). The collapsed reads were then compared against an in-house database that included 47,696 archaeal and bacterial genome assemblies from the NCBI Assembly database (A; 2018) using MEGAN Alignment Tool (MALT, v 0.3.8) using the following parameters: Weighted-LCA=80%, minimum bitscore=44, minimum E-value=0.01, minimum support percent=0.01 (Herbig et al.; 2016). To obtain the taxonomic information of eukaryotic DNA, the collapsed reads were also compared to the NCBI non-redundant database (updated on December 2017) using DIAMOND (v 0.9.13) (Buchfink et al.; 2015), as MALT required more RAM than was current available at the Unviersity of Adelaide. The generated DIAMOND alignment archive (DAA) files were annotated using the daa-meganizer script included in MEGAN6 (parameters used: Weighted-LCA=80%, minimum bitscore=44, min-

imum E-value=0.01, minimum support percent=0.01). Resulting data were visualized and analysed in MEGAN6 (v 6.13.1) (Huson et al.; 2016).

4.2.5 Competitive mapping and draft genome assembly

The reference sequences used for competitive mapping consisted of five *Actinomyces* species: *A. slackii* ATCC 49928 (GCF_000428685.1), *A. ruminicola* (GCF_900103565.1), *A. glycerinotolerans* G10 (GCF_900098745.1), *A. succiniciruminis* (GCA_900002405), and *A. oris* (GCF_001553935.1). The genomes of these *Actinomyces* species were concatenated with a spacer of 100 Ns to create one concatenated reference sequence. The collapsed reads were then mapped to the concatenated reference sequence using BWA v.0.7.13 backtrack algorithm with the following parameters: -l 1024, -n 0.01 (Li and Durbin; 2009). Duplicate reads and mapped reads with quality lower than a Phred score of 30 were removed using SAMtools v.1.3.1 (Li et al.; 2009). The resulting bam files were then visualized and analyzed using QualiMap v.2.2.1 (Okonechnikov et al.; 2015). The draft genomes were then constructed by separating the uniquely mapped sequences according to the best mapped species, creating five draft genomes for each sample. The majority of the reads were mapped to the one of the five genomes (Figure 4.4), the draft genomes have the most mapped reads were retained for the downstream analysis. The consensus sequence of assembled genomes of *Actinomyces* strains was generated using Geneious (Kearse et al.; 2012). For the ambiguous bases, the consensus was called with a threshold of 85%. The coverage of each draft genomes was calculated using SAMtools (v 1.3.1) bedcov on a window size of 2.5kb. The GC content was calculated using a python script GCcalc.py with a window size of 2.5kb (https://github.com/WenchaoLin/GCcalc/blob/master/GCcalc.py). The GC skew was calculated using a perl script gcSkew.pl with a window size of 2.5kb (https://github.com/Geo-omics/scripts/blob/master/AssemblyTools/gcSkew.pl). The coverage of the bam files were exported from Geneious and then fed into an inhouse awk script to generate a bed file the included the range of mapped regions for each genome. The results were visualized using Circos (v 0.67-7) (Kearse et al.; 2012; Krzywinski et al.; 2009). Nucleotide misincorporation and fragmentation patterns were calculated using MapDamage2.0 (Jónsson et al.; 2013).

4.2.6 Phylogenetic analysis

The assembled draft genomes of two ancient *A. ruminicola* bison strains (Strain A19865 and Strain A19885), one *A. slackii* strain from a modern bison (Strain MMB1519), and the reference sequences of *A. slackii* ATCC 49928 (GCF_000428685.1), *A. ruminicola* (GCF_900103565.1), *A. glycerinotolerans* G10 (GCF_900098745.1),

A. succiniciruminis (GCA 900002405), *A. oris* (GCF 001553935.1), *A. gaoshouyii* (GCF 002076915.1), *A. liubingyangii* (GCF 001907245.1), and *A.vulturis* (GCF 001 687305) were aligned using mauveAligner algorithm in Geneious (Kearse et al.; 2012). The locally collinear blocks (LCBs) were concatenated with a spacer of 100 Ns. The resulting alignment was then used to build a maximum likelihood phylogenetic tree using RAxML in Geneious (parameters used: GTR GAMMA model, 1000 bootstrap replicates) (Stamatakis; 2014). *A. liubingyangii* was used as an outgroup to root the tree.

4.2.7   Mapping sequences to the bison mitochondrial genome

Collapsed reads from all libraries were mapped to *Bison priscus* mitochondrial genome (GenBank ID: KX269121.1) using the BWA v.0.7.13 backtrack algorithm (with options: -l 1024, -n 0.01) (Li and Durbin; 2009). Mapped reads with quality lower than a Phred score 30 were removed using SAMtools v.1.3.1 (Li et al.; 2009). The resulting bam files passed the quality filtering were imported into Geneious and the number of mapped reads were counted (Kearse et al.; 2012).

4.3   Results

4.3.1   Oral microbiome signal in ancient bison oral samples

The microbial composition of the bison cementum and plaque samples was examined by comparing the microbiome of ancient and modern bison oral samples to environmental and laboratory control samples by performing PCoA on Bray Curtis distances on species level (Figure 4.3A and B). The microbiome from most ancient and modern bison samples are distinct from those of environmental samples and controls (p<0.01). A proportion of the bison oral samples overlap with controls, suggesting varied preservation. Irrespective of decontamination method, we observed the infiltration of the environmental taxa in both ancient and modern bison oral samples. Specifically, the ancient bison samples maintained bacteria indicative of soil microbes and were dominated by Burkholderiales (19.34%), Rhizobiales (14.5%), and Micrococcales (6.4%). These taxa were identified in modern bison samples as well (Burkholderiales: 5.0%, Rhizobiales: 1.8%, and Micrococcales: 16.7%). These dominant, environmental taxa were detected in both ancient and modern samples, irrespective of decontamination method, suggesting that the decontamination methods had minimal effects on reducing environmental species. Nevertheless, we decided to filter out the presence of environmental species to better examine the oral microbial signal within the data set. Next, we examined the presence of laboratory

contaminant species within the data set. We first examined the impact of laboratory contamination with respect to the type of decontamination method applied to cementum samples, as no modern plaque samples were available for comparison. Laboratory contamination significantly contributed to the alpha-diversity observed in cementum, as a higher alpha diversity observed in UV-EDTA treated samples compared to UV-bleach group (p=0.02639) was largely driven by laboratory contamination. The removal of laboratory taxa identified in extraction blank controls ameliorated this effect (p=0.08842). Therefore, species identified in the EBCs from all samples were also removed from bison cementum and plaque for downstream analysis.

We then explored the presence of oral bacteria in bison oral samples. As the bison oral microbiome has not yet been examined, we explored the presence of three different origins of microorganisms in the oral cavity: 1) microbes in the oral cavity of other mammals, including key genera required for the primary formation of dental plaque in human, such as *Actinomyces* and *Streptococcus* species (Alfano et al.; 2015; Hyde et al.; 2014; Bik et al.; 2016; Kolenbrander et al.; 2010); 2) microbes present in the bovine nasopharyngeal mucosa, such as *Pasteurella, Neisseria,* and *Psychrobacter* (Timsit et al.; 2016); and 3) ruminal microorganisms, such as methanogens, as the rumen fluid containing diverse microorganisms can be transferred into ruminants' mouth through eructation and rumination (Sirohi et al.; 2010; DePeters and George; 2014). The microbial composition of all modern and ancient bison oral samples was then classified at bacterial order and genera level (Table S4.2 and Table S4.3). In cementum, modern samples contained higher abundances of oral microbes relative to that of the ancient samples. In modern cementum samples, *Streptococcus* (25.28%) represents the most abundant bacterial genus detected, *Psychrobacter* (5.56%) and *Actinomyces* (3.92%) also detected with relatively high abundance. *Streptococcus* (0.97%), *Psychrobacter* (0.10%), and *Actinomyces* (0.15%) were also detected from ancient cementum samples, while the abundance is lower than modern samples. *Fusobacterium* (0.03%) was detected from seven ancient cementum samples. Ancient plaque seems preserve better oral microbial signal than cementum. In ancient plaque samples, *Actinomyces* (21.28%) is the most abundant bacterial genus, other oral bacteria, such as *Streptococcus* (2.54%), *Fusobacterium* (0.76%), *Neisseria* (0.69%), *Porphyromonas* (0.57%) were also detected. We then explored the presence of other ancient oral taxa on the species level (Figure 4.3C and 3D). Although only limited oral microbial signal was detected from the majority of ancient cementum samples, cementum samples from one ancient bison (A16193(#11), 33 kyr) preserves good oral signal, with 4 *Streptococcus* species, 3 *Psychrobacter* species, 11 *Olsenella* species, and 5 *Actinomyces* species identified; ruminal methanogens, including 5 *Methanobrevibacter* species, were also detected from these samples. The

ancient plaque contained more oral species, which was dominated by *Actinomyces* species compared to species within other genera (Figure 4.3D).

Overall, we show that oral microorganisms that likely originate from the bison oral cavity are preserved in ancient cementum and plaque samples, although these samples are intermixed with environmental taxa. The ancient dental plaque preserved more oral signal compared to the cementum, although further examination of bison dental plaque is needed across a wide range of individuals.

4.3.2   Draft genome of two 33kyr old *Actinomyces ruminicola* strains

As *Actinomyces* species are among the most diverse and abundant oral taxa that we observed in our ancient bison oral samples, we reconstructed the draft genomes of two *Actinomyces* strains from the oral samples in two 33kyr bison individuals, as well as two modern bison samples for the comparison. We employed competitive mapping to avoid cross-mapping of DNA fragments of closely related *Actinomyces* species. The reference genome consisted of the genomes of the five most abundant *Actinomyces* species identified from bison oral samples: *A. slackii* ATCC 49928 (GCF_000428685.1), *A. ruminicola* (GCF_900103565.1), *A. glycerinotolerans* G10 (GCF_900098745.1), *A. succiniciruminis* (GCA_900002405), and *A. oris* (GCF_001553935.1). The first four reference species were isolated from ruminant dental plaque or rumen fluid (Dent and Williams; 1986; NA et al.; 2016; An et al.; 2006), and the latter was present in the mammalian oral cavity (Henssge et al.; 2009). The majority of reads from two ancient samples uniquely mapped to *A. ruminicola*. In the modern specimens, one individual (MMB1511) was dominated by *A. ruminicola*, while the other (MMB1519) was dominated by *A. slackii* (Figure 4.4). As the abundance of *Actinomyces* species of other samples is not enough for competitive mapping, it is unclear if the oral cavity of modern and ancient bison was dominated by different *Actinomyces* species. We then assembled the draft genomes of two ancient *A. ruminicola* strains using only uniquely mapped sequences obtained during competitive mapping (Figure 4.5). The draft genome of the first ancient *A. ruminicola* strain (ancient strain A19865) covered 72.1% of the reference genome with a mean depth of 2.2×. The second draft genome (ancient strain A19885) covered 30.0% of the reference genome with a mean depth of 0.5×. The GC content of the reference genome (*A. ruminicola*) is 69.5% was also similar to what was seen in both ancient genomes (ancient strain A19865 and A19885 are 68.8% and 69.3%, respectively).

The two ancient strains showed signs of typical ancient DNA damage (fragmentation and deamination) that was consistent with the age of the samples (A19865:
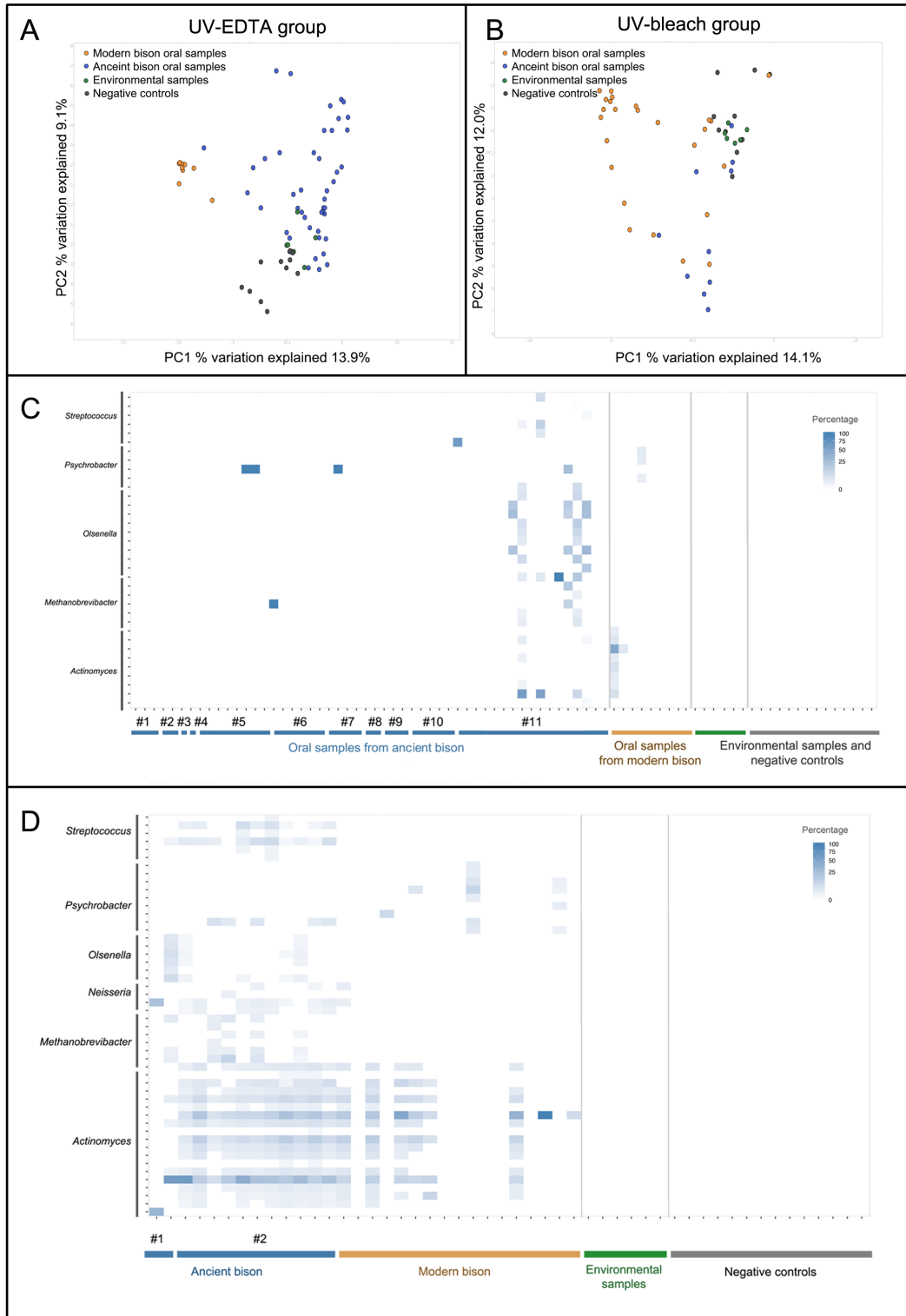
Figure 4.3: The oral microbiome of ancient and modern bison. A and B: PCoA plot (Bray-Curtis distance) of cementum (A) and plaque (B) microbiome. C and D: The presence of typical bovid oral microbiome taxa in bison cementum samples (C) and plaque samples (D). Each column represents one sample. Multiple samples were collected from ancient bison individuals, and individuals were indicated using #.

mean fragment length: 64 bp, 3′ G-to-A: 0.14; 5′ C-to-T: 0.07; A19885: mean fragment length: 102 bp; 3′ G-to-A: 0.16; 5′ C-to-T: 0.15), supporting their ancient origin (Figure S4.1) (Briggs et al.; 2007). Interestingly, the modern sample showed typical DNA fragmentation that resulting mainly from depurination, while the C-to-T deamination substitutions were minimal. This suggests that depurination and therefore DNA fragmentation occurs rapidly (less than 1 year) after death, while the deamination of cytosine occurs at a slower rate, similar to previous results.

We then constructed a maximum likelihood tree using the consensus sequences of two ancient bison oral strains, one modern oral strain, and all the available genomes of *Actinomyces* strains that have been isolated from animal oral cavity or gut. The second *Actinomyces* strain (MMB1511) was excluded from further analysis due to its limited coverage. The maximum likelihood tree was rooted using *A. liubingyangii*, which is a strain isolated from vulture gut (Meng et al.; 2017). *Actinomyces* strains isolated from mammals fall within two distinct clades: strains isolated from mammal gastrointestinal tract and those isolated from mammal oral cavity. Within the species typical found within the mammalian gut, the phylogeny of these strains was similar to that of their hosts (pika, cattle, sheep and bison), suggesting a deep co-evolutionary relationship between *Actinomyces* and their hosts. The two ancient *A. ruminicola* strains fell within the oral clade and clustered together with modern *A. ruminicola* (An et al.; 2006). In contrast, the modern *A. slckaii* strain MMB 1519 obtained from modern bison clustered together with the *A. slckaii* strain isolated from cattle plaque (Dent and Williams; 1986).

4.3.3   Invertebrate and plant DNA in bison oral samples

The presence of eukaryotic DNA in the bison oral samples was explored by comparing the sequencing data to the NCBI non-redundant protein database. On average, 0.20% to 3.05% of the total sequences were assigned to non-vertebrate eukaryotes (Table S4.4). The modern samples (3%) contained more non-vertebrate eukaryotic DNA than ancient samples (<1%) and were identified in as Alveolates (Apicomplexa), fungi (Ascomycota), arthropods (Arthropoda), roundworms (Nematoda), and plants (Streptophyta). We were also able to detect the presence of several species, including mammalian parasites. *Trichinella* were detected from seven ancient bison oral samples and seven modern bison oral samples. *Plasmodium* was present in three ancient samples, but was absent in any modern samples. Nevertheless, the presence of these parasites needs to be further assessed (*e.g.* themorphological identification of the parasite eggs or key marker genes). Several environmental eukaryotic organisms were also identified, including *Aspergillus* and *Penicillium* (0.07% to 2.24%), *Bombyx moriowere* and *Tribolium castaneum* (0.04% to 0.42%),
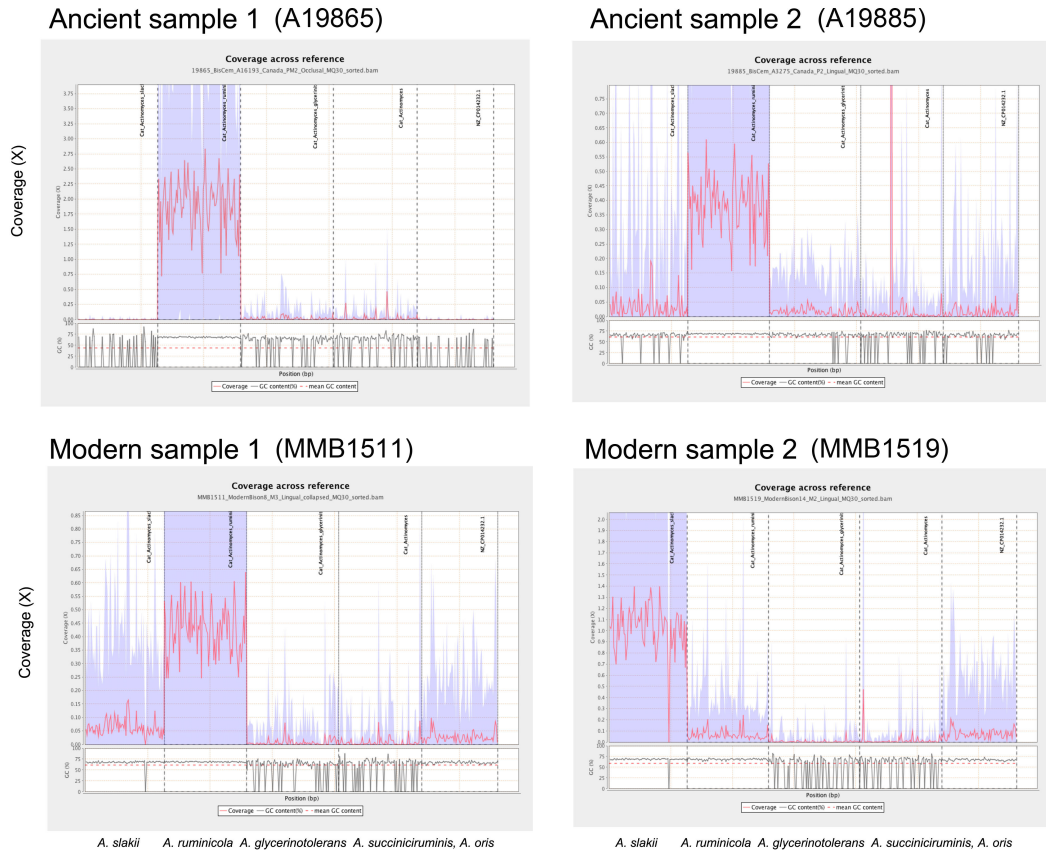
Figure 4.4: Comparative mapping of bison oral samples to *Actinomyces* species. The X axis represent the linear reference genome. Five *Actinomyces* species (from left to right: *A. slackii*, *A. ruminicola*, *A. glycerinotolerans*, *A. succiniciruminis*, and *A. oris*) genomes were concatenated and used as reference genomes. Coverage is indicated using red lines.
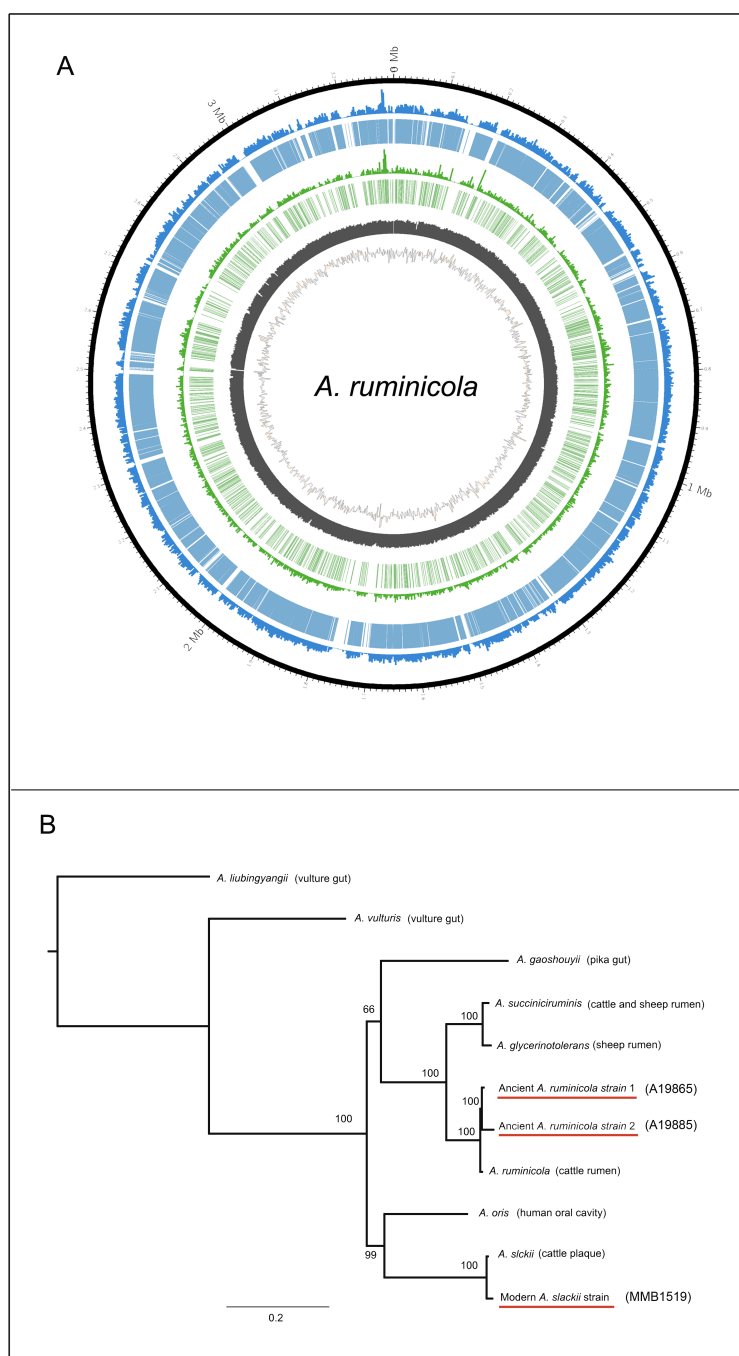
Figure 4.5: Draft genomes and phylogenetic tree of ancient *A. ruminicola* strains. A. Draft genomes of two ancient *A. ruminicola* strains. From inside out: GC skew of the reference genome, GC content of the reference genome, coverage of ancient *A. ruminicola* strain A19865, depth of ancient strain A19865, coverage of ancient *A. ruminicola* strain A19885, depth of ancient strain A19885, and genome scale. B. RAxML tree of *Actinomyces* strains isolated from animal gut or oral cavity. The tree was rooted using *A. liubingyangii* (isolated from vulture gut). Strains assembled from this study are underlined.

Prunus and *Trifolium subterraneum* (0.15% to 0.31%) across modern and ancient samples.

4.3.4 Host DNA in ancient bison oral samples

To estimate the presence and the abundance of host mitochondrial DNA (mtDNA) in the bison oral samples, we mapped all DNA sequences to a bison mitochondrial genome. Bison mtDNA was detected in most of the samples (81% of the ancient bison cementum samples and 85% of the modern bison cementum samples). On average, 0.03% of the total collapsed reads of ancient cementum samples were mapped to the host genome. Bison mtDNA was also detected in ancient plaque samples, although the proportion was lower than in cementum samples; only 0-0.006% of the collapsed reads could be mapped to bison mtDNA from any of the plaque samples. The proportion of bison DNA in modern cementum samples is slightly higher than the ancient samples, with on average 0.15% of collapsed reads mapped to bison mtDNA. Despite the low proportion, the presence of host mtDNA in cementum and plaque samples nonetheless offers an alternative solution for the non-invasive taxonomic identification of valuable samples.

4.4 Discussion

Ancient microbiome research in past non-human animals can reveal information about the commensal microorganisms, bacterial evolution, microbiome-host interactions involved in environmental adaptation and evolution, and unveil information about the past behaviors, diets, and diseases of extinct species (Weyrich et al.; 2015, 2017; Tito et al.; 2012). Currently, ancient oral microbiome studies are limited to human dental calculus, due to the prevalence of dental calculus on ancient human remains and extensive research into modern human oral microbiomes (Weyrich et al.; 2015; Wade; 2013). Obtaining reliable non-human oral microbiome information requires the exploration of sub-fossil materials that are resilient to taphonomy and easily accessible in the paleontological and archeological collections. While the soft tissues can rapidly decomposed after the death, compacted tissues, such as bone and teeth, can preserve the host and microbial DNA for thousands of years (Hofreiter et al.; 2001; Rohland and Hofreiter; 2007; Mann et al.; 2018). This is the first study to explore the survival of oral microbial DNA in ancient non-human specimens.

By applying protocols used for the analyses of the ancient human oral microbiome, we describe for the first time several oral microorganisms that can be recovered from ancient bison cementum and plaque samples. From ancient bison teeth, we were able to obtain microbes that likely originate from several sources, including

1) typical oral taxa, including taxa that mediate the formation of biofilms in the oral cavity (Kolenbrander et al.; 2010), such as *Streptococcus* and *Actinomyces*; 2) species that inhabit the bovine pharyngeal mucosal (Timsit et al.; 2016), such as *Pasteurella* and *Psychrobacter*; and 3) ruminal methanogens (Sirohi et al.; 2010), such as *Methanobrevibacter* and *Methanococcales*. Using competitive mapping, we were also able to reconstruct the two 33kyr old *A. ruminicola* draft genome, suggesting that these oral species can survive for millennia and likely have deep co-evolutionary histories with their hosts. Ancient bacterial genomes of commensal species in calculus can serve as an unique source of past oral microorgansisms to examine the co-evolutionary history of non-human mammals and their microbes and the implications for extinction, health, and physiology.

The recovery of ancient oral microbiome information has important implications for better understanding animal conservation, evolution, and health (Guerrero et al.; 2013; Warinner et al.; 2015). For instance, it is not clear how a population bottleneck of host species would impact the relationship with their commensal microbes, and if these changes had subsequent physiological consequences. In this study, we found 18 *Actinomyces* species in ancient bison, while two separate species were identified in modern bison – one of these species was phylogenetically more similar to that found in modern cattle. A further comparison of the microorganisms from many modern animal and their ancient microbial counterparts is needed to better address these questions. Furthermore, the unique digestive habits of ruminants – the eructation and rumination – make them a special case which both the oral and stomach microorganisms are present in their oral cavity (DePeters and George; 2014), and thus be potentially preserved in ancient ruminant teeth. As a result, ancient ruminant teeth may serve as a unprecedented material for the recovery of both the oral and gut microbiomes. In addition, the ability to detection of eukaryotic DNA, including the host mtDNA, putative dietary DNA (plants), and putative parasites DNA, from the ancient samples is also an unprecedented opportunity to investigation of the diets, parasitology, and behavior of the past animals.

We identified several limitations in the application of current paleomicrobiology methods to ancient ruminant teeth. Despite robust oral microbiome information obtained from dental calculus of ancient humans (Adler et al.; 2013; Weyrich et al.; 2017, 2015), we found that the approach applied to human dental calculus was not always appropriate when assessing ancient ruminant teeth. This is mainly due to the differences in tooth structure and oral biochemistry between human and ruminants. First, it remains unclear why the cementum sample preserved an oral signal. Is this presence of oral microorganisms observed because they can bind or fuse with cementum or does the cementum itself incorporates microbial DNA? Second, the "plaque-like" tissue seems to preserve strong oral microbiome signal, but the condi-

tions that allow the formation and encourage preservation of that material remain unknown. In addition, the lack of calcified dental plaque on these specimens is also unclear. Future studies should examine how the microstructures present in and on bison teeth are subject to taphonomic processes. Third, decontaminating oral samples using 3% bleach does not appear to be suitable for all ancient oral samples, as we observed the disintegration of several ancient plaque samples with bleach treatment. We found that a milder decontamination approach (an EDTA wash instead of bleach) was more appropriate for all sample types. Lastly, a large proportion of the metagenomic information obtained from ancient samples has environmental origins (*e.g.* DNA from bacteria that are widely distributed in soil, such as Burkholderiales and Rhizobiales), despite the use of decontamination methods. As human dental calculus is more robust, the bison cementum and plaque samples examined here are likely to be more susceptible to contamination. Further research and the novel application of common ancient DNA methods, such as hybridization enrichment, may improve the current limitations in assessing the microbial content of non-calculus oral specimens.

This study investigated the ancient DNA damage signatures from bison oral samples experience different taphonomic processes, ranging from one-year of environmental exposure to >50 kyrs. We found that some samples contained no oral microbial signal within a period as short as one year, while oral signals were routinely obtained from ˜33 kyr samples, suggesting that other factors, and not age, likely play an important role in the preservation of the oral microbial DNA. Among the oral species that were well preserved, *Actinomyces* species maintain a relatively high GC content (˜69%). As DNA molecules with high GC content tend have higher thermostabilities than GC-poor molecules (Vinogradov; 2003), the high GC content of specific microorganisms could aid in their long-term preservation. We also examined the DNA damage of three reconstructed *Actinomyces* genomes (Figure mapdamage). Within the 33 kyr samples, patterns consistent with ancient DNA damage (fragmentation, depurination, and deamination) were observed in two draft bacterial genomes. In contrast, the modern bacterial genome only contained evidence of DNA fragmentation and depurination, with minimal deamination. While previous studies have observed DNA fragmentation in samples that only a decade old (Sawyer et al.; 2012), we show that one-year exposure can lead to DNA fragmentation, whereas extensive deamination seems to occur with longer exposure to the environment. This offers some clues for distinguishing recently introduced contaminant DNA from authentic ancient DNA.

In conclusion, our results show that typical oral, mucosal, and ruminal microbes, dietary information, and host DNA can be recovered from ancient bison cementum and plaque specimens. Therefore, ancient ruminant oral samples serves as a promis-

ing new proxy for recovering information on past behavior, disease, diet, and host information from non-human animals. This research in the future will improve our understanding the evolutionary processes and signatures that drive microbial evolution and co-adaptation within hosts.

4.5 Acknowledgement

We would like to thank Dr.Michael C. Wilson for his kind help with the morphological identification of ancient bison tooth structure.

Bibliography

A, E. R. (2018). *New and Refined Tools and Guidelines to Expand the Scope and Improve the Reproducibility of Palaeomicrobiological Research*, PhD thesis, School of Biological Sciences, The University of Adelaide. An optional note.

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A., Alt, K. W. et al. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the neolithic and industrial revolutions, *Nature genetics* **45**(4): 450.

Alfano, N., Courtiol, A., Vielgrader, H., Timms, P., Roca, A. L. and Greenwood, A. D. (2015). Variation in koala microbiomes within and between individuals: effect of body region and captivity status, *Scientific reports* **5**: 10189.

An, D., Cai, S. and Dong, X. (2006). Actinomyces ruminicola sp. nov., isolated from cattle rumen, *International journal of systematic and evolutionary microbiology* **56**(9): 2043–2048.

Bik, E. M., Costello, E. K., Switzer, A. D., Callahan, B. J., Holmes, S. P., Wells, R. S., Carlin, K. P., Jensen, E. D., Venn-Watson, S. and Relman, D. A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea, *Nature communications* **7**: 10516.

Bowen, W. H., Burne, R. A., Wu, H. and Koo, H. (2018). Oral biofilms: pathogens, matrix, and polymicrobial interactions in microenvironments, *Trends in microbiology* **26**(3): 229–242.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. et al. (2007). Patterns of damage in genomic dna sequences from a neandertal, *Proceedings of the National Academy of Sciences* **104**(37): 14616–14621.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adler, C. J., Richards, S. M., Der Sarkissian, C., Ganslmeier, R., Friederich, S. et al. (2013). Neolithic mitochondrial haplogroup h genomes and the genetic origins of europeans, *Nature communications* **4**: 1764.

Buchfink, B., Xie, C. and Huson, D. H. (2015). Fast and sensitive protein alignment using diamond, *Nature methods* **12**(1): 59.

Cooper, A. and Poinar, H. N. (2000). Ancient dna: do it right or not at all, *Science* **289**(5482): 1139–1139.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. and Bradshaw, C. J. (2015). Abrupt warming events drove late pleistocene holarctic megafaunal turnover, *Science* **349**(6248): 602–606.

Dent, V. and Williams, R. (1986). Actinomyces slackii sp. nov. from dental plaque of dairy cattle, *International Journal of Systematic and Evolutionary Microbiology* **36**(3): 392–395.

DePeters, E. and George, L. (2014). Rumen transfaunation, *Immunology letters* **162**(2): 69–76.

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W.-H., Lakshmanan, A. and Wade, W. G. (2010). The human oral microbiome, *Journal of bacteriology* **192**(19): 5002–5017.

Dewhirst, F. E., Klein, E. A., Thompson, E. C., Blanton, J. M., Chen, T., Milella, L., Buckley, C. M., Davis, I. J., Bennett, M.-L. and Marshall-Jones, Z. V. (2012). The canine oral microbiome, *PloS one* **7**(4): e36067.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns, *Proceedings of the National Academy of Sciences* **107**(26): 11971–11975.

Farrer, A. G. (2016). *Ancient DNA studies of dental calculus*, PhD thesis, School of Biological Sciences, The University of Adelaide. An optional note.

Gates, C. C., Freese, C. H., Gogan, P. J. and Kotzman, M. (2010). *American bison: status survey and conservation guidelines 2010*, IUCN.

Guerrero, R., Margulis, L., Berlanga, M. et al. (2013). Symbiogenesis: the holobiont as a unit of evolution, *Int Microbiol* **16**(3): 133–143.

Hartl, G. B. and Pucek, Z. (1994). Genetic depletion in the european bison (bison bonasus) and the significance of electrophoretic heterozygosity for conservation, *Conservation biology* **8**(1): 167–174.

Henssge, U., Do, T., Radford, D. R., Gilbert, S. C., Clark, D. and Beighton, D. (2009). Emended description of actinomyces naeslundii and descriptions of actinomyces oris sp. nov. and actinomyces johnsonii sp. nov., previously identified as actinomyces naeslundii genospecies 1, 2 and wva 963, *International journal of systematic and evolutionary microbiology* **59**(Pt 3): 509.

Herbig, A., Maixner, F., Bos, K. I., Zink, A., Krause, J. and Huson, D. H. (2016). Malt: Fast alignment and analysis of metagenomic dna sequence data applied to the tyrolean iceman, *BioRxiv* p. 050559.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M. and Pääbo, S. (2001). ancient dna, *Nature Reviews Genetics* **2**(5): 353.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J. and Tappu, R. (2016). Megan community edition-interactive exploration and analysis of large-scale microbiome sequencing data, *PLoS computational biology* **12**(6): e1004957.

Hyde, E. R., Luk, B., Cron, S., Kusic, L., McCue, T., Bauch, T., Kaplan, H., Tribble, G., Petrosino, J. F. and Bryan, N. S. (2014). Characterization of the rat oral microbiome and the effects of dietary nitrate, *Free Radical Biology and Medicine* **77**: 249–257.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. and Orlando, L. (2013). mapdamage2. 0: fast approximate bayesian estimates of ancient dna damage parameters, *Bioinformatics* **29**(13): 1682–1684.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics* **28**(12): 1647–1649.

Kemp, B. M. and Smith, D. G. (2005). Use of bleach to eliminate contaminating dna from the surface of bones and teeth, *Forensic science international* **154**(1): 53–61.

Koenigswald, W. V. (2011). Diversity of hypsodont teeth in mammalian dentitions—construction and classification, *Palaeontogr. Abt. A* **294**: 63–94.

Kolenbrander, P. E., Palmer Jr, R. J., Periasamy, S. and Jakubovics, N. S. (2010). Oral multispecies biofilm development and the key role of cell–cell distance, *Nature Reviews Microbiology* **8**(7): 471.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics, *Genome research* **19**(9): 1639–1645.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform, *bioinformatics* **25**(14): 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). The sequence alignment/map format and samtools, *Bioinformatics* **25**(16): 2078–2079.

Lyman, R. L. (2018). Dental enamel hypoplasias in holocene bighorn sheep (ovis canadensis) in eastern washington state, usa, *Canadian Journal of Zoology* **96**(5): 460–465.

Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Yates, J. A. F., Salazar-García, D. C. et al. (2018). Differential preservation of endogenous human and microbial dna in dental calculus and dentin, *Scientific reports* **8**.

Meng, X., Lu, S., Lai, X.-H., Wang, Y., Wen, Y., Jin, D., Yang, J. and Xu, J. (2017). Actinomyces liubingyangii sp. nov. isolated from the vulture gypaetus barbatus, *International journal of systematic and evolutionary microbiology* **67**(6): 1873–1879.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing, *Cold Spring Harbor Protocols* **2010**(6): pdb–prot5448.

Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z. and Dominguez-Bello, M. G. (2015). The infant microbiome development: mom matters, *Trends in molecular medicine* **21**(2): 109–117.

NA, S. P., Pristaš, P., Hrehová, L., Javorskỳ, P., Stams, A. J. and Plugge, C. M. (2016). Actinomyces succiniciruminis sp. nov. and actinomyces glycerinitolerans sp. nov., two novel organic acid-producing bacteria isolated from rumen, *Systematic and applied microbiology* **39**(7): 445–452.

Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2015). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data, *Bioinformatics* **32**(2): 292–294.

Rohland, N. and Hofreiter, M. (2007). Ancient dna extraction from bones and teeth, *Nature protocols* **2**(7): 1756.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and dna fragmentation in ancient dna, *PloS one* **7**(3): e34131.

Schubert, M., Lindgreen, S. and Orlando, L. (2016). Adapterremoval v2: rapid adapter trimming, identification, and read merging, *BMC research notes* **9**(1): 88.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J. et al. (2004). Rise and fall of the beringian steppe bison, *Science* **306**(5701): 1561–1565.

Sirohi, S., Pandey, N., Singh, B. and Puniya, A. (2010). Rumen methanogens: a review, *Indian journal of microbiology* **50**(3): 253–262.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y., Kosintsev, P., Lee, M. S., Baryshnikov, G. et al. (2016). Early cave art and ancient dna record the origin of european bison, *Nature communications* **7**: 13158.

Stamatakis, A. (2014). Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics* **30**(9): 1312–1313.

Stromberg, C. A. (2006). Evolution of hypsodonty in equids: testing a hypothesis of adaptation, *Paleobiology* **32**(2): 236–258.

Timsit, E., Holman, D. B., Hallewell, J. and Alexander, T. W. (2016). The nasopharyngeal microbiota in feedlot cattle and its role in respiratory health, *Animal Frontiers* **6**(2): 44–50.

Tito, R. Y., Knights, D., Metcalf, J., Obregon-Tito, A. J., Cleeland, L., Najar, F., Roe, B., Reinhard, K., Sobolik, K., Belknap, S. et al. (2012). Insights from characterizing extinct human gut microbiomes, *PloS one* **7**(12): e51146.

Vinogradov, A. E. (2003). Dna helix: the importance of being gc-rich, *Nucleic acids research* **31**(7): 1838–1844.

Wade, W. G. (2013). The oral microbiome in health and disease, *Pharmacological research* **69**(1): 137–143.

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S. et al. (2014). Pathogens and host immunity in the ancient human oral cavity, *Nature genetics* **46**(4): 336.

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): 20130376.

Weyrich, L., Farrer, A. G., Eisenhofer, R., Arriola, L. A., Young, J., Selway, C. A., Handsley-Davis, M., Adler, C., Breen, J. and Cooper, A. (2018). Laboratory contamination over time during low-biomass sample analysis, *BioRxiv* p. 460212.

Weyrich, L. S., Dobney, K. and Cooper, A. (2015). Ancient dna analysis of dental calculus, *Journal of Human Evolution* **79**: 119–124.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V. et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus, *Nature* **544**(7650): 357.

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., Paczian, T., Trimble, W. L., Bagchi, S., Grama, A. et al. (2015). The mg-rast metagenomics database and portal in 2015, *Nucleic acids research* **44**(D1): D590–D594.

Wilson, M. C. (1988). Bison dentitions from the henry smith site, montana: evidence for seasonality and paleoenvironments at an avonlea bison kill, *Avonlea Yesterday and Today: Archaeology and Prehistory, Saskatchewan Archaeological Society, Saskatoon* pp. 203–225.

4.6 Supplementary materials

4.6.1 Figure S4.1

A. Ancient *A. ruminicola* strain 1 (A19865)
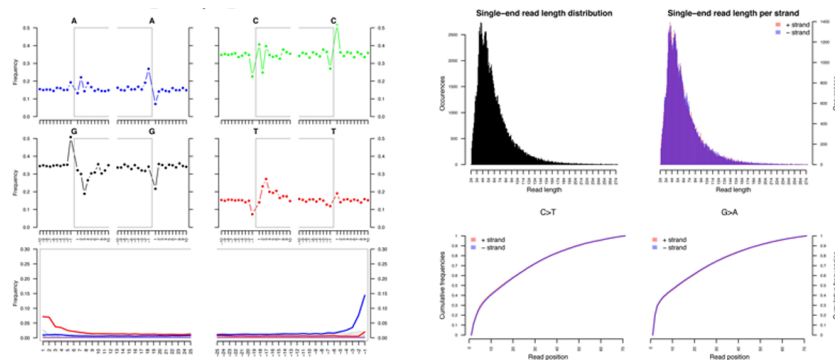
B. Ancient *A. ruminicola* strain 2 (A19885)
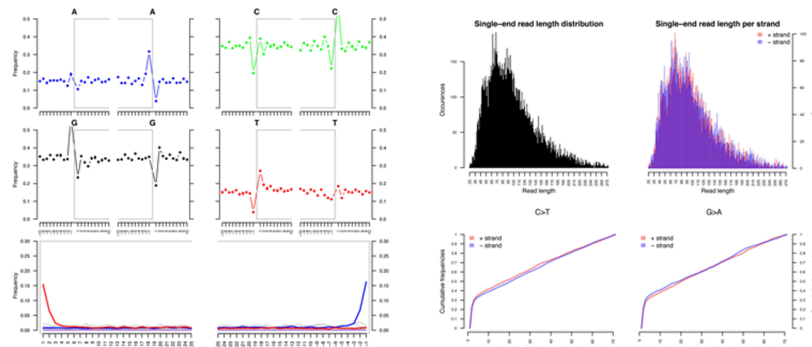
C. Modern *A. slackii* strain (MMB1519)
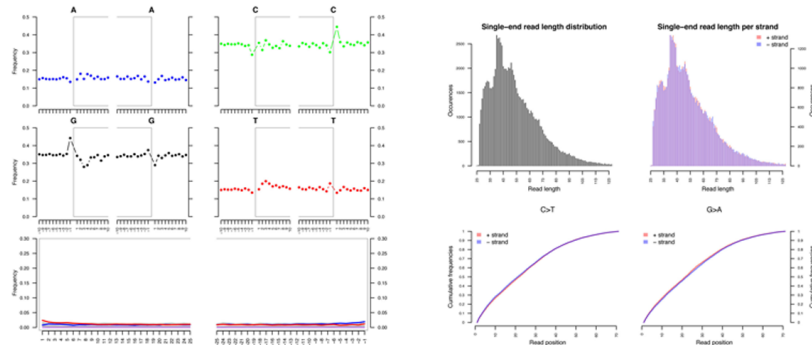
The MapDamage plots of the assembled *Actinomyces* genomes. A and B. Ancient *A. ruminicola* strains. The C-to-T and G-to-A substitution rates increased at the end of the sequences. The length of the majority of the mapped sequences is <100 bp. C. The modern *A. slackii* strain. No evident C-to-T and G-to-A substitutions was observed, while the sequences are highly fragmented, with an average length about 50 bp.

4.6.2 Table S4.1

Sample details.

#SampleID	Decon method	Category		Tooth	Surface	Age	Indivdual	Sampling location	
A19366	UV-bleach	Cementum	Ancient	NA	Buccal	infinite	A626	Canada	Yukon
A19372	UV-bleach	Cementum	Ancient	NA	Occlusal	infinite	A626	Canada	Yukon
A19373	UV-bleach	Cementum	Ancient	NA	Buccal	6k	A3033	Canada	Yukon
A19374	UV-bleach	Cementum	Ancient	NA	Occlusal	6k	A3033	Canada	Yukon
A19375	UV-bleach	Cementum	Ancient	NA	Occlusal	6k	A3033	Canada	Yukon
A19376	UV-bleach	Cementum	Ancient	NA	Lingual	3k	A1538	Canada	Alberta
A19377	UV-bleach	Cementum	Ancient	NA	Occlusal	3k	A1538	Canada	Alberta
A19378	UV-bleach	Cementum	Ancient	NA	Occlusal	3k	A1538	Canada	Alberta
A19380	UV-bleach	Cementum	Ancient	NA	Occlusal	6k	A3030	Canada	Yukon
A19382	UV-bleach	Cementum	Ancient	NA	Occlusal	7k	A3021	USA	NA
A19383	UV-bleach	Cementum	Ancient	NA	Interproximal	7k	A3021	USA	NA
A19385	UV-bleach	Cementum	Ancient	NA	NA	2k	A1465	Canada	NA
A19391	UV-bleach	Cementum	Ancient	NA	Interproximal	infinite	A626	Canada	Yukon
A19841	UV-bleach	Cementum	Ancient	NA	NA	24k	A3130	Canada	Yukon
A19842	UV-bleach	Cementum	Ancient	M1	Lingual	24k	A3130	Canada	Yukon
A19843	UV-bleach	Cementum	Ancient	M3	Buccal	24k	A3130	Canada	Yukon
A19844	UV-bleach	Cementum	Ancient	M2	Lingual	24k	A3130	Canada	Yukon
A19845	UV-bleach	Cementum	Ancient	NA	NA	24k	A3130	Canada	Yukon
A19846	UV-bleach	Cementum	Ancient	M2	Buccal	24k	A3130	Canada	Yukon
A19847	UV-bleach	Cementum	Ancient	M3	Lingual	24k	A3130	Canada	Yukon
A19848	UV-bleach	Cementum	Ancient	NA	NA	49k	A3131	Canada	Yukon
A19849	UV-bleach	Cementum	Ancient	M2	Lingual	49k	A3131	Canada	Yukon
A19850	UV-bleach	Cementum	Ancient	M3	Lingual	49k	A3131	Canada	Yukon
A19851	UV-bleach	Cementum	Ancient	M3	Lingual	49k	A3131	Canada	Yukon
A19852	UV-bleach	Cementum	Ancient	M1	Interproximal	49k	A3131	Canada	Yukon
A19853	UV-bleach	Cementum	Ancient	NA	NA	33k	A16193	Canada	Yukon
A19854	UV-bleach	Cementum	Ancient	M3	Lingual	33k	A16193	Canada	Yukon
A19855	UV-bleach	Cementum	Ancient	M3	Buccal	33k	A16193	Canada	Yukon
A19857	UV-bleach	Cementum	Ancient	M2	Buccal	33k	A16193	Canada	Yukon
A19858	UV-bleach	Cementum	Ancient	M1	Buccal	33k	A16193	Canada	Yukon
A19861	UV-bleach	Cementum	Ancient	PM2	Buccal	33k	A16193	Canada	Yukon
A19862	UV-bleach	Cementum	Ancient	PM3	Buccal	33k	A16193	Canada	Yukon
A19863	UV-bleach	Cementum	Ancient	PM1	Buccal	33k	A16193	Canada	Yukon
A19864	UV-bleach	Cementum	Ancient	PM2	Occlusal	33k	A16193	Canada	Yukon
A19865	UV-bleach	Cementum	Ancient	PM2	Occlusal	33k	A16193	Canada	Yukon
A19867	UV-bleach	Cementum	Ancient	M1	Lingual	33k	A16193	Canada	Yukon
A19868	UV-bleach	Cementum	Ancient	M3	Occlusal	33k	A16193	Canada	Yukon
A19869	UV-bleach	Cementum	Ancient	M2	Occlusal	33k	A16193	Canada	Yukon
A19870	UV-bleach	Cementum	Ancient	PM2	Occlusal	33k	A16193	Canada	Yukon
A19871	UV-bleach	Cementum	Ancient	PM3	Occlusal	33k	A16193	Canada	Yukon
A19872	UV-bleach	Cementum	Ancient	M2	Lingual	33k	A16193	Canada	Yukon
A19873	UV-bleach	Cementum	Ancient	NA	Buccal	30k	A16186	Canada	Yukon
A19874	UV-bleach	Cementum	Ancient	NA	Lingual	30k	A16186	Canada	Yukon
A19875	UV-bleach	Cementum	Ancient	NA	Occlusal	30k	A16186	Canada	Yukon
A19876	UV-bleach	Cementum	Ancient	NA	Buccal	30k	A16186	Canada	Yukon
A19877	UV-bleach	Cementum	Ancient	NA	Root	30k	A16186	Canada	Yukon
A19878	UV-bleach	Cementum	Ancient	NA	Buccal	30k	A16186	Canada	Yukon
A19879	UV-bleach	Cementum	Ancient	NA	Buccal	28k	A16192	Canada	Yukon
A19881	UV-bleach	Cementum	Ancient	NA	Lingual	28k	A16192	Canada	Yukon
A19882	UV-bleach	Cementum	Ancient	NA	Interproximal	28k	A16192	Canada	Yukon
A19883	UV-bleach	Cementum	Ancient	NA	Interproximal	28k	A16192	Canada	Yukon
A19884	UV-bleach	Cementum	Ancient	M3	Lingual	33k	A16193	Canada	Yukon

A19880	UV-EDTA	Plaque	Ancient	P2	Lingual	33k	A3275	Canada	Yukon
A19885	UV-EDTA	Plaque	Ancient	P2	Lingual	33k	A3275	Canada	Yukon
A19886	UV-EDTA	Plaque	Ancient	M1	Lingual	33k	A3275	Canada	Yukon
A19887	UV-EDTA	Plaque	Ancient	PM3	Lingual	33k	A3275	Canada	Yukon
A19888	UV-EDTA	Plaque	Ancient	M2	Lingual	33k	A3275	Canada	Yukon
A19889	UV-EDTA	Plaque	Ancient	M1	Buccal	33k	A3275	Canada	Yukon
A20106	UV-EDTA	Plaque	Ancient	PM3	Lingual	33k	A3275	Canada	Yukon
A20107	UV-EDTA	Plaque	Ancient	M1	Lingual	33k	A3275	Canada	Yukon
A20108	UV-EDTA	Plaque	Ancient	M2	Lingual	33k	A3275	Canada	Yukon
A20109	UV-EDTA	Plaque	Ancient	M3	Lingual	33k	A3275	Canada	Yukon
A20110	UV-EDTA	Plaque	Ancient	M3	Buccal	33k	A3275	Canada	Yukon
1501B	UV-bleach	Cementum	Modern	M2	Lingual	0	MB1	USA	South Dakota
1502B	UV-bleach	Cementum	Modern	M2	Interproximal	0	MB1	USA	South Dakota
1509B	UV-bleach	Cementum	Modern	M2	Buccal	0	MB7	USA	South Dakota
1512B	UV-bleach	Cementum	Modern	M1M2	Interproximal	0	MB8	USA	South Dakota
1513B	UV-bleach	Cementum	Modern	M2	Buccal	0	MB9	USA	South Dakota
1514B	UV-bleach	Cementum	Modern	M2	Buccal	0	MB10	USA	South Dakota
1515B	UV-bleach	Cementum	Modern	M1M2	Interproximal	0	MB11	USA	South Dakota
1516B	UV-bleach	Cementum	Modern	M3	Lingual	0	MB11	USA	South Dakota
1523B	UV-bleach	Cementum	Modern	M1	Buccal	0	MB18	USA	South Dakota
MMB1501	UV-EDTA	Cementum	Modern	M2	Lingual	0	MB1	USA	South Dakota
MMB1504	UV-EDTA	Cementum	Modern	M2	Lingual	0	MB2	USA	South Dakota
MMB1505	UV-EDTA	Cementum	Modern	PM2	Lingual	0	MB3	USA	South Dakota
MMB1506	UV-EDTA	Cementum	Modern	M3	Lingual	0	MB4	USA	South Dakota
MMB1507	UV-EDTA	Cementum	Modern	M1	Buccal	0	MB6	USA	South Dakota
MMB1508	UV-EDTA	Cementum	Modern	PM1	Lingual	0	MB7	USA	South Dakota
MMB1511	UV-EDTA	Cementum	Modern	M3	Lingual	0	MB8	USA	South Dakota
MMB1513	UV-EDTA	Cementum	Modern	M2	Buccal	0	MB9	USA	South Dakota
MMB1514	UV-EDTA	Cementum	Modern	M2	Buccal	0	MB10	USA	South Dakota
MMB1516	UV-EDTA	Cementum	Modern	M3	Lingual	0	MB11	USA	South Dakota
MMB1517	UV-EDTA	Cementum	Modern	M1	Buccal	0	MB12	USA	South Dakota
MMB1518	UV-EDTA	Cementum	Modern	M1	Buccal	0	MB13	USA	South Dakota
MMB1519	UV-EDTA	Cementum	Modern	M3	Lingual	0	MB14	USA	South Dakota
MMB1520	UV-EDTA	Cementum	Modern	M2	Buccal	0	MB15	USA	South Dakota
MMB1521	UV-EDTA	Cementum	Modern	M2	Lingual	0	MB16	USA	South Dakota
MMB1522	UV-EDTA	Cementum	Modern	M3	Buccal	0	MB17	USA	South Dakota
MMB1523	UV-EDTA	Cementum	Modern	M1	Buccal	0	MB18	USA	South Dakota
A19392	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A19393	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A19893	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A19894	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A19899	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A19900	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A20115	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A20115 E	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A20116	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A20116 E	NA	EBC	NA	NA	NA	NA	NA	NA	NA
MMB1524	NA	EBC	NA	NA	NA	NA	NA	NA	NA
MMB1525	NA	EBC	NA	NA	NA	NA	NA	NA	NA
MMB1526	NA	EBC	NA	NA	NA	NA	NA	NA	NA
MMB1527	NA	EBC	NA	NA	NA	NA	NA	NA	NA
A750	NA	Bone	Ancient	NA	NA	NA	NA	NA	Canada
A757	NA	Bone	Ancient	NA	NA	NA	NA	NA	Alaska
A770	NA	Bone	Ancient	NA	NA	NA	NA	NA	Canada
Air.mgm451(	NA	Environmental :	Modern	NA	NA	NA	NA	NA	NA
BraccishWat(	NA	Environmental :	Modern	NA	NA	NA	NA	NA	NA
FreshGround	NA	Environmental :	Modern	NA	NA	NA	NA	NA	NA
GrasslandSoi	NA	Environmental :	Modern	NA	NA	NA	NA	NA	NA

4.6.3 Table S4.2

The 50 most abundant bacterial orders detected from ancient and modern bison oral samples.

UV-EDTA						UV-Bleach					
Ancient			Modern			Ancient			Modern		
Bacterial order	Mean	SD	Bacterial order	Mean	SD	Bacterial order	Mean	SD	Bacterial order	Mean	SD
Actinomycetales	15.80%	16.68%	Micrococcales	20.43%	20.10%	Burkholderiales	22.57%	19.53%	Lactobacillales	19.37%	12.99%
Pasteurellales	11.97%	10.92%	Actinomycetales	15.00%	18.67%	Rhizobiales	16.89%	16.07%	Micromonosporales	14.46%	13.54%
Micrococcales	7.81%	8.03%	Pseudomonadales	7.93%	10.62%	Streptomycetales	13.71%	23.81%	Pseudomonadales	9.87%	10.00%
Clostridiales	6.84%	5.59%	Burkholderiales	6.18%	5.46%	Pseudomonadales	7.73%	11.51%	Micrococcales	9.59%	8.93%
Burkholderiales	6.30%	4.90%	Corynebacteriales	4.82%	4.90%	Micrococcales	6.13%	10.50%	Bacillales	9.10%	5.36%
Pseudomonadales	6.28%	7.74%	Lactobacillales	4.80%	4.71%	Propionibacteriales	2.37%	2.03%	Actinomycetales	6.54%	16.97%
Rhizobiales	4.97%	3.41%	Micromonosporales	4.73%	8.33%	Flavobacteriales	2.29%	2.36%	Enterobacterales	5.74%	3.38%
Flavobacteriales	3.41%	2.50%	Bacillales	3.84%	3.42%	Corynebacteriales	1.99%	3.18%	Fusobacteriales	5.45%	3.63%
Propionibacteriales	3.23%	2.34%	Propionibacteriales	3.49%	4.54%	Myxococcales	1.97%	7.21%	Burkholderiales	2.80%	3.58%
Streptomycetales	3.09%	4.33%	Pasteurellales	2.86%	6.11%	Sphingomonadales	1.63%	1.10%	Xanthomonadales	2.71%	5.27%
Lactobacillales	2.90%	1.82%	Enterobacterales	2.51%	3.17%	Lactobacillales	1.43%	3.77%	Corynebacteriales	2.57%	1.02%
Bacteroidales	2.81%	3.57%	Rhizobiales	2.43%	2.42%	Bacillales	1.28%	1.54%	Clostridiales	2.20%	0.87%
Corynebacteriales	2.25%	1.89%	Streptomycetales	1.64%	1.09%	Pseudonocardiales	1.20%	0.91%	Pasteurellales	1.85%	5.30%
Sphingomonadales	2.00%	2.61%	Xanthomonadales	1.62%	3.77%	Rhodospirillales	1.19%	0.72%	Propionibacteriales	1.48%	3.88%
Coriobacteriales	1.98%	4.67%	Caulobacterales	1.55%	2.63%	Clostridiales	1.13%	1.54%	Pseudonocardiales	1.48%	1.26%
Enterobacterales	1.85%	1.66%	Cytophagales	1.35%	1.76%	Rhodobacterales	1.12%	0.80%	Caulobacterales	1.26%	2.08%
Cytophagales	1.38%	2.75%	Fusobacteriales	1.23%	1.59%	Micromonosporales	1.02%	0.93%	Cytophagales	0.84%	2.01%
Methanobacteriales	1.30%	1.41%	Deinococcales	1.10%	2.50%	Xanthomonadales	0.96%	0.62%	Rhizobiales	0.79%	1.19%
Fusobacteriales	1.17%	1.52%	Pseudonocardiales	1.10%	1.07%	Caulobacterales	0.88%	0.99%	Streptomycetales	0.50%	0.75%
Bacillales	1.04%	0.60%	Clostridiales	1.05%	1.46%	Nitrosomonadales	0.77%	1.44%	Flavobacteriales	0.39%	0.78%
Pseudonocardiales	0.92%	0.78%	Sphingobacteriales	0.97%	3.08%	Desulfuromonadales	0.76%	1.54%	Chlamydiales	0.22%	0.65%
Neisseriales	0.89%	0.76%	Sphingomonadales	0.83%	1.41%	Enterobacterales	0.71%	0.42%	Neisseriales	0.21%	0.41%
Caulobacterales	0.85%	1.48%	Neisseriales	0.77%	1.51%	Rhodocyclales	0.70%	0.69%	Sphingomonadales	0.19%	0.38%
Synechococcales	0.74%	2.44%	Streptosporangiales	0.75%	1.25%	Bacteroidales	0.56%	1.16%	Solirubrobacterales	0.18%	0.53%
Micromonosporales	0.73%	0.82%	Flavobacteriales	0.74%	1.21%	Streptosporangiales	0.48%	0.32%	Deinococcales	0.12%	0.35%
Xanthomonadales	0.61%	0.65%	Rhodobacterales	0.69%	0.92%	Desulfovibrionales	0.46%	1.00%	Rhodobacterales	0.10%	0.31%
Rhodospirillales	0.53%	0.41%	Myxococcales	0.65%	1.92%	Coriobacteriales	0.45%	2.52%	Campylobacterales	0.01%	0.01%
Eggerthellales	0.48%	0.88%	Nitrosomonadales	0.54%	2.28%	Actinomycetales	0.44%	1.67%	Rickettsiales	0.00%	0.00%
Rhodobacterales	0.46%	0.30%	Geodermatophilales	0.53%	1.14%	Cytophagales	0.38%	0.38%	Acidobacteriales	0.00%	0.00%
Pelagibacterales	0.42%	1.39%	Rhodospirillales	0.50%	0.85%	Campylobacterales	0.35%	0.62%	Holophagales	0.00%	0.00%
Streptosporangiales	0.39%	0.25%	Bacteroidales	0.36%	0.81%	Oceanospirillales	0.35%	1.01%	Calditrichales	0.00%	0.00%
Campylobacterales	0.31%	0.67%	Solirubrobacterales	0.35%	0.71%	Chromatiales	0.30%	0.29%	Deferribacterales	0.00%	0.00%
Sphingobacteriales	0.28%	0.43%	Desulfuromonadales	0.32%	1.12%	Chitinophagales	0.30%	0.35%	Bacteroidetes Order II. I	0.00%	0.00%
Myxococcales	0.28%	0.27%	Campylobacterales	0.28%	1.05%	Sphingomonadales	0.29%	0.39%	Bacteroidales	0.00%	0.00%
Planctomycetales	0.26%	0.42%	Nostocales	0.27%	1.19%	Desulfobacterales	0.29%	0.56%	Marinilabiliales	0.00%	0.00%
Desulfovibrionales	0.26%	0.39%	Rhodocyclales	0.17%	0.42%	Planctomycetales	0.29%	0.31%	Chitinophagales	0.00%	0.00%
Nitrosomonadales	0.20%	0.30%	Opitutales	0.11%	0.52%	Neisseriales	0.28%	0.34%	Sphingobacteriales	0.00%	0.00%
Chitinophagales	0.20%	0.23%	Desulfobacterales	0.10%	0.33%	Solirubrobacterales	0.27%	0.26%	Chlorobiales	0.00%	0.00%
Geodermatophilales	0.19%	0.30%	Oscillatoriales	0.09%	0.35%	Nitrospirales	0.26%	0.56%	Ignavibacteriales	0.00%	0.00%
Solirubrobacterales	0.19%	0.20%	Frankiales	0.09%	0.40%	Syntrophobacterales	0.25%	0.65%	Gemmatimonadales	0.00%	0.00%
Frankiales	0.18%	0.23%	Desulfovibrionales	0.08%	0.30%	Methanobacteriales	0.23%	0.73%	Nitrospirales	0.00%	0.00%
Legionellales	0.15%	0.36%	Verrucomicrobiales	0.08%	0.36%	Selenomonadales	0.21%	0.91%	Pelagibacterales	0.00%	0.00%
Vibrionales	0.15%	0.13%	Oceanospirillales	0.07%	0.26%	Alteromonadales	0.19%	0.57%	Rhodospirillales	0.00%	0.00%
Alteromonadales	0.14%	0.14%	Chitinophagales	0.07%	0.21%	Frankiales	0.18%	0.18%	Nitrosomonadales	0.00%	0.00%
Chromatiales	0.12%	0.23%	Erysipelotrichales	0.07%	0.35%	Acidobacteriales	0.17%	0.30%	Rhodocyclales	0.00%	0.00%
Nitrospirales	0.12%	0.18%	Thermales	0.07%	0.26%	Pasteurellales	0.17%	0.81%	Desulfarculales	0.00%	0.00%
Rhodocyclales	0.12%	0.15%	Synechococcales	0.06%	0.21%	Anaerolineales	0.16%	0.90%	Desulfobacterales	0.00%	0.00%
Selenomonadales	0.11%	0.12%	Planctomycetales	0.06%	0.18%	Geodermatophilales	0.15%	0.18%	Desulfovibrionales	0.00%	0.00%
Deinococcales	0.11%	0.36%	Bacteroidetes Order I	0.06%	0.25%	Methylococcales	0.14%	0.20%	Desulfuromonadales	0.00%	0.00%
Oceanospirillales	0.10%	0.21%	Jiangellales	0.05%	0.18%	Fusobacteriales	0.14%	0.27%	Myxococcales	0.00%	0.00%

The 50 most abundant bacterial genera detected from ancient and modern bison oral samples.

UV-EDTA						UV-Bleach					
Modern			Ancient			Modern			Ancient		
Bacterial genus	Frequency	Proportion	Bacterial genus	Frequency	Proportion	Bacterial genus	Frequency	Proportion	Bacterial genus	Frequency	Proportion
Anaplasma	2,022,814	14.32%	Actinomyces	899,332	21.28%	Streptococcus	575,798	25.28%	Streptomyces	3,204,175	18.18%
Actinomyces	1,895,025	13.41%	Flavobacterium	394,737	9.34%	Enterococcus	339,518	14.91%	Pseudomonas	1,856,422	10.53%
Streptococcus	1,002,547	7.10%	Actinobacillus	219,913	5.20%	Escherichia	269,385	11.83%	Curvibacter	1,132,950	6.43%
Arthrobacter	949,454	6.72%	Arthrobacter	187,647	4.44%	Micromonospora	155,824	6.84%	Rhodomicrobium	1,106,620	6.28%
Kocuria	935,959	6.62%	Pseudomonas	177,138	4.19%	Psychrobacter	126,704	5.56%	Bradyrhizobium	681,433	3.87%
Psychrobacter	888,556	6.29%	Curtobacterium	146,032	3.46%	Bacillus	102,692	4.51%	Acinetobacter	590,689	3.35%
Pseudomonas	685,628	4.85%	Sphingomonas	136,070	3.22%	Enterobacter	97,744	4.29%	Acidovorax	554,949	3.15%
Enterococcus	607,729	4.30%	Streptococcus	107,507	2.54%	Actinomyces	89,227	3.92%	Arthrobacter	530,453	3.01%
Escherichia	474,191	3.36%	Mannheimia	99,389	2.35%	Klebsiella	76,448	3.36%	Janthinobacterium	487,644	2.77%
Micromonospora	354,144	2.51%	Desulfosporosinus	95,341	2.26%	Acinetobacter	63,446	2.79%	Herminiimonas	472,176	2.68%
Acinetobacter	293,038	2.07%	Acinetobacter	83,165	1.97%	Streptobacillus	51,226	2.25%	Polaromonas	345,562	1.96%
Hymenobacter	233,748	1.65%	Hymenobacter	77,322	1.83%	Stenotrophomonas	49,179	2.16%	Nocardioides	303,379	1.72%
Microbacterium	232,828	1.65%	Clostridium	73,918	1.75%	Pseudomonas	36,541	1.60%	Flavobacterium	284,419	1.61%
Bradyrhizobium	219,066	1.55%	Streptomyces	70,997	1.68%	Citrobacter	32,699	1.44%	Rhodoferax	272,539	1.55%
Rhodococcus	218,746	1.55%	Frigoribacterium	67,376	1.59%	Brevundimonas	24,697	1.08%	Massilia	190,612	1.08%
Bacillus	203,274	1.44%	Brevundimonas	58,956	1.40%	Alcaligenes	20,755	0.91%	Hydrogenophaga	183,881	1.04%
Enterobacter	196,356	1.39%	Haemophilus	51,606	1.22%	Clostridium	20,259	0.89%	Anaeromyxobacter	175,456	1.00%
Actinobacillus	190,116	1.35%	Microbacterium	47,547	1.13%	Arthrobacter	13,672	0.60%	Streptococcus	171,852	0.97%
Brevundimonas	189,437	1.34%	Olsenella	45,873	1.09%	Saccharomonospora	13,436	0.59%	Rhizobium	156,044	0.89%
Streptomyces	182,619	1.29%	Pelagibacter	41,296	0.98%	Kocuria	12,220	0.54%	Variovorax	147,721	0.84%
Deinococcus	173,189	1.23%	Bradyrhizobium	41,240	0.98%	Rhodococcus	10,780	0.47%	Mycobacterium	137,684	0.78%
Stenotrophomonas	168,979	1.20%	Propionibacterium	37,380	0.88%	Fusobacterium	10,732	0.47%	Brevundimonas	133,504	0.76%
Klebsiella	153,251	1.08%	Corynebacterium	36,546	0.86%	Microbacterium	10,561	0.46%	Comamonas	120,298	0.68%
Corynebacterium	142,006	1.00%	Methanobrevibacter	33,494	0.79%	Actinobacillus	9,346	0.41%	Mesorhizobium	119,995	0.68%
Mannheimia	118,732	0.84%	Fusobacterium	32,157	0.76%	Campylobacter	9,061	0.40%	Hyphomicrobium	116,951	0.66%
Sphingomonas	115,633	0.82%	Chryseobacterium	31,651	0.75%	Aeromicrobium	6,172	0.27%	Actinoplanes	112,806	0.64%
Nocardioides	109,418	0.77%	Phycicoccus	30,108	0.71%	Massilia	5,881	0.26%	Enterococcus	108,800	0.62%
Lautropia	105,507	0.75%	Neisseria	29,215	0.69%	Hymenobacter	5,373	0.24%	Burkholderia	105,470	0.60%
Aeromicrobium	79,819	0.56%	Janibacter	27,268	0.65%	Mannheimia	5,317	0.23%	Bosea	103,647	0.59%
Streptobacillus	76,023	0.54%	Mycobacterium	26,392	0.62%	Streptomyces	4,478	0.20%	Geobacter	102,013	0.58%
Brevibacterium	75,013	0.53%	Enterococcus	25,875	0.61%	Nocardioides	4,315	0.19%	Noviherbaspirillum	84,181	0.48%
Curtobacterium	69,944	0.50%	Porphyromonas	24,061	0.57%	Mycolicibacterium	4,257	0.19%	Halomonas	80,946	0.46%
Neisseria	66,866	0.47%	Bibersteinia	23,929	0.57%	Chlamydia	3,605	0.16%	Olsenella	78,034	0.44%
Dietzia	62,643	0.44%	Nocardioides	23,346	0.55%	Brevibacterium	3,109	0.14%	Sphingomonas	76,991	0.44%
Desulfocarbo	60,614	0.43%	Kineococcus	22,294	0.53%	Haemophilus	2,841	0.12%	Methylobacterium	74,804	0.42%
Massilia	56,063	0.40%	Aggregatibacter	21,696	0.51%	Pontibacter	1,615	0.07%	Escherichia	73,420	0.42%
Haemophilus	53,521	0.38%	Bacillus	21,326	0.50%	Leucobacter	1,516	0.07%	Devosia	73,194	0.42%
Geodermatophilus	53,428	0.38%	Polaromonas	21,226	0.50%	Bibersteinia	1,422	0.06%	Micromonospora	72,491	0.41%
Methylobacterium	51,229	0.36%	Escherichia	20,670	0.49%	Aggregatibacter	1,392	0.06%	Microbacterium	68,788	0.39%
Paracoccus	50,829	0.36%	Ilumatobacter	18,048	0.43%	Chryseobacterium	1,255	0.06%	Dechloromonas	59,039	0.33%
Pontibacter	50,044	0.35%	Aeromicrobium	17,651	0.42%	Mycobacterium	1,156	0.05%	Collimonas	57,423	0.33%
Macrococcus	47,271	0.33%	Actinobaculum	16,157	0.38%	Deinococcus	661	0.03%	Bacillus	56,008	0.32%
Tessaracoccus	44,231	0.31%	Bacteroides	16,038	0.38%	Neisseria	488	0.02%	Rhizobacter	55,318	0.31%
Mycobacterium	43,867	0.31%	Cyanobium	15,683	0.37%	Polaromonas	475	0.02%	Amycolatopsis	53,119	0.30%
Brachybacterium	43,658	0.31%	Prevotella	14,723	0.35%	Brachybacterium	437	0.02%	Phycicoccus	52,233	0.30%
Blastococcus	42,975	0.30%	Micromonospora	13,593	0.32%				Herbaspirillum	51,919	0.29%
Fusobacterium	42,482	0.30%	Acidovorax	12,806	0.30%				Stenotrophomonas	51,399	0.29%
Citrobacter	41,849	0.30%	Pasteurella	12,633	0.30%				Sinorhizobium	51,066	0.29%
Variovorax	38,313	0.27%	Massilia	12,218	0.29%				Rhodopseudomonas	49,390	0.28%
Janibacter	38,013	0.27%	Methanosphaera	12,019	0.28%				Bordetella	47,096	0.27%

4.6.5 Table S4.4

Sequences assigned to eukaryotic phyla (%).

Eukaryotic phylum	UV-Bleach (modern)		UV-Bleach (ancient)		UV-EDTA (modern)		UV-EDTA (ancient)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Apicomplexa	1.80%	0.86%	0.02%	0.05%	0.59%	0.70%	0.00%	0.01%
Ascomycota	0.62%	1.14%	0.28%	0.78%	1.45%	2.04%	0.07%	0.08%
Basidiomycota	0.01%	0.03%	0.02%	0.04%	0.16%	0.26%	0.01%	0.03%
Mucoromycota	0.00%	0.00%	0.01%	0.02%	0.00%	0.01%	0.00%	0.00%
Nematoda	0.15%	0.12%	0.02%	0.12%	0.06%	0.09%	0.00%	0.02%
Arthropoda	0.42%	0.19%	0.05%	0.07%	0.16%	0.17%	0.02%	0.04%
Cnidaria	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.02%
Bacillariophyta	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.07%
Chlorophyta	0.00%	0.00%	0.01%	0.02%	0.01%	0.02%	0.01%	0.01%
Streptophyta	0.03%	0.06%	0.31%	1.14%	0.47%	1.26%	0.05%	0.10%

Statement of Authorship

Title of Paper	Reconstruction of the bison methylome history over the past 50,000 years
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	In preperation for submission

Principal Author

Name of Principal Author (Candidate)	Yichen Liu
Contribution to the Paper	Data analysis, data interpretation, writing the manuscript.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 05/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    i.   the candidate's stated contribution to the publication is accurate (as detailed above);

    ii.   permission is granted for the candidate in include the publication in the thesis; and

    iii.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Graham Gower
Contribution to the Paper	Processed methylation data up to methylation calling. Calculated ACF for DMR smoothing.
Signature	Date 22/01/2019

Name of Co-Author	Holly Heiniger
Contribution to the Paper	Performed experiments.
Signature	Date 1/3/19

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

    iv.   the candidate's stated contribution to the publication is accurate (as detailed above);

    v.   permission is granted for the candidate in include the publication in the thesis; and

    vi.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Bastien Llamas		
Contribution to the Paper			
Signature		Date	04/03/19

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Designed the project, data interpretation		
Signature		Date	01/03/2019

Chapter 5

Reconstruction of the bison methylome history over the past 50,000 years

Abstract

DNA methylation is a major regulator in many critical cellular processes, including cell differentiation and gene silencing. Emerging evidence suggests a role for DNA methylation in mammal evolution, making the generation of ancient DNA methylation data essential for a better investigation of this hypothesis. However, no existing method can recover ancient methylomes at a single-base resolution. In this study, we optimised a hairpin-based DNA library protocol and used the gold standard bisulfite treatment to retrieve high-quality methylome data from highly degraded DNA samples. We applied the method to a range of ancient and modern bison specimens and were able to reconstruct 11 ancient and 14 modern methylomes. We show that the hairpin method can increase the mappability of short bisulfite-treated DNA fragments and our results are consistent with a previously reported approach that uses ancient DNA damage as a proxy to recover the methylation pattern from ancient samples. Comparison of ancient and modern bison methylomes suggests tissue-specific methylation patterns can be preserved within ancient specimens. By constructing a 50,000-year methylome history, we were able to identify potential methylation hotspots responding to mammal-environment interactions. Our study demonstrates the great potential of the hairpin method in revealing novel dynamics in mammal microevolution.

Key words: Paleomethylome, ancient DNA, epigenetics, bison

5.1 Introduction

DNA methylation is an important mechanism of regulation of gene expression, and is involved in various essential cellular processes including cell differentiation, suppression of transposable elements, and genomic imprinting (Jones; 2012). In mammals, DNA methylation is the addition of a methyl group to the 5th carbon position of cytosines. DNA methylations affects cytosines mainly in a CpG context (cytosine followed by a guanine), which in humans represent 1% of all DNA bases and 70–80% of all CpG dinucleotides (Bird; 2002). DNA methylation can respond to environmental cues, leading to phenotypic variation (Jirtle and Skinner; 2007). These phenotypic outcomes can be maintained over multiple generations and are subject to selection in laboratory mice (Cropley et al.; 2012). Consequently, DNA methylation has been suggested to play a role in animal adaptation to the environment (Cropley et al.; 2012; Skinner; 2015). However, supporting empirical data remains scarce, as studies using modern animals can only infer what happened over microevolutionary timescales. The ability to recover DNA methylation information from ancient animals would provide an alternative and powerful approach to investigate the role of DNA methylation in animal adaptation.

Although DNA decays over time, methylation data have been successfully recovered from ancient remains using several experimental and computer-based approaches (Llamas et al.; 2012; Gokhman et al.; 2014; Pedersen et al.; 2014; Smith et al.; 2015; Seguin-Orlando et al.; 2015). The first approach uses damage patterns of ancient DNA to infer its methylation profile (Hanghøj et al.; 2016). In ancient DNA, deamination of cytosines (C) into uracils (U) (leading to C-to-T substitutions in the sequencing data) and methylated cytosines (5mC) into thymines (T) is a characteristic post-mortem damage (Briggs et al.; 2009). After enzymatically removing U during sequencing library preparation, the remaining C-to-T substitutions in the sequence data can thus be used as a proxy to infer methylation (Briggs et al.; 2009). Alternatively, it is possible to use a high-fidelity DNA polymerase that cannot amplify U. Again, C-to-T substitutions observed in the sequence data result from the deamination of C (Briggs et al.; 2009). Ancient human and archaic hominin methylomes have been reconstructed and lineage-specific methylation identified using these methods (Gokhman et al.; 2014; Pedersen et al.; 2014). A second approach uses bisulfite conversion, the gold standard for modern methylation studies (Frommer et al.; 1992). Sodium bisulfite converts C into U and leaves 5mC intact. After *in vitro* DNA amplification and sequencing, any original 5mC is detected as C in the data, while original C are detected as T. This method has been applied to ancient bison and humans, and single-base-resolution methylation patterns have been obtained from a limited number of candidate loci (Llamas et al.; 2012; Smith

et al.; 2015). A third approach, Methylated Binding Domains (MBD)-based enrichment, has also been tested on ancient DNA, but the performance is overall limited by DNA preservation and leads to strong biases toward the amplification of long, GC-rich, and damage-free DNA fragments (Seguin-Orlando et al.; 2015).

So far, obtaining methylome-wide data from ancient samples at single-base resolution has not been achieved. Because DNA damage is a random and low probability process, not all 5mC at a particular genomic position will be affected by deamination. Therefore, methylation levels can only be inferred with enough statistical power across a sliding genomic window of several tens of nucleotides (Gokhman et al.; 2014; Pedersen et al.; 2014; Hanghøj et al.; 2016). Therefore, the application of this method is largely limited to the identification of differentially methylated regions (DMR) (Hanghøj et al.; 2016). Only bisulfite conversion would allow to characterise differentially methylated loci (DML) at high resolution. However, the conversion of C into T during bisulfite treatment decreases the complexity of the sequences from four possible bases (ATGC) to mostly three possible bases (ATG). This issue is critical for the very short ancient DNA fragments since decreased complexity can lead to a major reduction in mappability, rendering the short converted fragments useless. One potential way to bypass this problem is hairpin-bisulfite PCR (Polymerase Chain Reaction), which was originally used to investigate the methylation symmetry between complementary DNA strands (Laird et al.; 2004). Specifically, a hairpin adapter is ligated to double-stranded DNA fragments, so the complementary strands are physically linked and can be sequenced together at once. After bisulfite conversion, amplification, and sequencing, the methylation state of both strands and the original four-base coded sequences can be obtained by "folding" the hairpin and the two complementary strands sequences bioinformatically. This method has been subsequently adapted for High Throughput Sequencing (Zhao et al.; 2014), but not yet explored in ancient specimens.

Steppe bison represent one of the best model species for the investigation of interaction between mammals and the environment (Shapiro et al.; 2004). This species experienced and survived drastic climate fluctuations previously to and during the last ice age (Shapiro et al.; 2004). In addition, a large number of bison sub-fossil samples are available, which makes them good candidates for the reconstruction of a methylome history. Here, we optimised the hairpin-bisulfite PCR for ancient DNA (referred as "hairpin method" in the following text) and applied it to a range of ancient and modern bison samples. The performance of the hairpin method was assessed and a single-base-resolution methylome history was reconstructed from 11 ancient bison samples.

5.2   Methods

5.2.1   Sample details

Ancient samples

In order to investigate the impact of climate changes on bison methylomes, we collected 11 ancient bison (*Bison priscus*) bone samples from North America, the age of which ranged from 50ky BP to 1ky BP (Figure 5.1, Table S5.1). These samples include three females and eight males. Sample A3020 is a radius bone, and the remaining samples are petrosals (petrous part of the temporal bone). Prior to PCR amplification, all the ancient samples were stored and processed in an ancient DNA laboratory at the Australian Centre for Ancient DNA (ACAD) at The University of Adelaide.

Modern samples

We also included 14 modern North American bison (*Bison bison*) samples from 3 females and 6 males (Table S5.1). The modern samples are from tissues derived from the paraxial mesoderm, including 6 radius samples, 6 petrosals, and 2 muscle samples. All the modern samples were stored and processed in a modern DNA laboratory at The University of Adelaide.

5.2.2   DNA extraction and fragmentation

Ancient DNA was extracted using a protocol described previously (Rohland and Hofreiter 2007; Brotherton, et al. 2013). Modern samples were extracted using a DNeasy Blood Tissue Kit (Qiagen) following the manufacturer's instructions for soft tissues, and the modified protocol for bones available from the Qiagen website ("Purification of total DNA from compact animal bone using the DNeasy Blood Tissue Kit"). Before constructing bisulfite sequencing libraries, 2 μg of DNA extracted from modern samples was fragmented down to 200 bp using an S220 Focused-ultrasonicator (Covaris). Ancient DNA was not fragmented since it is typically highly fragmented.

5.2.3   Illumina sequencing library construction

Fully repaired double-stranded sequencing library were prepared for 3 extinct steppe bison (see Table S5.2), as previously published (Meyer and Kircher; 2010). The final amplification of the sequencing library was performed using the Phusion polymerase (Thermo Fisher Scientific). Data from this repaired library were used for the characterisation of methylation based on ancient DNA patterns.
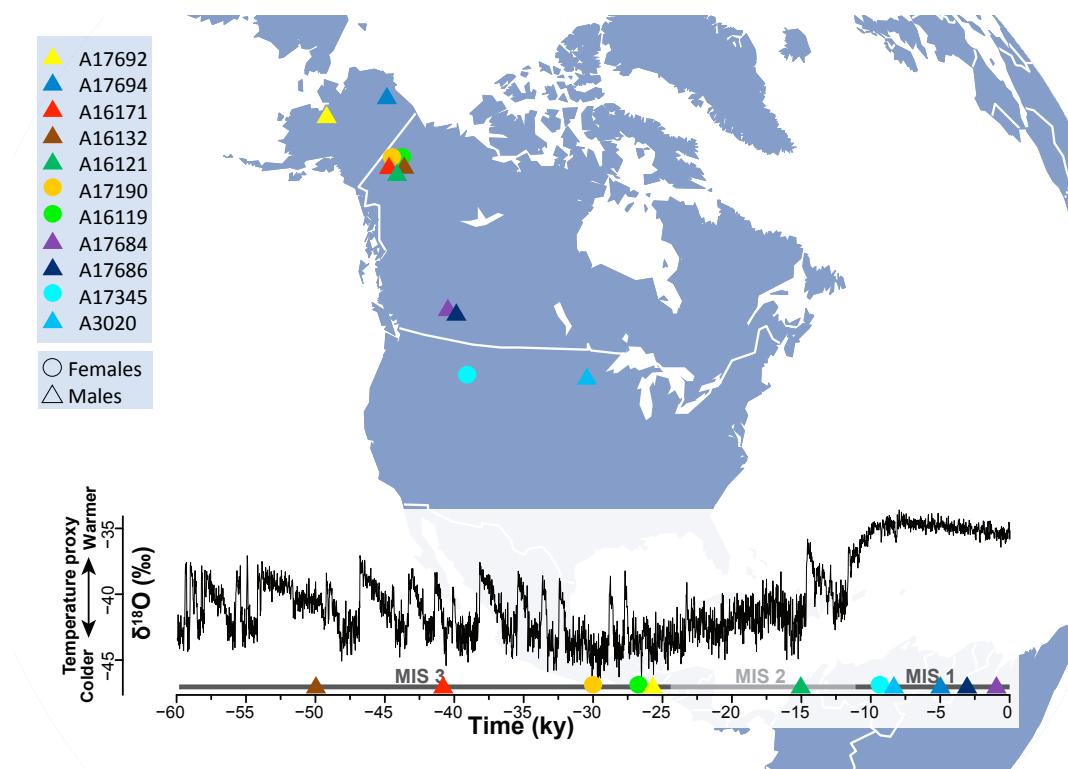
Figure 5.1: Distribution of ancient bison bone samples. Sampling locations were mapped to North America continent. The age of samples were displayed on a timeline, on top of which is the estimated temperature fluctuations. Each individual were labelled using different colour, and the circles represent females and the triangles represent males.

5.2.4 Bisulfite sequencing library construction

A schematic flowchart of the hairpin protocol is provided in Figure 5.2. The hairpin sequence and modifications (5′ phosphorylation, biotinylated T in the loop sequence) were same as in Zhao et al. (2014), except that all C were methylated. Table S5.2 lists samples for which hairpin bisulfite libraries were generated and regular whole-genome bisulfite sequencing (WGBS) was performed.

Size selection, polishing, and A-tailing of modern DNA

Following sonication, large DNA fragments were removed using $0.6\times$ of Ampure XP beads (Beckman Coulter) following the manufacturer's instructions. Then short DNA fragments were removed using $0.9\times$ of Ampure XP beads. We spiked in λ-DNA—previously fragmented and purified similarly to the modern bison DNA—at 0.5% (wt/wt). The λ-DNA serves as unmethylated control sequences from which to determine the efficiency of the sodium bisulfite treatment. The modern hairpin libraries were prepared using the Truseq Nano kit (Illumina), following the manufacturer's Low Sample protocol.

Damage repair, polishing, and A-tailing of ancient DNA

The ancient hairpin libraries were prepared using the NEBNext library preparation kit (New England Biolabs, or NEB) with several modifications. A total of 25 μL of DNA extract was polished and repaired with 6 μL 10x Buffer Tango (NEB), 0.2 μL of dNTP (25 mM each), 6 μL of 10 mM ATP, 5.2 μL of ddH2O, 3ul of 10 U/μL polynucleotide 5'-hydroxyl-kinase (PNK; NEB), and 3.6 μL of USER enzyme mix (NEB) for 60 min at 37°C. Then 3.6 μL of 2 U/μL Uracil Glycosylase Inhibitor (UGI; NEB) was added to the reaction and incubated at 37°C for 30 min and then 12°C for 1 min. Subsequently, 3 μL of NEBNext Ultra II End Prep Enzyme mix and 7 μL of NEBNext Ultra II End prep reaction Buffer was added to the reactions and incubated under following conditions: 20°C for 30 min and then 65°C for 30 min.

Ligation of modern DNA with adapters

The ligation procedure was slightly modified from the Illumina Truseq Nano Low Sample protocol. We used 1.25 μL of 15 μM DNA Adapter Indices and 1.25 μL of 15 μM phosphorylated hairpin adapter in the ligation reaction. At the end of the incubation, reactions were cleaned up using a MinElute Reaction Cleanup kit (Qiagen) and eluted in 20 μl of elution buffer (10 mM Tris-Cl, pH 8.5 containing 0.05% Tween).

Ligation of ancient DNA with adapters

A total of 30 μL of polished DNA was mixed with 1 μL of NEBNext Ligation Enhancer, 1.25 μL of 15 μM NEBNext methylated adaptor and 1.25 μL of 15 μM phosphorylated hairpin adapter, and incubated at 20°C for 15 min. Then 3 μL of USER enzyme was added to the reaction mix and incubated at 37°C for 15 min. Reactions were cleaned up using a MinElute Reaction Cleanup kit (Qiagen) and eluted in 20 μl of elution buffer (10 mM Tris-Cl, pH 8.5 containing 0.05% Tween).

Streptavidin beads preparation and isolation of DNA ligated with the hairpin adapter

A total of 20 μL of MyOne Streptavidin C1 magnetic beads were washed twice with 400 μL of binding buffer (1 M NaCl; 10 mM Tris-HCl pH 7.5; 1 mM EDTA) at room temperature. Then the supernatant was discarded and the beads were incubated in 400 μl of binding buffer and 40 μg of yeast tRNA for 30 min on a rotor at room temperature, in order to saturate all non-specific DNA binding sites on the beads. The residual yeast tRNA was removed through washing the beads twice with 400 μL of binding buffer. Then 20 μL of the resulting DNA from previous step and 200 μL of binding buffer were added to the beads and incubated at room temperature for 30 min, following by two washes with 400 μL of binding buffer. The beads were then resuspended in 20 μL of binding buffer, incubated at 95°C for 5 min in a thermocycler, placed on an ice block for 10 min, and brought back up to room temperature for the next step.

Bisulfite conversion and PCR amplification

A total of 130 μL CT Conversion Reagent (EZ DNA Methylation-Gold kit, ZYMO Research) was added to 20 μl of DNA from previous step and then bisulfite converted under the following conditions: 98°C for 10 min; 64°C for 2.5 hours. Resulting reactions were purified following the manufacturer's instructions. Then 2 μL of bisulfite converted DNA was mixed with 12.5 μL of 2× KAPA HiFi uracil+ Readymix (Kapa Biosystems), 8.5 μL of ddH$_2$O, 2 μl of Illumina PCR Primers Cocktail from the NEBNext kit, and amplified under the following conditions: 95°C for 2 min; 98°C for 30 sec; 5–19 cycles of 98°C for 15 sec, 60°C for 30 sec, 72°C for 1 min; 72°C for 10 min. The resulting reactions were purified using Ampure XP beads and DNA was eluted in elution buffer (10 mM Tris-Cl, pH 8.5 containing 0.05% Tween).

Whole genome bisulfite sequencing (WGBS) library construction

The WGBS libraries were prepared as the hairpin libraries with two modifications. First, in the ligation step (4.2), we used 2.5 μL of 15 μM NEBNext methylated

adapter instead of a mixture of NEBNext adapter and hairpin adapter. Second, step "Streptavidin beads preparation and isolation of DNA ligated with the hairpin adapter" was skipped, so the resulting ligation product were directly put through to the bisulfite conversion (4.4).

5.2.5 Sequencing

The constructed libraries, each with unique indexes, were pooled at equimolar concentrations prior to 300-cycle paired-end sequencing on an Illumina HiSeq XTen platform.

5.2.6 Methylome data analyses

Sequencing data processing

Fastq files were demultiplexed by the sequencing provider using CASAVA v1.8. Four-base DNA sequences were reconstructed from the demultiplexed fastq files, mapped to the *Bos taurus* reference (UMD 3.1.1), and methylation status called at each C in the reference, using a newly developed and validated bioinformatic pipeline (https://github.com/grahamgower/PP5mC). Upon examination of M-bias plots, which inform about the empirical frequency of C and 5mC along the reads (Hansen et al.; 2012), we trimmed methylation data from the first 10 nucleotides of each folded sequence.

Reconstruction of ancient methylomes using epiPALEOMIX

The raw sequencing data from the three fully repaired genomic libraries were processed using the pipeline Paleomix v1.0.1 (Schubert et al.; 2014), which 1) trims residual adapter sequences and low quality sequences, collapses overlapping paired reads, and filters reads shorter than 25 bp after trimming using AdapterRemoval v2.2.0 (Schubert et al.; 2016); and 2) maps collapsed and paired reads against UMD3.1.1 using the MEM algorithm with default parameters in BWA v.0.7.13 (Li and Durbin; 2009). Methylomes were then inferred from ancient DNA damage patterns with the pipeline epiPALEOMIX (Hanghøj et al.; 2016), following the developer's instructions available at https://bitbucket.org/khanghoj/epipaleomix/wiki/Home and using default parameters.

Simulations

To examine the mappability of 4-base coded short reads, 20 subsets of reads were extracted from the genome sequencing data of a wood bison mapped to the *Bos taurus* reference genome UMD 3.1.1. Each subset contains n = 100,000 uniquely
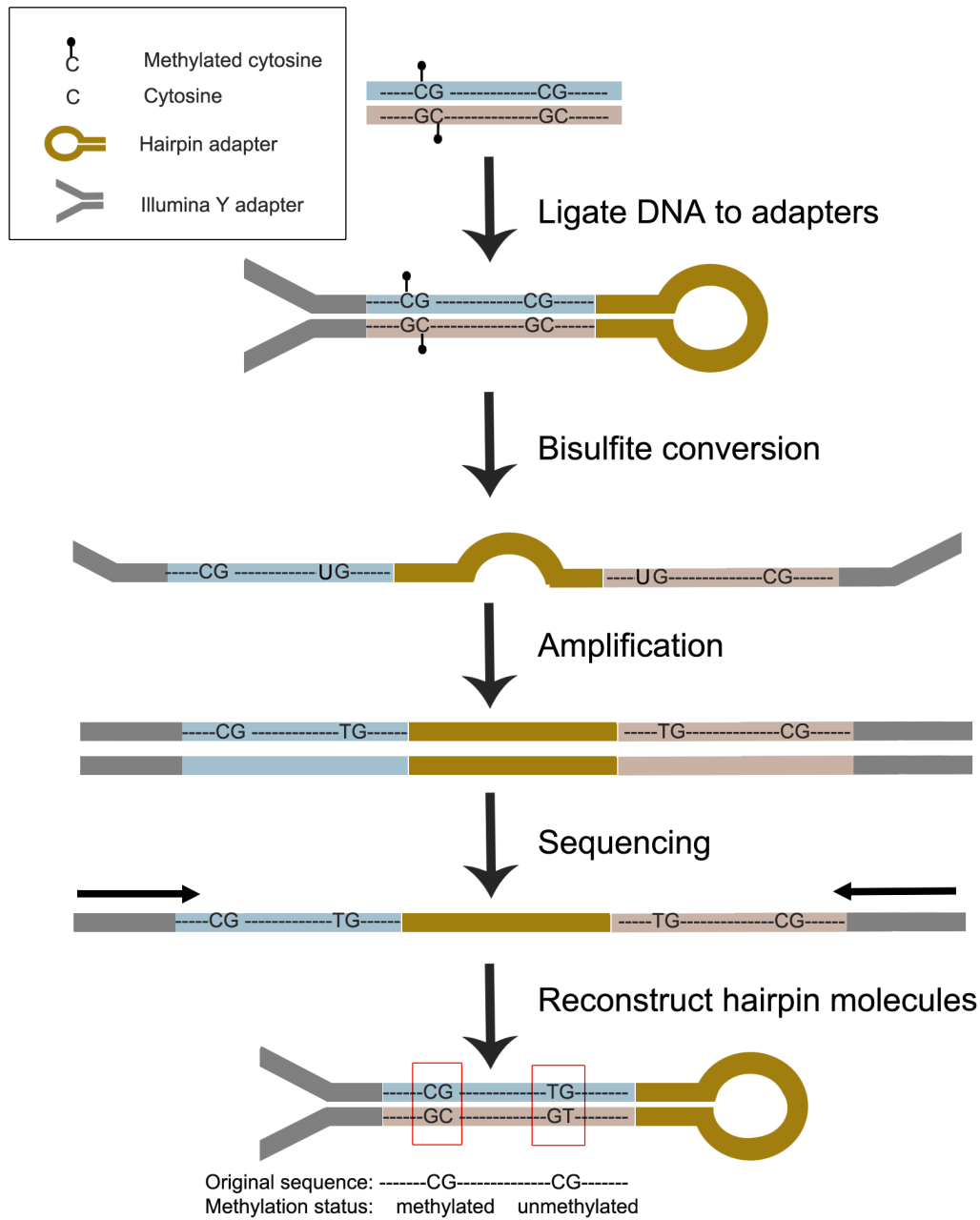
Figure 5.2: Schematic flowchart of the hairpin protocol. Briefly, DNA molecules were ligated to a hairpin adapter and an Illumina Y-adapter, so that the two complementary strands were physically linked. Then the ligation products were bisulfite converted, which converted unmethylated cytosines to uracils and denatured DNA. The bisulfite converted DNA was then amplified and sequenced. The sequencing data was used for the reconstruction of original sequences, and a C-G pair indicate a methylated cytosine, while a T-G pair indicate a unmethylated cytosine.

mapped reads that were 150 bp in length. These reads were randomly fragmented *in silico* into varied sizes ranging from 15bp to 100 bp (n = 100,000 for each size) and then mapped to UMD 3.1.1. The proportion of uniquely mapped reads was calculated. The mappability of 3-base coded short reads were evaluated the same way except all the C in the reads and UMD3.1.1 were converted to T *in silico* prior to mapping.

To explore the distribution of DNA fragment lengths within an ancient DNA library, 13 subsets of reads were simulated using SimWreck (https://github.com/mtrw/simwreck). Each subset contained 100,000 reads with a length ranging from 25 bp to 155 bp. Different shape parameters (α and β) were used to simulate fragment distributions with varied mean fragment length (Table S5.3).

Principle components analysis (PCA)

PCA was performed on the beta values using prcomp function in R (Team et al.; 2013). The beta value of a given CpG locus is equal to the number of C divided by the sum of C and 5mC. Proportion of variance explained by PC1 and PC2 was visualised using ggplot2 in R (Wickham; 2016).

Adjustment for deamination-induced decrease in global methylation level

As DNA damage can decrease the apparent methylation levels of ancient samples due to the deamination of C into T, the methylation levels were adjusted using the ComBat function from the package sva (Leek et al.; 2018). ComBat can be used to correct global differences between data generated from different batches (Johnson et al.; 2007). As the deamination rate is homogeneous across the genome in ancient DNA (Gokhman, et al. 2014), we reasoned that this package could be used for the correction of damage in ancient methylomes. No significant difference (F-test, p = 0.59) was observed between the variances of ancient and modern samples. Therefore, we grouped ancient methylomes as one batch and modern samples as the second batch, and adjusted data using parametric adjustments (par.prior = TRUE); the mean of the batch effect was corrected and the scales were adjusted (mean.only = FALSE).

DMR (differentially methylated region) calling

As BSmooth performs well on low-coverage data (Hansen et al.; 2012; Rackham et al.; 2017), it was employed to detect DMR in the methylome data. To estimate the proper window size for smoothing the data, an in-house python script was used to calculate the autocorrelation function (ACF) taking into account the distance between CpG sites. It calculates the correlation for 10-bp bins, up to a 10-kb

distance between CpG sites. ACF was calculated over a 100k-CpG region of the methylome data obtained from a modern and an ancient petrosal sample (A17199 and A17345). The results showed that the correlation drops to a base level at around 2 kb in both cases (Figure S5.1). Therefore, we used a 2-kb region (parameters used in BSmooth: ns = 50, h = 2000, and maxGap = 100,000,000) to smooth the data. After smoothing, CpGs with little or no coverage were removed. The remaining CpGs were at least covered by two samples with a depth greater than $2\times$ within each of the group. Then t-statistics were calculated using the following parameters: estimate.var = c("same"), local.correct = TRUE, qSd = 0.75, and k = 101. DMRs were identified by using an FDR corrected P value of 0.05(q cutoff: 0.025 and 0.975). DMRs that had less than 5 CpGs or a difference in methylation between the two groups less than 0.1 were removed. The genomic features of the DMRs were identified by comparing their genomic positions with UCSC Genome Browser annotation tracks for the reference UMD3.1.1 using BEDTools (Quinlan and Hall; 2010).

5.3 Results

5.3.1 Sequencing data and mapping results

For the bisulfite sequencing libraries prepared using the hairpin method, 85 to 404 million reads were generated for each sample (Table S5.2). On average, the sequences of complementary strands could be recovered from 57.8% of the reads by 'folding' *in silico* the hairpin molecules and collapse the plus and minus strand sequences (as showed in the last step of Figure 5.2: reconstruction of the hairpin molecule). The reads that could not be folded are likely from the molecules that failed to ligate to the hairpin adapter. The average length of sequence fragments after folding the complementary strands and trimming the adapter sequences was 54 bp for ancient samples and 151 bp (*i.e.* the maximum read length) for modern samples. The small fragment sizes of the ancient DNA libraries are characteristic of DNA degradation. Using the *Bos taurus* genome (UMD3.1.1) as a reference, 23 to 137 million folded reads were uniquely mapped. The coverage on autosomes was 13–78% with a depth of 0.17–1.99$\times$. To compare the mappability of the data generated using the hairpin method and WGBS, a well-preserved sample, A16121, was also prepared using the WGBS protocol. A total of 34 million collapsed reads were mapped to the *Bos taurus* genome (UMD3.1.1), and 32.97% of the genome was covered with a mean depth of 2.54X. For the comparison of the methylome profiles generated using the hairpin method and epiPALOEMIX, we also shotgun sequenced three ancient libraries (A16121, A16171, A3020) and obtained 302–917 million raw

reads (table S4); 112–515 million reads were uniquely mapped to the *Bos taurus* genome (UMD3.1.1) and the coverage was 78–89% with a depth of 2.73–11.99×.

5.3.2 The hairpin method improves the mappability of small fragments

We first examined the mappability of data generated using the hairpin method and WGBS (Figure 5.3). The main difference between the hairpin and WGBS data is the former is ATCG-coded, while the latter is largely ATG-coded. To explore the differences in mappability between the data generated using these two methods, we used simulated data to compare the proportion of ATCG-coded and ATG-coded reads that could be uniquely mapped to the *Bos taurus* genome (UMD 3.1.1) with varied read lengths (Figure 5.3A). The mappability of ATCG-coded reads greatly decreases when the read length is smaller than 25 bp, while a similar reduction in mappability occurs when ATG-coded reads are smaller than 35 bp. This means that theoretically the large majority of reads with a length between 25 and 35 bp is lost during mapping of bisulfite-converted data.

We then examined the mappability of empirical data generated from an ancient bison petrosal sample (A16121) using the hairpin method and WGBS (Figure 5.3B). The method-dependent difference in mappability is most prominent in fragments with a size between 30 and 60 bp, where the hairpin method provided a higher ratio of mappable data. In this case, the proportion of uniquely mapped reads with a length between 30 and 60 bp increased by 21.0% in the hairpin data. Compared to simulated data, the size of fragments exhibiting method-dependent differential mappability shifted to a larger and wider range in the empirical data. One possible explanation is that simulated reads were generated from a subsample of uniquely mapped reads, while the empirical data contains also DNA fragments with low complexity (*e.g.*, those from repetitive regions). As the fraction of short fragments tends to increase with decreased mean fragment size, the power of the hairpin method is likely to increase with highly degraded samples (Table S5.3, Figure S5.2).

5.3.3 Limited effects of sample age (radiocarbon date) on the methylation levels of ancient samples

To further explore the effect of DNA damage on the methylation levels of ancient samples, we analysed genome-wide methylation levels of all the petrosal samples (except an outlier, A16132, see below and figure S5.2). We also excluded the radius sample from analyses because 1) the mean methylation level is different in different tissues, meaning including samples from different tissues would have inflated the
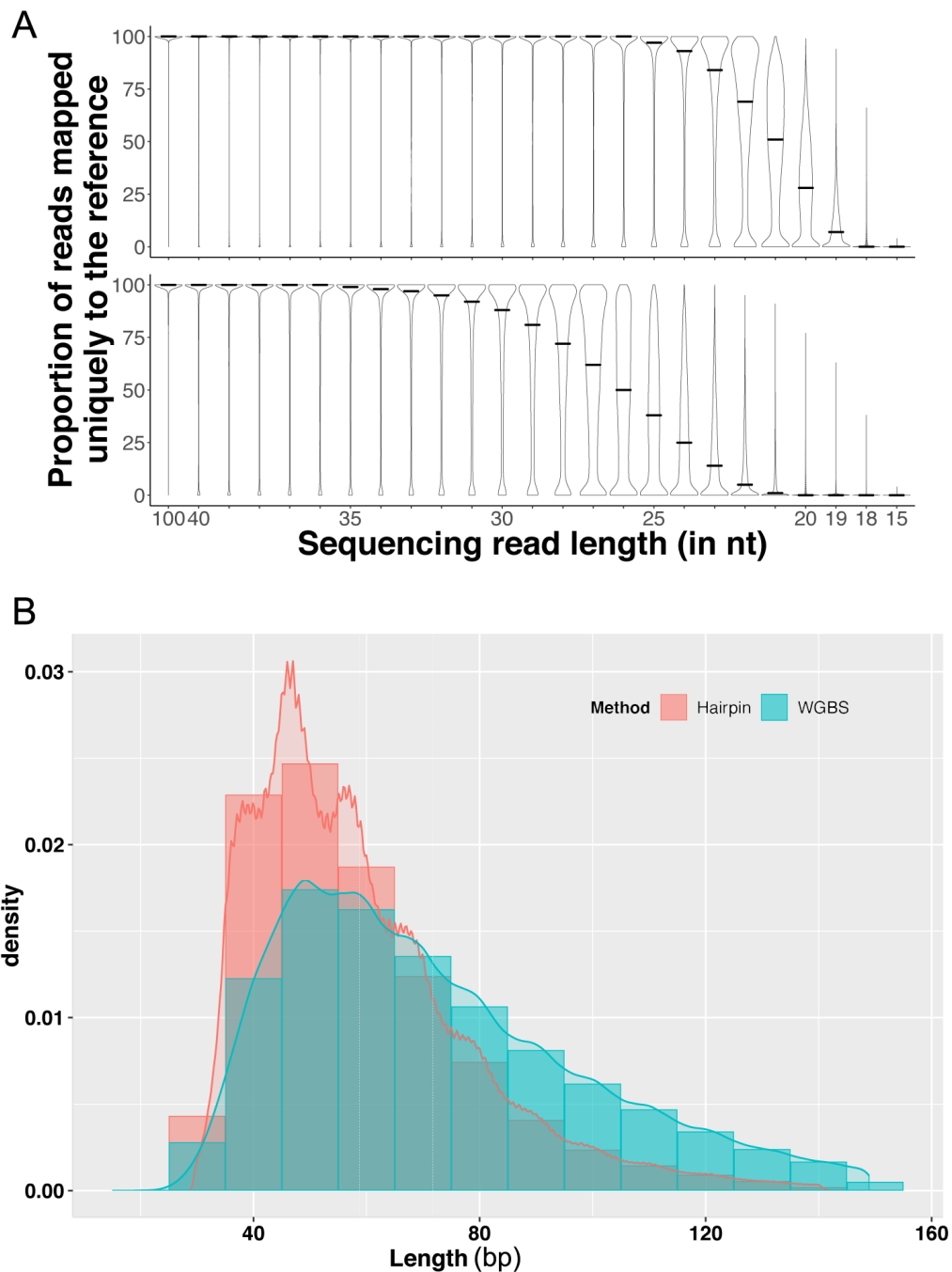
Figure 5.3: Comparison of mappability of data generated using hairpin method and WGBS. The proportion of uniquely mapped reads was used as a proxy to evaluate the mappability. A. The proportion of 4-base coded (upper) and 3-base coded (lower) reads mapped uniquely to cow genome (UMD 3.1.1) with varied read length. The x-axis represented the sequencing read length, and the y-axis reprinted the proportion of uniquely mapped reads. B. The proportion of uniquely mapped reads of data generated using hairpin method (pink) and WGBS (cyan). The curves showed the proportion of uniquely mapped reads from 30 bp to 150 bp in every base pair. The boxes showed the average proportion of uniquely mapped reads in every 10 bp.

methylation level variation observed in ancient samples, and 2) we only had one ancient radius sample, which could not provide enough power to perform statistical tests for comparison with modern radius samples. The ancient petrosal samples showed a global hypomethylation compared to modern samples (t-test, p=3.37e-07). This is likely because 5mC can be deaminated and ultimately read as unmethylated in sequencing data, thus lowering methylation levels in ancient samples. We also observed a decrease in the fraction of fully methylated loci (beta value = 1) and an increase of partially methylated loci (beta value <1) in ancient samples (Figure 5.4), supporting this hypothesis.

To determine if the ancient samples were affected by taphonomy in an age-dependent manner, we tested the correlation between the age and methylation levels of ancient petrosal samples. Only a negligible correlation was detected (Pearson correlation coefficient=-0.11, p=0.77) (Figure 5.5). In addition, the variation in methylation levels of ancient petrosals did not significantly differ from that of modern petrosals (F-test, p=0.59), suggesting minimal variance caused by the age of samples.

5.3.4 Data generated using the hairpin method is consistent with that generated using the damage-based method

We compared the methylation levels calculated using hairpin data and the damage pattern with the same samples (Figure 5.3). The methylation levels were calculated across nine different genomic features: CpG islands and flanking 3′-shores, 3′-shelves, 5′-shores, and 5′-shelves, promoters with high, intermediate, and low CpG density, and mitochondrial genomes. Since cytosine methylation is likely to be absent from mitochondrial genomes in mammals (Mechta, et al. 2017), the positive methylation scores for mitochondrial genomes from data generated using the damage-based method likely resulted from leftover uracils that UDG treatment failed to remove. Likewise, methylation scores for the mitochondrial genome from data generated using the hairpin method likely stem from C that were not converted by bisulfite. Apart from the mitochondrial genome, CpG islands were poorly methylated compared to flanking shores and shelves, as expected (Weber et al.; 2007).

CpG island shelves exhibited the highest methylation score, followed by the CpG island shores. Little difference in the methylation scores of CpG island shores and shelves with different directionality (3′ and 5′) was observed. For promoters, low-CpG density ones appeared to be the most methylated, while the high-CpG density ones were the least methylated, which is in line with the previously observed inverse correlation between methylation levels and CpG density and GC content (Ball et al.; 2009). The methylation scores generated using the hairpin method and the damage-
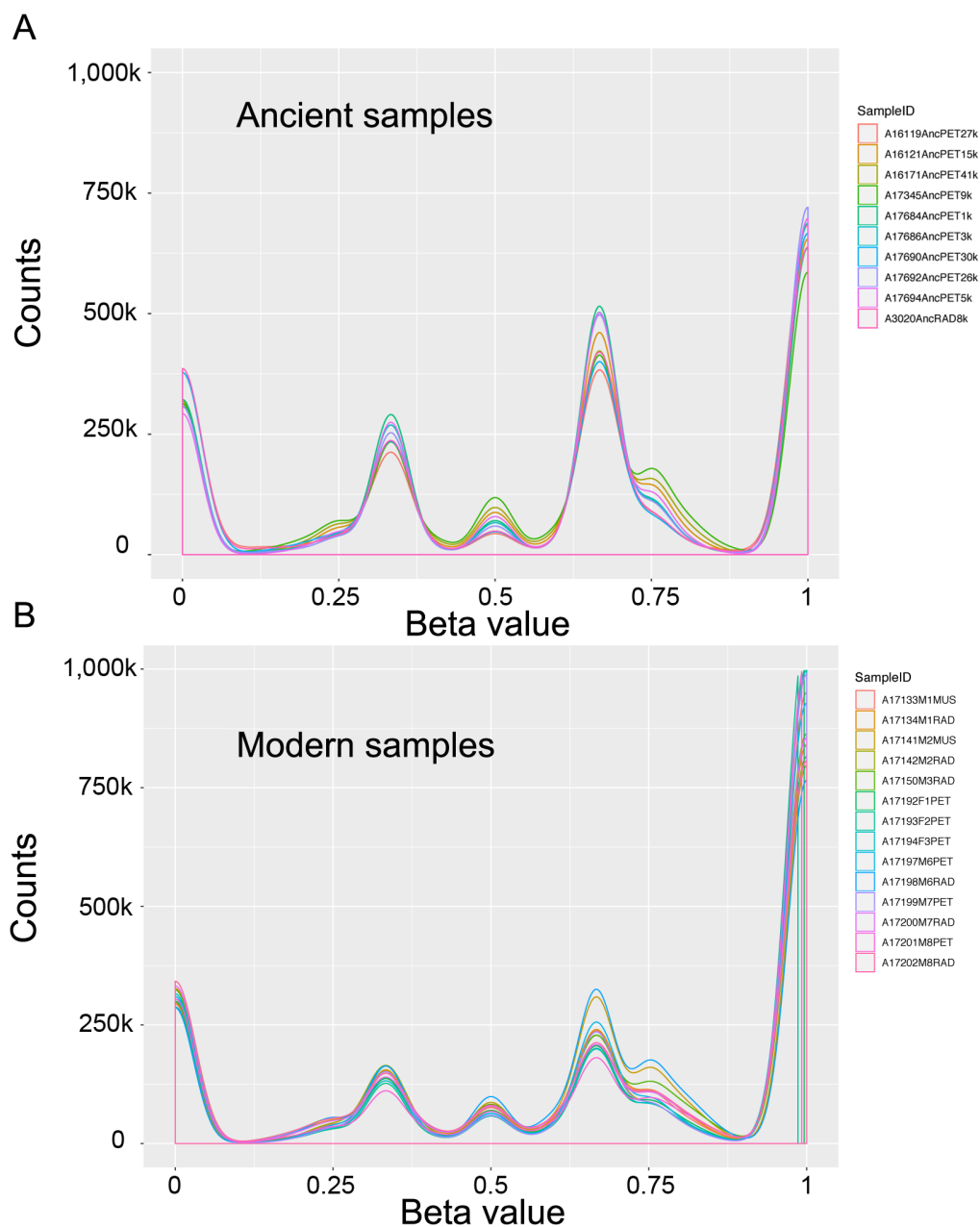
Figure 5.4: The beta-value distribution of ancient (A) and modern (B) methylomes. Y-axis represented beta value and X-asis represented the frequency of corresponding beta-value observed in 150k CpG sites. A beta value equals to 1 means the loci is full methylated and 0 means the site is unmethylated. Note that figure A and B used the same scale on X and Y axis.
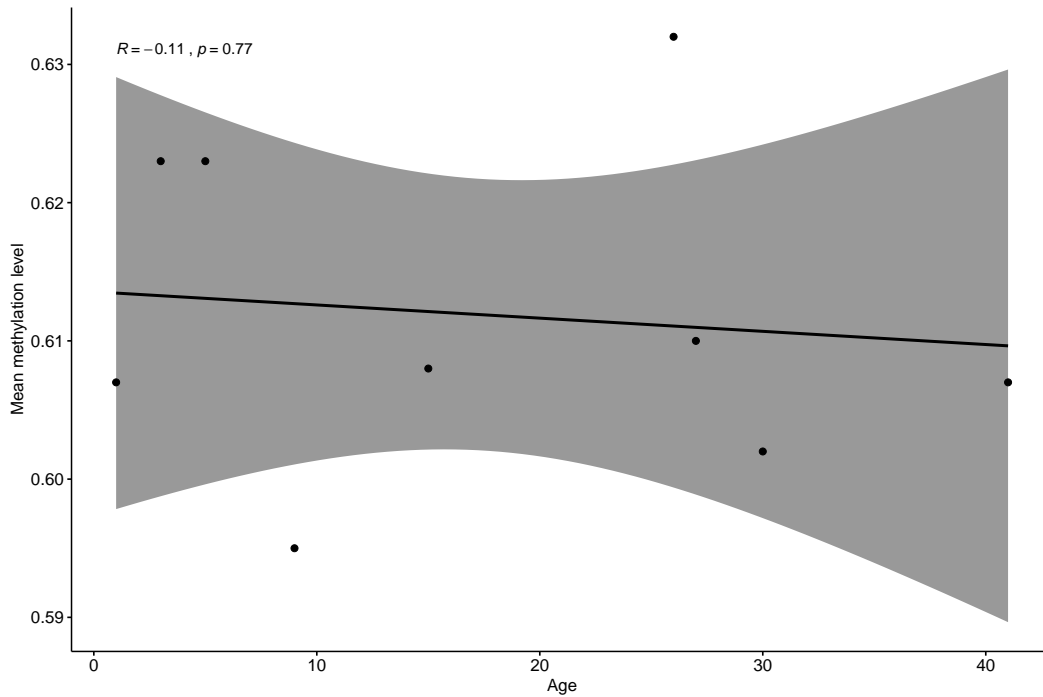
Figure 5.5: The correlation between the global methylation levels and the sample age. X-axis represent sample age (ky) and Y–axis represent the correlation. A Pearson correlation test detect a very weak correlation with an R equals to -0.11.

based method were highly comparable, with Pearson correlation coefficients of 0.999 (A16121), 0.994 (A16171), and 0.981 (A3020), and the coefficients are even higher (0.999, 0.997, and 0.994 respectively) after removal of the mitochondrial genome data. Additionally, the methylation distributions are very similar across the three ancient samples (pairwise Pearson correlation coefficients >0.99), despite the great difference in the sample age (8, 15, and 41ky BP), suggesting the age of the sample minimally affects the methylation patterns.

5.3.5 Base-resolution ancient methylomes preserve tissue-specific signal comparable to modern data

We compared the methylation levels of CpG sites covered by all samples. Sample A16132 appeared to be an outlier: even after filtering sites with low coverage, this sample still clustered away from the rest of the samples (Figure S5.2). Additionally, the coverage of sample A16132 was the lowest, and the coverage of the second lowest covered sample was 1.5 times higher than A16132 (Table S5.2). To avoid biases caused by missing data and abnormal methylation level, this sample was excluded from downstream analyses. PCA showed a separation between ancient and modern samples on PC1 (Figure 5.7), which is likely caused by DNA damage in the ancient samples. Separation on PC2 seemed to be driven by tissue specificity, as petrosal
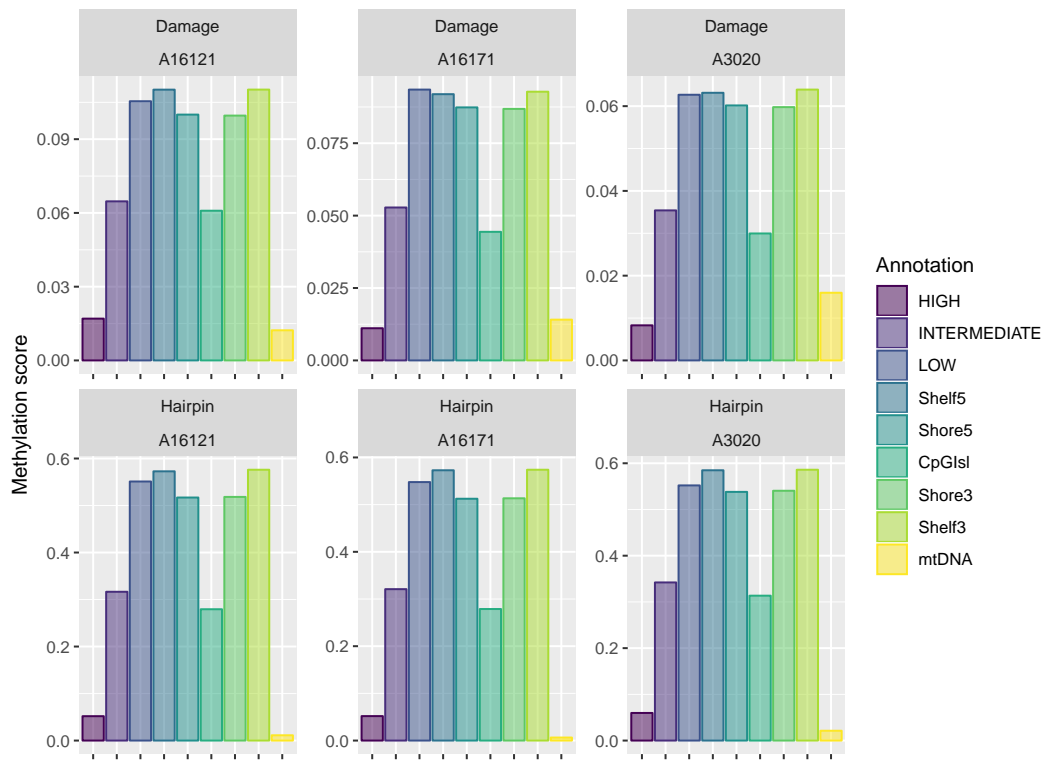
Figure 5.6: Mean DNA methylation score of different genomic features. Top, data generated using damage-based method (ratio of C-to-T substitution); bottom, data generated using hairpin method. Abbreviations: mtDNA: mitochondrial DNA, High: regions with a CG content higher 0.55, intermediate: regions with a CG content between 0.48 and 0.55; low: everything else. CpGIsl: CpG islands. Shore3 and Shore5 are the 2kb flanking the 3′- and 5′- CpG islands, and shelves are the next 2kb.

methylomes were clustered away from radius methylomes, although the tissue speci-
ficity seems less clear within muscle samples. The influence of DNA damage on the
methylation level was likely due to the deamination of 5mC in the ancient samples,
which can lead to a decrease in the overall methylation level (similar pattern was
observed in Figure 5.4). Therefore, we adjusted the global methylation levels of the
ancient samples so that they were comparable to modern samples. After the ad-
justment, no separation between ancient samples and modern samples was observed
on PC1. Instead, ancient petrosal samples were clustered with modern petrosal
samples, and the ancient radius sample was clustered closely with modern radius
samples, suggesting the ancient samples preserve tissue-specific signals comparable
to those found in modern samples. We did not observe sex-specificity on PC1 or
PC2, likely due to small effects compare to tissue specificity or lack of coverage on
sex-specific methylated loci.

5.3.6 Genes involves in development were actively methy-lated in ancient bison from different time periods

We compared the methylomes of modern samples and adjusted ancient samples. No
DMR was detected from this comparison, likely because these datasets had a lim-
ited number of overlapping CpG sites. We then compared the ancient samples from
different geological epochs. The late Pleistocene period is characterized by dramatic
climate events, while the Holocene climate is relatively stable (Figure 5.1) (Cooper
et al.; 2015). We hypothesized that the difference between the late Pleistocene and
Holocene bison methylomes would reflect an epigenetic response to climate changes.
The ancient samples were split into two groups: a Late Pleistocene group (11 ky BP)
and a Holocene group (11-0 ky BP). To avoid potential bias caused by tissue-specific
methylation, the only radius sample (A3020) was excluded and all the remaining
samples were petrosals. Both groups include males and females; thus, the influence of
sex-specific methylation should be minimal. Six DMRs of length between 59 and 778
bp were identified (Table S4). Three DMRs overlapped with the shores and shelves
of CpG islands; one DMR did not overlap with any know genomic feature, and two
DMRs overlapped with gene bodies [glutathione S-transferase, theta 4 (GSTT4)
and tumor necrosis factor receptor superfamily, member 10d (TNFRSF10D)]. Glu-
tathione S-transferase protects cellular components from damage by catalyzing the
conjugation of glutathione with electrophilic substrates, thus playing an important
role in stress responses (Hayes and Pulford; 1995; Eaton and Bammler; 1999). TN-
FRSF10D has been found to be dynamically regulated in pre- and postnatal skeletal
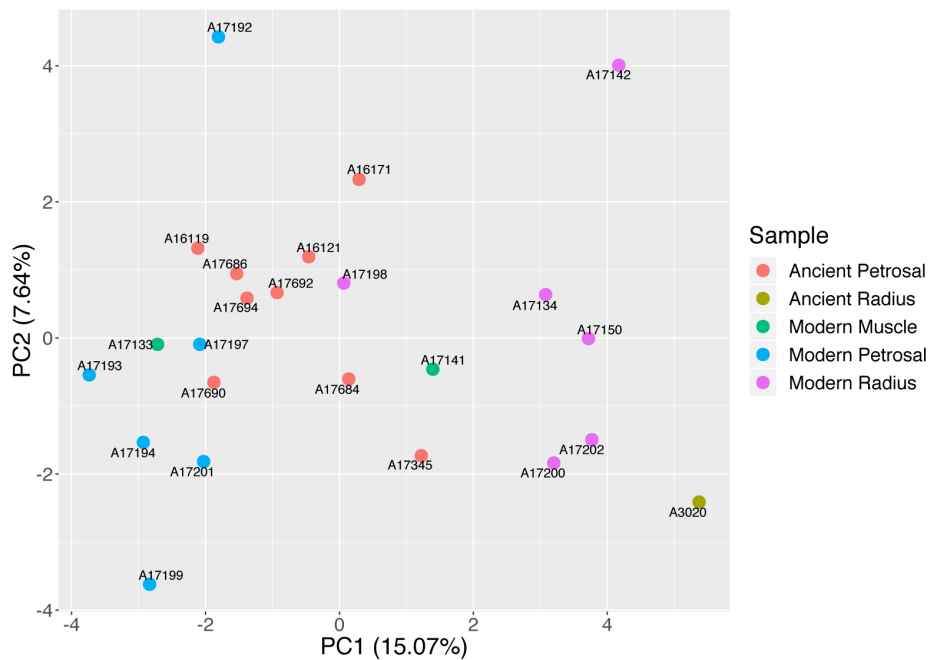muscle development in pigs (YANG et al.; 2014).
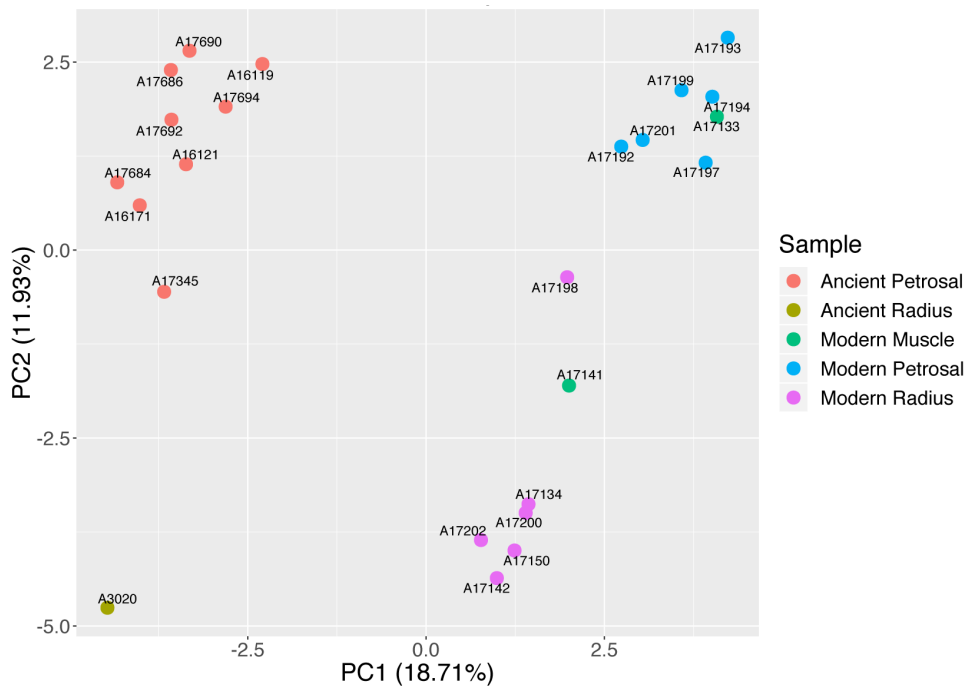
Figure 5.7: PCA on the methylation level of 3352 CpG sites. A: Before the adjustment for DNA damage, ancient samples were clustered away from modern samples. B. After the adjustment, ancient samples clustered closely with modern samples according to tissue type.

5.4 Discussion

5.4.1 Robustness of the hairpin method in recovering methylome data from highly degraded samples

DNA recovered from ancient samples is typically highly fragmented, which has limited the application of WGBS to ancient samples to retrieve single-base-resolution methylomes (Briggs, et al. 2007; Llamas, et al. 2012) . To address this limitation, we developed a hairpin method, which was optimized for ancient samples so that the complexity of DNA sequences can be maintained after bisulfite conversion. This method was applied to a range of samples and base-resolution methylomes were reconstructed from 11 ancient bison bones samples. To our knowledge, these data represent the first single-base-resolution methylomes recovered from ancient mammals. The hairpin method allows the original sequences to be reconstructed after bisulfite conversion and sequencing (Figure 5.2), which offers advantages in two aspects: 1) the reconstructed reads are coded by four bases, which greatly increases the mappability of short fragments, especially in hypomethylated regions; 2) both genomic and methylome information can be recovered using one hairpin library. Increased mappability of short fragments was observed from empirical data and simulated data, verifying the robustness of the hairpin method. As this method is powerful in increasing the mappability, it could potentially be applied to a range of difficult cases, such as degraded clinical samples and DNA from regions with low complexity (*e.g.*, repetitive regions). Our results demonstrate the power of the hairpin method in increasing the quality and quantity of methylome data from highly degraded samples.

5.4.2 Marginal correlation between deamination and sample ages

Deamination of 5mC in ancient samples can lead to a decrease in overall methylation level, as this process transforms 5mC into T, which are read as an unmethylated locus (Briggs et al.; 2009). It has been suggested that cytosine methylation decays over time, with a half-life of 700 years (Smith et al.; 2014). However, although we observed a genome-wide hypomethylation in the ancient methylomes, only marginal correlations (R = -0.11) were detected between the methylation levels and sample ages. Furthermore, the methylation levels and the distribution of methylation on varied genomic features are highly consistent among three samples with ages ranging from 8ky BP to 41ky BP. However, we trimmed methylation data in the first 10 nucleotides of all reads because of known biased methylation levels at the start of sequences (M-bias) (Bock; 2012). These are also the regions that are most affected

by deamination, with deamination rates decreasing exponentially a few bases from the termini of the reads (Briggs et al.; 2007). It is possible that the deamination rates within regions that are distant from molecule ends were accumulated in a less time-dependent manner. It can also relate to preservation conditions, as our samples were all collected from permafrost, which can alleviate the degradation of DNA. However, it is difficult to evaluate the preservation conditions, as 1) our samples are much older than the samples described in (Smith et al.; 2014), and 2) we used ancient mammal bone samples and they used archaeological plant seeds. While the deamination of 5mC in ancient samples needs to be further investigated, our data suggest removal of the end of reads can be an efficient way to minimise the biases caused by deamination.

5.4.3 Ancient methylomes maintain high fidelity and resolution

The fidelity of the ancient methylomes reconstructed in this study was evaluated in two ways. First, we compared the data generated using the hairpin method and an ancient DNA damage-based method—*i.e.* epiPALEOMIX (Hanghøj et al.; 2016). The two datasets were generated independently: the hairpin results were obtained from bisulfite-converted DNA, and the epiPALEOMIX outputs were generated through estimation of the deamination rate from shotgun sequencing data. The two methods are potentially subject to opposite biases, although both biases result from DNA damage. In the bisulfite-conversion-based method (hairpin method), false negative results come the random deamination of 5mC into T, looking like unmethylated C that are also converted into T after bisulfite treatment. On the other hand, the damage-based method (epiPALEOMIX) relies on the deamination of 5mC into T, and this time false negatives are due to the incomplete removal of deaminated C (*i.e.* U). Surprisingly, although the scale of the methylation scores needed to be adjusted before doing any direct comparison (random low-frequency deamination leads to low methylation score values in epiPLAEOMIX, see Figure 5.6), the data generated using the two methods showed high consistency. This finding can be extremely valuable for paleo-methylome research, as it provides a way to validate results without using sample replicates. Indeed, it is a great challenge to retrieve from the paleontological/archaeological record multiple specimens of a particular taxon from the very same historical/pre-historical period. Furthermore, our results also suggest the outputs generated from the two methods seem compatible at least at the regional methylation level and could be potentially combined in future studies.

Second, we compared the ancient and modern methylome data. Although preliminary results suggested that taphonomy drives the ancient methylomes away from

the modern samples, consistent tissue specificity was observed after adjustment of data to account for hypomethylation of ancient samples. These results suggest that the biological signals could be maintained in ancient methylome at single-base resolution. Our results show that the single-base-resolution signal and biological features can be preserved in and recovered from sub-fossil mammal specimens, allowing the investigation of the methylome history over microevolutionary timescales.

5.4.4 Potential roles of methylation in bison-environment interactions

In order to investigate the epigenetic response to environmental changes, methylomes of ancient bison from two distinct geological epochs were interrogated and DMRs in two genes were detected. Interestingly, these two genes are involved in stress responses and muscle development, which are in line with the population history of steppe bison: they experienced dramatic climate oscillations, which were an environmental stressor, during the late Pleistocene (Shapiro et al.; 2004; Cooper et al.; 2015); and the fossil records of steppe bison exhibit extremely high morphological diversity without any known genetic explanation (McDonald; 1981; Shapiro et al.; 2004; Martin et al.; 2018). It is possible that when animals undergo rapid environmental changes, epigenetic mechanisms may play a role that precedes or complements genetic adaptation. Of note, the hairpin method generates both the methylome and the underlying genomic sequence. It means that should sample size be adequate, it would be possible to dissect molecular evolutionary processes behind adaptation to changing environment by decoupling signals from cytosine methylation and the underlying genetic sequence (Teschendorff and Relton; 2018).

We noted that only a limited number of DMRs were detected, which is likely due to two main reasons. First, the coverage of ancient methylomes is relatively low, as only 1/3 to 2/3 of the genome were covered, and the coverage was further decreased after filtering out loci with low coverage. Consequently, only a small proportion of CpG loci were retained for downstream analyses. It is highly possible that more DMRs will be identified with increased sequencing effort. Second, both the Holocene and late Pleistocene groups contain methylomes of bison that span a large temporal range (the minimal difference in age between any two samples within a group is 1 ky). The statistical power of DMR identification was greatly decreased with the inflated within-group variance. Also, it is unclear how gene expression would respond to the different methylation levels in bison; future studies on modern animals are necessary to verify the semi-stable states of the methylation profile of these two genes and subsequent consequences. While functional studies are necessary to characterise the roles of these two DMRs overlapping with genes,

it is conceivable that the expression of the genes was actively regulated in response to environmental stressors via dynamic methylation.

In conclusion, this study reconstructed a mammalian methylome history spanning two geological epochs by utilising a novel method that allows the recovery of methylation profiles from highly degraded DNA. As the methylation states of a single CpG site can significantly affect gene expression levels and animal phenotypes (Nile et al.; 2008; Claus et al.; 2012), our method has great potential to reveal fine epigenetic changes that may underly mammalian adaptation. We acknowledge that the sequencing depth of the ancient methylome data is relatively shallow, and further functional validation of the regulatory role of the DMRs is necessary. However, we demonstrated the power and potential of aDNA techniques in revealing animal evolution from an epigenetic perspective, and the hairpin method presented in this study represents a robust tool for obtaining high-resolution methylome data from various ancient specimens in the future.

Bibliography

Ball, M. P., Li, J. B., Gao, Y., Lee, J.-H., LeProust, E. M., Park, I.-H., Xie, B., Daley, G. Q. and Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, *Nature biotechnology* **27**(4): 361.

Bird, A. (2002). Dna methylation patterns and epigenetic memory, *Genes & development* **16**(1): 6–21.

Bock, C. (2012). Analysing and interpreting dna methylation data, *Nature Reviews Genetics* **13**(10): 705.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. et al. (2007). Patterns of damage in genomic dna sequences from a neandertal, *Proceedings of the National Academy of Sciences* **104**(37): 14616–14621.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M. and Pääbo, S. (2009). Removal of deaminated cytosines and detection of in vivo methylation in ancient dna, *Nucleic acids research* **38**(6): e87–e87.

Claus, R., Lucas, D. M., Stilgenbauer, S., Ruppert, A. S., Yu, L., Zucknick, M., Mertens, D., Bühler, A., Oakes, C. C., Larson, R. A. et al. (2012). Quantitative dna methylation analysis identifies a single cpg dinucleotide important for zap-70 expression and predictive of prognosis in chronic lymphocytic leukemia, *Journal of Clinical Oncology* **30**(20): 2483.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. and Bradshaw, C. J. (2015). Abrupt warming events drove late pleistocene holarctic megafaunal turnover, *Science* **349**(6248): 602–606.

Cropley, J. E., Dang, T. H., Martin, D. I. and Suter, C. M. (2012). The penetrance of an epigenetic trait in mice is progressively yet reversibly increased by selection and environment, *Proceedings of the Royal Society B: Biological Sciences* **279**(1737): 2347–2353.

Eaton, D. L. and Bammler, T. K. (1999). Concise review of the glutathione s-transferases and their significance to toxicology., *Toxicological sciences: an official journal of the Society of Toxicology* **49**(2): 156–164.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L. and Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual dna strands., *Proceedings of the National Academy of Sciences* **89**(5): 1827–1831.

Gokhman, D., Lavi, E., Prüfer, K., Fraga, M. F., Riancho, J. A., Kelso, J., Pääbo, S., Meshorer, E. and Carmel, L. (2014). Reconstructing the dna methylation maps of the neandertal and the denisovan, *Science* **344**(6183): 523–527.

Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J. S., Willerslev, E. and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epipaleomix, *Molecular biology and evolution* **33**(12): 3284–3298.

Hansen, K. D., Langmead, B. and Irizarry, R. A. (2012). Bsmooth: from whole genome bisulfite sequencing reads to differentially methylated regions, *Genome biology* **13**(10): R83.

Hayes, J. D. and Pulford, D. J. (1995). The glut athione s-transferase supergene family: regulation of gst and the contribution of the lsoenzymes to cancer chemoprotection and drug resistance part i, *Critical reviews in biochemistry and molecular biology* **30**(6): 445–520.

Jirtle, R. L. and Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility, *Nature reviews genetics* **8**(4): 253.

Johnson, W. E., Li, C. and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical bayes methods, *Biostatistics* **8**(1): 118–127.

Jones, P. A. (2012). Functions of dna methylation: islands, start sites, gene bodies and beyond, *Nature Reviews Genetics* **13**(7): 484.

Laird, C. D., Pleasant, N. D., Clark, A. D., Sneeden, J. L., Hassan, K. A., Manley, N. C., Vary, J. C., Morgan, T., Hansen, R. S. and Stöger, R. (2004). Hairpin-bisulfite pcr: assessing epigenetic methylation patterns on complementary strands of individual dna molecules, *Proceedings of the National Academy of Sciences* **101**(1): 204–209.

Leek, J. T., Johnson, W. E., Parker, H. S., Fertig, E. J., Jaffe, A. E., Storey, J. D., Zhang, Y. and Torres, L. C. (2018). *sva: Surrogate Variable Analysis*. R package version 3.28.0.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform, *bioinformatics* **25**(14): 1754–1760.

Llamas, B., Holland, M. L., Chen, K., Cropley, J. E., Cooper, A. and Suter, C. M. (2012). High-resolution analysis of cytosine methylation in ancient dna, *PloS one* **7**(1): e30226.

Martin, J. M., Mead, J. I. and Barboza, P. S. (2018). Bison body size and climate change, *Ecology and evolution* **8**(9): 4564–4574.

McDonald, J. N. (1981). *North American bison: their classification and evolution*, Univ of California Press.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing, *Cold Spring Harbor Protocols* **2010**(6): pdb–prot5448.

Nile, C. J., Read, R. C., Akil, M., Duff, G. W. and Wilson, A. G. (2008). Methylation status of a single cpg site in the il6 promoter is related to il6 messenger rna levels and rheumatoid arthritis, *Arthritis & Rheumatism* **58**(9): 2686–2693.

Pedersen, J. S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S. et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome, *Genome research* **24**(3): 454–466.

Quinlan, A. R. and Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* **26**(6): 841–842.

Rackham, O. J., Langley, S. R., Oates, T., Vradi, E., Harmston, N., Srivastava, P. K., Behmoaras, J., Dellaportas, P., Bottolo, L. and Petretto, E. (2017). A bayesian approach for analysis of whole-genome bisulfite sequencing data identifies disease-associated changes in dna methylation, *Genetics* **205**(4): 1443–1458.

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M. et al. (2014). Characterization of ancient and modern genomes by snp detection and phylogenomic and metagenomic analysis using paleomix, *Nature protocols* **9**(5): 1056.

Schubert, M., Lindgreen, S. and Orlando, L. (2016). Adapterremoval v2: rapid adapter trimming, identification, and read merging, *BMC research notes* **9**(1): 88.

Seguin-Orlando, A., Gamba, C., Der Sarkissian, C., Ermini, L., Louvel, G., Boulygina, E., Sokolov, A., Nedoluzhko, A., Lorenzen, E. D., Lopez, P. et al. (2015). Pros and cons of methylation-based enrichment methods for ancient dna, *Scientific reports* **5**: 11826.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J. et al. (2004). Rise and fall of the beringian steppe bison, *Science* **306**(5701): 1561–1565.

Skinner, M. K. (2015). Environmental epigenetics and a unified theory of the molecular aspects of evolution: a neo-lamarckian concept that facilitates neo-darwinian evolution, *Genome biology and evolution* **7**(5): 1296–1302.

Smith, O., Clapham, A. J., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2014). Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process, *Scientific reports* **4**: 5559.

Smith, R. W., Monroe, C. and Bolnick, D. A. (2015). Detection of cytosine methylation in ancient dna from five native american populations using bisulfite sequencing, *PloS one* **10**(5): e0125344.

Team, R. C. et al. (2013). R: A language and environment for statistical computing.

Teschendorff, A. E. and Relton, C. L. (2018). Statistical and integrative system-level analysis of dna methylation data, *Nature Reviews Genetics* **19**(3): 129.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter dna methylation in the human genome, *Nature genetics* **39**(4): 457.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.
URL: *http://ggplot2.org*

YANG, Y.-l., Yan, L., LIANG, R.-y., Rong, Z., Hong, A., MU, Y.-l., Yang, S.-l., Kui, L. and TANG, Z.-l. (2014). Dynamic expression of microrna-127 during porcine prenatal and postnatal skeletal muscle development, *Journal of Integrative Agriculture* **13**(6): 1331–1339.

Zhao, L., Sun, M.-a., Li, Z., Bai, X., Yu, M., Wang, M., Liang, L., Shao, X., Arnovitz, S., Wang, Q. et al. (2014). The dynamics of dna methylation fidelity during mouse embryonic stem cell self-renewal and differentiation, *Genome research* **24**(8): 1296–1307.

5.5 Supplementary materials

5.5.1 Figure S5.1
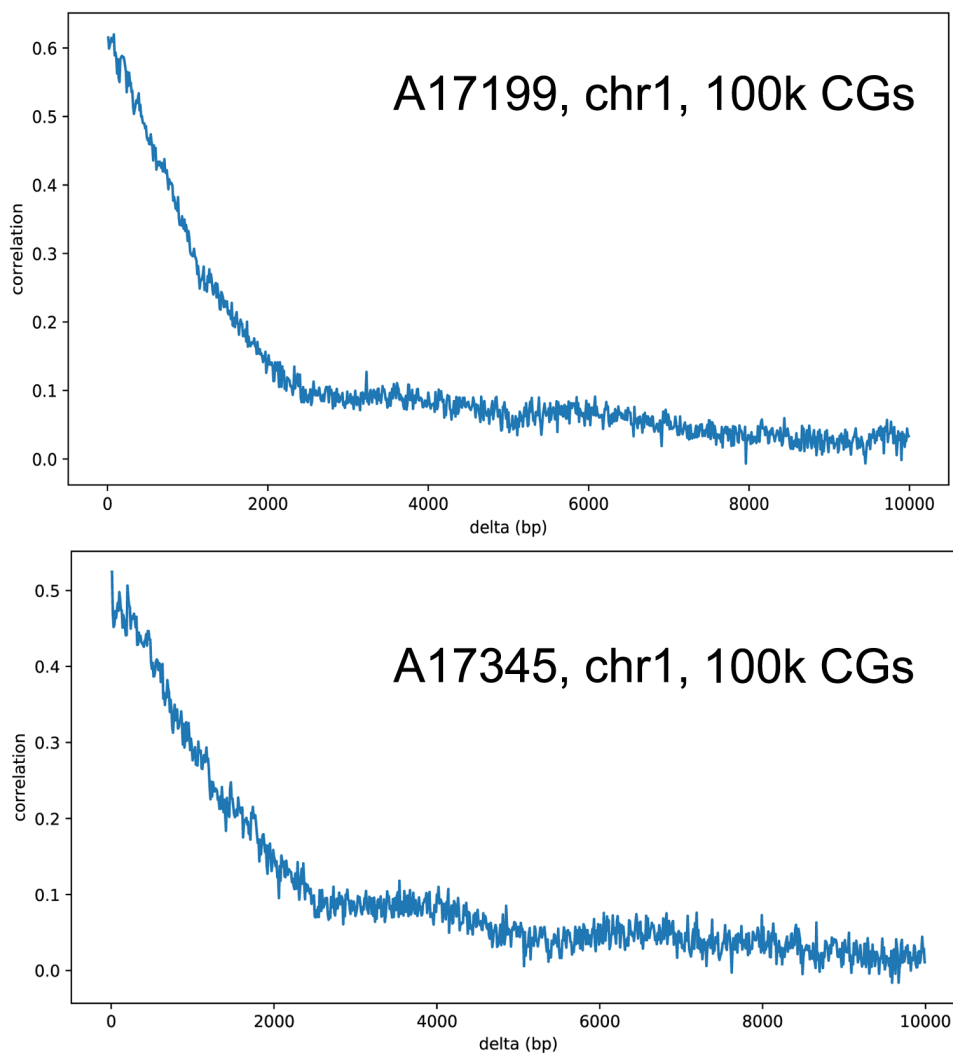
The correlation of CpG sites across varied distances of a modern (top) and an ancient (bottom) petrosal sample. The correlation reaches to a baseline at about 2kb.

5.5.2 Figure S5.2
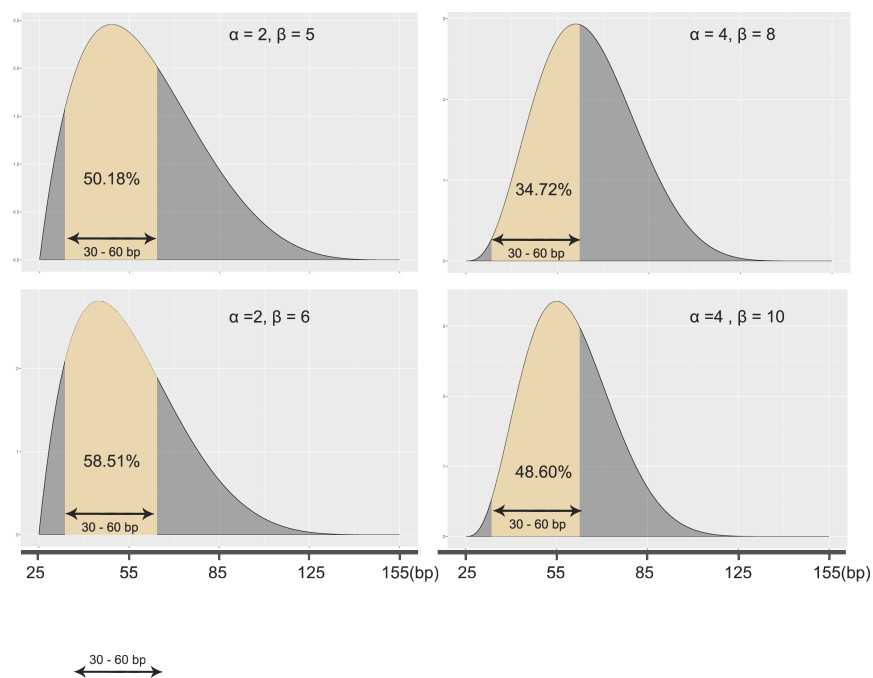
Examples of simulated ancient DNA fragment length distribution and the proportion of short fragments (30-60 bp).

5.5.3   Figure S5.3
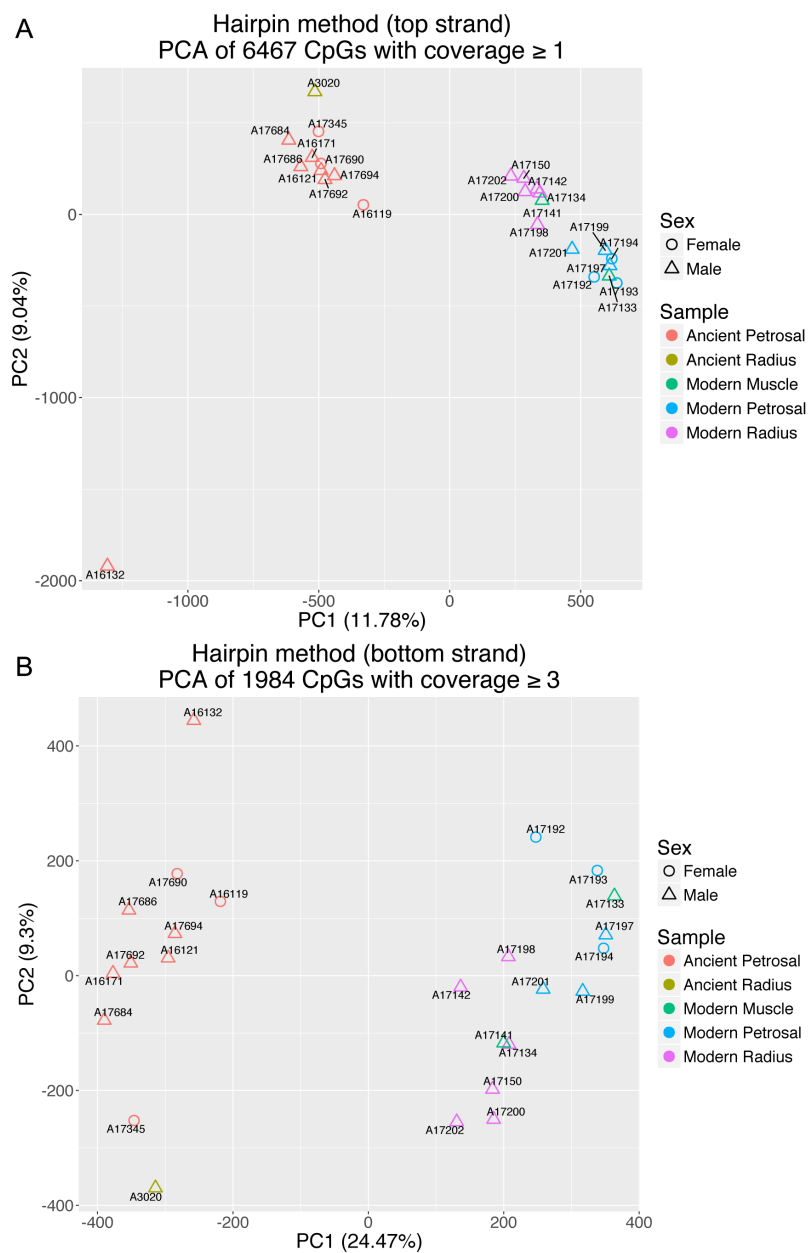
PCA on the methylation level of the CpG sites overlapped by all the samples. A showed the PCA on all the CpG sites, and B only include the sites with a coverage >2. Note that sample A16132 appeared to be an outliner and clustered away from the rest of ancient samples in both figure A and B.

5.5.4 Table S5.1

Sample description

ACAD ID	Other ID	Genus	Species	Sex	Tissue	Type	Age (ky)	Group	Country	State/province
A16119	N/A	*Bison*	*priscus*	Female	petrosal	ancient	27	LGM	Canada	Yukon
A17345	N/A	*Bison*	*priscus*	female	petrosal	ancient	9	postLGM	USA	Idaho
A17690	N/A	*Bison*	*priscus*	female	petrosal	ancient	30	preLGM	USA	Alaska
A16121	N/A	*Bison*	*priscus*	male	petrosal	ancient	15	LGM	Canada	Yukon
A16132	N/A	*Bison*	*priscus*	male	petrosal	ancient	50	preLGM	Canada	Yukon
A16171	N/A	*Bison*	*priscus*	male	petrosal	ancient	41	preLGM	Canada	Yukon
A17684	N/A	*Bison*	*priscus*	male	petrosal	ancient	1	postLGM	Canada	Alberta
A17686	N/A	*Bison*	*priscus*	male	petrosal	ancient	3	postLGM	Canada	Alberta
A17692	N/A	*Bison*	*priscus*	male	petrosal	ancient	26	LGM	USA	Alaska
A17694	N/A	*Bison*	*priscus*	male	petrosal	ancient	5	postLGM	USA	Alaska
A3020	N/A	*Bison*	*priscus*	male	radius	ancient	8	postLGM	USA	Minnesota
A17192	Bison1	*Bison*	*bison*	female	petrosal	modern	N/A	Modern	USA	Kentucky
A17193	Bison2	*Bison*	*bison*	female	petrosal	modern	N/A	Modern	USA	Kentucky
A17194	Bison3	*Bison*	*bison*	female	petrosal	modern	N/A	Modern	USA	Kentucky
A17133	Bison-01 92827	*Bison*	*bison*	male	muscle	modern	N/A	Modern	USA	Kentucky
A17134	Bison-01 92827	*Bison*	*bison*	male	radius	modern	N/A	Modern	USA	Kentucky
A17141	Bison-02 92828	*Bison*	*bison*	male	muscle	modern	N/A	Modern	USA	Kentucky
A17142	Bison-02 92828	*Bison*	*bison*	male	radius	modern	N/A	Modern	USA	Kentucky
A17150	Bison-03 92829	*Bison*	*bison*	male	radius	modern	N/A	Modern	USA	Kentucky
A17197	Bison6	*Bison*	*bison*	male	petrosal	modern	N/A	Modern	USA	Kentucky
A17198	Bison6	*Bison*	*bison*	male	radius	modern	N/A	Modern	USA	Kentucky
A17199	Bison7	*Bison*	*bison*	male	petrosal	modern	N/A	Modern	USA	Kentucky
A17200	Bison7	*Bison*	*bison*	male	radius	modern	N/A	Modern	USA	Kentucky
A17201	Bison8	*Bison*	*bison*	male	petrosal	modern	N/A	Modern	USA	Kentucky
A17202	Bison8	*Bison*	*bison*	male	radius	modern	N/A	Modern	USA	Kentucky

5.5.5 Table S5.2

Sequencing details

ACAD ID	Sample ID	Reference	Reads	Folded	Foldratio	Average lenth (bp)	Mapped	Unique	Coverage	Method
A3020	A3020-AncRAD8k	Cow_UMD3_1	147,624,653	96,741,037	65.53%	53.12	48,650,097	20,875,656	0.50	Hairpin
A16119	A16119-AncPET27k	Cow_UMD3_1	142,186,200	96,577,697	67.92%	57.64	54,429,610	19,476,471	0.53	Hairpin
A16121	A16121-AncPET15k	Cow_UMD3_1	202,885,707	127,430,061	62.81%	51.30	78,159,320	45,353,142	1.11	Hairpin
A16132	A16132-AncPET50k	Cow_UMD3_1	140,043,994	107,465,734	76.74%	59.65	47,486,241	6,268,208	0.17	Hairpin
A16171	A16171-AncPET41k	Cow_UMD3_1	243,644,648	107,942,001	44.30%	53.88	68,287,514	52,493,779	1.37	Hairpin
A17345	A17345-AncPET9k	Cow_UMD3_1	404,281,429	246,950,035	61.08%	53.21	137,135,582	70,599,569	1.71	Hairpin
A17684	A17684-AncPET1k	Cow_UMD3_1	122,797,128	77,925,440	63.46%	55.63	42,168,159	28,211,646	0.67	Hairpin
A17686	A17686-AncPET3k	Cow_UMD3_1	116,821,015	75,750,289	64.84%	52.10	40,970,498	29,537,261	0.67	Hairpin
A17690	A17690-AncPET30k	Cow_UMD3_1	142,776,142	40,626,756	28.45%	51.30	23,260,120	17,260,428	0.43	Hairpin
A17692	A17692-AncPET26k	Cow_UMD3_1	130,511,294	104,148,096	79.80%	51.72	65,373,567	29,598,877	0.66	Hairpin
A17694	A17694-AncPET5k	Cow_UMD3_1	138,689,446	61,581,415	44.40%	54.35	39,836,971	33,229,959	0.83	Hairpin
A17133	A17133-M1MUS	Cow_UMD3_1	110,290,913	58,869,714	53.38%	143.97	48,901,475	23,531,701	1.26	Hairpin
A17134	A17134-M1RAD	Cow_UMD3_1	146,968,405	96,551,611	65.70%	150.11	80,881,757	26,256,480	1.40	Hairpin
A17141	A17141-M2MUS	Cow_UMD3_1	158,205,559	98,192,151	62.07%	137.15	79,212,429	39,815,893	2.00	Hairpin
A17142	A17142-M2RAD	Cow_UMD3_1	155,079,848	87,526,305	56.44%	155.57	73,259,729	18,353,746	0.98	Hairpin
A17150	A17150-M3RAD	Cow_UMD3_1	164,491,746	103,920,421	63.18%	138.21	86,210,842	35,172,282	1.84	Hairpin
A17192	A17192-F1PET	Cow_UMD3_1	133,735,300	57,492,355	42.99%	147.02	48,567,254	21,641,871	1.14	Hairpin
A17193	A17193-F2PET	Cow_UMD3_1	85,021,739	60,070,422	70.65%	135.84	50,146,648	20,105,674	1.02	Hairpin
A17194	A17194-F3PET	Cow_UMD3_1	153,158,717	69,030,586	45.07%	157.80	57,595,551	22,111,415	1.14	Hairpin
A17197	A17197-M6PET	Cow_UMD3_1	101,881,661	44,517,025	43.69%	137.19	36,500,452	17,751,400	0.95	Hairpin
A17198	A17198-M6RAD	Cow_UMD3_1	211,202,762	94,955,272	44.96%	228.65	78,324,630	41,090,244	2.09	Hairpin
A17199	A17199-M7PET	Cow_UMD3_1	119,634,164	59,001,937	49.32%	143.63	48,800,619	17,468,605	0.89	Hairpin
A17200	A17200-M7RAD	Cow_UMD3_1	196,833,572	116,519,874	59.20%	173.19	95,196,576	30,785,501	1.61	Hairpin
A17201	A17201-M8PET	Cow_UMD3_1	144,900,507	103,178,837	71.21%	141.59	85,207,562	31,775,783	1.68	Hairpin
A17202	A17202-M8RAD	Cow_UMD3_1	217,559,780	128,208,341	58.93%	128.83	104,710,304	34,494,657	1.76	Hairpin
A16121	A16121-AncPET15k	Cow_UMD3_1	54,352,396	N/A	N/A	74.59	34,025,834	33,706,735	0.33	WGBS
A16121	A16121-AncPET15k	Cow_UMD3_1	915,674,432	N/A	N/A	60.23	553,504,609	515,903,243	0.89	Shotgun sequencing
A16171	A16171-AncPET41k	Cow_UMD3_1	330,310,184	N/A	N/A	68.81	189,849,805	184,985,236	0.87	Shotgun sequencing
A3020	A3020-AncRAD8k	Cow_UMD3_1	302,137,632	N/A	N/A	65.21	115,446,712	111,897,801	0.78	Shotgun sequencing

5.5.6   Table S5.3

Simulated length distribution of aDNA fragments (25-150bp)

Simulation parameters		Simulation results		
alpha	beta	Number of simulated reads	Mean fragment length (bp)	Reads with a length of 30-60bp (%)
2	3	100,000	77.3	29.00%
2	4	100,000	36.7	39.99%
2	5	100,000	62.3	50.18%
2	6	100,000	57.7	58.51%
3	3	100,000	90.5	12.62%
3	4	100,000	81.3	20.11%
3	6	100,000	68.7	37.30%
3	8	100,000	60.8	53.18%
4	4	100,000	90.5	9.13%
4	6	100,000	77.2	20.76%
4	8	100,000	68.6	34.72%
4	10	100,000	62.4	48.60%
4	12	100,000	57.7	61.09%

5.5.7 Table S5.4

DMRs identified between Holocene and Pleistocene bison methylomes.

Chr	Start	End	Number of CGs	Width (bp)	MeanDiff	Pleistocene mean	Holocene mean	Genomic features
chr3	8098565	8098998	11	434	10.17%	54.78%	44.60%	5' shelf (chr3: 8098185-8100185); 3' shore (chr3: 8097119-8099119)
chr3	8100599	8100996	11	398	10.22%	63.67%	53.45%	3' shelf (chr3: 8099119-8101119); 5' shore (chr3: 8100185-8102185)
chr6	6773018	6773757	9	740	10.23%	60.88%	50.64%	NA
chr8	71017416	71018193	6	778	-10.08%	49.65%	59.73%	NM_001102327, Bos taurus tumor necrosis factor receptor superfamily, member 10d, (chr8: 70908188-71192762)
chr17	73294290	73295019	18	730	17.89%	74.25%	56.36%	NM_001206623; Bos taurus glutathione S-transferase, theta 4 (GSTT4) (chr17:73292732-73297355)
chr25	32377049	32377077	7	59	16.78%	32.16%	15.38%	3' shelf (chr25:32376255-32378255)

Statement of Authorship

Title of Paper	Recovery of RNA from a 30,000-year-old bison bone
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	In prepevation for submission

Principal Author

Name of Principal Author (Candidate)	Yichen Liu
Contribution to the Paper	Lab work, data analyses, and wrote the manusript
Overall percentage (%)	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 05/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Steve Richards
Contribution to the Paper	Protocol design and optimisation
Signature	Date 01-03-2019

Name of Co-Author	Alan Cooper
Contribution to the Paper	Data interpretation
Signature	Date 01/03/2019

Please cut and paste additional co-author panels here as required.

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

iv.   the candidate's stated contribution to the publication is accurate (as detailed above);

v.   permission is granted for the candidate in include the publication in the thesis; and

vi.   the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Data interpretation; reviewed the manuscript		
Signature		Date	04/03/19

Chapter 6

Recovery of RNA from a 30,000-year-old bison bone

Abstract

The ability to recover non-genetic information such as proteomes, transcriptomes and epigenomes from ancient animals is important for tracking how they adapted and responded to past environment changes. Currently, ancient RNA (aRNA) has been recovered only from a few exceptionally well-preserved biological specimens, including viruses, plants, humans and canids. Here, we extracted RNA from a ~30,000-year-old bison bone, pushing the empirical time limit of RNA preservation by a further 15,000 years compared to previous reports. We evaluated the potential contamination by co-extracted DNA and the necessity of a RNA molecule end repair step for sequencing library construction. Our results also suggested that damages other than the absence of 5′ phosphate are likely present in aRNA and could hinder sequencing library construction. This study highlighted the feasibility of recovering aRNA from sub-fossil bones—which represent most of the animal sub-fossil record—and the necessity for a better understanding of aRNA properties.

Key words: Ancient RNA, bison, RNA extraction

6.1 Introduction

DNA is now routinely recovered from sub-fossil animals, plants, and sediments, allowing the examination of ancient animal genomes, commensal microbes, and even chemical modifications of DNA such as cytosine methylation (Slon et al.; 2017; Miller et al.; 2008; Adler et al.; 2013; Gokhman et al.; 2014; Estrada et al.; 2018). Proteins can also be recovered from sub-fossil remains and it is arguably be better preserved than DNA in some cases (Cappellini et al.; 2018). Together, these multiple sources of ancient biomolecular information provide important insights into past population demography, evolution, diet, and behaviour (Larson et al.; 2010; Welker et al.; 2015; Adler et al.; 2013; Soubrier et al.; 2016; Weyrich et al.; 2017; Hendy et al.; 2018). However, the third macromolecule of the central dogma of molecular biology, RNA, was only recovered from a handful of sub-fossil samples (Fordyce et al.; 2013; Ng et al.; 2014; Smith et al.; 2014; Keller et al.; 2017; Smith et al.; 2017, 2019), despite the crucial role of RNA in reflecting (messenger RNA; mRNA) and regulating (non-coding RNA; ncRNA) gene expression levels (Smirnova et al.; 2005; de Sousa Abreu et al.; 2009). Namely, ancient RNA (aRNA) has been recovered from plant seed endosperm, which is suggested to favour the preservation of nucleic acids (Rollo et al.; 1994; Fordyce et al.; 2013). RNA has also been recovered from exceptionally well-preserved animal tissues such as in the case of the Tyrolean Iceman (Keller et al.; 2017), as well as a few historical wolf skins and a 14,300-year-old Pleistocene canid liver (Smith et al.; 2019). One major challenge is that ribonucleases (RNase) can be released during post-mortem tissue autolysis and are generally ubiquitously present in the environment (Girija and Sreenivasan; 1966; D'Alessio and Riordan; 1997).

It seems that RNA can survive for thousands of years in animal tissues, but our understanding of the property of aRNA is still very limited. Unlike extraction procedures optimised for ancient DNA (aDNA) (Rohland and Hofreiter; 2007), RNA extraction from ancient remains was largely adapted from RNA extraction protocols that are designed for fresh tissues. However, aRNA extraction suffers from more challenges than modern RNA extraction. A major concern is the amount of RNA that survived in ancient tissues is likely extremely low, which requires the RNA extraction protocol to be highly sensitive. Issues in modern RNA extraction (Gayral et al.; 2011), such as DNA contamination, can also occur during aRNA extraction, as the specimens used for aRNA extraction are usually well-preserved and may contain a high proportion of endogenous DNA. Furthermore, aDNA damage is well characterised (Dabney et al.; 2013), but current knowledge on aRNA damage is minimal: increased cytosine deamination levels at the ends of aRNA molecules was observed, and total aRNA has an overall higher GC content than aRNA that mapped

to coding regions (likely transcripts) (Smith et al.; 2019). However, unlike for aDNA, we are still far from leveraging any RNA damage pattern for the authentication of aRNA or from preferential enrichment of damaged aRNA from ancient specimens.

Here, we successfully obtained RNA from a 30,000-year-old bison astragalus bone. Stringent controls were included to reduce potential contamination and to assess RNA extraction efficiency. We evaluated the proportion of contaminating DNA in the RNA extracts and the necessity of including an end repair step to increase the yield during RNA library construction for high throughput sequencing.

6.2 Methods

6.2.1 Sample description

A bison astragalus (ACAD ID: A3133) was collected in the Yukon Territory, Canada. This specimen was housed within the Australian Centre for Ancient DNA (ACAD) at The University of Adelaide and stored at –20°C until the RNA extraction. A3133 was dated to 26,360 ±220 uncalibrated radiocarbon years before present using a ^{14}C half-life of 5568 years, as described previously Llamas et al. (2012). After calibration using OxCal13, the mean age of the sample is 30,612 calBP (calibrated years before present), with a 95.4% probability that it is in the range 31.044–30,092 calBP.

6.2.2 Sample preparation

Sample A3133 was processed at the ultra-clean ancient DNA laboratory at ACAD. This sample was prepared in two ways: (1) ˜100 mg of dense cortical bone was put into a clean stainless-steel canister containing a stainless-steel ball and pulverised using a Mikro-Dismembrator at a frequency of 1,000 rpm for 1 min and then cooled down at 4°C for 3 min. This step was repeated 3–5 times until the bone was fully pulverised; (2) as the pulverisation step can heat the sample and thus potentially lead to RNA degradation, we fragmented another ˜100 mg of dense cortical bone into small pieces (2×2×0.5 mm) using a hammer.

6.2.3 RNA extraction

To test the efficiency of the RNA extraction protocol, we included three bone powder samples (˜100 mg each), three bone fragments samples (100mg each), four empty tubes as extraction negative controls, and 2 μL of microRNA Control from the NEXTflex Small RNA-Seq Kit (Bioo Scientific) as an extraction positive control (Table 6.1). All the bone samples were subsamples from specimen A3133, and all the controls were treated as if they were bone samples. Additional bone samples

were extracted for the library construction (see 6.2.5 Sequencing library preparation) (Table 6.1).

Samples were lysed in a 2-ml tube that contained 500 μL of 0.5 M ethylenediaminetetraacetic acid (EDTA, pH=8), 0.25 μL of Tween, and 100 units of RNase inhibitor and left to rotate at room temperature for 20 hours. Then the lysates were centrifuged at 4500 rpm for 5 min and each supernatant was transferred into a new tube. RNA was extracted from the supernatants using miRNeasy Serum/Plasma Kit (Qiagen) following the manufacturer's instructions. Then the RNA extracts were quantified using a Bioanalyzer Small RNA Analysis kit (Agilent).

Table 6.1 Sample used for RNA extraction and library construction

Specimen	Discription	Weight (mg)	DNase digestion	PNK	Results
N/A	Negative control 1	N/A	N/A	N/A	
A3133	Bone fragments 1	100	N/A	N/A	
A3133	Bone fragments 2	110	N/A	N/A	
A3133	Bone fragments 3	100	N/A	N/A	
N/A	Negative control 2	N/A	N/A	N/A	Figure 6.1
N/A	Negative control 3	N/A	N/A	N/A	
A3133	Bone power 1	100	N/A	N/A	
A3133	Bone power 2	100	N/A	N/A	
A3133	Bone power 3	110	N/A	N/A	
N/A	Negative control 4	N/A	N/A	N/A	
N/A	Negative control	N/A	N	N	
A3133	Bone power 4	110	Yes	Yes	
A3133	Bone power 5	120	Yes	N	
A3133	Bone power 6	110	N	Yes	Figure 6.3
A3133	Bone power 7	110	N	N	
N/A	Positive control 1	N/A	Yes	N	
N/A	Positive control 2	N/A	N	N	
N/A	Library poitive control	N/A	N/A	N	

6.2.4 Verification for the presence of RNA

In general, DNA may be present in ancient animal bones and could be co-extracted with RNA, and the Bioanalyzer Small RNA Analysis kit cannot accurately distinguish RNA from DNA (Figure S6.1). In fact, DNA extracted from A3133 has been used in two published studies (Llamas et al.; 2012; Richards et al.; 2019). Therefore, we further tested the presence of DNA and RNA in the extracts using RNase and DNase digestion.

RNase digestion: 1 μL of RNA extract was mixed with 0.5 μL RNase A (10 mg/mL); the no-treatment control was 1 μL of RNA extract mixed with 0.5 μL ddH$_2$O. The reactions were incubated at 37 °C for 15 min.

DNase digestion: 1 μL of RNA extract was mixed with 3 μL RNase-free DNase I (2,000 units/ml, NEB), 0.5 μL 10X DNase buffer, and 0.5 μL RNase inhibitor; the no-treatment control was 1 μL of RNA extract mixed with 3.5 μL RNase-free water,

and 0.5 μL RNase inhibitor. The reactions were incubated at room temperature for 20 min.

The resulting reactions from DNase and RNase digestion were quantified using a Bioanalyzer Small RNA Analysis kit.

6.2.5 Sequencing library preparation

To investigate if ancient RNA is compatible with high throughput sequencing (HTS) techniques, we prepared sequencing libraries using the RNA extracts. For the construction of the sequencing libraries, four additional bone power samples, an extraction negative control, and two extraction positive controls (2 μL of microRNA Control from the NEXTflex Small RNA-Seq Kit, Bioo Scientific) were extracted as described in 6.2.3 RNA extraction (Table 6.1).

To examine the presence of DNA in the RNA extracts and its potential influence on library construction, an additional on-column DNase digestion step was performed on two of the RNA extracts and one extraction positive control before the final elution step during RNA extraction (Table 1). Specifically, 20 μL of DNase solution (containing 1 unit of DNase) was added to the column and incubated at room temperature for 15 min. Then the RNA was washed using 700 μL of Buffer RWT (miRNeasy Serum/Plasma Kit (Qiagen)), 500 μL of Buffer RPE (miRNeasy Serum/Plasma Kit (Qiagen)), and 500 μL of 80% ethanol. Purified RNA was eluted using 14 μL of RNase-free water.

As the RNA extracted from ancient remains may be damaged and impact negatively the ligation of sequencing adapters, we performed an end repair step on two of the RNA extracts using T4 polynucleotide kinase (T4 PNK; New England Biolabs, NEB) before the library preparation (Table 6.1). Each repair reaction contained 3 μL of 10x NEB2 buffer, 3 μL of 10 mM ATP, 9.75 μL of ddH$_2$O, 0.75 of RNase inhibitor (40 unit/μL), 1.5 μL of PNK (10 unit/μL), and 12 μL of RNA extract. The reactions were incubated at 25°C for 15 min. Then each reaction was loaded to an RNeasy MinElute spin column and washed using 700 μL of Buffer RWT, 500 μL of Buffer RPE, and 500 μL of 80% ethanol. Spin columns and buffers were from the miRNeasy Serum/Plasma Kit (Qiagen). Purified RNA was eluted using 14 μL of RNase-free water.

In addition to the RNA extracts and controls, a library positive control (0.5 μL of microRNA Control from the NEXTflex Small RNA-Seq Kit) was included during library preparation. Then sequencing libraries were prepared using the NEXTflex Small RNA-Seq Kit following the manufacturer's instructions. An additional 20 cycles of PCR amplification was performed at the end of libraries preparation. Each PCR reaction contained: 5 μL of 10X Gold buffer (Thermo Fisher), 0.5 μL of Am-

pliTaq Gold enzyme (Thermo Fisher), 5 μL of 25 mM magnesium chloride (MgCl2), 1.5 μL of Illumina IS5 primer, 1.5 μL of Illumina IS6 primer, and 3 μL of prepared libraries (Meyer and Kircher 2010). The reactions were performed under the following conditions: 95°C for 5 minutes; 20 cycles of 95°C for 30s, 57°C for 30s, 72°C for 45s; and 72°C for 5 minutes. PCR products were cleaned using Ampure (NEB) to remove the primers and enzymes, and DNA concentration was measured using TapeStation (Agilent).

6.3   Results

6.3.1   RNA extraction from ancient bison bone samples

RNA was detected in all bone extractions regardless of the preparation method (pulverisation or fragmentation), while no visible RNA was detected from extraction negative controls (Figure 6.1). The size of the RNA extracted from the extraction positive control was consistent with the input miRNA (22nt). The RNA extracted from the ancient bison bone were very short, with the lengths of the majority of the RNA fragments ranging from 10 nt to 40 nt. The average concentration of RNA extracted from the bone fragments (~100 mg) was 1441 pg/μL with a standard deviation of 261 pg/μL, and that of the bone powder (~100mg) was 4373 pg/μL with a standard deviation of 369 pg/μL. We were concerned that the heat generated from the pulverisation step may lead to RNA degradation. However, the result suggested that despite the potential heat, extracts from the pulverised samples contained more RNA than extracts from the fragmented samples. The higher amount of RNA released from bone powder was likely due to the increased effective contact area between the sample and the EDTA solution.

6.3.2   Test for the presence of RNA and DNA

DNA can be co-extracted with RNA during the RNA extraction process, and fluorescent dyes can bind to both DNA and RNA and thus DNA can be detected by a Bioanalyzer RNA chip (Figure S6.1). Therefore, we tested for the presence of DNA and RNA using DNase and RNase digestion (Figure 6.2). The concentration of RNA in the extract before and after RNase digestion was 1067 pg/μL and 172 pg/μL, respectively (Figure 6. 2A), suggesting the majority (84.0%) of nucleic acids in the extract was RNA while the remaining 16.0% was most likely DNA. We then treated some of the RNA extract with DNase. RNA concentration before and after DNase treatment was 876 pg/μL and 674 pg/μL, suggesting that 76.0% of the nucleic acids in the RNA extract is made of RNA, while 24.0% (the missing fraction after DNase treatment) is DNA (Figure 6. 2B). Overall, these results suggest that
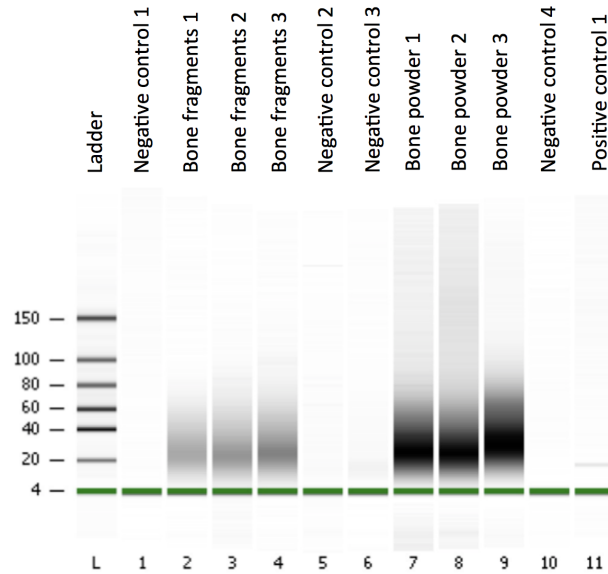
Figure 6.1: RNA extracted from bone fragments, bone powder, and positive and negative controls.

some DNA (16–24%) was co-extracted with RNA, and thus a DNase digestion step is necessary for the removal of DNA contamination.

6.3.3 End-repair is required during library construction using aRNA

To evaluate the potential impact of RNA damage during library preparation, we compared the RNA sequencing library yield of RNA extracts with or without an end repair step using T4 PNK (Figure 6.3). Regardless of whether a DNase digestion step was performed or not, the library yields of end-repaired RNA were 5 to 33 times higher than those without end repair (Figure 6. 3, B vs C and D vs E). These results suggest the absence of 5' phosphate in the majority of the extracted RNA molecules and an end repair step can greatly increase the efficiency of sequencing library construction when using aRNA. The effect of DNase digestion is less clear, as the library yield was higher in the DNase-digested sample when end repair was performed (Figure 6.3, B and D), while the library yield was lower in the DNase-digested sample when end repair was skipped (Figure 6.3, C and E). It is possible that DNase digestion had minimal effects on the library yield, thus the observation was mainly driven by stochastic variances. Further investigation (*e.g.*, include replicates) is required to provide more clues on the impact of DNase digestion on aRNA library construction. The negative control also yielded small yields after library preparation, suggesting it is necessary to control and monitor potential contamination, especially when the concentration of input RNA is extremely low.
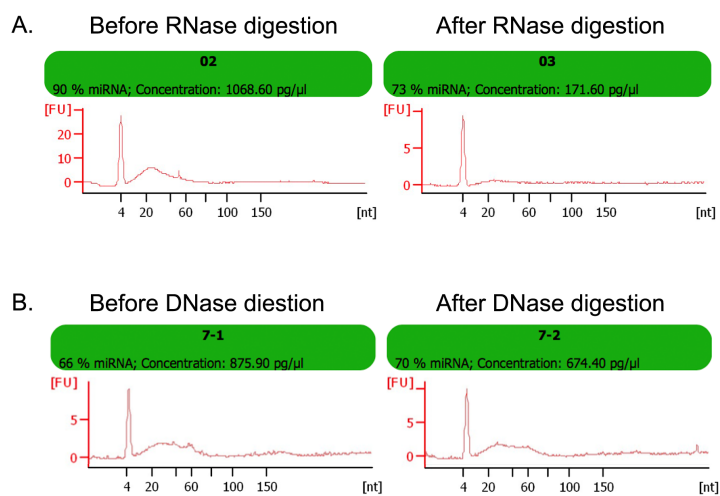
Figure 6.2: A. Concentration and fragment size distribution of the RNA extract without (left) or with (right) RNase digestion. B. Concentration and fragment size distribution of the RNA extract without (left) or with (right) DNase digestion.
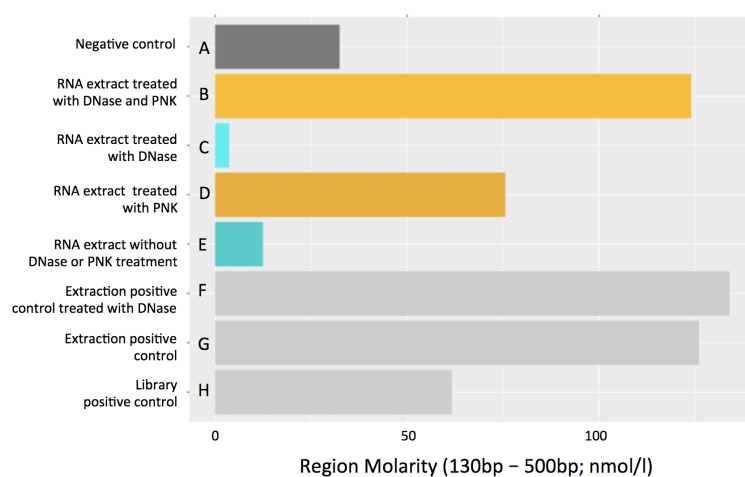

Figure 6.3: Concentration of libraries prepared using RNA extracts and controls.

6.4 Discussion

In this study, we show that RNA can be recovered from a 30-ky-old mammal bone. The presence of RNA was tested in multiple ways and the feasibility of constructing sequencing libraries using aRNA was validated. One of the initial concerns from this study is the potential contamination of DNA. This is mainly because DNA can be co-extracted with RNA, and DNA contamination may be exacerbated by the likely extremely low concentration of aRNA in ancient specimens. Staining nucleic acids using fluorescent dye (*e.g.*, Agilent Bioanalyzer system) can provide accurate quantification when the input sample is pure DNA or RNA. However, the fluorescent dyes only have limited specificity to DNA and RNA: DNA can be stained by a RNA dye (Bioanalyzer High Sensitivity RNA kit) and RNA can be stained by a DNA dye (Bioanalyzer High Sensitivity DNA kit). As our aRNA extracts were potentially a mixture of aDNA and aRNA, fluorescent dye staining could not provide accurate quantification of either nucleic acid. Therefore, we used RNase or DNase digestion to test the presence of DNA and RNA. The RNase and DNase digestions showed consistent results: the majority of the extract was RNA (~80%), with a relatively small amount of DNA contamination (~20%). This result validated the recovery of aRNA from ancient specimen, while also highlighting the necessity for monitoring potential DNA contamination, such as by using DNase digestion to remove residual DNA or using RNase digestion to evaluate DNA contamination. However, DNase may be contaminated by RNase (Grady, Campbell et al. 1980), and it has been reported that DNase digestion can potentially degrade ancient RNA (Smith, Dunshea et al. 2019). In that case, it is possible that our aRNA extracts are pure and do not contain aDNA, and that the quantification results only reflect the undesirable effects of both DNase and RNase.

This study also provided new insights into aRNA damage. Given the prevalence and resilience of RNase in animal tissues and the environment, it is highly likely that the majority of RNA in animal tissues is rapidly degraded after the death of the organism. Indeed, our aRNA yields are low and fragment length is short: the best yield in this study was ~50 ng of aRNA isolated from 100 mg of bone powder (while the RNA yield from ~100 mg fresh tissue can be higher than 100 μg, e.g., (Reno, Marchuk et al. 1997)), and the mean fragment length of isolated aRNA was ~20 nt. The aRNA also showed signs of damage, as end-repair could greatly increase the yields of constructed libraries. Although the end repair is not a routine step for RNA library construction, here we highly recommend including this step to increase the library yield when using aRNA. We note that the concentrations of RNA extracts from bone samples were much higher than the extraction positive control (Figure 6. 1), while the library yields for the two experiments were similar even in the case of

aRNA end repair (Figure 6. 3). This is likely due to other types of RNA damage that hindered library construction. For example, cross-linking, a type of damage identified in aDNA extracts (Pääbo 1989), could happen between RNA molecules as well (Harris and Christian 2009). The cross-linked RNA possibly present in aRNA could decrease the library construction efficiency. Additionally, depurination of RNA can result in abasic sites that hamper the amplification (Barbieri, Ferreras et al. 1992). Current understanding of aRNA damage is overall very limited, and the preservation and damage of RNA in sub-fossils remains needs to be further investigated.

This study represents one of few that successfully recovered RNA from extremely old mammal specimens. We show that optimised sample preparation methods can improve the RNA yield, and an end-repair step is crucial for sequencing library construction. Our result also suggested that the processing of aRNA into sequencing libraries is likely limited by unknown RNA damage and further optimisation of the protocol is required to maximise the retrieval of aRNA information. Future efforts will be placed on the analyses and interpretation of obtained sequencing data. In conclusion, we highlight the survival of RNA in ancient mammal remains and the potential to recover partial transcriptome information from ancient mammals.

Bibliography

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A., Alt, K. W. et al. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the neolithic and industrial revolutions, *Nature genetics* **45**(4): 450.

Cappellini, E., Prohaska, A., Racimo, F., Welker, F., Pedersen, M. W., Allentoft, M. E., de Barros Damgaard, P., Gutenbrunner, P., Dunne, J., Hammann, S. et al. (2018). Ancient biomolecules and evolutionary inference, *Annual review of biochemistry* **87**: 1029–1060.

Dabney, J., Meyer, M. and Pääbo, S. (2013). Ancient dna damage, *Cold Spring Harbor perspectives in biology* **5**(7): a012567.

D'Alessio, G. and Riordan, J. F. (1997). *Ribonucleases: structures and functions*, Academic Press.

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. and Vogel, C. (2009). Global signatures of protein and mrna expression levels, *Molecular BioSystems* **5**(12): 1512–1526.

Estrada, O., Breen, J., Richards, S. M. and Cooper, A. (2018). Ancient plant dna in the genomic era, *Nature plants* **4**(7): 394.

Fordyce, S. L., Avila-Arcos, M. C., Rasmussen, M., Cappellini, E., Romero-Navarro, J. A., Wales, N., Alquezar-Planas, D. E., Penfield, S., Brown, T. A., Vielle-Calzada, J.-P. et al. (2013). Deep sequencing of rna from ancient maize kernels, *PLoS One* **8**(1): e50961.

Gayral, P., Weinert, L., Chiari, Y., Tsagkogeorga, G., Ballenghien, M. and Galtier, N. (2011). Next-generation sequencing of transcriptomes: a guide to rna isolation in nonmodel animals, *Molecular Ecology Resources* **11**(4): 650–661.

Girija, N. and Sreenivasan, A. (1966). Characterization of ribonucleases and ribonuclease inhibitor in subcellular fractions from rat adrenals, *Biochemical Journal* **98**(2): 562.

Gokhman, D., Lavi, E., Prüfer, K., Fraga, M. F., Riancho, J. A., Kelso, J., Pääbo, S., Meshorer, E. and Carmel, L. (2014). Reconstructing the dna methylation maps of the neandertal and the denisovan, *Science* **344**(6183): 523–527.

Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C. and Collins, M. J. (2018). A guide to ancient protein studies, *Nature ecology & evolution* p. 1.

Keller, A., Kreis, S., Leidinger, P., Maixner, F., Ludwig, N., Backes, C., Galata, V., Guerriero, G., Fehlmann, T., Franke, A. et al. (2017). mirnas in ancient tissue specimens of the tyrolean iceman, *Molecular biology and evolution* **34**(4): 793–801.

Larson, G., Liu, R., Zhao, X., Yuan, J., Fuller, D., Barton, L., Dobney, K., Fan, Q., Gu, Z., Liu, X.-H. et al. (2010). Patterns of east asian pig domestication, migration, and turnover revealed by modern and ancient dna, *Proceedings of the National Academy of Sciences* **107**(17): 7686–7691.

Llamas, B., Holland, M. L., Chen, K., Cropley, J. E., Cooper, A. and Suter, C. M. (2012). High-resolution analysis of cytosine methylation in ancient dna, *PloS one* **7**(1): e30226.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., Tomsho, L. P., Packard, M. D., Zhao, F., Sher, A. et al. (2008). Sequencing the nuclear genome of the extinct woolly mammoth, *Nature* **456**(7220): 387.

Ng, T. F. F., Chen, L.-F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P. D., Varsani, A., Kondov, N. O., Wong, W., Deng, X. et al. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch, *Proceedings of the National Academy of Sciences* **111**(47): 16842–16847.

Richards, S. M., Hovhannisyan, N., Gilliham, M., Ingram, J., Skadhauge, B., Heiniger, H., Llamas, B., Mitchell, K. J., Meachen, J., Fincher, G. B. et al. (2019). Low-cost cross-taxon enrichment of mitochondrial dna using in-house synthesised rna probes, *PloS one* **14**(2): e0209499.

Rohland, N. and Hofreiter, M. (2007). Ancient dna extraction from bones and teeth, *Nature protocols* **2**(7): 1756.

Rollo, F., Venanzi, F. M. and Amici, A. (1994). Dna and rna from ancient plant seeds, *Ancient DNA*, Springer, pp. 218–236.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., De la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R. et al. (2017). Neandertal and denisovan dna from pleistocene sediments, *Science* **356**(6338): 605–608.

Smirnova, L., Gräfe, A., Seiler, A., Schumacher, S., Nitsch, R. and Wulczyn, F. G. (2005). Regulation of mirna expression during neural cell specification, *European Journal of Neuroscience* **21**(6): 1469–1477.

Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2014). A complete ancient rna genome: identification, reconstruction and evolutionary history of archaeological barley stripe mosaic virus, *Scientific reports* **4**: 4003.

Smith, O., Dunshea, G., Sinding, M.-H. S., Fedorov, S., Germonpre, M., Gilbert, M. et al. (2019). Ancient rna from late pleistocene permafrost and historical canids shows tissue-specific transcriptome survival, *BioRxiv* p. 546820.

Smith, O., Palmer, S. A., Clapham, A. J., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2017). Small rna activity in archeological barley shows novel germination inhibition in response to environment, *Molecular biology and evolution* **34**(10): 2555–2562.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y., Kosintsev, P., Lee, M. S., Baryshnikov, G. et al. (2016). Early cave art and ancient dna record the origin of european bison, *Nature communications* **7**: 13158.

Welker, F., Collins, M. J., Thomas, J. A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S. T., Reguero, M., Gelfo, J. N., Kramarz, A. et al. (2015). Ancient proteins resolve the evolutionary history of darwin's south american ungulates, *Nature* **522**(7554): 81.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V. et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus, *Nature* **544**(7650): 357.
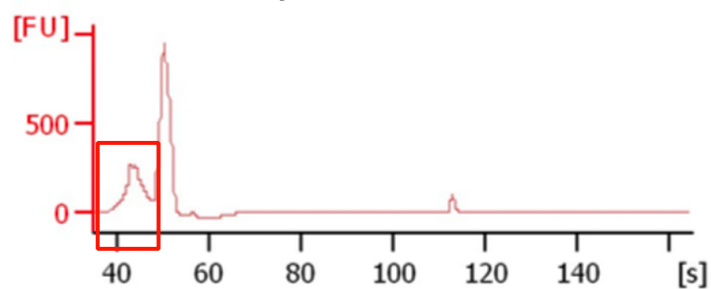
6.5 Supplementary materials

6.5.1 Figure S1

A. DNA ladder was detected by BioAnalyzer small RNA chip
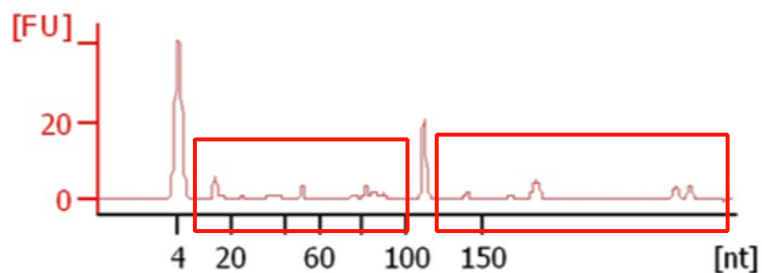
B. Yeast tRNA was detected by BioAnalyzer High sensitivity DNA chip

Fluorescent Staining cannot distinguish DNA and RNA.

Chapter 7

Discussion

7.1 Overview

Integrating natural selection, Mendelian inheritance, and genetic variation, the Modern Synthesis laid the cornerstone for contemporary evolutionary biology (Huxley; 1942). However, well-studied phenotypes, such as height and weight, cannot be fully explained by genetics (Gibson; 2018), and accumulating evidence suggests that non-genetic mechanisms play key roles in animal evolution (Alberdi et al.; 2016; Skinner; 2015). In 2007, the 'extended evolutionary synthesis' was proposed, calling for scientists to rethink and redefine the Modern Synthesis with non-genetic mechanisms (Pigliucci; 2007). Non-genomic mechanisms, such as epigenetics and microbiomes, tend to be highly dynamic and susceptible to internal (*i.e.*, genetics) and external factors (*i.e.*, environmental cues), but the evidence from modern animal models remains scarce and controversial (Laland et al.; 2014; Horsthemke; 2018; Van Opstal and Bordenstein; 2015). In this thesis, I employed advanced ancient DNA (aDNA) techniques to explore the role of non-genetic mechanisms over an evolutionary timescale. I identified two cases where adaption to the environment appeared to be assisted by the gut microbiome (Chapter 3) or epigenetics (Chapter 6). In addition to these case studies, I also participated in the development and optimisation of protocols to retrieve ancient epigenetic and microbial information from extremely degraded and contaminated DNA sources; I also explored the taphonomy of these ancient DNA molecules under varied conditions. In this Discussion, I will discuss the implications of these case studies and suggest future applications of paleomicrobiome and paleoepigenetics approaches, highlighting the potential of aDNA techniques to resolve long-standing evolutionary questions.

7.2 Case studies suggest the epigenome and microbiome are potential adaptive mechanisms

In Chapter 3 and Chapter 5, I used two bovine species as models to investigate animal adaptation to changing environments. Chapter 3 examined how an extinct cave goat (*Myotragus balearicus*) adapted to consuming a toxic plant that was widely distributed throughout their habitat. I found that the gut microbiome of this extinct goat was highly specialised to degrade toxins and promote the host gut function, which may play a key role in the ability to tolerate immunogenic stimuli from the toxin. Chapter 5 investigated the methylome history of the extinct steppe bison (*Bison priscus*) and I found that several genes involved in stress response and development were actively methylated over the late Pleistocene and Holocene, which may be a response to climate fluctuations and explain the extreme morphological diversity of steppe bison. When environments fluctuate in a drastic and rapid manner (*e.g.*, the late Pleistocene), it would be of great advantage to have fast response mechanisms to increase fitness and facilitate adaptation, other than genetic adaptations that take many generations (Cooper et al.; 2015; Gotthard and Nylin; 1995). In both cases studied in this thesis, large mammals experienced and adapted to adverse environments. The microbiome and methylome evidence suggest that these non-genetic mechanisms very likely play a complementary role that contributed to rapid adaptation to the environment.

In a natural setting, the environment is complex and consists of numerous biotic

and abiotic factors, including nutrient availability, temperature, and other stressors. Therefore, it is very difficult to directly link any microbiome and epigenome alterations that occurred in the evolutionary histories of animals to specific environmental variables. However, I showed that paleomicrobiome and paleoepigenome studies can identify microbial species or genomic loci that are highly likely to be critical for animal adaption, which provides a short list of candidates that should be explored further and validated using modern animal models.

Furthermore, a large number of sub-fossils are available for diverse large mammalian species, and each of these species has its unique evolutionary history: some have gone extinct, some survived extreme environments, some experienced severe population bottlenecks, while others were domesticated (Cooper et al.; 2015; Lorenzen et al.; 2011). The fossil and molecular records of large ancient mammals enable not only a thorough investigation of mammalian evolution from many different angles, but provide also an overview of the dynamic interactions between large mammals, their genomes, epigenomes, and microbiomes, as well as the environment in an ecosystem over millennia. Such evolutionary research can hardly be done in the time scales dictated by the generation time of living wild mammals, but I showed in this thesis that studies on ancient mammals have great potential to reveal novel insights in evolutionary biology.

7.3 New protocols for the recovery of paleoepigenetics and paleomicrobiome data

In order to obtain high-quality data from ancient animal remains, I report in this thesis the development and optimisation of three protocols for obtaining data from highly degraded nucleic acids. Specifically, the protocol initially designed for the recovery of the oral microbiome information from ancient human dental calculus (Adler et al.; 2013a) was adapted to the analysis of ruminant cementum (Chapter 4); a hairpin bisulfite sequencing protocol (Laird et al.; 2004; Zhao et al.; 2014) was optimised for aDNA, allowing retrieval of paleomethylome data at single-base resolution (Chapter 5); and an RNA extraction protocol was modified for extracting ancient RNA (Chapter 6).

7.3.1 Extraction of microbial DNA from ruminant cementum

Oral microbiome information can be used to infer diet, behaviour, and microbiome-host interactions of ancient mammals. Despite the fact that existing oral microbial DNA extraction and analysis protocols were designed for human dental calculus (Weyrich et al.; 2017, 2015; Adler et al.; 2013b), I adjusted and tested these protocols on ancient bison teeth and obtained oral microbial signals in Chapter 4. I also identified several experimental limitations, such as the decontamination procedures. Although further optimisation is required before applying the protocol to a wide range of samples, I show that it is possible to recover information from some oral microbes from non-human mammal teeth.

7.3.2 Hairpin protocol

The differential damage patterns of cytosine and methylated cytosine in aDNA have been leveraged to obtain methylation information from ancient DNA (Pedersen et al.; 2014). Although methylome data have been successfully obtained from ancient humans using damage patterns (Hanghøj et al.; 2016; Gokhman et al.; 2014; Pedersen et al.; 2014), the resolution of this method is limited. In Chapter 5, I report the optimisation of a hairpin bisulfite sequencing protocol (Laird et al.; 2004; Zhao et al.; 2014) to obtain DNA sequences and methylation information simultaneously. This method allowed not only the retrieval of methylation information from ultra-short aDNA fragments, but also the examination of the asymmetric methylation loci. The hairpin protocol allowed the reconstruction of single-base resolution methylomes from highly degraded DNA samples for the first time.

7.3.3 RNA extraction protocol

Due to the prevalence and resilience of RNases, RNA has only been obtained from historical plants, faecal remains, and several well-preserved soft tissues from ancient mammals (Pedersen et al.; 2014; Smith et al.; 2014b; Ng et al.; 2014; Smith et al.; 2017; Fordyce et al.; 2013; Smith et al.; 2014a). However, I was able to extract RNA from a 30ky-old bison bone sample. As ancient RNA (aRNA) is likely to be highly degraded, I adapted a protocol for micro RNA (miRNA) extraction, and added further measures controlling for potential contamination with co-extracted DNA. This protocol showed that aRNA could be recovered from sub-fossil bones, which represent most of the animal sub-fossil record, and thus pave the way for future study on ancient transcriptomes on a large scale.

7.3.4 Methodological Considerations for damaged and fragmented nucleic acids

The aforementioned protocols were developed and optimised for ancient nucleic acids with two major considerations in mind. First, ancient nucleic acid molecules are damaged, thus they need to be repaired before the sequencing library construction step. Typical repair steps can include the removal of uracils (resulting from the deamination of cytosines) and the phosphorylation of the $5'$ end of nucleic acid molecules (DNA and RNA), as the presence of a phosphate group at the $5'$ end of a DNA molecule is necessary for the ligation of sequencing adapters. Sometime protocols do not remove uracils (*e.g.*, uracils were not removed in the microbial DNA extraction in Chapter 3 and Chapter 4), as it is an important characteristic of aDNA and can be used for the authentication of the results. If this is the case, then the enzyme used for amplification must be able to read uracils (such as KAPA HiFi DNA Polymerase), or else the amplification of these damaged molecules will fail. Further, While the absent of $5'$ phosphate group is well described for ancient DNA, this is not described for aRNA. In Chapter 6, I showed that a phosphorylation step greatly improves the yield of the RNA sequencing libraries.

In addition to being damaged, ancient nucleic acid molecules are also highly fragmented, so aDNA protocols usually need to be optimised for short molecules. For example, the RNA extraction protocol (Chapter 6) was adapted from a (short) miRNA extraction protocol rather than total RNA extraction protocols. Note that

extremely short fragments can also pose challenges in the downstream bioinformatic analysis, so it is important to account for the potential limitations caused by fragment lengths. In Chapter 5, the primary reason for the introduction of the hairpin adapter is to improve the mappability of short low-complexity DNA molecules after bisulfite conversion.

7.4 Potential of paleomicrobiome and paleomethylome data mining

With the recent and ongoing advances in HTS techniques, as well as the optimisation of protocols for the analysis of ancient nucleic acids, aDNA data are rapidly accumulating. However, I showed in this thesis that aDNA data could be further explored in many ways to retrieve additional layers of information. In Chapter 3, functional profiling revealed that the gut microbiome of the extinct cave goat (*Myotragus balearicus*) was enriched with detoxification genes, which were likely an essential factor for the goat to consume an otherwise toxic plant. While a limited number of functions can be predicted from taxonomic information, functional information is usually largely obscured in taxonomic profiles. This is mainly because bacteria have short generation times, and they evolve rapidly through mutation, recombination, and horizontal gene transfer (Thomas and Nielsen; 2005; Ochman et al.; 2000). Sometimes beneficial genes can even be transferred from distantly related species (Jain et al.; 1999). Consequently, the functions of ancient bacteria would be hard to predict using only taxonomic information, as multiple recombination and horizontal gene transfer events would likely occur during their evolutionary histories. In Chapter 3 and Chapter 4, I found that even with relatively high coverage (*e.g.*, an ancient Romboutsia strain was reconstructed with a mean depth of $12\times$ in Chapter 3), there were regions that could not be reconstructed using modern bacterial genomes as references. Such regions are shared by multiple ancient strains, suggesting these resulted from genomic differences between the ancient strains and reference genomes, rather than stochastic DNA degradation.

Although it is difficult to reconstruct bacterial genomic regions that have undergone recombination and horizontal gene transfer events, functional profiling can still infer functions from the protein sequences determined from DNA fragments without knowing which strain it came from. While much of the current paleomicrobiome data were generated using a shotgun sequencing approach and both the taxonomic and functional information are available, the functional information in the datasets remains mostly under-explored due to a lack of bioinformatics tools. An additional obstacle for exploring functional information is the lack of a way to authenticate the ancient origin of the functional data, but alternative methods (*e.g.*, 'change point', further discussed in the next section) could be applied and the origins of functional data could be assessed.

Chapter 3 and Chapter 4 also demonstrated that ancient bacterial genomes can be reconstructed using several different approaches. For instance, iterative mapping (Hahn et al.; 2013), a method that combines *de novo* assembly and reference-guided mapping, can reduce gaps in the assembled genomes and thus increase the completeness of reconstructed genomes (Chapter 3). In Chapter 4, competitive mapping was applied to identify reads that uniquely mapped to a given reference genome,

thereby minimising potential cross-mapping from closely related species. These reconstructed genomes offer information for the direct inference of the evolution of commensal bacteria. These methods can be also applied to modern metagenomic data. For example, iterative mapping can increase the completeness of assembled genomes, especially for novel strains, while competitive mapping can help to distinguish dominant strains from closely related ones.

The observation of a possible phylosymbiotic relationship between bacteria and mammalian hosts suggests commensal bacterial very likely co-evolved with their hosts. Although currently the resolution of this relationship is limited due to a lack of available genomes of bacteria isolated from modern mammals, ancient bacteria genomes represent a promising source for investigating the co-evolutionary relationship between commensal bacteria and animal hosts. The additional information that can be obtained from commensal bacteria that shared an evolutionary history with their hosts might provide clues to infer the evolution of mammals in cases where genetic data only are not informative.

One of the major concerns in paleomicrobiome studies is the potential bias caused by contamination from modern microorganisms (Weyrich et al.; 2015). Although it is impossible to completely eliminate contaminants from paleomicrobiome data, I showed that such bias could be carefully monitored and minimised. In Chapter 3 and Chapter 4, I showed that a comparison between paleomicrobiome data and environmental data could provide a method to rapidly evaluate whether the obtained ancient data were dominated by environmental signals or not. Utilising negative controls (such as extraction blank controls) can help to monitor laboratory contaminants and cross-contamination between samples. Damage patterns can also help to determine if the obtained data originated from an ancient source. Another potential bias, the alteration of the bacterial community after death of the host animal—*i.e.* taphonomy, is even more challenging to monitor, as the understanding of the taphonomic process of microbiomes is limited. However, comparing paleomicrobiome and the corresponding modern microbiome, as well as bacteria that are susceptible to *in vitro* multiplications (Amir et al.; 2017), can offer some clues to assess this bias.

The paleomethylome study (Chapter 5) in this thesis is still at an exploratory stage, which is mainly due to the novel development of this protocol. It is, therefore, of paramount importance to validate the accuracy of the obtained data before further interpretation. As there are limited replicates available for many aDNA studies, one possible way to verify results is to compare data generated using different methods. In Chapter 5, the methylome data generated using the hairpin method are based on experimental work in the laboratory, whereas that of the damage-based method are obtained bioinformatically. I compared results generated using the hairpin method and the damage-based method and found that the two were highly consistent. This is strong evidence to suggest that the new hairpin approach is robust. In addition, I also showed in Chapter 5 that, after adjustment for DNA damage, the paleomethylome data were comparable to modern methylome data. This increased my confidence in the robustness of the paleomethylome data, but also demonstrated the feasibility of conducting large-scale paleomethylome studies combining different methods and/or modern and ancient methylome data. A better understanding of data biases is however crucial before merging datasets with confidence.

Overall, I showed that paleomicrobiome and paleomethylome research possess great potential and should be further explored, as the existing data can be explored

from additional angles and the emergence of new laboratory protocols and bioinformatics tools offers new ways to mine data from existing ancient materials.

7.5 Insights into the impacts of ancient DNA damage on paleomicrobiome and paleomethylome data

DNA damage is an integral part of aDNA. Better understanding DNA damage is therefore a prerequisite for further protocol development, data analysis, and interpretation for aDNA results. Protocols designed for damaged DNA have been developed to enrich for aDNA, such as single-stranded DNA extraction and selective enrichment for damaged DNA (U selection) (Gansauge and Meyer; 2014, 2013; Gansauge et al.; 2017). Additionally, damage patterns of aDNA have been used to distinguish authentic aDNA from modern contamination. Substitution plots showing an accumulation of C-to-T (and eventually reverse complement G-to-A substitutions) at the end of sequencing reads are a gold standard for the authentication of aDNA (Jónsson et al.; 2013). Consequently, sometimes only transversion mutation (G-C and T-A mutations) are used for downstream phylogenetic and population genetics analyses (Prüfer et al.; 2010). However, those protocols and standards were mainly developed for paleogenomic data. In this thesis, I offered new insights into the potential application of DNA damage pattern analysis in paleomicrobiome and paleomethylome data.

Paleomicrobiome data are even more susceptible to contamination than genomic data. This is mainly because microbiome data cannot be mapped to a single reference. This can lead to two major problems: 1) DNA damage cannot be characterised at the microbiome level using programs like MapDamage (Jónsson et al.; 2013), which relies on mapping data against one genome; and 2) for genomic data, a large proportion of contaminants can be removed by filtering out reads that failed to map to the reference genome; however, for paleomicrobiome data, as there is no specific 'target genome', it is impossible to filter unmapped data to remove unwanted information. These limitations make the assessment of the authenticity of paleomicrobiome data very difficult. To deal with this issue, I co-developed an algorithm ('change point') for the rapid assessment of damage patterns in paleomicrobiome data (Chapter 3). This algorithm examined the proportions of adenine (A) and thymine (T) as a proxy of DNA damage, as DNA damage can lead to an increase of A and T at the end of the reads in the sequencing data. The changes of A/T proportions along the reads were calculated, and the estimated statistical significance served as a proxy for the identification of the ancient origin of the sequences. This algorithm offers a fast assessment of significant damage patterns in a microbiome data set, which can be used as a preliminary authentication of a sample. In addition, I also found that short-term post-mortem exposure to the environment (one year in the case studied in the Chapter 4) can lead to DNA fragmentation similar to that observed in DNA as old as tens of thousand years, while deamination of cytosines is minimal. This observation offers a means to distinguish authentic aDNA from recently introduced contamination.

DNA damage affects paleomethylome data in a different manner. Methylome data can be mapped to a specific reference genome, and as a result, issues caused by

microbial contamination are less concerning than for paleomicrobiome data. However, deamination of methylated cytosines represents a major challenge in paleomethylome data. In paleogenomics data or paleomicrobiome data, deamination-induced C-to-T substitutions can reduce the mapping quality value, but the large proportion of reads is still correctly mapped to the reference sequence (paleogenomics data) or assigned to a microbial genus or family (paleomicrobiome data). However, deamination of methylated cytosines into thymines can lead to a decrease in the observed global methylation level in bisulfite sequencing data, because the net result of bisulfite treatment in sequencing data is the substitution of unmethylated cytosines into thymines. This experimental bias can greatly influence the analysis and interpretation of paleomethylome data. In Chapter 5, I investigated the deamination rates of methylated cytosines of samples with different radiocarbon dates. I found that, once several bases (10 bp) were soft-trimmed at the end of the reads before methylation calling, the deamination-induced bias seems to be alleviated in ancient samples. Significant variations in methylation levels were not detected from samples with a difference in radiocarbon dates of up to about 33ky, and after adjusting global methylation levels, ancient methylome data appeared to be comparable to modern methylome data.

7.6   Limitations and future directions

Although I carefully monitored and avoided potential pitfalls in my studies, this thesis can be further improved from several aspects.

7.6.1   A better understanding of microbiome and epigenetics in non-human mammals

Paleomicrobiome and paleoepigenetics research can hardly be done without a greater understanding of their modern counterparts. Data obtained from modern animals is critical to establish a baseline to interpret ancient data. This includes assessing the preservation of biological information and the identification of unique features within ancient data. The study of microbiome and epigenome information in modern animals is far from exhaustive, which creates limitations to aDNA research (*i.e.*, paleogenome, paleomicrobiome, and paleoepigenome studies). For example, oral microbiome studies on modern mammals are very limited. To my knowledge, the oral microbiome data were only available for a few mammalian species, such as koala (Alfano et al.; 2015), dog (Dewhirst et al.; 2012), mouse (Rautava et al.; 2015), and a few marine mammals (Bik et al.; 2016). Moreover, while the methylomes of human and mouse are extensively studied, the understanding of that of other mammals and related phenotypic consequences is minimal.

To tackle this limitation, I used data from closely related species (Chapter 3) and directly generated data for modern animals in my studies (Chapter 4 and 5). The quantity of modern data generated in this thesis was still restricted by the manpower and financial resources that could be allocated to my studies. However, these data can be combined with related future studies into larger datasets, and thus can be of use for future modern animal research as well as paleomicrobiome and paleoepigenetics research.

Additionally, my findings are only molecular evidence, but validation of these findings using modern animals can provide further insights regarding the detailed physiological response to microbiome and epigenetics modifications and potential applications to other related research. Here, I advocate that more effort should be put into the investigation of microbiome and epigenetics in non-human mammals to advance overall microbiome and epigenetics research, validate the findings from aDNA research, and integrate insights from both modern and ancient animals to gain a better understanding of animal evolution.

7.6.2 Further investigation into the taphonomy of nucleic acids

Despite that the presence of DNA damage is well documented (Briggs et al.; 2007; Lamers et al.; 2009), aDNA damage can be further explored, especially in regard to the DNA damage patterns in paleoepigenome and paleomicrobiome data. In terms of paleomethylome data, it has been proposed that the fraction of cytosine that undergo deamination increases over time (Sawyer et al.; 2012), but I only identified a marginal correlation between the levels of deamination and time (Chapter 5). The taphonomy of microbial DNA is more complex, as many characteristics of bacteria, including structure of the cell envelope, ability to sporulate, and physiologic types, may affect the post-mortem alteration of microbial community and the preservation of microbial DNA (Mann et al.; 2018). A better understanding of taphonomy is critical to distinguish genuine features of paleomicrobiomes and paleoepigenomes, and thus reconstruct the microbiome and methylome of ancient mammals with high confidence and accuracy. DNA damage is usually examined in aDNA studies on a case-by-case basis. However, for a comprehensive understanding of DNA taphonomy, research using modern material is necessary, as it would be much easier to control, manipulate, and monitor various variables.

7.6.3 Integration of approaches and analytical tools from other fields to address intrinsic challenges in aDNA research

Ancient DNA studies suffer from several intrinsic challenges, including temporal and geographical distribution of the samples, sample deposition, and the quantity and quality of data. For example, the identification of differentially methylated regions in Chapter 5 was underpowered by a lack of ancient replicates, and even with reasonable sequencing effort, only a limited proportion of data could be retained after the removal of potential contaminants and low-quality data. Although there is no easy solution to completely counter such issues, increasing sequencing effort and statistical adjustment of data can reduce potential biases caused by missing data.

It is also important to be aware that these challenges may not be exclusive to the aDNA field. Sometimes it is possible to seek solutions that are transferable to aDNA studies directly or with minor modifications from other fields. For instance, susceptibility to contamination and DNA degradation are also common issues when handling clinical samples (Cukier et al.; 2009; Matranga et al.; 2014), and single-cell sequencing techniques have been designed to obtain genomic information from extremely low-biomass sources (Gawad et al.; 2016). Many important insights from

those fields, such as limiting biases caused by amplification (Proserpio and Lönnberg; 2016) and statistically dealing with missing data (Little et al.; 2012; Kang; 2013), are highly compatible with aDNA research. Additionally, the comparison between ancient and modern methylome data is a typical case of comparing a heterogeneous (the methylation patterns vary between samples) dataset to a relatively homogenous dataset (the methylation patterns are similar between samples). This is also a problem in cancer epigenetics research, which requires the comparison of cancer cell methylomes, a heterogeneous dataset, to healthy cell methylomes, a homogenous dataset (Hansen et al.; 2011). Conversely, aDNA techniques can also result in advances for other research fields. For example, the highly optimised and efficient aDNA extraction protocols can be applied to forensic samples, and the stringent controls widely employed in aDNA studies are suitable for the study of low-biomass samples, such as for profiling the lung or placental microbiomes (Eisenhofer et al.; 2018).

7.7 Concluding remarks

This thesis represents one of the most comprehensive studies of the evolution of non-human mammals from a paleomicrobiome and paleoepigenetic perspective. This thesis starts with an overview of existing evidence of the potential roles of the microbiome and epigenetics in animal evolution and a proposal for using advanced aDNA techniques to further investigate these roles (Chapter 1 and 2), followed by the development of novel experimental and bioinformatics approaches that are optimised for ancient nucleic acids, and two case studies on bovid animals (Chapter 3-6). I demonstrate that paleomicrobiome and paleoepigenetics approaches possess great power and potential in providing new insights into fundamental questions about animal evolution. Much like utilising a time machine, studying the information retrieved from ancient animals allows us to delve into the past, thereby capturing the emergence of new microbiome and epigenome patterns and the subtle alterations that have occurred in their evolutionary history. Ultimately, this allows us to reflect on how the animal microbiome, epigenome, and genome can evolve over millennia.

Bibliography

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A., Alt, K. W. et al. (2013a). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the neolithic and industrial revolutions, *Nature genetics* **45**(4): 450.

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A., Alt, K. W. et al. (2013b). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the neolithic and industrial revolutions, *Nature genetics* **45**(4): 450.

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M. L. and Gilbert, M. T. P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation?, *Trends in ecology & evolution* **31**(9): 689–699.

Alfano, N., Courtiol, A., Vielgrader, H., Timms, P., Roca, A. L. and Greenwood, A. D. (2015). Variation in koala microbiomes within and between individuals: effect of body region and captivity status, *Scientific reports* **5**: 10189.

Amir, A., McDonald, D., Navas-Molina, J. A., Debelius, J., Morton, J. T., Hyde, E., Robbins-Pianka, A. and Knight, R. (2017). Correcting for microbial blooms in fecal samples during room-temperature shipping, *MSystems* **2**(2): e00199–16.

Bik, E. M., Costello, E. K., Switzer, A. D., Callahan, B. J., Holmes, S. P., Wells, R. S., Carlin, K. P., Jensen, E. D., Venn-Watson, S. and Relman, D. A. (2016). Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea, *Nature communications* **7**: 10516.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. et al. (2007). Patterns of damage in genomic dna sequences from a neandertal, *Proceedings of the National Academy of Sciences* **104**(37): 14616–14621.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. and Bradshaw, C. J. (2015). Abrupt warming events drove late pleistocene holarctic megafaunal turnover, *Science* **349**(6248): 602–606.

Cukier, H. N., Pericak-Vance, M. A., Gilbert, J. R. and Hedges, D. J. (2009). Sample degradation leads to false-positive copy number variation calls in multiplex real-time polymerase chain reaction assays, *Analytical biochemistry* **386**(2): 288–290.

Dewhirst, F. E., Klein, E. A., Thompson, E. C., Blanton, J. M., Chen, T., Milella, L., Buckley, C. M., Davis, I. J., Bennett, M.-L. and Marshall-Jones, Z. V. (2012). The canine oral microbiome, *PloS one* **7**(4): e36067.

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R. and Weyrich, L. S. (2018). Contamination in low microbial biomass microbiome studies: Issues and recommendations, *Trends in microbiology* .

Fordyce, S. L., Avila-Arcos, M. C., Rasmussen, M., Cappellini, E., Romero-Navarro, J. A., Wales, N., Alquezar-Planas, D. E., Penfield, S., Brown, T. A., Vielle-Calzada, J.-P. et al. (2013). Deep sequencing of rna from ancient maize kernels, *PLoS One* **8**(1): e50961.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A. and Meyer, M. (2017). Single-stranded dna library preparation from highly degraded dna using t4 dna ligase, *Nucleic acids research* **45**(10): e79–e79.

Gansauge, M.-T. and Meyer, M. (2013). Single-stranded dna library preparation for the sequencing of ancient or damaged dna, *Nature protocols* **8**(4): 737.

Gansauge, M.-T. and Meyer, M. (2014). Selective enrichment of damaged dna molecules for ancient genome sequencing, *Genome research* **24**(9): 1543–1549.

Gawad, C., Koh, W. and Quake, S. R. (2016). Single-cell genome sequencing: current state of the science, *Nature Reviews Genetics* **17**(3): 175.

Gibson, G. (2018). Population genetics and gwas: A primer, *PLoS biology* **16**(3): e2005485.

Gokhman, D., Lavi, E., Prüfer, K., Fraga, M. F., Riancho, J. A., Kelso, J., Pääbo, S., Meshorer, E. and Carmel, L. (2014). Reconstructing the dna methylation maps of the neandertal and the denisovan, *Science* **344**(6183): 523–527.

Gotthard, K. and Nylin, S. (1995). Adaptive plasticity and plasticity as an adaptation: a selective review of plasticity in animal morphology and life history, *Oikos* pp. 3–17.

Hahn, C., Bachmann, L. and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach, *Nucleic acids research* **41**(13): e129–e129.

Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J. S., Willerslev, E. and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epipaleomix, *Molecular biology and evolution* **33**(12): 3284–3298.

Hansen, K. D., Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D. et al. (2011). Increased methylation variation in epigenetic domains across cancer types, *Nature genetics* **43**(8): 768.

Horsthemke, B. (2018). A critical view on transgenerational epigenetic inheritance in humans, *Nature communications* **9**(1): 2973.

Huxley, J. (1942). *Evolution the modern synthesis*, George Allen and Unwin.

Jain, R., Rivera, M. C. and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis, *Proceedings of the National Academy of Sciences* **96**(7): 3801–3806.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. and Orlando, L. (2013). mapdamage2. 0: fast approximate bayesian estimates of ancient dna damage parameters, *Bioinformatics* **29**(13): 1682–1684.

Kang, H. (2013). The prevention and handling of the missing data, *Korean journal of anesthesiology* **64**(5): 402.

Laird, C. D., Pleasant, N. D., Clark, A. D., Sneeden, J. L., Hassan, K. A., Manley, N. C., Vary, J. C., Morgan, T., Hansen, R. S. and Stöger, R. (2004). Hairpinbisulfite pcr: assessing epigenetic methylation patterns on complementary strands of individual dna molecules, *Proceedings of the National Academy of Sciences* **101**(1): 204–209.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E. et al. (2014). Does evolutionary theory need a rethink?, *Nature News* **514**(7521): 161.

Lamers, R., Hayter, S. and Matheson, C. D. (2009). Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial dna, *Journal of molecular evolution* **68**(1): 40–55.

Little, R. J., D'agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A. et al. (2012). The prevention and treatment of missing data in clinical trials, *New England Journal of Medicine* **367**(14): 1355–1360.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R. et al. (2011). Species-specific responses of late quaternary megafauna to climate and humans, *Nature* **479**(7373): 359.

Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Yates, J. A. F., Salazar-García, D. C. et al. (2018). Differential preservation of endogenous human and microbial dna in dental calculus and dentin, *Scientific reports* **8**.

Matranga, C. B., Andersen, K. G., Winnicki, S., Busby, M., Gladden, A. D., Tewhey, R., Stremlau, M., Berlin, A., Gire, S. K., England, E. et al. (2014). Enhanced methods for unbiased deep sequencing of lassa and ebola rna viruses from clinical and biological samples, *Genome biology* **15**(11): 519.

Ng, T. F. F., Chen, L.-F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P. D., Varsani, A., Kondov, N. O., Wong, W., Deng, X. et al. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch, *Proceedings of the National Academy of Sciences* **111**(47): 16842–16847.

Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation, *nature* **405**(6784): 299.

Pedersen, J. S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S. et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome, *Genome research* **24**(3): 454–466.

Pigliucci, M. (2007). Do we need an extended evolutionary synthesis?, *Evolution: International Journal of Organic Evolution* **61**(12): 2743–2749.

Proserpio, V. and Lönnberg, T. (2016). Single-cell technologies are revolutionizing the approach to rare cells, *Immunology and cell biology* **94**(3): 225–229.

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J. and Green, R. E. (2010). Computational challenges in the analysis of ancient dna, *Genome biology* **11**(5): R47.

Rautava, J., Pinnell, L. J., Vong, L., Akseer, N., Assa, A. and Sherman, P. M. (2015). Oral microbiome composition changes in mouse models of colitis, *Journal of gastroenterology and hepatology* **30**(3): 521–527.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and dna fragmentation in ancient dna, *PloS one* **7**(3): e34131.

Skinner, M. K. (2015). Environmental epigenetics and a unified theory of the molecular aspects of evolution: a neo-lamarckian concept that facilitates neo-darwinian evolution, *Genome biology and evolution* **7**(5): 1296–1302.

Smith, O., Clapham, A. J., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2014a). Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process, *Scientific reports* **4**: 5559.

Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2014b). A complete ancient rna genome: identification, reconstruction and evolutionary history of archaeological barley stripe mosaic virus, *Scientific reports* **4**: 4003.

Smith, O., Palmer, S. A., Clapham, A. J., Rose, P., Liu, Y., Wang, J. and Allaby, R. G. (2017). Small rna activity in archeological barley shows novel germination inhibition in response to environment, *Molecular biology and evolution* **34**(10): 2555–2562.

Thomas, C. M. and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria, *Nature reviews microbiology* **3**(9): 711.

Van Opstal, E. J. and Bordenstein, S. R. (2015). Rethinking heritability of the microbiome, *Science* **349**(6253): 1172–1173.

Weyrich, L. S., Dobney, K. and Cooper, A. (2015). Ancient dna analysis of dental calculus, *Journal of Human Evolution* **79**: 119–124.

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V. et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient dna in dental calculus, *Nature* **544**(7650): 357.

Zhao, L., Sun, M.-a., Li, Z., Bai, X., Yu, M., Wang, M., Liang, L., Shao, X., Arnovitz, S., Wang, Q. et al. (2014). The dynamics of dna methylation fidelity during mouse embryonic stem cell self-renewal and differentiation, *Genome research* **24**(8): 1296–1307.