Doing DNA differently: A bioarchaeological secondary exploration of *Brucella* in ancient metagenomes in the NCBI SRA

by

Aparajita Bhattacharya

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Public Issues Anthropology

Waterloo, Ontario, Canada, 2024

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The use of biomolecular methods in bioarchaeological studies of health and disease offer novel insights into the dynamics of disease presence and prevalence in the past, such as pathogen evolution, human–pathogen–environmental interactions, and contexts of disease transmission. However, a growing awareness among public stakeholders and anthropologists of the ethical imperative to preserve human remains wherever possible has given rise to non- and minimally-destructive methods for biomolecular research. Metagenomic approaches represent one such avenue for research when applied to secondary analyses of previously sequenced aDNA. To this end, this study screened for the presence of *Brucella* aDNA in archaeological human metagenomes published in the NCBI Sequence Read Archive (SRA) using three methods—one alignment-based and two alignment-free. The results suggest the possible but still unconfirmed presence of *Brucella* or related sequences in a set of sequencing runs from two Late/Final Jomon individuals from the Sanganji Shell Mound site, Japan (*ca.* 2994 ± 19 BP and *ca.* 3061 ± 19 BP). Given the inconclusive results, alternate explanations are also explored and future analyses in this regard are proposed. In situating the utility of bioinformatics approaches and tools within a research framework inspired by biocultural theory, this study presents a heuristic approach to integrating non-destructive secondary analyses of mined metagenomic data with anthropological insights.

Acknowledgements

I am eternally thankful to my supervisor, Dr. Alexis Dolphin, without whose kind support, insight, and guidance this thesis could not have been completed. Dr. Dolphin's faith in my ability to carry out this project has inspired and motivated me in ways that extend well beyond the scope of this work.

Thank you to Dr. Andrew Doxey and Dr. Robert Stark, who graciously agreed to serve on my thesis committee and offered invaluable feedback about all aspects of my work. To Harold Hodgins of the Doxey Lab, who made metagenomics make sense and took the time to help me train in the software and tools necessary for my research—thank you.

For their generous financial support of my work, I extend my gratitude to the families of Iris Yuzdepski and Sally Weaver, as well as to the James Downey Graduate Scholarship fund, the J. Armand Bombardier fund, the Government of Ontario, and the Social Sciences and Humanities Research Council of Canada. Thank you as well to the Department of Anthropology, the Faculty of Arts, and the University of Waterloo Interdisciplinary Trailblazer Fund.

Thank you to Dr. Adrienne Lo and Dr. Seçil Dağtaş, as well as to my graduate student cohort, who introduced me to new ways of exploring old ideas.

The Department of Anthropology here at Waterloo has been my academic 'home' since 2019 and has supported me through both my undergraduate and graduate careers. It has been an honour to get to know the professors, instructors, staff, and students in the Department—both as a student and as a teaching assistant. I will cherish my time here for all I have learned and experienced.

To my loving and supportive family and friends, who always had an ear ready to listen to me talk about my project, I owe much gratitude. Thank you to my parents and my sister. And to all the others—you know who you are—thank you.

Table of contents

Author's declaration ..... **ii**

Abstract ..... **iii**

Acknowledgements ..... **iv**

List of tables ..... **vii**

List of abbreviations ..... **viii**

Chapter 1: Introduction **1**

1.1 Research design and scope 1

1.2 Biocultural approaches in biomolecular research 2

1.3 Sampling human remains: Ethics in practice 5

Chapter 2: Background **9**

2.1 Ancient pathogens research: Metagenomic approaches 9

2.1.1 Applications in ancient pathogens research 10

2.2 *Brucella*: Brucellosis 12

2.2.1 Palaeopathology and palaeoepidemiology of brucellosis 14

2.2.2 *Brucella* lineages and phylogenetics: Implications for metagenomic approaches 17

2.3 Jomon populations of the Late/Final periods 20

2.3.1 The archaeology of the Jomon period 20

2.3.2 Bioarchaeological evidence of health and disease 22

Chapter 3: Methods **24**

3.1 Sample selection 24

3.1.1 Data set selection 25

3.1.2 Reference sequence selection 26

3.2 Alignment-based assessment 27

3.3 Alignment-free assessments 28

Chapter 4: Results **30**
4.1 STAT *k*-mer counts: An overview 30
4.1.1 STAT analysis, stage 1 results 30
4.1.2 STAT analysis, stage 2 results 32
4.1.3 Data set selection 33
4.2 Selected data set: Original study and context 34
4.3 BLASTn results 36
4.4 Mash results 38
Chapter 5: Discussion **42**
5.1 Considerations in sample selection and preservation 42
5.2 Reference sequences and *Brucella* lineages 43
5.3 *Brucella* in the Jomon period 45
5.4 Other limitations 46
Chapter 6: Conclusions **48**
6.1 Conclusions 48
6.2 Future directions 49
References **52**
Appendix **73**

List of tables

Table 1: Basic definitions for assessment scores for alignment-based analysis (BLASTn)..... 28
Table 2: Selected STAT predicted taxonomic results for *k*-mer total and self-counts for *Brucella* 31
Table 3: Selected STAT predicted taxonomic results for *k*-mer total and self-counts for *Brucella melitensis*..... 33
Table 4: Maximum BLASTn scores for *Brucella melitensis* for selected sequencing runs from PRJDB4223..... 37
Table 5: Maximum BLASTn scores for *Brucella ovis* for selected sequencing runs from PRJDB4223..... 37
Table 6: Maximum BLASTn scores for *Brucella abortus* for selected sequencing runs from PRJDB4223..... 37
Table 7: Mash distance estimation scores for *Brucella melitensis* for selected sequencing runs from PRJDB4223..... 39
Table 8: Mash distance estimation scores for *Brucella ovis* for selected sequencing runs from PRJDB4223..... 39
Table 9: Mash distance estimation scores for *Brucella ovis* for selected sequencing runs from PRJDB4223..... 40
Table 10: Maximum BLASTn scores for *Brucella melitensis* for all sequencing runs from PRJDB4223..... 73
Table 11: Maximum BLASTn scores for *Brucella ovis* for all sequencing runs from PRJDB4223 73
Table 12: Maximum BLASTn scores for *Brucella abortus* for all sequencing runs from PRJDB4223..... 74

List of abbreviations

aDNA	Ancient DNA
ANI	Average nucleotide identity
BCE	Before common era
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BP	Years before present
ca.	*circa*
CE	Common era
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDH	DNA–DNA hybridisation
DRAC	Digital Research Alliance of Canada
EBA	Early Bronze Age
HTS	High-throughput sequencing
LEH	Linear enamel hypoplasia
ma	Million years ago
mtDNA	Mitochondrial DNA
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
STAT	SRA Taxonomy Analysis Tool
WGS	Whole genome sequencing

Chapter 1: Introduction

1.1 Research design and scope

The study of pathogens using ancient DNA (aDNA) has relied on advancements in data acquisition and analysis, many of which have been developed as part of research in the broad fields of bioinformatics and molecular biology (Key et al., 2017). As a result of this disciplinary slant, many approaches in bioinformatics and molecular biology often prioritise biological mechanisms and processes in the development of research questions and tools, and in analysis, which are then adopted into aDNA research. As such, only a limited number of studies have considered how complex biocultural interactions between human hosts and microbiota affect which pathogens are present, their abundances, and the nature of their evolutionary and ecological histories. This study contends that, as a field of anthropology, bioarchaeology is well-placed to tackle the challenges of studying human–pathogen interactions in the past by integrating tools in molecular biology—specifically, metagenomics—by framing research questions in holistic terms and situating human behaviour and experiences within past social, cultural, and environmental contexts.

This study attempts to unite these apparently disparate avenues of research by incorporating anthropological insights into an analysis of ancient pathogen DNA in metagenomic sequencing data. The research presented herein involves a secondary analysis of metagenomic sequencing data from ancient samples published in the NCBI Sequence Read Archive (SRA), a public bioinformatics repository of sequencing data containing the 'short reads' characteristic of high-throughput sequencing (HTS). Bioinformatics databases such as the SRA have thus far remained an unexplored avenue in anthropological research and represent an area of future research growth. Primarily, the goal of this study is to demonstrate how an anthropological and

bioarchaeological interest in understanding disease presence and prevalence, transmission, and evolution, as well as human–pathogen interaction in the past may employ bioinformatics approaches and tools in a secondary metagenomic analysis of aDNA.

Secondarily, this study aims to practically demonstrate how ancient pathogen DNA may be detected within ancient metagenomic data from the SRA by using a heuristic approach and incorporating anthropological insights and perspectives in the research design and analysis of the results. *Brucella*, a clade of largely pathogenic bacteria selected as the organisms of interest for this study, was queried against the SRA database in order to select and retrieve data sets containing a possible pathogenic component. As a result, sequencing data from two Late/Final Jomon period individuals (radiocarbon dated to *ca.* 2994 ± 19 BP and 3061 ± 19 BP) from the Sanganji Shell Mound, Japan, form the basis of subsequent investigation. Further alignment-based and alignment-free computational methods were used to verify these initial findings. The results were then interpreted in light of current archaeological and bioarchaeological data about health and disease in the Jomon contexts of interest. Inspired by biocultural anthropology, this study emphasises the associations between health and disease emergence and transmission, human behavioural ecology, subsistence economics, population dynamics, ecological contexts, and human–animal–environment interactions.

1.2 Biocultural approaches in biomolecular research

Beginning in the latter half of the twentieth century, biocultural anthropology as a theoretical framework in its various iterations primarily emerged as a result of several decades of research across the major subfields of anthropology, including biological anthropology and bioarchaeology (Blakey, 1998, 2008; Goodman et al., 1988; Zuckerman & Armelagos, 2011),

sociocultural anthropology (Schell, 1997), medical anthropology (Dressler, 1995; Ulijaszek & Lofink, 2006), and psychological anthropology (Hruschka et al., 2005). As an established theoretical and methodological stance in research designs across many areas of anthropological inquiry, biocultural theory in its present form owes its development to dialogical interactions between, on the one hand, cultural anthropology research focused on the role of power and political–economic perspectives, and on the other, the work of conventional biological anthropologists (Hoke & Schell, 2020).

Interdisciplinary research designs must contend with the diverse research goals, priorities, and practices of the researchers involved; in particular, the incorporation of new biotechnologies in anthropological research about human health in the past both enables new directions and raises challenges in negotiating theoretical and practical differences in collaborative work. While few bioinformatics-driven studies in metagenomics have explicitly attempted to interrogate the theoretical foundations of research in that field, theoretical commitments and priorities are often implicitly built into research questions, designs, workflows, and analyses. For example, a seminal metagenomic study investigating the function and diversity of organisms in the human microbiome based on specimens collected from various body habitats from 242 adult research participants investigated the relationship between microbial clade and metabolism and "host phenotype" (*e.g.*, age, gender, ethnicity, body mass index) (Human Microbiome Project Consortium, 2012, p. 211). While "phenotypic meta data" were not found to significantly impact variation of the microbiome, the authors speculated that short- and long-term diet, physical activity, daily cycles, and host genetics may instead correlate with microbiome structure (Human Microbiome Project Consortium, 2012). In such cases, an anthropological approach may be well-positioned to broaden the scope of the research to investigate, for example, the synergistic

interactions between ostensibly 'biological' characteristics (or 'phenotypes') and social, cultural, and ecological environments.

Biocultural theory makes legible, and provides theoretical reference for, anthropological perspectives and approaches to scholars across disciplines, opening new directions in research and subverting intra-scientific insularity. Such an approach may be enabled by the analysis of sociocultural, environmental, epidemiological, and historical factors associated with disease prevalence, load, and transmission, and pathogen–host co-evolution and adaptation (Glencross, 2011; Marciniak & Poinar, 2019; Schell, 1997). Most importantly, biocultural anthropology considers the human body as an agent in a mutually interactive relationship with its environment—including the ecological, social, cultural, historical, and political context—which is both transformed by, and transforms, human biological and social experience (Hoke & Schell, 2020; Wiley, 2020). Accordingly, this study incorporates data and evidence from across fields of research in the social and biological sciences.

A biocultural approach to analysing health and bodily experience must emphasise the biologically, culturally, and environmentally co-produced nature of the human body, attending to the recursive and synthetic interaction of these forces (Blakey, 1998). Goodman and Leatherman (1998) have demonstrated how a bioanthropological political economy may focus on the social relations involved in the allocation of resources and control of labour, as well as in the creation of local histories that connect separate communities through historical and political–economic processes. Furthermore, these processes, realised through local histories, have biological consequences in their ability to shape inequalities and exploitation, subsistence strategies and practices, exposure to health and disease risks, food allocation, living conditions, and inter- and intra-community interactions (Goodman & Leatherman, 1998).

A final aspect of a biocultural orientation toward studying health and disease in the past must consider the concept of adaptation. Rather than relying on a reading of data as realities in and of themselves, studies investigating adaptive responses in organisms should consider them the tangible results of biological, social, and ecological processes operating in historical time (Goodman & Leatherman, 1998; Mazess, 1975). Thus, biological features that at first sight appear to be adaptive responses must be evaluated according to the ecological, social, and political–economic contexts in which they exist, the organism or individual affected, and the point in time in which the feature appears—before being cited as an adaptation (Goodman & Leatherman, 1998). This study is inspired by such elements of biocultural theory in developing a unique research framework and process.

1.3 Sampling human remains: Ethics in practice

Conventional analyses of human skeletal remains for evidence of pathological conditions, primarily relying on observations of lesions on bones or teeth, have greatly expanded anthropological understanding of health and disease, and the conditions thereof, in the past (Buikstra, 2010; Eshed et al., 2010; Larsen, 2010; Waldron, 2009). Though somewhat still in its nascency, analyses of pathogen aDNA detected in human remains offer the potential to confirm assessments of infectious diseases that have otherwise been made based on palaeopathological evidence (Rose, 2017). As with most methods for reconstructing past health and disease, palaeopathological analysis of skeletal lesions is both useful and subject to a unique set of limitations. Acute illness is rarely diagnosable skeletally, and a great many health conditions and infectious diseases do not provoke responses in the skeleton. As such, biomolecular methods may complement conventional bioarchaeological or palaeopathological evidence by confirming the suspected presence of pathogens where skeletal data may be limited or inconclusive.

Previous studies of ethics in bioarchaeological research have argued that the destructive sampling of human remains for DNA analysis, radiocarbon dating, and isotope analysis all raise issues of ethical concern (Jones & Harris, 1998; Squires & García-Mancuso, 2021). While the particularities of such issues are specific to the locational and temporal contexts of relevance to the research project, a number of scholars writing of various regional contexts have emphasised that the imperative to preserve human remains is a matter of interest to both public stakeholders and the scientific community (Jones & Harris, 1998; Squires et al., 2019; Walker, 2008). In the UK, projects involving the use and sampling of human remains often require the completion of ethics approval procedures at universities and other research institutions—although this is by no means a standard practice globally (Squires et al., 2019). Furthermore, rigorous palaeopathological analysis using conventional, non-destructive methods is highly recommended in cases where infection is suspected, before biomolecular techniques are attempted (Wilbur et al., 2009). When research does involve aDNA sequencing, however, such projects often neglect to address the ethical implications of destructive sampling and, in some cases, may also fail to prioritise them during research development and design (Squires et al., 2019).

Bioarchaeologists have also exhorted the need to avoid destructive sampling as part of exploratory projects; ideally, sampling ought to be undertaken with valid and reasonable justification, where osteological and palaeopathological evidence support a hypothesis of infection with the potential to be confirmed by biomolecular methods. For example, Wilbur et al. (2009) have identified a study by Hershkovitz et al. (2008) involving the destructive sampling of human remains for the PCR detection of tuberculosis (*Mycobacterium tuberculosis*) aDNA—a disease for which the authors contend there are no known pathognomonic lesions and whose diagnostic criteria are nonspecific. Although tuberculosis has been successfully diagnosed

skeletally elsewhere, in this case Wilbur et al. (2009) note that the failure to perform a rigorous palaeopathological analysis of the remains resulted in the destruction of approximately two grams of osseous material from a region adjacent to a pathological lesion, as well as of a control sample of the same size taken from elsewhere on the skeleton. While the stance taken by Wilbur et al. (2009) represents a considerably more conservative approach to destructive sampling—wherein such techniques are only to be used as a last resort and other, minimally-destructive methods are preferred—other scholars (Kaestle, 2010) have suggested that destruction of osseous material for biomolecular research is justifiable for methods-testing work, or in cases where the presence of pathogens of known detectability are being tested. Indeed, researchers in recent years have proposed various minimally destructive methods for biomolecular research, such as DNA extraction from dental cementum (Harney et al., 2021) and the use of chemical methods that do not alter the integrity of the bone or dental sample (Bolnick et al., 2011).

As such, the research presented herein proceeds from the assertion made by Squires et al. (2019) that, rather than enabling exploratory studies, destructive sampling should be restricted to research designs that seek to answer specific and focused questions. This study aims to demonstrate how exploratory research in bioarchaeology may make use of previously sequenced aDNA made available by other researchers investigating other research questions. Thus, when destructive sampling must be performed for aDNA extraction, such secondary approaches can ensure that samples and associated data are utilised to their full potential.

Finally, the recent florescence of sequencing and bioinformatics technologies offers opportunities for biological anthropologists and archaeologists to synthetically integrate questions of biological concern into the broader study of past human societies. Understanding pathogen abundance and tracking disease processes using their genetic signatures has the

potential to transform not only studies of contemporary diseases and epidemics, but also the human cultural and social responses thereby precipitated. Faced with the long-term impacts of pandemics and epidemics, now more than ever it is incumbent upon researchers to approach disease and health as synergistically co-produced by biological and cultural processes, in both ancient past and contemporary contexts.

Chapter 2: Background

2.1 Ancient pathogens research: Metagenomic approaches

Conventional methods in microbiology have typically relied on obtaining a pure bacterial culture from an environmental or biological sample, which omits information about the natural diversity of microbes—a phenomenon known as the 'plate count anomaly' (*i.e.*, the discrepancy in the number of microorganisms identified by microscopy versus colony counts) (Coughlan et al., 2015). Culture-independent techniques, such as metagenomics, enable microorganisms to be studied without requiring a pure culture to be produced, thus permitting the totality of microbial DNA in a given environmental sample to be characterised. Metagenomic tools permit the sequencing of the entire biological content of samples, enabling species identification and the elucidation of metabolic processes and functional roles; unlike single organism genome studies, clonal culturing of microorganisms is not required (Coughlan et al., 2015; Wooley et al., 2010).

On the other hand, metagenomic data, often comprising data from thousands of microorganisms in the environmental sample, risks being incomplete, noisy, and of high volume (Whatmore, 2014; Wooley et al., 2010). This has given rise to developments in computational analysis over the last two decades. In general, 'metagenomics' refers to the functional and sequence-based analysis of the collective genomes in an environmental sample through the construction of DNA libraries (sometimes referred to as eDNA libraries, zoolibraries, soil DNA libraries, and recombinant environmental libraries) (Riesenfeld et al., 2004). Since its first published uses in 2004 (Tyson et al., 2004; Venter et al., 2004), metagenomic approaches have made possible, *inter alia*, investigations of complex microbiomes (Human Microbiome Project Consortium, 2012), the discovery of endosymbiotic behaviour in environmental bacteria (Brown

et al., 2015), the tracking of human pathogens (Loman et al., 2013), and the elucidation of interactions and relationships between the viral and bacterial fractions of the human microbiome (Norman et al., 2015; Quince et al., 2017). The development of computational tools capable of handling large amounts of metagenomic data is associated with the emergence of high-throughput sequencing (HTS) technologies and pipelines. In contrast to Sanger sequencing methods, which are restricted by throughput limitations and cost, HTS refers to technologies that enable DNA and RNA to be sequenced rapidly, in large volumes (*e.g.*, multiple DNA molecules simultaneously), and efficiently in terms of both cost and computational resources; these techniques include template preparation and the construction of a DNA library, clonal amplification, and parallel sequencing (Ambardar et al., 2016; Paul et al., 2018). HTS tools (sometimes referred to as next-generation sequencing [NGS]) also enable shotgun metagenomics—the 'untargeted' sequencing of all genomes in the sample to profile taxonomic compositions, functions, and recover whole genomes (Quince et al., 2017). This technique can be used, for example, to characterise microorganisms present in a given sample in microbiome studies or as a screening tool in pathogenomics (Warinner et al., 2017).

2.1.1 Applications in ancient pathogens research

Research in metagenomics has largely been confined to the study of modern laboratory or environmental samples, as part of the broader fields of genetics and molecular biology (Liang et al., 2021; Martin & Uroz, 2016; Zink et al., 2002). Research initiatives in molecular biology often aim to characterise the microbial composition of the metagenomic mixture, including aspects such as taxonomy, diversity, function, and interaction (Martin & Uroz, 2016), while considerably less attention is devoted to the processes and contexts that result in their uptake into or presence in the human microbiome, or, as in the case of pathogenic microorganisms, their

transmission and evolution within and alongside human communities. Furthermore, sample selection protocols in molecular biology—even when applied to human specimens—are primarily concerned with maximising microbial yield rather than identifying specimens of contextual significance.

As a result, relatively few standard protocols have been developed for the study, analysis, and characterisation of ancient microorganisms in metagenomes, and thus, bioarchaeologists and ancient DNA researchers, to their credit, have often had to develop unique workflows (*i.e.*, series of computation or data processing tasks to transform raw data into interpretable results) and techniques to deal with the equally unique challenges that arise in this field. Many of these challenges result from the peculiarities of doing bioarchaeological research—the question of exogenous contamination from the depositional environment, for example, is a key concern—but bioarchaeologists have also aimed to incorporate anthropological insights into research design and development. In particular, the selection of appropriate and relevant metagenomic data sets and software pipelines must account for differential pathogen recoverability (that pathogens vary in the degree to which they are recoverable from ancient tissues) (Marciniak & Poinar, 2019), low concentrations of ancient microbial DNA in specimens, and the fact that exogenous contamination from the depositional (burial) environment may impact the yield of post-analysis data and interpretation thereof.

Research investigating the presence of ancient pathogen molecules in human remains have typically relied on PCR (polymerase chain reaction) detection which, since its development, has been the workhorse of molecular methods in biological anthropology (Adler et al., 2011; de la Fuente et al., 2013; O'Rourke, 2010; Pääbo et al., 2004). PCR methods effectively enabled the amplification of DNA sequences through the creation of millions to billions of copies of a

particular segment of DNA (Pääbo et al., 2004). When searching for ancient bacterial DNA for analysis, in particular, species-specific PCR primers may be used for extraction and amplification (Adler et al., 2011). On the other hand, the introduction of metagenomic methods in aDNA analysis, primarily via HTS, have broadened the interdisciplinary scope of ancient pathogens research in bioarchaeology by expanding the range of questions able to be investigated (Der Sarkissian et al., 2021). Metagenomics can permit, among other things, the modelling of metagenome dynamics and community structure (Der Sarkissian et al., 2014), and the exploration of microbial diversity and evolution using *de novo* assembly (Granehäll et al., 2021). Whole genome sequencing has also elucidated relationships between the modern and ancient bacterial genomes of *Mycobacterium tuberculosis* and *Mycobacterium leprae* (Donoghue, 2013). More recently, Hodgins et al. (2023) have investigated the presence of ancient *Clostridium* DNA in the NCBI SRA and identified new lineages and neurotoxigenic variants of *C. tetani*. While few secondary studies in the exploration of ancient metagenomes have been performed, the availability of public repositories for HTS data, such as the NCBI SRA, may make it possible for anthropologists to ask more complex questions about human–disease interactions in the past, as well as potentially identify pathogens in previously unexplored contexts.

2.2 Brucella: *Brucellosis*

The following section presents a review of human brucellosis, and its causative bacterial agents, which form the pathogen of interest selected for this study. While past biomolecular studies of microbial aDNA have explored the presence of mycobacterial pathogens such as tuberculosis and leprosy, the possibility of the inadvertent detection of environmental (rather than pathogenic) mycobacteria using computationally efficient but imprecise tools for initial sample selection in the SRA, such as STAT (see Chapter 3.1), was considered when selecting a

pathogen of interest (Raoult & Drancourt, 2008). As a result, *Brucella* was chosen as an appropriate target pathogen due to its potential to be diagnosed skeletally, high pathogenicity, and a low likelihood of being present in the environment which could otherwise result in the contamination of the skeletal samples from the burial context, and thus, bias the post-analysis data.

Brucellosis, a highly infectious zoonosis of generally minimal mortality (Shakir, 2021), is caused by a group of nine Gram-negative, non-spore-forming coccobacillus species of the genus *Brucella*—of which four (*B. melitensis*, *B. abortus*, *B. suis*, and *B. canis*) are known to infect humans. *Brucella* bacteria behave like facultative intracellular parasites, capable of avoiding phagocytic destruction by growing inside macrophages, dendritic cells, placental trophoblasts, and epithelial cells (D'Anastasio et al., 2011; Leon-Sicairos et al., 2015; Shakir, 2021; Sulayman et al. 2020). *Brucella* are most often transmitted via the consumption of unpasteurised milk or milk products from infected animals; however, those who work in close proximity to animals, such as veterinarians, slaughterhouse workers, meat packers, hunters, and laboratory workers can become infected if aerosols containing bacteria are inhaled or exposure occurs through cuts or abrasions on the body (D'Anastasio et al., 2011; Shakir, 2021). *Brucella* species are also known to infect other animals: *B. melitensis* in small ruminants, *B. abortus* in cattle; *B. suis* in pigs, and *B. canis* in dogs (Sulayman et al., 2020). In modern human clinical settings, brucellosis presents as symptomatically nonspecific and variable, causing, *inter alia*, fever, nausea, weakness, muscular pain, increased perspiration, and liver inflammation (Sulayman et al., 2020; Yagupsky et al., 2019).

Brucellosis remains the most prevalent zoonosis worldwide (Buzgan et al., 2010) and is a public health concern in Mediterranean regions—namely, in Turkey—where farmers, laboratory

personnel, and veterinary clinicians have been assessed to represent the majority of infection cases (Pourbagher et al., 2006). In other regions of the world, the consumption of unpasteurised milk and milk products is reported to be the most common source of infection (Delam et al., 2022; Pourbagher et al., 2006; Qureshi et al., 2024). Laine et al. (2023) have reported annual incidence rates ranging from between 1.6 to 2.1 million cases globally per annum, with most of these cases accounted for by Asia (approximately 1.2–1.6 million) and Africa (approximately 500 million). These data do not include cases of misdiagnosis or underdiagnosis, which is a significant concern with this disease; indeed, in regions where malaria is endemic, 21%–50% of human brucellosis cases are reported to have been misdiagnosed as malaria—possibly due to a lack of reliable laboratory support (Laine et al., 2023; Njeru et al., 2016). Thus, while such challenges with clinical diagnosis of the disease is likely to result in the underreporting of cases, epidemiological estimates indicate a sustained risk of infection with brucellosis for many populations across all regions of the world.

2.2.1 Palaeopathology and palaeoepidemiology of brucellosis

Skeletally, the most common manifestations of brucellosis in human remains include sacroiliitis and monoarticular arthritis, with approximately a third of untreated infections resulting in the involvement of the lumbar spine (Waldron, 2009). Destructive lesions are also known to occur on the superior and inferior surfaces of the vertebral bodies, and deeper in cases of severe infection (Waldron, 2009). In modern clinical settings, the acute and subacute stages of infection are associated with the narrowing of intervertebral disk spaces and spondylodiskitis in the lumbar spine, often with the greatest involvement in the lower thoracic to upper lumbar region (approximately, T11–L2) (Pourbagher et al., 2006). In both modern patients and ancient remains, the cervical spine appears to exhibit the least involvement with this disease. On the

other hand, the similarities in skeletal response between tuberculosis and brucellosis infections have long been observed; spondylodiskitis in modern cases of acute brucellosis notably resembles tuberculous vertebral destruction (Pourbagher et al., 2006). In human remains, brucellosis is differentiated from tuberculosis by the presence of new bone formation and sclerosis around lesions (Waldron, 2009). The similarities in skeletal response to infection has resulted in the suggestion that some diagnosed cases of tuberculosis in human remains were, in fact, brucellosis—particularly where new bone formation is the most pronounced osseous response (Waldron, 2009).

It is generally believed that brucellosis was relatively common in the ancient past (D'Anastasio et al., 2011; Fournié et al., 2017; Waldron, 2009), and is suggested to have emerged as a sustained risk by the Early Neolithic (*ca.* 11,000–10,000 B.P.) in the Near East due to the increased contact between humans and animals necessitated by early agropastoralism and animal domestication (Bendrey & Fournié, 2021). Activities involved in animal husbandry and farming, such assisting cows in labour or the consumption of infected milk, may have heightened the risk of infection for certain segments of Early Neolithic populations (Bendrey & Fournié, 2021; Price et al., 2018; Waldron, 2009). Zooarchaeological modelling of Neolithic sheep and goat populations—the natural reservoir of *B. melitensis*—indicates that the pathogen could have been sustained, even at low rates of transmission, within the domestic goat populations of the Fertile Crescent and West Asia (Fournié et al., 2017).

One early case of possible brucellosis involving the lumbar spine has been differentially diagnosed in the partial remains of an *Australopithecus africanus* individual (2.3–2.5 ma) from Sterkfontein, South Africa (D'Anastasio et al., 2009); such an early case of infection suggests the possibility of random and isolated disease transmission from infected animals, rather than an

endemic presence of the disease within these Pliocene populations. Nevertheless, this assessment has been disputed by Haeusler (2019), who suggests that this pathology may instead represent a disk herniation. Osteoarthritic changes to the bodies and facet joints of the cervical vertebrae of La-Chapelle-aux-Saints 1, a Middle-to-Late Pleistocene *Homo neanderthalensis* individual assessed to be male and of > 60 years of age, has been assessed as a case of brucellosis, and Rothschild and Haeusler (2021) suggest the infection was contracted by butchering or consuming prey meat.

Another early case of brucellosis has been identified in the thoracic and lumbar spine of an adult male individual from the Neolithic site of Ganj Dareh, in the Zagros mountains of Iran, found alongside zooarchaeological evidence of early goat husbandry (Fournié et al., 2017). Cases of brucellosis have been diagnosed based on vertebral body lesions in four individuals buried in the EBA shaft tombs at Bâb edh-Dhrâ, part of a collection in which tuberculosis has also been identified (Ortner & Frohlich, 2007). Brucellosis has also been assessed, based on sacroiliitis on an innominate, from the remains of an ancient Egyptian individual from *ca*. 750 BCE (25th Dynasty) (Hodgkins, 2003; Moreno, 2014). Aubin (2004) reports a brucellosis frequency of 2.11% based on vertebral lesions in the ancient Nubian population at Semna South, Sudan (Meroitic, 350 BCE–1200 CE) and attributes this trend to the archaeological presence of the primary disease vectors, goats and sheep, in the region. Based on a study of skeletal remains from ancient Herculaneum by Capasso (1999), by *ca*. 79 CE 17.4% of adults excavated from the city exhibited signs of brucellar spondylitis; the author thereby suggests that brucellosis would have been endemic in ancient Rome.

2.2.2 Brucella *lineages and phylogenetics: Implications for metagenomic approaches*

Phylogenetic analyses of *B. melitensis* performed based on single nucleotide polymorphisms (SNPs) have revealed several distinct lineages of the pathogen, including a basal 'Mediterranean' lineage, and Asian, European, and African strains (Tan et al., 2015). While it is still unclear how these various lineages interact to shape the global spread of brucellosis, it has been suggested that ancient transmission of *Brucella* was patterned along the lines of international (intra-Mediterranean and cross-continental) trade during the Bronze Age and, perhaps, during the medieval period (Tan et al., 2015). Analyses of the complete genomes of *B. melitensis*, *B. suis*, and *B. abortus* reveal high degrees of similarity and a small number of truly unique sequences among these species (Halling et al., 2005). Following from this, phylogenetic reconstructions of the evolutionary relationships between these genomes using SNP data suggest that *B. melitensis* and *B. abortus* are closely related, whereas *B. suis* is more closely related to *B. abortus* than to *B. melitensis* (Halling et al., 2005). Additionally, some studies have suggested that *B. suis* may be considered a paraphyly of the main *Brucella* lineage, only forming a monophyly when *B.* canis is included in the clade (Foster et al., 2009).

Nevertheless, whole-genome sequencing (WGS) found more than 20,000 orthologous SNPs shared in all *Brucella* lineages and one *Ochrobactrum* lineage (Foster et al., 2009). Importantly, molecular clock dating of *Brucella* lineages indicated a split from *B. ovis*, the common ancestor, within the past 296,000 to 86,000 years. As one of the most commonly used techniques for dating evolutionary events, the molecular clock method involves the calculation of the divergence of two species in geological time based on the assumption of a constant rate of change in the nucleotide and amino acid sequences (dos Reis et al., 2016; Douzery et al., 2006). Although Foster et al. (2009) attribute the origins of brucellosis in animals such as pigs, goats,

and cattle to contact with sheep infected with *B. ovis*, this early date of divergence from the basal lineage actually suggests that the emergence of most *Brucella* species predates the domestication of their livestock hosts. As a result, *Brucella* species are not believed to have co-evolved with their respective hosts and, according to this model, the disease was endemic within wildlife populations rather than emerging due to domestication (Foster et al., 2009).

Together with the genera *Pseudochrobactrum*, *Crabteella*, *Paenochrobactrum*, *Ochrobactrum*, and *Mycoplana*, *Brucella* belong to the alphaproteobacteria class, with members of the genus *Ochrobactrum* representing the closest phylogenetic relations, sharing approximately 97%–98% identity with the consensus sequence of the *Brucella* 16S rRNA gene (Ryan & Pembroke, 2020; Yagupsky et al., 2019). At least two species of *Ochrobactrum* (*O. anthropi* and *O. intermedium*) display higher identity with *Brucella* species than others of their own genus (Yagupsky et al., 2019). As a result, conventional blood culture isolates of *O. anthropi* are known to have been misidentified as *B. melitensis* at times (Elsaghir & James, 2003). *Ochrobactrum* are a group of generally low-virulence species increasingly found in clinical settings to be the cause of serious infections in humans; various species are found in environments throughout the world, although the most definitive cases of infection are reported from North America, Southeast Asia, and Europe (Ryan & Pembroke, 2020).

Little is known about the skeletal response to infection with *Ochrobactrum*, but osteomyelitis (*i.e.*, of the lateral cuneiform bone of the foot) has been reported (Gigi et al., 2017), although in at least one case, brucellar osteomyelitis was incorrectly diagnosed as an *Ochrobactrum* infection (Trêpa et al., 2018). The resulting similarities between members of *Ochrobactrum* and *Brucella* have implications for their palaeopathological and biomolecular identification in ancient samples; while *Brucella* are entirely zoonotic, *Ochrobactrum* can be

found in animals as well as in a variety of natural environments, including water, soil, and plants (Ryan & Pembroke, 2020). This situation complicates the detection of *Brucella* sequences, since the high identity shared with *Ochrobactrum*, an environmental pathogen, means that uptake into bone samples from the burial environment is possible, potentially obfuscating the results of metagenomic analysis.

Whereas brucellosis is diagnosed moderately frequently using palaeopathological methods, palaeogenomic evidence for human *Brucella* infection has been limited to mostly ancient Near Eastern Neolithic or Early Bronze Age (EBA) contexts, and few other contexts have received much attention in this regard. Nevertheless, at the North Caucasus EBA kurgan groups Chekon, Natukhaevskaya, and Katusvina-Krivitsa (at Krasnodar Krai, Russia), HTS microbial reads in an aDNA sample from one individual from the Klady kurgan (near the Novosvobodnaya settlement) revealed infection with *B. abortus* (Sokolov et al., 2016). A more recent example is provided by a fourteenth-century CE male skeleton from Geridu, Italy, where a novel *B. melitensis* lineage was retrieved from a sample taken from a calcified nodule and analysed using shotgun metagenomics techniques (Kay et al., 2014).

Despite such identifications, little genomic evidence for brucellosis has been identified in other ancient contexts, even where diagnoses have been made based on skeletal lesion patterning. Thus, given that *Brucella* genomes are of known detectability and since brucellae are not known to be environmental bacteria (unlike *Ochrobactrum*), the potential for false detection of *Brucella* sequences remains reasonably low under normal conditions. Thus, the pathogen represents an opportunity for research into its potential presence in putatively pre- or non-agricultural contexts, such as the Neolithic Jomon populations of the Japanese archipelago. The

following section reviews the Jomon culture and, based on the results of the data analysis carried out in the SRA (see Chapter 4), forms the geographic and archaeological basis of this study.

2.3 Jomon populations of the Late/Final periods

Given the case presented in this study, this section will review the archaeological and bioarchaeological scholarship of the Jomon period in Japan, a broad archaeological horizon characterised by diverse forager and early agriculturalist populations following the Pleistocene/Holocene transition. Chapter 4 presents the results of an analysis of metagenomic sequencing data from two Late/Final Jomon individuals; as such, evidence of Jomon subsistence economics, animal domestication, population dynamics, and health and disease will be explored here in order to situate the results of this study within their proper geographic context. In this section, particular attention will be focused on the Late and Final Jomon sub-periods (4,500–2,360 BP), as this represents the most likely phase of origin for the Jomon skeletal remains under analysis.

2.3.1 The archaeology of the Jomon period

The Jomon period in Japanese history spans from *ca*. 14,000 BP to 2,360 BP (calibrated radiocarbon dates), with four key sub-periods: (a) Incipient (*ca.* 13,750–11,200 BP) and Initial (*ca.* 11,200–7,250 BP), concurrent with the Late Glacial warm period; (b) Early (*ca.* 7,250–5540 BP) and Middle (*ca. 5*,540–4,500 BP); (c) Late (*ca.* 4,500–3,180 BP); (d) and Final (*ca.* 3,180–2,360 BP) (Kuzmin & Keally, 2001; Natsuki, 2022; Noshiro & Sasaki, 2014; Pearson, 2006). The earliest expressions of Incipient Jomon culture are believed to be found in the southern Kagoshima Prefecture of Kyushu Island, where thousands of pottery sherds, arrowheads, polished and flaked axes, and querns and grinding stones have been recovered at seasonally-

occupied sites such as Kakoinohara (radiocarbon dated to *ca.* 13,500 BP) (Kobayashi, 2003; Pearson, 2006). By the time of Initial Jomon sites in the region, such as Kakuriyama (*ca.* 9,800 BP), occupation is believed to have been year-round and populations in the south of Kyushu displayed signs of increasing sedentism (Pearson, 2006; Shibutani, 2009). In general, the chronology of the Jomon period relies on an extensive ceramic typology known from hundreds of sites in the archipelago—namely, the persistence of 'cord-marked' Jomon pottery—and some radiocarbon dates, although there exists much internal temporal and regional variation in ceramic technologies, social structure, and subsistence economics (Kobayashi, 2003).

Populations of the Jomon period have been conventionally characterised by archaeologists as constituting many large, sedentary forager communities of low social complexity (Bleed & Matsui, 2010). The apparently oldest Jomon sites, accidentally discovered during railway and expressway construction projects between the late 1970s up to the mid-1990s, displayed signs of sedentism and settlement—including small groups of permanent dwellings, heavy pottery, stone tools, and ritual objects (Kawashima, 2010; Pearson, 2006). These materials differed appreciably from the smaller and mobile artefactual remains at earlier Palaeolithic sites and from those found in Palaeolithic occupation phases at sites which also contained Jomon phases. In the Jomon phases at Sojiyama, heavy adzes, potentially for building permanent houses and dugout canoes, as well as the presence of ventilated hearths, are starkly contrasted against lighter Palaeolithic stone tool technologies (Pearson, 2006).

Zooarchaeological evidence in relation to subsistence economics is of note when considering zoonotic disease transmission. While only limited evidence of animal remains is available, it appears that the zooarchaeological evidence at Jomon sites—particularly shell mounds of the Kanto and Tohoku districts—comprise primarily sika deer (*Cervus nippon*), wild

boar (*Sus scrofa leucomystax*), and the Japanese macaque (*Macaca fuscata*), alongside numerous bones of several species of fish (Matsui & Kanehara, 2006). In order to better understand the origins of a zoonosis such as brucellosis, it is pertinent to turn to the bioarchaeological evidence of disease in the Jomon context.

2.3.2 Bioarchaeological evidence of health and disease among the Jomon

Although the Jomon period now represents a key area of scholarship in Japanese archaeology, bioarchaeological studies of Jomon populations have seen limited growth and, where existent, interest is focused on studies of prehistoric and historic population genetics. Studies of health and disease in Jomon contexts have thus far been limited compared to other regions, perhaps because, as Bleed and Matsui (2010) observe, conventional reconstructions of Jomon lifeways have often emphasised wellness, bountiful environments, and the 'natural affluence' of the Japanese archipelago, and modern archaeological studies are frequently oriented toward understanding ancient population genetics, demographics, and migration patterns (Hudson, 2020). Nevertheless, a shift in focus toward the natural environment as an independent factor influencing human behaviour and as a lens through which to interpret archaeological evidence, in line with an analogous theoretical turn in archaeological scholarship elsewhere, has been observed in recent years (Hudson, 2020).

Analyses of linear enamel hypoplasia (LEH) and dental caries on maxillary first incisors from four eastern and four western/inland Japanese sites spanning the Middle/Late and Late/Final phases have revealed higher frequencies of hypoplasia in the western/inland group, with a temporal trend toward increasing rates in the Late/Final samples (Temple, 2007). Rates of dental caries at the intra-regional scale have been observed to increase during the Final period

(Temple, 2007, 2014). LEH frequencies have been interpreted according to a resource-stress model, attributing systematic physiological stress to increased plant dependence arising from seasonal marine resource depletion in the western sites during the later Jomon phases; caries prevalence may have been associated with shifts in subsistence strategies in response to climatic cooling during the Final Jomon period (Temple, 2007).

More recently, a limited number of aDNA studies using metagenomic approaches have attempted to elucidate the presence of pathogens in human remains from Jomon sites. Nishimura et al. (2021) have reported on the reconstruction of a complete sequence of the *Siphovirus* viral genome from metagenomic sequences extracted from the dental pulp of five individuals (one Initial, one Middle, and three Late Jomon). By using CRISPR loci to detect homologous spacer sequences, Nishimura et al. (2021) determined that the host of the siphovirus would be a species of bacterium closely related to *Schaalia meyeri* (*Actinomyoces meyeri*), a rare pathogen which, when contracted, is responsible for pleural infection of the lungs (Shimoda et al., 2021). On the other hand, no cases of brucellosis in Jomon populations have been reported; as an infectious disease, brucellosis represents a possible avenue for further investigation in the Jomon context. The generally limited biomolecular understanding of disease patterns and transmission during the Jomon period presents opportunities for anthropological study in a previously understudied context.

Chapter 3: Methods

This study aimed to detect and comment on the presence of *Brucella* in ancient metagenomic sequences from the NCBI Sequence Read Archive (SRA). The approach taken in sampling and analysing the metagenomic sequencing data that forms the basis of this study was primarily heuristic; that is, methods and tools were selected according to the availability and quality of sequencing data at each stage with a loosely defined workflow. Techniques used include alignment-based methods to detect regions of similarity in query and subject sequences, and alignment-free methods to estimate metagenome and genome distance. This study presents a methodologically inexhaustive approach to metagenome analysis and microbial identification: more stringent tools for sample selection, authentication procedures, and data analysis and modelling may further elucidate the *Brucella* component in the selected metagenomic data set if infection was indeed present (see Chapter 6.2). Special considerations and challenges associated with the use of public sequencing repositories as well as the selection and analysis of sequences from ancient samples are discussed in Chapter 5.

3.1 Sample selection

Sample selection was carried out on the available short read data in the SRA, the largest repository of HTS data available for public access. The SRA was selected over other databases of its type due to its large size and the likelihood of retrieving usable metagenomic data. As of August 2023, the NCBI SRA contained 26,780,597.471 gigabytes of data in 415,004 BioProjects, comprising 27,277,884 sequencing runs. A BioProject record provides links to all forms of original biological data, including sequencing runs, for a single research initiative deposited by the original study authors. Of these, the ambiguous organism type-label ‘fossil

metagenome' may be applied by study authors to indicate BioProjects whose samples are of some antiquity—although parameters for inclusion are not made clear and the use of this and other meta data terms is at the discretion of submitters (see Chapter 5.1).

3.1.1 Data set selection

In the absence of standardised meta data search terms to isolate BioProjects by attributes such as sample age or geographic region, a Boolean search query for 'fossil metagenome' was run against the SRA, which retrieved 934 samples. Google Cloud computing API was then used to generate a STAT data set. STAT (SRA Taxonomy Analysis Tool) calculates the taxonomic distribution of reads in a next-generation sequencing (NGS) run by mapping reads to a taxonomic hierarchy using a precomputed *k*-mer dictionary containing diagnostic *k*-mers (*i.e.*, a sub-sequence of nucleotides of length *k*) for each organism (Ayling et al., 2020). Katz et al. (2021) have shown that STAT results are proportional to the size of the sequenced genomes, so that mixed samples containing multiple organisms with genomes of varying sizes are expected to identify more reads originating from the larger genomes. As a result, along with sample composition, STAT counts are likely to also reflect genome sizes and the total genomic complexity of the sample (Katz et al., 2021). The use of Google Cloud computing tools allowed quick queries to be run for the taxonomic IDs *Brucella* (tax ID = 234), and *Brucella melitensis* (tax ID = 29459) to isolate potential data sets with the most appropriate composition.

The STAT generated data set was then filtered by organism type and *k*-mer count (total and self-counts). A *k*-mer threshold for *Brucella* of 100 was established based on standards and best practices outlined by Margaryan et al. (2018) and Kay et al. (2014) to account for deamination resulting from DNA degradation and other damage signals, as well as the fact that

DNA from ancient samples is expected to contain a low endogenous fraction. In addition, a low threshold was deliberately selected upon an initial review of the STAT data set, which revealed that higher *k*-mer counts were correlated with modern, rather than ancient, samples (*i.e.*, runs from BioProjects mislabelled 'fossil metagenome'). In stage 1, 36 SRA BioProjects were identified, including both truly ancient and possibly mislabelled 'fossil metagenomes', containing sequencing runs meeting or exceeding the *Brucella k*-mer threshold; further analysis of the SRA database revealed that several data sets represented sequences extracted from non-human animal hosts or were of modern origin, which were then removed from the selection pool, leaving seven possible BioProjects for selection. In stage 2, the STAT data set was filtered based on *k*-mer counts for *B. melitensis*; based on an initial review of the data which suggested that counts for this taxonomic ID were significantly lower than those for *Brucella*, no definitive count threshold was used.

These samples were then filtered further based on the degree of appropriateness vis-à-vis the defined research goals, and samples representing non-human animals were removed from consideration. Original publications and studies associated with each BioProject and data set were identified before any selections were made in order to ensure the availability and viability of information regarding sample type and species, site, context, and primary research questions. BioProject PRJDB4223 was selected for this study (see Chapter 4.2), with a total of 18 SRA experiments containing one sequencing run each, constituting 113 Gbases of data.

3.1.2 Reference sequence selection

Reference sequences for three *Brucella* species were obtained from the NCBI RefSeq database. Whole genome shotgun sequences for *B. ovis* (type strain NCTC10512, accession

NZ_UFUD01000002) of 1164212 bp, were cultured and sequenced at the Pathogen Informatics facility at the Wellcome Trust Sanger Institute, U.K. *B. melitensis* (accession NZ_QWAI00000000) whole genome shotgun sequences of 3353513 bp were derived from a sheep host and sequenced at the Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', Italy. Finally, *B. abortus* (strain I-181, accession NZ_WNZF00000000) whole genome shotgun sequences of 3252762 bp were sequenced at the Stavropol Plague Control Research Institute. Reference sequences for all three species were extracted from modern samples and selected for this study based on genome size and completeness, and the availability of meta data regarding host species and culturing conditions, although it was not possible to control for such parameters given their differing origins and the preparation methods involved.

3.2 Alignment-based assessment

In bioinformatics, sequence comparison algorithms have typically relied on the positioning of two or more sequences of DNA, RNA, or proteins in alignment to identify regions of similarity and elucidate functional, structural, or evolutionary relationships (Zielezinski et al., 2017). A minimum of two sequences are required for alignment-based comparisons: a subject, or database of interest, and a query, or a metagenomic input sequence that is being compared to others in the database. After sequencing runs were downloaded using the SRA Toolkit, the NCBI Standard Nucleotide BLAST (Basic Local Alignment Search Tool) algorithm was run on Digital Research Alliance of Canada (DRAC) infrastructure to compare query nucleotide sequences of each of the 18 sequencing runs from the selected BioProject against subject sequences from three *Brucella* species (see Chapter 3.1.2). Following BLASTn default guidelines (Wheeler & Bhagwat, 2007), expected (*E*) values were set to 10 to ensure no biologically significant alignments were missed. Alignment scores, *E* values, identity scores, and coverage for each

Brucella subject sequence were then assessed in advance of further analysis (see Chapter 4.3). For a list of basic definitions of BLASTn assessment scores, see Table 1 (Wheeler & Bhagwat, 2007).

Table 1: *Basic definitions for assessment scores for alignment-based analysis (BLASTn)*

Total score	A score computed by assigning a value to each aligned pair of letters/gaps in the sequences being compared, then summing these values over the whole alignment. Higher scores represent better alignments. When total score is the same as max score, there is one global alignment between the two sequences being compared.
Max score	Highest score calculated from matches and mismatches found in local alignments; higher max scores indicate closer alignments.
***E* value**	Known as the 'expect value,' this score represents the probability that an alignment identified in the sequences being compared would be expected to occur by chance given the size of the database being searched. Thus, the *E* value is expected to increase when searching a larger database.
Coverage	Measures the organisation and length of the alignment of the sequences being compared.
Percent identity	Percentage of nucleotides or amino acids that are identical between the sequences being compared. When considered alongside the *E*-value and similarity in conserved regions, may serve as a measure of the relatedness of two sequences, or their evolutionary distance.

3.3 Alignment-free assessment

Analyses of the metagenomic sequencing data sets from PRJDB4223 were largely performed using two types of alignment-free assessments. Alignment-free sequence comparison methods may be a particularly efficient means of processing HTS data as these are less computationally intensive, are resistant to shuffling and recombination events, and may provide

an alternative when alignment-based methods cannot handle low sequence conservation. Whereas the BLASTn program uses nucleotide sequences to find and compare local regions of similarity or alignment, alignment-free analyses (*i.e.*, STAT, Mash) use *k*-mers, a subsequence of a given length (*k*) composed of nucleotides, to reduce or compress large sequences for comparison. Mash, a toolkit developed for *k*-mer based genomic distance estimation, uses the MinHash technique to create sketch representations reduced from large sequences, independent of the size of the genome, and rapidly estimate the similarity of the input sequences (Ondov et al., 2016). The Mash program was run on DRAC infrastructure to sketch the fasta coding files for each sequencing run from the selected BioProject and the three *Brucella* reference genomes before distance estimations were made. Chapter 4.3 presents the results of these analyses alongside their corresponding average nucleotide identity (ANI) scores calculated from Mash distance reports.

Chapter 4: Results

4.1 STAT k-*mer counts: An overview*

The following section presents the results for the first stage of analysis—the review of the STAT predicted taxonomic data set—culminating in the selection of a BioProject and sequencing runs to be analysed further. This process was carried out heuristically in two stages; after each search of the STAT data, original study publications associated with each BioProject was manually identified and retrieved, and an examination of the literature was undertaken before the appropriate data sets were selected.

4.1.1 STAT analysis, stage 1 results

Sequencing runs with significant total and self-counts of *Brucella* (tax_ID = 234) *k*-mers were selected based on a review of counts across the complete data set. This process yielded 36 BioProjects with one or more sequencing runs meeting or exceeding a threshold of 100 *k*-mers (total count). Of these 36, sequencing runs representing samples taken from non-human organisms were removed from consideration; 'human' is defined under the scope of this study as originating from samples designated anatomically modern *Homo sapiens* (tax_ID = 9606). Sequencing runs from two BioProjects were identified as belonging to Neanderthal samples (tax_ID = 63221 [*Homo sapiens neanderthalensis*]), however, these fell outside of the scope of this study and were not considered. Several sequencing runs of apparently 'modern' origin were also not considered; within the scope of this study, 'modern' is defined as samples pre-dating 1800 CE.

This stage of analysis yielded seven possible BioProjects containing sequencing runs with *Brucella* total counts meeting or exceeding 100 *k*-mers and of truly ancient, human origin

(Table 2). Thus, at this stage, 29 of the 36 BioProjects which met the *k*-mer threshold (total count = ≥ 100) but were removed from consideration included 12 from modern samples, six samples whose primary study meta data was unable to be definitively identified, nine of non-human origin, and two ancient samples whose sample extraction techniques may have rendered the data inadmissible (*e.g.*, due to an elevated risk of exogenous contamination). Counts for the removed BioProjects ranged from 1 to 4652 *k*-mers. Sequencing runs of modern origin invariably exhibited some of the highest counts, but this is to be expected given the better preservation of modern samples and the low endogenous fraction of ancient ones.

Table 2: *Selected STAT predicted taxonomic results for* k*-mer total and self-counts for Brucella*[1]

BioProject ID	**Accession**	**Total count**	**Self-count**
PRJEB19769	ERR1880925	126	126
	ERR1880926	142	142
	ERR1880929	188	188
	ERR1880933	125	125
	ERR1883866	55	55
	ERR1883873	98	98
	ERR1883883	96	94
	ERR1883899	26	26
	ERR1883924	51	51
	ERR1883973	117	117
	ERR1884001	69	69
	ERR1884135	180	180
	ERR1884136	394	394
	ERR1884142	429	424
	ERR1884143	171	171
	ERR1884147	96	95
	ERR1884149	87	87
	SRR1884214	273	273
	ERR1884215	31	31
	ERR1884219	187	187
PRJDB4223	DRR046398	162	162
	DRR046399	179	178

[1] This table presents BioProjects in which at least one sequencing run has a total count ≥ 100 for *Brucella*. This table does not include sequencing runs where total counts = <10.

	DRR046400	336	335
	DRR046401	159	158
	DRR046402	276	276
	DRR046412	301	290
	DRR046413	279	270
PRJNA200950	SRR847052	69	69
PRJNA302605	SRR5581853	43	43
	SRR5581857	24	24
	SRR5581858	44	44
PRJNA395646	SRR5875719	40	36
	SRR5881850	102	102
PRJNA48862	SRR7774472	50	49
	SRR7774473	44	44
	SRR5581851	125	125
PRJNA657304	SRR12455961	44	44

4.1.2 STAT analysis, stage 2 results

A search of the STAT predicted taxonomy results for *Brucella melitensis* (tax ID = 29459) further reduced the number of data sets appropriate for selection (Table 3). Only 16 sequencing runs across six BioProjects contained *k*-mers for *B. melitensis*. Similar to stage 1, sequencing runs from non-human and/or modern samples were removed from consideration. As such, based on study parameters (*i.e.*, sample age and sample organism, availability of information regarding original study context and sampling procedures), six BioProjects were identified as possible candidates for further analysis. All but one of the seven BioProjects identified in stage 1 (Table 2) also contained *k*-mers for *B. melitensis* but in significantly lower counts.

Table 3: *Selected STAT predicted taxonomic results for* k*-mer total and self-counts for* Brucella melitensis[2]

BioProject ID	Accession	Total count	Self-count
PRJEB19769	ERR1880926	26	22
	ERR1880929	12	12
	ERR1884135	15	15
	ERR1884136	33	32
	ERR1884142	33	32
	ERR1884214	27	15
	ERR1884219	10	9
PRJDB4223	DRR046400	36	36
	DRR046401	19	19
	DRR046402	79	54
	DRR046412	33	32
	DRR046413	22	22
PRJNA302605	SRR5581853	12	12
PRJNA395646	SRR5881850	15	8
PRJNA48862	SRR5581851	13	13
PRJNA657304	SRR12455961	10	4

4.1.3 Data set selection

In general, the highest *k*-mer counts for *Brucella* from truly ancient samples belonged to PRJEB19769, with *Brucella* total counts ranging from 31 to 429 and representing the largest range in variation between sequencing runs. This BioProject represented the genomic DNA of *Mycobacterium leprae* from five mediaeval European and seven modern skeletons (Krause-Kyora et al., 2018). However, only limited information was available about the original study context, and a review of the original literature associated with the BioProject revealed insufficient information about aDNA authentication measures. BioProject PRJDB4223 also displayed high *k*-mer counts for *B. melitensis* relative to other sequencing runs, ranging from 22 to 79—the latter representing the highest self-count for that species in the STAT data. None of

[2] This table does not include sequencing runs where total counts = <10.

the sequencing runs identified in stage 2 had counts at or exceeding the threshold of 100 set for stage 1. At this stage, PRJDB4223 was identified as the most appropriate data set following a review of the original study literature. In particular, per the scope of this study, sample site, age, and extraction methods, aDNA authentication, geographic context, and original research questions were particularly significant (see Chapter 4.2). The BioProject was also selected in consideration of the presently limited understanding of health and disease dynamics among the Jomon population (see Chapter 2.3.2), and the potential to elucidate the presence of infectious disease in this context.

4.2 Selected data set: Original study and context

The selected sequencing data (PRJDB4223), produced in the Department of Anthropology at the National Museum of Nature and Science, Tokyo, and published by Kanzawa-Kiriyama et al. (2017), represents the mostly complete metagenomes of two Late/Final Jomon individuals from the Sanganji Shell Mound site in Shinchi, Fukushima Prefecture, Japan. The study continues from a previous metagenomic study performed by Kanzawa-Kiriyama et al. (2013), in which the mitochondrial DNA (mtDNA) of four Sanganji Jomon individuals were genotyped using extracts from dental pulp; two of these extracts and one novel extract were assessed for the new study (Kanzawa-Kiriyama et al., 2017). Using the three molar extracts, Kanzawa-Kiriyama et al. (2017) aimed to analyse the origins of modern Japanese individuals through genome sequence comparisons with the aforementioned ancient Jomon individuals; in particular, it was suggested that genome-wide comparisons, rather than mtDNA, would provide a more accurate and efficient means of inferring population origins. The study found that the Jomon samples revealed the greatest genetic affinities with populations of the Japanese archipelago, rather than other Eurasian populations, thereby confirming the widely held

conception that the genetics of modern Japanese populations are at least partly the result of admixture of indigenous Jomon and later migrant populations exhibiting genetic affinities with modern Northeast Asians (Kanzawa-Kiriyama et al., 2017).

Consistent with morphological observation of the remains of the two individuals, Sanganji 131421-3 (represented by samples A1 and A2) and Sanganji 131464 (represented by sample B) were assessed as male and female, respectively, using the ratio of sequence reads mapped to the X and Y chromosomes (Kanzawa-Kiriyama et al., 2017). Three DNA libraries were prepared from the three molar extracts: for sample A1, the GAIIx platform was used to generate 120 bp paired-end sequence reads, while HiSeq2000 was used to generate 100 bp paired-end sequence reads for samples A2 and B (Kanzawa-Kiriyama et al., 2017). Radiocarbon dating of the left upper molar roots (M1 and M2) of Sanganji 131421-3 (samples A1 and A2) yielded calibrated dates of 2994 ± 19 BP, while the right upper second molar of Sanganji 131464 (sample B) indicated a date of 3061 ± 19 BP (Kanzawa-Kiriyama et al., 2017). Both dates correspond to the Late and Final Jomon periods (Kanzawa-Kiriyama et al., 2017; Kuzmin & Keally, 2001). Authentication of mapped sequence reads was carried out in the original study based on low observed ratios of endogenous human DNA as well as characteristic aDNA degradation signals, including C to T misincorporation in the 3' and 5' ends and depurination at the sequence read termini, suggesting that these sequence reads are more likely to contain endogenous Jomon DNA (Kanzawa-Kiriyama et al., 2017). No palaeopathological analysis of the two individuals were reported.

Archaeological reporting from the Sanganji Shell Mound site is limited, including evidence of burial goods or material culture. In general, shell mounds are thus named due to the abundance of mollusc shells (Pearson, 2006), and appear to have flourished during the Late and

Final Jomon periods. Mounds were characteristically constructed as either ring- or horseshoe-shaped; in both cases, mounds had a depression in the centre and revealed evidence of long-term occupation structures such as houses and pits (Kawashima, 2010). Burials at such shell mounds are known to have been either primary (individual) or secondary (collective), and both types are notably found near houses and other structures identified as domestic dwellings (Minagawa & Kondo, 2023). Bioarchaeological analyses of other burials in this mound mainly include trauma, such as an 'arrow wound' in a human vertebra and ulna from two different individuals and an arrowhead embedded in the anterior iliac spine of a third individual (Suzuki, 1958).

Eighteen sequencing runs from PRJDB4223 were downloaded from the SRA and reviewed using the SRA Run Browser to explore the composition of the identified reads. Fourteen sequencing runs in the BioProject were extracted from the Sanganji 131421-3 skeleton (with ten and four sequencing runs, respectively, for samples A1 and A2), and four sequencing runs from the skeleton Sanganji 131464 (sample B).

4.3 BLASTn results

Nucleotide BLAST searches were carried out using reference genomic sequences from *B. melitensis*, *B. ovis*, and *B. abortus* from the RefSeq database (see Chapter 3.1.2). Tables 4, 5, and 6 present maximum score data from BLASTn searches run against sequencing runs DRR046398, DRR046399, DRR046400, DRR046401, DRR046402, DRR046412, DRR046413 from BioProject PRJDB4223. Per the STAT results presented in Chapter 4.1 (Table 2), these seven runs were identified as the most likely to contain *Brucella* alignments. BLASTn searches were also performed on the remaining 11 sequencing runs in the BioProject (see Tables 10–12 in Appendix), although, as expected, results revealed few alignments to *Brucella* reference

sequences, with generally low query coverage (*i.e.*, < 20%) relative to the sequencing runs presented in Tables 4–6.

Table 4: *Maximum BLASTn scores for* Brucella melitensis *for selected sequencing runs from PRJDB4223*

Accession	Sample ID	Max score	Total score	Query coverage (%)	*E* value	Percent identity
DRR046398	Sanganji 131421-3 (A1)	360.65	385.94	26.66	7e-25	86.12
DRR046399	Sanganji 131421-3 (A1)	199.05	223.71	35.18	4e-23	91.13
DRR046400	Sanganji 131421-3 (A1)	361.11	381.72	21.43	7e-25	92.5
DRR046401	Sanganji 131421-3 (A1)	133.45	174.36	18.98	1e-29	87.16
DRR046402	Sanganji 131421-3 (A1)	214	249.19	28.7	1e-20	81.8
DRR046412	Sanganji 131464 (B)	356.5	395.11	36.19	4e-23	93.37
DRR046413	Sanganji 131464 (B)	318.67	346.53	40.65	3e-20	71.2

Table 5: *Maximum BLASTn scores for* Brucella ovis *for selected sequencing runs from PRJDB4223*

Accession	Sample ID	Max score	Total score	Query coverage (%)	*E* value	Percent identity
DRR046398	Sanganji 131421-3 (A1)	297.01	334.28	26.78	7e-25	72.15
DRR046399	Sanganji 131421-3 (A1)	200.15	240.54	33.45	6e-23	85.38
DRR046400	Sanganji 131421-3 (A1)	361.98	388.11	21.7	5e-26	93.46
DRR046401	Sanganji 131421-3 (A1)	125.55	132.17	26.48	2e-24	90.06
DRR046402	Sanganji 131421-3 (A1)	261.12	312.59	30.11	1e-20	89.54
DRR046412	Sanganji 131464 (B)	398.88	455.33	39.55	3e-27	92.42
DRR046413	Sanganji 131464 (B)	315.9	391.81	39.32	3e-21	75.97

Table 6: *Maximum BLASTn scores for* Brucella abortus *for selected sequencing runs from PRJDB4223*

Accession	Sample ID	Max score	Total score	Query coverage (%)	*E* value	Percent identity
DRR046398	Sanganji 131421-3 (A1)	312.6	330.05	33.15	6e-27	84.12
DRR046399	Sanganji 131421-3 (A1)	203.06	264.19	43.58	5e-27	93.74
DRR046400	Sanganji 131421-3 (A1)	363.8	383.7	27.61	5e-27	93.26
DRR046401	Sanganji 131421-3 (A1)	154.81	188.51	36.99	2e-24	84.14

DRR046402	Sanganji 131421-3 (A1)	274.3	327.77	38.87	2e-23	91.78
DRR046412	Sanganji 131464 (B)	387.39	444.83	40.01	1e-24	92.87
DRR046413	Sanganji 131464 (B)	325.88	387.97	42.48	3e-22	73.65

A review of the BLASTn results for each of the seven sequencing runs revealed max scores in the range of 133.45 (DRR046401) to 356.5 (DRR046412) for *B. melitensis*, 125.55 (DRR046401) to 398.88 (DRR046412) for *B. ovis*, and 154.82 (DRR046401) to 387.39 for *B. abortus* (DRR046412). In general, sequencing run DRR046412 consistently produced the highest max scores and high coverage and percent identity scores relative to other runs for all three subject sequences. *E* values across the results are consistent with the low query coverage scores; these are not unexpected given the age and degraded nature of the sample and the low probability of successfully detecting *Brucella* sequences in the selected runs. The possibility of evolutionary divergence between bacterial sequences (if present) in the sample and their modern equivalents, as well as the probability of detecting homologs, must also be considered. Thus, further alignment-free analysis was undertaken in order to generate more potential data for consideration.

4.4 Mash results

Mash, a MinHash-based, alignment-free assessment tool for rapid genome and metagenome distance estimation using *k*-mers was employed as a final method to confirm the possible presence of *Brucella* in the seven sequencing runs predicted to contain the reference sequences. In general, Mash distances (*D*) strongly correlate to average nucleotide identity (ANI), a commonly used measure of sequence similarity, so that $D \approx 1 - \text{ANI}$ (Ondov et al., 2016). Thus, a Mash distance of ≤ 0.05 equates to an ANI of $\geq 95\%$, which is approximately the identity threshold used to distinguish similar species genomes (Ondov et al., 2016). As such,

although not provided by Mash, this ANI score (calculated as $D = 1 - \text{ANI}$) is provided here to contextualise the Mash distances for a more efficient analysis, but must be considered alongside all data presented, including *P* values and matching hashes. Sketch sizes were kept at $s = 1000$, the default set by the program and recommended by the developers, to enable the most accurate distance estimates to be made (Ondov et al., 2016). Tables 7, 8, and 9 present the results of the Mash analysis, including Mash distances, *P* values, matching hashes, and ANI values.

Table 7: *Mash distance estimation scores for* Brucella melitensis *for selected sequencing runs from PRJDB4223*

Accession	Sample ID	Mash distance (*D*)	*P* value	Matching hashes	ANI (%)
DRR046398	Sanganji 131421-3 (A1)	0.29872	0.0620588	1/1000	70.128
DRR046399	Sanganji 131421-3 (A1)	0.3466	0.0634898	2/1000	65.34
DRR046400	Sanganji 131421-3 (A1)	0.17293	0.089384	1/1000	82.707
DRR046401	Sanganji 131421-3 (A1)	0.278931	0.061	1/1000	72.0169
DRR046402	Sanganji 131421-3 (A1)	0.298616	0.06742	2/1000	70.1384
DRR046412	Sanganji 131464 (B)	0.137653	0.06473	1/1000	86.2347
DRR046413	Sanganji 131464 (B)	0.219321	0.056869	1/1000	78.0697

Table 8: *Mash distance estimation scores for* Brucella ovis *for selected sequencing runs from PRJDB4223*

Accession	Sample ID	Mash distance (*D*)	*P* value	Matching hashes	ANI (%)
DRR046398	Sanganji 131421-3 (A1)	0.28867	0.053947	1/1000	71.133
DRR046399	Sanganji 131421-3 (A1)	0.333311	0.050532	2/1000	66.6689
DRR046400	Sanganji 131421-3 (A1)	0.20008	0.0502632	1/1000	79.992
DRR046401	Sanganji 131421-3 (A1)	0.256739	0.050328	1/1000	74.3261
DRR046402	Sanganji 131421-3 (A1)	0.365561	0.053829	2/1000	63.4439
DRR046412	Sanganji 131464 (B)	0.093726	0.04783	2/1000	90.6274
DRR046413	Sanganji 131464 (B)	0.218364	0.05128	1/1000	78.1636

Table 9: *Mash distance estimation scores for* Brucella abortus *for selected sequencing runs from PRJDB4223*

Accession	Sample ID	Mash distance (*D*)	*P* value	Matching hashes	ANI (%)
DRR046398	Sanganji 131421-3 (A1)	0.263728	0.053947	1/1000	73.6272
DRR046399	Sanganji 131421-3 (A1)	0.335833	0.050532	2/1000	66.4167
DRR046400	Sanganji 131421-3 (A1)	0.202847	0.0502632	1/1000	79.7153
DRR046401	Sanganji 131421-3 (A1)	0.14738	0.050328	1/1000	85.262
DRR046402	Sanganji 131421-3 (A1)	0.346381	0.053829	2/1000	65.3619
DRR046412	Sanganji 131464 (B)	0.093997	0.04783	3/1000	90.6003
DRR046413	Sanganji 131464 (B)	0.229837	0.05128	1/1000	77.0163

Across reference sequences for all three *Brucella* species, Mash distances were greater than the threshold suggested by Ondov et al. (2016) of $D = \leq 0.05$, while scores for matching hashes were distinctly low (approximately 1–3 out of a sketch size of $s = 1000$). In general, Mash scores (distance, *P* values, and matching hashes) were relatively consistent across *Brucella* species, while exhibiting greater internal variation across the selected sequencing runs. The highest ANI values were produced by sequencing runs DRR046400 and DRR46412. Notably, sequencing run DRR046412 exhibited the highest ANI values across all three reference genomes, generally consistent with the BLASTn results presented in Chapter 4.3 (Tables 4–6). *P* values for distance estimation analysis against *B. abortus* and *B. ovis* more closely approximated the standard $P = < 0.05$ threshold (Ondov et al., 2016), while this value was less statistically significant when *B. melitensis* sequencing runs were queried. Taken together, these scores may reflect the sketch size ($s = 1000$), in which case, further analyses with larger sketch sizes may be more fruitful, especially since larger sketch sizes may generate more accurate Mash distances, particularly for larger genomes (Ondov et al., 2016).

It must also be taken into account that the Mash algorithm is better suited for the analysis of modern samples where microbial genomes are expected to be less degraded. Additionally, the

use of modern reference sequences for *Brucella* may result in a reduced ability to detect similarity across the sequences, particularly given that modern *Brucella* genomes may have diverged from ancient ones; if so, it may be possible that otherwise conserved sequences of the ancient *Brucella* genomes were not present in the sample sequencing runs, and were therefore not detectable (see Chapter 5.2). In general, the sample sequencing runs detected to contain *Brucella* using STAT, and further analysed using BLASTn and Mash, were recovered from sample A1 (Sanganji 131421-3), whereas two sequencing runs originated in sample B (Sanganji 131464). While not conclusive, these data may suggest that, rather than pathogenic *Brucella* being detected in the samples (since infection with *Brucella* for two individuals within a data set appears unlikely), it is environmental *Ochrobactrum* molecules—possibly from the burial context—that have instead been detected (see Chapter 5.4). Mash distance estimations were not performed on the 11 remaining sequencing runs that form part of PRJDB4223, and this may represent an area of future analysis, particularly to establish a control group and compare results across sequencing runs predicted and not predicted to contain *k*-mers for *Brucella*. Additionally, further analysis may be required to determine whether potential evolutionary divergences in *Brucella* lineages may have emerged and, if so, the kinds of changes that may be expected within this data set.

Chapter 5: Discussion

5.1 Considerations in sample selection and preservation

The application of anthropological approaches in metagenomics entails the incorporation of particular perspectives and priorities at every stage of the research design process. The numerous heuristic concerns associated with such secondary aDNA studies may be broadly described as issues of sample selection and the availability of meta data. Overall, metagenomic data sets in online repositories number few and sequencing databases like the NCBI SRA offer only limited functionality for searching project meta data to enable the user to distinguish ancient from modern specimens. Furthermore, sequencing runs from ancient specimens may be labelled 'fossil metagenome' at the uploader's discretion, yet no inclusion criteria guidelines are made available. Thus, sequences from both relatively modern and truly ancient samples are often labelled 'fossil metagenome', and researchers conducting secondary analyses of the SRA must manually retrieve primary project information in order to confirm specimen age and origin. This is particularly salient for sequencing runs from ancient samples containing apparently high bacterial content. For example, in this study at least one BioProject identified by STAT with high *Brucella k*-mer counts was found to have been sampled from modern horses; further examination using the SRA Run Browser revealed that bacteria represented more than 42% of the identified reads from these sequencing runs, thus increasing the likelihood of bacterial DNA detection.

Sequencing data published in the SRA is indexed by the Entrez database system, which allows users to access integrated nucleotide and protein sequence data from across the NCBI's molecular and literature databases (National Center for Biotechnology Information, 2006). The Entrez system supports the use of Boolean terms, query translation, automatic term mapping, and

fielded searching (National Center for Biotechnology Information, 2006). For example, a user may construct a Boolean query which integrates search terms for species, protein, sequence length, and publication date while also using filters to restrict particular kinds of molecular records. While Entrez does provide a suite of search options for technically-specific queries (*e.g.*, searching according to nucleotide, protein, rRNA, mDNA, SNP, genome, *etc*.), the system does not index fields related to project parameters or specimen information such as geographic region, time period, organism, or original study; thus, relatively fewer options are available when searching for such information, and users are generally required to employ key query terms related to their subject of interest to ensure that the appropriate data sets are retrieved.

Finally, a lack of standardised terms in project meta data published on the archive poses a significant challenge when aiming to retrieve all samples or data sets that match the contextual criteria, and as a result, complex Boolean queries may need to be generated to account for possible variations in terminology and the degree of specificity. These limitations necessarily impact the kinds of searches that can be performed and limits the range of sampling techniques for bioarchaeologists and anthropologists interested in exploring the metagenomic data available in repositories. These limitations in search functionality result in an inherent selection bias in the study, such that SRA sequencing data sets that are published with detailed and accurate meta data are more likely to be selected to the exclusion of other potentially appropriate data sets.

5.2 Reference sequences and **Brucella** *lineages*

For this study, the STAT predicted taxonomy results presented Chapter 4.1 were analysed using only *Brucella* and *Brucella melitensis* taxonomic ID tags, while the following analyses employed genomic reference sequences for three *Brucella* species, including *B. melitensis*. Given

the high degree of genomic similarity, shared DNA fragments, and orthologs between *B. melitensis*, *B. abortus*, and *B. ovis* (Halling et al., 2005), these species were selected for further downstream analysis, while *B. suis* was not selected for analysis given its putative paraphyletic status, divergence from other brucellae, and high degree of intraspecific genetic diversity (Foster et al., 2019). *B. melitensis*, as the most common human pathogen among the brucellae (Gomez et al., 2013), was selected for all three analyses due to the higher likelihood of infection by that species in an individual human sample. Future advances related to this study may benefit from the inclusion of *B. suis* reference sequences to maximise the possibility of identifying reads from paraphyletic *Brucella* clades.

Genomic reference sequences for *Brucella* were retrieved from the NCBI RefSeq database, namely, *B. melitensis*, *B. abortus*, and *B. ovis*. Reference sequences for the three organisms were sequenced in different laboratories and derived from different hosts; limited information is available about host organisms and culturing conditions (see Chapter 3.1.4), however, all are modern samples. As Weiß et al. (2020) note, nucleotide divergence between sample and reference sequences may result in difficulties assessing typical age-associated deamination signals or damage patterns and thereby hinder authentication. Other concerns associated with the use of modern reference sequences may involve the degree to which the reference sequences are representative of the described ancient species and strains. A notable example of such a challenge is presented by Sereno et al. (2018), who were able to detect using NGS the presence of an infectious bacterium, *Leishmania*, in three pre-Columbian Andean mummies; however, given the absence of the species subtype in their reference database, the authors suggested the possibility of an infection by an ancient *Leishmania* species currently unknown to researchers or of an extinct lineage. The potential genetic or evolutionary distance

between ancient infectious pathogens and their modern reference counterparts is important to consider given the age of the metagenomic data sets under analysis and the degree of DNA damage in the form of deamination.

5.3 Brucella *in the Jomon period*

The evolutionary origins of *Brucella* lineages remain somewhat enigmatic, with palaeopathological evidence suggesting the sporadic (if not endemic) presence of brucellosis by the late Neolithic and Early Bronze Ages (see Chapter 2.2). Molecular clock evidence suggests that brucellae may have been endemic by about 296,000 to 86,000 BP in wildlife populations—prior to animal domestication and the subsequent emergence of domesticated livestock hosts (Foster et al., 2009); additionally, at least one putative assessment has been made for an *Australopithecus africanus* specimen from South Africa (D'Anastasio et al., 2009). Taken together, this evidence suggests that brucellosis may have been sporadically contracted by individuals in past communities regardless of the presence of livestock domesticates. While the genomic characteristics of these early, wildlife-associated brucellae are unclear, there exists the possibility that human brucellosis may have been transmitted by means other than close human–livestock contact, such as close contact with wild animals or their products, perhaps during hunting or other food procurement or processing activities.

Little is known about the possible prevalence of brucellosis—or, indeed, other infectious diseases—in Jomon populations, and circumstantial evidence may instead be considered. Brucellosis has been diagnosed in the remains of pre-contact Indigenous peoples of Alaska pre-dating the presence of animal domesticates (*e.g.*, sheep, cows, pigs) in the region (Shephard & Rode, 1996). Oxenham et al. (2013) observe the epidemiological similarities between pre-contact

Indigenous Alaskan groups and subarctic Japanese populations during the historic period—in particular, with regard to the marine component of the diet, environmental conditions, and the common presence of diseases such as tuberculosis. Nevertheless, limited evidence exists for the presence of brucellosis as a disease in communities without animal domesticates, and palaeoepidemiological and anthropological models continue to associate the disease with the emergence of animal domestication and agriculture (Schug et al., 2023). Thus, care must be taken when inferring the presence of brucellosis within a non-agricultural population—or one exhibiting limited evidence of agricultural practices—such as the Late/Final Jomon groups represented in this study.

5.4 Other limitations

The close phylogenetic relationship between *Ochrobactrum* and *Brucella* presents challenges for the identification of the latter in metagenomic sequences. Indeed, despite differences in structure, physiology, genomic traits, clinical presentation, pathogenicity, as well as disease epidemiology between the two bacterial groups (Moreno et al., 2023), studies have attempted to reclassify *Ochrobactrum* under the genus *Brucella* on a cladistic, rather than evolutionary, basis (Hördt et al., 2020). This—in addition to the fact that *Brucella* strains are identifiable only as a single species when using DNA–DNA hybridisation (DDH) techniques—has led researchers to report that the overall genomic divergence and diversity of the *Brucella*–*Ochrobactrum* clade are lower than in other, single-genus bacterial clades (Hördt et al., 2020).

Such classifications remain a matter of taxonomic and clinical debate (Moreno et al., 2022; Moreno et al., 2023), although from a genomic perspective Scholz et al. (2008) report that 16S rRNA gene-based comparative sequence analyses (*rrs* and *recA*) have yielded similarities of

96.2% and 85.5% within *Brucella*, complicating inter-species differentiation on a genomic level. Furthermore, the same study reports that at the protein level, *recA* sequences within *Ochrobactrum* and between *Ochrobactrum* and *Brucella* are highly similar and exhibit only a small number of amino acid substitutions (Scholz et al., 2008). Thus, it has been recommended that *recA* and *rrs* gene-analysis should be performed for accurate species identification when subtyping *Ochrobactrum* or *Brucella* (Scholz et al., 2008). Given the presence of *Ochrobactrum* in the environment and its status as an opportunistic pathogen of humans, the distinction between *Brucella* and *Ochrobactrum* in epidemiological and pathological terms is significant. In this study, five of the seven sequencing runs detected as possibly containing *Brucella* originated in sample A1 (Sanganji 131421-3), while two sequencing runs were extracted from sample B (Sanganji 131464). Thus, given the fact that the affected samples represent two likely non-contemporaneous skeletons and the low probability of infection by the same pathogen in both individuals at the time of death, the possibility of environmental or exogenous contamination by *Ochrobactrum* or its homologs, rather than infection by *Brucella*, must be considered. As a result, future developments of this study may benefit from further analyses such as metagenome assembly and gene-based comparative sequence analyses to accurately subtype *Ochrobactrum* and *Brucella* considering their pathogenicity and low levels of intra-clade diversity.

Chapter 6: Conclusions

6.1 Conclusions

Data mining techniques—namely, the processing and analysis of large volumes of data—have long been a staple of bioinformatics (Wang et al., 2005) but represent a somewhat newer introduction to anthropology (Paff, 2021). Anthropological fieldwork and research both incorporates and debates the role of local, particularistic, and 'bottom-up' approaches vis-à-vis 'top-down', seemingly objective approaches (Paff, 2021). However, quantitative methods need not be 'top-down' (Paff, 2021); both quantitative and qualitative data are regularly successfully integrated into bioarchaeological investigations of human health and disease in the past at various scales of analysis from the individual to the population, and to this end this study offers a potential starting point for biomolecular research that incorporates basic bioinformatics tools and techniques for quantitative analysis integrated with anthropological and archaeological knowledge. Inspired by biocultural theory, this project emphasises a holistic examination of the presently available archaeological and bioarchaeological data on the Late/Final Jomon context under consideration.

This study has introduced a potential approach to mining the NCBI Sequence Read Archive, a public repository of high-throughput sequencing data using bioinformatics tools. Following a STAT search of the database, two forms of analysis—BLASTn and Mash—were employed to confirm, and further elucidate, the possible presence of the candidate *Brucella* species (*B. melitensis, B. ovis, B. abortus*) in the sequencing runs selected for deeper study from BioProject PRJDB4223. BLASTn and Mash analyses revealed that sequencing runs DRR046400 and DRR046412 exhibited the highest scores and the most promising potential for a *Brucella*

fraction to be identified. The results of these analyses remain inconclusive insofar as further analysis is needed for more accuracy, but they suggest that sequences similar to the reference *Brucella* species may be present at varying rates within seven of the 18 runs in the data set. These results, however, must be interpreted with reference to anthropological and archaeological knowledge of Late/Final Jomon populations. The absence of concrete evidence of animal domesticates, but the possibility of infection via wildlife vectors, leaves the question of brucellosis in the Jomon context unclear.

6.2 Future directions

The analyses and results presented herein offer insight into the potential presence of infectious disease in the Jomon context—a heretofore unexplored area of bioarchaeological research. Further analysis of the selected sequencing data will enable the findings in this study to be confirmed more stringently; in particular, metagenome assembly and taxonomic classification and binning may offer precise insight into the nature of the *Brucella* lineage, strain, or species present, and potentially gene function and pathogenicity. In contrast to analyses performed on unassembled metagenomes, assembled metagenomes offer greater sensitivity where small microbial genomes are suspected. Metagenome co-assembly, in particular, may permit the recovery of genes present in low abundances that may otherwise be undetectable (Delgado & Andersson, 2022). Given the close phylogenetic relationship between, and low level of diversity within, the putative *Brucella–Ochrobactrum* clade, greater sensitivity in analysis would enable distinctions to be made at the intra-clade level. Finally, given the environmental presence of *Ochrobactrum* species, future studies may benefit from analyses of metagenomic sequencing data from comparable archaeological contexts as a control against which the findings of this study may be compared to determine the likelihood of exogenous contamination.

Future studies of SRA data may benefit from changes to database submission functionality in two key areas: the standardisation of meta data terminology and the expansion of searchability. Firstly, when made available to uploaders during the submission process (for example, as part of a user's guide), standardised terms and well-defined meta data tags that cover a broad range of contextual and molecular information (*e.g.*, sample type, origin, date, organism type) would enable greater consistency in the indexing and searchability of data in the Archive. Challenges in achieving a consensus in standardised terminology among scientists and Archive administrators may be expected due to variations in term use within and across the sciences; thus, replacing subjective and imprecise terms such as 'fossil metagenome' with precise date/age range tags for each record may be the most effective means of reducing ambiguity. Importantly, however, such standardisation would allow SRA data to be more efficiently and accurately screened during the initial stages of a secondary study and reduce downstream bias by decreasing the likelihood of missed or overlooked records due to mislabelling.

Secondly, as an important service provided by the SRA to its users, expanded searchability using native meta data terms is important for any initial queries of the database. A simple solution is the incorporation of standardised terms into the SRA's Boolean search functionality, thereby allowing users to quickly identify data indexed under the appropriate tags. This process could be integrated into the Archive's current search function, but future modifications may benefit from the addition of filtering or sorting options. Such adaptations would not only expand the availability of data to researchers across the sciences—in particular, in fields where bioinformatics data mining has seen little involvement—but also broaden the scope of questions able to be asked of the metagenomic data in the SRA.

The integration of metagenomic approaches in anthropology makes available new lines of inquiry into past human experience. Human health and disease studies form an integral aspect of bioarchaeological studies (Glencross, 2011) and this, coupled with ethical concerns about destructive analysis, has resulted in a growing call from public stakeholders and the broader scientific community for anthropology to develop new, minimally destructive methods in the analysis of human skeletal remains (Jones & Harris, 1998). If data collection methods are rigorous and raw sequences and post-analysis data are made available when possible, secondary analyses of mined data can provide opportunities for ancient genomes and metagenomes to be further explored while new techniques and methods suitable for analysing ancient specimens are developed. This 'recycling' of genomic or metagenomic data presents a potential path forward for anthropologists concerned about the impact of destructive biomolecular analysis on the skeletons being sampled, and more broadly, the ethical imperative to preserve archaeological human remains wherever possible. By integrating methods and tools from across the social and biological sciences, anthropology stands to benefit not only by expanding the horizons of research and supporting the respectful stewardship of human remains, but especially by establishing meaningful connections based on shared interests between the scientific community and the public.

References

Adler, C. J., Haak, W., Donlon, D., & Cooper, A. (2011). Survival and recovery of DNA from ancient teeth and bones. *Journal of Archaeological Science*, *38*(5), 956–964. https://doi.org/10.1016/j.jas.2010.11.010

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High throughput sequencing: An overview of sequencing chemistry. *Indian Journal of Microbiology*, *56*(4), 394–404. https://doi.org/10.1007/s12088-016-0606-4

Aubin, M. M. (2004). Brucellosis in ancient Nubia: Morbidity in biocultural perspective through time at Semna South, Sudan. *American Journal of Physical Anthropology*, *56*(S37).

Ayling, M., Clark, M. D., & Leggett, R. M. (2020). New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics*, *21*(2), 584–594. https://doi.org/10.1093/bib/bbz020

Blakey, M. L. (1998). Beyond European Enlightenment: Toward a critical and humanistic human biology." In A. H. Goodman & T. L. Leatherman (Eds.), *Building a new biocultural synthesis: Political-economic perspectives on human biology* (pp. 379–405). The University of Michigan Press.

Blakey, M. L. (2008). An ethical epistemology of publicly engaged biocultural research. In J. Habu, C. Fawcett, & J. M. Matsunaga (Eds.), *Evaluating multiple narratives* (pp. 17–28). Springer New York. https://doi.org/10.1007/978-0-387-71825-5_2

Bleed, P., & Matsui, A. (2010). Why didn't agriculture develop in Japan? A consideration of Jomon ecological style, niche construction, and the origins of domestication. *Journal of Archaeological Method and Theory*, *17*(4), 356–370. https://doi.org/10.1007/s10816-010-9094-8

Bolnick, D. A., Bonine, H. M., Mata-Míguez, J., Kemp, B. M., Snow, M. H., & LeBlanc, S. A. (2011). Nondestructive sampling of human skeletal remains yields ancient nuclear and mitochondrial DNA. *American Journal of Physical Anthropology*, *147*(2), 293–300. https://doi.org/10.1002/ajpa.21647

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., Wilkins, M. J., Wrighton, K. C., Williams, K. H., & Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*, *523*(7559), 208–211. https://doi.org/10.1038/nature14486

Buikstra, J. E. (2010). Paleopathology: A contemporary perspective. In C. S. Larsen (Ed.), *A companion to biological anthropology* (pp. 395–411). Wiley-Blackwell.

Buzgan, T., Karahocagil, M. K., Irmak, H., Baran, A. I., Karsen, H., Evirgen, O., & Akdeniz, H. (2010). Clinical manifestations and complications in 1028 cases of brucellosis: A retrospective evaluation and review of the literature. *International Journal of Infectious Diseases*, *14*(6), e469–e478. https://doi.org/10.1016/j.ijid.2009.06.031

Capasso, L. (1999). Brucellosis at Herculaneum (79 AD). *International Journal of Osteoarchaeology*, *9*(5), 277–288. https://doi.org/10.1002/(SICI)1099-1212(199909/10)9:5%3C277::AID-OA489%3E3.0.CO;2-0

Coughlan, L. M., Cotter, P. D., Hill, C., & Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontiers in Microbiology*, *6*(672). https://doi.org/10.3389/fmicb.2015.00672

D'Anastasio, R., Staniscia, T., Milia, M. L., Manzoli, L., & Capasso, L. (2011). Origin, evolution and paleoepidemiology of brucellosis. *Epidemiology and Infection*, *139*(1), 149–156. https://doi.org/10.1017/S095026881000097X

D'Anastasio, R., Zipfel, B., Moggi-Cecchi, J., Stanyon, R., & Capasso, L. (2009). Possible brucellosis in an early hominin skeleton from Sterkfontein, South Africa. *PLoS ONE*, *4*(7), e6439. https://doi.org/10.1371/journal.pone.0006439

Delgado, L. F., & Andersson, A. F. (2022). Evaluating metagenomic assembly approaches for biome-specific gene catalogues. *Microbiome*, *10*(1), 72. https://doi.org/10.1186/s40168-022-01259-2

de la Fuente, C., Flores, S., & Moraga, M. (2013). DNA from human ancient bacteria: A novel source of genetic evidence from archaeological dental calculus. *Archaeometry*, *55*(4), 767–778. https://doi.org/10.1111/j.1475-4754.2012.00707.x

Delam, H., Keshtkaran, Z., Rezaei, B., Soufi, O., & Bazrafshan, M.-R. (2022). Changing patterns in epidemiology of brucellosis in the south of Iran (2015–2020): Based on Cochrane-Armitage trend test. *Annals of Global Health*, *88*(1), 11. https://doi.org/10.5334/aogh.3474

Der Sarkissian, C., Ermini, L., Jónsson, H., Alekseev, A. N., Crubezy, E., Shapiro, B., & Orlando, L. (2014). Shotgun microbial profiling of fossil remains. *Molecular Ecology*, *23*(7), 1780–1798. https://doi.org/10.1111/mec.12690

Der Sarkissian, C., Velsko, I. M., Fotakis, A. K., Vågene, Å. J., Hübner, A., & Fellows Yates, J. A. (2021). Ancient metagenomic studies: Considerations for the wider scientific community. *mSystems*, *6*(6), e01315-21. https://doi.org/10.1128/msystems.01315-21

Donoghue, H. D. (2013). Insights into ancient leprosy and tuberculosis using metagenomics. *Trends in Microbiology*, *21*(9), 448–450. https://doi.org/10.1016/j.tim.2013.07.007

dos Reis, M., Donoghue, P. C. J., & Yang, Z. (2016). Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics*, *17*(2), 71–80. https://doi.org/10.1038/nrg.2015.8

Douzery, E. J. P., Delsuc, F., & Philippe, H. (2006). Molecular dating in the genomic era. *Médecine Sciences*, *22*(4), 374–380. https://doi.org/10.1051/medsci/2006224374

Dressler, W. W. (1995). Modeling biocultural interactions: Examples from studies of stress and cardiovascular disease. *American Journal of Physical Anthropology*, *38*(S21), 27–56. https://doi.org/10.1002/ajpa.1330380604

Elsaghir, A., & James, E. (2003). Misidentification of *Brucella melitensis* as *Ochrobactrum anthropi* by API 20NE. *Journal of Medical Microbiology*, *52*, 441–442. https://doi.org/10.1099/jmm.0.05153-0

Eshed, V., Gopher, A., Pinhasi, R., & Hershkovitz, I. (2010). Paleopathology and the origin of agriculture in the Levant. *American Journal of Physical Anthropology*, *143*(1), 121–133. https://doi.org/10.1002/ajpa.21301

Foster, J. T., Beckstrom-Sternberg, S. M., Pearson, T., Beckstrom-Sternberg, J. S., Chain, P. S. G., Roberto, F. F., Hnath, J., Brettin, T., & Keim, P. (2009). Whole-genome-based phylogeny and divergence of the genus *Brucella*. *Journal of Bacteriology*, *191*(8), 2864–2870. https://doi.org/10.1128/JB.01581-08

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

Gigi, R., Flusser, G., Kadar, A., Salai, M., & Elias, S. (2017). *Ochrobactrum anthropi*-caused osteomyelitis in the foot mimicking a bone tumor: Case report and review of the literature. *The Journal of Foot and Ankle Surgery*, *56*(4), 851–853. https://doi.org/10.1053/j.jfas.2017.02.008

Glencross, B. (2011). Skeletal injury across the life course: Towards understanding social agency. In S. C. Agarwal & B. A. Glencross (Eds.). *Social bioarchaeology* (pp. 390–409). Wiley-Blackwell.

Gomez, G., Adams, L. G., Rice-Ficht, A., & Ficht, T. A. (2013). Host–*Brucella* interactions and the *Brucella* genome as tools for subunit antigen discovery and immunization against brucellosis. *Frontiers in Cellular and Infection Microbiology*, *3*, 17. https://doi.org/10.3389/fcimb.2013.00017

Goodman, A. H., Brooke Thomas, R., Swedlund, A. C., & Armelagos, G. J. (1988). Biocultural perspectives on stress in prehistoric, historical, and contemporary population research. *American Journal of Physical Anthropology*, *31*(S9), 169–202. https://doi.org/10.1002/ajpa.1330310509

Goodman, A. H., & Leatherman, T. L. (1998). Traversing the chasm between biology and culture: An introduction. In A. L. Goodman & T. L. Leatherman (Eds.), *Building a new biocultural synthesis: Political-economic perspectives on human biology* (pp. 3–41). The University of Michigan Press.

Granehäll, L., Huang, K. D., Tett, A., Manghi, P., Paladin, A., O'Sullivan, N., Rota-Stabelli, O., Segata, N., Zink, A., & Maixner, F. (2021). Metagenomic analysis of ancient dental calculus reveals unexplored diversity of oral archaeal *Methanobrevibacter*. *Microbiome*, *9*(1), 197. https://doi.org/10.1186/s40168-021-01132-8

Haeusler, M. (2019). Evolutionary origin of musculoskeletal problems. In E. Been, A. Gómez-Olivencia, & P. A. Kramer (Eds.), *Spinal evolution: Morphology, function, and pathology of the spine in hominoid evolution* (pp. 213–245). Springer. https://doi.org/10.1007/978-3-030-19349-2

Halling, S. M., Peterson-Burch, B. D., Bricker, B. J., Zuerner, R. L., Qing, Z., Li, L.-L., Kapur, V., Alt, D. P., & Olsen, S. C. (2005). Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. *Journal of Bacteriology*, *187*(8), 2715–2726. https://doi.org/10.1128/JB.187.8.2715-2726.2005

Harney, É., Cheronet, O., Fernandes, D. M., Sirak, K., Mah, M., Bernardos, R., Adamski, N., Broomandkhoshbacht, N., Callan, K., Lawson, A. M., Oppenheimer, J., Stewardson, K., Zalzala, F., Anders, A., Candilio, F., Constantinescu, M., Coppa, A., Ciobanu, I., Dani, J., … Pinhasi, R. (2021). A minimally destructive protocol for DNA extraction from ancient teeth. *Genome Research*, *31*(3), 472–483. https://doi.org/10.1101/gr.267534.120

Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y.-C., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L., Lemma, E., Bar-Gal, G. K., & Spigelman, M. (2008). Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the eastern Mediterranean. *PLoS ONE*, *3*(10), e3426–e3426. https://doi.org/10.1371/journal.pone.0003426

Hodgins, H. P., Chen, P., Lobb, B., Wei, X., Tremblay, B. J. M., Mansfield, M. J., Lee, V. C. Y., Lee, P.-G., Coffin, J., Duggan, A. T., Dolphin, A. E., Renaud, G., Dong, M., & Doxey, A. C. (2023). Ancient *Clostridium* DNA and variants of tetanus neurotoxins associated with human archaeological remains. *Nature Communications*, *14*(1), 5475. https://doi.org/10.1038/s41467-023-41174-0

Hodgkins, J. M. (2003). The tale that tail bones tell about the antiquity of the human disease brucellosis. *American Journal of Physical Anthropology*, *120*(S36), 115. https://doi.org./10.1002/ajpa.10249

Hoke, M. K., & Schell, L. M. (2020). Doing biocultural anthropology: Continuity and change. *American Journal of Human Biology*, *32*(4), e23471. https://doi.org/10.1002/ajhb.23471

Hördt, A., López, M. G., Meier-Kolthoff, J. P., Schleuning, M., Weinhold, L.-M., Tindall, B. J., Gronow, S., Kyrpides, N. C., Woyke, T., & Göker, M. (2020). Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of Alphaproteobacteria. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00468

Hruschka, D. J., Lende D. H., & Worthman, C. M. (2005). Biocultural dialogues: Biology and culture in psychological anthropology. *Ethos*, *33*(1), 1–19. httpss://doi.org/10.1525/eth.2005.33.1.001

Hudson, M. J. (2020). Slouching toward the Neolithic: Complexity, simplification, and resilience in the Japanese archipelago. In G. R. Schug (Ed.), *The Routledge Handbook of the Bioarchaeology of Climate and Environmental Change* (1st ed.) (pp. 379–395). Routledge. https://doi.org/10.4324/9781351030465

Human Microbiome Project Consortium. (2012). Structure, function, and diversity of the healthy human microbiome. *Nature*, *486*, 207–214. https://doi.org/10.1038/nature11234

Inoue, N., Kuo, C. H., Ito, G., Kamegai, T. (1981). Dental diseases in Japanese skeletal remains: II. Later Jomon period. *Journal of the Anthropological Society of Nippon*, *89*(3), 363–378.

Jones, D. G., & Harris, R. J. (1998). Archeological human remains: Scientific, cultural, and ethical considerations. *Current Anthropology*, *39*(2), 253–264. https://doi.org/10.1086/204723

Kaestle, F. A. (2010). Palaeogenetics: Ancient DNA in anthropology. In C. S. Larsen (Ed.), *A companion to biological anthropology* (pp. 427–442). Wiley-Blackwell.

Kanzawa-Kiriyama, H., Kryukov, K., Jinam, T. A., Hosomichi, K., Saso, A., Suwa, G., Ueda, S., Yoneda, M., Tajima, A., Shinoda, K.-I., Inoue, I., & Saitou, N. (2017). A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. *Journal of Human Genetics*, *62*(2), 213–221. https://doi.org/10.1038/jhg.2016.110

Kanzawa-Kiriyama, H., Saso, A., Suwa, G, & Saitou, N. (2013). Ancient mitochondrial DNA sequences of Jomon teeth samples from Sanganji, Tohoku district, Japan. *Anthropological Science*, *121*, 89–103. https://doi.org/10.1537/ase.121113

Katz, K. S., Shutov, O., Lapoint, R., Kimelman, M., Brister, J. R., & O'Sullivan, C. (2021). STAT: A fast, scalable, MinHash-based k-mer tool to assess Sequence Read Archive next-generation sequence submissions. *Genome Biology*, *22*, 270. https://doi.org/10.1186/s13059-021-02490-0

Kawashima, T. (2010). Mounds and rituals in the Jomon period. *Documenta Praehistorica*, *37*, 185–194. https://doi.org/10.4312/dp.37.16

Kay, G. L., Sergeant, M. J., Giuffra, V., Bandiera, P., Milanese, M., Bramanti, B., Bianucci, R., & Pallen, M. J. (2014). Recovery of a medieval *Brucella melitensis* genome using shotgun metagenomics. *mBio*, *5*(4), e01337-14. https://doi.org/10.1128/mBio.01337-14

Key, F. M., Posth, C., Krause, J., Herbig, A., & Bos, K. I. (2017). Mining metagenomic data sets for ancient DNA: Recommended protocols for authentication. *Trends in Genetics*, *33*(8), 508–520. https://doi.org/10.1016/j.tig.2017.05.005

Kobayashi, T. (2003). *Jomon reflections: Forager life and culture in the prehistoric Japanese archipelago* (S. Kaner & O. Nakamura, Eds.). Oxbow Books. https://doi.org/10.2307/j.ctv2p7j5rc

Krause-Kyora, B., Nutsua, M., Boehme, L., Pierini, F., Pedersen, D. D., Kornell, S.-C., Drichel, D., Bonazzi, M., Möbus, L., Tarp, P., Susat, J., Bosse, E., Willburger, B., Schmidt, A. H., Sauter, J., Franke, A., Wittig, M., Caliebe, A., Nothnagel, M., … Nebel, A. (2018). Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nature Communications*, *9*, 1569. https://doi.org/10.1038/s41467-018-03857-x

Kusaka, S., Uno, K. T., Nakano, T., Nakatsukasa, M., & Cerling, T. E. (2015). Carbon isotope ratios of human tooth enamel record the evidence of terrestrial resource consumption during the Jomon period, Japan. *American Journal of Physical Anthropology*, *158*(2), 300–311. https://doi.org/10.1002/ajpa.22775

Kuzmin, Y. V., & Keally, C. T. (2001). Radiocarbon chronology of the earliest Neolithic sites in East Asia. *Radiocarbon*, *43*(2B), 1121–1128. https://doi.org/10.1017/S0033822200041771

Laine, C. G., Johnson, V. E., Scott, H., & Arenas-Gamboa, A. M. (2023). Global estimate of human brucellosis incidence. *Emerging Infectious Diseases*, *29*(9), 1789–1797. https://doi.org/10.3201/eid2909.230052.

Larsen, C. S. (2010). Introduction. In C. S. Larsen (Ed.), *A companion to biological anthropology* (pp. 1–10). Wiley-Blackwell.

Leon-Sicairos, N., Reyes-Cortes, R., Guadrón-Llanos, A. M., Madueña-Molina, J., Leon-Sicairos, C., & Canizalez-Román, A. (2015). Strategies of intracellular pathogens for obtaining iron from the environment. *BioMed Research International*, *2015*, 1–17. https://doi.org/10.1155/2015/476534

Liang, R., Li, Z., Lau Vetter, M. C. Y., Vishnivetskaya, T. A., Zanina, O. G., Lloyd, K. G., Pfiffner, S. M., Rivkina, E. M., Wang, W., Wiggins, J., Miller, J., Hettich, R. L., & Onstott, T. C. (2021). Genomic reconstruction of fossil and living microorganisms in ancient Siberian permafrost. *Microbiome*, *9*(1), 110. https://doi.org/10.1186/s40168-021-01057-2

Loman, N. J., Constantinidou, C., Christner, M., Rohde, H., Chan, J. Z.-M., Quick, J., Weir, J. C., Quince, C., Smith, G. P., Betley, J. R., Aepfelbacher, M., & Pallen, M. J. (2013). A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA*, *309*(14), 1502. https://doi.org/10.1001/jama.2013.3231

Marciniak, S., & Poinar, H. N. (2019). Ancient pathogens through human history: A paleogenomic perspective. In C. Lindqvist & O. P. Rajora (Eds.), *Paleogenomics: Genome-scale analysis of ancient DNA* (pp. 115–138). Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-04753-5

Margaryan, A., Hansen, H. B., Rasmussen, S., Sikora, M., Moiseyev, V., Khoklov, A., Epimakhov, A., Yepiskoposyan, L., Kriiska, A., Varul, L., Saag, L., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2018). Ancient pathogen DNA in human teeth and petrous bones. *Ecology and Evolution*, *8*(6), 3534–3542. https://doi.org/10.1002/ece3.3924

Martin, F., & Uroz, S. (Eds.). (2016). *Microbial Environmental Genomics* (Vol. 1399). Springer New York. https://doi.org/10.1007/978-1-4939-3369-3

Matsui, A., & Kanehara, M. (2006). The question of prehistoric plant husbandry during the Jomon period in Japan. *World Archaeology*, *38*(2), 259–273.

Mazess, R. B. (1975). Biological adaptation: Aptitudes and acclimatization. In E. S. Watts, F. E. Johnston, & G. W. Lasker (Eds.), *Biosocial interrelations in population adaptation* (pp. 9–18). Mouton Publishers.

Minagawa, M., & Kondo, O. (2023). Taphonomic observation of Jomon human skeletal remains from a collective burial of the Gongenbara shell-mound, Chiba Prefecture, Japan. *Anthropological Science*, *131*(2), 89–105. https://doi.org/10.1537/ase.230223

Moreno, E. (2014). Retrospective and prospective perspectives on zoonotic brucellosis. *Frontiers in Microbiology*, *5*, 213. https://doi.org/10.3389/fmicb.2014.00213

Moreno, E., Blasco, J. M., Letesson, J. J., Gorvel, J. P., & Moriyón, I. (2022). Pathogenicity and its implications in taxonomy: The *Brucella* and *Ochrobactrum* case. *Pathogens*, *11*(3), 377. https://doi.org/10.3390/pathogens11030377

Moreno, E., Middlebrook, E. A., Altamirano-Silva, P., Al Dahouk, S., Araj, G. F., Arce-Gorvel, V., Arenas-Gamboa, Á., Ariza, J., Barquero-Calvo, E., Battelli, G., Bertu, W. J., Blasco, J. M., Bosilkovski, M., Cadmus, S., Caswell, C. C., Celli, J., Chacón-Díaz, C., Chaves-Olarte, E., Comerci, D. J., … Moriyón, I. (2023). If you're not confused, you're not paying attention: *Ochrobactrum* is not *Brucella*. *Journal of Clinical Microbiology*, *61*(8), e00438-23. https://doi.org/10.1128/jcm.00438-23

National Center for Biotechnology Information. (2006). *Entrez help*. National Library of Medicine: Bookshelf. Retrieved July 23, 2023, from https://www.ncbi.nlm.nih.gov/books/NBK3837/

Natsuki, D. (2022). Migration and adaptation of Jomon people during Pleistocene/Holocene transition period in Hokkaido, Japan. *Quaternary International*, *608–609*, 49–64. https://doi.org/10.1016/j.quaint.2021.01.009

Nishimura, L., Sugimoto, R., Inoue, J., Nakaoka, H., Kanzawa-Kiriyama, H., Shinoda, K., & Inoue, I. (2021). Identification of ancient viruses from metagenomic data of the Jomon people. *Journal of Human Genetics*, *66*(3), 287–296. https://doi.org/10.1038/s10038-020-00841-6

Njeru, J., Wareth, G., Melzer, F., Henning, K., Pletz, M. W., Heller, R., & Neubauer, H. (2016). Systematic review of brucellosis in Kenya: Disease frequency in humans and animals and risk factors for human infection. *BMC Public Health*, *16*(1), 853. https://doi.org/10.1186/s12889-016-3532-9

Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., Kambal, A., Monaco, C. L., Zhao, G., Fleshner, P., Stappenbeck, T. S., McGovern, D. P. B., Keshavarzian, A., Mutlu, E. A., Sauk, J., Gevers, D., Xavier, R. J., Wang, D., Parkes, M., & Virgin, H. W. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*, *160*(3), 447–460. https://doi.org/10.1016/j.cell.2015.01.002

Noshiro, S., & Sasaki, Y. (2014). Pre-agricultural management of plant resources during the Jomon period in Japan—A sophisticated subsistence system on plant resources. *Journal of Archaeological Science*, *42*, 93–106. https://doi.org/10.1016/j.jas.2013.11.001

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1), 132. https://doi.org/10.1186/s13059-016-0997-x

O'Rourke, D. H. (2010). Human molecular genetics: The DNA revolution and variation. In C. S. Larsen (Ed.), *A companion to biological anthropology* (pp. 88–103). Wiley-Blackwell.

Ortner, D. J., & Frohlich, B. (2007). The EB IA tombs and burials of Bâb edh-Dhrâ, Jordan: A bioarchaeological perspective on the people. *International Journal of Osteoarchaeology*, *17*(4), 358–368. https://doi.org/10.1002/oa.907

Oxenham, M., Matsumura, H., & Drake, A. (2013). Trauma and infectious disease in Northern Japan. In K. Pechenkina & M. Oxenham (Eds.), *Bioarchaeology of East Asia* (pp. 399–416). University Press of Florida. https://doi.org/10.5744/florida/9780813044279.003.0016

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*(1), 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Paff, S. (2021). Anthropology by data science. *Annals of Anthropological Practice*, *46*(1), 7–18. https://doi.org/10.1111/napa.12169

Paul, F., Otte, J., Schmitt, I., & Dal Grande, F. (2018). Comparing Sanger sequencing and high-throughput metabarcoding for inferring photobiont diversity in lichens. *Scientific Reports*, *8*, 8624. https://doi.org/10.1038/s41598-018-26947-8

Pearson, R. (2006). Jomon hot spot: Increasing sedentism in south-western Japan in the Incipient Jomon (14,000-9250 cal. BC) and Earliest Jomon (9250-5300 cal. BC) periods. *World Archaeology*, *38*(2), 239–258. https://doi.org/10.1080/00438240600693976

Pourbagher, A., Pourbagher, M. A., Savas, L., Turunc, T., Demiroglu, Y. Z., Erol, I., & Yalcintas, D. (2006). Epidemiologic, clinical, and imaging findings in brucellosis patients with osteoarticular involvement. *American Journal of Roentgenology*, *187*(4), 873–880. https://doi.org/10.2214/AJR.05.1088

Price, M. D., Makarewicz, C. A., & Chesson, M. S. (2018). Domestic animal production and consumption at Tall al-Handaquq South (Jordan) in the Early Bronze III. *Paléorient*, *44*(1), 75–91. https://doi.org/10.3406/paleo.2018.5786

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Qureshi, K. A., Parvez, A., Fahmy, N. A., Abdel Hady, B. H., Kumar, S., Ganguly, A., Atiya, A., Elhassan, G. O., Alfadly, S. O., Parkkila, S., & Aspatwar, A. (2023). Brucellosis: Epidemiology, pathogenesis, diagnosis and treatment–a comprehensive review. *Annals of Medicine*, *55*(2), 2295398. https://doi.org/10.1080/07853890.2023.2295398

Raoult, D., & Drancourt, M. (2008). Molecular detection of past pathogens. In D. Raoult & M. Drancourt (Eds.), *Paleomicrobiology: Past human infections* (pp. 55–68). Springer. https://doi.org/10.1007/978-3-540-75855-6_4

Riesenfeld, C. S., Schloss, P. D., & Handelsman, J. (2004). Metagenomics: Genomic analysis of microbial communities. *Annual Review of Genetics*, *38*(1), 525–552. https://doi.org/10.1146/annurev.genet.38.072902.091216

Rose, J. C. (2017). History of and recent trends in bioarcheological research in the Nile valley and the Levant. *Bioarchaeology of the Near East*, *11*, 7–28.

Rothschild, B., & Haeusler, M. (2021). Possible vertebral brucellosis infection in a Neanderthal. *Scientific Reports*, *11*(1), 19846. https://doi.org/10.1038/s41598-021-99289-7

Ryan, M. P., & Pembroke, J. T. (2020). The genus *Ochrobactrum* as major opportunistic pathogens. *Microorganisms*, *8*(11), 1797. https://doi.org/10.3390/microorganisms8111797

Schell, L. M. (1997). Culture as a stressor: A revised model of biocultural interaction. *American Journal of Physical Anthropology*, *102*(1), 67–77. https://doi.org/10.1002/(SICI)1096-8644(199701)102:1<67::AID-AJPA6>3.0.CO;2-A

Scholz, H. C., Al Dahouk, S., Tomaso, H., Neubauer, H., Witte, A., Schloter, M., Kämpfer, P., Falsen, E., Pfeffer, M., & Engel, M. (2008). Genetic diversity and phylogenetic relationships of bacteria belonging to the *Ochrobactrum–Brucella* group by recA and 16S rRNA gene-based comparative sequence analysis. *Systematic and Applied Microbiology*, *31*(1), 1–16. https://doi.org/10.1016/j.syapm.2007.10.004

Schug, G. R., Buikstra, J. E., DeWitte, S. N., Baker, B. J., Berger, E., Buzon, M. R., Davies-Barrett, A. M., Goldstein, L., Grauer, A. L., Gregoricka, L. A., Halcrow, S. E., Knudson, K. J., Larsen, C. S., Martin, D. L., Nystrom, K. C., Perry, M. A., Roberts, C. A., Santos, A. L., Stojanowski, C. M., … Zakrzewski, S. R. (2023). Climate change, human health, and resilience in the Holocene. *Proceedings of the National Academy of Sciences*, *120*(4), e2209472120. https://doi.org/10.1073/pnas.2209472120

Sereno, D., Dorkeld, F., Akhoundi, M., & Perrin, P. (2018). Pathogen species identification from metagenomes in ancient remains: The challenge of identifying human pathogenic species of Trypanosomatidae via bioinformatic tools. *Genes*, *9*(8), 418. https://doi.org/10.3390/genes9080418

Shakir, R. (2021). Brucellosis. *Journal of the Neurological Sciences*, *420*, 117280. https://doi.org/10.1016/j.jns.2020.117280

Shephard, R. J., & Rode, A. (1996). *The health consequences of "modernization": Evidence from circumpolar peoples*. Cambridge University Press.

Shibutani, A. (2009). Late Pleistocene to early Holocene plant movements in Southern Kyushu, Japan. *Archaeologies*, *5*(1), 124–133. https://doi.org/10.1007/s11759-009-9094-z

Shimoda, M., Tanaka, Y., Kokutou, H., Furuuchi, K., Osawa, T., Morimoto, K., Yano, R., Yoshimori, K., & Ohta, K. (2021). *Actinomyces meyeri* pleural infection that was difficult to treat due to delayed culture: A case report and literature review of 28 cases. *Respiratory Medicine Case Reports*, *34*, 101530. https://doi.org/10.1016/j.rmcr.2021.101530

Sokolov, A. S., Nedoluzhko, A. V., Boulygina, E. S., Tsygankova, S. V., Sharko, F. S., Gruzdeva, N. M., Shishlov, A. V., Kolpakova, A. V., Rezepkin, A. D., Skryabin, K. G., & Prokhortchouk, E. B. (2016). Six complete mitochondrial genomes from Early Bronze Age humans in the North Caucasus. *Journal of Archaeological Science*, *73*, 138–144. https://doi.org/10.1016/j.jas.2016.07.017

Squires, K., Booth, T., & Roberts, C. A. (2019). The ethics of sampling human skeletal remains for destructive analyses. In K. Squires, D. Errickson, & N. Márquez-Grant (Eds.), *Ethical approaches to human remains: A global challenge in bioarchaeology and forensic anthropology* (pp. 265–297). Springer International Publishing. https://doi.org/10.1007/978-3-030-32926-6

Squires, K., & García-Mancuso, R. (2021). Ethical challenges associated with the study and treatment of human remains in anthropological sciences in the 21st century. *Revista argentina de antropología biológica*, *23*(2). https://doi.org/ 10.24215/18536387e034

Sulayman, S. M. A., Bora, R. S., Sabir, J. S. M., & Ahmed, M. M. M. (2020). Brucellosis: Current status of the disease and future perspectives. *Advancements of Microbiology*, *59*(4), 337–344. https://doi.org/10.21307/PM-2020.59.4.25

Suzuki, H. (1958). A prehistoric human ilium penetrated by an arrow head. *Journal of the Anthropological Society of Nippon*, *66*(3), 112–115. https://doi.org/10.1537/ase1911.66.112

Tan, K.-K., Tan, Y.-C., Chang, L.-Y., Lee, K. W., Nore, S. S., Yee, W.-Y., Mat Isa, M. N., Jafar, F. L., Hoh, C.-C., & AbuBakar, S. (2015). Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics*, *16*(1), 93. https://doi.org/10.1186/s12864-015-1294-x

Temple, D. H. (2007). Dietary variation and stress among prehistoric Jomon foragers from Japan. *American Journal of Physical Anthropology*, *133*(4), 1035–1046. https://doi.org/10.1002/ajpa.20645

Temple, D. H. (2014). Plasticity and constraint in response to early-life stressors among Late/Final Jomon period foragers from Japan: Evidence for life history trade-offs from incremental microstructures of enamel. *American Journal of Physical Anthropology*, *155*(4), 537–545. https://doi.org/10.1002/ajpa.22606

Trêpa, J., Mendes, P., Gonçalves, R., Chaves, C., Brás, A. M., Mesa, A., Ramos, I., Sá, R., & da Cunha, J. G. S. (2018). *Brucella* vertebral osteomyelitis misidentified as an *Ochrobactrum anthropi* infection. *IDCases*, *11*, 74–76. https://doi.org/10.1016/j.idcr.2018.01.010

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., & Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, *428*(6978), 37–43. https://doi.org/10.1038/nature02340

Ulijaszek, S. J., & Lofink, H. (2006). Obesity in biocultural perspective. *Annual Review of Anthropology*, *35*(1), 337–60. https://doi.org/10.1146/annurev.anthro.35.081705.123301

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., … Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, *304*(5667), 66–74. https://doi.org/10.1126/science.1093857

Waldron, T. (2009). *Palaeopathology*. Cambridge University Press.

Walker, P. L. (2008). Bioarchaeological ethics: A historical perspective on the value of human remains. In M. A. Katzenberg & S. R. Saunders (Eds.), *Biological anthropology of the human skeleton* (pp. 3–40). Wiley.

Wang, J. T. L., Zaki, M. J., Toivonen, H. T. T., & Shasha, D. (2005). Introduction to data mining in bioinformatics. In X. Wu, L. Jain, J. T. L. Wang, M. J. Zaki, H. T. T. Toivonen, & D. Shasha (Eds.), *Data mining in bioinformatics* (pp. 3–8). Springer. https://doi.org/10.1007/1-84628-059-1_1

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., Orlando, L., & Krause, J. (2017). A robust framework for microbial archaeology. *Annual Review of Genomics and Human Genetics*, *18*(1), 321–356. https://doi.org/10.1146/annurev-genom-091416-035526

Weiß, C. L., Gansauge, M.-T., Aximu-Petri, A., Meyer, M., & Burbano, H. A. (2020). Mining ancient microbiomes using selective enrichment of damaged DNA molecules. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-020-06820-7

Whatmore, A. M. (2014). Ancient-pathogen genomics: Coming of age? *mBio*, *5*(5), e01676-14. https://doi.org/10.1128/mBio.01676-14

Wheeler, D., & Bhagwat, M. (2007). BLAST QuickStart: Example-driven web-based BLAST tutorial. *Methods in Molecular Biology*, *395*, 149–176.

Wilbur, A. K., Bouwman, A. S., Stone, A. C., Roberts, C. A., Pfister, L.-A., Buikstra, J. E., & Brown, T. A. (2009). Deficiencies and challenges in the study of ancient tuberculosis DNA. *Journal of Archaeological Science*, *36*(9), 1990–1997. https://doi.org/10.1016/j.jas.2009.05.020

Wiley, A. S. (2020). Continuity and change in biocultural anthropology. *American Journal of Human Biology*, *32*(4), e23464. https://doi.org/10.1002/ajhb.23464

Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology*, *6*(2), e1000667. https://doi.org/10.1371/journal.pcbi.1000667

Yagupsky, P., Morata, P., & Colmenero, J. D. (2019). Laboratory diagnosis of human brucellosis. *Clinical Microbiology Reviews*, *33*(1), e00073-19. https://doi.org/10.1128/CMR.00073-19

Zielezinski, A., Vinga, S., Almeida, J., & Karlowski, W. M. (2017). Alignment-free sequence comparison: Benefits, applications, and tools. *Genome Biology*, *18*(1), 186. https://doi.org/10.1186/s13059-017-1319-7

Zink, A. R., Reischl, U., Wolf, H., & Nerlich, A. G. (2002). Molecular analysis of ancient microbial infections. *FEMS Microbiology Letters*, *213*(2), 141–147. https://doi.org/10.1111/j.1574-6968.2002.tb11298.x

Zuckerman, M. K, & Armelagos, G. J. (2011). The origins of biocultural dimensions in bioarchaeology. In S. C. Agarwal & B. A. Glencross (Eds.). *Social bioarchaeology* (pp. 15–43). Wiley-Blackwell.

Appendix

Table 10: *Maximum BLASTn scores for* Brucella melitensis *for all sequencing runs from PRJDB4223*

Accession	Sample ID	Max score	Total score	Query coverage (%)	*E* value	Percent identity
DRR046398	Sanganji 131421-3 (A1)	360.65	385.94	26.66	7e-25	86.12
DRR046399	Sanganji 131421-3 (A1)	199.05	223.71	35.18	4e-23	91.13
DRR046400	Sanganji 131421-3 (A1)	361.11	381.72	21.43	7e-25	92.5
DRR046401	Sanganji 131421-3 (A1)	133.45	174.36	18.98	1e-29	87.16
DRR046402	Sanganji 131421-3 (A1)	214	249.19	28.7	1e-20	81.8
DRR046403	Sanganji 131421-3 (A1)	9.92	16.52	1.18	2e-15	55.61
DRR046404	Sanganji 131421-3 (A1)	11.56	19.85	0.55	7e-12	52.19
DRR046405	Sanganji 131421-3 (A1)	10.01	20.33	1.37	6e-25	46.75
DRR046406	Sanganji 131421-3 (A1)	32.1	49.07	1.98	4e-16	28.51
DRR046407	Sanganji 131421-3 (A1)	14.7	38.66	0.94	3e-24	21.08
DRR046408	Sanganji 131421-3 (A2)	19.45	43.8	1.01	4e-22	58
DRR046409	Sanganji 131421-3 (A2)	25.83	51.12	4.67	3e-12	38.17
DRR046410	Sanganji 131421-3 (A2)	14.44	38.45	2.07	3e-17	29.64
DRR046411	Sanganji 131421-3 (A2)	18.54	49.71	3.86	2e-12	33.24
DRR046412	Sanganji 131464 (B)	356.5	395.11	36.19	4e-23	93.37
DRR046413	Sanganji 131464 (B)	318.67	346.53	40.65	3e-20	71.2
DRR046414	Sanganji 131464 (B)	34.19	53.17	6.39	1e-24	31.53
DRR046415	Sanganji 131464 (B)	44.63	65.9	15.87	1e-18	49.41

Table 11: *Maximum BLASTn scores for* Brucella ovis *for all sequencing runs from PRJDB4223*

Accession	Sample ID	Max score	Total score	Query coverage (%)	*E* value	Percent identity
DRR046398	Sanganji 131421-3 (A1)	297.01	334.28	26.78	7e-25	72.15
DRR046399	Sanganji 131421-3 (A1)	200.15	240.54	33.45	6e-23	85.38
DRR046400	Sanganji 131421-3 (A1)	361.98	388.11	21.7	5e-26	93.46
DRR046401	Sanganji 131421-3 (A1)	125.55	132.17	26.48	2e-24	90.06
DRR046402	Sanganji 131421-3 (A1)	261.12	312.59	30.11	1e-20	89.54
DRR046403	Sanganji 131421-3 (A1)	9.84	16.78	1.78	2e-15	51.5
DRR046404	Sanganji 131421-3 (A1)	13.56	29.57	1.45	6e-17	48.97
DRR046405	Sanganji 131421-3 (A1)	17.06	39.49	3.39	6e-23	45.77
DRR046406	Sanganji 131421-3 (A1)	36.84	66.3	3.44	4e-15	35.98
DRR046407	Sanganji 131421-3 (A1)	16.27	34.53	2.96	3e-24	20
DRR046408	Sanganji 131421-3 (A2)	19.01	35.62	1.2	3e-17	43.19
DRR046409	Sanganji 131421-3 (A2)	33.26	52.28	5.28	3e-10	56.73

DRR046410	Sanganji 131421-3 (A2)	16.57	43.48	3.45	3e-23	23.64
DRR046411	Sanganji 131421-3 (A2)	15.64	57.69	4.03	1e-21	32
DRR046412	Sanganji 131464 (B)	398.88	455.33	39.55	3e-27	92.42
DRR046413	Sanganji 131464 (B)	315.9	391.81	39.32	3e-21	75.97
DRR046414	Sanganji 131464 (B)	38.12	50.19	4.89	1e-24	23.13
DRR046415	Sanganji 131464 (B)	43.01	60.51	15.13	2e-19	38.92

Table 12: *Maximum BLASTn scores for* Brucella abortus *for all sequencing runs from PRJDB4223*

Accession	**Sample ID**	**Max score**	**Total score**	**Query coverage (%)**	***E* value**	**Percent identity**
DRR046398	Sanganji 131421-3 (A1)	312.6	330.05	33.15	6e-27	84.12
DRR046399	Sanganji 131421-3 (A1)	203.06	264.19	43.58	5e-27	93.74
DRR046400	Sanganji 131421-3 (A1)	363.8	383.7	27.61	5e-27	93.26
DRR046401	Sanganji 131421-3 (A1)	154.81	188.51	36.99	2e-24	84.14
DRR046402	Sanganji 131421-3 (A1)	274.3	327.77	38.87	2e-23	91.78
DRR046403	Sanganji 131421-3 (A1)	10.46	17.89	3.68	3e-16	51.25
DRR046404	Sanganji 131421-3 (A1)	18.42	36.19	2.95	4e-20	40.66
DRR046405	Sanganji 131421-3 (A1)	20.34	44.88	1.13	6e-27	43.21
DRR046406	Sanganji 131421-3 (A1)	37.9	62.13	0.88	3e-23	39.34
DRR046407	Sanganji 131421-3 (A1)	18.75	39.79	0.96	3e-26	18.87
DRR046408	Sanganji 131421-3 (A2)	21.32	47.55	0.67	4e-24	65.2
DRR046409	Sanganji 131421-3 (A2)	36.85	56.9	6.54	3e-18	43.09
DRR046410	Sanganji 131421-3 (A2)	14.74	37.15	3.41	4e-25	28.46
DRR046411	Sanganji 131421-3 (A2)	18.8	49.23	5.43	1e-21	31.87
DRR046412	Sanganji 131464 (B)	387.39	444.83	40.01	1e-24	92.87
DRR046413	Sanganji 131464 (B)	325.88	387.97	42.48	3e-22	73.65
DRR046414	Sanganji 131464 (B)	43.17	60.74	9.91	1e-22	30.08
DRR046415	Sanganji 131464 (B)	44.09	68.38	16.18	3e-16	22.6