Phylogenomics and Zoonotic Spillover of *Mycobacterium leprae* in the Pacific Islands and Brazil

by

Adele E. Crane

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

Approved June 2024 by the
Graduate Supervisory Committee:

Anne Stone, Chair
Keolu Fox
Arvind Varsani
Melissa Wilson

ARIZONA STATE UNIVERSITY

August 2024

ABSTRACT

Leprosy, or Hansen's disease, is often relegated to antiquity, yet it remains a modern public health concern with around 200,000 new cases reported annually around the world (World Health Organization, 2023). Most leprosy cases in humans are caused by *Mycobacterium leprae*, but a small number of cases are now known to be caused by *Mycobacterium lepromatosis*, which is found mostly in Mexico and the Caribbean (Han et al., 2008; 2012). Recent work has improved our understanding of the pattern of genomic variation in *M. leprae* strains, however certain regions remain understudied. Additionally, ongoing surveillance has identified pockets of hyperendemicity, which could act as hotspots for the spread of drug-resistant strains. Despite increased surveillance, there has been limited genomic research in regions like the Pacific and few surveys incorporating a broad species range in endemic areas. To address this, *M. leprae* genomes were isolated from clinical formalin-fixed, paraffin-embedded (FFPE) samples from the Pacific Islands and used in phylogenetic and phylogeographic analyses. 21 novel *M. leprae* strains from these samples were sequenced and dating analyses determined that *M. leprae* strains have been circulating in the Pacific since the original peopling of the region. Furthermore, recent radiation events among the Pacific Islands, as well as a rise in drug-resistant *M. leprae* strains, were identified and described. Additionally, a survey of animals in an endemic state in Brazil resulted in the first *M. leprae* genome isolated from a big cat. This body of research leverages genomic data to characterize novel diversity of *M. leprae* strains, identify drug-resistant strains in these regions, and determine how this pathogen spreads through space and time. The results of

this work will aid in our understanding of the history of leprosy and improve public health responses to this disease.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my friends and family who have acted as a constant support throughout this process. First and foremost, my partner Bobby, who has supported me from day one. We have been on many adventures together and with our dog Ella, which have made this journey worth it. I must also attribute my work to my parents, Ned and Virginia, who taught me about the natural world and how to always be curious. My siblings – Judy, Susan, Harry, and Andrew – and my amazing niblings have provided joy and support every step of the way. I would also like to thank my extended family, including my Mère, my aunts, uncles, and cousins, who have always sent me love and encouragement from afar. Finally, a special thanks to Granddad and Grandma Cox, who welcomed me to Arizona and the natural wonders this state contains.

I must thank my friends in Phoenix, as well as those from Grinnell College, Arizona State University, the California coast, and beyond. This work is not done overnight or in a social vacuum. Maintaining support inside and outside of the university helped me survive several lab moves and navigating a degree during a global pandemic. In addition to my friends, I extend my sincere gratitude to my former colleagues at the Smithsonian Institution. Without their mentorship and friendship, I would not have pursued this degree.

I would also like to thank my committee, made up of Anne Stone, Keolu Fox, Arvind Varsani, and Melissa Wilson. They provided excellent direction, advice, collaboration, and patience in support of my research. I must thank my current and former collaborators, especially Christopher Lum, Kanako Furuta, Kelly Blevins, and

Luciano Nakazato, who have been essential to this work. I am deeply grateful as well for the mentorship from former Stone Lab graduate students and postdoctoral researchers.

Finally, I must acknowledge the location and cultural significance of my research. As a student of Arizona State University, I acknowledge that the Tempe campus of the university sits on the ancestral homelands of indigenous peoples who have inhabited this place for centuries, including the Akimel O'odham (known as Pima) and Piipaash (known as Maricopa) peoples and their ancestors. Furthermore, this research was completed on a disease deeply connected to stigma and trauma, which has a long history in the Pacific Islands. I recognize that the samples used in my work come from the ancestral homelands of indigenous peoples who have inhabited these islands and larger territory. This includes research done through the support of institutions and peoples who inhabit the ancestral and current homelands of Indigenous Hawaiians, the CHamoru (Chamorro) people and their descendants, the Samoan people, the people of Belau (Palau), the Marshallese, and many more

Finally, I wish to dedicate this dissertation to the late Jonathan "Jackie" Brown. He taught me that to be a biologist is to also engage in history, culture, and philosophy, all of which shape - and are shaped by - modern biological thought.

TABLE OF CONTENTS

APPENDIX Page

APPENDIX Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figures Page

CHAPTER 1
INTRODUCTION

1.1 Background

Leprosy, also known as Hansen's disease, is an infectious disease caused by the obligate, intracellular pathogens *Mycobacterium leprae* and the more recently discovered *Mycobacterium lepromatosis* (Han et al., 2012; 2008). Although *M. leprae* strains are widely dispersed, *M. lepromatosis* has been identified mostly in the Caribbean and Mexico, with isolated cases reported in Singapore and Canada (Vera-Cabrera et al., 2011; Jessamine et al., 2012). Leprosy is a neglected tropical disease, so defined by an outsized impact on impoverished countries, its presence in the tropics, the potential for stigma or lack of access to treatment, and less research or money going towards eradication (Jayashankar et al., 2024). There are around 200,000 new cases of leprosy reported each year, with the highest case numbers in India, Brazil, and Indonesia (WHO, 2023). Leprosy damages peripheral nerves and, if left untreated, can cause blindness, limb deformities, and death (WHO, 2023).

Although the mechanism of leprosy transmission is not fully understood, the most likely route is through droplets during coughing or sneezing, as well as via skin-to-skin contact (Avanzi et al., 2020c). The clinical presentation of the disease is described by the Ridley-Jopling classification system, which depends on the individual's cell-mediated immune response and spectrum of symptoms (Ridley & Jopling, 1966). The classification system ranges from tuberculoid, which consists of a small number of lesions and a low bacillary load, to lepromatous, which involves numerous infiltrated lesions, more significant nerve damage, and a higher bacillary load (Britton, 1993). Borderline cases

exist on a spectrum between tuberculoid and lepromatous leprosy. The WHO has also recommended an additional classification system based on bacillary load, from paucibacillary to multibacillary (WHO, 2017). The paucibacillary and multibacillary classifications correlate with the two ends of the Ridley-Jopling system, from tuberculoid to lepromatous.

1.2 Molecular Biology and Genetics of Causative Agents

M. leprae is a slow-growing, Gram-positive, and acid-fast bacillus, which is around 1-8 μm long and has a diameter of 0.3 μm (Draper 1983). It is nonmotile and non-spore-forming, with straight or slightly curved rods (Scollard et al., 2006). A major lipid in the cell wall is phenolic glycolipid-1 (PGL-1), which is important for serological diagnosis via the detection of IgM antibodies in humans (Scollard et al., 2006; Stefani, 2008). Furthermore, *M. leprae* displays strong tropism towards Schwann cells and macrophages and grows very slowly, with a doubling time of 14 days. (Scollard et al., 2006). Although *M. leprae* cannot be cultivated in traditional media, it can be grown in mouse-footpads and nine-banded armadillos (*Dasypus novemcinctus*) (Stefani, 2008). Microbiological analysis of *M. lepromatosis* has not been performed to the same extent.

Genomic Diversity and Categorization

The genome of *M. leprae* is compact, at ~3.27 Mb, and has undergone significant reductive evolution with nearly half of the genome's length made up of pseudogenes (Cole et al., 2001). The genomic diversity among *M. leprae* strains is relatively low—there are around 300 single-nucleotide polymorphisms (SNPs) differentiating among geographically distant strains (Monot et al., 2009; Schuenemann et al., 2018). *M. leprae*

genomes are grouped by shared informative SNPs and variable number tandem repeats (VNTRs) into four SNP types and 16 SNP subtypes (1 – 4 and A – P, respectively) (Monot et al., 2005; 2009). More recently, phylogenies generated from whole genome data have been used to define six branches (branches 0 – 5), although branch 1 and the paraphyletic branch 2 belong to the same clade (Benjak et al., 2018; Schuenemann et al., 2013; 2018). The branches and SNP types have strong geographic associations, with branch 0 primarily found in Eastern Asia and the Pacific, branch 1 largely in Southern and Eastern Asia, branch 2 in the Near East and South Asia, branch 3 in the Americas, branch 4 in West Africa and South America, and branch 5 in the Pacific Islands (Benjak et al., 2018; Monot et al., 2009; Schuenemann et al., 2018).

M. leprae and *M. lepromatosis* diverged an estimated 13.9 Mya, which was likely after a significant amount of reductive evolution in the ancestral lineage had occurred (Han et al., 2012; Singh et al., 2015). The *M. lepromatosis* genome is small (~3.3 Mb) and shares 93% nucleotide sequence identity in coding regions and 82% pseudogene nucleotide sequence identity with *M. leprae* (Singh et al., 2015). Between 20-40% of leprosy cases reported in Mexico are due to *M. lepromatosis* and, of those, cases involving coinfection with both causative agents have been reported (Avanzi et al., 2020c; Singh et al., 2015).

Intrahost Diversity

The study of intrahost diversity in *M. leprae* reveals genomic variation within individual hosts, however most research is limited to comparing SNP and VNTR differences between different sample types from the same individual (Lima et al., 2016;

Tio-Coma et al., 2020a). Longitudinal studies on leprosy are challenging, because there may not be enough variation for differentiation when only SNP types are analyzed over time. Additionally, minimally invasive sampling methods like slit-skin smears and nasal swabs are easier to collect but contain less bacterial DNA compared to tissue biopsies, which provide more genomic material (Avanzi et al. 2020c). Despite these barriers to intrahost studies, research comparing SNP and VNTR types across different sample types has shown some differences, suggesting the presence of different strains within one host (Lima et al., 2016; Tio-Coma et al., 2020a). Furthermore, Avanzi et al. (2020c) suggest that nasal swabs may capture passive carriage of *M. leprae* from individuals in close contact with leprosy patients. Araujo et al. (2016) observed that while nasal carriage of *M. leprae* was not predictive of disease onset, seropositivity and blood presence increased the risk of leprosy among individuals living with leprosy patients, highlighting the need for further exploration of intrahost dynamics and disease risk analysis. These findings underscore the complexity of leprosy infections and the hidden genetic diversity within hosts, posing challenges for traditional phylogenetic studies and the development of effective treatment strategies. Understanding these dynamics is crucial for managing drug resistance and improving leprosy control efforts.

1.3 Epidemiology and Spillover of *M. leprae*

M. leprae is difficult to study due to the pathogen's unculturable nature and low genetic diversity, particularly when focusing solely on SNP/VNTR typing (Hall & Salipante, 2010). There is a growing body of work devoted to characterizing the molecular epidemiology of *M. leprae*, with the goal of understanding transmission routes

and predicting community disease risk (Avanzi et al. 2020c; Hall & Salipante, 2010). There is clearly a higher risk for leprosy among household contacts and those in close proximity to active cases (Romero-Montoya et al., 2017; Araujo et al., 2016), though multiple strains may coexist within households or communities in hyperendemic areas (Avanzi et al., 2016a). In addition to more abiotic factors, such as proximity, genetic susceptibility also plays a large role in disease risk (Das et al., 2020). Transmission risk is further complicated by the long incubation period of leprosy, which lasts around 3-10 years, although it can last up to 30 years (Avanzi et al., 2020c).

In addition to humans, *M. leprae* has also been detected in naturally infected nine-banded armadillos (*D. novemcinctus*) (Sharma et al., 2015; Truman et al., 2011), six-banded armadillos (*Ephractus sexcinctus*) (Frota et al., 2012), British red squirrels (*Sciurus vulgaris*) (Avanzi et al., 2016b), captive but wild-born non-human primates, including *Pan troglodytes*, *Macaca fasicularis*, and *Cercocebus atys* (Honap et al., 2018), and wild populations of western chimpanzees (*Pan troglodytes verus*) (Hockings et al., 2021). *M. lepromatosis* has also been isolated and characterized from British red squirrels (Avanzi et al., 2016b), but has yet to be isolated from other non-human hosts. Modern animal strains for *M. leprae* largely correspond to the SNP types of human strains circulating in the same region (Honap et al., 2018; Ploemacher et al., 2020; Sharma et al., 2015).

Spillover of leprosy to and from animals is an ongoing public health concern that is listed as a barrier to eradication in the WHO Global Hansen's Disease Strategy 2021-2030 (WHO, 2023). An estimated two thirds of autochthonous cases of leprosy in

humans in the United States are armadillo-associated strains (Sharma et al., 2015). Brazil has also seen direct animal transmission from armadillos, with da Silva et al. (2018) identifying spillover in communities that hunt and consume the animals. Although similar direct transmission has not been seen between other animals and humans, ongoing surveillance in animal populations in endemic regions of leprosy will further clarify the role natural animal reservoirs play in disease prevalence (Ploemacher et al., 2020).

A wide array of alternative vectors for *M. leprae* has been investigated, including in free-living amoeba (Wheat et al., 2014), kissing bugs (da Silva Neumann et al., 2016), and ticks (Ferreira et al., 2018). Although *M. leprae* can remain viable in these organisms for a short length of time, there is little evidence to support transmission via these vectors. Viable *M. leprae* has also been identified in soil samples outside of households with active infections (Lavania et al., 2008, Turankar et al., 2016; 2022). Additionally, *M. leprae* may persist in water sources near communities with active infections (Mohanty et al., 2016). As with the potential vectors, more work is needed to interrogate the risk of transmission and length of *M. leprae* viability in the environment.

1.4 Leprosy in History and Modern Challenges

Identification of leprosy in skeletal remains can be difficult as distinguishing characteristics of the disease, such as truncation of the phalanges and metacarpals or changes in the rhinomaxillary region, may also be caused by syphilis or various types of joint diseases (Donoghue et al., 2015). Despite this difficulty, well-supported osteological evidence for leprosy has been identified as early as 2000 BCE at the Balathal site in India, with textual references in the region beginning in 600 BCE (Robbins et al., 2009).

Other skeletal evidence has been found extensively in Europe, including some of the earliest cases dating to 400 BCE in modern day Italy (Mariotti et al., 2005) and around 400 CE in UK sites (Donoghue et al., 2015). Beyond Europe and the Near East, skeletal remains with signs of leprosy have been identified in Southeast Asia, which date to around 300 to 200 BCE (Tayles & Buckley, 2004) and later in Japan between 1400 to 1700 CE (Suzuki et al., 2014). Potential evidence for leprosy in the Pacific Islands has been found in the Mariana Islands, dating to around 800 CE, however these remains only exhibit the destruction of the metatarsals, metacarpals, and phalanges (Trembly, 1995).

The oldest sequenced *M. leprae* genome dates to the 2nd century BCE and was isolated from an Egyptian mummy excavated from Abusir el-Meleq, located in Middle Egypt (Neukamm et al., 2020). Molecular support has also been identified in remains from the 1st century CE in Jerusalem (Matheson et al., 2009), and more extensive genome sequencing has focused on remains from 850 CE on in Europe (Schuenemann et al., 2013; 2018). Schuenemann et al. (2018) proposed two models to explain the high genetic diversity observed in medieval Europe: either *M. leprae* originated in Western Eurasia, or diverse strains with different geographical origins were introduced to Europe before the Medieval period. There is a bias towards European strains in our current understanding of ancient *M. leprae* genomic diversity. Dating methods have revealed that the most recent common ancestor of all *M. leprae* strains was circulating between 2000 and 4000 BCE (Hockings et al., 2021; Schuenemann et al., 2018). Branch 0 is the most basal clade and branch 0 strains are now found in the Pacific Islands, as well as in skeletal remains from medieval Denmark (Blevins et al., 2020; Schuenemann et al., 2018).

Effective treatment for leprosy began in the 1940s with dapsone, and the implementation of multi-drug therapy (MDT) by the WHO in 1982 significantly reduced the global disease burden (Smith et al., 2017). Despite this, certain countries are still endemic for the disease, with most new cases occurring in India, Brazil, and Indonesia (WHO, 2019). Further complicating eradication efforts is the rise of drug resistance in *M. leprae* strains. Key genes implicated in resistance to primary drugs in MDT (rifampicin, dapsone, and ofloxacin) include *rpoB, folp1*, and *gyrA*/*gyrB* (Williams and Gillis, 2012; Benjak et al., 2018). Cambau et al. (2018) reported that up to 5% of relapsing leprosy cases worldwide could be linked to emerging drug resistance. Most surveillance is ongoing in hyperendemic regions, however drug-resistant strains have been identified in non-endemic regions and require further study (Blevins et al., 2020; Cambau et al., 2018; Williams and Gillis, 2012).

1.5 Research Scope

Since a significant focus of leprosy in history has been on Europe and the Near East (*e.g.*, Donoghue et al., 2015 and Schuenemann et al., 2018), there are large gaps in our understanding of the origin and modern genetic diversity of *M. leprae* in other regions. Specifically, the Pacific Islands have very poor representation in the *M. leprae* phylogeny and in paleopathological studies. Chapter 2 of this dissertation aims to determine the timing of the origin of *M. leprae* into the Pacific, as well as characterize modern strains circulating among the islands. The former question of origin is of particular interest as historians trace the presence of leprosy in the Pacific to the 1800s, due to the appearance of the disease in contemporary public health records and

administrative accounts (Inglis, 2014). Dating analyses of the most basal branches of the *M. leprae* phylogeny show that branch 0 - a branch largely represented by strains in the Pacific - diverged much earlier at around 2000-3000 BP (Schuenemann et al., 2018). Thus, Chapter 2 addresses this discrepancy by performing phylogenetic and dating analyses on nine novel strains isolated from the Pacific Islands.

Chapter 3 of this dissertation furthers the goals of Chapter 2 to both sequence even more *M. leprae* strains circulating in the Pacific, as well as to characterize the spatial dynamics and timing of transmission among the islands. These aims are explored through leveraging novel *M. leprae* genomes isolated from the Pacific Islands and phylogeographic approaches to determine how specific strains may have spread. Both chapters not only enhance our understanding of the modern genomic diversity of *M. leprae* but are also the first to characterize the spread of leprosy in the Pacific Islands using phylogenomic techniques.

Chapter 4 of this dissertation analyzes the spillover risk of animal-associated strains of *M. leprae* in Brazil, which has the second highest incidence of new cases reported each year (WHO, 2023). Although ongoing surveillance has identified *M. leprae* strains and spillover in local armadillos (da Silva et al., 2018; Frota et al., 2012), a broader range of species has not been analyzed beyond PCR-based analyses (Maruyama et al., 2018). Chapter 4 describes the surveillance of mammals collected from the Amazonian side of Peru and the state of Mato Grosso in Brazil. A low-coverage genome of *M. leprae* was successfully isolated from a puma (*P. concolor*), which raises intriguing questions about the potential host range for *M. leprae* and hidden genomic diversity present in animal populations.

Ultimately, this dissertation aims to address unknown genetic diversity of *M. leprae* by expanding our knowledge of strains circulating in specific regions or species. Phylogenetic, phylogeographic, and Bayesian approaches are used to quantify this genetic diversity, as well as to infer the origin and spread of leprosy. The results of this dissertation have implications for the global health response to leprosy, including an increased need for surveillance of natural reservoirs and drug resistance.

CHAPTER 2

EVOLUTIONARY HISTORY OF *MYCOBACTERIUM LEPRAE* IN THE PACIFIC ISLANDS

2.1 Abstract

As one of the oldest known human diseases, leprosy, or Hansen's disease, remains a public health concern around the world with over 200,000 new cases in 2018. Most human leprosy cases are caused by *Mycobacterium leprae*, but a small number of cases are now known to be caused by *M. lepromatosis*, a sister taxon of *M. leprae*. The global pattern of genomic variation in *M. leprae* is not well defined. Particularly in the Pacific Islands, the origins of leprosy are disputed. Historically, it has been argued that leprosy arrived on the islands during 19th century colonialism, but some oral traditions and paleopathological evidence suggest an older introduction. To address this, as well as investigate patterns of pathogen exchange across the Pacific Islands, we extracted DNA from 39 formalin-fixed paraffin embedded biopsy blocks dating to 1992-2016. Using whole-genome enrichment and Next-Generation Sequencing, we produced nine *M. leprae* genomes dating to 1998-2015 and ranging from 4-63x depth of coverage. Phylogenetic analyses indicate that these strains belong to basal lineages within the *M. leprae* phylogeny, specifically falling in branches 0 and 5. The phylogeographic patterning and evolutionary dating analysis of these strains support a pre-modern introduction of *M. leprae* into the Pacific Islands.

2.2 Introduction

Leprosy, or Hansen's disease, is a chronic and progressive infectious disease caused by the obligate intracellular pathogen *Mycobacterium leprae* and the more recently discovered *Mycobacterium lepromatosis* (Han et al., 2008; 2012). The *M. leprae* genome is compact and demonstrates a low level of diversity, with less than 300 SNPs segregating geographically disparate strains (Schuenemann et al., 2018; Benjak et al., 2018; Cole et al., 2001). In initial comparative genetic studies, Monot et al. (2005) compared the reference TN (India) genome sequence to 142 kb of sequence from a Brazilian strain (Br4923) and found five SNPs; of these, three SNPs were used to identify and define *M. leprae* types (SNP types 1-4) in a broader range of strains. Subsequent comparative analyses resulted in the identification of additional SNP subtypes (A-P) (Monot et al., 2009) that have been used for broader genotyping studies (*e.g*., Lavania et al., 2015; Reibel et al., 2015; Singh et al., 2011).

More recently, phylogenies generated from genome data were used to define six branches (denoted by 0-5) (Benjak et al., 2018; Schuenemann et al., 2013; 2018). As with the SNP subtypes, the branches have geographic associations, with branch 0 found mainly in Eastern Asia, branch 1 found mainly in Southern and Eastern Asia, the paraphyletic branch 2 found in the Near East and South Asia, branch 3 primarily found in North and Latin America, branch 4 found mainly in West Africa and South America, and branch 5 found in the Pacific Islands (Schuenemann et al., 2018; Monot et al., 2009). Monot et al. (Monot et al., 2009) suggest that the geographic patterning of strains reflects human migration patterns and trade along early and more modern routes, including the Silk Road. Despite close geographic associations of branches, new genomic data, and

recent work adding to the understanding of ancient genetic diversity (Schuenemann et al., 2013; 2018), the origin and dispersal of *M. leprae* is still not well understood.

Leprosy was endemic in many Western Pacific and Pacific Island countries prior to 2010 (World Health Organization, 2008; 2011). Today, the success of public health campaigns and effectiveness of multidrug therapy are evident; now most of the Pacific Islands have prevalence rates below the elimination threshold (WHO, 2016; WHO, 2019). The antiquity of the disease in the region, however, is not well documented by historical or paleopathological sources. While cases caused by *M. lepromatosis* have not been identified in the Pacific Islands, two previously published whole *M. leprae* genomes from the region belong to the deepest lineages 0 and 5 (Benjak et al., 2018), and genotyping studies have identified the presence of all *M. leprae* SNP types in New Caledonia (Monot et al., 2005; Reibel et al., 2015). Unfortunately, genomic data are limited for strains present in the region, precluding a clear understanding of the diversity and evolutionary history of this pathogen.

Antiquity of M. leprae Globally and in the Pacific Islands

The oldest probable skeletal evidence of leprosy comes from Balathal, India, dating to 2000 BCE (Robbins et al., 2009), although there is a skeleton from Europe with possible rhinomaxillary signs of leprosy dating to 1500 years earlier (Köhler et al., 2017). The oldest written record of leprosy is described in the *Suchruta samhita*, an Indian text from around 600 BCE (described in Skinsnes, 1973; Dharmendra, 1947). Later evidence of leprosy in South-East Asia comes from a skeleton in Thailand, dating to 200 BCE - 300 CE (Tayles & Buckley, 2004). Skeletal and biomolecular evidence of leprosy from around the first to third century CE has also been identified in Central Asia (Blau &

Yagodin, 2005; Taylor et al., 2009). Lack of molecular evidence from these early skeletal cases, however, makes confirmation of the disease difficult.

More recent discussion of *M. leprae* population history has focused on entry into regions, specifically western Eurasia, rather than global patterning. Donoghue et al. (2015) argue that westward migration of groups from Central Asia in the first millennium may have introduced different *M. leprae* strains to Europe, whereas Schuenemann et al. (2018) offer two models for the high level of genetic diversity in medieval Europe. Either *M. leprae* originated in Western Eurasia or diverse strains of *M. leprae* with different geographical origins were introduced to Europe prior to the Medieval period (Schuenemann et al., 2018). Recent efforts to fill in the history of *M. leprae* have introduced many more ancient lineages to the family tree, however these ancient isolates are largely limited to Western Europe and Eurasia (Schuenemann et al., 2013; 2018).

Despite the considerable number of paleopathological investigations in the Pacific Islands, particularly on the US territory of Guam (see Pietrusewsky & Douglas, 2012), there is little published evidence of premodern skeletal leprosy. This could be due to its misdiagnosis as treponemal disease (Heathcote et al., 1998), compounded by the poor preservation of hand and foot bones - important for identifying skeletal leprosy - in tropical environments. The oldest convincing skeletal evidence of leprosy from the Pacific Islands comes from the Mariana Islands. Trembly (1995) describes six skeletons with destruction of the metatarsals, metacarpals, and manual and pedal phalanges, including concentric atrophy of the diaphyses, indicating probable leprosy infection (Møller-Christensen, 1961). One of these individuals' dates to 830 ± 170 CE, and the three others for which radiocarbon dating was performed date between 1140 and 1370

CE, suggesting the disease was present in the Pacific Islands prior to Portuguese and Spanish contact in the 1500s.

The earliest skeletal evidence of leprosy in Japan is later, with the oldest known skeletal case dated to the medieval period (1200-1400 CE) (Hirata et al., 2000). Suzuki et al. (2014; 2015) present skeletal and biomolecular evidence of leprosy infection from later time periods (1400-1800 CE). Historical evidence of leprosy in Japan, however, can be traced back to the 8th century CE in the *Chronicles of Japan* (*Nihon Shoki*) (described in Burns, 2019). Called *rai*, the disease description matches that of leprosy and was said to have been introduced by a migrant from the Korean peninsula (Burns, 2019, p. 20).

Following the WHO's regional country designations (WHO, 2019), the earliest historical and skeletal evidence of leprosy from the South-East Asia region and the continental portion of the Western Pacific region predates the earliest evidence from the Western Pacific Islands. While there is no published skeletal evidence of leprosy from premodern China, a passage from the medical text *Lingshu* describes signs consistent with leprosy, such as collapse of the nose, leg deformity, eyebrow loss, and hoarseness (Skinsnes, 1980); this text has been used to establish the presence of leprosy in China before the first century CE (Leung, 2009, pp. 19-25).

Based on the lack of diagnostic evidence and unclear descriptions of signs and symptoms, some historians and physicians have concluded that leprosy was introduced to the Pacific Islands during the 19th century, first in Hawaii and then to other Pacific Islands by Chinese migrants (McCarthy & Numa, 1962; Montgomerie, 1988). Native Hawaiians report that a mysterious disease arrived in the Hawaiian Islands from China, brought by Chinese sailors or by Native Hawaiians who had been sent to China on trade

missions. They call the disease *ma'i pake*, meaning "Chinese disease" and *m'ai ali'i*, meaning "chief's or royal disease" (Silva & Fernandez, 2006; Bindord, 1936). It has been argued that an uptick in leprosy cases occurred across the Pacific Islands during the 1800s (Luker & Buckingham, 2017); this is likely a bias introduced by Western physicians diagnosing the disease and geopolitical pressure to isolate those affected on island colonies, for example Kalaupapa on Molokai or Makogai in Fiji (Edmond, 2007). Because the convincing skeletal evidence of leprosy on the Mariana Islands by 800 CE (1995; 2002) is limited, it remains unresolved whether *M. leprae* were introduced during the initial island migrations or later during subsequent periods of colonialism and imperialism (Lange, 2017).

To further the understanding of the origins of *M. leprae* in the Pacific Islands as well as the phylogeography and genetic diversity of strains in an under sampled region, we present phylogenetic and evolutionary dating analyses of nine *M. leprae* genomes from Samoa (n=2), Hawaii (n=5), and Guam (n=2) isolated from formalin-fixed paraffin-embedded (FFPE) tissue samples. Such archival tissue specimens are valuable resources for clinical genomics; however, fragmentation and low DNA yield can make sufficient recovery difficult (Do & Dobrovic, 2015). We applied techniques used more commonly on ancient DNA, and after analyses of the genomic data, we determined that these isolate lineages fall into branches 0 and 5, which form the deepest lineages of the *M. leprae* family tree. These data also support a strong geographic association of modern lineages, with a temporal and geographic distribution that indicates an introduction of the pathogen during the peopling of the Pacific Islands.

2.3 Materials and Methods

Sampling

Data and samples were retrospectively collected from the pathology database at Hawaii Pathologists Laboratory, Honolulu, HI. Biopsies were pathologically diagnosed with leprosy between 1991 and 2016. In order to focus on the Pacific Islander population, presumptive Hawaiian or Polynesian ethnicities were determined by the patient's first or last name. Thirty-nine formalin-fixed paraffin-embedded (FFPE) samples from Pacific Islanders were obtained for this study. Patient samples were received throughout the Austronesian region; Oahu (n=18), Hawaii neighbor islands (n=6), Guam (n=3), Saipan (n=2), American Samoa (n=6), Marshall Islands (n=2) and Palau (n=2). There were 21 males and 18 females. The study population ranged from 9 to 79 years-of-age with an average age of 34 years-old.

DNA Extraction and qPCR analyses

Thirty-nine FFPE tissue biopsies from 34 people were sectioned using a microtome; one to three 40-46μm slices were removed from the FFPE block. In some cases, almost all of the sample was used. Other genetic studies using FFPE blocks removed 2-10 slices at 5-20μm per slice. We chose to use fewer but thicker slices so that we could manually excise the tissue sample from the paraffin, thereby reducing the amount of paraffin in the extraction reaction and inhibition in downstream PCR. Excess paraffin was manually removed from each slice using a scalpel blade, and extractions were performed directly on the excised tissue samples. The microtome blade and scalpel were sterilized with a 10% bleach solution, followed by water and 70% ethanol between sampling of each FFPE block.

DNA extractions were performed on 28 FFPE block samples following Purification of Total DNA from Animal Tissues protocol from the Qiagen DNeasy Blood and Tissue (DBT) kit with no FFPE-specific pre-treatments (following Do & Dobrovic, 2015) and some modifications: lysis occurred for 24 hours with intermittent vortexing, Qiagen MinElute spin columns were used in place of the DBT spin columns, and DNA was eluted twice in 100ul of Tris EDTA buffer with 0.05% Tween-20 (TET). Due to low yields, further extractions were performed using a protocol developed for extracting ancient DNA with a modified digest step (DAB) (Duggan et al., 2016; Dabney et al., 2013). DAB was used to extract DNA from 19 samples, including 11 new samples and eight of the samples previously extracted using DBT. To improve yields further, the DAB protocol was modified to include a 15-minute heat treatment at 98 °C during digestion and prior to the addition of proteinase K, following Gilbert et al. (2007).

All extracts, 1:10 dilutions of the extracts, and extraction blanks were screened for *M. leprae* DNA in triplicate using TaqMan qPCR assays designed to target two *M. leprae*-specific elements: 85B and RLEP. 85B is a single-copy gene, and RLEP is a multi-copy gene found in *M. leprae* (Martinez et al., 2006; Truman et al., 2008). Samples were chosen for whole-genome capture based on how many replicates amplified for both assays.

Library preparation and whole-genome capture

Fragment shearing was not necessary due to the already short fragment lengths (<300 bp). The 8 DBT and 11 DAB extracts that had the best qPCR signals were converted into double-indexed libraries following Meyer and Kircher (2010). Nucleotide misincorporation patterns were identified in the DBT sequencing data, so the 11 DAB

extracts were treated with uracil-DNA glycosylase (UDG) during library preparation. For the DBT libraries, 35 - 290 ng were enriched for the *M. leprae* genome using an Arbor Biosciences myBaits kit V 3.02. The in-solution capture reaction used biotinylated baits prepared from *M. leprae* Br4923, Thai53, and NHDP strains. After a 48-hour hybridization reaction, capture products were amplified for 14 cycles using AccuPrime™ Pfx DNA Polymerase. The DBT enriched libraries, extraction blank, and library blank were sequenced on the Illumina HiSeq2500 with 2 x 100 cycles. For the DAB libraries, 542-2890 ng were enriched for the *M. leprae* genome using the myBaits kit V 3.02 with the same bait preparation as above. The enriched DAB libraries, extraction blank, and library blank were sequenced with 2 x 75 cycles across two Illumina Mi-Seq runs.

Data acquisition and processing

Due to their low depth of coverage after DBT extraction, Samples 511, 515, 519, and 523A1 were re-extracted using DAB, re-captured, and re-sequenced. The quality of the raw data for these four samples from both sequencing runs was visualized using fastqc v0.11.7 (Andrews, 2018) and determined to be of high enough and consistent quality to concatenate, thus increasing genome coverage and depth of coverage. Paired-end reads from all samples were trimmed and merged using AdapterRemoval v2 v2.2.3, with the following parameters --minquality 20 --minlength 30 (Schubert et al., 2016). The resulting merged reads were mapped to the *M. leprae* TN strain (NCBI NC_002677.1) using Burrows-Wheeler Alignment tool (BWA) v0.7.17 with the following commands and parameters: aln samse -l 1000 -n 0.1 (Li & Durbin, 2009). Unmapped reads were removed using samtools view and PCR duplicates were removed using samtools markdup -r v1.9 (Li et al., 2009). Given that samples not treated with UDG exhibited

terminal misincorporations (Supplementary Figures 1 and 2), MapDamage 2.0 (v2.0.9) was used to rescale the quality scores for terminal bases prior to variant calling (Okonechnikov et al., 2016). Coverage and mapping quality estimates were generated using QualiMap 2 (v2.2.1) (54).

For comparative purposes, genomic data for 164 modern and ancient *M. leprae* samples were downloaded from the NCBI Sequence Read Archive in FASTQ and FASTA format (Supplementary Table 1). Raw fastq data were downloaded using SRA Toolkit v2.8.0 (https://github.com/ncbi/sra-tools) fastq-dump command. The 164 comparative samples used in this analysis do not include published samples with an average depth of coverage of <5x. Samples that have been identified as hypermutators were included in the maximum likelihood and maximum parsimony phylogenetic trees but not in the BEAST analysis (Benjak et al., 2018). Ancient sample data were processed exactly as the study samples. Modern sample data were processed similarly to the above samples, except that paired-end reads were not merged, and the mapping qualities were not rescaled. Assembled genomes in FASTA format, TN, Br4923, Kyoto-2, and *M. lepromatosis* JRPY were mapped to the reference genome using bwa mem with default parameters (Li, 2013).

Variant calling and analysis

SNPs were called for all samples using GATK UnifiedGenotyper v3.5 with default parameters except for --out_mode EMIT_ALL_SITES (DePristo et al., 2011). The resulting VCF files were filtered for homozygous SNPs with a coverage depth of at least 5x and a GATK genotyping quality of at least 30 and aligned using MultiVCFAnalyzer v0.85.1 (https://github.com/alexherbig/MultiVCFAnalyzer/releases)

(Bos et al., 2014). Also using MultiVCFAnalyzer, positions that are in known repeat regions and that were covered in the SK12 negative control were excluded, following Honap et al. (2018) and Schuenemann et al. (2013). SNPs in the *M. lepromatosis* outgroup were only called if they were also present in *M. leprae*. For the full genomes, the VCF files were manually edited for a coverage depth of >5x and a genotyping quality of at least 30 so that they could be filtered simultaneously with the raw fastq data. An alignment of 3521 SNPs was generated; after removing sites with less than 95% coverage across the dataset, a final alignment of 2736 SNPs was used for phylogenetic analyses. After removing genomes previously identified as hypermutators (Schuenemann et al., 2018) and sites with less than 95% site coverage, a final alignment of 2184 SNPs was used for BEAST (Bouckaert et al., 2014) analysis.

To investigate the effects of the unique SNPs in our samples that are shared among the Pacific Island strains, the VCF files for the samples generated in this study were processed using SnpEff v3.1 to annotate variants and determine their functional effects (Cingolani et al., 2012). Default parameters were used, except the upstream and downstream interval size was set to 100. SnpEff v3.1 was used to ensure compatibility with the MultiVCFAnalyzer SNP table output.

Phylogenetic Analysis

To examine consistency in topology across different methods, two phylogenetic trees were made. A maximum likelihood tree was constructed using the GTR nucleotide substitution model with the GAMMA model of rate heterogeneity (-m GTRGAMMA) with 1000 bootstrap replicates using RAxML v8.2.12 (Stamatakis, 2014). A maximum parsimony (MP) tree was made using the subtree-pruning-regrafting inference model and

a bootstrap test of phylogeny with 1000 replicates in MEGA7 (Kumar et al., 2016). All trees were rooted using *M. lepromatosis* as the outgroup.

We also estimated the time to Most Recent Common Ancestor (tMRCA) and substitution rates through application of Bayesian methods using BEAST 2.4.5 (Bouckaert et al., 2014). A concatenated SNP alignment of 2184 informative sites was used in the analysis. Additionally, counts of invariant sites shared by all strains and the reference were included to avoid ascertainment bias of only using variant sites. The alignment was run through the jModelTest2 tool (Darriba et al., 2012; Guindon & Gascuel, 2003) to determine a best-fit model of nucleotide substitution. The alignment was analyzed using a Bayesian Skyline model, assuming a variable population size (Bouckaert et al., 2014) with a relaxed clock and a GTR substitution model (Drummond et al., 2005; 2006). The alignment was also run again using a strict clock with a GTR substitution model and a lognormal relaxed clock with an HKY substitution model. Each run was completed with 300,000,000 iterations and a 30,000,000 burn-in. Calibrated radiocarbon dates were used for ancient strain tip dates and sampling/isolation dates were used for modern samples.

2.4 Results

DNA extraction and Screening

DNA extraction for DBT and DAB (Duggan, et al., 2016; Dabney et al., 2013) methods yielded between <0.50 ng/µl and 18.2 ng/µl. For the initial 28 DBT extractions, the yields were between 0.0005 and 0.7 ng/µl. Due to these low yields, further extractions were performed on 19 samples, eight of which had also been extracted using DBT. The

DAB extraction yields were between 0.516 - 18.2 ng/µl. All samples and 1:10 dilutions of each sample underwent qPCR assays in triplicate for 85B and RLEP regions. Samples with all three replicates amplifying for at least one assay were considered for whole-genome capture, including assays of 1:10 diluted extracts (Table 1). In addition to the number of replicates amplified, the mean number of cycles of amplification, DNA quantity, and leprosy type were evaluated before a subset of samples was selected for whole-genome capture and sequencing (Supplementary Table 2).

Table 1. Summary of DNA extraction concentrations and qPCR assay results for samples sequenced in this study. The number of successfully amplified replicates out of three is given for each assay.

Sample ID	Type	DNA Extraction Type	DNA Concentration (ng/ul)	85B	85B 1:10	RLEP	RLEP 1:10
511	LL	DBT	0.414	3/3	2/3	3/3	3/3
		DAB	6.68	0/3	0/3	3/3	3/3
515	LL	DBT	0.198	3/3	1/3	3/3	3/3
		DAB	9.96	0/3	0/3	1/3	3/3
516	LL	DBT	<0.01	3/3	3/3	3/3	3/3
		DAB	2.56	0/3	0/3	3/3	3/3
517	LL	DBT	0.436	3/3	3/3	3/3	3/3
		DAB	6.9	0/3	0/3	3/3	3/3
518	BL	DBT	0.67	3/3	3/3	3/3	3/3
		DAB	1.56	0/3	0/3	0/3	3/3
519	BL	DBT	0.26	3/3	3/3	3/3	3/3
		DAB	6.06	0/3	0/3	0/3	3/3
520	BL	DBT	0.138	3/3	3/3	3/3	3/3
		DAB	2.46	0/3	0/3	0/3	3/3
523A1	BL	DBT	<0.01	3/3	3/3	3/3	3/3
		DAB	1.75	0/3	0/3	0/3	3/3
536	LL	DAB	18.2	0/3	0/3	0/3	3/3
537	LL	DAB	6.3	3/3	0/3	3/3	3/3
538	LL	DAB	1.23	1/3	1/3	0/3	3/3
539	BL	DAB	0.516	3/3	0/3	3/3	3/3
540	BL	DAB	3.28	0/3	0/3	0/3	3/3
542	BT	DAB	0.952	0/3	1/3	0/3	3/3
543	BT or BB	DAB	2.98	0/3	0/3	0/3	3/3

Genome-wide Analyses

Of the 15 samples chosen for whole-genome sequencing, nine have sufficient coverage for whole-genome analysis (Table 2). Of these samples, the depth of coverage ranges from 4x to 63x, with the percent of the reference genome being covered at 5x ranging from 43% to 98%. The mapping statistics of the comparative data and a comparison of the mapping statistics for the four samples that underwent extraction using both methods are available in Supplementary Tables 2 and 3, respectively.

The average fragment lengths of merged reads mapped to the *M. leprae* reference range from 38 to 71 bp for all 15 samples, and from 52 to 71 bp for the nine samples with an average depth of coverage > 4x . For the eight samples that were sequenced without UDG treatment, fragment misincorporation ranged from 1.5% to 3.5% at the terminal 3' base and from 1.8% to 3.7% at the terminal 5' base (Supplementary Table 4). A visual comparison of a sample that was sequenced with and without UDG treatment can be seen in Supplementary Figures 1 and 2.

Table 2. Whole-genome analysis summary of samples sequenced in this study. Note that reads generated from separate extractions were concatenated for samples 511, 515, 519, and 523A1.

Sample	Origin	Leprosy form	Non-duplicate, uniquely mapped reads	Average depth of coverage (X)	% reference covered >= 1X	% reference covered >= 5X	Average fragment length	Endogenous content* %	Branch	Number of SNPs	Non-synonymous coding SNPs
511	Samoa	LL	576,690	9.8	96	77	56	8.7	0	141	38
515	Hawai'i	LL	635,824	10.1	95	73	52	15	5	101	20
516	Samoa	LL	635,186	13.1	98	94	67	64.2	0	172	54
517	Guam	LL	1,574,103	31.4	98	97	65	93.2	5	154	32
518	Hawai'i	BL	2,899,280	63.2	98	98	71	80.5	0	184	57
519	Hawai'i	BL	545,816	10.3	97	87	62	42.8	0	152	42
520	Hawai'i	BL	1,616,046	32.7	98	97	66	94	0	183	56
523A1	Hawai'i	BL	242,303	4.3	94	43	58	15.2	5	31**	7**
536	Guam	LLp	374,295	6.1	91	52	53	20	5	69**	12**
537	Northern Mariana Islands	LL	6,308	0.1	8	0	44	0.4	-	-	-
538	Samoa	LL	1,365	0	2	0	38	11	-	-	-
539	Northern Mariana Islands	BL	20,388	0.3	30	0	57	6.2	-	-	-
540	Northern Mariana Islands	BL	4,691	0.1	7	0	54	1.3	-	-	-
542	Palau	BT	1,589	0	2	0	53	5.7	-	-	-
543	Palau	BT or BB	1,889	0	3	0	49	1.3	-	-	-

*# of reads after mapping before duplicate removal/number of trimmed and merged reads
**The low number of SNPs in samples 523A1 and 536 reflects the overall low coverage of those genomes at ≥ 5x depth of coverage

Using SnpEff, a range of 31 to 184 SNPs was identified in the new genomes presented here (Table 2). Additionally, samples have between 7 and 57 non-synonymous SNPs in coding regions. A summary of SNPs and their effects can be found in

Supplementary Table 5. Only five samples have unique SNPs located within genes, 515 (n=1), 516 (n=2), 517 (n=3), 518 (n=2), and 520 (n=1); details of these unique SNPs can be found in Supplementary Table 6.

The genomes generated in this study were found to have all previously identified branch-defining SNPs for either branch 0 or 5 (Supplementary Table 7) (Schuenemann et al., 2018). Additionally, the Pacific Island clade of four genomes within branch 5 (515, 523A1, 536, US57) was found to share four non-synonymous SNPs, and the Pacific Island clade of six genomes within branch 0 (511, S9-96008, 516, 518, 519, 520) was found to share 20 non-synonymous SNPs, including one in a gene associated with drug resistance (*rpoB*).

Phylogenetic analyses

Phylogenetic analyses show that the new *M. leprae* genomes fall within the most basal branches, branches 0 and 5. Specifically, the genomes from Samoa (n = 2) fall onto branch 0, the genomes from Guam (n = 2) fall onto branch 5, and genomes from Hawaii fall onto branch 0 (n = 3) and branch 5 (n = 2) (Figure 1). All of our phylogenetic trees reveal a topology that is consistent with previously published studies (Benjak et al., 2018; Schuenemann et al., 2013; 2018), as well as a consistent branch placement for the nine new genomes we introduce here (Figure 2 and Supplementary Figure 3).

The inclusion of the Pacific Island genomes creates geographically associated substructure within branch 0. The Pacific Island genomes, including S9-96008 from New Caledonia, form a clade, as do the East Asian lineages and the medieval European lineages. Additionally, the branch 0 Pacific Island genomes are more closely related to a strain isolated from a naturally infected crab-eating macaque from the Philippines (Honap

et al., 2018) than they are to other branch 0 strains. In branch 5, the two genomes from Hawaii (515 and 523A1) and one genome from Guam (536) are closely related to a strain that was isolated from the Marshall Islands (US57), while one genome from Guam (517) is more closely related to a strain previously isolated from the Ryukyu islands of Japan (Ryukyu-2).

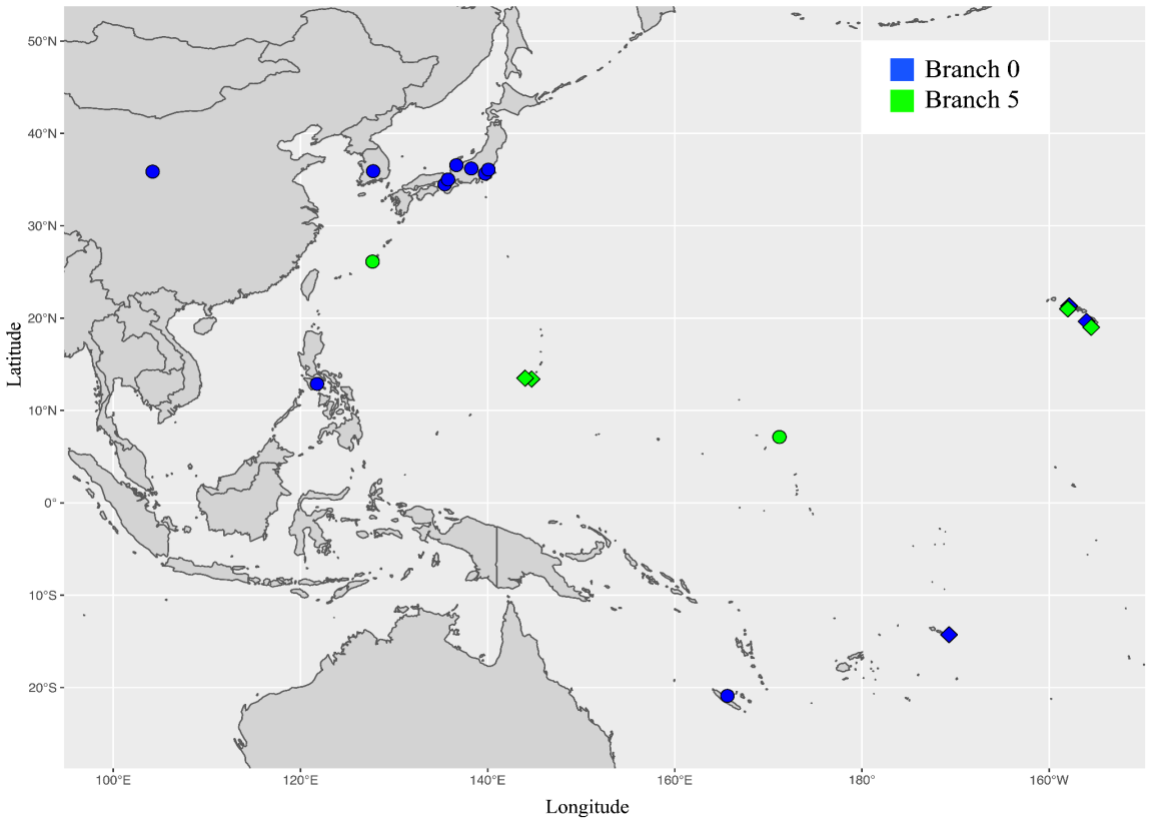


Figure 1. Map of the Pacific showing source locations of novel and previously sequenced *M. leprae* strains from branch 0 and branch 5. Novel strains from this study all fall within these branches and are shown as diamonds. Comparative data are shown as circles.

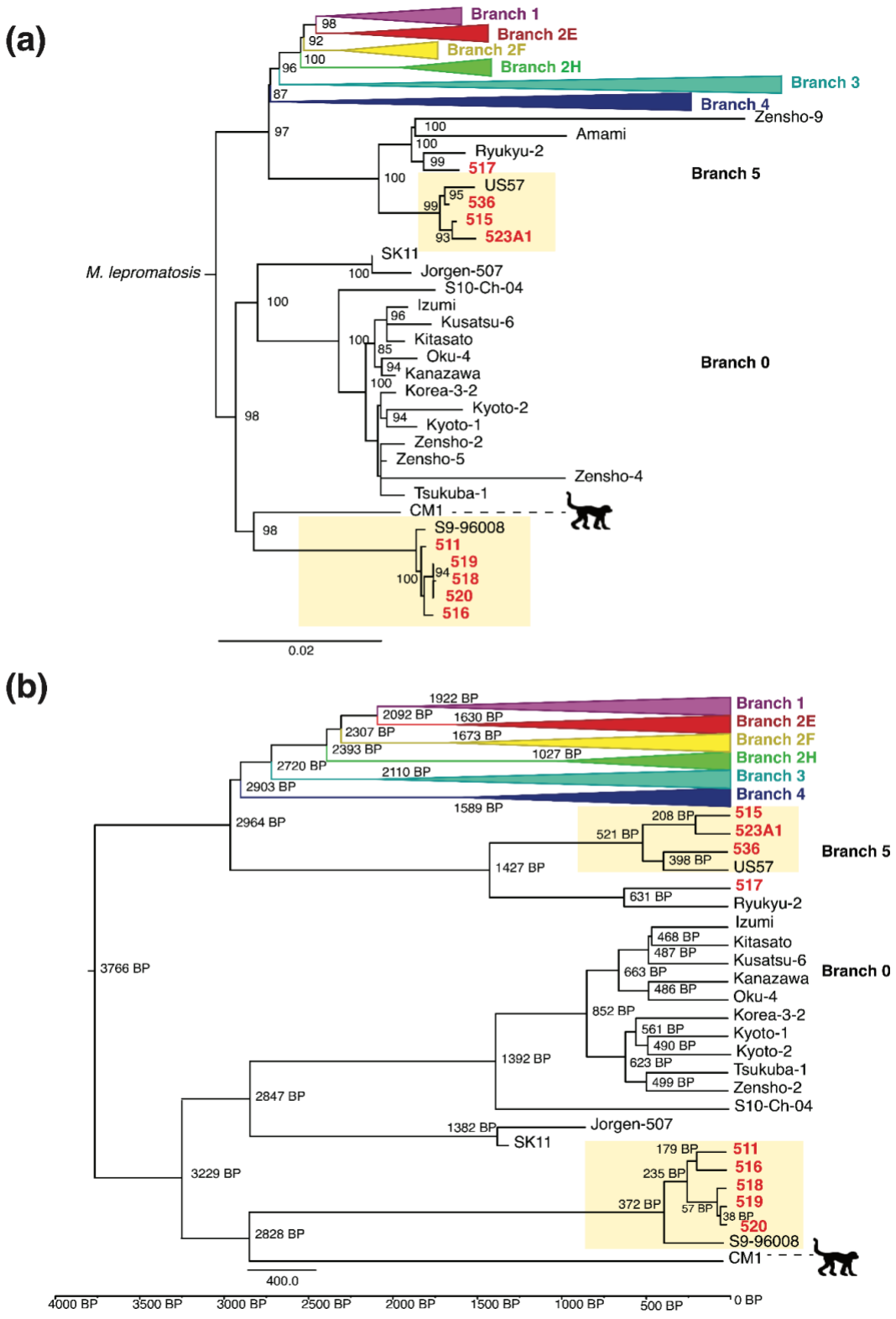

(a)
Branch 1
Branch 2E
Branch 2F
Branch 2H
Branch 3
Branch 4
Zensho-9
Amami
Ryukyu-2
517
US57
536
515
523A1
Branch 5
M. lepromatosis
SK11
Jorgen-507
S10-Ch-04
Izumi
Kusatsu-6
Kitasato
Oku-4
Kanazawa
Korea-3-2
Kyoto-2
Kyoto-1
Zensho-2
Zensho-5
Zensho-4
Tsukuba-1
CM1
S9-96008
511
519
518
520
516
Branch 0
0.02
(b)
1922 BP
2092 BP
1630 BP
2307 BP
1673 BP
2393 BP
1027 BP
2720 BP
2110 BP
2903 BP
1589 BP
2964 BP
208 BP
521 BP
398 BP
1427 BP
631 BP
3766 BP
468 BP
487 BP
663 BP
486 BP
852 BP
561 BP
490 BP
623 BP
499 BP
1392 BP
2847 BP
1382 BP
3229 BP
179 BP
235 BP
372 BP
57 BP
38 BP
2828 BP
400.0
4000 BP
3500 BP
3000 BP
2500 BP
2000 BP
1500 BP
1000 BP
500 BP
0 BP

Figure 2. (a) Maximum likelihood phylogenetic tree created using 2736 informative SNPs from an alignment of 164 comparative samples and the nine new genomes presented here; branch lengths are proportional to the number of substitutions, and the bootstrap values above 85 are present at nodes. Novel strains from this study are shown bold and in red. Pacific Island-specific clades are designated with beige boxes (b) Bayesian phylogenetic tree based on 2184 informative SNPs as well as shared invariant sites calculated with BEAST 2.4.5 (Bouckaert et al., 2014). Median divergence times before present (BP) shown on main branch nodes. Novel strains from this study are shown bold and in red. Pacific Island-specific clades are designated with beige boxes

BEAST Analyses

The time to the most recent common ancestor (tMRCA) for *M. leprae* strains was estimated to be 3766 y BP (3011 y – 4572 y 95% Highest Posterior Density (HPD)) under the Bayesian Skyline model with a relaxed lognormal clock and a GTR substitution model (Figure 2b). The mutation rate for *M. leprae* was estimated to be 3.255×10^{-9} ($2.706 \times 10^{-9} – 3.821 \times 10^{-9}$, 95% HPD) substitutions per site per year, which is slightly slower than previous estimates (Schuenemann et al., 2018; Benjak et al., 2018; Honap et al., 2018). The Bayesian Skyline analysis run under a strict clock model with a GTR substitution model produced a tMRCA of 3878 y BP (3284 y - 4510 y BP 95% HPD) with a substitution rate of 3.101×10^{-9} ($2.654 \times 10^{-9} – 3.554 \times 10^{-9}$, 95% HPD) substitutions per site per year. The Bayesian skyline analysis run under a relaxed clock model with an HKY substitution model produced a tMRCA of 3765 y BP (2973 - 4623

BP 95% HPD) with a substitution rate of 3.221×10^{-9} ($2.643 \times 10^{-9} – 3.834 \times 10^{-9}$, 95% HPD) substitutions per site per year. Further analyses run using two demographic models using alignments without novel samples, as well as without low coverage 523A1 and 536 samples, had similar results (Supplementary Table 8).

2.5 Discussion

Here we report nine *M. leprae* genomes from the Pacific Islands. Phylogenetic analyses show that they all belong to branches 0 and 5, which previously have been identified in modern contexts only in the Western Pacific, including Japan, China, and the Philippines, and, interestingly, in medieval Hungary and Denmark (Schuenemann et al., 2018; Benjak et al., 2018; Honap et al., 2018). Prior to this study, only two whole genomes from the Pacific Islands had been published, with Sample US57 (Marshall Islands) falling within branch 5 and Sample S9-96008 (New Caledonia) falling within branch 0 (Schuenemann et al., 2018; Benjak et al., 2018). In addition to creating geographically associated substructure in branch 0, the Pacific Island genomes we sequenced tripled the number of whole-genome samples within branch 5.

Our tMRCA estimate of 3766 y BP (3011 y – 4572 y 95% HPD) is very close to that of Benjak et al. (2018) and about 300 - 750 years younger than the Schuenemann et al. (2018) estimate. Analyses completed with different demographic and substitution models also have estimates of tMRCA within decades of our initial estimate. The estimated rates of substitution, however, differ slightly, both with different demographic and substitution models and between other studies. The estimated substitution rate per site per year of 3.255×10^{-9} ($2.706 \times 10^{-9} – 3.821 \times 10^{-9}$, 95% HPD) is similar but slightly

slower than those estimated previously (Avanzi et al., 2016b; Benjak et al., 2018; Schuenemann et al., 2013; 2018), with non-overlapping 95% HPD ranges in some cases (Avanzi et al., 2016b; Benjak et al., 2018; Schuenemann et al., 2018). Our analyses suggest a marginally younger tMRCA and slower substitution rate that is within the same magnitude as previous estimates. Despite these differences with previously published studies, our overlapping HPD ranges obtained when using multiple models and different combinations of samples indicate that our data are robust and informative (Supplementary Table 8).

Our data suggest that *M. leprae* may have been in the Pacific Islands since the initial peopling of Remote Oceania and was possibly re-introduced during newly recognized subsequent migrations (Posth et al., 2018; Lipson et al., 2018). Remote Oceania was occupied beginning around 3000 BP by Austronesian-speaking people who originally expanded out of East Asia (Posth et al., 2018; Spriggs, 2011; Petchey et al., 2017; Skoglund et al., 2016). Ancient DNA data suggest that another migration wave from New Guinea nearly replaced these original inhabitants by 2300 BP (Posth et al., 2018; Lipson et al., 2018). Because the Pacific Island *M. leprae* genomes form a clade within branch 0, and the date of the LCA of branch 0 clades is 3229 BP, it is possible that the Austronesian migration brought this basal lineage to Remote Oceania. This scenario also aligns with a South Asian origin of leprosy, which has been suggested elsewhere (Monot et al., 2005). Since branch 5 strains share a common ancestor with branches 1-4 at 2964 BP and only consist of genomes from Pacific Islands, it is possible that these younger lineages were introduced during subsequent migrations into Remote Oceania.

The Pacific Island clade within branch 0 is differentiated from other branch 0 strains by a high number of non-synonymous SNPs, in contrast to the Pacific Island clade within branch 5 and the overall low genetic diversity of *M. leprae*, which likely reflects the older age of this clade. The low diversity within the branch 0 Pacific Island clade could reflect an introduction and radiation of these strains within the Pacific Islands around 350-400 years ago. The low divergence among these lineages, however, could be the result of sampling bias or genetic drift during infection and transmission across the Polynesian Islands during this time. Multiple lines of evidence, such as oral histories (Cachola-Abad, 1993), geochemical analysis of traded stone tools (Weisler, et al., 2016; Collerson & Weisler, 2007), and population genetics of the Polynesian rat (Matisoo-Smith et al., 1998), demonstrate that voyaging among islands was common after initial settlement until approximately 1600 CE (Weisler et al., 2016; Rolett, 2002). This cessation of inter-island voyaging roughly aligns with the age of the branch 0 Pacific Island clade and may have served as a bottleneck for *M. leprae* diversity. Nonetheless, the tight geographic clustering of clades, particularly within branch 0, coupled with the old age of the last common ancestor of Japanese and Pacific Island strains, strongly supports a more ancient introduction.

Despite paleopathological evidence for leprosy in the Pacific Islands prior to 1000 CE, some researchers have suggested that leprosy was absent in the Pacific Islands until the 19th century CE and was introduced as a result of European and Japanese imperialism or large-scale Chinese migration throughout the region (Luker & Buckingham, 2017; Lange, 2017; Browne, 1985; Akeli, 2007; D'Arcy, 2014). However, if strains were introduced during the Japanese imperial occupations beginning in 1890, we would not

expect to see such deep divergence between Japanese and Pacific Island strains. Further, we would expect the Pacific Island genomes to be more closely related to the genome from China (S10-Ch-04), instead of having diverged over 3,000 years ago. Likewise, if European colonists introduced *M. leprae* to the Pacific Islands, we would expect at least some modern infections to be caused by branch 3 lineages, since (1) this branch was suspected to have been introduced to the Americas by European colonists, (2) strains falling within branch 3 have been isolated from modern squirrels in the UK, and (3) members of the branch have been identified in high frequency, along with branch 2, in medieval European skeletons (Schuenemann et al., 2018; Benjak et al., 2018; Drummond et al., 2005). Additionally, if strains were introduced in the 16th century during European exploration, then we would expect the Pacific Island strains in branch 0 to be more closely related to the ancient European strains within branch 0 (SK11 and Jorgen 507).

Although the timeline we propose based on the phylogeographic and evolutionary dating analysis is not corroborated by the archaeological record, this is likely due to the poor preservation on tropical islands, as well as the paucity of large documented skeletal assemblages from these earlier time periods across Asia. Additionally, it is possible that these earlier lineages of *M. leprae* did not elicit the same skeletal response as they do today, obscuring the identification of the disease in the archaeological record. We also have a poor understanding of what animal reservoirs exist in this region and the extent of exchange among species. Though the Pacific Island branch 0 strains are closely related to one isolated from a crab-eating macaque from the Philippines (Honap et al., 2018), it is unclear when and in which direction the exchange occurred. Further sampling is clearly needed to expand our understanding of genetic diversity; however, in regions without

sufficient archaeological evidence of leprosy and few surveys of potential animal reservoirs, clinical specimens, such as FFPE samples, are valuable resources for characterizing modern diversity of *M. leprae* in humans.

For this study FFPE samples allowed us to expand upon our knowledge of modern standing diversity in branches 5 and 0. Such samples have been successfully used in past *M. leprae* research (Benjak et al., 2018; Stefani et al., 2017), although both studies used a DNA extraction kit more specific to FFPE DNA. We found that DNA extraction protocols commonly used for ancient DNA were also effective. Our small comparison between the DAB and DBT extraction techniques found marginal improvements in coverage and mapping quality in the ancient DNA-specific DAB methods, but the limited sample size prevents a conclusive comparison. Interestingly, the DAB extracts had an increased extraction quantification of 2-50-fold over the DBT extracts but had fewer replicates amplifying for the qPCR assays. This could reflect the inclusion of more inhibitors through the DAB extraction or, since all 85B assays on the DAB extracts were performed during the same experiment, an inefficient qPCR run. The small sample size of this study precludes definitive conclusions about which extraction method is best for FFPE material. The added heat treatment step and extended proteinase K treatment targeted DNA crosslinks and protein-DNA crosslinks, which are commonly induced during the formalin fixation step (Mcdonough et al., 2019; Carrick et al., 2015). DAB methods and downstream bioinformatics pipelines can also address extensive DNA fragmentation found in FFPE samples, as well as the duplicates accumulated during the multiple rounds of amplification required for targeting low-concentration endogenous DNA (Do & Dobrovic, 2015; Dabney et al., 2013). Despite the difficulties of dealing

with FFPE samples, methods are available to address their shortcomings and add a useful resource to answer further questions regarding *M. leprae* evolutionary history.

2.6 Conclusion

The phylogeographic and evolutionary dating analyses of nine new *M. leprae* genomes from the Pacific Islands suggest that *M. leprae* may have been introduced to Remote Oceania during the first human migrations at approximately 3000 BP, with re-introduction during subsequent migrations. Although a better understanding of *M. leprae* genomic diversity in Near Oceania and China is needed to support this hypothesis further, the data presented here strongly suggest a premodern introduction of leprosy into the Pacific Islands and refute its initial introduction either during European exploration or 19th century imperialism and colonialism.

2.6 Ethics

The Dermatopathology services at Hawaii Pathology's Laboratory approved of the ethical use of this formalin fixed paraffin embedded tissue for the exploration of population genetic inquiries into the diaspora of *Mycobacteria leprae* for genomics studies in the Pacific. We believe that this research will allow us to create high resolution diagnoses of tropical diseases and chart the future development of predictive diagnoses in remote communities.

2.7 Data, code, and materials

All raw sequencing data generated for this project, including samples, extraction blanks, and library preparation blanks, can be found under the NCBI BioProject accession PRJNA603842. For convenience, NCBI SRA accession numbers by sample can be found in Supplementary Table 1. Scripts are available at https://github.com/ACStoneLab/Mleprae

CHAPTER 3

GENETIC DIVERSITY AND SPATIAL DYNAMICS OF *MYCOBACTERIUM LEPRAE* IN THE PACIFIC

3.1 Abstract

Leprosy, or Hansen's disease, is a modern global health concern, with the WHO reporting approximately 200,000 new cases each year. Recent work has improved our understanding of the pattern of genomic variation in *M. leprae* strains; however, certain regions remain understudied and unrepresented in the phylogeny. There is especially poor representation from *M. leprae* isolates in the Pacific, which have only seen recent efforts in whole-genome sequencing. Furthermore, obtaining sequences from less severe cases of leprosy has been difficult given the lower bacillary load and common clinical sampling methods. To address these challenges, we extracted and sequenced DNA from 79 formalin-fixed paraffin-embedded (FFPE) tissue samples from Guam, Hawaii, the Marshall Islands, Palau, Pohnpei, and Samoa collected between 1992 and 2016. Of that number, 20 samples from mostly paucibacillary leprosy cases were selected for a pairwise comparison of two DNA extraction methods developed for FFPE analysis. Comparative and phylogenetic analyses of the sequences revealed distinct allelic changes in intrahost samples during treatment, clustering of strains with other Pacific and East Asian strains, and a potential distribution history of these stains. We also characterize the movement of *M. leprae* strains through the Pacific with phylogeographic approaches to clarify the geographic patterning and evolutionary history of leprosy.

3.2 Introduction

Leprosy, or Hansen's Disease, is caused by the intracellular, obligate pathogen *Mycobacterium leprae* and the more recently discovered *M. lepromatosis*, which is endemic to the Caribbean and Mexico (Han et al., 2012; 2008). The *M. leprae* genome underwent reductive evolution and is compact (3.3 Mb) with 1614 protein coding genes and 1300 pseudogenes, which make up approximately half of the genome's length (Cole et al., 2001). The genome is highly conserved, with <300 single-nucleotide polymorphisms (SNPs) segregating among geographically distant strains (Monot et al., 2009; Schuenemann et al., 2018).

M. leprae strains are commonly grouped by shared informative SNPs and six single-base indels into four SNP types and 16 SNP subtypes (1 – 4 and A – P respectively) (Monot et al., 2005; 2009). Phylogenies generated from whole genome data have also been used to define six branches (branches 0 – 5) (Benjak et al., 2018; Schuenemann et al., 2013; 2018). The branches and SNP types have strong geographic associations, with branch 0 primarily found in Eastern Asia and the Pacific, branch 1 largely in Southern and Eastern Asia, the paraphyletic branch 2 in the Near East and South Asia, branch 3 in North and Latin America, branch 4 mainly in West Africa and South America, and branch 5 in the Pacific Islands (Monot et al., 2009; Schuenemann et al., 2018). Interestingly, recent work on *M. leprae* lineages isolated from European skeletal samples has shown that extremely diverse strains were circulating in medieval Europe, spanning branches 2, 3, 4, and 0 (Schuenemann et al., 2018).

Although the transmission route for leprosy has not fully been established, there is a higher risk for leprosy among household contacts and those in close proximity to active cases (Romero-Montoya et al., 2017; Araujo et al., 2016). Research is ongoing regarding animal reservoirs (Oliveira et al. 2019) and potential vector or environmental transmission (Ploemacher et al., 2020). The most likely route is through infectious aerosols when coughing or sneezing, as well as skin-to-skin contact with active cases (Gama et al., 2018). The clinical presentations of leprosy depend upon immune response and have been classically described by Ridley and Jopling (1966) as a spectrum ranging between lepromatous leprosy to tuberculoid leprosy, with various borderline states in between. Tuberculoid leprosy is characterized by a small number of skin lesions with low bacillary load, also commonly referred to using the more recent WHO classification of paucibacillary (Britton, 1993; WHO, 2017). The lepromatous form is characterized by numerous infiltrated skin lesions with high bacillary loads, or "multibacillary" under WHO classification. This is also associated with impaired peripheral nerves and possible involvement of internal organs (WHO, 2017; Walker & Lockwood, 2007). Paucibacillary and multibacillary classifications now cover the two ends of the leprosy manifestation spectrum, including the borderline states in between tuberculoid and lepromatous.

The epidemiology of leprosy is difficult given its unculturable nature and low genetic diversity, especially when analyzing SNP/VNTR typing, which is more common and affordable in *M. leprae* studies (Hall & Salipante, 2010). Spatial clustering analyses of leprosy suggest that there is strong familial aggregation but in hyperendemic areas, multiple strains may exist in an individual, household, or community (Avanzi et al., 2016a). Both genetic susceptibility and social factors, such as contact with infected

individuals, contribute to the possibility of developing leprosy (Shields et al., 1987; Das et al., 2020). This area of research has focused on disease risk analysis, GWAS studies for susceptibility in humans, and phylogeographic approaches (Das et al., 2020; Faber et al., 2023). Epidemiological studies that incorporate whole genome data focus on generating many strains of an outbreak (*e.g.,* Avanzi et al., 2020a) and may infer much more about local spatial dynamics of the disease. Phylogenetic studies can incorporate transmission dynamics to a lesser extent, but may be less resource-intensive, easily incorporate historic and ancient strain data, and still add to our understanding of genetic diversity and disease movement in under sampled outbreaks or understudied regions.

Here, we leverage newly sequenced *M. leprae* genomic data and a phylogeographic approach to characterize the spread of the disease through the Pacific. We hypothesize that transmission patterns in the Pacific will trace the route of the initial peopling of the Pacific between 3000 BP and 2300 BP (Posth et al., 2018; Lipson et al., 2018), with later novel strain inflow from European and Asian contact during recent centuries.

3.3 Materials and Methods

Sample Selection

79 formalin-fixed, paraffin-embedded (FFPE) tissue samples were collected from 63 individuals undergoing treatment for leprosy between 1995-2015 in Chuuk State (Micronesia), Guam, Hawaii, the Marshall Islands, Palau, Pohnpei (Micronesia), Saipan, and Samoa. Some individuals had multiple tissue biopsies collected over the course of

treatment, and those time-series samples were included in our study. Of that number, 68 cases were classified as borderline lepromatous or lepromatous leprosy, and 11 cases were classified as borderline or borderline tuberculoid leprosy. Tissue samples were manually excised from the paraffin blocks and diced to increase surface area. Previous work (Blevins et al., 2020, Stefani et al., 2017) used a microtome to section FFPE samples; however, we have found that manual excision more easily removes more wax from tissue without the extra use of xylene or other organics often recommended (Gilbert et al., 2007).

DNA Extraction Methods

Initial DNA extractions of 59 FFPE tissue samples (borderline lepromatous or lepromatous leprosy cases) were done using the DNeasy Blood and Tissue Kit (Qiagen). As per Gilbert et al. (2007) recommendations, an extended digestion step of 96h at 56C was included. Due to low initial DNA concentrations as determined by Qubit 3.0 Fluorometer (Invitrogen), 20 more FFPE tissue samples (borderline tuberculoid, borderline, borderline lepromatous, and lepromatous cases) were selected for a paired analysis of two FFPE-specific DNA extraction methods. Each sample was divided and underwent DNA extraction using the Hot-Alkaline Lysis Phenol-Chloroform protocol (Campos & Gilbert, 2012; Hahn et al., 2021) and the QIAamp FFPE DNA Kit (Qiagen), as they were both developed for FFPE tissue DNA extraction and recommended in recent methods papers (Hahn et al., 2021; McDonough et al., 2019; Sarnecka et al., 2019). DNA extraction modifications for paired samples are fully described in supplementary data (Supplementary Document 1).

All DNA extractions, 1:10 dilutions of the extracts, and extraction blanks were screened for *M. leprae* DNA using a TaqMan qPCR assay designed to target two *M. leprae* specific elements RLEP and 85B (Truman et al., 2008; Martinez et al., 2006). Samples were run in triplicate and those that had at least two replicates amplify moved on to library preparation. Paired samples with at least one or both assays showing the presence of RLEP or 85B went on to library preparation.

Library Preparation, Enrichment, and Sequencing

DNA extracts underwent fragment analysis using the TapeStation and those exceeding 300 bp in length were sheared using a Covaris M220 Ultrasonicator (ASU Genomics Facility) for a target insert size of 200 bp. Unpaired extracts (HL1-HL59) were converted into double-indexed, single-stranded libraries using the BEST protocol optimized for ancient and degraded DNA (Kapp et al., 2021), whereas paired extracts (HL60-HL98) underwent double-indexed, double-stranded library preparation following Meyer and Kircher (2010). The Meyer and Kircher (2010) was used due to poor initial results with the BEST protocol. Both library preparation procedures included a partial uracil-DNA glycosylase (UDG) treatment step. UDG treatment is commonly used in ancient DNA sequencing to reduce nucleotide misincorporation and has been shown to dramatically reduce non-reproducible sequence artifacts seen in FFPE tissues (Berra et al., 2019; Rohland et al., 2015). Libraries were enriched for the *M. leprae* genome using a Daicel Arbor Biosciences myBaits kit v5 with biotinylated baits prepared from *M. leprae* strains Thai53, and NHDP. Capture products were then amplified using the KAPA HiFi DNA Polymerase (Roche). Enriched libraries, extraction blanks, and library blanks were sequenced on the Illumina HiSeq (Fulgent Genomics, California, United States). Due to

poor sequencing results, enriched libraries and blanks were re-sequenced on the Illumina NovaSeq X Plus (Admera Health, New Jersey, United States).

Data Analysis

Raw sequence data were processed using the EAGER v2 pipeline (Fellows Yates et al., 2021), which uses Nextflow v22.10.6 (Di Tommaso et al., 2017) to scale the pipeline up for high-throughput data analysis. For the purposes of this study, bioinformatics tools used in EAGER v2 included FastQC v0.11.9 (Andrews, 2018), AdapterRemoval v2.3.2 using default parameters (Schubert et al., 2016), and the Burrows-Wheeler Aligner (BWA) mem v2.2.1 (Li, 2013) for mapping against the *M. leprae* TN reference strain (NCBI NC_002677.1). Filtering of mapped files in EAGER v2 was completed with SAMtools v1.20 (Li et al., 2009) and the BamUtil v1.01.15 (Breese & Liu, 2013) clipOverlap function, which removes potential overlap between paired reads, which reduces the bias in genotype likelihoods. Coverage and mapping quality estimates will be generated using QualiMap v2.2.1 (Okonechnikov et al., 2016) and damage patterns were ascertained with MapDamage v2.2.1 (Jónsson et al. 2013). Variants were called within EAGER 2 through the GATK Unified Genotyper v3.5 (Van der Auwera & O'Connor, 2020).

For comparative purposes, we used publicly available genomes from modern and ancient *M. leprae* sequences on the NCBI Sequence Read Archive (SRA) (Supplementary Table 9). Raw fastq data were downloaded using SRA Toolkit v2.8.0 (https://github.com/ncbi/sra-tools) fastqdump command and processed through the same workflow as the novel sequences we generated. For finished *M. leprae* genomes, fasta

files were aligned against the reference *M. leprae* TN strain and BAM files were generated with default parameters with LAST (Kielbasa et al., 2013). The BAM files were then processed through EAGER v2 with other comparative data.

A concatenated SNP alignment was generated from VCF data using MultiVCFanalyzer v0.85.2 (Bos et al., 2014) with a minimum genotyping quality of 30, a minimum coverage for base calls of 5x, and minimum allele frequencies for homozygous calls and heterozygous calls of 0.8 respectively, as per recommendations by Benjak et al., (2018). MultiVCFanalyzer was also used to filter out insertions and deletions, as well as positions in known repeat regions, rRNA, and positions covered by the SK12 negative control sample often included in *M. leprae* comparative data (Honap et al., 2018). Finally, the concatenated SNP alignment was run through SnpEff v4.3.1t (Cingolani et al. 2013) with the *M. leprae* TN annotation as a reference to annotate variants and determine functional effects.

Phylogenetic Analysis

To examine potential geographic associations with strains, we built phylogenetic trees with several methods. A maximum likelihood (ML) tree was constructed using RAxML v8.2.12 (Stamatakis, 2014) with a TVM-GAMMA model and 1000 bootstrap replicates. This model was selected based on the modeltest function within RAxML to determine the best fit model for our data. We examined the robustness of this topology by constructing a neighbor-joining tree in MEGA11 (Tamura et al. 2021). We estimated the time to the Most Recent Common Ancestor (tMRCA) and substitution rates through the application of Bayesian methods using BEAST v1.10.4 (Suchard et al., 2018). SNPs from

known hypermutator strains (Benjak et al., 2018) were excluded from the BEAST analysis. Invariant sites were manually included in the .xml file to avoid ascertainment bias of only using variant sites. These sites were determined through the number of sites shared fully by all comparative strains and the reference. We implemented generalized stepping-stone sampling (GSS) to infer a best-fit model of nucleotide substitution (Baele et al., 2016). Based on this, we used an HKY nucleotide substitution model with a relaxed clock. The alignment was also run using an GTR nucleotide substitution model with a relaxed clock to determine the robustness of data. Calibrated radiocarbon dates were used for ancient strain tip dates and sampling/isolation dates were used for modern strain tip dates.

Spatial Analysis

To explore the distribution of cases, we implemented a Bayesian discrete phylogenetic approach. A discrete space is appropriate for when samples are clustered based on geographic location in which spread within a unit is freer and more spread between units may be hindered by geographical barriers (*e.g*., Pacific Ocean). This approach was run through BEAST v1.10.4 (Suchard et al., 2018) with the addition of sequence collection location as part of the trait input. The nucleotide substitution model was selected using previously determined inferences for the dating analysis. We specified a symmetric substitution model which allows for a discrete state for ancestral reconstruction. Results were visualized in the software SpreaD3, or Spatial Phylogenetic Reconstruction of EvolutionAry Dynamics using Data Driven Documents (Bielejec et al., 2016), which visualizes the output of Bayesian phylogeographic analyses.

3.4 Results

Sample Screening and Sequencing Results

79 FFPE tissue samples underwent DNA extraction, with 20 paired samples undergoing two DNA extractions for comparative purposes. DNA concentrations ranged from 0.372 ng/µl to 56.5 ng/µl. Further DNA extraction concentrations, qPCR screening, and sample selection for library preparation are described in Supplementary Table 2.

Twelve novel *M. leprae* genomes were sequenced from 11 individuals with mean coverage ranging from 1.05x - 25.97x (Table 1). The sequences were isolated from FFPE tissue samples collected between 1999 - 2015 from individuals from Chuuk State (n=1), Guam (n=3), Hawaii (n=5), and Palau (n=2). HL15 and HL18 are time-series samples collected three months apart from the same individual during treatment. Branch specific SNPs were checked manually using IGV and a table with branch-specific SNPs from Schuenemann et al. (2013).

Two low coverage genomes (HL27 with 2.1x and HL53 with 2.58x) have an undetermined SNP type and were removed from further analysis as they were too poor quality to confidently determine branch placement, use for dating analyses, or scan for the presence of SNPs associated with drug resistance. Phylogenetic analyses show that isolates HL15 (Hawaii), HL18 (Hawaii), and HL65 (Palau) fall into branch 0. Samples HL26 (Chuuk State), HL39d (Hawaii), HL56 (Guam), HL57 (Guam), HL73 (Palau), and HL96 (Guam) fall into branch 5. The HL98 strain falls into branch 1 and is the first from Hawaii to be placed into this branch (Figure 1).

Table 1: Sample metadata and sequencing results for 12 novel *M. leprae* genomes isolated from 11 individuals. The pathology type is listed per Ridley & Jopling (1966) classification. Mapping statistics completed with QualiMap v2.2.1 (Okonechnikov et al., 2016).

Sample	Date	Location	Type	DNA Extraction Method	DNA Concentration (ng/ul)	RLEP Ct Avg.	85B Ct Avg.	Branch Placement	Mean Read Length (bp)	Mean Coverage	Percent Genome covered >1x	Percent Genome covered >5x	% Endogenous DNA	# SNPS Called
HL15	1999	Honolulu, Hawai'i	LL	DNeasy Blood and Tissue Kit (Qiagen)	3.46	28.20	29.34	0	64.38	5.3x	93.44	53.32	97.58	83
HL18	1999	Waianae, Hawai'i	LL	DNeasy Blood and Tissue Kit (Qiagen)	70	26.12	28.02	0	118.22	5.38x	85.08	43.44	88.63	108
HL26	2002	Weno, Chuuk State	LLp	DNeasy Blood and Tissue Kit (Qiagen)	0.212	30.18	33.13	5	71.66	1.49x	57.49	7.49	98.23	12
HL27	2003	Waianae, Hawai'i	LL	DNeasy Blood and Tissue Kit (Qiagen)	0.212	29.48	30.58	Undetermined	66.85	2.1x	77.78	11.29	98.78	4
HL39d	2006	Honolulu, Hawai'i	LLs	DNeasy Blood and Tissue Kit (Qiagen)	1.48	25.15	26.90	5	73.21	8.87x	97.02	81.39	92.09	125
HL53	2013	Waipahu, Hawai'i	LL	DNeasy Blood and Tissue Kit (Qiagen)	6.26	35.59	34.29	Undetermined	56.38	2.58x	80.89	17.64	99.29	9
HL56	2015	Dededo, Guam	LL	DNeasy Blood and Tissue Kit (Qiagen)	3.94	31.27	30.04	5	68.65	13.8x	96.87	85.11	99.43	128
HL57	2015	Hagâtña, Guam	LL	DNeasy Blood and Tissue Kit (Qiagen)	2.52	35.53	34.72	5	57.12	1.05x	59.96	1.19	99.51	0
HL65	2002	Koror, Palau	LLp	QIAamp FFPE DNA Kit (Qiagen) HA Phenol-Chloroform (Campos & Gilbert, 2014)	2.04 7.88	29.48 21.06	35.85 28.19	5	69.80	12.16x	96.85	87.89	76.82	186
HL73	2006	Koror, Palau	LL	QIAamp FFPE DNA Kit (Qiagen) HA Phenol-Chloroform (Campos & Gilbert, 2014)	0.362 8	30.43 21.87	35.96 29.05	5	74.13	25.97x	97.46	97.17	53.36	152
HL96	2013	Tamuning, Guam	BT	QIAamp FFPE DNA Kit (Qiagen) HA Phenol-Chloroform (Campos & Gilbert, 2014)	1.42 3.67	34.24 31.98	38.19 38.91	5	65.75	9.23x	97.10	83.09	29.43	132
HL98	1999	Honolulu, Hawai'i	LL	QIAamp FFPE DNA Kit (Qiagen) HA Phenol-Chloroform (Campos & Gilbert, 2014)	15.3 5.8	26.31 33.47	31.80 45.83	1	67.71	14.51x	97.15	91.51	43.99	57

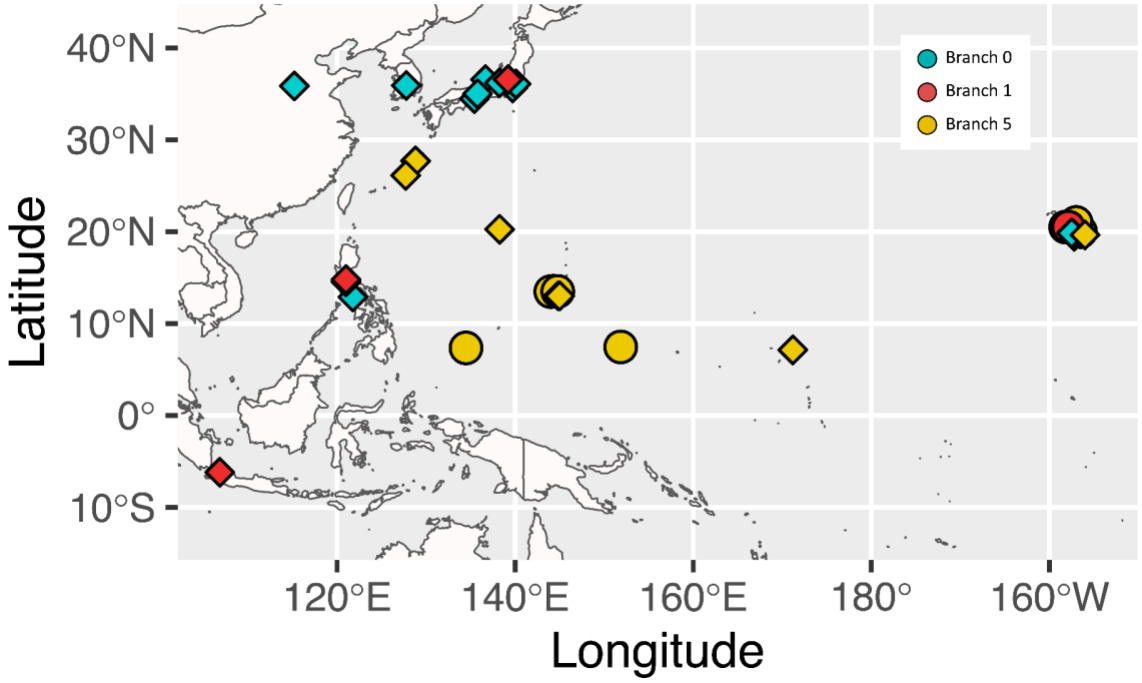


Figure 1: Map of novel strains and comparative strains from Pacific Islands or East Asia in branch 0, branch 5, and branch 1 visualized with ggplot2 (Wickham, H., 2016) in R (R Core Team, 2023).

Phylogenetic and Dating Analysis

A maximum-likelihood tree was generated from a SNP alignment using 4002 sites and 253 comparative strains with a bootstrap test of phylogeny with 1000 replicates (Figure 2a). RAxML's model testing function was used to determine the best-fit model for this alignment, and the analysis was then run with TVM-GAMMA for this tree with *M. lepromatosis* as an outgroup. Only strains with a mean coverage of >5x were used for phylogenetic analyses. To determine the robustness of the alignment, a neighbor-joining tree was also generated in MEGA11, which supported our maximum-likelihood

phylogeny (Supplementary Figure 6). The phylogenetic results supported the branch placement of the 10 genomes based on branch-informative SNPs as previously described. HL15 and HL18, both isolated from the same individual, clustered together in the Pacific Island subclade in branch 0. HL65, HL73, HL39d, HL96, and HL56 cluster in branch 5. HL98 is clustered with strains Ph15, Ph17, and the *M. leprae* reference strain TN.

BEAST v1.10.4 (Suchard et al., 2018) was used to estimate tMRCA with an alignment of 3496 sites and 249 strains, omitting known hypermutator strains (Benjak et al., 2018). GSS analysis supported the use of HKY as a best-fit model of nucleotide substitution. HL18 was removed from this analysis as its inclusion was found to act as a hypermutator strain, which skews the estimated tMRCA and substitution rate (see discussion in Benjak et al., 2018). The analyses were run with 300,000,000 interactions and a 30,000,000 burn-in. Calibrated tree annotation (Figure 2b) shows tMRCA to be 4453 y BP for *M. leprae* (3575 y BP - 5188 y BP 95% Highest Posterior Density [HPD]), 1297.5 y BP for the tMRCA of branch 1, 3998.8 y BP for branch 0, and 3329.8 y BP for branch 5. The substitution rate was estimated to be 7.87×10^{-9} (5.9×10^{-9} – 9.84×10^{-9} 95% HDP) substitutions per site per year.

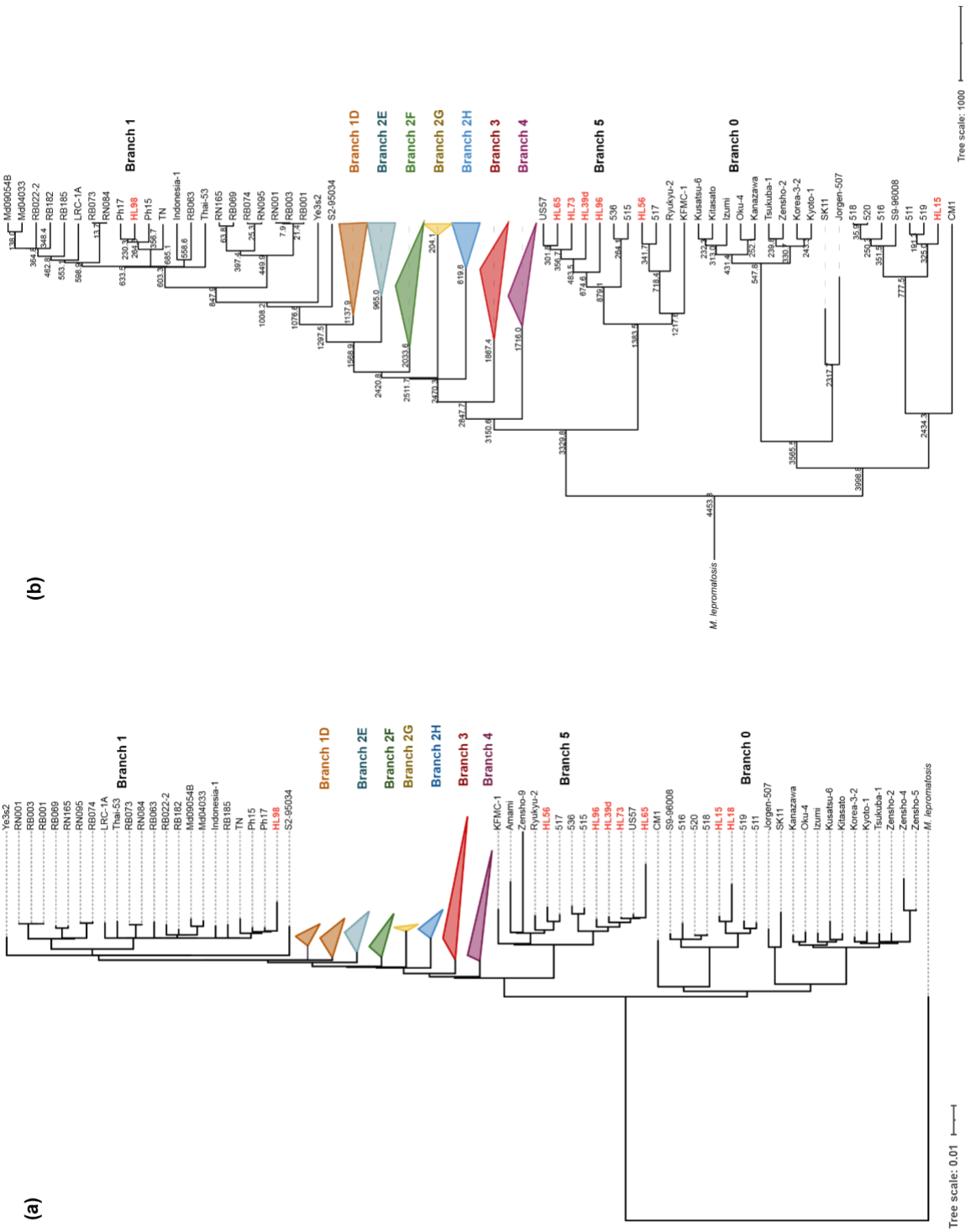


Figure 2: (a) Maximum-likelihood tree generated in RAxML v8.2.12 (Stamatakis, 2014) with 253 comparative strains, 4002 informative sites, and a bootstrap test of phylogeny with 1000 replicates. Clades without novel strains included in this study collapsed. The

tree was run with a TVM-GAMMA best-fit model with *M. lepromatosis* as an outgroup. (b) Bayesian phylogenetic tree based on 3496 informative SNPs as well as shared invariant sites calculated with BEAST v1.10.4 (Suchard et al., 2018). Median divergence times before present (BP) shown on main branch nodes. Novel strains from this study are shown bold and in red.

Spatial Analysis

To visualize the distribution of isolates through time, we implemented a Bayesian discrete phylogenetic approach, which used the phylogenetic tree generated for dating analyses through BEAST v1.10.4 (Suchard et al., 2018). A KML file (Keyhole Markup Language) was generated in SpreaD3 (Bielejec et al., 2016) with latitudes and longitudes used at the country or state level. Resolution at the city level was lost, but the results were visualized in R (R Core Team, 2023) with ggplot2 (Wickham, H., 2016). From these results, hotspots of leprosy spread near the Pacific are seen in Hawaii, Guam, and Bangladesh (Figure 3). Although directionality is not determined, direct movement of *M. leprae* strains is identified between Hawaii and Japan, Guam, Samoa, and New Caledonia. Furthermore, transmission is identified from Hawaii to Hungary, which shows the connection between branch 0 strains in Hawaii and the medieval Hungarian strains SK11 (Schuenemann et al., 2018), which also falls in branch 0. In the Pacific, the Marshall Islands, New Caledonia, and Samoa have limited connections to the rest of this network, while Guam and Hawaii are connected to multiple islands in the Pacific and regions in East Asia.

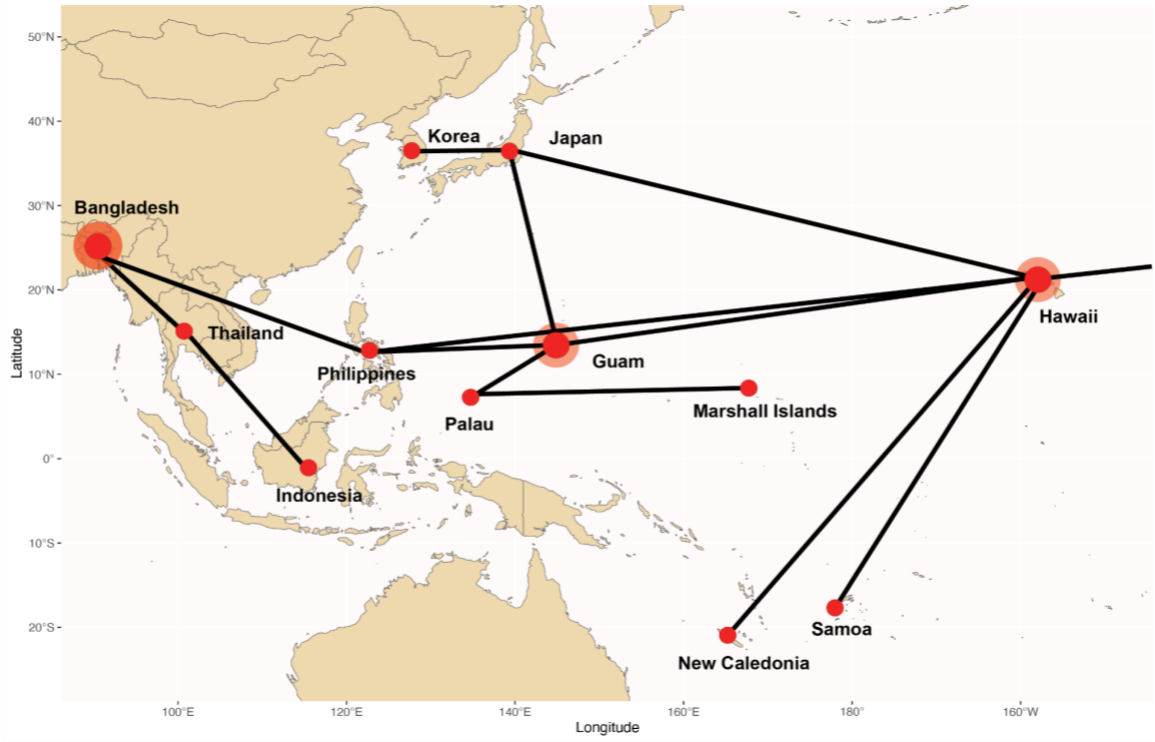


Figure 3: Transmission map using BEAST v1.10.4 (Suchard et al., 2018) phylogenetic tree (see dating analyses) and SpreaD3 (Bielejec et al., 2016). The map shows locations of *M. leprae* isolates with black lines indicating transmission but not direction. Large circles indicate sites with many points of potential transmission to other regions.

In the analysis of functional effects, we used SnpEff to scan for nonsynonymous and synonymous SNPs in coding regions and focused on a list of 9 genes specifically associated with drug resistance (Benjak et al., 2018). All 12 genomes contained between 2-36 nonsynonymous SNPs in coding regions (Table 2), and each isolate except for HL98 contained SNPs in regions associated with drug resistance. All *gyrA* mutations, which were present in HL15, HL18, HL39d, HL53, HL73 and HL96, were nonsynonymous. For the time-series isolates HL15 and HL18, the latter contains 10 more nonsynonymous

SNPs, including one in the *rpoB* region. HL18 was collected three months after HL15, during the course of the patient's treatment. None of the Pacific strains contained any SNPs in the regions *fadD9*, *ribD*, *ethA*, or *pks4*, and HL57 lacked any SNPs in any of the regions listed in Benjak et al. (2018).

Table 2: Presence of SNPs in regions associated with drug-resistance (Benjak et al., 2018) with synonymous SNPs shown by a black "X" and nonsynonymous SNPs shown by a red "X".

Sample	# Nonsynonymous SNPs	*folP1*	*rpoB*	*gyrA*	*gyrB*	*fadD9*	*nth*
HL15	26	X		**X**	X		
HL18	36		**X**	**X**	X		X
HL26	4	X					
HL27	2				X		
HL39d	3	X		**X**	X		
HL53	10					X	
HL56	29	X	X	**X**	X		
HL57	5						
HL65	10	X				X	
HL73	33	X	X	**X**	X		
HL96	27	X	X	**X**	X		
HL98	11						

3.5 Discussion

Our study aimed to explore the phylogenetic and geographic distribution of *M. leprae* in the Pacific, leveraging newly sequenced and published genomes to characterize transmission patterns and genetic diversity. We generated 12 novel *M. leprae* genomes

from FFPE tissue samples collected from leprosy patients during treatment in Chuuk State, Guam, Hawaii, and Palau (1999-2015). Of that number, 10 genomes were suited for further phylogenetic, dating, and spatial analyses. This work not only adds valuable genomic data to the limited pool of *M. leprae* sequences from the Pacific but also provides insights into the historical and recent spread of leprosy in this understudied area.

Phylogenetic and Dating Analysis

The phylogenetic analysis revealed that the novel genomes predominantly belong to branches 0 and 5, with a single genome (HL98 from Hawaii) being placed into branch 1. These results align with previous findings that suggest a strong geographic association of *M. leprae* lineages, in which branch 0 is typically found in Eastern Asia and the Pacific, and branch 5 is primarily found in the Pacific Islands (Monot et al., 2009; Schuenemann et al., 2018). HL26 is the first *M. leprae* whole genome isolated from Chuuk State. These data nearly double the known whole genome sequences from the Pacific Islands, which include strains isolated between 1996 - 2015 from the Marshall Islands (Benjak et al., 2018), New Caledonia (Schuenemann et al., 2013), and genomes generated by Blevins et al., (2020) from Guam, Hawaii, the North Mariana Islands, Palau, and Samoa. Finally, Honap et al. (2018) sequenced an isolate CM1 from a naturally infected cynomolgus macaque (*Macaca fascicularis*) in the Philippines, which is to date the only animal-associated strain in branch 0.

The tMRCA for *M. leprae* strains was estimated to be 4453 y BP for *M. leprae* (3575 y BP - 5188 y BP 95% HPD) with a relaxed lognormal clock and an HKY substitution model (Figure 2b). This estimate is earlier than previous estimates of

tMRCA, but the 95% HPD ranges overlap (Hockings et al., 2021; Blevins et al., 2020; Schuenemann et al., 2018). The substitution rate was estimated to be 7.87×10^{-9} (5.9×10^{-9} – 9.84×10^{-9} 95% HDP) substitutions per site per year, which is comparable to previous estimates (Schuenemann et al., 2018; Benjak et al., 2018; Honap et al., 2018). The similarity in estimates and overlapping 95% HPD ranges suggest that this alignment is robust; however, an earlier tMRCA may be due to the addition of several novel strains with long branch lengths due to the presence of drug-resistance associated SNPS. These isolates with potentially highly mutated genes associated with drug resistance could interfere with the clock models in Bayesian inference and therefore tMRCA estimation (Benjak et al., 2018).

The presence of branch 1 in Hawaii, a lineage more commonly associated with Southern and Southeast Asia (Monot et al., 2009) may indicate either recent travel of the individual to those areas or a more historic introduction of branch 1 strains to Hawaii through earlier migration or trade interaction. The dating results suggest that HL98 diverged from its most recent common ancestor with Ph15 in 230.3 y BP and the slightly more diverged Ph17 in 264.3 y BP. These strains were isolated from the Philippines between 1990 - 2008 (Avanzi et al., 2020a), whereas HL98 was collected in 1999 in Hawaii. The divergence time suggests that the strain was brought over to Hawaii in the 1700s, possibly due to migration to the island or due to trade. Alternatively, without the presence of other branch 1 genomes circulating in Hawaii, this strain may be due to recent travel history of the patient. The demographic information we have for the patients in treatment was limited, however the host patient of HL98 is known to be a Hawaiian resident and part of the local Filipino community. We do not have travel information,

however the hypothesis of recent travel to the Philippines as a source of origin for HL98 is better supported by known patient information. This the first branch 1 genome isolated from Hawaii, but branch 1 strains have also been identified in other areas of the Pacific and Southeast Asia, such as New Caledonia (Monot et al., 2005), Indonesia (Benjak et al., 2018), the Philippines (Avanzi et al., 2020a), and Japan (Singh et al., 2014).

Recent and Historic Spread of Leprosy

Public health campaigns and effective multidrug therapy implementation in the Pacific Island countries have reduced prevalence of leprosy to below the elimination threshold (WHO, 2019), which means that it is no longer considered endemic in the region. Despite this, leprosy care is still impaired by the stigma surrounding the disease and a lack of rapid diagnostics at early disease stages (Avanzi et al., 2020c). The past decade has seen fluctuations in the rates of new cases in the Pacific Islands, which could be attributed to targeted external case identification and SNP typing (Rahevar et al., 2012; Reibel et al., 2015) or to the general decrease in reporting during the COVID-19 pandemic (WHO, 2023).

The antiquity of the disease in the region was explored in Blevins et al., (2020), which found that *M. leprae* likely spread with the initial peopling of the Pacific and in subsequent waves, from 3000 BP to 2300 BP (Posth et al., 2018; Lipson et al., 2018). The presence of branch 0 and 5 strains across multiple geographically distant Pacific locations suggests that early migrations introduced distinct and multiple *M. leprae* strains, which spread and experienced more recent radiation within and among islands in the past three centuries (Blevins et al., 2020). Although this sheds some light on broader transmission

of the disease, further inferences of ancient strains are difficult as there is very limited skeletal evidence in the region (Trembly, 1995).

The 19th century saw a shift in how leprosy was seen in the Pacific. European and American colonial presence in the Pacific Islands brought leprosy to the forefront of contemporary public health measures and shaped the governance of colonies. This focus on leprosy and the timing of other disease introductions by Europeans led to modern historians tracing the presence of the disease in the Pacific to the late 18th or early 19th century (Luker & Buckingham, 2017; Inglis, 2014). In 1865, the government of the Kingdom of Hawai'i instituted quarantine for individuals infected with leprosy, and by the overthrow of the monarchy in 1893 and later annexation of Hawaii by the US in 1898, over 5000 people with leprosy had been exiled to the island of Moloka'i (Amundson & Oakaokalani Ruddle-Miyamato, 2010). In 1938, an injectable version of the traditional ayurvedic remedy chaulmoogra oil (as designed by Alice Ball) was recommended for use in leprosy treatment (Inglis, 2014). This treatment had limited success for the symptoms, but the implementation of dapsone for treatment in the 1940s significantly improved patient outcome (Inglis, 2014). By 1969, the quarantine laws were abolished for Molokai exile, and multidrug therapy (MDT) was recommended by the World Health Organization in 1983 (WHO, 2016; Amundson & Oakaokalani Ruddle-Miyamato, 2010). Such quarantining and exile were seen in other islands of the Pacific in the late 19th century and early 20th century (Luker & Buckingham, 2017), which has interesting implications for the evolution of *M. leprae* during this time. In addition to *M. leprae* strains largely endemic to the Pacific, input of novel European strains and trade between colonies may have brought in more diverse strains.

Although we are unable to infer within-island nuance with these data, among-island interactions have been characterized to a limited extent. Within the Pacific, Hawaii and Guam seem to act as important regions for transmission towards other islands (Figure 3). This analysis also shows interesting spatial clustering with Hawaii, Samoa, and New Caledonia, as well as between Palau and the Marshall Islands. Despite branch 5 strains identified in all locations sampled, the Marshall Islands, New Caledonia, and Samoa act as very limited origin points for *M. leprae* strains. Although the Marshall Islands and Samoa are more remote, New Caledonia has been identified as an area with diverse *M. leprae* strains among humans (Monot et al., 2005; Schuenemann et al., 2013). This discrepancy may be due to more migration of strains into New Caledonia and a longer time circulating in that region. An important caveat to this map is that strains from Bangladesh are overrepresented in the *M. leprae* phylogeny, and within the Pacific Island genomes, strains from Guam and Hawaii also outnumber those from other islands. This drawback is part of a larger issue of using phylogenetic data to infer epidemiological relationships. Although we may infer disease hotspots to a degree, overrepresentation of strains makes it difficult to determine which locations may act as points of disease transmission and which locations are overrepresented. Despite this limitation, phylogeographic methods are useful for a broader view of strain relationships.

Implications for Public Health and Future Research

The identification of multiple strains within a single individual or potential rapid evolution in a single strain (HL15 and HL18) over a short time span highlights the dynamic nature of *M. leprae* infections and ongoing intrahost evolution. There were 10

additional nonsynonymous SNPs present in HLP18, one of which was present in *rpoB*, which confers rifampin resistance (Williams & Gillis, 2004). This underscores the importance of monitoring genomic changes in clinical settings, particularly for understanding drug resistance and treatment efficacy. Additionally, the detection of SNPs associated with drug resistance in the novel genomes, especially those documented in *gyrA* in six newly sequenced isolates, highlights the potential for circulation of drug resistant strains over a wide expanse of the Pacific. Despite better documentation of drug resistant cases and associated SNPs in *M. leprae* genomes, there is very little work on resistance to second line drugs (Avanzi et al., 2020c) and the role drug resistance plays in relapse cases (Cambau et al., 2018). Ultimately, we call for further surveillance in the Pacific and a push for integrating whole-genome sequencing, historical context, and a phylogeographic approach to address leprosy treatment and eradication.

3.6 Conclusion

Leprosy care is still significantly limited by stigmatization of infected individuals, lack of resources in certain regions to diagnose or treat, and an incomplete view of global genomic diversity (Avanzi et al., 2020c). This study demonstrates the importance of integrating whole-genome sequencing and phylogeographic approaches to understand evolutionary history of *M. leprae*. These efforts also clarify how human migration, travel, and even policy may affect pathogen evolution. Further work that combines genomics and a historical approach is necessary to understand the nuances of human-pathogen co-evolution.

CHAPTER 4

ISOLATION OF LOW COVERAGE *MYCOBACTERIUM LEPRAE* GENOME FROM *PUMA CONCOLOR* IN BRAZIL

4.1 Abstract

Brazil has the second highest incidence of new cases of leprosy reported each year, with the highest number of relapse cases. Ongoing surveillance has identified pockets of hyperendemicity, which could act as hotspots for the spread of drug-resistance strains. Despite this, there have been limited surveys for leprosy in animals in the region. Most research in Brazil on animal-associated *M. leprae*, one of the causative agents of leprosy, focuses on armadillos. Here, we present a low coverage *M. leprae* genome isolated from a puma (*Puma concolor*) in the state of Mato Grosso, Brazil. Phylogenetic analysis suggests that strains circulating in local human populations are also potentially transferring to animal populations. Although this finding does not clarify the transmission route of this case or leprosy more broadly, it strongly supports both the need to increase surveillance of animals in endemic regions and a larger host range for *M. leprae*.

4.2 Introduction

Leprosy (Hansen's disease) is a chronic infectious disease that has long been recognized as a significant public health concern. Despite a primary research focus on the disease in humans, a growing body of literature on animal-associated leprosy raises intriguing questions about zoonotic reservoirs, spillover concerns, and new approaches to eradication (Avanzi et al., 2020c; Urban et al., 2021).

Leprosy in humans is largely caused by the obligate, intracellular bacterium *Mycobacterium leprae*, as well as the more recently discovered *M. lepromatosis* (Han et al., 2008; 2012), which is limited in geographic range to Mexico and the Caribbean. *M. leprae* and *M. lepromatosis* diverged around 13.9 Mya and have both experienced extensive reductive evolution and gene decay (Han et al., 2008; Cole et al., 2001). *M. leprae* has a small genome of ~3.27 Mb, and approximately half of the genome is made up of pseudogenes (Monot et al., 2005; 2009). Using genetic typing, *M. leprae* strains were initially categorized into four single nucleotide polymorphism (SNP) types (1-4) and 16 SNP subtypes (A-P) based on a series of informative SNPs and variable number of tandem repeats (VNTRs) (Monot et al., 2005; Scheunemann et al., 2018). Whole genome sequencing has allowed for more comprehensive phylogenetic analyses, and six branches have been characterized that generally correspond to the SNP types and subtypes. *M. leprae* branches and SNP types have strong geographic associations: branches 0 and 5 were identified in Eastern Asia and the Pacific, branch 1 in Southern and Eastern Asia, the paraphyletic branch 2 in the Near East and South Asia, branch 3 in the Americas, and branch 4 in West Africa and South America (Blevins et al., 2020; Monot et al., 2009; Schuenemann et al., 2013; 2018).

M. leprae has been characterized as the causative agent of leprosy in naturally infected nine-banded armadillos (*Dasypus novemcinctus*) (Sharma et al., 2015; Truman et al., 2011), six-banded armadillos (*Ephractus sexcinctus*) (Frota et al., 2012), British red squirrels (*Sciurus vulgaris*) (Avanzi et al., 2016), captive non-human primates, including *Pan troglodytes*, *Macaca fasicularis*, and *Cercocebus atys* (Honap et al., 2018), and wild populations of western chimpanzees (*Pan troglodytes verus*) (Hockings et al., 2021). *M. lepromatosis* has also been isolated from British red squirrels (Avanzi et al., 2016), but has yet to be identified in naturally occurring cases in other animal species. In the southern United States, around two thirds of autochthonous human cases are armadillo associated strains (Sharma et al., 2015). Genomic analysis reveals that these animal strains correspond to the SNP types of human strains in same geographic region, with the interesting exception of *M. lepromatosis* isolated from red squirrels in the UK (Avanzi et al., 2016).

Animal-associated strains and spillover are of particular interest in countries like Brazil, which has the second highest incidence of new cases per year and pockets of hyperendemicity (>40 cases/100,000 people per year) (World Health Organization, 2023). Ongoing surveillance has focused on disease prevalence in the Amazonian states of Brazil, with recent increased rates found in the states of Mato Grosso and Tocantins (WHO, 2023, Maruyama et al., 2018; Montiero et al., 2015). Work examining disease trends in Brazil during the last several decades suggests that leprosy rates increased due to urbanization and deforestation, as well as the construction of the national highway in the 1970s and 1980s (Montiero et al., 2015; Schaub et al., 2020). As in the United States,

spillover from armadillos in Brazil into humans has been connected to hunting and food preparation (da Silva et al., 2018; Frota et al., 2012).

Further animal surveys in Brazil have yielded mixed results. Maruyama et al. (2018) ran PCR assays targeting the repeat RLEP region of *M. leprae* in 69 captive and free wild animals and found positive results in a captive margay (*Leopardus wiedii*), one captive and one free-ranging lowland tapir *(Tapirus terrestris*), free-ranging capuchin monkeys (*Sapajus apella*), and an owl monkey (*Aotus trivirgatus*). Similar surveys of armadillos and other mammals native to the region vary wildly in PCR-based identification of *M. leprae*, from 0% prevalence in endemic regions (Ploemacher et al., 2020; Araújo Stefani et al., 2019) to high prevalence in armadillos near infection hotspots (Cardona-Castro et al., 2009; Frota et al., 2012).

Given such a wide range of potential reservoirs and endemic areas, we sought to both expand the types of mammals examined and broaden the geographical range of study. In addition to including Brazil cases in our analyses, we also incorporated samples from the Amazonian side of Peru to capture more species in an area of potential spillover. To this end, we analyzed 108 mammals from the endemic state of Mato Grosso in Brazil and the Madre de Dios region of Peru. We hypothesized mammals from the state of Mato Grosso would have higher prevalence of *M. leprae* circulating than in Peru, and that those strains identified would be closely related to strains circulating in local human populations. If strains in animals reflect the diversity of strains also circulating in humans, we also expect to find multidrug resistance strains, which remain low globally for *M. leprae*, but are on the rise in Brazil (Rosa et al., 2020). Here, we present results of

this survey, which includes the first *M. leprae* genome isolated from a puma (*Puma concolor*).

4.3 Materials and Methods

Brazil Mammal Screening

Necropsies were completed by veterinary staff at the Federal University of Mato Grosso (Universidade Federal de Mato Grosso) on roadkill in the state of Mato Grosso, Brazil. Animals showing skin ulcers were selected for further analysis. Spleen tissue samples (0.5g) were homogenized in 4mL of lysis buffer with proteinase K. This mixture was incubated overnight at 56°C. A 1mL aliquot was then processed for phenol-chloroform extraction and resuspended in nuclease-free, molecular grade water. DNA extracts were purified with the PureLink PCR Purification Kit (Invitrogen). All extracts underwent PCR analysis for RLEP, a multicopy element in *M. leprae*, using protocols as recommended in Woods and Cole (1990).

In total, 14 DNA extracts that tested positive for RLEP at the Federal University of Mato Grosso were sent to the Molecular Laboratory of Anthropology at Arizona State University for further screening and processing. The DNA extracts processed at ASU came from a range of mammals (Table 1) and were quantified using a Qubit 3.0 Fluorometer (Invitrogen). All DNA extracts and 1:10 dilutions of the extracts were screened for *M. leprae* DNA using a TaqMan quantitative polymerase chain reaction (qPCR) assay. All DNA extractions, 1:10 dilutions of the extracts, and extraction blanks were screened for *M. leprae* DNA with a TaqMan qPCR assay designed to target two *M.*

leprae specific elements RLEP and 85B (Truman et al., 2008; Martinez et al., 2006). Samples were run in triplicate and those that had at least two replicates amplified moved on to library preparation.

Peru Mammal DNA Extraction and Screening

In the Madre de Dios region of Peru, 94 small mammals (including rodents, marsupials, and lagomorphs, see table X) were trapped for minimally invasive sampling of tissues and ectoparasite collection. At ASU, ear and tail biopsies from these animals underwent DNA extraction using the DNeasy Blood and Tissue kit (Qiagen). Following kit recommendations for mammal ear and tail samples, the proteinase K digestion time was increased to three hours. DNA extracts were quantified using a Qubit 3.0 Fluorometer (Invitrogen). Out of 94 extracts, 68 extracts had DNA concentrations >0.01 ng/µl and those underwent screening for *M. leprae* DNA using the above TaqMan qPCR assay for RLEP and 85B (Truman et al., 2008; Martinez et al., 2006). DNA extracts, 1:10 dilutions of extracts, and extraction blanks were screened in triplicate. Samples with amplification of at least two replicates during qPCR assays moved on to library preparation.

Library Preparation and Enrichment

Samples that passed qPCR screening were sheared using a Covaris (ASU Genomics Core) and converted into double-indexed libraries (Meyer & Kircher, 2010). Based on pre-library qPCR screening results, post-library qPCR quantification results, and fragment analysis using the Agilent 4200 TapeStation (ASU Genomics Core), we proceeded to targeted enrichment with 8 samples. The indexed libraries were enriched for

the *M. leprae* genome using an Arbor Biosciences myBaits kit v5 with biotinylated baits prepared from *M. leprae* strains Thai53 and NHDP (as per the recommended protocol). Capture products were then amplified using the Kapa Hifi HotStart ReadyMix (Roche). We sequenced the enriched libraries on the Illumina MiSeq at ASU's Genomics Core (Arizona, United States). After initial poor sequencing results but high complexity as determined for certain libraries with Preseq (Daley & Smith, 2013), we resequenced the enriched libraries on the Illumina NovaSeq X Plus (Admera Health, New Jersey, United States).

Data Analysis

Raw sequencing data were processed using EAGER v2, a bioinformatics best-practice analysis pipeline, originally developed for ancient DNA data analysis (Fellows Yates et al., 2021). The EAGER v2 pipeline uses Nextflow v22.10.6 (Di Tommaso et al., 2017), which is a workflow tool that connects a variety of bioinformatic tools and can scale for large-batch sequence analysis for data generated using different methods across different platforms. For the purposes of this study, EAGER v2 was used to visualize raw sequence quality with FastQC v0.11.9 (Andrews 2018), trim and merge sequences with AdapterRemoval v2.3.2 using default parameters (Schubert et al., 2016), and map to the *M. leprae* TN reference strain (NCBI NC_002677.1) using the Burrows-Wheeler Aligner (BWA) aln v0.7.17 (Li & Durbin, 2009). BAM files were filtered with SAMtools v1.20 (Li et al., 2009) and variants were called within EAGER v2 through the GATK Unified Genotyper v3.5 (Van der Auwera & O'Connor, 2020).

For comparative purposes, publicly available genomic data from modern and ancient *M. leprae* sequences on the NCBI Sequence Read Archive (SRA) were downloaded as raw fastq data using the SRA Toolkit v2.8.0 (https://github.com/ncbi/sra-tools) fastqdump command. These data were processed through the same pipeline as the novel sequences, except ancient genomes underwent extra processing to generate coverage and mapping quality estimates through QualiMap v2.2.1 (Okonechnikov et al., 2016). For finished *M. leprae* genomes, fasta files were aligned against the reference *M. leprae* TN strain and BAM files were generated with default parameters with LAST (Kielbasa et al., 2013). The BAM files were then processed through EAGER v2 with other comparative data.

VCFs for novel data and comparative data were merged using MultiVCFanalyzer v0.85.2 (Bos et al., 2014) with a minimum genotyping quality of 25, a minimum coverage for base calls of 1x, and minimum allele frequencies for homozygous calls and heterozygous calls of 0.8 respectively, as per recommendations by Benjak et al. (2018). MultiVCFanalyzer was also used to filter out insertions and deletions, as well as positions in known repeat regions, rRNA, and positions covered by the SK12 negative control sample often included in *M. leprae* comparative data (Honap et al., 2018).

SNP and Phylogenetic Analyses

Variant annotation and functional effect prediction were performed using SnpEff 5.2 (Cingolani et al., 2012). Variants were checked manually in IGV (Robinson et al, 2011) due to the low mean coverage of the genome. The results were cross-checked

against known drug-resistance associated regions in the *M. leprae* genome (Benjak et al., 2018).

A concatenated SNP alignment of 4123 informative sites was used to perform phylogenetic analyses in MEGA11 (Tamura et al., 2021). Evolutionary history was inferred using the neighbor-joining method (Saitou & Nei, 1987) with positions 80% site coverage or more included. The tree was run with a bootstrap test of phylogeny of 1000 replicates. This alignment was also used to generate a pairwise distance matrix with strains isolated in Brazil using MEGA11. This analysis was completed using a bootstrap test with 1000 replicates and a GAMMA model of rate heterogeneity. To determine robustness of the topology, phylogenetic analyses were performed in MEGA11 with minimum-evolution methods.

4.4 Results

DNA Extraction and Sample Screening

Brazil mammal DNA extracts were quantified at ASU and had a large range in concentrations (0.01-94.2 ng/µl). Given the low number of samples and positive PCR assays performed by the veterinary team in Brazil, 8 samples that had at least one replicate pass for either the RLEP or 85B qPCR assay were chosen for library preparation (Table 1). We successfully extracted DNA from 68/94 samples from Peruvian mammals with a quantification range of 2.16-30 ng/µl. None of these DNA extracts tested positive for the RLEP or 85B qPCR assays for any replicate and were removed from further analyses (Supplementary Table 10).

Table 1: Sample metadata, DNA extraction information, and qPCR screening results for the Brazil mammal samples. DNA concentrations of <0.001 ng/µl fell below the lower Qubit limit.

Sample ID	Host	Location	DNA Concentration (ng/uL)	RLEP Avg. Ct	85B Avg. Ct
Br2021-1	South American coati (*Nasua nasua*)	Cuiabá, Brazil	2.08	n/a	14
Br2021-2	Brazilian guinea pig (*Cavia aperea*)	Barrão de Melgaço, Brazil	4.18	n/a	n/a
Br2021-3	Black-tailed marmoset (*Mico melanurus*)	Cuiabá, Brazil	2.94	n/a	16
Br2021-4	Capybara (*Hydrochoerus hydrochaeris*)	Cuiabá, Brazil	<0.001	n/a	n/a
Br2021-5	Gray brocket (*Mazama gouazoubira*)	Cuiabá, Brazil	2.12	n/a	26
Br2021-6	Southern tamandua (*Tamandua tetradactyla*)	Cuiabá, Brazil	5.9	40	n/a
Br2021-7	Puma (Puma concolor)	Cuiabá, Brazil	15.7	n/a	23*
Br2021-8	Black-tailed marmoset (*Mico melanurus*)	Cuiabá, Brazil	94.2	28*	44
Br2021-9	Six-banded armadillo (*Euphractus sexcinctus*)	Chapada dos Guimarães, Brazil	<0.001	38	32
Br2021-10	Azara's night monkey (*Aotus infulatus*)	Cuiabá, Brazil	<0.001	28*	33*
Br2021-11	Southern tamandua (*Tamandua tetradactyla*)	Cuiabá, Brazil	6.78	30	22
Br2021-12	Puma (*Puma concolor*)	Cuiabá, Brazil	<0.001	n/a	n/a
Br2021-13	Crab-eating fox (*Cerdocyon thous*)	Poconé, Brazil	3.72	n/a	14
Br2021-14	Tayra (*Eira barbara*)	Cuiabá, Brazil	6.78	n/a	n/a

Denotes at least 2/3 replicates amplifying during qPCR assay

Sequencing Results

We recovered one low coverage genome (1.3x, "Br2021-7") with ~44% of the genome covered >1x, a mean length of mapped reads of 236bp, and a high number of duplicate reads from a puma from Mato Grosso, Brazil. We analyzed library complexity on all samples sequenced, and Br2021-7 has a highly complex library, which suggests that re-sequencing may provide more informative sequence data for this sample.

Given the novel host and lack of serological data, we sought to verify the host of this isolated strain to ensure there was no contamination from within the lab environment. In addition to aligning sequence data against *M. leprae* TN, the data were also aligned

using the EAGER v2 pipeline against the reference genome for *P. concolor* mtDNA (GCF_003327715.1) and *Homo sapiens* mtDNA (GCF_000001405.40). These reference genomes were selected because the most likely source of contamination was *M. leprae* strains from human hosts, and mtDNA analysis would better identify a host from *M. leprae*-enriched samples. The *P. concolor* mtDNA alignment resulted in a mean genome coverage of 4.03x, whereas the *Homo sapiens* mtDNA alignment resulted in a mean genome coverage of 0.017x. Metagenomic screening using Kraken2 (Wood et al., 2019) found that most unmapped reads cluster with the *M. leprae* Kyoto2 strain (GCF_003584725.1) and the reference *M. leprae* TN strain (NCBI NC_002677.1).

Veterinary Pathology Results

The *P. concolor* host was an adult male (29.7kg) that was found on a federal highway in the Tangara da Serra municipality in Mato Grosso, Brazil. The cause of death was attributed to six bullet wounds (three in the head and three in the abdomen). Multiple skin ulcers were identified, but unfortunately, skin was not collected for histological analysis. Tissue samples were collected from the brain, lung, liver, heart, spleen, and kidney for further analysis.

SNP and Phylogenetic Results

Branch-specific SNPs as described in Scheunemann et al. (2018) were manually checked in IGV (Robinson et al., 2011) with limited success. The low coverage meant that there were no reads covering certain informative sites. Despite this drawback, we determined that Br2021-7 did not fall into branches 1 or 2 based on branch-specific SNPs. A concatenated SNP alignment of 4123 informative sites was used to perform

phylogenetic analyses in MEGA11 (Tamura et al., 2021). Evolutionary history was inferred using the neighbor-joining method (Saitou & Nei, 1987) with positions 80% site coverage or more included (Figure 1). The tree was run with a bootstrap test of phylogeny of 1000 replicates. This alignment was also used to generate a pairwise distance matrix with strains isolated in Brazil using MEGA11 with a bootstrap test with 1000 replicates, a GAMMA model of rate heterogeneity, and elimination of any site with less than 80% covered (Supplementary Figure 1).

Based on these results, Br2021-7 falls within Branch 4, which hosts strains from west Africa and South America. We were unable to resolve the SNP type of Br2021-7, so further strain clustering was determined with the distance matrix (Supplementary Figure 7). Br2021-7 clustered closely with Br2016-141 (SNP type 1D), Br2016-16 (SNP type 4N), and Br2016-45 (3I-1), which fall in branches 1, 4 and 3 respectively.

SNP analysis with SNPeff identified three variants in regions associated with drug resistance as defined in Benjak et al. (2018). The regions *rpoB* and *gyrA* both contain nonsynonymous variants, but the variant identified for *gyrB* is synonymous.

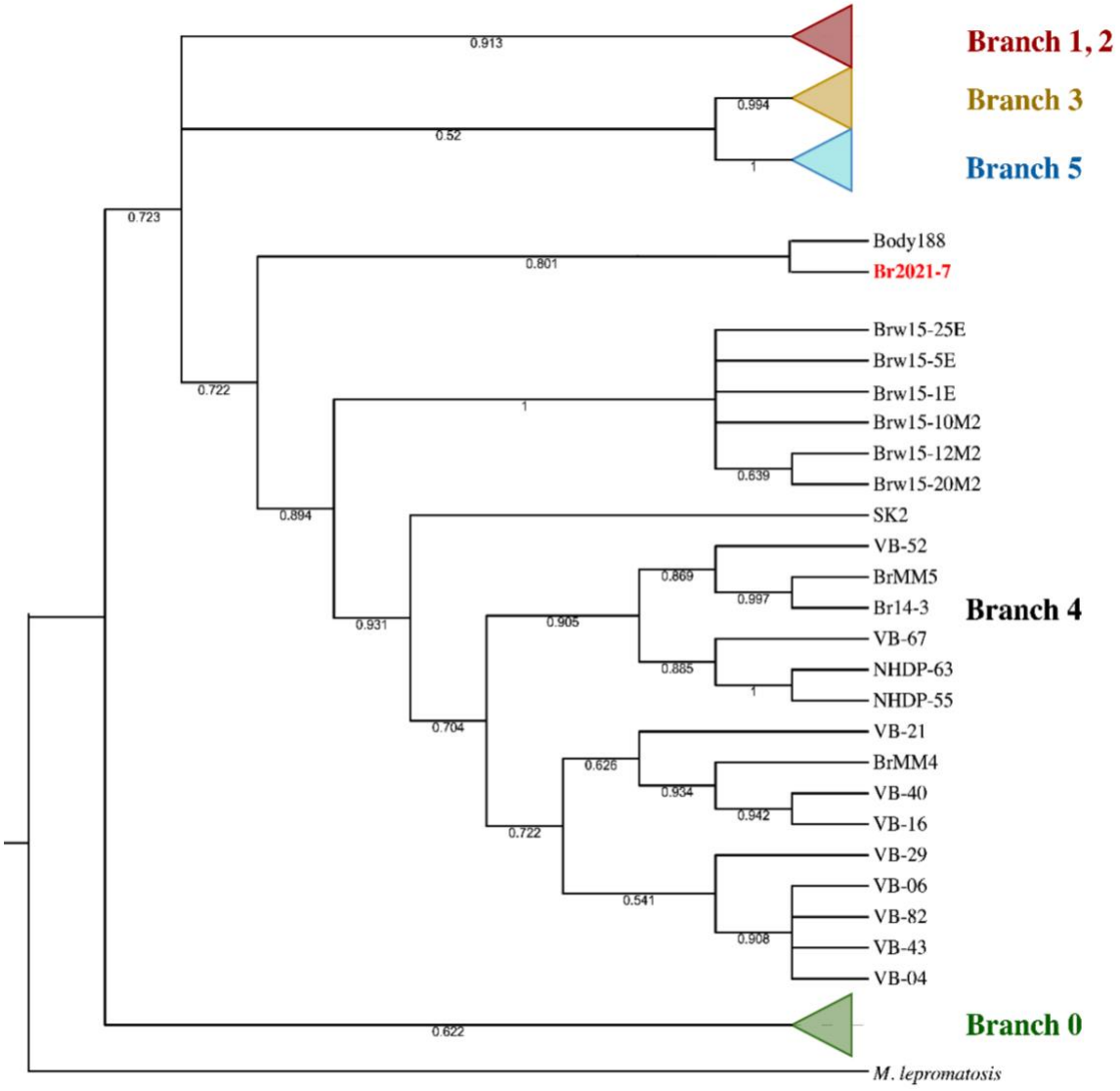


Figure 1: Neighbor-joining tree generated in MEGA11 (Tamura et al., 2021) with alignment of 4123 informative SNPs and 80% site coverage or more included. A bootstrap test of 1000 replicates was performed and results shown out of 1 on branches.

4.5 Discussion

Br2021-7 is the first *M. leprae* genome isolated from a puma or any large cat species. The low coverage of 1.3x but high library complexity suggests that this sample should be re-sequenced. Despite the poor quality of the genome, we could place the strain in branch 4, which contains strains isolated from West Africa and South America. *M. leprae* genomes in Brazil reflect the country's and the continent's history of slave trade, imported laborers, and colonization, with strains largely clustering into branches 3 and 4, as well as sporadically into branch 1 (Fontes et al., 2017; 2009).

In a pairwise genetic distance analysis, Br2021-7 was closer to three Brazil stains that fell into branches 1, 3, and 4 (Supplementary Figure 1). This result reflects both the expected strain type in the country, as well as our difficulty in placing Br2021-7 due to low mean coverage. Within the country, branch 3 strains were found circulating in humans in the states of Rio de Janeiro and São Paulo, both in southeastern Brazil (Fontes et al., 2009). In northeastern states, branch 4 strains and a small number of branch 1 strains have been identified (Fontes et al., 2012). Branch 1 and the paraphyletic branch 2 are part of the same larger clade; however, those strains in branch 2 (SNP types 2E, 2G, 2H) are found in eastern Africa, the Arabian Peninsula, and central Asia instead of the Americas (Avanzi et al., 2016a; Benjak et al., 2018). There are few whole genomes of *M. leprae* isolated from animals, which are published. However, within the Americas, W09 was sequenced from a nine-banded armadillo (*D. novemcinctus*) (Truman et al., 2011). W09 is in branch 3 along with other human-associated Brazil strains, but it was collected from the southeastern United States. Branch 4 contains other animal-associated strains, both of which come from non-human primates in western Africa (Honap et al., 2018).

These strains belong to SNP subtype 4N/O within branch 4, which are found in humans in Brazil and the non-human primates in West Africa (Benjak et al., 2018, Honap et al., 2018). Brazil strains and their branches are visualized in Figure 2.

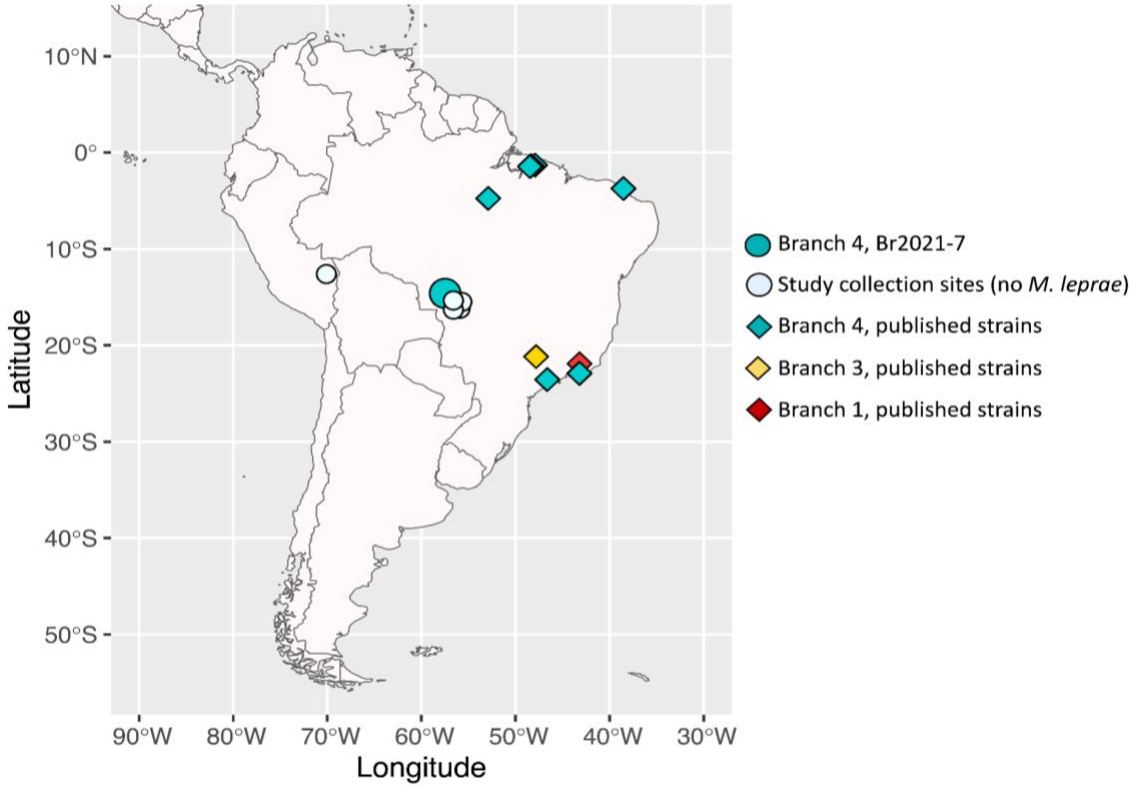


Figure 2: Map showing Brazil and Peru mammals samples used in this study and 21 locations of origin for published samples in Brazil with their branch information. Strains in this area without region- or city-specific location information were not included.

Br2021-7 clusters with the expected local strain types found in humans and has mutations associated with drug resistance. Despite relatively low rates of drug resistance, areas of hyperendemicity in Brazil are a particular concern for public health efforts. In a former

leprosy colony in the Brazilian Amazon, rates of the disease were as high as 12.8% and relapse cases showed multidrug resistance to rifampicin and dapsone (Lazaro et al., 2010; Rosa et al, 2020). Although these numbers are very high compared to the rest of the country (WHO, 2023), these pockets are primed for circulation and spread of multidrug resistant strains (Bouth et al., 2023). The Br2021-7 nonsynonymous variants in the *rpoB* and *gyrA* regions are associated with rifampicin and fluoroquinolone resistance respectively (Williams & Gillis, 2012). The presence of this strain potentially circulating in wild populations suggests that rates of drug resistance in Brazil may be higher than previously thought.

Infection of Host and Reservoir Potential

Isolation of Br2021-7 from the spleen suggests that the animal was infected by *M. leprae*. Unlike the well documented animal reservoir *D. novemcinctus*, we do not have enough information to determine if *P. concolor* is a natural reservoir for leprosy. We suggest that the host of Br2021-7 became infected through consumption of already infected prey. Spillover from armadillos to humans in the United States and Brazil occurs through handling, hunting, and consumption (Truman et al., 2011; da Silva et al., 2018; Frota et al., 2012). In lieu of further histological and survey evidence in *P. concolor*, we find a limited spillover event due to predator-prey interaction to be the most parsimonious explanation.

The range of *P. concolor* extends across the Americas, from Canada to the southern tip of South America (Castilho et al., 2011). Although severe habitat loss and illegal hunting in southern Brazil have caused genetic bottlenecks in *P. concolor*

populations, they maintain a wide range across forested and agricultural areas in the country (Castilho et al, 2011; Magioli et al., 2014). Their prey includes armadillos and other small and midsize mammals, such as capybaras (Magioli et al., 2014). The puma range and food choice overlap with known animal reservoirs of *M. leprae*. The *D. novemcinctus* range extends to South America, where the other 20 extant armadillo species are also widely distributed (Superina et al. 2013). Of those species, only the six-banded armadillo (*E. sexcinctus*) has shown positive *M. leprae* PCR results, though such surveys are extremely limited (Frota et al., 2012). Around 70% of studies on armadillos and *M. leprae* focus only on nine-banded armadillo (*D. novemcinctus*), which is the only armadillo species in North America (Deps et al., 2020). Surveys of *M. leprae* in *D. novemcinctus* show that prevalence in Brazilian armadillos could be as high as 10% of the population; however, wide variation in assay type (*e.g.*, PCR assays and Ziehl-Neehlsen staining), sample size, and tissue type suggests that this number is not accurate (Deps et al., 2020; da Silva Ferreira et al., 2020). Despite such uncertainty with *M. leprae* prevalence in armadillos and other potential prey, the ranges of these species support our hypothesis that transmission to the puma occurred through hunting of infected prey.

Finally, we must consider the potential of puma to act as a natural reservoir. Aside from positive *M. leprae* PCR results from the margay (*L. wiedii*) (Maruyama et al., 2018), a small wild cat native to Central and South America, there have been no *M. leprae* strains identified in domesticated or wild cats. This may reflect the lack of cat species in the Americas studied for *M. leprae*, but we find no evidence to support that the pathogen is permanently maintained in *P. concolor* or that the *P. concolor* can actively transmit *M. leprae* to other species. These key traits help define a reservoir species

(Viana et al., 2014). Additionally, so few animal-associated *M. leprae* genomes have been sequenced, that it is difficult to determine if spillover is limited to specific hyperendemic regions. *M. leprae* is particularly difficult to study in wild animal populations as infected animals may die of other causes (*i.e.*, Br2021-7 puma host), the pathology in such populations may be different from known reservoirs and therefore difficult to identify, and frequency of testing (as well as pathogen load) is low. Much more research is needed on South American mammals and *M. leprae* to adequately understand the extent to which animal reservoirs may affect leprosy eradication efforts.

In addition to surveys of mammals in endemic areas, researchers have also investigated alternative vectors and the presence of *M. leprae* in the environment. Viable *M. leprae* has been found to persist in free-living amoebae for up to 8 months in specific species (Wheat et al., 2014), kissing bugs for 20 days (da Silva Neumann et al., 2016), and in the tick midgut and ovaries (Ferreira et al., 2018). The latter two studies found that small mammals could be successfully inoculated using kissing bugs and ticks, respectively. There is also great interest in the persistence of viable *M. leprae* bacilli in the environment. Multiple studies have identified *M. leprae* RNA in soil samples in endemic areas and outside of households with active infections (Lavania et al., 2008, Turankar et al., 2016; 2022). Mohanty et al. (2016) identified a quarter of soil and water samples collected in endemic areas positive for 16S rRNA during screening. Tio-Coma et al. (2019) expanded these analyses by investigating soil samples across Bangladesh, Suriname, and the British Isles by analyzing patients' households, armadillo holds, and red squirrel habitats. As with previous soil studies, *M. leprae* does persist in the environment for a limited time, however given the obligate, intracellular nature of the

pathogen, more work is needed to determine actual transmission versus viability by RNA presence. Although these potential sources for Br2021-7 are intriguing, there is not enough support for vector- or environmentally-mediated transmission as a primary mechanism of transmission for this strain.

One Health Approach to Leprosy

Despite its historical association with human populations, the spillover of *M. leprae* into animal reservoirs poses novel challenges to our understanding of the epidemiology and evolutionary aspects of leprosy. The WHO Global Hansen's Disease Strategy 2021-2030 report noted that global eradication of the disease would be difficult or infeasible due to zoonotic transmission (WHO, 2023). Advocates for a One Health approach to leprosy have sought to address this problem by leveraging ancient and modern genetics of *M. leprae* with veterinary and public health approaches (Deps et al, 2020; Urban et al., 2021; 2024). There is a clear need for an integrated approach to reducing the transmission and impact of leprosy in Brazil as pockets of hyperendemicity, "hidden cases" in isolated populations, and unknown animal reservoirs harm local public health efforts.

4.6 Conclusion

Here we present the first whole genome *M. leprae* strain sequenced from a puma (*P. concolor*). This isolate, Br2021-7, was collected from a puma in the state Mato Grosso in Brazil, which is a high-risk region for leprosy and zoonotic spillover of the disease (WHO, 2023). Although the mean low coverage of the genome prevented SNP typing, we were able to place Br2021-7 in branch 4 of the *M. leprae* phylogeny, which is

one of three common branches represented in Brazil *M. leprae* strains. The presence of Br2021-7 in a wild puma suggests the need for broad surveillance of mammal populations for *M. leprae* or *M. lepromatosis* to characterize natural reservoirs of these pathogens and better address the global health burden of leprosy.

CHAPTER 5

CONCLUSION

Leprosy is one of the oldest known human diseases and remains a modern global health problem. The global patterning of genomic variation of *M. leprae*, along with zoonotic spillover risk and transmission routes, are not well defined but have been increasingly studied with the advent of whole-genome sequencing (Avanzi et al., 2020c; Schuenemann et al., 2018). This dissertation aims to address gaps in our understanding of *M. leprae* origin and dispersal in the Pacific, as well as to characterize strains in novel host species in Brazil. The background on leprosy, its causative agents, and their molecular biology and epidemiology have been outlined in Chapter 1.

Chapter 2 addresses questions regarding the origin of leprosy into the Pacific, as well as characterizing unknown genetic diversity of *M. leprae* in this region. Through whole-genome sequencing, 9 novel *M. leprae* genomes were added to the phylogeny, tripling the number of sequenced strains in the Pacific Islands. Phylogenetic analyses placed these novel strains with other *M. leprae* genomes in branches 0 and 5, which cluster with other isolates from the Pacific and East Asia. Furthermore, this chapter supports and expands limited dating analyses already done on the divergence times of branch 0 strains into the Pacific. The entry of *M. leprae* in this region was determined to be around 3000 BP, which occurred around the initial peopling of the Pacific Islands.

Chapter 3 expands upon Chapter 2 by incorporating additional whole-genome data and phylogeographic analyses to determine how various *M. leprae* strains are circulating among the Pacific Islands. To this end, 12 novel *M. leprae* genomes were sequenced from Guam, Hawaii, the Marshall Islands, Palau, Pohnpei, and Samoa.

Comparative and phylogenetic analyses of the sequences revealed distinct allelic changes in intrahost samples during treatment, clustering of strains with other Pacific and East Asian strains, and a potential distribution history of these stains. The movement and timing of *M. leprae* strains through the Pacific were characterized with a phylogeographic approach that incorporated Bayesian methods to clarify the geographic patterning and evolutionary history of this pathogen.

Finally, chapter 4 describes a new, low-coverage *M. leprae* genome isolated from a puma in Mato Grosso, Brazil. Brazil has the second highest incidence of new cases of leprosy reported each year and has high genetic diversity in circulating strains (WHO, 2023). Additionally, ongoing surveillance has identified likely zoonotic spillover from armadillos into humans, as well as a rise in drug-resistant strains (Frota et al., 2012). Therefore, this chapter not only presents an animal-associated *M. leprae* strain from a novel host, but also suggests the need to improve surveys of a wider range of mammal species in endemic regions. These cases also reveal more about the evolutionary mechanisms of zoonotic diseases and how they might be transmitted between species.

Ultimately, this dissertation addresses significant gaps in our understanding of genomic diversity of *M. leprae* in the Pacific and identifies the need for further surveillance in understudied animal populations. The long incubation period of leprosy, along with the unculturable nature of the pathogen, present unique challenges to tracking the spread of the disease. Through a more robust body of genomic data of *M. leprae* strains and a phylogenetic approach that accounts for movement across time and space, we may better inform global health efforts in addressing leprosy.

REFERENCES

Akeli, S. (2007). *Leprosy in Samoa 1890 to 1922: Race, Colonial Politics and Disempowerment*.

Amundson, R., & Oakaokalani Ruddle-Miyamoto, A. (2010). A Wholesome Horror: The Stigmas of Leprosy in 19th Century Hawaii. *Disability Studies Quarterly*, *30*(3/4). https://dsq-sds.org/index.php/dsq/article/view/1270/1300

Andrews, S. (2018). *FastQC: A quality control tool for high throughput sequence data* (0.11.7). http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Araujo, S., Freitas, L. O., Goulart, L. R., & Goulart, I. M. B. (2016). Molecular Evidence for the Aerial Route of Infection of *Mycobacterium leprae* and the Role of Asymptomatic Carriers in the Persistence of Leprosy. *Clinical Infectious Diseases*, *63*(11), 1412–1420. https://doi.org/10.1093/CID/CIW570

Araújo Stefani, M. M., Rosa, P. S., Costa, M. B., Mendes Schetinni, A. P., Manhães, I., Andrade Pontes, M. A., Costa, P., Vincenzi Fachin, L. R., Foschiani Dias Batista, I. M., Virmond, M., Pereira, E., Fernandes Penna, M. L., & Penna, G. O. (2019). Leprosy survey among rural communities and wild armadillos from Amazonas state, Northern Brazil. *PLOS ONE*, *14*(1), e0209491. https://doi.org/10.1371/JOURNAL.PONE.0209491

Avanzi, C., Busso, P., Benjak, A., Loiseau, C., Fomba, A., Doumbia, G., Camara, I., Lamou, A., Sock, G., Drame, T., Kodio, M., Sakho, F., Sow, S. O., Cole, S. T., & Johnson, R. C. (2016). Transmission of Drug-Resistant Leprosy in Guinea-Conakry Detected Using Molecular Epidemiological Approaches. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America*, *63*(11), 1482–1484. https://doi.org/10.1093/CID/CIW572

Avanzi, C., del-Pozo, J., Benjak, A., Stevenson, K., Simpson, V. R., Busso, P., McLuckie, J., Loiseau, C., Lawton, C., Schoening, J., Shaw, D. J., Piton, J., Vera-Cabrera, L., Velarde-Felix, J. S., McDermott, F., Gordon, S. v, Cole, S. T., & Meredith, A. L. (2016). Red squirrels in the British Isles are infected with leprosy bacilli. *Science*, *354*(6313), 744 LP – 747. https://doi.org/10.1126/science.aah3783

Avanzi, C., Lécorché, E., Rakotomalala, F. A., Benjak, A., Rapelanoro Rabenja, F., Ramarozatovo, L. S., Cauchoix, B., Rakoto-Andrianarivelo, M., Tió-Coma, M., Leal-Calvo, T., Busso, P., Boy-Röttger, S., Chauffour, A., Rasamoelina, T., Andrianarison, A., Sendrasoa, F., Spencer, J. S., Singh, P., Dashatwar, D. R., … Cole, S. T. (2020). Population Genomics of *Mycobacterium leprae* Reveals a New Genotype in Madagascar and the Comoros. *Frontiers in Microbiology*, *11*, 523573. https://doi.org/10.3389/FMICB.2020.00711/BIBTEX

Avanzi, C., Maia, R. C., Benjak, A., Nery, J. A., Sales, A. M., Miranda, A., Duppre, N. C., Fontes, A. N. B., da Silva, T. P., Pinheiro, R. O., Neves-Manta, F., Moreira, S. J. M., Busso, P., Sarno, E. N., Suffys, P. N., Cole, S. T., & Moraes, M. O. (2020). Emergence of *Mycobacterium leprae* rifampin resistance evaluated by whole-genome sequencing after 48 years of irregular treatment. *Antimicrobial Agents and Chemotherapy*, *64*(7). https://doi.org/10.1128/AAC.00330-20/SUPPL_FILE/AAC.00330-20-S0001.PDF

Avanzi, C., Singh, P., Truman, R. W., & Suffys, P. N. (2020). Molecular epidemiology of leprosy: An update. *Infection, Genetics and Evolution*, *86*, 104581. https://doi.org/10.1016/J.MEEGID.2020.104581

Baele, G., Lemey, P., & Suchard, M. A. (2016). Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. *Systematic Biology*, *65*(2), 250–264. https://doi.org/10.1093/SYSBIO/SYV083

Bakker, M. I., Hatta, M., Kwenang, A., van Mosseveld, P., Faber, W. R., Klatser, P. R., & Oskam, L. (2006). Risk factors for developing leprosy – a population-based cohort study in Indonesia. *Leprosy Review*, *77*(1), 48–61. https://doi.org/10.47276/LR.77.1.48

Benjak, A., Avanzi, C., Singh, P., Loiseau, C., Girma, S., Busso, P., Fontes, A. N. B., Miyamoto, Y., Namisato, M., Bobosha, K., Salgado, C. G., Silva, M. B., Bouth, R. C., Frade, M. A. C., Filho, F. B., Barreto, J. G., Nery, J. A. C., Bührer-Sékula, S., Lupien, A., … Cole, S. T. (2018). Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. *Nature Communications*, 1–11. http://dx.doi.org/10.1038/s41467-017-02576-z

Benjak, A., Honap, T. P., Avanzi, C., Becerril-Villanueva, E., Estrada-García, I., Rojas-Espinosa, O., Stone, A. C., & Cole, S. T. (2017). Insights from the genome sequence of *Mycobacterium lepraemurium*: Massive gene decay and reductive evolution. *MBio*, *8*(5). https://doi.org/10.1128/MBIO.01283-17/SUPPL_FILE/MBO005173527S1.DOCX

Berra, C. M., Torrezan, G. T., de Paula, C. A., Hsieh, R., Lourenço, S. v., & Carraro, D. M. (2019). Use of uracil-DNA glycosylase enzyme to reduce DNA-related artifacts from formalin-fixed and paraffin-embedded tissues in diagnostic routine. *Applied Cancer Research*, *39*(1). https://doi.org/10.1186/s41241-019-0075-2

Bielejec, F., Baele, G., Vrancken, B., Suchard, M. A., Rambaut, A., & Lemey, P. (2016). SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Molecular Biology and Evolution*, *33*(8), 2167–2169. https://doi.org/10.1093/MOLBEV/MSW082

Binford, C. H. (1936). The History and Study of Leprosy in Hawaii. *Public Health Reports (1896-1970)*, *51*(15), 415–423. https://doi.org/10.2307/4581802

Blau, S., & Yagodin, V. (2005). Osteoarchaeological evidence for leprosy from western Central Asia. *American Journal of Physical Anthropology*, *126*(2), 150–158. https://doi.org/10.1002/AJPA.20121

Blevins, K. E., Crane, A. E., Lum, C., Furuta, K., Fox, K., & Stone, A. C. (2020). Evolutionary history of *Mycobacterium leprae* in the Pacific Islands. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1812). https://doi.org/10.1098/rstb.2019.0582

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Steadman, D. L. W., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., … Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, *514*(7253), 494–497. https://doi.org/10.1038/nature13591

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003537

Bouth, R. C., Gobbo, A. R., Barreto, J. G., do Carmo Pinto, P. D., Bittencourt, M. S., Frade, M. A. C., Nascimento, A. C., Bandeira, S. S., da Costa, P. F., Conde, G. A. B., Avanzi, C., Ribeiro-dos-Santos, Â., Spencer, J. S., da Silva, M. B., & Salgado, C. G. (2023). Specialized active leprosy search strategies in an endemic area of the Brazilian Amazon identifies a hypermutated *Mycobacterium leprae* strain causing primary drug resistance. *Frontiers in Medicine*, *10*, 1243571. https://doi.org/10.3389/FMED.2023.1243571/BIBTEX

Breese, M. R., & Liu, Y. (2013). NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics*, *29*(4), 494–496. https://doi.org/10.1093/BIOINFORMATICS/BTS731

Britton, W. J. (1993). 3. Immunology of leprosy. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *87*(5), 508–514. https://doi.org/10.1016/0035-9203(93)90066-Y

Browne, S. (1985). The history of leprosy. In R. Hastings (Ed.), *Leprosy* (pp. 1–31).

Burns, S. (2019). *Kingdom of the sick: a history of leprosy and Japan*. University of Hawai‘i .

CA, J., P, V. B. K., Kandi, V., N, G., K, S., Dharshini, D., Batchu, S. V. C., & Bhanu, P. (2024). Neglected Tropical Diseases: A Comprehensive Review. *Cureus*, *16*(2). https://doi.org/10.7759/CUREUS.53933

Cachola-Abad, C. (1993). Evaluating the orthodox dual settlement model for the Hawaiian Islands: an analysis of artefact distribution and Hawaiian oral traditions. In M. Graves & R. Green (Eds.), *The evolution and organisation of prehistoric society in Polynesia* (pp. 13–32). Archaeological Association.

Cambau, E., Saunderson, P., Matsuoka, M., Cole, S. T., Kai, M., Suffys, P., Rosa, P. S., Williams, D., Gupta, U. D., Lavania, M., Cardona-Castro, N., Miyamoto, Y., Hagge, D., Srikantam, A., Hongseng, W., Indropo, A., Vissa, V., Johnson, R. C., Cauchoix, B., … Nanba, Y. (2018). Antimicrobial resistance in leprosy: results of the first prospective open survey conducted by a WHO surveillance network for the period 2009–15. *Clinical Microbiology and Infection*, *24*(12), 1305–1310. https://doi.org/10.1016/j.cmi.2018.02.022

Campos, P. F., & Gilbert, T. M. P. (2012). DNA Extraction from Formalin-Fixed Material. *Methods in Molecular Biology*, *840*, 81–85. https://doi.org/10.1007/978-1-61779-516-9_11

Cardona-Castro, N., Beltrán, J. C., Ortiz-Bernal, A., & Vissa, V. (2009). Detection of *Mycobacterium leprae* DNA in nine-banded armadillos (*Dasypus novemcinctus*) from the Andean region of Colombia. *Leprosy Review*, *80*(4), 424–431.

Carrick, D. M., Mehaffey, M. G., Sachs, M. C., Altekruse, S., Camalier, C., Chuaqui, R., Cozen, W., Das, B., Hernandez, B. Y., Lih, C. J., Lynch, C. F., Makhlouf, H., McGregor, P., McShane, L. M., Rohan, J. P., Walsh, W. D., Williams, P. M., Gillanders, E. M., Mechanic, L. E., & Schully, S. D. (2015). Robustness of Next Generation Sequencing on Older Formalin-Fixed Paraffin-Embedded Tissue. *PLOS ONE*, *10*(7), e0127353. https://doi.org/10.1371/JOURNAL.PONE.0127353

Castilho, C. S., Marins-Sá, L. G., Benedet, R. C., & Freitas, T. O. (2011). Landscape genetics of mountain lions (*Puma concolor*) in southern Brazil. *Mammalian Biology*, *76*(4), 476–483. https://doi.org/10.1016/J.MAMBIO.2010.08.002

Chokkakula, S., Shui, T., Jiang, H., Yang, J., Li, X., He, J., Shen, L., Liu, J., Wang, D., Suryadevara, N. C., Pathakumari, B., Wang, L., Chen, Y., Shi, Y., Zhang, W., Wang, H., Chen, H., Kuang, Y., Li, B., … Wang, H. (2020). Genotyping of *Mycobacterium leprae* for understanding the distribution and transmission of leprosy in endemic provinces of China. *International Journal of Infectious Diseases*, *98*, 6–13. https://doi.org/10.1016/J.IJID.2020.06.032

Cingolani, P., Platts, A., Wang, L. L. L., Coon, M., Nguyen, T., Wang, L. L. L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*, 80–92. https://doi.org/10.1016/S1877-1203(13)70353-7

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honoré, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., … Barrell, B. G. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*(6823), 1007–1011. https://doi.org/10.1038/35059006

Collerson, K. D., & Weisler, M. I. (2007). Stone age compositions and the extent of ancient Polynesian voyaging and trade. *Science*, *317*(5846), 1907–1911. https://doi.org/10.1126/SCIENCE.1147013/SUPPL_FILE/COLLERSON.SOM.PDF

da Silva Ferreira, J., de Carvalho, F. M., Vidal Pessolani, M. C., de Paula Antunes, J. M. A., de Medeiros Oliveira, I. V. P., Ferreira Moura, G. H., Truman, R. W., Peña, M. T., Sharma, R., Duthie, M. S., de Paula Souza e Guimarães, R. J., Nogueira Brum Fontes, A., NoelSuffys, P., & McIntosh, D. (2020). Serological and molecular detection of infection with *Mycobacterium leprae* in Brazilian six banded armadillos (Euphractus sexcinctus). *Comparative Immunology, Microbiology and Infectious Diseases*, *68*, 101397. https://doi.org/10.1016/J.CIMID.2019.101397

da Silva, M. B., Portela, J. M., Li, W., Jackson, M., Gonzalez-Juarrero, M., Hidalgo, A. S., Belisle, J. T., Bouth, R. C., Gobbo, A. R., Barreto, J. G., Minervino, A. H. H., Cole, S. T., Avanzi, C., Busso, P., Frade, M. A. C., Geluk, A., Salgado, C. G., & Spencer, J. S. (2018). Evidence of zoonotic leprosy in Pará, Brazilian Amazon, and risks associated with human contact or consumption of armadillos. *PLoS Neglected Tropical Diseases*, *12*(6). https://doi.org/10.1371/journal.pntd.0006532

da Silva Neumann, A., de Almeida Dias, F., da Silva Ferreira, J., Fontes, A. N. B., Rosa, P. S., Macedo, R. E., Oliveira, J. H., de Figueiredo Teixeira, R. L., Pessolani, M. C. V., Moraes, M. O., Suffys, P. N., Oliveira, P. L., Sorgine, M. H. F., & Lara, F. A. (2016). Experimental Infection of Rhodnius prolixus (Hemiptera, Triatominae) with *Mycobacterium leprae* Indicates Potential for Leprosy Transmission. *PLOS ONE*, *11*(5), e0156037. https://doi.org/10.1371/JOURNAL.PONE.0156037

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763. http://www.pnas.org/cgi/doi/10.1073/pnas.1314445110

Daley, T., & Smith, A. D. (2013). Predicting the molecular complexity of sequencing libraries. *Nature Methods*, *10*(4), 325–327. https://doi.org/10.1038/nmeth.2375

D'arcy, P. (2014). The Chinese Pacifics: A Brief Historical Review. *The Journal of Pacific History*, *49*(4), 396–420. http://www.jstor.org/stable/24644647

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *36*(2), 716–723. https://doi.org/10.1038/nmeth.2109

Das, M., Diana, D., Wedderburn, A., Rajan, L., Rao, S., Horo, I., & Vedithi, S. C. (2020). Molecular epidemiology and transmission dynamics of leprosy among multicase families and case-contact pairs. *International Journal of Infectious Diseases*, *96*, 172–179. https://doi.org/10.1016/j.ijid.2020.04.064

de Sousa Oliveira, J. S., dos Reis, A. L. M., Margalho, L. P., Lopes, G. L., da Silva, A. R., de Moraes, N. S., & Xavier, M. B. (2019). Leprosy in elderly people and the profile of a retrospective cohort in an endemic region of the Brazilian Amazon. *PLOS Neglected Tropical Diseases*, *13*(9), e0007709. https://doi.org/10.1371/JOURNAL.PNTD.0007709

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. v, Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498. http://dx.doi.org/10.1038/ng.806

Deps, P., Antunes, J. M., Santos, A. R., & Collin, S. M. (2020). Prevalence of *Mycobacterium leprae* in armadillos in Brazil: A systematic review and meta-analysis. *PLOS Neglected Tropical Diseases*, *14*(3), e0008127. https://doi.org/10.1371/JOURNAL.PNTD.0008127

Dharmendra. (1947). Leprosy in ancient Indian medicine. *Int J Lepr*, *15*, 424–430.
di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Do, H., & Dobrovic, A. (2015). Sequence artifacts in DNA from formalin-fixed tissues: Causes and strategies for minimization. In *Clinical Chemistry* (Vol. 61, Issue 1, pp. 64–71). American Association for Clinical Chemistry Inc. https://doi.org/10.1373/clinchem.2014.223040

Donoghue, H. D., Michael Taylor, G., Marcsik, A., Molnár, E., Pálfi, G., Pap, I., Teschler-Nicola, M., Pinhasi, R., Erdal, Y. S., Velemínsky, P., Likovsky, J., Belcastro, M. G., Mariotti, V., Riga, A., Rubini, M., Zaio, P., Besra, G. S., Lee, O. Y. C., Wu, H. H. T., … Spigelman, M. (2015). A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. *Infection, Genetics and Evolution*, *31*, 250–256. https://doi.org/10.1016/j.meegid.2015.02.001

Draper, P. (1983). The bacteriology of *Mycobacterium leprae*. *Tubercle*, *64*(1), 43–56. https://doi.org/10.1016/0041-3879(83)90050-8

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. *PLOS Biology*, *4*(5), e88. https://doi.org/10.1371/JOURNAL.PBIO.0040088

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192. https://doi.org/10.1093/MOLBEV/MSI103

Duggan, A. T., Perdomo, M. F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Emery, M. v., Buchmann, J. P., Duchêne, S., Jankauskas, R., Humphreys, M., Golding, G. B., Southon, J., Devault, A., Rouillard, J.-M., Sahl, J. W., Dutour, O., Hedman, K., Sajantila, A., Smith, G. L., … Poinar, H. N. (2016). 17th Century Variola Virus Reveals the Recent History of Smallpox. *Current Biology*, *26*(24), 3407–3412. https://doi.org/10.1016/j.cub.2016.10.061

Edmond, R. (2007). *Leprosy and Empire: A Medical and Cultural History*. Cambridge University Press.

Faber, W. R., Sewpersad, K., Menke, H., Avanzi, C., Geluk, A., Verhard, E. M., Coma, M. T., Chan, M., & Pieters, T. (2023). Origin and spread of leprosy in Suriname. A historical and biomedical study. *Front. Trop. Dis*, *4*. https://doi.org/10.3389/fitd.2023.1258006

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J., & Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ*, *9*, e10947. https://doi.org/10.7717/peerj.10947

Ferreira, J. da S., Souza, D. A., Santos, J. P., Ribeiro, C. C. D. U., Baêta, B. A., Teixeira, R. C., Neumann, A. da S., Rosa, P. S., Pessolani, M. C. V., Moraes, M. O., Bechara, G. H., Oliveira, P. L. de, Sorgine, M. H. F., Suffys, P. N., Fontes, A. N. B., Bell-Sakyi, L., Fonseca, A. H., & Lara, F. A. (2018). Ticks as potential vectors of *Mycobacterium leprae*: Use of tick cell lines to culture the bacilli and generate transgenic strains. *PLOS Neglected Tropical Diseases*, *12*(12), e0007001. https://doi.org/10.1371/JOURNAL.PNTD.0007001

Fontes, A. N. B., Lima, L. N. G. C., Mota, R. M. S., Almeida, R. L. F., Pontes, M. A., Gonçalves, H. de S., Frota, C. C., Vissa, V. D., Brennan, P. J., Guimaraes, R. J. P. S., Kendall, C., Kerr, L. R. F. S., & Suffys, P. N. (2017). Genotyping of *Mycobacterium leprae* for better understanding of leprosy transmission in Fortaleza, Northeastern Brazil. *PLOS Neglected Tropical Diseases*, *11*(12), e0006117. https://doi.org/10.1371/JOURNAL.PNTD.0006117

Fontes, A. N. B., Sakamuri, R. M., Baptista, I. M. F. D., Ura, S., Moraes, M. O., Martínez, A. N., Sarno, E. N., Brennan, P. J., Vissa, V. D., & Suffys, P. N. (2009). Genetic diversity of *Mycobacterium leprae* isolates from Brazilian leprosy patients. *Leprosy Review*, *80*(3), 302–315.

Frota, C. C., Lima, L. N. C., Rocha, A. da S., Suffys, P. N., Rolim, B. N., Rodrigues, L. C., Barreto, M. L., Kendall, C., & Kerr, L. R. S. (2012). *Mycobacterium leprae* in six-banded (Euphractus sexcinctus) and nine-banded armadillos (*Dasypus novemcinctus*) in Northeast Brazil. *Memorias Do Instituto Oswaldo Cruz*, *107*(SUPPL.1), 209–213. https://doi.org/10.1590/S0074-02762012000900029

Gama, R. S., Leite, L. A., Colombo, L. T., & Fraga, L. A. de O. (2020). Prospects for new leprosy diagnostic tools, a narrative review considering ELISA and PCR assays. *Revista Da Sociedade Brasileira de Medicina Tropical*, *53*, e20200197. https://doi.org/10.1590/0037-8682-0197-2020

Gilbert, M. T. P., Haselkorn, T., Bunce, M., Sanchez, J. J., Lucas, S. B., Jewell, L. D., van Marck, E., & Worobey, M. (2007). The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues-Which Methods Are Useful When? *PLoS ONE*, *2*(6). https://doi.org/10.1371/journal.pone.0000537

GitHub - ncbi/sra-tools: SRA Tools. (n.d.). Retrieved August 6, 2020, from https://github.com/ncbi/sra-tools

Guan, Q., Almutairi, T. S., Alhalouli, T., Pain, A., & Alasmari, F. (2020). Metagenomics of Imported Multidrug-Resistant *Mycobacterium leprae*, Saudi Arabia, 2017. *Emerging Infectious Diseases*, *26*(3), 615. https://doi.org/10.3201/EID2603.190661

Guindon, S., & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

Hahn, E. E., Alexander, M. R., Grealy, A., Stiller, J., Gardiner, D. M., & Holleley, C. E. (2021). Unlocking inaccessible historical genomes preserved in formalin. *Molecular Ecology Resources 1–18*. https://doi.org/10.1111/1755-0998.13505

Hall, B. G., & Salipante, S. J. (2010). Molecular epidemiology of *Mycobacterium leprae* as determined by structure-neighbor clustering. *Journal of Clinical Microbiology*, *48*(6), 1997–2008. https://doi.org/10.1128/JCM.00149-10/SUPPL_FILE/SUPPLEMENTARY_MATERIALS.ZIP

Han, X. Y., Seo, Y. H., Sizer, K. C., Schoberle, T., May, G. S., Spencer, J. S., Li, W., & Nair, R. G. (2008). A new *Mycobacterium* species causing diffuse lepromatous leprosy. *American Journal of Clinical Pathology*, *130*(6), 856–864. https://doi.org/10.1309/AJCPP72FJZZRRVMM

Han, X. Y., Sizer, K. C., Velarde-Félix, J. S., Frias-Castro, L. O., & Vargas-Ocampo, F. (2012). The leprosy agents *Mycobacterium lepromatosis* and *Mycobacterium leprae* in Mexico. *International Journal of Dermatology*, *51*(8), 952–959. https://doi.org/10.1111/j.1365-4632.2011.05414.x

Han, X. Y., Zhang, J., & Li, L. (2015). Leprosy agents *Mycobacterium lepromatosis* and *Mycobacterium leprae* in Mexico: A clarification. In *Journal of Clinical Microbiology* (Vol. 53, Issue 10, pp. 3387–3388). American Society for Microbiology. https://doi.org/10.1128/JCM.01588-15

Heathcote, G. M., Stodder, A. L. W., Buckley, H. R., Hanson, D. B., Douglas, M. T., Underwood, J. H., Taisipic, T. F., & Diego, V. P. (1998). On Treponemal Disease in the Western Pacific: Corrections and Critique. *Https://Doi.Org/10.1086/204745*, *39*(3), 359–368. https://doi.org/10.1086/204745

Hirata, K., Oku, C., & Morimoto, I. (2000). A case of leprosy in a medieval Japanese. In *Anthropological Science* (p. 114).

Hockings, K. J., Mubemba, B., Avanzi, C., Pleh, K., Düx, A., Bersacola, E., Bessa, J., Ramon, M., Metzger, S., Patrono, L. v., Jaffe, J. E., Benjak, A., Bonneaud, C., Busso, P., Couacy-Hymann, E., Gado, M., Gagneux, S., Johnson, R. C., Kodio, M., … Leendertz, F. H. (2021). Leprosy in wild chimpanzees. *Nature 2021 598:7882*, *598*(7882), 652–656. https://doi.org/10.1038/s41586-021-03968-4

Honap, T. P., Pfister, L.-A., Housman, G., Mills, S., Tarara, R. P., Suzuki, K., Cuozzo, F. P., Sauther, M. L., Rosenberg, M. S., & Stone, A. C. (2018). *Mycobacterium leprae* genomes from naturally infected nonhuman primates. *PLoS Neglected Tropical Diseases*, *12*(1), e0006190. https://doi.org/10.1371/journal.pntd.0006190

Inglis, K. (2014). "One's Molokai Can Be Anywhere": Global Influence in the Twentieth-Century History of Hansen's Disease. *Journal of World History*, *25*(4), 611–627. https://www.jstor.org/stable/43818467

Jessamine, P. G., Desjardins, M., Gillis, T., Scollard, D., Jamieson, F., Broukhanski, G., Chedore, P., & McCarthy, A. (2012). Leprosy-like illness in a patient with *Mycobacterium lepromatosis* from Ontario, Canada. *Journal of Drugs in Dermatology : JDD*, *11*(2), 229–233. http://www.ncbi.nlm.nih.gov/pubmed/22270208

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/BIOINFORMATICS/BTT193

Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research*, *21*(3), 487–493. https://doi.org/10.1101/GR.113985.110

King, F. (2023). *Comparison of Two DNA Extraction Methods for Isolating Mycobacterium leprae DNA from FFPE Tissue Collected in the Pacific Islands Region*. ProQuest (Master's Thesis).

Köhler, K., nia Marcsik, A., Zádori, P., Biro, G., Szeniczey, T., Fábiá, S., bor Serlegi, G., Marton, T., Donoghue, H. D., & Hajdu, T. (2017). Possible cases of leprosy from the Late Copper Age (3780-3650 cal BC) in Hungary Introduction Archaeological background. *PLoS One*. https://doi.org/10.1371/journal.pone.0185966

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Lange, R. (2017). Leprosy in the Cook Islands, 1890–1925. *Journal of Pacific History*, *52*(3), 302–324. https://doi.org/10.1080/00223344.2017.1379117

Lavania, M., Katoch, K., Katoch, V. M., Gupta, A. K., Chauhan, D. S., Sharma, R., Gandhi, R., Chauhan, V., Bansal, G., Sachan, P., Sachan, S., Yadav, V. S., & Jadhav, R. (2008). Detection of viable *Mycobacterium leprae* in soil samples: Insights into possible sources of transmission of leprosy. *Infection, Genetics and Evolution*, *8*(5), 627–631. https://doi.org/10.1016/J.MEEGID.2008.05.007

Lavania, M., Singh, I., Turankar, R. P., Ahuja, M., Pathak, V., Sengupta, U., Das, L., Kumar, A., Darlong, J., Nathan, R., & Maseey, A. (2018). Molecular detection of multidrug-resistant *Mycobacterium leprae* from Indian leprosy patients. *Journal of Global Antimicrobial Resistance*, *12*, 214–219. https://doi.org/10.1016/j.jgar.2017.10.010

Lázaro, F. P., Werneck, R. I., Mackert, C. C. O., Cobat, A., Prevedello, F. C., Pimentel, R. P., Macedo, G. M. M., Eleutério, M. A. M., Vilar, G., Abel, L., Xavier, M. B., Alcaïs, A., & Mira, M. T. (2010). A Major Gene Controls Leprosy Susceptibility in a Hyperendemic Isolated Population from North of Brazil. *The Journal of Infectious Diseases*, *201*(10), 1598–1605. https://doi.org/10.1086/652007

Leung, A. (2009). *Leprosy in China: a history*. Colombia University Press.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics*. http://arxiv.org/abs/1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lima, L., Fontes, A. N. B., Li, W., Suffys, P. N., Vissa, V., Mota, R. M. S., Almeida, R. L. F., Gonçalves, H. S., Frota, C. C., Rodrigues, L. C., Kendall, C., & Kerr, L. R. S. (2016). Intrapatient comparison of *Mycobacterium leprae* by VNTR analysis in nasal secretions and skin biopsy in a Brazilian leprosy endemic region. *Leprosy Review*, *87*, 486–500.

Lipson, M., Skoglund, P., Spriggs, M., Mentzer, A. J., Pinhasi, R., & Reich, D. (2018). Population Turnover in Remote Oceania Shortly after Initial Settlement In Brief. *Current Biology*, *28*. https://doi.org/10.1016/j.cub.2018.02.051

Luker, V., & Buckingham, J. (2017). The Journal of Pacific History ISSN: (Print) (Histories of Leprosy: Subjectivities, Community and Pacific Worlds. *The Journal of Pacific History*, *52*(3), 265–286. https://doi.org/10.1080/00223344.2017.1379124

Magioli, M., Moreira, M. Z., Ferraz, K. M. B., Miotto, R. A., de Camargo, P. B., Rodrigues, M. G., da Silva Canhoto, M. C., & Setz, E. F. (2014). Stable Isotope Evidence of *Puma concolor* (*Felidae*) Feeding Patterns in Agricultural Landscapes in Southeastern Brazil. *Biotropica*, *46*(4), 451–460. https://doi.org/10.1111/BTP.12115

Mariotti, V., Dutour, O., Belcastro, M. G., Facchini, F., & Brasili, P. (2005). Probable early presence of leprosy in Europe in a Celtic skeleton of the 4th-3rd century BC (Casalecchio di Reno, Bologna, Italy). *International Journal of Osteoarchaeology*, *15*(5), 311–325. https://doi.org/10.1002/oa.775

Martinez, A. N., Britto, C. F. P. C., Nery, J. A. C., Sampaio, E. P., Jardim, M. R., Sarno, E. N., & Moraes, M. O. (2006). Evaluation of real-time and conventional PCR targeting complex 85 genes for detection of *Mycobacterium leprae* DNA in skin biopsy samples from patients diagnosed with leprosy. *Journal of Clinical Microbiology*, *44*(9), 3154–3159. https://doi.org/10.1128/JCM.02250-05

Maruyama, F. H., Morgado, T. O., Pacheco, R. C., Nakazato, L., & Dutra, V. (2018). Molecular detection of *Mycobacterium leprae* by Polymerase Chain Reaction in captive and free-ranging wild animals. *Brazilian Journal of Infectious Diseases*, 22(5). https://doi.org/10.1016/j.bjid.2018.09.003

Matheson, C. D., Vernon, K. K., Lahti, A., Fratpietro, R., Spigelman, M., Gibson, S., Greenblatt, C. L., & Donoghue, H. D. (2009). Molecular exploration of the first-century tomb of the shroud in Akeldama, Jerusalem. *PLoS ONE*, *4*(12). https://doi.org/10.1371/JOURNAL.PONE.0008319

Matisoo-Smith, E., Roberts, R. M., Irwin, G. J., Allen, J. S., Penny, D., & Lambert, D. M. (1998). Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15145–15150. https://doi.org/10.1073/PNAS.95.25.15145/ASSET/55D03863-C976-4A33-A409-A5679830982D/ASSETS/GRAPHIC/PQ2583927002.JPEG

Matsuoka, M. (2010). Drug Resistance in Leprosy. *Japanese Journal of Infectious Diseases*, *63*(1), 1–7. https://doi.org/10.7883/YOKEN.63.1

McCarthy, D., & Numa, J. (1962). Leprosy in the Cook Islands. *N Z Med J*, *61*(354), 78–79.

McDonough, S. J., Bhagwate, A., Sun, Z., Wang, C., Zschunke, M., Gorman, J. A., Kopp, K. J., & Cunningham, J. M. (2019). Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. *PLoS ONE*, *14*(4), e0211400. https://doi.org/10.1371/journal.pone.0211400

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6). http://www.cshprotocols.org/cgi/doi/10.1101/pdb.prot5448

Mohanty, P. S., Naaz, F., Katara, D., Misba, L., Kumar, D., Dwivedi, D. K., Tiwari, A. K., Chauhan, D. S., Bansal, A. K., Tripathy, S. P., & Katoch, K. (2016). Viability of *Mycobacterium leprae* in the environment and its role in leprosy dissemination. *Indian Journal of Dermatology, Venereology and Leprology*, *82*(1), 23. https://doi.org/10.4103/0378-6323.168935

Moller-Christensen, V. (1961). *Bone Changes in Leprosy*. Munksgaard.

Monot, M., Honoré, N., Garnier, T., Araoz, R., Coppée, J., Lacroix, C., Sow, S., Spencer, J. S., Truman, R. W., Williams, D. L., Gelber, R., Virmond, M., Flageul, B., Cho, S., Ji, B., Paniz-Mondolfi, A., Convit, J., Young, S., Fine, P. E., … Cole, S. (2005). On the Origin of Leprosy. *Science (New York, N.Y.)*, *308*(5724), 1040–1042. papers3://publication/doi/10.1126/science/1109759

Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., Matsuoka, M., Taylor, G. M., Donoghue, H. D., Bouwman, A., Mays, S., Watson, C., Lockwood, D., Khamispour, A., Dowlati, Y., Jianping, S., Rea, T. H., Vera-Cabrera, L., Stefani, M. M., … Cole, S. T. (2009). Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nature Genetics*, *41*(12), 1282–1289. http://www.nature.com/doifinder/10.1038/ng.477

Monteiro, L. D., Martins-Melo, F. R., Brito, A. L., Alencar, C. H., & Heukelbach, J. (2015). Spatial patterns of leprosy in a hyperendemic state in Northern Brazil, 2001-2012. *Revista de Saúde Pública*, *49*, 84. https://doi.org/10.1590/S0034-8910.2015049005866

Montgomerie, J. (1988). Leprosy in New Zealand. *J Polyn Soc.*, *97*(2), 115–152.

Neukamm, J., Pfrengle, S., Molak, M., Seitz, A., Francken, M., Eppenberger, P., Avanzi, C., Reiter, E., Urban, C., Welte, B., Stockhammer, P. W., Teßmann, B., Herbig, A., Harvati, K., Nieselt, K., Krause, J., & Schuenemann, V. J. (2020). 2000-year-old pathogen genomes reconstructed from metagenomic analysis of Egyptian mummified individuals. *BMC Biology*, *18*(1), 1–18. https://doi.org/10.1186/S12915-020-00839-8/FIGURES/4

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566

Petchey, F., Clark, G., Winter, O., O'day, P., & Litster, M. (2017). Colonisation of Remote Oceania: New dates for the Bapot-1 site in the Mariana Islands. *Archaeology in Oceania*, *52*(2), 108–126. https://doi.org/10.1002/ARCO.5108

Pietrusewsky, M., & Douglas, M. T. (2015). History of Paleopathology in the Pacific. *The Global History of Paleopathology*, 594–616. https://doi.org/10.1093/ACPROF:OSOBL/9780195389807.003.0063

Ploemacher, T., Faber, W. R., Menke, H., Rutten, V., & Pieters, T. (2020). Reservoirs and transmission routes of leprosy; A systematic review. *PLOS Neglected Tropical Diseases*, *14*(4), e0008276. https://doi.org/10.1371/JOURNAL.PNTD.0008276

Posth, C., Nägele, K., Colleran, H., Valentin, F., Bedford, S., Kami, K. W., Shing, R., Buckley, H., Kinaston, R., Walworth, M., Clark, G. R., Reepmeyer, C., Flexner, J., Maric, T., Moser, J., Gresky, J., Kiko, L., Robson, K. J., Auckland, K., … Powell, A. (2018). Language continuity despite population replacement in Remote Oceania. *Nature Ecology & Evolution 2018 2:4*, *2*(4), 731–740. https://doi.org/10.1038/s41559-018-0498-2

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. Https://Www.R-Project.Org/}.

Reibel, F., Chauffour, A., Brossier, F., Jarlier, V., Cambau, E., & Aubry, A. (2015). New Insights into the Geographic Distribution of *Mycobacterium leprae* SNP Genotypes Determined for Isolates from Leprosy Cases Diagnosed in Metropolitan France and French Territories. *PLOS Neglected Tropical Diseases*, *9*(10), e0004141-10. https://dx.plos.org/10.1371/journal.pntd.0004141

Ridley, D. S., & Jopling, W. H. (1966). Classification of leprosy according to immunity. A five-group system. *International Journal of Leprosy*, *34*(3), 255–273.

Robbins, G., Mushrif Tripathy, V., Misra, V. N., Mohanty, R. K., Shinde, V. S., Gray, K. M., & Schug, M. D. (2009). Ancient Skeletal Evidence for Leprosy in India (2000 B.C.). *PLOS ONE*, *4*(5), e5669. https://doi.org/10.1371/JOURNAL.PONE.0005669

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology 2011 29:1*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil – DNA – glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660). https://doi.org/10.1098/rstb.2013.0624

Rolett, B. v. (2002). *Voyaging and Interaction in Ancient East Polynesia*.

Romero-Montoya, M., Beltran-Alzate, J. C., & Cardona-Castro, N. (2017). Evaluation and Monitoring *of Mycobacterium leprae* Transmission in Household Contacts of Patients with Hansen's Disease in Colombia. *PLOS Neglected Tropical Diseases*, *11*(1), e0005325. https://doi.org/10.1371/JOURNAL.PNTD.0005325

Rosa, P. S., D'Espindula, H. R. S., Melo, A. C. L., Fontes, A. N. B., Finardi, A. J., Belone, A. F. F., Sartori, B. G. C., Pires, C. A. A., Soares, C. T., Marques, F. B., Branco, F. J. D., Baptista, I. M. F. D., Trino, L. M., Fachin, L. R. V., Xavier, M. B., Floriano, M. C., Ura, S., Diório, S. M., Delanina, W. F. B., … Mira, M. T. (2020). Emergence and transmission of drug-/multidrug-resistant *Mycobacterium leprae* in a former leprosy colony in the Brazilian amazon. *Clinical Infectious Diseases*, *70*(10), 2054–2061. https://doi.org/10.1093/cid/ciz570

Saitou, N., & Nei, M. (n.d.). *The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees'*. Retrieved May 23, 2024, from https://academic.oup.com/mbe/article/4/4/406/1029664

Sarnecka, A. K., Nawrat, D., Piwowar, M., Ligęza, J., Swadźba, J., & Wójcik, P. (2019). DNA extraction from FFPE tissue samples – a comparison of three procedures. *Wspolczesna Onkologia*, *23*(1), 52–58. https://doi.org/10.5114/wo.2019.83875

Schaub, R., Avanzi, C., Singh, P., Paniz-Mondolfi, A., Cardona-Castro, N., Legua, P., Crespo, L., Sewpersad, K., Dávila, J. J., Barreto, J., Dwivedi, P., Morris-Wilson, H., Larrea, M. P., Talhari, C., Lahiri, R., Truman, R. W., Gozlan, R. E., Couppié, P., & de Thoisy, B. (2020). Leprosy Transmission in Amazonian Countries: Current Status and Future Trends. *Current Tropical Medicine Reports*, *7*(3), 79–91. https://doi.org/10.1007/S40475-020-00206-1/TABLES/2

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, 1–7. papers3://publication/doi/10.1186/s13104-016-1900-2

Schuenemann, V. J., Avanzi, C., Krause-Kyora, B., Seitz, A., Herbig, A., Inskip, S., Bonazzi, M., Reiter, E., Urban, C., Dangvard Pedersen, D., Taylor, G. M., Singh, P., Stewart, G. R., Velemínský, P., Likovsky, J., Marcsik, A., Molnár, E., Pálfi, G., Mariotti, V., … Krause, J. (2018). Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. *PLOS Pathogens*, *14*(5), e1006997. https://doi.org/10.1371/journal.ppat.1006997

Schuenemann, V. J., Singh, P., Mendum, T. A., Krause-Kyora, B., Jäger, G., Bos, K. I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J. L., Kjellström, A., Wu, H., Stewart, G. R., Taylor, G. M., Bauer, P., Lee, O. Y. C., Wu, H. H. T., … Krause, J. (2013). Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science (New York, N.Y.)*, *341*(6142), 179–183. http://www.sciencemag.org/cgi/doi/10.1126/science.1238286

Scollard, D. M., Adams, L. B., Gillis, T. P., Krahenbuhl, J. L., Truman, R. W., & Williams, D. L. (2006). The Continuing Challenges of Leprosy. *CLINICAL MICROBIOLOGY REVIEWS*, *19*(2), 338–381. https://doi.org/10.1128/CMR.19.2.338-381.2006

Sharma, R., Singh, P., Loughry, W. J., Lockhart, J. M., Inman, W. B., Duthie, M. S., Pena, M. T., Marcos, L. A., Scollard, D. M., Cole, S. T., & Truman, R. W. (2015). Zoonotic Leprosy in the Southeastern United States. *Emerging Infectious Diseases*, *21*(12), 2127–2134. https://doi.org/10.3201/eid2112.150501

Shields, E. D., Russell, D. A., & Pericak-Vance, M. A. (1987). Genetic epidemiology of the susceptibility to leprosy. *The Journal of Clinical Investigation*, *79*(4), 1139–1143. https://doi.org/10.1172/JCI112930

Silva, N., & Fernandez, P. (2006). Mai Ka ’ Aina O Ka ’ Eha ’ eha Mai : Testimonies of Hansen’s Disease Patients in Hawai’i, 1866-1897. *Hawaii J Hist*, *40*, 75–97.

Singh, P., Benjak, A., Carat, S., Kai, M., Busso, P., Avanzi, C., Paniz-Mondolfi, A., Peter, C., Harshman, K., Rougemont, J., Matsuoka, M., & Cole, S. T. (2014). Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3. *Clinical Microbiology and Infection*, *20*(10), O619–O622. https://doi.org/10.1111/1469-0691.12609

Singh, P., & Cole, S. T. (2011). Mycobacterium leprae: genes, pseudogenes and genetic diversity. *Future Microbiology*, *6*(1), 57–71. https://doi.org/10.2217/fmb.10.153

Skinsnes, O. (1973). Notes from the history of leprosy. *Int J Lepr*, *41*(2), 220–233.

Skinsnes, O. K. (1980). Leprosy in archeologically recovered bamboo book in China. *International Journal of Leprosy and Other Mycobacterial Diseases : Official Organ of the International Leprosy Association*, *48*(3), 333.

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G. R., Reepmeyer, C., Petchey, F., Fernandes, D., Fu, Q., Harney, E., Lipson, M., Mallick, S., Novak, M., Rohland, N., Stewardson, K., Abdullah, S., Cox, M. P., … Reich, D. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature 2016 538:7626*, *538*(7626), 510–513. https://doi.org/10.1038/nature19844

Smith, C. S., Aerts, A., Saunderson, P., Kawuma, J., Kita, E., & Virmond, M. (2017). Multidrug therapy for leprosy: a game changer on the path to elimination. *The Lancet Infectious Diseases*, *17*(9), e293–e297. https://doi.org/10.1016/S1473-3099(17)30418-8

Spriggs, M. (2011). Archaeology and the Austronesian expansion: where are we now? *Antiquity*, *85*(328), 510–528. https://doi.org/10.1017/S0003598X00067910

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu033

Stefani, de A. (2008). Challenges in the post genomic era for the development of tests for leprosy diagnosis Desafios na era pós genômica para o desenvolvimento de testes laboratoriais para o diagnóstico da hanseníase. *Revista Da Sociedade Brasileira de Medicina Tropical*, *41*, 89–94.

Stefani, M. M. A., Avanzi, C., Bührer-Sékula, S., Benjak, A., Loiseau, C., Singh, P., Pontes, M. A. A., Gonçalves, H. S., Hungria, E. M., Busso, P., Piton, J., Silveira, M. I. S., Cruz, R., Schetinni, A., Costa, M. B., Virmond, M. C. L., Diorio, S. M., Dias-Baptista, I. M. F., Rosa, P. S., … Penna, G. O. (2017). Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases. *PLoS Neglected Tropical Diseases*, *11*(6). https://doi.org/10.1371/journal.pntd.0005598

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). https://doi.org/10.1093/VE/VEY016

Superina, M., Pagnutti, N., & Abba, A. M. (2014). What do we know about armadillos? An analysis of four centuries of knowledge about a group of South American mammals, with emphasis on their conservation. *Mammal Review*, *44*(1), 69–80. https://doi.org/10.1111/MAM.12010

Suzuki, K., Saso, A., Hoshino, K., Sakurai, J., Tanigawa, K., Luo, Y., Ishido, Y., Mori, S., Hirata, K., & Ishii, N. (2014). Paleopathological Evidence and Detection of *Mycobacterium leprae* DNA from Archaeological Skeletal Remains of Nabe-kaburi (Head-Covered with Iron Pots) Burials in Japan. *PLOS ONE*, *9*(2), e88356. https://doi.org/10.1371/JOURNAL.PONE.0088356

Suzuki, K., Takigawa, W., Tanigawa, K., Nakamura, K., Ishido, Y., Kawashima, A., Wu, H., Akama, T., Sue, M., Yoshihara, A., Mori, S., & Ishii, N. (2010). Detection of *Mycobacterium leprae* DNA from Archaeological Skeletal Remains in Japan Using Whole Genome Amplification and Polymerase Chain Reaction. *PLOS ONE*, *5*(8), e12422. https://doi.org/10.1371/JOURNAL.PONE.0012422

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Tayles, N., & Buckley, H. R. (2004). Leprosy and tuberculosis in Iron Age Southeast Asia? *American Journal of Physical Anthropology*, *125*(3), 239–256. https://doi.org/10.1002/AJPA.10378

Taylor, G. M., Blau, S., Mays, S., Monot, M., Lee, O. Y. C., Minnikin, D. E., Besra, G. S., Cole, S. T., & Rutland, P. (2009). *Mycobacterium leprae* genotype amplified from an archaeological case of lepromatous leprosy in Central Asia. *Journal of Archaeological Science*, *36*(10), 2408–2414. https://doi.org/10.1016/J.JAS.2009.06.026

Tió-Coma, M., Avanzi, C., Verhard, E. M., Pierneef, L., van Hooij, A., Benjak, A., Roy, J. C., Khatun, M., Alam, K., Corstjens, P., Cole, S. T., Richardus, J. H., & Geluk, A. (2020). Genomic Characterization of *Mycobacterium leprae* to Explore Transmission Patterns Identifies New Subtype in Bangladesh. *Frontiers in Microbiology*, *11*, 1220. https://doi.org/10.3389/fmicb.2020.01220

Tió-Coma, M., Wijnands, T., Pierneef, L., Schilling, A. K., Alam, K., Roy, J. C., Faber, W. R., Menke, H., Pieters, T., Stevenson, K., Richardus, J. H., & Geluk, A. (2019). Detection of *Mycobacterium leprae* DNA in soil: multiple needles in the haystack. *Scientific Reports 2019 9:1*, *9*(1), 1–7. https://doi.org/10.1038/s41598-019-39746-6

Trembly, D. (2002). Perspectives on the history of leprosy in the Pacific. In C. Roberts, M. Lewis, & K. Manchester (Eds.), *The past and present of leprosy: archaeological, historical, paleopathological, and clinical approaches* (pp. 233–238). Archaeopress.

Trembly, D. L. (1995). On the antiquity of leprosy in western Micronesia. *International Journal of Osteoarchaeology*, *5*(4), 377–384. https://doi.org/10.1002/OA.1390050408

Truman, R. W., Andrews, P. K., Robbins, N. Y., Adams, L. B., Krahenbuhl, J. L., & Gillis, T. P. (2008). Enumeration of *Mycobacterium leprae* using real-time PCR. *PLoS Neglected Tropical Diseases*, *2*(11). https://doi.org/10.1371/journal.pntd.0000328

Truman, R. W., Singh, P., Sharma, R., Busso, P., Rougemont, J., Paniz-Mondolfi, A., Kapopoulou, A., Brisse, S., Scollard, D. M., Gillis, T. P., & Cole, S. T. (2011). Probable Zoonotic Leprosy in the Southern United States. *New England Journal of Medicine*, *364*(17), 1626–1633. https://doi.org/10.1056/NEJMoa1010536

Turankar, R. P., Lavania, M., Singh, M., Sengupta, U., Siva Sai, K., & Jadhav, R. S. (2016). Presence of viable *Mycobacterium leprae* in environmental specimens around houses of leprosy patients. *Indian Journal of Medical Microbiology*, *34*(3), 315–321. https://doi.org/10.4103/0255-0857.188322

Turankar, R. P., Singh, V., Lavania, M., Singh, I., Sengupta, U., & Jadhav, R. S. (2022). Existence of viable *Mycobacterium leprae* in natural environment and its genetic profiling in a leprosy endemic region. *Frontiers in Tropical Diseases*, *3*. https://doi.org/10.3389/fitd.2022.972682

Urban, C., Blom, A. A., Avanzi, C., Walker-Meikle, K., Warren, A. K., White-Iribhogbe, K., Turle, R., Marter, P., Dawson-Hobbis, H., Roffey, S., Inskip, S. A., & Schuenemann, V. J. (2024). Ancient *Mycobacterium leprae* genome reveals medieval English red squirrels as animal leprosy host. *Current Biology*, *34*(10), 2221-2230.e8. https://doi.org/10.1016/J.CUB.2024.04.006/ATTACHMENT/BF6CAE55-0BE2-4CAE-9C48-354FA96FC996/MMC7.XLSX

Urban, C., Blom, A. A., Pfrengle, S., Walker-Meikle, K., Stone, A. C., Inskip, S. A., & Schuenemann, V. J. (2021). One Health Approaches to Trace *Mycobacterium leprae*'s Zoonotic Potential Through Time. *Frontiers in Microbiology*, *12*, 762263. https://doi.org/10.3389/FMICB.2021.762263/BIBTEX

van der Auwera, G., & O'Connor, B. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (1st Edition). O'Reilly.

Vera-Cabrera, L., Escalante-Fuentes, W. G., Gomez-Flores, M., Ocampo-Candiani, J., Busso, P., Singh, P., & Cole, S. T. (2011). Case of diffuse lepromatous leprosy associated with "*Mycobacterium lepromatosis*." *Journal of Clinical Microbiology*, *49*(12), 4366–4368. https://doi.org/10.1128/JCM.05634-11

Viana, M., Mancy, R., Biek, R., Cleaveland, S., Cross, P. C., Lloyd-Smith, J. O., & Haydon, D. T. (2014). Assembling evidence for identifying reservoirs of infection. *Trends in Ecology & Evolution*, *29*(5), 270–279. https://doi.org/10.1016/J.TREE.2014.03.002

Walker, S. L., & Lockwood, D. N. J. (2007). Leprosy. *Clinics in Dermatology*, *25*(2), 165–172. https://doi.org/10.1016/j.clindermatol.2006.05.012

Weisler, M. I., Bolhar, R., Ma, J., Pierre, E. S., Sheppard, P., Walter, R. K., Feng, Y., Zhao, J. X., & Kirch, P. v. (2016). Cook Island artifact geochemistry demonstrates spatial and temporal extent of pre-European interarchipelago voyaging in East Polynesia. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8150–8155. https://doi.org/10.1073/PNAS.1608130113/SUPPL_FILE/PNAS.1608130113.SD03.XLSX

Wheat, W. H., Casali, A. L., Thomas, V., Spencer, J. S., Lahiri, R., Williams, D. L., McDonnell, G. E., Gonzalez-Juarrero, M., Brennan, P. J., & Jackson, M. (2014). Long-term Survival and Virulence of *Mycobacterium leprae* in Amoebal Cysts. *PLOS Neglected Tropical Diseases*, *8*(12), e3405. https://doi.org/10.1371/JOURNAL.PNTD.0003405

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Williams, D. L., & Gillis, T. P. (2012). Drug-resistant leprosy: monitoring and current status. *Leprosy Review*, *83*(3), 269–281.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 1–13. https://doi.org/10.1186/S13059-019-1891-0/FIGURES/2

Woods, S. A., & Cole, S. T. (1990). A family of dispersed repeats in *Mycobacterium leprae*. *Molecular Microbiology*.

World Health Organization. (2004). *Status report on leprosy in the WHO Western Pacific Region, 2002*. http://www.who.int/lep/resources/SWG01.pdf

World Health Organization. (2011). *Epidemiological review of leprosy in the western pacific region 2008-2010*.

World Health Organization. (2016). *Global leprosy strategy 2016-2020*. https://doi.org/978-92-9022-509-6

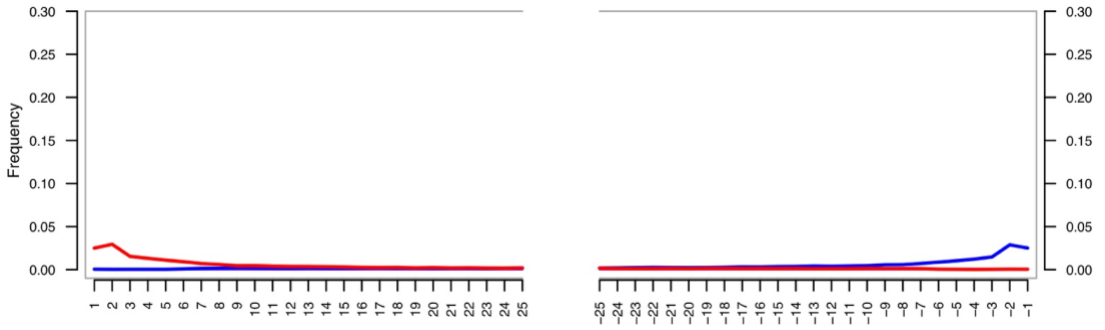
World Health Organization. (2017). *Global leprosy update, 2017: reducing the disease burden due to leprosy*.

World Health Organization. (2019). Global leprosy update, 2018: moving towards a leprosy free world. *Weekly Epidemiological Record*, *35/36*(94), 389–412. https://apps.who.int/iris/bitstream/handle/10665/326775/WER9435-36-en-fr.pdf

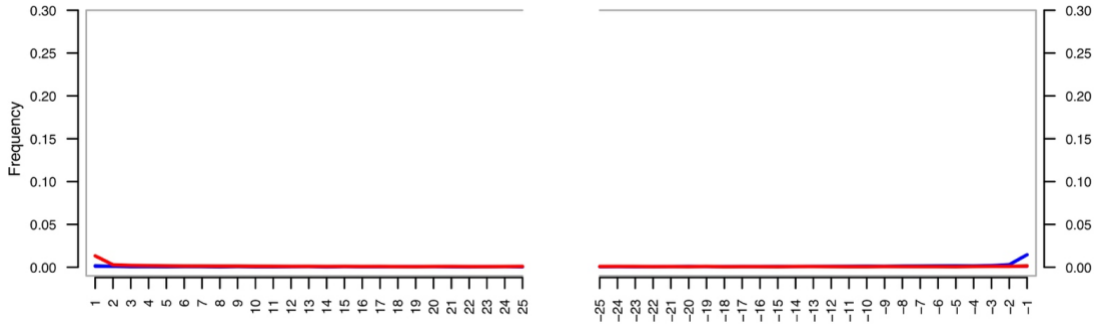
World Health Organization. (2023). Global leprosy (Hansen disease) update, 2022: new paradigm - control to elimination. *WHO Weekly Epidemiological Record*, *37*, 409–430.

APPENDIX A

SUPPLEMENTARY FIGURES 1 AND 2: UDG AND NON-UDG FRAGMENT ANALYSIS



SI Figure 1. Sample 511 DBT extraction without UDG treatment fragment misincorporation plot showing the frequency of A -> G (left) and C -> T (right) transitions



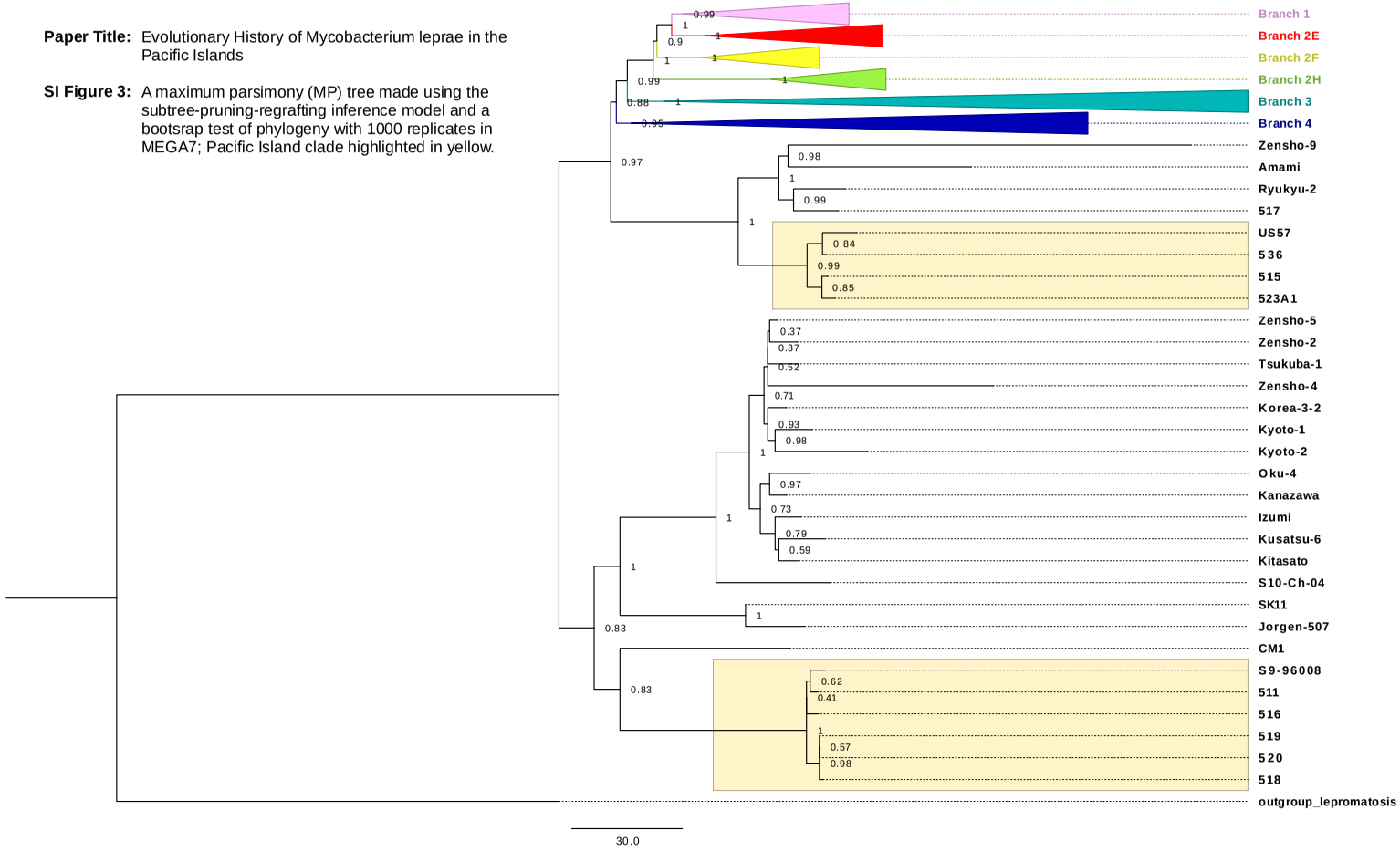
SI Figure 2. Sample 511 DAB extraction with UDG treatment fragment misincorporation plot showing the frequency of A -> G (left) and C -> T (right) transitions

APPENDIX B

SUPPLEMENTARY FIGURE 3: MAXIMUM PARSIMONY TREE

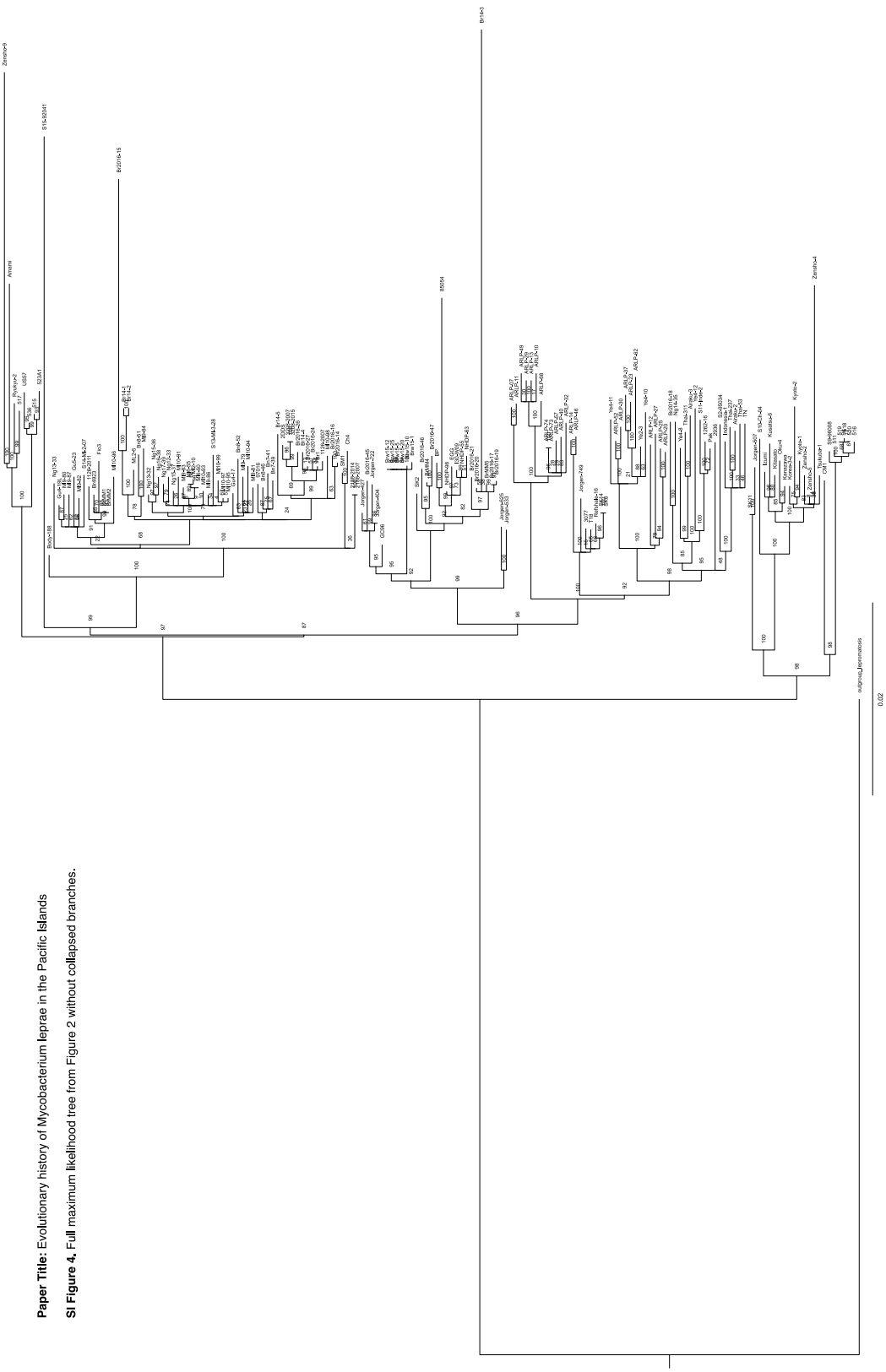
Paper Title: Evolutionary History of Mycobacterium leprae in the Pacific Islands

SI Figure 3: A maximum parsimony (MP) tree made using the subtree-pruning-regrafting inference model and a bootsrap test of phylogeny with 1000 replicates in MEGA7; Pacific Island clade highlighted in yellow.



APPENDIX C

SUPPLEMENTARY FIGURE 4: MAXIMUM LIKELIHOOD TREE



Paper Title: Evolutionary history of *Mycobacterium* leprae in the Pacific Islands

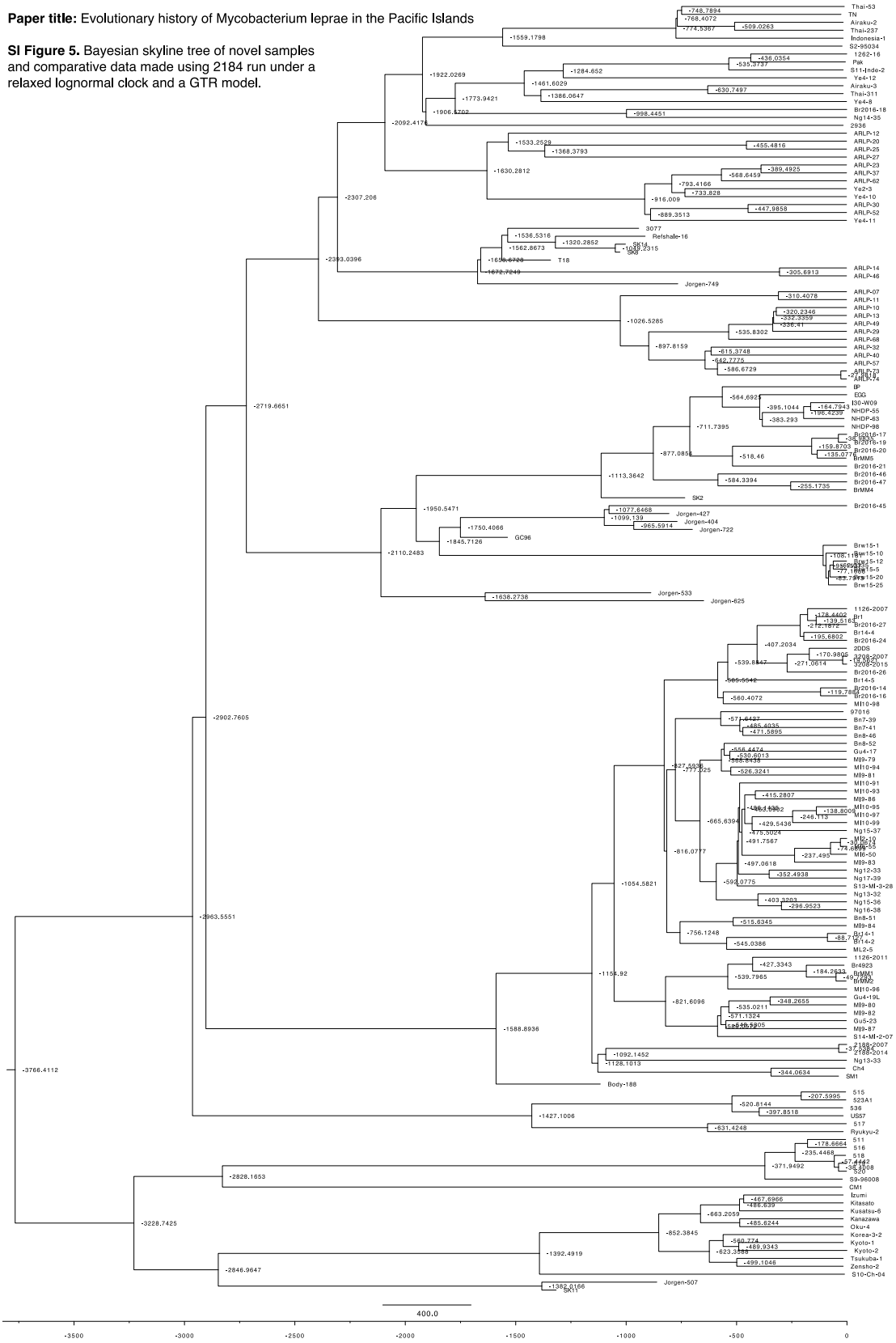
SI Figure 4. Full maximum likelihood tree from Figure 2 without collapsed branches.

APPENDIX D

SUPPLEMENTARY FIGURE 5: BAYESIAN SKYLINE TREE WITH GTR MODEL

Paper title: Evolutionary history of Mycobacterium leprae in the Pacific Islands

SI Figure 5. Bayesian skyline tree of novel samples and comparative data made using 2184 run under a relaxed lognormal clock and a GTR model.



APPENDIX E

SUPPLEMENTARY TABLE 1: COMPARATIVE SAMPLE INFORMATION

[CONSULT ATTACHED FILE]

APPENDIX F

SUPPLEMENTARY TABLE 2: CHAPTER 2 SCREENING INFORMATION

[CONSULT ATTACHED FILE]

APPENDIX G

SUPPLEMENTARY TABLE 3: METHODS COMPARISON FOR CHAPTER 2

METHODS

[CONSULT ATTACHED FILE]

APPENDIX H

SUPPLEMENTARY TABLE 4: SUMMARY OF DNA DAMAGE PATTERNS FOR CHAPTER 2

Supplementary Table 4. Summary of DNA damage patterns for all samples that were sequenced without UDG treatment and samples that were sequenced with and without having been UDG treated

Sample	Extraction type	G to A misincorporation 1st Base 3'	G to A misincorporation 2nd Base 3'	C to T misincorporation 1st Base 5'	C to T misincorporation Base 5'
511	DBT	0.025	0.028	0.025	0.029
	DAB + UDG treatment	0.014	0.003	0.013	0.002
515	DBT	0.035	0.034	0.037	0.037
	DAB + UDG treatment	0.022	0.004	0.022	0.004
519	DBT	0.022	0.023	0.023	0.023
	DAB + UDG treatment	0.025	0.004	0.027	0.004
523A1	DBT	0.015	0.016	0.016	0.018
	DAB + UDG treatment	0.034	0.012	0.034	0.012
516	DBT	0.02	0.019	0.02	0.018
517	DBT	0.031	0.032	0.031	0.033
518	DBT	0.026	0.026	0.026	0.026
520	DBT	0.024	0.023	0.024	0.023

APPENDIX I

SUPPLEMENTARY TABLE 5: CONCATENATED SNPEFF OUTPUTS.

[CONSULT ATTACHED FILE]

APPENDIX J

SUPPLEMENTARY TABLE 6: UNIQUE SNP AND SNP-EFFECT LIST

[CONSULT ATTACHED FILE]

APPENDIX K

SUPPLEMENTARY TABLE 7: BRANCH DEFINING SNPS

[CONSULT ATTACHED FILE]

APPENDIX L

SUPPLEMENTARY TABLE 8: SUBSTITUTION RATES AND TMRCA ESTIMATES ACROSS SNP ALIGNMENTS

[CONSULT ATTACHED FILE]

APPENDIX M

SUPPLEMENTARY DOCUMENT 1: PAIRED DNA METHODS AND ANALYSES AS DESCRBIED IN KING (2023, MASTER'S THESIS, ASU)

This document contains modified methods for Chapter 3 DNA extractions as used in the thesis King (2023, ProQuest) for the project "Comparison of Two DNA Extraction Methods for Isolating *Mycobacterium leprae* DNA from FFPE Tissue Collected in the Pacific Islands Region". The methods of DNA extraction listed are for the n=20 paired samples used in the Chapter 3 analysis and therefore only include modifications to the QIAmp DNA FFPE Kit (Qiagen) and the Hot Alkaline Lysis Phenol-Chloroform procedure (Campos & Gilbert 2014 with recommendations by Hahn et al., 2021). The modifications were designed by King (2023, ProQuest) as per manufacturer or source article recommendations.

Protocol 1: QIAmp DNA FFPE Kit (Qiagen)

Manufacturer recommendations as seen in the QIAmp Handbook, 2020 (Qiagen) were followed with several modifications. Residual paraffin was removed from the tissue sample within the FFPE block by centrifugation with xylene. Samples were pelleted with 80% ethanol and then underwent an extended incubation of 56°C for three hours with proteinase K. Lysis was not complete at this point, so 180 uL of Buffer ATL (Qiagen) was added and incubation at 56°C was extended overnight for a total of 20 hours. A heat step of 90°C for 1 hour occurred after the overnight incubation. Kit protocols were followed for extraction with a final elution volume of 60 uL.

Protocol 2: Hot Alkaline Lysis Phenol-Chloroform

The Hot Alkaline Lysis Phenol-Chloroform procedure (HA-Lysis) was originally developed by Campos and Gilbert (2014) and further modified in Hahn et al. (2021). The samples processed in King (2023, ProQuest) and used for analysis in Chapter 3 in this dissertation followed recommendations made by Hahn et al. (2021). The HA-Lysis procedure was optimized for tissue slices with 3-10 μm thickness, however samples here were manually excised with a scalpel and diced carefully to increase surface area. Following the autoclave lysis, extractions were conducted directly following Campos and Gilbert (2014). There was an optional solution listed in the original protocol for pelleting, however this was not used. Pellets were resuspended in 60 uL of TE buffer.

APPENDIX N

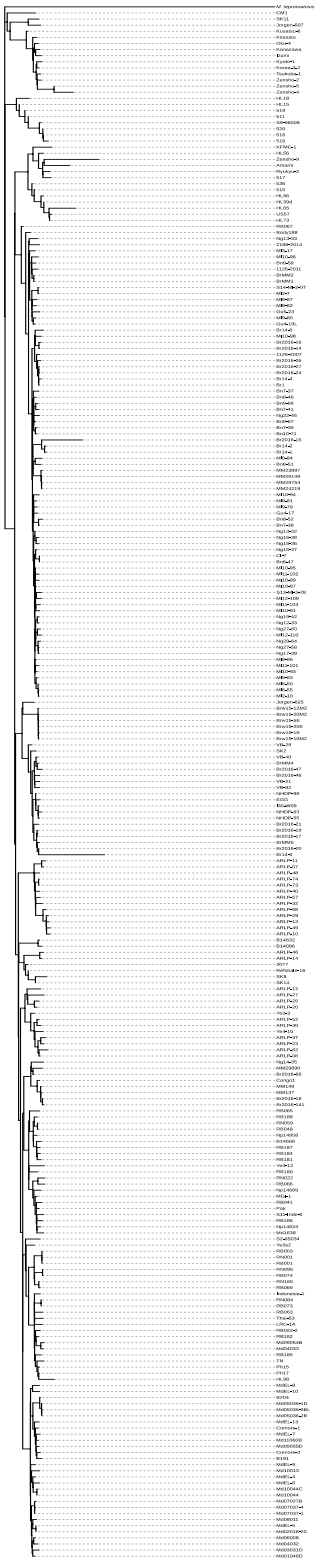
SUPPLEMENTARY TABLE 9: SAMPLE METADATA FOR CHAPTER 3 SCREENING

[CONSULT ATTACHED FILE]

APPENDIX O

SUPPLEMENTARY FIGURE 6: NEIGHBOR-JOINING TREE

Tree scale: 0.1



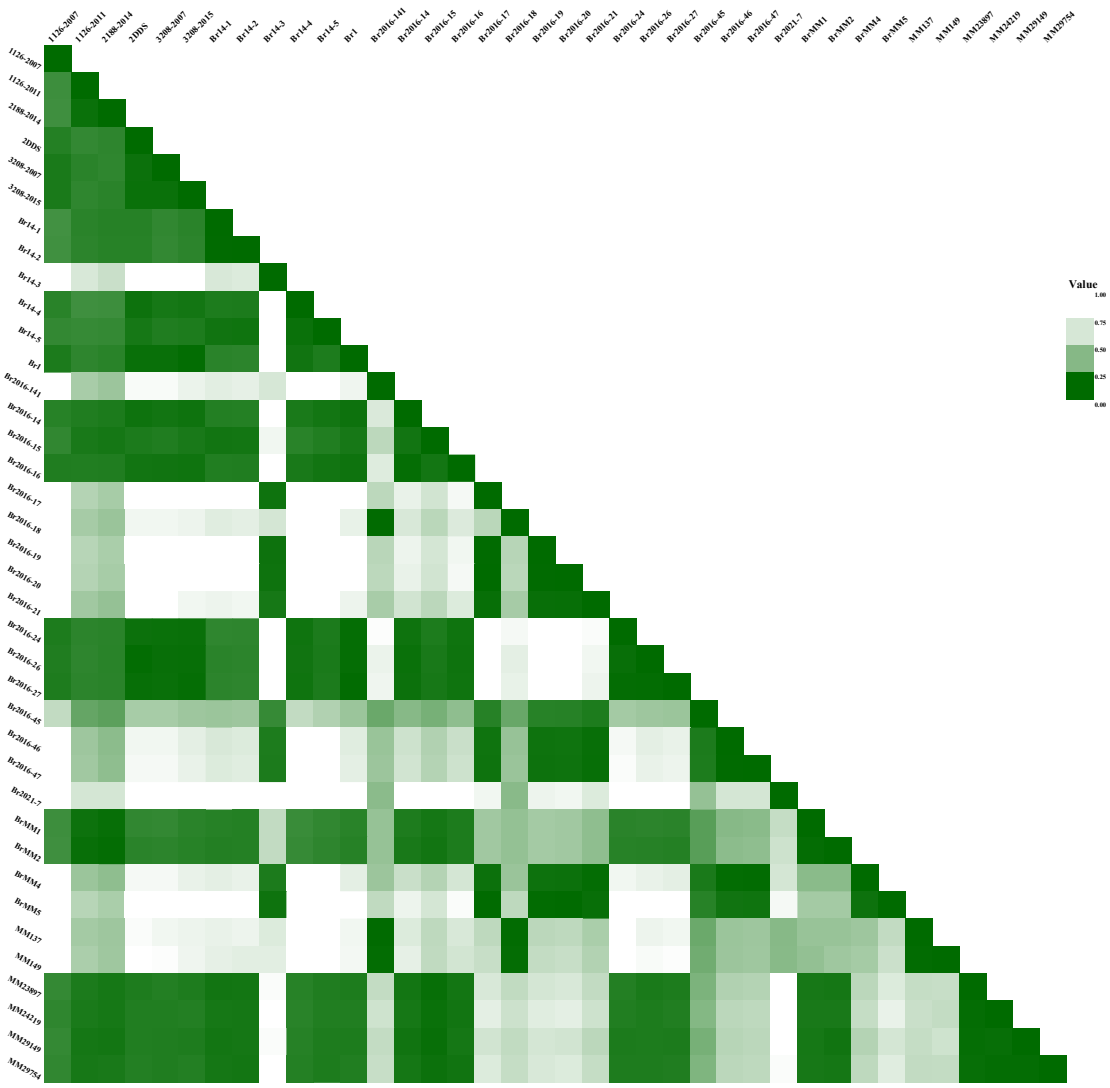
APPENDIX P

SUPPLEMENTARY TABLE 10: PERU SAMPLE METADATA AND *M. LEPRAE* SCREENING

[CONSULT ATTACHED FILE]

APPENDIX Q

SUPPLEMENTARY FIGURE 7: DISTANCE MATRIX OF BRAZIL STRAINS IN CHAPTER 4


1126-2007
1126-2011
2188-2014
2DDS
3208-2007
3208-2015
Br14-1
Br14-2
Br14-3
Br14-4
Br14-5
Br1
Br2016-141
Br2016-14
Br2016-15
Br2016-16
Br2016-17
Br2016-18
Br2016-19
Br2016-20
Br2016-21
Br2016-24
Br2016-26
Br2016-27
Br2016-45
Br2016-46
Br2016-47
Br2021-7
BrMM1
BrMM2
BrMM4
BrMM5
MM137
MM149
MM23097
MM24219
MM29149
MM29754
Value
1.00
0.75
0.50
0.25
0.00