Comparison of Two DNA Extraction Methods for Isolating *Mycobacterium leprae* DNA from FFPE Tissue Collected in the Pacific Islands Region

by

Felicia King

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2023 by the
Graduate Supervisory Committee:

Anne Stone, Chair
Melissa Wilson
Ken Buetow

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

Mycobacterium leprae, the causative agent of Hansen's disease (leprosy), has plagued humans and other animal species for millennia and remains of concern to public health throughout the world today. Recent research into the expanded use of medical tissues preserved as formalin-fixed, paraffin-embedded samples (FFPE), opened the door for the study of M. leprae DNA from preserved skin samples. However, problems persist with damage to the DNA including fragmentation and cross linkage. This study evaluated two methods commonly used for the recovery of host DNA from FFPE samples for their efficacy in extracting pathogen DNA (hot alkaline lysis protocol and QIAGEN QIAamp FFPE DNA kit). Twenty FFPE skin samples collected from 1995-2015 from human subjects in the Pacific Islands suffering from M. leprae infection, each exhibiting a range of bacillary loads, were analyzed to determine which extraction method was most successful in terms of ability to consistently yield reliable, robust traces of M. leprae infection. This study further examined these samples to understand the phylogeny of leprosy in the region, where gaps in the evolutionary history of M. leprae persist.

DNA recovery from paired samples was similar using either method. However, by extending the incubation time of post-paraffin removal sample lysis, both protocols were more likely to yield positive traces of M. leprae, with this enhancement being especially evident in paucibacillary samples with low bacterial presence. The qPCR assay findings suggest that the hot alkaline procedure is most likely to yield positive identification of infection in these traditionally challenging samples.

ACKNOWLEDGEMENTS

The authors acknowledge the Hawaiian Pathologists Laboratory for the use of FFPE samples from their facilities, as well as the ASU Genomics Facilities which allowed the use of their machines. Additional thanks go to the Research Computing at Arizona State University for providing HPC and storage resources that have contributed to the research results reported within this paper.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

INTRODUCTION

Though sometimes regarded as a disease from antiquity, leprosy (or Hansen's Disease) remains a serious public health concern today. The World Health Organization reported 127,558 new cases in 2020 (WHO Leprosy Fact Sheet, 2023) and leprosy is classified as a Neglected Tropical Disease- a diverse group of diseases disproportionately affecting impoverished nations in the tropics that have largely remain understudied. While multi-drug therapy strategies have been developed and implemented with considerable success (Smith et al., 2017), leprosy remains endemic in regions worldwide and is particularly prevalent in low socioeconomic areas where treatment is hampered by the extreme stigma of the disease and lack of healthcare access (Pescarini et al., 2018).

The primary causative agent for leprosy, *Mycobacterium leprae*, is an obligate, intracellular pathogen known to target the upper respiratory mucosa and peripheral nervous system of its host (Britton & Lockwood, 2004). Some cases, however, can be attributed to the less common *M. lepromatosis* which is largely found in Mexico, the Caribbean (in humans), and the UK (in red squirrels; Han et al., 2008, Avanzi et al., 2016). The two *Mycobacterium* species diverged from one another approximately 13 million years ago and are both strong examples of reductive evolution as their genomes have decreased to only around 3.2 million base pairs. An estimated 1,614 coding genes have been identified, while nearly half the total genome of both species is composed of pseudogenes (Singh et al., 2011). With an overall G+C content of 57.8%, the leprosy

bacillus is also the most A+T rich of the Mycobacterium, which can result in complications with downstream analysis.

In clinical settings, leprosy is diagnosed using the Ridley-Jopling scale (Ridley & Jopling, 1966), a system that ranks severity from tuberculoid (TT) to borderline (BB) to lepromatous (LL), with intermediate classifications between each. The least severe form, tuberculoid, is characterized by strong cell mediated immune response with a small number of bacilli-containing granulomas (Avanzi et al., 2016). Conversely, the lepromatous form refers to the extensive presence of diffuse granuloma with large quantities of bacilli. The more recent WHO classification of leprosy (1982) refers to the bacillary load, ranging on a gradient scale from the multibacillary lepromatous cases to the paucibacillary tuberculoid cases. While the tuberculoid cases are clinically favorable, the lower quantity of bacteria leads to increased difficulty with sequencing. On the other hand, lepromatous cases often result in more favorable conditions for genetic research and sequencing.

The *M. leprae* phylogeny is characterized by six primary branches (1, 2, 3, 4, 5, and 0) containing sixteen subtypes (A-O), differentiated by a varied set of diagnostic nucleotide polymorphisms (SNPs). General phylogenetic trends broadly associate Branch 1 with South Asia, Branch 2 with East Africa, Branch 3 in Europe, Branch 4 with West Africa, and Branch 0 in East Asia (Schuenemann et al., 2018, Monot et al. 2005; 2009). While the understanding of dispersal has improved due to widespread recent studies, geographical gaps such as in the Pacific Islands, persist. Blevins et al. 2020 sought to

address this disparity by adding nine novel strains from the Pacific. These strains fell into Branches 0 and 5 which are associated with existing Pacific and East Asians strains.

Neither species of leprosy causing *Mycobacterium* have been successfully cultured in artificial, laboratory settings and, as a result, the study of *M. leprae* primarily consists of culturing the bacterium in mouse footpads over a six-month period or clinical sampling from patients. Modern methods of sampling include slit-skin smears (SSS), nasal swabs (NS), and tissue sampling from leprosy lesions (Avanzi, 2016). While SSS and NS are far less invasive methods, DNA extractions from these samples typically yield very low quantities of bacterial DNA, which limits their use to lepromatous presentations. The use of NS is further complicated by nasal carriage of *M. leprae* in individuals who are not infected, which can result in false positive identification (Beyene et al., 2003) of infection. Questions remain as to whether nasal carriage is a true asymptomatic infection or a passive carriage from infected household members. Because of these shortcomings, leprosy lesion sampling, though invasive, remains the most efficient way to capture quality strains.

Clinical tissues collections are often fixed in formalin and embedded in paraffin wax to prevent the decay of the infected tissue. These formalin-fixed paraffin-embedded (FFPE) tissue samples are useful for cell morphology and immunohistochemistry research (Donoghue et al., 2015). Such samples are an excellent resource for host DNA and RNA research (Gaffney et al., 2018) and have been studied extensively in cancer evolution research (Ghagwate et al., 2019, McDonough et al., 2019 Arreaza et al., 2017). However, the use of FFPE for the study of pathogen DNA from host samples, particularly

M. leprae, is still largely novel due to the damage the fixation process causes to the DNA. Specifically, FFPE preservation often results in deamination, cross-linkage, and fragmentation of the DNA. Existing research has attempted to identify a method of extracting that degraded DNA with mixed results (Sarnecka et al. 2019, Janecka et al., 2015).

In this research, we seek to identify an effective method for DNA extraction and isolation of pathogen DNA from FFPE tissue samples by comparing two different extraction methods: the manual hot alkaline lysis phenol-chloroform method (HA) (Campos & Gilbert, 2014, Hahn et al., 2021) and the guanidine-based QIAamp FFPE DNA Kit (Qiagen). These methods were selected based on their successful extraction of host DNA from FFPE tissues from the literature. Several other DNA extraction methods were considered such as the Covaris FFPE NA Ultra Kit and a modified procedure for ancient DNA extractions. However, the selected materials represent two of the higher performing methods that are also time and cost effective. We compared these methods by examining commonly quantified metrics such as DNA yield, extract quality, sequence quality, and overall extraction cost per sample across different leprosy tissue types, from tuberculoid to lepromatous. Additionally, the data obtained from this comparison aid in the effort to improve the phylogenetic representation of *M. leprae* in the Pacific. A total of 40 paired FFPE tissue subsamples, collected from 20 individuals from Hawaii (n=6), Guam (n=6), Palau (n=6), and Pohnpei (n=2) treated from 1995 to 2011, were extracted. While the QIAamp kit showed merit in the literature (McDonough et al., 2019, Sarnecka et al., 2019), the HA protocol is expected to outperform the commercial test kit, due to its

specificity towards low quantity DNA and its inherent protections against common errors in FFPE due to cross linkage and DNA degradation (Gilbert et al., 2007).

METHODS

(1) Sampling:

20 FFPE tissue samples embedded in 1x1x0.5cm paraffin blocks were obtained from the Hawaiian Pathology Laboratory and selected for paired extraction method analysis. These 20 samples came from individuals who were diagnosed with leprosy from Palau (n=6), Hawaii (n=6), Guam (n=6), and Pohnpei (n=2) with pathological presentation of LL (n=7), BL (n=2), BB (N=3), and BT (n=8). Samples were manually excised from the paraffin blocks using metal dissection picks and single use scalpels. The sample was then separated into two sub-samples with an equal mass (+/- 0.1 mg) such that the same input tissue mass was used for each extraction method for each paired sample. The input tissue mass between individuals ranged from 0.5 mg to 5.2 mg. As FFPE samples are highly susceptible to contamination due to their fragmented nature, all sub-sampling, extractions, and preliminary quality tests were conducted in an ancient DNA Class 10000 cleanroom with regular UV and bleach sterilization of all materials and surfaces.

(2) DNA Extraction

Qiagen QIAamp DNA FFPE Kit:

The manufacturer recommended protocols for extracting DNA from FFPE tissue using the QIAamp kit (Qiagen; QIAamp Handbook, 2020) were followed with the following modifications: after residual paraffin was removed by centrifugation with xylene and the samples were centrifuged with 80% ethanol to produce a pellet, the samples underwent an extended incubation of 56C for three hours with proteinase K to degrade the proteins left in solution. At the three-hour mark when samples were still not lysed, an additional 180uL of Buffer ATL (Qiagen) was added and samples were incubated overnight, for a total of 20 hours at 56°C. A final heat step of 90°C for 1hr was applied and kit protocols were followed as instructed with a final elution volume of 60µL.

Hot Alkaline Lysis Manual Procedure:

For the second extraction method, the protocol developed by Campos and Gilbert (2014) was implemented as per the recommendations of Hahn et al. 2021, with the following modifications. The original protocol has several variable steps, depending on preference and sampling. Although the HA Lysis procedure was optimized for 3-10µm tissue thickness, samples here were manually excised rather than excised via microtome. To increase surface area, each subsample was diced with a scalpel to reduce tissue size. Following the autoclave lysis, the phenol: chloroform: isopropanol extractions were conducted following protocols outlined in Campos and Gilbert (2014). An optional carrier solution was listed for pellet visualization, but none was used in this project. After rinsing and decanting, the pellets were resuspended with 60µL of TE buffer.

(3) Extract Quantification and *M. leprae* DNA Identification

To assess the quantity of DNA in the extracts from both extraction procedures, a Qubit version 3.0 Fluorometer (Invitrogen) double stranded, high sensitivity protocol (Qubit User Guide, 2015) was followed. All 40 subsamples moved on for further stages. Two quantitative Polymerase Chain Reaction (qPCR) assays were used to identify and quantify *M. leprae* DNA using TaqMan *M. leprae* specific probes of RLEP and 85B (Truman et al., 2008; Martinez et al.,2006). Extracts, 1:10 dilutions of extracts, and blanks were run in triplicate for each assay and positive results were identified when two out of the three replicates amplified.

(4) Library Preparation, Indexing, Capture, and Sequencing

Based on the Qubit and qPCR results, four samples were removed from downstream analysis. These samples were removed specifically because of low DNA quantities across both extractions as well as replicate failure in the *M. leprae* DNA screening process. The remaining 32 subsamples and 2 blanks from each extraction proceeded to library preparation. To ensure that the samples were appropriately sized for the library procedure, all samples were sheared using a Covaris Ultrasonicator with a target fragment size of 200bp. Double stranded libraries were built following Meyer & Kircher (2010) with an added partial uracil-DNA glycosylase (UDG) treatment step. While most utilized for ancient DNA sequencing for the reduction of nucleotide misincorporation, the UDG treatment was utilized due to its ability to reduce non-reproducible sequence artifacts like those often seen in FFPE tissues (Berra et al.,2019; Rohland et al.,2015). Next, an Amplitaq Gold indexing reaction was run with a total of

20 cycles, after which the samples were purified through bead cleanup. The purified samples were analyzed via Qubit to identify samples with the highest level of success.

The indexed libraries were then enriched for *M. leprae* DNA using a targeted, capture-based enrichment following the myBaits Hybridization Capture for Targeted NGS v5.00 protocol for high sensitivity. An additional tapestation run was used to assess the quality of the capture products. Of the 40 original subsamples, all paired samples that had at least one extraction method successful produce an amplified capture product moved on in the project (n=20). These remaining samples were sequenced on a MiSeq v2 2x150. Due to poor initial sequencing results, paired capture products that both had low yields were re-amplified, then all samples were sequenced again using an Illumina HiSeq 2x150.

(5) Data Analysis

To compare the extract quality and quantity, data was first normalized by gram of subsample starting mass, such that the concentration of DNA was reported in ng/μL/g. Although it was determined via Levene test that variance in the dataset for extract DNA concentration was approximately equally and well distributed, Shapiro testing indicated the data exhibited non-normal distribution within this dataset. While this distribution could be attributed to chance due to the small sample size, log normalization was used to bring this data into normal distribution for the use of a two-factor ANOVA. This test used the Qubit DNA extraction data to assess significance of the DNA concentration, extraction method, and leprosy type. To examine the results of the qPCR assays, the ANOVA test could not be applied due to the presence of zeros from samples which failed

to amplify and non-numerical variables for pass/fail comparisons with method type and bacillary load. Instead, a series of Kruskal-Wallis tests for identical variables were used.

Sequence Data

To assess the quality of the sequenced samples, the EAGER pipeline (Peltzer et al., 2016) will be run on the Arizona State University's High Performance Computing Cluster, which trims adapters, merges paired ends reads, and maps the sequences to a reference. The pipeline then reports information regarding the sequence and mapping quality. To call the SNP variants and examine and functional effects, the MultiVCFAnalyzer (Bos et al., 2014) will be used. Finally, a GTR nucleotide substitution will be performed with GAMMA modeling to create a Maximum Likelihood Tree (Rogers, 2001) and the BEAST protocol for Bayesian modeling (Suchard et al., 2018) will be used to estimate the most recent common ancestor.

Cost Analysis

As educational discounts are not available in all facilities, the material costs were collected directly from supplier catalogs for the public pricing. The all-in-one QIAamp kit was purchased as a single comprehensive item while the HA protocol required individually sourced reagents and materials. Cost per sample was calculated based on supplier costs listed on the Sigma Aldrich website and QIAamp costs (accessed March 2023), though prices vary based on supplier and accessibility. Time was also considered in this comparison but may differ between labs as extra time was spent following cleanroom best practices.

RESULTS

Table 1: Summary of DNA extraction concentration and qPCR assays for each paired sample in the study

Samples	Type	Location	Method	Qubit Conc (ng/ul)	Norm Conc (ng/ul/g)	rLep Pass	rLep Ct	85B Pass	85 Ct
HL-62	LLs	KOROR, PALAU	Q	0.358	0.398	3/3	29.54	3/3	35.93
			HA	0.118	0.131	3/3	27.16	3/3	34.46
HL-63	BL	KOROR, PALAU	Q	0	0.000	3/3	37.41	0/3	0.00
			HA	0	0.000	3/3	36.11	0/3	0.00
HL-64	BL	KOROR, PALAU	Q	0.604	0.183	0/3	0.00	0/3	0.00
			HA	0	0.000	0/3	0.00	0/3	0.00
HL-65	LLp	KOROR, PALAU	Q	2.04	0.453	3/3	29.48	3/3	35.85
			HA	7.88	1.751	3/3	21.06	3/3	28.19
HL-69	BT	KOROR, PALAU	Q	0.124	0.310	1/3	38.93	0/3	0.00
			HA	0.104	0.260	1/3	39.34	0/3	0.00
HL-70	LL	KOROR, PALAU	Q	0.106	0.046	3/3	36.90	0/3	0.00
			HA	0.926	0.403	3/3	33.75	0/3	0.00
HL-72	LLp	KOROR, PALAU	Q	0.162	0.324	3/3	34.89	2/3	38.36
			HA	0.222	0.444	3/3	29.32	3/3	36.48
HL-73	LL	KOROR, PALAU	Q	0.362	0.213	3/3	30.43	3/3	35.96
			HA	8	4.706	3/3	21.87	3/3	29.05
HL-74	LLp	KOROR, PALAU	Q	2.5	1.389	3/3	31.69	3/3	37.73
			HA	0.56	0.311	3/3	27.54	3/3	35.35
HL-76	BT	HONOLULU, HAWAII	Q	7.68	5.120	3/3	34.14	2/3	38.81
			HA	3.96	2.640	3/3	29.65	3/3	36.52
HL-77	BL or BT	HONOLULU, HAWAII	Q	10	4.167	2/3	37.66	0/3	0.00
			HA	3.64	1.517	3/3	30.15	3/3	37.76
HL-78	BB	HONOLULU, HAWAII	Q	1.07	0.446	1/3	45.99	0/3	0.00
			HA	0	0.000	0/3	0.00	0/3	0.00
HL-82	BT	KOROR, PALAU	Q	0.496	0.138	2/3	38.40	0/3	0.00
			HA	1.64	0.456	3/3	31.10	3/3	40.93
HL-87	BT	KEALAKEKUA, HAWAII	Q	2.16	0.450	2/3	37.89	0/3	0.00
			HA	4.56	0.950	3/3	33.97	1/3	39.20
HL-89	BB	HAGATNA, GUAM	Q	0.214	0.074	2/3	37.67	0/3	0.00
			HA	0.372	0.128	3/3	34.13	0/3	0.00
HL-92	BT	MILILANI, HAWAII	Q	0.172	0.101	1/3	38.27	0/3	0.00
			HA	0.276	0.162	0/3	19.83	0/3	0.00
HL-94	BT	TAMUNING, GUAM	Q	0.232	0.097	3/3	32.65	1/3	38.45
			HA	0.262	0.109	3/3	24.79	3/3	31.73
HL-96	BT	TAMUNING, GUAM	Q	1.42	1.183	3/3	34.24	1/3	38.19
			HA	3.67	3.058	3/3	31.98	3/3	38.91
HL-97	BT	KOLONIA, POHNPEI (MICRONESIA)	Q	2.54	0.488	1/3	38.95	0/3	0.00
			HA	2.22	0.427	3/3	24.31	3/3	31.06
HL-98	LL	HONOLULU, HAWAII	Q	15.3	13.909	3/3	26.31	3/3	31.80
			HA	5.8	5.273	3/3	33.47	0/3	45.83

(1) DNA Quantification Comparison

Four samples (HA n=3, QIAamp n=1) were too low to acquire a Qubit reading and were therefore removed from analysis. The remaining 36 extract raw concentrations ranged from 0.104 ng/ml to 8.000 ng/ml for the HA protocol and from 0.106 ng/ml to 15.300 ng/ml for the QIAamp protocol. The raw data was normalized to account for the differences in subsample starting mass. Initial data quantification analyses were performed using the normalized data results. As shown in Figure 1, five samples were identified as outliers (HA n=2, QIAamp n=3). While the median concentration is very similar between the two extractions, the HA procedures had a greater interquartile interval and maximum. However, the ANOVA statistics showed no significant differences between the methods and the concentration yield (p=0.3949).

No significant differences between the extraction method and the pathological classification when using both the Ridley-Jopling classification and the broader multi-/pauci-bacillary classification (Table 2). However, there was a significant difference in the concentration of the DNA extract and on the year in which the sample was fixed in formalin (p-value=0.02331) (Figure 2).

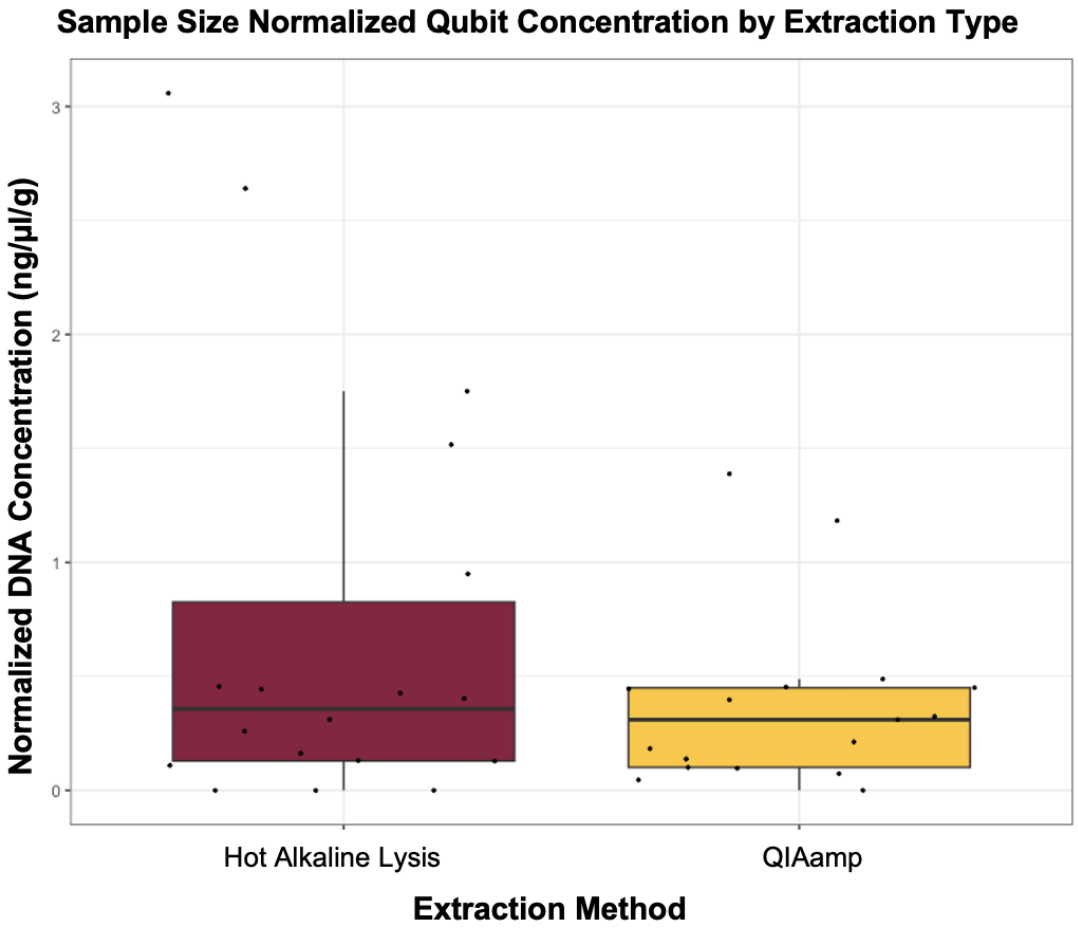

Figure 1: Box plot of post extraction DNA concentrations for both extraction types, hot alkaline lysis (n=20) and QIAamp (n=20), normalized by sample mass to account for differential starting tissue mass.

Table 2: ANOVA table with results comparing the log normalized concentration of DNA extracts to the extraction method and the pathogen type.

2-Factor ANOVA for Log. Normalized Concentration					
	Df	Sum Sq	Mean Sq F	F Value	Pr(>F)
Method	1	0.2955	0.29547	0.72	0.4033
Type	5	0.9008	0.18016	0.439	0.8174
Method:Type	5	0.6153	0.12306	0.2999	0.9088
Residuals	28	11.491	0.41039		

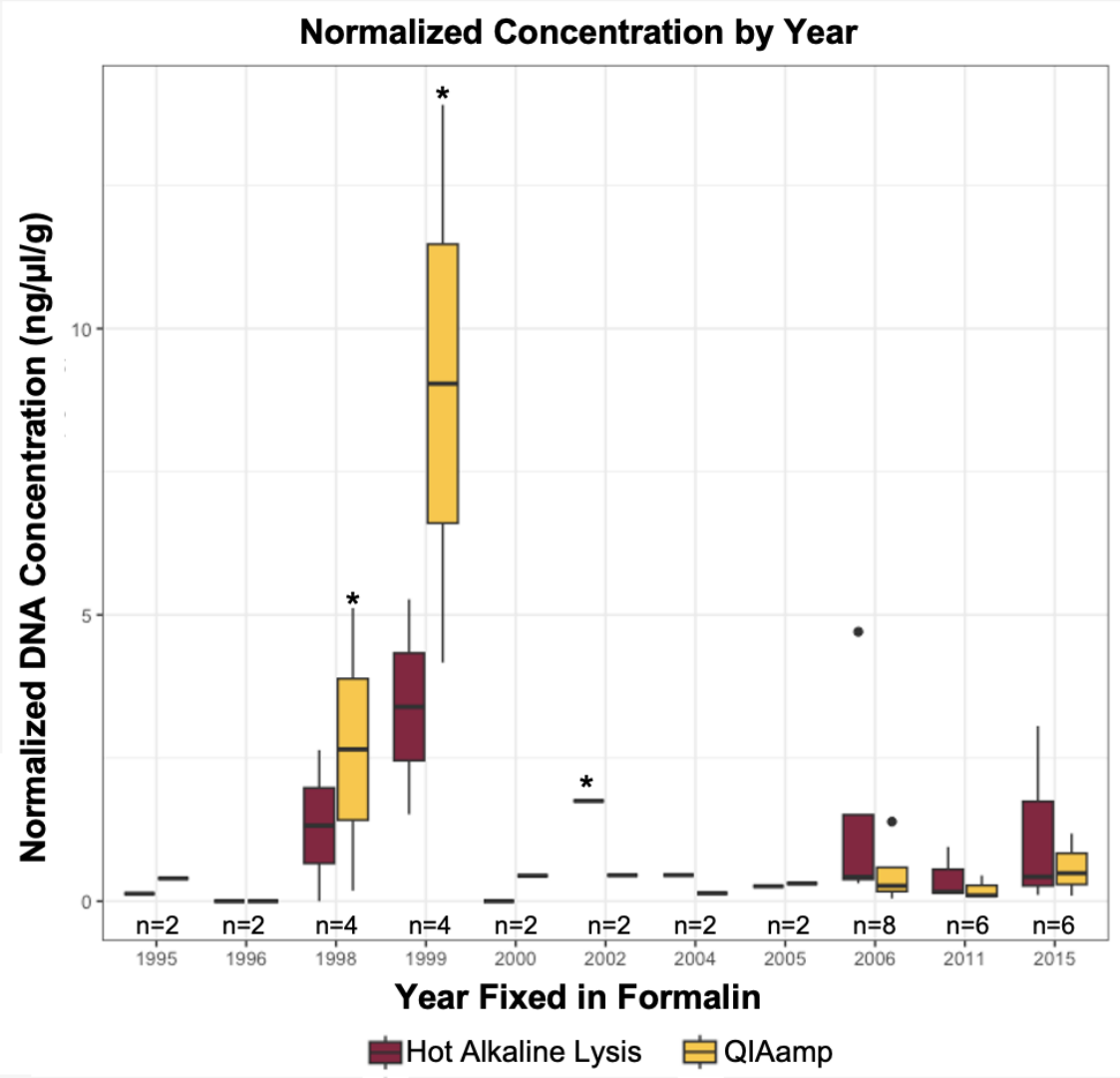

Figure 2: Box plot of DNA concentration normalized by sub-sample starting mass group by extraction method and the year of formalin fixation. Years with a single line depict those with only a single set of paired subsamples in the study.

(2) qPCR Assay

Analysis of the qPCR assay results required firstly a measure of success. Subsamples in which two or three of the triplicate reactions amplified were considered passing, samples in which only one reaction amplified were subject to the success of their

paired sample, and samples with no amplification were considered failing. The RLEP assay consisted of 16 passing HA samples and 15 passing QIAamp, while these numbers declined in the 85B assay, with 11 passing HA and 7 passing QIAamp (Figure 3). Both the RLEP and the 85B assays displayed significance in the number of passing samples, but not due to the extraction method. Rather, significance was shown by the Kruskal-Wallis test to be the result of Ridley-Jopling classification.

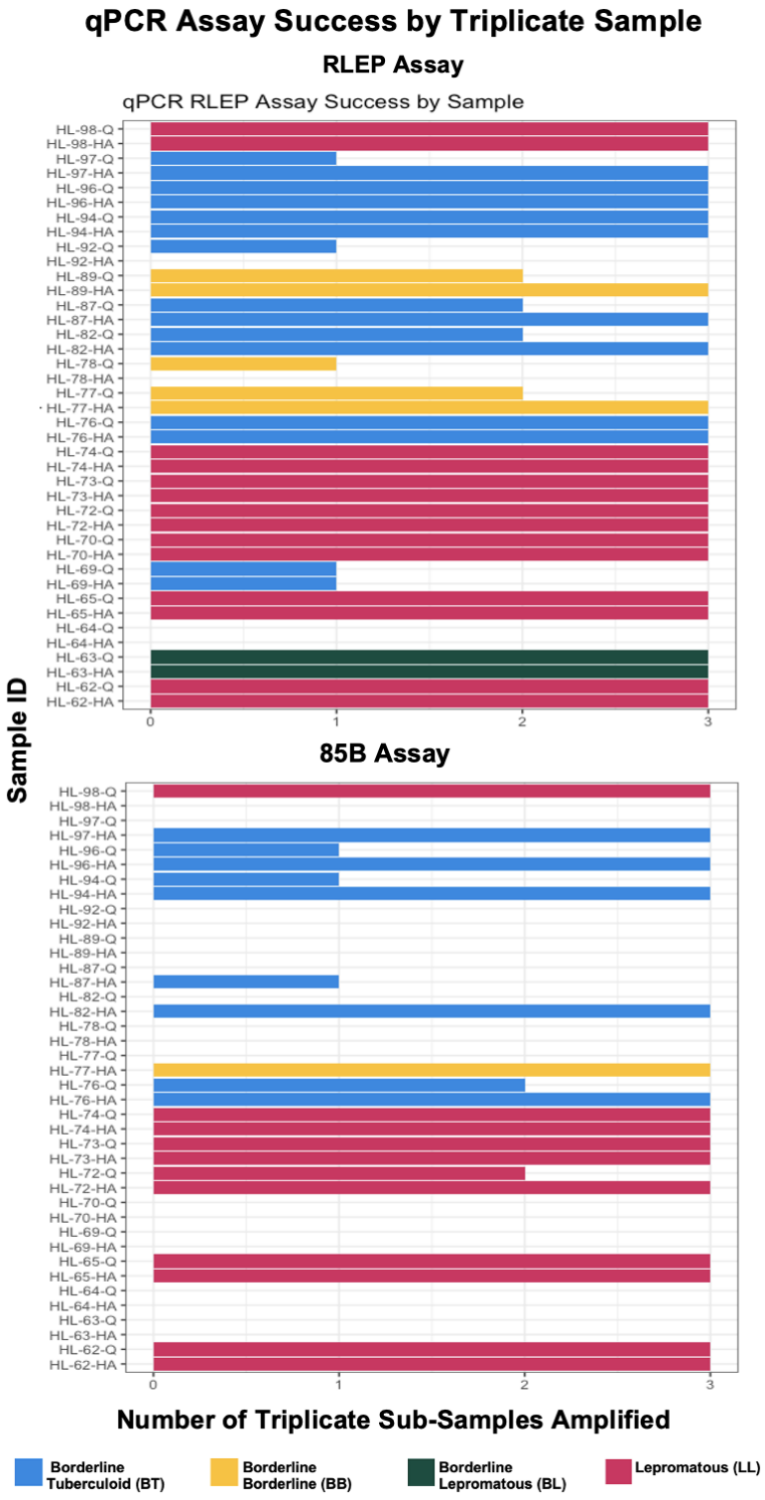

Figure 3: Bar graph on qPCR assay success defined by the number of amplified samples in triplicate with Ridley Jopling pathogen type classification.

Another method of comparison for qPCR, the Ct was examined next, which refers to the cycle at which the amplicons of a subsample pass a system threshold for amplification, with the average being the combination of all three replicates for that subsample. In the RLEP assay, passing Ct averages ranged between 21.06 cycles and 38.39 cycles from both methods. The HA samples amplified on average at 29.42 cycles whereas the QIAamp samples amplified at an average of 35.34 cycles. In the 85B assay, the passing Ct averages were between 28.19 and 40.9 across the methods with the average amplifying cycle 35.80 for HA, and 36.79 for QIAamp. Additional Kruskal-Wallis tests showed that while the method was not significant in the 85B assay, it was significantly different in the RLEP assay (**Figure 4**).

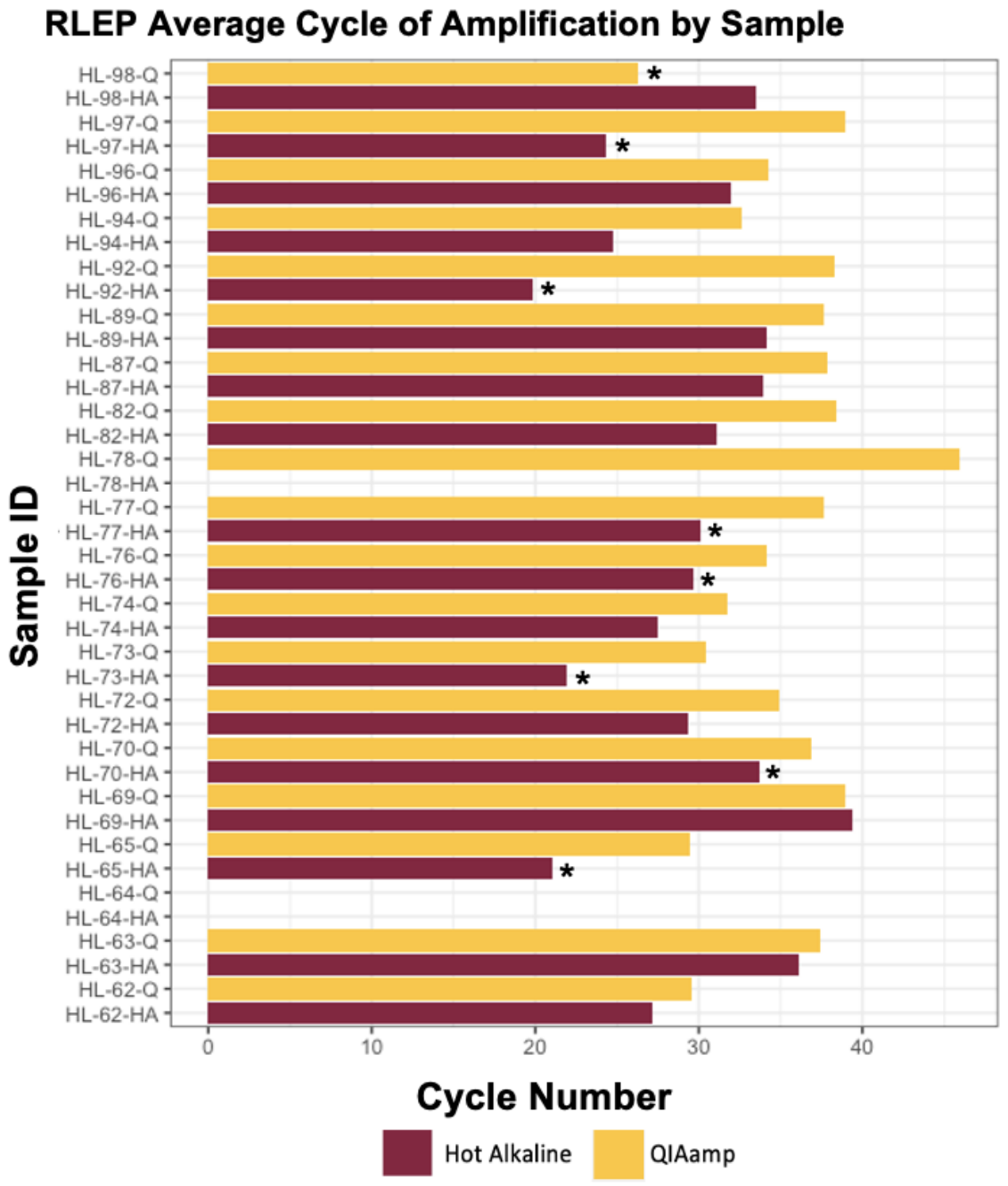

Figure 4: qPCR RLEP Assay amplification cycle of the triplicate results averaged for each sample.

(3) Sequencing

No results are presently available due to delays in returning the HiSeq read data from Fulgent, however further analysis will be performed upon reception of the sequences.

(4) Cost Comparison

As the lowest purchase quantity for the QIAamp kit was 50 samples, this count was used as the baseline for cost comparison, alongside the individual cost per sample. While the QIAamp kit is all in one, a lab utilizing the HA protocol from scratch would require several materials and reagents that can be costly to purchase but are sold in bulk, leaving left-over materials after the 50-sample batch. Thus, the total cost and the batch calculated cost are included in Table 3.

Table 3: Monetary comparison of total batch cost for 50 samples and individual sample cost for each extraction method

	QIAGEN QIAamp Kit	**Hot Alkaline Procedure**
Total Cost	$321	$772.97
		2ml O-ring screw cap tubes *$187/1000 → 50 needed=$9.35*
		1.5mL LoBind Microcentrifuge tubes *$47/250 →100 needed=$18*
		NaOH 1M *$33/L →2.5ml needed= $0.08*
		25:24:1 Phenol: Chloroform: Isoamyl *$284/400ml →25ml needed= $17.75*
		Chloroform *$116/L →25ml needed= $2.90*
		Isopropanol *$34.97/L →2.5ml needed= $0.09*
		Sodium Acetate *$38/L →.25ml needed=$0.01*
		TE Buffer *$33/100ml →2ml needed= $=0.99*
Cost per 50 Samples	($321)	$49.17
Cost per 1 Sample	**$7.20**	**$0.94**
Specific Equipment	Microcentrifuge Thermomixer	Autoclave Microcentrifuge

DISCUSSION

To determine the significance of variation within the DNA extract concentrations, a two-factor ANOVA test of the logarithmically normalized concentration, extraction method, and pathogen type was performed after confirming variation and distribution were normal. The p-values (Table 3) were greater than the alpha (α=.05), so the null hypothesis of no significant differences in the concentration from the extraction method was accepted. Both methods successfully extracted DNA from the problematic FFPE samples in both multibacillary and paucibacillary samples and are recommended for use in future *M. leprae* FFPE studies. However, the scope of our results is limited by the sample size (n=20).

While the difference between the extraction methods and the concentration was not significant, the variable that did suggest significant difference, year of formalin fixation, warranted additional review. A significant increase in concentration of QIAamp extracts was found from the years 1998 (n=4), 1999 (n=4), while 2004 (n=2) had a significantly higher concentration in the HA extracts (Figure 2). The increase in concentration from these two years may be purely random or it may indicate a difference in sample preparation, time before fixation, or other factor not included in the available sample metadata. Every category of pathogen type from the study was represented within this significant 10 samples, so the Ridley-Jopling characterization was not responsible for the difference in concentration from these years. Likewise, these samples originated in both Hawaii and Palau, so they are not specific to an individual location.

The overall results of the qPCR are similar. Both assays were consistently more successful on the Pass/Fail metric with the lepromatous samples, though still strong when used on the tuberculoid samples. Interestingly, the samples which had the least success were the borderline lepromatous and borderline borderline which typically have higher volumes of pathogen DNA present for detection, though this may be related to conditions outside the scope of this project, such as time between collection and fixation, storage condition, or other extraneous variables. The RLEP assay, which primarily pertains to the repetitive element within the *M. leprae* DNA and is known to be the more sensitive to the mycobacterium's presence (Yan et al., 2014), was the only analysis in this project to display significant difference with regards to both the pathology type and the extraction method. Higher Ct averages typically occur when less initial DNA is present, as it takes more amplification rounds for DNA detection across the threshold using fluorescence. Thus, the significantly lower Ct average of the HA protocol suggests it may be a better method for detecting low presence of *M. leprae* DNA, even in the notoriously difficult paucibacillary cases. Further study with increased sample number would be beneficial in confirming this significance.

Due to the failure of the initial MiSeq sequencing run, the sequencing results were significantly delayed and have not been received for analysis. The HiSeq sequences, once returned, will still undergo comparisons based on the findings of the EAGER pipeline and undergo phylogenetic assessments, but this work was not completed prior to this paper. The samples will undergo this stage of analysis once the sequencing is complete.

After the comparative cost analysis, the HA procedure was deemed more cost effective per sample, despite a higher overall cost of high-volume reagents, but some may prefer the ease of acquisition and comprehensive nature of the QIAamp kit. A secondary goal of this research is to increase accessibility of DNA research. As such, the methods that were used present an opportunity for affordable, low-tech extraction. The HA procedure required a higher up-front cost, an approximated total of $772.97 for all required materials, when compared with the all-inclusive $321 of the QIAamp kit for 50 samples. However, that total cost covers far greater volumes of reagents than needed for a 50-sample batch, bringing the batch cost down to $49.17. Simplified, that equates to $0.94/sample for the HA procedure and $7.20/sample for the QIAamp kit. When preparing to extract from large sample numbers, investment in the HA protocol is monetarily favorable, but for small sample size projects, the QIAamp kit simplifies the purchasing of materials at a higher cost.

Regarding methods selection, the two extractions differed from one another on several major levels. The first variation was the paraffin removal and lysis procedure, in which the QIAamp kit utilized a xylene stage followed by proteinase K and buffer ATL while the HA procedure used an alkaline buffer of NaOH and SDS to break down the tissue. Additionally, the method of isolating the DNA varied, with QIAamp indicating the use of additional buffers, ethanol, and MinElute Columns and HA requiring a phenol: chloroform method. While neither method was particularly laborious, the procedures have several key differences regarding ease of use. One consideration when selecting a protocol regards the potential risks. Both methods use chemicals that can be hazardous to

the researcher's health. The QIAamp kit uses guanidine salts in the DNA extraction, which can produce chlorine case when combined with bleach, which is commonly used as a cleaning agent. Caution must be taken to ensure a bleach free cleanup. Likewise, the HA extraction requires a phenol: chloroform: isoamyl mixture. As potential risks of phenol exposure have been presented including potential harm to the livers and kidney after prolonged use (University of Pennsylvania, 2022), there are those who avoid phenol chloroform extractions on principle. However, if safety precautions are taken to placate the risks, the HA lysis procedure presented no outstanding concern. Another consideration between the two methods is the time required for extraction. While the QIAamp kit was advertised to take only 30 minutes after lysis to extract 50 samples, the extended incubation increased the overall extraction time to 26 hours, where the HA Lysis extraction was approximately 4 hours total.

CONCLUSIONS

While neither method outperformed the other, the findings of this study identified two successful means of using archival FFPE samples for the study of pathogen DNA within the preserved host tissue samples. The study of DNA from tissue is inherently destructive, but both the hot alkaline lysis procedure and the QIAGEN QIAamp FFPE DNA kit produced quality extracts and sequences from minimal starting mass, which is beneficial for the preservation of tissue records in cases where limited samples are available. Likewise, both methods successfully extracted DNA from the borderline tuberculoid, paucibacillary samples which have previously presented issues in extraction,

with results suggesting the hot alkaline lysis may higher quality extracts of these samples. Based on the findings of study, we recommend both extraction methods for future use, with the similarity of the results offering future studies the freedom to select based on time, material accessibility, and individual preference.

REFERENCES

World Health Organization. (2023, January 11). *Leprosy*. World Health Organization. Retrieved October 24, 2022, from https://www.who.int/news-room/fact-sheets/detail/leprosy

Centers for Disease Control and Prevention. (2022, March 7). *CDC - Neglected Tropical Diseases - diseases*. Centers for Disease Control and Prevention. Retrieved March 7, 2023, from https://www.cdc.gov/globalhealth/ntd/diseases/index.html

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). https://doi.org/10.1016/s1473-3099(17)30418-8

Pescarini, J. M., Strina, A., Nery, J. S., Skalinski, L. M., Andrade, K. V. F., Penna, M. L. F., Brickley, E. B., Rodrigues, L. C., Barreto, M. L., & Penna, G. O. (2018). Socioeconomic risk markers of leprosy in high-burden countries: A systematic review and meta-analysis. *PLoS neglected tropical diseases*, *12*(7), e0006622. https://doi.org/10.1371/journal.pntd.0006622

Britton, W. J., & Lockwood, D. N. (2004). Leprosy. The Lancet, 363(9416), 1209-1219. https://doi.org/10.1016/s0140-6736(04)15952-7

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

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), 744–747. https://doi.org/10.1126/science.aah3783

WHO (1982) Chemotherapy of leprosy for control programmes. Technical report series 675. World Health Organization, Geneva

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

Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. *Int J Lepr Other Mycobact Dis.* 1966; 34:255–273.

Cole, S. T., Eiglmeier K James KD, Thomson NR, Wheeler PR, Honore N, P. J., Garnier T Harris D, Mungall K, Basham D, Brown D, Chillingworth T, C. C., Connor R

Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd, D. R. M., S Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail, H. T., MA Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, R. M. A., Skelton J Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, S. R., & BG., B. (2001). Massive gene decay in the leprosy bacillus. *Nature*, *409*, 1007–1011. www.nature.com

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: M. leprae in the Pacific. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1812). https://doi.org/10.1098/rstb.2019.0582rstb20190582

Beyene, D., Aseffa, A., Harboe, M., Kidane, D., Macdonald, M., Klatser, P. R., Bjune, G. A., & Smith, W. C. (2003). Nasal carriage of Mycobacterium leprae DNA in healthy individuals in Lega Robi village, Ethiopia. *Epidemiology and infection*, *131*(2), 841–848. https://doi.org/10.1017/s0950268803001079

Schuenemann VJ, Avanzi C, Krause-Kyora B, Seitz A, Herbig A, Inskip S, et al. ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. PLoS Pathog. 2018;14(5): e1006997 https://doi.org/10.1371/journal.ppat.1006997.

Monot, M., Honoré, N., Garnier, T., Araoz, R., Coppée, J. Y., Lacroix, C., Sow, S., Spencer, J. S., Truman, R. W., Williams, D. L., Gelber, R., Virmond, M., Flageul, B., Cho, S. N., Ji, B., Paniz-Mondolfi, A., Convit, J., Young, S., Fine, P. E., Rasolofo, V., … Cole, S. T. (2005). On the origin of leprosy. *Science (New York, N.Y.)*, *308*(5724), 1040–1042. https://doi.org/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., Khamesipour, A., Dowlati, Y., Jianping, S., Rea, T. H., Vera-Cabrera, L., Stefani, M. M., Banu, S., … Cole, S. T. (2009). Comparative genomic and phylogeographic analysis of Mycobacterium leprae. *Nature genetics*, *41*(12), 1282–1289. https://doi.org/10.1038/ng.477

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

EF Gaffney, PH Riegman, WE Grizzle & PH Watson (2018) Factors that drive the increasing use of FFPE tissue in basic and translational cancer research, Biotechnic & Histochemistry, 93:5, 373-386, DOI: 10.1080/10520295.2018.1446101

Bhagwate, A.V., Liu, Y., Winham, S.J. *et al.* Bioinformatics and DNA-extraction strategies to reliably detect genetic variants from FFPE breast tissue samples. *BMC Genomics* **20**, 689 (2019). https://doi.org/10.1186/s12864-019-6056-8

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). https://doi.org/10.1371/journal.pone.0211400

Arreaza, G., Qiu, P., Pang, L., Albright, A., Hong, L. Z., Marton, M. J., & Levitan, D. (2016). Pre-analytical considerations for successful next-generation sequencing (NGS): Challenges and opportunities for formalin-fixed and paraffin-embedded tumor tissue (FFPE) samples. *International Journal of Molecular Sciences, 17*(9). https://doi.org/10.3390/ijms17091579

Gilbert, M. T., Haselkorn, T., Bunce, M., Sanchez, J. J., Lucas, S. B., Jewell, L. D., Marck, E. V., & 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

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. *Współczesna Onkologia*, *23*(1), 52–58. https://doi.org/10.5114/wo.2019.83875

Janecka A., Adamczyk A., Gasińska A., Comparison of eight commercially available kits for DNA extraction from formalin-fixed paraffin-embedded tissues, Analytical Biochemistry, Volume 476, 2015, Pages 8-10, ISSN 0003-2697, https://doi.org/10.1016/j.ab.2015.01.019.

Campos, P. F., & Gilbert, T. M. P. (2014). DNA Extraction from Formalin-Fixed Material. January 2012, 10–15. https://doi.org/10.1007/978-1-61779-516-9

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, 22(6), 2130–2147. https://doi.org/10.1111/1755-0998.13505

QIAamp® DNA FFPE Tissue Handbook . Qiagen. (2020, February). Retrieved August 1, 2021, from https://qiastock.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-ffpe-tissue-kit/

ThermoFisher. (2015). *User guide: QUBIT DSDNA HS Assay Kits - Thermo Fisher Scientific*. Invitrogen. Retrieved August 1, 2021, from https://tools.thermofisher.com/content/sfs/manuals/Qubit_dsDNA_HS_Assay_UG.pdf

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), e328. https://doi.org/10.1371/journal.pntd.0000328

Martinez, A. N., Britto Constança F., Nery José A., Sampaio, E. P., Jardim Márcia 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

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, *2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

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

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 of London. Series B, Biological sciences*, *370*(1660), 20130624. https://doi.org/10.1098/rstb.2013.0624

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome Biol. 2016;17(1):60 https://doi.org/10.1186/s13059-016-0918-z.

Bos, Harkins, Herbig, Coscolla *et al.* 'Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis' *Nature* 514, 494–497 (23 October 2014) https://doi.org/10.1038/nature13591

Rogers J. S. (2001). Maximum likelihood estimation of phylogenetic trees is consistent when substitution rates vary according to the invariable sites plus gamma distribution. *Systematic biology*, *50*(5), 713–722. https://doi.org/10.1080/106351501753328839

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ & Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10 *Virus Evolution* **4**, vey016. https://doi.org/10.1093/ve/vey016

University of Pennsylvania. (2022, January 18). *Fact sheet: Phenol-chloroform extraction*. Fact Sheet: Phenol-Chloroform Extraction | PennEHRS. Retrieved March 3, 2023, from https://ehrs.upenn.edu/health-safety/lab-safety/chemical-hygiene-plan/fact-sheets/fact-sheet-phenol-chloroform-extraction#:~:text=Chemical%20Hazards&text=Systemic%20effects%20from%20prolonged%20or,human%20carcinogen%20and%20reproductive%20hazard.

Yan, W., Xing, Y., Yuan, L. C., De Yang, R., Tan, F. Y., Zhang, Y., & Li, H. Y. (2014). Application of RLEP real-time PCR for detection of M. leprae DNA in paraffin-embedded skin biopsy specimens for diagnosis of paucibacillary leprosy. *The American journal of tropical medicine and hygiene*, *90*(3), 524–529. https://doi.org/10.4269/ajtmh.13-0659