**Non-destructive extraction of DNA from preserved tissues in medical collections**

**Supplementary Information**

Enrique Rayo 1,2\*, Giada Ferrari 1,3\*, Judith Neukamm 1,4, Gülfirde Akgül 1, Abagail M. Breidenstein 1, Martyn Cooke 5, Carina Phillips 5, Abigail S. Bouwman1, Frank J. Rühli1, Verena J. Schuenemann1

1. Institute of Evolutionary Medicine (IEM), University of Zürich, Zürich, Switzerland

2. Plant Ecology Genomics Group, Institute of Integrative Biology (IBZ), ETH Zürich

3. Royal Botanical Garden Edinburgh, Edinburgh, United Kingdom

4. Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany

5. Museums and Archives, The Royal College of Surgeons of England, London, United Kingdom

\* Both authors contributed equally. Correspondence: enrique.rayo@iem.uzh.ch, gferrari@rbge.org.ukted equally. Correspondence: enrique.rayo@iem.uzh.ch, gferrari@rbge.org.uk

# **Supplementary Note 1 - Laboratory workflow**

## **Samples used in this study**

Samples were obtained from the Hunterian Museum collections at the Royal College of Surgeons in London (RCS) (UK) and are summarized in Table SI.1. We collected tissue and fixative samples from ten specimens (fourteen in total with some replicates subsampled) dated between 1760 and 1886 CE. The selected individuals were diagnosed with tuberculosis, leprosy, syphilis, cancer, or skin conditions. All samples are fully encoded and are older than 70 years (post-mortem). Therefore, they do not require additional approval under Swiss law (Swiss Federal Act on Research involving human beings. Human Research Act, HRA art.1 and art.36; RS 810.30) according to the responsible ethics committee (Kantonale Ethikkommission Zürich, Switzerland).

## **Processing of tissue samples**

All DNA extractions and sequencing library preparations were performed in an ancient DNA clean laboratory [[1]](https://paperpile.com/c/vyj1k1/914vV) following standard anti-contamination protocols [[2–4]](https://paperpile.com/c/vyj1k1/MKx4q%2BRAacj%2BkyBWS) with parallel non-template controls, namely an Extraction Blank (EB) containing only reagents from the digestion step onwards, and a Library Blank (LB) containing only reagents from the library preparation step onwards.

The tissue was extracted following a protocol by Devault et al. [[5]](https://paperpile.com/c/vyj1k1/jG3P8) with the modifications described in [[6]](https://paperpile.com/c/vyj1k1/phbbx). In summary, 100 mg of tissue were homogenized using sterile scalpel blades and digested at 50oC for 24h in 25 mM Tris-HCl, 5 mM CaCl2, 25 mM sodium citrate, 2.5 mM ethylenediaminetetraacetic acid, 1% sodium dodecyl sulfate, 50 mM 1,4-dithiothreitol, 10 mM N-phenacylthiazolium bromide and 1.8 mg/mL proteinase K. digestion was performed at 50°C and nutated for 24 h. DNA was extracted with 25:24:1 phenol, chloroform, and isoamyl alcohol and purified using QIAquick spin columns (QIAGEN), with reduced centrifugation speed (6–10 krpm) to prevent the loss of short DNA fragments. Sequencing libraries were generated from 10 μl of extract or extraction blank following Meyer and Kircher [[7,8]](https://paperpile.com/c/vyj1k1/GaN3R%2BRymMU) and sequenced on one lane of a HiSeq2500 (Illumina) with paired-end 125 bp reads and v4 chemistry by the Functional Genomics Center Zürich (Switzerland).

## **Processing of fixative samples**

The fixative extraction was performed following a slight modification of the protocol by Dabney et al. [[9]](https://paperpile.com/c/vyj1k1/IkVAY). Approximately 1-2 mL of fixative was sampled from the embedding liquid of each sample, and transferred without prior digestion to 10 mL of binding buffer (5M GuHCL, 40 % Isopropanol, 400 µl sodium acetate 3 M; UV treated before use). The mixture was then transferred to a high pure extender assembly (Roche Applied Science, UV treated)  modified with a MinElute column (QIAGEN), centrifuged, and washed twice with 700 μl PE buffer before elution in 100 μl TET.

Sequencing libraries were generated following Meyer and Kircher [[7,8]](https://paperpile.com/c/vyj1k1/GaN3R%2BRymMU) with modifications. The re-amplification step was performed with 1 unit Herculase II DNA polymerase (Agilent), 5X Herculase II  reaction mix, 0.3 μM primers IS5 and IS6 [[7]](https://paperpile.com/c/vyj1k1/GaN3R) and 4-7 μl library template with the following thermal profile: initial denaturation at 95◦C for 2 min, 10 to 25 cycles of denaturation at 95°C for 30 sec, annealing at 60°C for 30 sec and elongation at 72°C for 30 sec, followed by a final elongation at 72°C for 5 min. Libraries were purified with MinElute spin columns (QIAGEN) following the manufacturer’s instructions. Quantitative PCR (qPCR) and analysis on an Agilent 2200 TapeStation were used to assess the quality and concentration of the libraries. Libraries were pooled equally and sequenced on one lane of a HiSeq2500 (Illumina) with paired-end 125 bp reads and v4 chemistry by the Functional Genomics Center Zurich (Switzerland).

## **In-solution capture with DNA baits**

After the shotgun screening, the samples were enriched by using a hybridization capture protocol with customs biotinylated baits for human mtDNA [[10]](https://paperpile.com/c/vyj1k1/97BwQ) and the pathogens *Mycobacterium leprae*, *Mycobacterium tuberculosis* [[11–14]](https://paperpile.com/c/vyj1k1/VFpGv%2BqkW82%2BDlSZh%2BjiPYd), and *Treponema pallidum*. For the syphilis capture four genes from the *TPR* gene family (*TprB, TprH, TprL, Tpp15*) (Table 2) were selected based on regions that have shown enough coverage for genome reconstruction of modern and extant *T. pallidum pallidum genomes* [[15]](https://paperpile.com/c/vyj1k1/F5CWo). Long-range PCR products of these genes were produced from modern *T. pallidum pallidum* DNA and were subsequently used as bait for the molecular enrichment through hybridization following the Maricic et al. protocol for long-range PCR bait production (Table SI.2) [[10]](https://paperpile.com/c/vyj1k1/97BwQ). The following primers were designed: TprB-FWD: CGTATATGCGCGTTTTGCGT, TprH-RVS: ACCATGTAGGGCGTTGGTTC, TprH-FWD: GAGTACGCTGTCCCTGTCTT, TprH-RVS: ACCATGTAGGGCGTTGGTTC, TprL-FWD: CAAGGACCTGACCGTTGATT, TprL-RVS: CGCACACAAAGGTGCCAAA, Tpp15-FWD: ATGGTGAAAAGAGGTGGCG, Tpp15-RVS: CTACCTGCTAATAATGGCTTCCT.

PCRs reactions were carried out using the Roche long-range PCR kit in 100 μl reactions (1μl DNA template, 1 unit 10× PCR buffer 2, 0.4 mg/mL BSA, 3μl DMSO, 3125 μM each dNTP, 7U polymerase and 0.3 mM each primer) with the following thermal profile: an initial denaturation at 92°C for 2 min, 10 cycles consisting of a denaturation at 92°C for 10 sec, an annealing at 66°C for 30 sec and a 9-min elongation at 68°C, followed by 30 cycles using the same thermal profile with an additional increase of the elongation time by 20 sec each cycle and a final elongation at 68°C for 7 min. After purification using MinElute spin columns and quantification by NanoDrop, the PCR products were sonicated to 300 bp using a S220 Covaris (Duty cycle 10%, peak incident power 140W, cycles per burst 200, time 120 sec). Products were ligated to biotinylated adapters and immobilized on streptavidin-coated magnetic beads.

Between 8 to 10 individually indexed libraries were pooled at a final concentration of 2 µg of DNA. For the blocking step, a blocking oligos mix composed of 500 µM of each blocking oligo primer Bio4, Bio6, Bio8 and Bio10 was added [[10]](https://paperpile.com/c/vyj1k1/97BwQ). Negative controls (extraction and library blanks) were pooled and enriched separately. After 48 h incubation at 65°C and several washing steps, the library molecules were eluted by NaOH melting and quantified by qPCR (LightCycler 480, Roche) following previous methods [[10]](https://paperpile.com/c/vyj1k1/97BwQ). Prior to sequencing, sample pools and negative controls were combined equimolarly to a final concentration of 10nM. The pools were sequenced using a NextSeq 500 Illumina platform at the Functional Genomic Center of Zurich (Switzerland).

# **Supplementary Note 2 - Analysis**

## **Metagenomic analysis**

There exist several metagenomic analysis tools available for bacterial composition reconstruction. Given its proven efficiency in the analysis of ancient DNA datasets we decided to use the sequence alignment and analysis tool MALT, specifically designed for processing high-throughput sequencing data in the context of metagenomics [[16]](https://paperpile.com/c/vyj1k1/TdAes). MALT uses the NCBI taxonomy to provide unique names and IDs for over 660,000 taxa, including approximately 25,000 prokaryotes. The raw reads were processed with AdapterRemoval2 [[17]](https://paperpile.com/c/vyj1k1/o9PgP), with reads trimmed for both Ns and low quality bases (where quality is lower than 20, both ends). Reads shorter than 30bp were discarded following trimming. Overlapping reads of at least 10 nucleotides were merged. The remaining reads (paired and unpaired) are used as input for MALT. We executed MALT with the following mapping parameters: reads with a minimum 85% identity (−−minPercentIdentity) were considered as a possible match to the reference; only nodes with minimum support of five reads (−−minSupport set to 5); BlastN mode and SemiGlobal alignment were applied and a top percent value of 1 (−−topPercent). The remaining parameters were set to default  [[16]](https://paperpile.com/c/vyj1k1/TdAes).

Results were visualized with MEGAN [[18]](https://paperpile.com/c/vyj1k1/bRKD2). Sample comparisons used normalized counts to control for variation in sequencing depth. Further visualization and figures were performed with the R package phyloseq [[19]](https://paperpile.com/c/vyj1k1/dehYq). Genera that were found in the blanks and that are known to be common residual contaminants in the reagents were not taken into account, mainly Pseudomonas and Burkholderia [[20]](https://paperpile.com/c/vyj1k1/Iny6o); one of the blanks had an abundance higher than 90% of the genus Plantactinospora, a very resistant spore-forming member of the Actinobacteria, but since it was absent in the other blanks we assumed it to be a punctual contamination.

The proportion of taxa present in each fixative sample that can be explained by tissue contribution or laboratory contamination was estimated by using SourceTracker2 [[21]](https://paperpile.com/c/vyj1k1/9VOnj). A BIOM1 file was generated from the MEGAN output, complemented with a mapping file where fixative samples were defined as ‘sink’ material and each tissue sample and laboratory blank as ‘source’. A sink rarefaction depth of 0 and source rarefaction depth of 0 were used to maximize taxa trace detection. The mixing proportions output was then plotted in a stacked bar chart graph using R [[22]](https://paperpile.com/c/vyj1k1/9G8Mj).

## **Endogenous DNA - Methods**

To assess the levels of endogenous DNA in the samples, we used the EAGER pipeline  [[23]](https://paperpile.com/c/vyj1k1/ndC6r) to perform an alignment against both mitochondrial and whole human genome. An important characteristic of ancient sequencing libraries is the occurrence of the

substitution of C by T at the fragment ends [[24]](https://paperpile.com/c/vyj1k1/4uXMt). This is due to the post-mortem decay of the DNA and can be used to authenticate ancient DNA. We selected the in-built function DamageProfiler in EAGER for an estimate of damage in 5’ and 3’ ends for ancient DNA authentication [[25]](https://paperpile.com/c/vyj1k1/DL7s7).  Reads were adapter clipped, merged (suited for pair-end datasets), and quality trimmed using AdapterRemoval2 [[17]](https://paperpile.com/c/vyj1k1/o9PgP). For those read pairs that could not be merged because the overlap region was shorter than 10 nucleotides, or for which the corresponding read was removed during the combined adapter clipping and quality filtering step, the respective single-end reads were first trimmed at the 3’ end such that all bases have a Phred quality score of at least 20 and then mapped individually. The resulting reads for all samples were mapped to the human mitochondrial reference genome hg19 (NC\_001807) treated as double end reads and using the CircularMapper in-built feature the EAGER pipeline, with an error rate (-n) of 0.2 to assure high specificity [[26]](https://paperpile.com/c/vyj1k1/B7xW0).  The enrichment factor for all enriched libraries was calculated by dividing the percentage of endogenous DNA after enrichment by the percentage of endogenous DNA in the shotgun sequencing (Tables SI.7, SI.8, SI.9, SI.10).

## **Pathogen DNA Capture - Methods**

To assess the presence of pathogen DNA in the samples, we used the EAGER pipeline  [[23]](https://paperpile.com/c/vyj1k1/ndC6r) to perform an BWA alignment against the whole genome of *Mycobacterium leprae*, *Mycobacterium tuberculosis*, and *Treponema pallidum.* We selected the in-built function DamageProfiler in EAGER for an estimate of damage in 5’ and 3’ ends for ancient DNA authentication [[25]](https://paperpile.com/c/vyj1k1/DL7s7).  Reads were adapter clipped, merged (suited for pair-end datasets), and quality trimmed using Clip&Merge [[27]](https://paperpile.com/c/vyj1k1/DE3Ey). For those read pairs that could not be merged because the overlap region was shorter than 10 nucleotides or for which the corresponding read was removed during the combined adapter clipping and quality filtering step, the respective single-end reads were first trimmed at the 3’ end such that all bases have a Phred quality score of at least 20 and then mapped individually. Duplicate reads were merged using DeDup. The resulting reads for all samples were mapped treated as double end reads and using the CircularMapper and BWA in-built features of the EAGER pipeline, with an error rate (-n) of 0.2 to assure high specificity [[26]](https://paperpile.com/c/vyj1k1/B7xW0). The enrichment factor for all enriched libraries was calculated by dividing the percentage of endogenous DNA after enrichment by the percentage of endogenous DNA in the shotgun sequencing.

# **Supplementary Note 3 - Results**

## **Endogenous DNA - Results**

For the shotgun mtDNA screening, nine out of fourteen tissue replicates showed detectable traces of mtDNA, eight with a 5X coverage above 98% and one above 24% (Table SI.2). These replicates demonstrate the characteristic misincorporation patterns for ancient DNA (elevated C to T substitutions at 5′-ends and elevated G to A substitutions at 3′-ends) in frequencies ranging from 10- 20% damage [[25]](https://paperpile.com/c/vyj1k1/DL7s7). For the liquid samples, only sample HA4.1 showed sufficient DNA for damage profiling and therefore ancient DNA validation (94.59% 5X coverage of mitogenome, 16% damage). The enrichment strategy [[10]](https://paperpile.com/c/vyj1k1/97BwQ) (SI, Notes 2) was powerful enough to increase the endogenous content of all the tissue samples (Table SI.3), with fold increments from 14 to 275, and also for eight of the ten liquid samples (Table SI.4), not including HA4.1. This sample had a remarkable increase with a final output of more than three million mapped ancient mtDNA reads, a 537-fold increase in the endogenous content (100% coverage), and retaining damage of 10%. Seven tissue samples and liquid sample HA4.1 were assigned to human mtDNA haplogroups, and the percentage of contaminated DNA was estimated (Table SI.5). The haplogroups identified were typically European (Table SI.2) [[28,29]](https://paperpile.com/c/vyj1k1/VWfsh%2BQz1n8).

## **Pathogen DNA Capture - Results**

For the tuberculosis samples, the tissue P1255 was the most successful with 1287 reads that corresponded to 0.25% endogenous content, and presented a damage pattern of 10%, although with a low coverage (0.92%) and a high cluster factor (2.34). The fixative sample for P1255, however, provided a very low number of reads - only 151 with a 2.7% endogenous content, 0.04% coverage, and cluster factor of 6.34, that were not enough to be discerned as ancient. For the leprosy samples, the tissue HA4.1 delivered the highest output with 1051 reads with a damage of 16%, (0.27% endogenous content, 0.93 coverage) but with an extremely high cluster factor of 25.11, indicating an elevated number of repetitive reads. For the fixative material from HA4.1, only 790 reads were detected, with a damage pattern of 22% (0.082% endogenous DNA, 0.3% coverage, and a cluster factor of 10.29), not providing enough material to allow for a finer analysis of the pathogen. An interesting finding was sample tissue P1245 that presented the second-best results for the leprosy enrichment with 917 reads, 20% damage pattern (0.34% endogenous content, 2.4% coverage and a cluster factor of 13), although it was supposedly associated with a syphilis diagnosis (Table SI.1).

# **Supplementary Note 4 - Extended Sample Information**

## **Notes on the History of HA4.1 (RCS Reference number: RCSPC/03240)**

by Martyn Cooke

Specimen HA 4.1 was donated to the RCS in 1886 and consists of the tongue and adjacent tissues of a man suffering from leprosy. The patient history indicates that the man died in the Trinidad Leper Asylum, so we must assume that the specimen was removed at post-mortem and placed in a tissue fixative of some sort of alcohol or ‘spirit’ in Trinidad. Spirit solutions had been in regular usage since the 1600’s and most anatomical or medical museums at that time would have been using a solution of Ethyl Alcohol. However, we can only guess at what was available in The Trinidad Leper Asylum to preserve specimen HA 4.1. It was clearly a suitable substance, as it appears to have minimized autolytic changes and decomposition during the long journey to the UK.

Transfer to Kaiserling III Preservative

At the beginning of the 1900’s, the Kaiserling method was one of many methods to replace Spirit preparations and become a standard method for many pathology museums throughout the world, based around the use of Formaldehyde as a direct replacement to Alcohol. Formaldehyde was first prepared in 1859 and by the time Kaiserling’s new method was introduced in 1896 [[30]](https://paperpile.com/c/vyj1k1/xSjeQ), Formaldehyde was already going into commercial production. Kaiserling made further changes to his method in 1897 [[31]](https://paperpile.com/c/vyj1k1/CUOVs) and again in 1900 [[32]](https://paperpile.com/c/vyj1k1/TGYSf). Formaldehyde fixation does initially cause a loss of colour due to oxidation, but this process can be partially reversed.

Kaiserling’s three stage procedure is as follows:

* **Kaiserling I (KI) solution:** The Fixative. Formaldehyde and buffering salts.
* **Kaiserling II (KII) solution:** Colour Restoration. Short immersion in Alcohol.
* **Kaiserling III (KIII) solution:** The Preservative. Glycerine and buffering salts.

At the same time Kaiserling was perfecting his method, many other scientists were developing their own solutions and modifications along very similar, if not identical lines. It is extremely difficult to determine the exact makeup and chemical ingredients of KIII preservatives that have been used over time. It is also very difficult to determine when specimen HA4.1 was transferred from Alcohol to KIII. Without any records to refer to, it is almost impossible to determine what happened to the specimen. Some records related to the purchase of Glycerine for specimen preparation and the earliest dated back to 1941, so we can say with a reasonable amount of accuracy that the RCS had the means to prepare and use Glycerine in Kaiserling III Preservative as early as the 1940’s.

In 1951 the RCS was actively using and promoting a variation of the Kaiserling technique and L. Proger, the Pathological Curator of the RCS produced a short article (Annals of the Royal College of Surgeons of England Volume 8, No 5, May 1951, pages 388-391). The chemical formula of the preservative (Described as the ‘Mounting Solution’) is as follows:

The Pick-Judah Method (Daukes’ Modification)

* Sodium Acetate 2.7 Kg
* Distilled Water 9 L
* Glycerine 5.4 L
* Camphor 50g dissolved in 200 mL Industrial Methylated Spirit.

This raises the concern that the ‘Kaiserling III’ preservative possibly used for preservation of specimen HA 4.1 may not even have been Kaiserling’s own method, but an inspired variant.

In 1988, the RCS introduced a number of modifications to the Kaiserling fixation/preservation process:

* The original Kaiserling I formaldehyde fixative was replaced by a histological formaldehyde fixative, containing alternate and improved buffering salts.
* The original Kaiserling II colour restoration in an alcohol bath was replaced by the addition of chemical (Reducing) agents into the preservative to restore colour
* The original Kaiserling III preservative was altered to give a higher percentage of Glycerine, which reduced the likelihood of mould growth and eliminated the need for Camphor or Formaldehyde additives.

This means that the original Kaiserling III preservative formula that was used in the 1950’s for specimen HA4.1 was not the same Kaiserling III preservative formula that was used in the 1990’s.

Previous Research Request for Sampling HA 4.1

During the 1990s, a team approached the RCS from St Mary’s Hospital, who were investigating new methods of obtaining DNA from *Mycobacterium,* and requested a sample from specimen HA 4.1. When sampled, it was found that the sample was mounted in pure Liquid Paraffin. Liquid Paraffin is a highly refined mineral oil derived from petroleum. It is also known as Medicinal Paraffin (due to its more historical use as a laxative), Paraffinum Liquidum, and Russian Mineral Oil. Liquid Paraffin is a colorless, odorless, inert substance, which has found favor as a tissue preservative in certain circumstances.

The RCS was possibly unique in using Liquid Paraffin almost exclusively for routine pathological specimens for a period of time, as it minimized discoloration of the preservative caused by pigmentation leaching out of the specimen. Histological examination showed Liquid Paraffin to be a very good preservative. However, the high refractive index had a tendency to cause muscle and skin to darken and become slightly transparent, which obscured many pathological lesions. For this reason, many paraffins preserved specimens were returned back to the Kaiserling III preservative, where the transparency was reversed.

Liquid Paraffin was used from 1978 onwards and started to be phased out for routine use by the 1990s. It is possible that Specimen HA 4.1 was transferred from Kaiserling III preservative to Liquid Paraffin around 1978 and following the first sampling session in the 1990s was returned back to Kaiserling III preservative, where it remained until the current research study sampling took place.

Brief Summary of HA4.1 fixative history

The current RCS Kaiserling III preservative formulae is as follows:

* Distilled Water 15,000mL
* Glycerine 10,000mL
* Sodium Acetate Trihydrate 3,500g

The timeline of fluid changes suggested is as follows:

* 1886 - Fixation, in an unknown Alcohol solution in Trinidad.
* 1886 - Preservation, transfer to 70-80% Ethyl Alcohol at RCS
* 1950’s - Preservation, transfer to Pick, Judah, Daukes preservative (Kaiserling III Modification)
* 1970’s - Preservation, transfer to pure Liquid Paraffin
* 1990’s - Preservation, transfer to Current RCS Kaiserling III preservative

# **Supplementary Tables**

**Table SI.1:** List of tissue and fixative samples utilized in this study. The range of 1760-1793 corresponds to the span when John Hunter (born 1728, died 1793) created his collection.

\* Based on the jar lid sealing, these samples have not undergone recent conservation, i.e. the fixative was not recently replaced.

| **Sample** |  **Diagnosis** | **Tissue** | **Fixative** | **Dating** |
| --- | --- | --- | --- | --- |
| HA4.1 | Leprosy | Lymph node | Kaiserling | 1886 |
| R15.3 | Syphilis | Larynx cartilage | 70%EtOH/silicone | 1877 |
| P906 | Tuberculosis | Femur | 70%EtOH/silicone | 1760-1793 |
| P1189 | Syphilis | Lung | 70%EtOH \* | 1760-1793 |
| P1245 | Syphilis | Larynx | 70%EtOH \* | 1760-1793 |
| P1255 | Tuberculosis | Lung | 70%EtOH \* | 1760-1793 |
| P1499 | Molluscum | Eyelid skin | 70%EtOH \* | 1760-1793 |
| P1505 | Wart | Genital wart | 70%EtOH/silicone | 1760-1793 |
| P1526 | Carcinoma | Lip carcinoma | 70%EtOH \* | 1760-1793 |
| S51.10 | Syphilis | Femur head | 70%EtOH \* | 1889 |

**Table SI.2:** Primer list: Primers used for the bait design of the *Treponema pallidum* hybridization-capture protocol.

| **Primer** | **Sequence** | **Fragment length** |
| --- | --- | --- |
| TprB fwd | CGTATATGCGCGTTTTGCGT | 1979 bp |
| TprB rev | TCACCACAGAACCTTACACGA | 1979 bp |
| TprH fwd | GAGTACGCTGTCCCTGTCTT | 2078bp |
| TprH rev | ACCATGTAGGGCGTTGGTTC | 2078bp |
| TprL fwd | CAAGGACCTGACCGTTGATT | 1541bp |
| TprL rev | CGCACACAAAGGTGCCAAA | 1541bp |
| Tpp15 fwd | ATGGTGAAAAGAGGTGGCG | 425bp |
| Tpp15 rev | CTACCTGCTAATAATGGCTTCCT | 425bp |

\*Fwd: forward; Rev: reverse

**Table SI.3:** Long-range PCR reaction for the generation of baits.

| **LR-PCR** |
| --- |
| **Kit Nr** | **Reagent** | **Concentration** | **Final** | **ul (per reaction)** |
| #2 (on ice) | Buffer 2 green | 5 x | 1x | 20 |
| #5 | DMSO | 100% | 3% | 3 |
| #6 (on ice) | dNTP mix | 10uM | 0,5ulM | 5 |
|  | BSA | 10mg/ml | 0,4mg/ml | 4 |
|  | Primer For | 10uM | 0,3uM | 3 |
|  | Primer Rev | 10uM | 0,3uM | 3 |
| #1 (on ice) | Enzyme mix | 5U/ul | 0,075U | 1.4 |
|  | H20 |  |  | 59.6 |
|  | Template ( Modern syphilis DNA) |  |  | 1 |
|  | Total |  |  | 100 |

**Table SI.4:** Temperature settings for the LR-PCR

| **Temperature profile** |
| --- |
| Temp | Time | Cycles |
| 92°C | 2min |  |
| 92°C | 10sec | 10 cycles |
| 66°C | 30sec |
| 68°C | 9min |
| 92°C | 10sec | 24 cycles |
| 66°C | 30sec |
| 68°C | 9min ( +20 sec/cycle) |
| 68°C | 7min |  |
| 8°C | Hold |  |

**Table SI.5:** Shotgun screening for mtDNA, indicating the cluster factor and the percentage of coverage (1X and 5X).

| **Sample** | **Cluster Factor** | **Coverage >= 1X in %** | **Coverage >= 5X in %** |
| --- | --- | --- | --- |
| HA4.1 (L) | 1.104 | 100.00 | 94.59 |
| HA4.1 (LN) | 1.194 | 31.12 | 0.00 |
| HA4.1 (T) | 1.194 | 99.99 | 98.39 |
| P1189 (L) | 3.00 | 0.59 | 0.00 |
| P1189 (T) | 1.758 | 100 | 99.6 |
| P1245 (L) | 6.167 | 1.79 | 0.00 |
| P1245 (T) | 1.539 | 100 | 100 |
| P1255 (T) | 1.102 | 100 | 99.97 |
| P1499 (L) | NA | NA | NA |
| P1499 (T) | 1.00 | 30.56 | 0.00 |
| P1505 (L) | NA | NA | NA |
| P1505 (T) | 1.257 | 76.52 | 24.87 |
| P1526 (L) | NA | NA | NA |
| P1526 (T) | 2.007 | 100 | 99.99 |
| P906 (L) | NA | NA | NA |
| P906 (B) | 1.00 | 0.87 | 0.00 |
| P906 (T) | 2.525 | 18.99 | 0.00 |
| R15.3 (L) | NA | NA | NA |
| R15.3 (C) | 1.12 | 100 | 100 |
| R15.3 (CS) | 1.233 | 100 | 100 |
| R15.3 (M) | NA | NA | NA |
| S51.10 (L) | 2.00 | 0.91 | 0.00 |
| S51.10 (T) | 1.098 | 100.00 | 99.89 |

L= Liquid; T = Tissue; LN = Lymph Node; B = Bone; C = Cartilage; CS = Cross Section; M= Mucosa

**Table SI.6:** Endogenous content of the mtDNA present in the tissue samples before and after enrichment, and the calculated fold increase.

| **Sample** | **Shotgun Endogenous DNA (%)** | **Enriched Endogenous DNA (%)** | **Fold increase** |
| --- | --- | --- | --- |
| HA.4.1LN | 0.006 | 0.207 | 34.500 |
| HA4.1T | 0.09 | 7.223 | 80.260 |
| P1189T | 0.220 | 41.593 | 198.060 |
| P1245T | 1.251 | 47.769 | 38.190 |
| P1255T | 0.135 | 28.512 | 211.200 |
| P1499T | 0.002 | 0.253 | 126.500 |
| P1505T | 0.021 | 4.884 | 232.570 |
| P1526T | 0.456 | 43.575 | 95.560 |
| P906B | 0.001 | 0.014 | 14.000 |
| P906T | 0.003 | 0.531 | 177 |
| R15.3CS | 0.839 | 25.064 | 29.900 |
| R15.3LN | 0.660 | 22.244 | 33.700 |
| S51.10T | 0.092 | 25.361 | 275.700 |

L= Liquid; T = Tissue; LN = Lymph Node; B = Bone; C = Cartilage; CS = Cross Section; M= Mucosa

**Table SI.7:** Endogenous content of the mtDNA present in the fixative samples before and after enrichment, and the calculated fold increase.

| **Sample** | **Shotgun Endogenous DNA (%)** | **Enriched Endogenous DNA (%)** | **Fold increase** |
| --- | --- | --- | --- |
| HA4.1 | 0.083 | 44.612 | 537.500 |
| P1189 | 0.000 | 1.621 | - |
| P1245 | 0.000 | 2.249 | - |
| P1255 | 0.002 | 18.639 | 9319.500 |
| P1499 | 0.000 | 0.000 | - |
| P1505 | 0.000 | 0.751 | - |
| P1526 | 0.000 | 1.987 | - |
| P906 | 0.000 | 0.732 | - |
| R15.3 | 0.000 | 0.000 | - |
| S51.10 | 0.000 | 1.000 | - |

**Table SI.8:** Mitochondrial DNA haplogroup estimation after the hybridization capture, indicating the quality of the estimate and percentage of the contamination, and percentage of damage in the reads.

| **Sample** | **Haplogroup** | **Quality (%)** | **Cont.Est** | **Damage (%)** |
| --- | --- | --- | --- | --- |
| HA4.1 (T) | H2a2 | 52.20 | 0.01 | 23 |
| HA4.1 (LN) | H2a2 | 51.69 | 0.01 (?) | failed |
| HA4.1 (L) | H | 91.44 | 0.04 (K1) | 10 |
| P906 (T.1) | K1 | 66.67 | 0.99 (?) | 9 |
| P906 (T.2) | failed | failed | failed | failed |
| P906 (L) | failed | failed | failed | failed |
| P1189 (T) | I2 | 94.77 | 0.01 (?) | 17 |
| P1189 (L) | failed | failed | failed | failed |
| P1245 (T) | U5a1a1e | 98.81 | 0.02 (?) | 23 |
| P1245 (L) | failed | failed | failed | failed |
| P1255 (T) | H1bb | 97.54 | 0.00 | failed |
| P1255 (L) | failed | failed | failed | failed |
| P1499 (T) | H1bb | 65.06 | 0.01 (?) | 20 |
| P1499 (L) | failed | failed | failed | failed |
| P1505 (T) | W3a1a2 | 76.55 | 0.01 (?) | 21 |
| P1505 (L) | failed | failed | failed | failed |
| P1526 (T) | H3 | 94.57 | 0.01 (?) | failed |
| P1256 (L) | failed | failed | failed | failed |
| R153 (C) | K1e1 | 94.19 | 0.00  | failed |
| R153 (L) | failed | failed | failed | failed |
| S51.10 (T) | failed | failed | failed | failed |
| S51.10 (L) | failed | failed | failed | failed |

L= Liquid; T = Tissue; LN = Lymph Node; B = Bone; C = Cartilage; (?) = unknown contaminant

**Table SI.9:** Endogenous content of the pathogen DNA present in the tissue samples before and after enrichment, and the calculated fold increase.

| **Sample** | **Shotgun DNA (%)** | **Enriched DNA (%)** | **Fold increase** |
| --- | --- | --- | --- |
| HA4.1LN (leprosy) | 0.003 | 0.008 | 2.660 |
| HA4.1T (leprosy) | 0.011 | 0.273 | 24.810 |
| P1189T (syphilis) | 0.003 | 0.005 | 1.666 |
| P1245T (syphilis) | 0.001 | 0.002 | 2 |
| R15.3CS (syphilis) | 0.000 | 0.000 | - |
| R15.3LN (syphilis) | 0.000 | 0.000 | - |
| S51.10T (syphilis) | 0.000 | 0.000 | - |
| P1255T (TB) | 0.013 | 0.248 | 19.076 |
| P906B (TB) | 0.004 | 0.004 | 2.000 |
| P906T (TB) | 0.006 | 0.017 | 2.830 |

**Table SI.10:** Endogenous content of the pathogen DNA present in the fixative samples before and after enrichment, and the calculated fold increase.

| **Sample** | **Shotgun DNA (%)** | **Enriched DNA (%)** | **Fold increase** |
| --- | --- | --- | --- |
| HA4.1 (leprosy) | 0.013 | 0.082 | 6.300 |
| P1189 (syphilis) | 0.002 | 0.006 | 3.000 |
| P1245 (syphilis) | 0.003 | 0.005 | 1.660 |
| R15.3 (syphilis) | 0.001 | 0.006 | 6.00 |
| S51.10 | 0.003 | 0.005 | 1.660 |
| P1255 (TB) | 0.019 | 0.027 | 1.410 |
| P906 (TB | 0.007 | 0.0352 | 5.020 |

# **Supplementary Figures**

**Figure SI.1:** Richness of bacteria expressed with alpha diversity indices observed (top) and Shannon Index (bottom), both separated based on the type of fixative (EtOH70 in green, Kaiserling in blue) and negative controls (Blank in red).

EB = Extraction Blank; LB = Library Blank; Lon = London; E = Ethanol (liquid) ; LN = Lymph Node; T = Tissue; B = Bone; C = Cartilage; CS = Cross Section; M = Mucosa.

**Figure SI.2:** Main 20 bacterial families present in the samples expressed in their percentage of abundance, separated by blank, liquid or tissue samples. EB = Extraction Blank; LB = Library Blank; Lon = London; E = Ethanol; LN = Lymph Node; T = Tissue; B = Bone; C = Cartilage; CS = Cross Section; M = Mucosa.

**References**

1. [Knapp M, Hofreiter M. Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. *Genes* . 1(2), 227–243 (2010).](http://paperpile.com/b/vyj1k1/914vV)

2. [Cooper A, Poinar HN. Ancient DNA: do it right or not at all. *Science*. 289(5482), 1139–1139 (2000).](http://paperpile.com/b/vyj1k1/MKx4q)

3. [Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I. Assessing ancient DNA studies. *Trends Ecol. Evol.* 20(10), 541–544 (2005).](http://paperpile.com/b/vyj1k1/RAacj)

4. [Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Science & Technology of Archaeological Research*. 3(1), 1–14 (2017).](http://paperpile.com/b/vyj1k1/kyBWS)

5. [Devault AM, Golding GB, Waglechner N, *et al.* Second-pandemic strain of Vibrio cholerae from the Philadelphia cholera outbreak of 1849. *N. Engl. J. Med.* 370(4), 334–340 (2014).](http://paperpile.com/b/vyj1k1/jG3P8)

6. [Ferrari G, Neukamm J, Baalsrud HT, *et al.* Variola virus genome sequenced from an eighteenth-century museum specimen supports the recent origin of smallpox. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375(1812), 20190572 (2020).](http://paperpile.com/b/vyj1k1/phbbx)

7. [Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010(6), db.prot5448 (2010).](http://paperpile.com/b/vyj1k1/GaN3R)

8. [Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40(1), e3 (2012).](http://paperpile.com/b/vyj1k1/RymMU)

9. [Dabney J, Knapp M, Glocke I, *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* 110(39), 15758–15763 (2013).](http://paperpile.com/b/vyj1k1/IkVAY)

10. [Maricic T, Whitten M, Pääbo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*. 5(11), e14004 (2010).](http://paperpile.com/b/vyj1k1/97BwQ)

11. [Schuenemann VJ, Singh P, Mendum TA, *et al.* Genome-wide comparison of medieval and modern Mycobacterium leprae. *Science*. 341(6142), 179–183 (2013).](http://paperpile.com/b/vyj1k1/VFpGv)

12. [Bos KI, Jäger G, Schuenemann VJ, *et al.* Parallel detection of ancient pathogens via array-based DNA capture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370(1660), 20130375 (2015).](http://paperpile.com/b/vyj1k1/qkW82)

13. [Bos KI, Harkins KM, Herbig A, *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*. 514(7523), 494–497 (2014).](http://paperpile.com/b/vyj1k1/DlSZh)

14. [Furtwängler A, Neukamm J, Böhme L, *et al.* Comparison of target enrichment strategies for ancient pathogen DNA. *Biotechniques*. 69(6), 455–459 (2020).](http://paperpile.com/b/vyj1k1/jiPYd)

15. [Weinstock GM, Hardham JM, McLeod MP, Sodergren EJ, Norris SJ. The genome of Treponema pallidum: new light on the agent of syphilis. *FEMS Microbiol. Rev.* 22(4), 323–332 (1998).](http://paperpile.com/b/vyj1k1/F5CWo)

16. [Vågene ÅJ, Herbig A, Campana MG, *et al.* Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol*. 2(3), 520–528 (2018).](http://paperpile.com/b/vyj1k1/TdAes)

17. [Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes*. 9, 88 (2016).](http://paperpile.com/b/vyj1k1/o9PgP)

18. [Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data [Internet]. Genome Research17(3), 377–386 (2007). Available from:](http://paperpile.com/b/vyj1k1/bRKD2) <http://dx.doi.org/10.1101/gr.5969107>[.](http://paperpile.com/b/vyj1k1/bRKD2)

19. [McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 8(4), e61217 (2013).](http://paperpile.com/b/vyj1k1/dehYq)

20. [Salter SJ, Cox MJ, Turek EM, *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12(1), 87 (2014).](http://paperpile.com/b/vyj1k1/Iny6o)

21. [Knights D, Kuczynski J, Charlson ES, *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods*. 8(9), 761–763 (2011).](http://paperpile.com/b/vyj1k1/9VOnj)

22. [Core Team R, Others. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. *Available*. (2013).](http://paperpile.com/b/vyj1k1/9G8Mj)

23. [Peltzer A, Jäger G, Herbig A, *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* 17, 60 (2016).](http://paperpile.com/b/vyj1k1/ndC6r)

24. [Briggs AW, Stenzel U, Johnson PLF, *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* 104(37), 14616–14621 (2007).](http://paperpile.com/b/vyj1k1/4uXMt)

25. [Neukamm J, Peltzer A, Nieselt K. DamageProfiler: Fast damage pattern calculation for ancient DNA [Internet]. bioRxiv, 2020.10.01.322206 (2020). Available from:](http://paperpile.com/b/vyj1k1/DL7s7) <https://www.biorxiv.org/content/10.1101/2020.10.01.322206v1>[.](http://paperpile.com/b/vyj1k1/DL7s7)

26. [Peltzer A. CircularMapper [Internet]. Github. Available from:](http://paperpile.com/b/vyj1k1/B7xW0) <https://github.com/apeltzer/CircularMapper>[.](http://paperpile.com/b/vyj1k1/B7xW0)

27. [Peltzer A. ClipAndMerge [Internet]. Github. Available from:](http://paperpile.com/b/vyj1k1/DE3Ey) <https://github.com/apeltzer/ClipAndMerge>[.](http://paperpile.com/b/vyj1k1/DE3Ey)

28. [Renaud G, Slon V, Duggan AT, Kelso J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16, 224 (2015).](http://paperpile.com/b/vyj1k1/VWfsh)

29. [Weissensteiner H, Pacher D, Kloss-Brandstätter A, *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44(W1), W58–63 (2016).](http://paperpile.com/b/vyj1k1/Qz1n8)

30. [Kaiserling C. Über die conservierung von sammlungspräpaten mit erhaltung der natürlichen farben. *Berliner Klinische Wochenschift*. 33((35)), 775–777 (1896).](http://paperpile.com/b/vyj1k1/xSjeQ)

31. [Kaiserling C. Weitere mittheilungen über die herstellung möglichst naturgefreuer sammlungspräpaten. *Virchows Arch. Pathol. Anat. Physiol. Klin. Med.* 147((3)), 389–417 (1897).](http://paperpile.com/b/vyj1k1/CUOVs)

32. [Kaiserling C. Über die Konservierung und aufstellung pathologisch-anatomischer präparate für schau- und lehrsammlungen. In: *Verhandlungen der Deutschen Pathologischen Boyle.*, Baltimore: John Hopkins University Press. (1900).](http://paperpile.com/b/vyj1k1/TGYSf)