



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2012

Genotype of a historic strain of *Mycobacterium tuberculosis*

Bouwman, Abigail S ; Kennedy, Sandra L ; Müller, Romy ; Stephens, Richard H ; Holst, Malin ; Caffell, Anwen C ; Roberts, Charlotte A ; Brown, Terence A

DOI: <https://doi.org/10.1073/pnas.1209444109>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-98491>

Journal Article

Originally published at:

Bouwman, Abigail S; Kennedy, Sandra L; Müller, Romy; Stephens, Richard H; Holst, Malin; Caffell, Anwen C; Roberts, Charlotte A; Brown, Terence A (2012). Genotype of a historic strain of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, 109(45):18511-18516.

DOI: <https://doi.org/10.1073/pnas.1209444109>

Genotype of a historic strain of *Mycobacterium tuberculosis*

Abigail S. Bouwman^{a,1,2}, Sandra L. Kennedy^{a,2}, Romy Müller^{a,2}, Richard H. Stephens^a, Malin Holst^b, Anwen C. Caffell^c, Charlotte A. Roberts^c, and Terence A. Brown^{a,3}

^aManchester Institute of Biotechnology, Faculty of Life Sciences, University of Manchester, Manchester M1 7DN, United Kingdom; ^bYork Osteoarchaeology Ltd., York YO42 1SR, United Kingdom; and ^cDepartment of Archaeology, Durham University, Durham DH1 3LE, United Kingdom

Edited by Richard G. Klein, Stanford University, Stanford, CA, and approved September 19, 2012 (received for review June 2, 2012)

The use of ancient DNA in paleopathological studies of tuberculosis has largely been restricted to confirmation of disease identifications made by skeletal analysis; few attempts at obtaining genotype data from archaeological samples have been made because of the need to perform different PCRs for each genetic locus being studied in an ancient DNA extract. We used a next generation sequencing approach involving hybridization capture directed at specific polymorphic regions of the *Mycobacterium tuberculosis* genome to identify a detailed genotype for a historic strain of *M. tuberculosis* from an individual buried in the 19th century St. George's Crypt, Leeds, West Yorkshire, England. We obtained 664,500 sequencing by oligonucleotide ligation and detection (SOLiD) reads that mapped to the targeted regions of the *M. tuberculosis* genome; the coverage included 218 of 247 SNPs, 10 of 11 insertion/deletion regions, and the repeat elements IS1081 and IS6110. The accuracy of the SOLiD data was checked by conventional PCRs directed at 11 SNPs and two insertion/deletions. The data placed the historic strain of *M. tuberculosis* in a group that is uncommon today, but it is known to have been present in North America in the early 20th century. Our results show the use of hybridization capture followed by next generation sequencing as a means of obtaining detailed genotypes of ancient varieties of *M. tuberculosis*, potentially enabling meaningful comparisons between strains from different geographic locations and different periods in the past.

biomolecular archaeology | paleopathology

Tuberculosis (TB) has afflicted the human population for at least the last 8,000 y and probably much longer. A disease of poverty, the first clear historical account of TB dates to 2700 B.C. in China (1), and the first archaeological evidence derives from Italy at 5,800 ± 90 B.P. (2). Older indications of TB are controversial (3, 4), but evolutionary analyses of the group of bacteria that causes the disease suggest that the main human pathogen, *Mycobacterium tuberculosis*, has an ancient origin and might have been in existence for as long as 3.0 million y (5–10). TB was widespread during the classical period, being described by Hippocrates, Aristotle, and Galen (11). It increased rapidly in Europe during the 17th century, giving rise to the White Plague (named because of the pallor associated with the disease) (12), which by the 19th century, was causing up to one-quarter of the deaths in London according to the Bills of Mortality (13–15). Its prevalence at this time was almost certainly promoted by the higher population densities associated with urbanization, which provide ideal conditions for transmission of an airborne pathogen (16). Modern vaccination programs and chemotherapies were thought to have brought the disease under control by the late 1980s (17), but its frequency began to rise again in the early 1990s to the extent that the World Health Organization declared TB to be a global emergency in 1993 (18). One-third of the world population has latent TB, and after HIV, it kills more people than any other infectious agent. In 2010, it was estimated that 8.8 million people contracted TB, and there were 1.3 million deaths (19).

M. tuberculosis is now globally distributed, and it displays biogeographical diversity (20–22) caused by genome sequence variations that result in strains that have differing virulence and immunological properties (9). The emergence of new strains and

changes in the geographical distributions of existing ones have been recorded during the modern clinical era (23), and similar events presumably occurred in the past. However, the link between past changes in the population genetics of the bacterium and the evolution of the disease in prehistoric and historic human populations is largely unexplored. The materials for such a study exist in the form of excavated and curated archaeological skeletons displaying lesions associated with TB (11). TB is caused by members of the *M. tuberculosis* complex (MTBC) of bacteria, which along with *M. tuberculosis*, includes the less common human pathogens *M. canettii* and *M. africanum* as well as *M. bovis*, *M. microti*, *M. caprae*, and *M. pinnipedii*; they cause TB in various mammals but only infrequently infect humans today, especially where milk pasteurization and disease control of animals are practiced. In 19th century England, *M. bovis* would have been more of a problem, because pasteurization was not introduced until the early 1900s: it has been suggested that between 1850 and 1950, there were up to 800,000 deaths from bovine TB in Britain (24, 25). TB is contracted by inhalation of droplets containing bacteria (usually *M. tuberculosis*) from an infected person or consumption of infected animal sources, such as food products (usually *M. bovis*). Some 3–5% of people infected with pulmonary TB develop skeletal TB when the bacteria spread through the blood and lymphatic systems from the lungs to the skeleton. The skeletal manifestations of the disease can also occur, possibly at a higher frequency, in individuals with gastrointestinal TB (26). The spine is the most affected part of the skeleton, and other less specific skeletal changes have also been described (27). Lesions resulting from destruction and limited remodeling of the infected areas, although they have a number of alternative diagnoses, can be recognized in archaeological skeletons, providing an indication of the prevalence of TB in the past (11). This paleopathological information can then be supplemented with genetic data if ancient DNA (aDNA) from the infecting bacteria is preserved in the lesions or other parts of the skeleton (28). The veracity of some of this work has been questioned (3, 21, 29), but convincing identifications of *M. tuberculosis* aDNA have been reported for several human skeletons and mummified remains up to 2,500 y in age (30–33).

Until recently, aDNA research, in general, has had limited scope because of the need to perform different PCRs for each locus being studied in an archaeological sample, which meant that the use of aDNA in studies of past TB has largely been limited to confirmation of paleopathological identifications. Typically, PCRs have been directed at loci such as the IS6110 and IS1081 repeat elements thought to be diagnostic for members of the MTBC

Author contributions: C.A.R. and T.A.B. designed research; A.S.B., S.L.K., R.M., M.H., and A.C.C. performed research; R.M., R.H.S., and T.A.B. analyzed data; and C.A.R. and T.A.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹Present address: Zentrum für Evolutionäre Medizin, Universität Zürich, CH-8057 Zürich, Switzerland.

²A.S.B., S.L.K., and R.M. contributed equally to this work.

³To whom correspondence should be addressed. E-mail: terry.brown@manchester.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209444109/-DCSupplemental.

(34, 35). A few projects have additionally targeted SNPs and other sequence features that distinguish *M. tuberculosis* from *M. bovis* and other members of the MTBC and enable modern antibiotic-resistant strains, which might contaminate aDNA extracts, to be recognized (31). Only limited attempts have been made to place an archaeological strain of *M. tuberculosis* into one of the groups recognized in present day populations (the largest number of informative SNPs typed in a single ancient sample is four) (36).

The advent of next generation sequencing (NGS) methods has opened up possibilities in aDNA research, particularly in the study of ancient pathogens. The ability of NGS to generate genome-wide sequence data from small amounts of starting material has already enabled the complete sequence of a historic *Yersinia pestis* strain to be assembled from aDNA from a 14th century A.D. skeleton, providing insights into the origins of the Black Death (37). Reconstruction of complete *M. tuberculosis* genome sequences from archaeological remains is feasible but complicated by the presence in bone extracts of contaminating mycobacteria from the burial environment (3). These ubiquitous, nonpathogenic species have sequence similarity with *M. tuberculosis*, and distinguishing NGS reads derived from these species from genuine *M. tuberculosis* aDNA sequences can be difficult. This problem can be avoided by adopting a more directed NGS approach using hybridization capture (38) directed at specific polymorphic regions of the *M. tuberculosis* genome. The captured aDNA fragments are then sequenced to type the polymorphisms. In this paper, we report the use of this approach to identify the detailed genotype of a historic strain of *M. tuberculosis* from a 19th century adolescent female skeleton who was buried in St. George's Crypt, Leeds, England.

Results

We extracted aDNA from a rib displaying surface bone formation possibly indicative of pulmonary TB (Fig. 1) and prepared a sequencing by oligonucleotide ligation and detection (SOLiD) library from fragments captured by a hybridization enrichment system comprising 551 baits (Dataset S1) targeting 260 regions of the *M. tuberculosis* genome (Table 1 and Table S1). A total of 726,848 sequence reads corresponding to the reference sequence was obtained. Of these reads, 664,500 reads mapped to the targeted regions of the *M. tuberculosis* genome; the coverage included 218 of 247 SNPs, eight of nine regions of difference (RDs), the *M. tuberculosis*-specific deletion TbD1, the mtp40 locus, and the insertion sequences IS1081 and IS6110 (Table 2).

The presence of reads corresponding to the IS1081 and IS6110 sequences (Fig. 2) showed that the St. George's Crypt skeleton contained DNA from the MTBC, because these elements are thought to be absent from other species (34, 35). Within the complex, individual species can be distinguished by identifying whether insertions are present or absent at the various RD loci (5). It was possible to type seven of nine RDs that were targeted, and the



Fig. 1. Ribs from the female adolescent skeleton 4006 from St. George's Crypt, Leeds. Bone formation possibly indicative of pulmonary TB is visible on the surface of the ribs within the area indicated by the boxes.

Table 1. Regions of the *M. tuberculosis* genome targeted by the hybridization enrichment system

Type of locus	Number targeted*	Refs.
SNPs	247 [†]	6, 39–42
RDs	9	5
<i>M. tuberculosis</i> -specific deletion (TbD1)	1	5
mtp40	1	43
Insertion sequences	2	34, 35

*Details are given in Table S1.

[†]SI Materials and Methods has information on choice of SNPs.

profile showed that the species was either *M. tuberculosis* or *M. canettii* (Table 3). The TbD1 deletion characterizes *M. tuberculosis* (5), but this locus could not be typed unambiguously from the sequence data. The mtp40 locus was identified, but although this insertion was initially reported to be present only in *M. tuberculosis* (43), there is now doubt about its specificity (44–46). However, among the SNP data, the presence of a G at *katG*1388 is diagnostic for *M. tuberculosis* (5, 39).

To compare the historic strain of *M. tuberculosis* with extant varieties, consensus sequences were determined for the regions surrounding each of the 218 SNPs that were typed, and these sequences were concatenated and compared with the equivalent regions of 17 completely sequenced MTBC genomes (Fig. 3). Among these modern strains, the closest similarity was with H37Rv, a member of the Europe/Americas clade of modern *M. tuberculosis* (22). The St. Georges sample differed from the published H37Rv sequence at six positions (403978, 906855, 1901491, 2228965, 2786950, and 3721802). Although different stocks of H37Rv show some sequence variability, none of these polymorphisms have previously been reported (47). The SNP data (Table S2) also enabled comparison with two phylogenies of modern MTBC, placing the St. Georges strain in lineage II and one of synonymous sequence types 2, 8, or 10 (Dataset S2) in the phylogeny in the work by Baker et al. (6) and SNP cluster group (SCG) 6b and SNP type (ST) 14 or 40 (Dataset S3) according to the work by Filliol et al. (40).

Thirteen loci were typed independently by the conventional PCR approach to confirm the accuracy of the NGS results. All 13 PCRs were successful with both the initial DNA extract and a second extract, which was prepared at a later date (Table 4). Control PCRs, which were set up with blank extracts or water instead of extract, always failed to give a product of the expected size. Most of the sequences obtained from the cloned PCR products

Table 2. Summary of sequencing results

Type of locus	Total sequence reads*	Unique sequence reads*
SNPs	0–8,235 (mean = 661)	0–45 (mean = 10.4)
RD1	5,856	21
RD4	16,436	175
RD7	570	26
RD8	22,896	202
RD9	1,290	30
RD10	2,573	84
RD12	26,695	308
RD13	1,169	54
RD14	0	0
TbD1	15,034	53
mtp40	1,573	32
IS1081	12	4
IS6110	689	130

*Data for SNPs give the range and mean for sequence reads covering the SNP positions for the 247 loci that were targeted. The individual SNPs that were identified at each of these loci are listed in Table S2. For the other loci, the numbers are the total reads captured by the relevant baits.

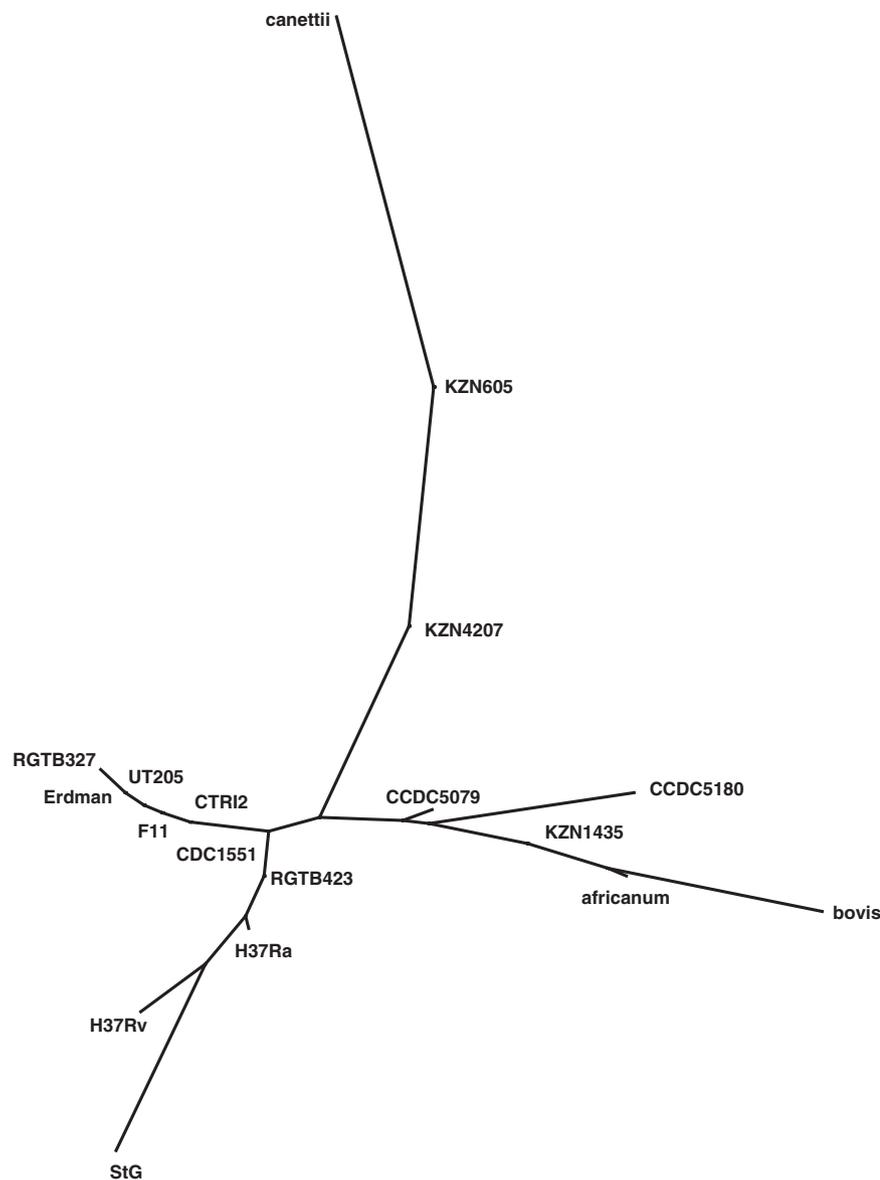


Fig. 3. Neighbor-joining tree comparing the concatenated consensus sequences obtained from the St. George's Crypt skeleton (StG) with the equivalent regions of 17 completely sequenced MTBC genomes. These genomes are *M. tuberculosis* strains H37Rv (National Center for Biotechnology Information reference sequence NC_000962.2), H37Ra (NC_009525.1), ATCC35801 str. Erdman (AP012340.1), CCDC5079 (NC_017523.1), CCDC5180 (NC_017522.1), CDC1551 (NC_002755.2), CTR12 (NC_017524.1), F11 (NC_009565.1), KZN605 (NC_018078.1), KZN1435 (NC_012943.1), KZN4207 (NC_016768.1), RGTB327 (NC_017026.1), RGTB423 (NC_017528.1), and UT205 (NC_016934.1); *M. bovis* bacillus Calmette–Guérin str. Mexico (NC_016804.1); *M. africanum* GM041182 (NC_015758.1); and *M. canettii* CIPT140010059 (NC_015848.1).

carryover of amplicons from previous PCRs in the form of aerosols generated when microfuge tubes are opened. With studies of *M. tuberculosis* aDNA, this problem is particularly pernicious, because it is difficult to detect; many of the routine PCRs, such as those PCRs testing for IS1081, IS6110, and the TbD1 deletion, give products with no sequence variability, which makes it impossible to be certain that a product derives from the sample under study and is not a false positive arising from a contaminating amplicon. To avoid this problem, we used precautions equivalent to those precautions required for human aDNA research (*SI Materials and Methods*). To remove external contamination from the rib sample, we scraped away the outer surface of the bone and then irradiated the resulting core, a method that we have previously shown to be effective in removing DNA contamination placed on the original bone surface by handling (52). This procedure, as well as all subsequent manipulations up to and including posthybridization amplification, was carried out in specially designed laboratories for aDNA analysis by personnel wearing full forensic clothing. Analysis of the sequence reads showed that these precautions were successful, because carryover of amplicons would be detectable by extensive clonality of sequence reads beginning at the two ends of the amplicon sequence. This result was not the case with the sequence reads for those loci, such as the IS1081

and IS6110 elements, that we have previously used as targets for conventional PCR. Instead, these reads, particularly the large number obtained for IS6110, were tiled along the region covered by the baits (Fig. 2), indicating that they were derived from genomic DNA rather than PCR amplicons.

A second contamination issue that is relevant in studies of *M. tuberculosis* aDNA is the possible presence of DNA from environmental mycobacteria that inhabited the burial environment and possibly infiltrated the bones of the skeleton. The genus *Mycobacterium* includes over 100 species, most of which are found in soil and water. The expectation is, therefore, that any archaeological material that has come into contact with soil and/or water is likely to contain environmental mycobacteria (3, 53) and that bone extracts will contain DNA from these species. Genome data are unavailable for the large majority of these species, and the extent to which they share sequence similarity with *M. tuberculosis* is unknown. The possibility exists, therefore, that sequence reads derived from environmental species might be mistaken for ones representing *M. tuberculosis*. When individual SNPs with known variability in the MTBC are being typed, the presence of sequence reads from non-MTBC mycobacteria is unlikely to cause confusion. In our case, the specificity of the enrichment system, combined with the discrimination of the process by which reads were

Table 4. Results of genotyping by conventional PCR

Locus*	Genotype
<i>gyrA284</i>	G
<i>katG1388</i>	G
<i>leuB</i> (3352929)	C
<i>oxyR37</i>	C
<i>oxyR285</i>	G
<i>qcrB</i> (2460626)	C
<i>recN</i> (1920118)	G
<i>rpoB2646</i>	T
<i>rpoB3243</i>	T
Rv0083 (92197)	T
Rv2802c (311473)	C
TbD1	Deletion was detected
<i>pks15/1</i>	7-bp deletion was detected

*Numbers in parentheses indicate the nucleotide positions of those SNPs studied in the work by Filliol et al. (40).

mapped on to the reference sequence, meant that all of the reads that were used to type polymorphisms could be clearly ascribed to MTBC. In fact, because the sample was taken from a crypt burial rather than soil inhumation, the opportunity for contamination with environmental mycobacteria was perhaps less substantial. The presence of contaminating reads would be a greater problem if attempts were made to reconstruct an entire *M. tuberculosis* genome from aDNA. It would be impossible to be certain of the authenticity of any SNP that is not known in modern MTBC, because such an SNP could derive from reads obtained from the homologous region of the genome of a contaminating non-MTBC species with a sequence, in that region, that is identical to the sequence of the MTBC members, except for the SNP position. Obtaining high coverage of the SNP would not solve the problem if the relevant region is absent or at low copy number in the aDNA fraction compared with a relatively high copy number of the contaminant.

We used the SNP data from the sample from the St. George's Crypt skeleton to classify this historic strain of *M. tuberculosis* in accordance with two schemes described for modern varieties of the MTBC (6, 40). This analysis placed the 19th century strain in lineage II and one of synonymous sequence types 2, 8, or 10 according to the work by Baker et al. (6) and ST14 or ST40 and SCG6b according to the work by Filliol et al. (40). SCG6b is uncommon among present day strains of *M. tuberculosis*. Only 6 of 219 strains examined in the work by Filliol et al. (40) were members of SCG6b, and in a later study of 428 antibiotic-resistant strains from Australia, Colombia, India, Mexico, Spain, and the United States, only 8 members of SCG6b were identified, with low prevalence in each of the six countries (54). SCG6b does, however, include the modern strain H37Rv, with a genotype that differs from the historic *M. tuberculosis* at six of the SNP positions that we typed. H37Rv is derived from an isolate, H37, collected from a TB patient in 1905 by Edward R. Baldwin, who at that time, was Director of the Saranac Lake Laboratory in upstate New York, United States. Although circumstantial, this finding suggests that strains belonging to SCG6b were not uncommon in North America in the early 20th century, and our discovery of a strain related to H37 in 19th century Yorkshire might not be unusual.

Materials and Methods

Source Material. The source material was a human rib from St. George's Church Crypt, a late 19th century burial site in Leeds, West Yorkshire, England. The site was excavated in 2009 by MAP Archaeological Consultancy Ltd. The human remains, buried extended and supine in coffins, were recovered from vaults adjacent to the south side of St. George's Church that had been in use between 1840 and 1911 (55) (*SI Material and Methods*). Ten articulated skeletons were recovered, and disarticulated bones were found in three contexts, representing a minimum of 30 individuals (22 adults and 8 nonadults). There was no evidence of definitive tuberculous lesions among the skeletal remains,

but four individuals displayed new bone formation on the visceral surfaces of their ribs (two adolescents and two adults), possibly caused by pulmonary TB, although the diagnoses are not pathognomonic (56). The two adolescents were small gracile skeletons of almost identical size reassembled from Vault 16. Both skeletons were female of similar age (16–18 y), and the patterns of dental crowding and caries were almost identical, suggesting that they may have been siblings, possibly even twins, who died at about the same time (55). A sample was taken from one of the ribs of skeleton 4006 (Fig. 1) under clean conditions by personnel wearing sterile gloves, face masks, hairnets, and protective outer clothing and on work surfaces treated with DNA-Away (Molecular Bioproducts). The sample was then placed in a DNA-free plastic bag and transferred to the aDNA facility at the University of Manchester, which comprises a suite of independent, physically isolated laboratories; each laboratory has an ultrafiltered air supply maintaining positive displacement pressure and a managed access system. Within these laboratories, DNA extracts and libraries were prepared in a class II biological safety cabinet, and PCRs were set up in a laminar flow hood.

Genotyping by SOLiD Sequencing. To remove external contamination from the bone sample, the outer 1–2 mm were scraped away with a sterile scalpel, and the remaining bone was UV-irradiated (254 nm, 120,000 $\mu\text{J cm}^{-2}$) for 2 \times 5 min, with 180° rotation between the two exposures (52). The sample was then sealed in a DNA-free plastic bag and crushed into powder, and 0.25 g were used for DNA extraction as described previously (57). Hybridization capture was carried out using the SureSelect Target Enrichment System for AB SOLiD Multiplexed Sequencing (Agilent Technologies) in accordance with the manufacturer's instructions with the following modifications. Because aDNA fragments are usually <200 bp in length (28), end repair was carried out with 30 μL unsharpened DNA extract, and after ligation of the P1 and A adaptors, two size fractions, of ~150 and 200 bp, were selected and pooled. The nick translation PCR was carried out for 20 rather than 12 cycles, and after the subsequent purification, the DNA was eluted in 25 rather than 50 μL Buffer E1. To maximize the amount of DNA in the hybridization stage, the RNase Block was diluted by adding directly to the DNA library, and hybridization to the capture library was carried out at 65 °C for 72 rather than 24 h. The capture library comprised 551 baits (*Dataset S1*) targeting 260 regions of the *M. tuberculosis* genome (Table 1 and *Table S1*). After posthybridization amplification, a paired-end read fragment library for SOLiD sequencing was prepared in accordance with the manufacturer's instructions (Applied Biosystems), and templated beads were sequenced using the SOLiD 5500 system.

A reference sequence was constructed from *M. tuberculosis* accession NC_000962.2 (58) supplemented with insertions reported in the *M. bovis* genome. Sequence reads were converted into FASTQ format and mapped onto the reference using the default settings of BWA 0.5.9-r15 (59) on the web-based platform Galaxy (60). The resulting datasets were converted to interval files and downloaded along with FASTA versions of the sequence reads, and summary data, including the identities of individual polymorphisms, were generated for each alignment. Consensus sequences for each alignment were concatenated in order based on their genomic coordinates and aligned with the equivalent regions of sequenced MTBC genomes using MegaBLAST (61). A neighbor-joining tree was produced by BLAST Tree View, exported in Newick tree format, and visualized with Dendroscope 3.2.2 (62). Sequence reads are deposited in the European Nucleotide Archive (project accession ERP001877).

Genotyping by PCR. A subset of loci was typed independently by the conventional PCR approach to confirm the accuracy of the capture array results. One set of PCRs was carried out with extracts prepared as described above (57), and a second duplicate set was carried out with extracts prepared using a silica suspension procedure (63, 64). For the latter procedure, 0.25 g bone powder were resuspended in 5 mL extraction buffer (450 mM Na₂EDTA, pH 8.0, 0.25 mg mL⁻¹ proteinase K) and agitated for 24 h in the dark. After centrifugation for 2 min at 5,000 \times g, the supernatant was transferred to a new tube, 2.5 mL binding buffer (5 M guanidinium thiocyanate, 300 mM sodium acetate, pH 5.2) and 100 μL silica suspension (64) were added, and the mixture agitated for an additional 3 h in the dark. The silica pellet was collected by centrifugation for 2 min at 5,000 \times g and resuspended in another 1 mL binding buffer. The silica was then washed, and the DNA was eluted as described in the work by Rohland and Hofreiter (64). PCRs were directed at 11 SNPs at the *qcrB*, *gyrA*, *katG*, *leuB*, *oxyR*, *recN*, *rpoB*, Rv0083, and Rv2802c loci as well as the TbD1 deletion and a deletion in the *pks15/1* gene. PCRs were set up in a final volume of 50 μL , comprising 3 μL extract, 1 \times GeneAmp PCR Gold buffer (Applied Biosystems), 2.0 mM MgCl₂, 200 μM each dNTP, 400 μM each primer, and 1.25 units AmpliTaq Gold DNA polymerase (Applied Biosystems). Cycling conditions were 7 min at 95 °C followed by 45 cycles of 1 min at x °C, 1 min at 72 °C, and 1 min at 94 °C followed by 2 min at x °C and 10 min at

72 °C, where x °C is the primer-specific annealing temperature (Table S3). PCR products were examined by electrophoresis in 2% agarose gels, and bands were purified using Qiaquick columns (Qiagen). DNA was cloned using the CloneJet PCR cloning kit (Fermentas) in *Escherichia coli* XL1-Blue cells (Agilent). Recombinant plasmid DNA was purified using Qiaquick columns and sequenced (GATC Biotech). Sequences were then aligned and polymorphisms were identified using Geneious Basic (<http://www.geneious.com/>).

1. Morse D (1967) Tuberculosis. *Diseases in Antiquity*, eds Brothwell D, Sandison AT (Charles Thomas, Springfield, IL), pp 249–271.
2. Canci A, Minozzi S, Tarli S (1996) New evidence of tuberculous spondylitis from Neolithic Liguria (Italy). *Int J Osteoarchaeol* 6(5):497–501.
3. Wilbur AK, et al. (2009) Deficiencies and challenges in the study of ancient tuberculosis. *J Archaeol Sci* 36(9):1990–1997.
4. Roberts CA, Pfister L-A, Mays S (2009) Letter to the editor: Was tuberculosis present in *Homo erectus* in Turkey? *Am J Phys Anthropol* 139(3):442–444.
5. Brosch R, et al. (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA* 99(6):3684–3689.
6. Baker L, Brown T, Maiden MC, Drobniewski F (2004) Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg Infect Dis* 10(9):1568–1577.
7. Mostowy S, Behr MA (2005) The origin and evolution of *Mycobacterium tuberculosis*. *Clin Chest Med* 26(2):207–216.
8. Gagneux S, et al. (2006) Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 103(8):2869–2873.
9. Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 7(5):328–337.
10. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV (2009) Myths and misconceptions: The origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 7(7):537–544.
11. Roberts CA, Buikstra JE (2008) *The Bioarchaeology of Tuberculosis: A Global View on a Reemerging Disease* (University Press of Florida, Gainesville, FL).
12. Ott K (1996) *Fevered Lives: Tuberculosis in American Culture Since 1870* (Harvard Univ Press, Cambridge MA).
13. Dubos RJ, Dubos J (1952) *The White Plague: Tuberculosis, Man, and Society* (Rutgers Univ Press, Piscataway, NJ).
14. Clarkson L (1975) *Death, Disease and Famine in Preindustrial England* (Gill and Macmillan, Dublin).
15. Roberts CA, Cox M (2003) *Health and Disease in Britain: From Prehistory to the Present Day* (Sutton Publishing, Stroud, Gloucestershire, UK).
16. Lönnroth K, Jaramillo E, Williams BG, Dye C, Raviglione M (2009) Drivers of tuberculosis epidemics: The role of risk factors and social determinants. *Soc Sci Med* 68(12):2240–2246.
17. Smith ER (1988) *The Retreat of Tuberculosis 1850–1950* (Croom Helm, London).
18. Anon (1993) WHO declares tuberculosis a global emergency. *Soz Praventimed* 38(4):251–252.
19. World Health Organization (2011) *Global Tuberculosis Control. WHO Report 2011* (WHO Press, Geneva).
20. Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* 4(11):e7815.
21. Achtman M (2008) Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 62:53–70.
22. Hershberg R, et al. (2008) High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 6(12):e311.
23. Tanaka MM, Francis AR (2006) Detecting emerging strains of tuberculosis by using spoligotypes. *Proc Natl Acad Sci USA* 103(41):15266–15271.
24. Atkins P (2000) Milk consumption and tuberculosis in Britain 1850–1950. *Order and Disorder: The Health Implications of Eating and Drinking in the 19th and 20th Centuries*, ed Fenton A (Tuckwell, Edinburgh), pp 83–95.
25. Pfeiffer DU (2008) Animal tuberculosis. *Clinical Tuberculosis*, eds Davies PDO, Barnes PF, Gordon SB (Hodder Arnold, London), 4th Ed, pp 519–528.
26. Jaffe HL (1972) *Metabolic, Degenerative and Inflammatory Diseases of Bones and Joints* (Lea and Febiger, Philadelphia).
27. Roberts CA (2012) Re-emerging infections: Developments in bioarchaeological contributions to understanding tuberculosis today. *A Companion to Palaeopathology*, ed Grauer A (Blackwell, Oxford), pp 434–457.
28. Brown TA, Brown KA (2011) *Biomolecular Archaeology: An Introduction* (Wiley-Blackwell, Malden, MA).
29. Willerslev E, Cooper A (2005) Ancient DNA. *Proc Biol Sci* 272(1558):3–16.
30. Salo WL, Aufderheide AC, Buikstra J, Holcomb TA (1994) Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci USA* 91(6):2091–2094.
31. Taylor GM, Goyal M, Legge AJ, Shaw RJ, Young D (1999) Genotypic analysis of *Mycobacterium tuberculosis* from medieval human remains. *Microbiology* 145(Pt 4):899–904.
32. Taylor GM, Young DB, Mays SA (2005) Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. *J Clin Microbiol* 43(5):2236–2240.
33. Taylor GM, Murphy E, Hopkins R, Rutland P, Chistov Y (2007) First report of *Mycobacterium bovis* DNA in human remains from the Iron Age. *Microbiology* 153(Pt 4):1243–1249.
34. Collins DM, Stephens DM (1991) Identification of insertion sequence, IS1081, in *Mycobacterium bovis*. *FEMS Microbiol Lett* 67(1):11–16.
35. Thierry D, et al. (1990) Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis. *J Clin Microbiol* 28(12):2668–2673.
36. Fletcher HA, Donoghue HD, Taylor GM, van der Zanden AG, Spigelman M (2003) Molecular analysis of *Mycobacterium tuberculosis* DNA from a family of 18th century Hungarians. *Microbiology* 149(Pt 1):143–151.
37. Bos KI, et al. (2011) A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478(7370):506–510.
38. Knapp M, Hofreiter M (2010) Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes* 1(2):227–243.
39. Sreevatsan S, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 94(18):9869–9874.
40. Filliol I, et al. (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: Insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol* 188(2):579–772.
41. Gutierrez MC, et al. (2005) Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog* 1(1):e5.
42. Huard RC, et al. (2006) Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J Bacteriol* 188(12):4271–4287.
43. Gori A, Franzetti F, Marchetti G, Catozzi L, Corbellino M (1996) Specific detection of *Mycobacterium tuberculosis* by *mtp40* nested PCR. *J Clin Microbiol* 34(11):2866–2867.
44. Liébana E, Aranaz A, Francis B, Cousins D (1996) Assessment of genetic markers for species differentiation within the *Mycobacterium tuberculosis* complex. *J Clin Microbiol* 34(4):933–938.
45. Weil A, Plikaytis BB, Butler WR, Woodley CL, Shinnick TM (1996) The *mtp40* gene is not present in all strains of *Mycobacterium tuberculosis*. *J Clin Microbiol* 34(9):2309–2311.
46. Koivuola T, Svenson SB, Källenius G (2002) The *mtp40* gene is not present in *Mycobacterium bovis*. *Tuberculosis (Edinb)* 82(4–5):183–185.
47. Iøerger TR, et al. (2010) Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. *J Bacteriol* 192(14):3645–3653.
48. Marmiesse M, et al. (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* 150(Pt 2):483–496.
49. Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* 289(5482):1139.
50. Gilbert MTP, Bandelt H-J, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends Ecol Evol* 20(10):541–544.
51. Donoghue HD, Spigelman M (2006) Comment. Pathogenic microbial ancient DNA: A problem or an opportunity? *Proc Biol Sci* 273(1587):641–642.
52. Bouwman AS, Chilvers ER, Brown KA, Brown TA (2006) Brief communication: Identification of the authentic ancient DNA sequence in a human bone contaminated with modern DNA. *Am J Phys Anthropol* 131(3):428–431.
53. Konomi N, Lebwahl E, Mowbray K, Tattersall I, Zhang D (2002) Detection of mycobacterial DNA in Andean mummies. *J Clin Microbiol* 40(12):4738–4740.
54. Brimacombe M, Hazbon M, Motiwala AS, Alland D (2007) Antibiotic resistance and single-nucleotide polymorphism cluster grouping type in a multinational sample of resistant *Mycobacterium tuberculosis* isolates. *Antimicrob Agents Chemother* 51(11):4157–4159.
55. Caffell AC, Holst M (2009) Osteological analysis. *St. George's Crypt, Leeds, West Yorkshire. Report No 0409* (York Osteoarchaeology Ltd., Yorkshire, UK).
56. Roberts CA, Lucy D, Manchester K (1994) Inflammatory lesions of ribs: An analysis of the Terry Collection. *Am J Phys Anthropol* 95(2):169–182.
57. Bouwman AS, Brown TA (2005) The limits of biomolecular palaeopathology: Ancient DNA cannot be used to study venereal syphilis. *J Archaeol Sci* 32(5):703–713.
58. Camus JC, Pryor MJ, Médigue C, Cole ST (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148(Pt 10):2967–2973.
59. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
60. Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.
61. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7(1–2):203–214.
62. Huson DH, Scornavacca C (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 10.1093/sysbio/sys062.
63. Rohland N, Siedel H, Hofreiter M (2010) A rapid column-based ancient DNA extraction method for increased sample throughput. *Mol Ecol Resour* 10(4):677–683.
64. Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. *Nat Protoc* 2(7):1756–1762.