Tuberculosis 137 (2022) 102181

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: www.elsevier.com/locate/tube

Metagenomic analysis reveals mixed *Mycobacterium tuberculosis* infection in a 18th century Hungarian midwife

Heidi Y. Jäger [a,*], Frank Maixner [a,**], Ildikó Pap [b,c,d], Ildikó Szikossy [b,c], György Pálfi [b,***,1], Albert R. Zink [a,1]

[a] Institute for Mummy Studies, Eurac Research, Viale Druso, 1, 39100, Bolzano, Italy
[b] Department of Biological Anthropology, Faculty of Science and Informatics, University of Szeged, 6726, Szeged, Közép Fasor 52, Hungary
[c] Department of Anthropology, Hungarian Natural History Museum, 1083, Budapest, Ludovika tér 2-6, Hungary
[d] Department of Biological Anthropology, Eötvös Loránd University, Faculty of Science, 1117, Budapest, Pázmány Péter sétány 1/c, Hungary

ARTICLE INFO

Keywords:
Ancient DNA
Metagenomics
Mixed infection
Single nucleotide polymorphisms
Phylogeny

ABSTRACT

The Vác Mummy Collection comprises 265 well documented mummified individuals from the late 16th to the early 18th century that were discovered in 1994 inside a crypt in Vác, Hungary. This collection offers a unique opportunity to study the relationship between humans and pathogens in the pre-antibiotic era, as previous studies have shown a high proportion of tuberculosis (TB) infections among the individuals. In this study, we recovered ancient DNA with shotgun sequencing from a rib bone sample of a 18th century midwife. This individual is part of the collection and shows clear skeletal changes that are associated with tuberculosis and syphilis. To provide molecular proof, we applied a metagenomic approach to screen for ancient pathogen DNA. While we were unsuccessful to recover any ancient *Treponema pallidum* DNA, we retrieved high coverage ancient TB DNA and identified a mixed infection with two distinct TB strains by detailed single-nucleotide polymorphism and phylogenetic analysis. Thereby, we have obtained comprehensive results demonstrating the long-time prevalence of mixed infections with the sublineages L4.1.2.1/Haarlem and L4.10/PGG3 within the local community in preindustrial Hungary and put them in context of sociohistorical factors.

1. Introduction

Tuberculosis (TB) is an ancient disease that co-evolved with humans many millennia ago [1,2]. Its evolutionary history has been of great interest because, to date, it remains a globally spread infectious disease caused by the members of the *Mycobacterium tuberculosis* complex (MTBC) and was responsible for approximately 1.4 million deaths in 2019 alone [3]. Over the past decade, vast achievements in high throughput sequencing technologies have drastically increased the amount of genomic pathogen data retrieved from historical human remains. The reconstruction of ancient pathogen genomes is highly valuable as it allows the assessment of pathogen occurrence in past populations, and provides important insight into host-pathogen evolution, host susceptibility, and host resistance [4–10]. Nevertheless, there is only a limited number of available ancient TB genomes, as Mycobacteria are highly clonal and historical specimens are often contaminated with environmental bacterial DNA that are hard to distinguish from authentic ancient tuberculous DNA [11].

The aim of this study was to extend the current knowledge of TB by supporting the number of ancient TB genomes of the Vác Mummy Collection (VMC). The VMC, housed at the Department of Anthropology of the Hungarian Natural History Museum in Budapest, Hungary, comprises the remains of 265 well documented individuals and allows a unique insight into the relationship between human populations and infectious diseases during the pre-urbanization era of 17th-18th century Central Europe. The collection came to light in 1994 during reconstruction work of the 18th-century Dominican "Church of Whites" (Fig. 1B), which is located in the city of Vác, about 70 km North from

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: heidi.jaeger@eurac.edu (H.Y. Jäger), frank.maixner@eurac.edu (F. Maixner), ildiko.pap.2@hotmail.com (I. Pap), szikossy@gmail.com (I. Szikossy), palfigy@bio.u-szeged.hu (G. Pálfi), albert.zink@eurac.edu (A.R. Zink).
[1] G.P. and A.R.Z. should be considered joint senior authors.

https://doi.org/10.1016/j.tube.2022.102181
Received 7 December 2021; Received in revised form 12 January 2022; Accepted 3 February 2022
Available online 21 February 2022

Budapest (Fig. 1C). Due to the cold and dry microclimate and continuous ventilation in the crypt, many of the bodies remained in different states of natural mummification for over 200 years. The pinewood coffins filled with pine shavings, often preserved the names and years of death of the deceased, revealing that the burial site was used from 1731 until 1838. In addition, archives or other contemporary registers offered further details such as the age of death, sex, family relationship, and sometimes occupation and cause of death of the individual. Several studies have been conducted using macromorphological and radiographic examination, polymerase chain reaction-based detection, and high-performance liquid chromatography approaches, confirming a very high proportion of tuberculosis (TB) infections within the deceased individuals from Vác [12–17]. TB was epidemic in Europe with a seasonal pattern from the 18th to 19th century. Denser populations, poor sanitation, and malnutrition caused by urbanization were driving forces for the spread of TB until its peak during the Industrial Revolution in the mid-19th century and then decline due to general improvements of living conditions [18,19]. Fletcher et al. and Kay et al. were the first to raise the epidemiological aspect of this collection, with the latter successfully reconstructing 14 ancient TB genome sequences from eight individuals and demonstrating a high prevalence of mixed infections (MIs) with the Euro-American Lineage 4 (L4) [14,20,21].

In this study, we present a thus far unpublished individual, the late Szabina Orlich, a 62-year-old midwife who died in 1755 in Vác (from here on referred to as body210) and whose partially mummified remains were discovered inside the crypt. Beside year of birth and death as well as occupation, macromorphological examination revealed palaeopathological evidence indicating a probable TB and syphilis co-infection. To provide molecular proof for such co-infection, we applied a metagenomic shotgun-sequencing approach to recover ancient DNA from a partial rib bone sample of body210 and screened for pathogen DNA. While we could not confirm any treponemal DNA, we retrieved high coverage ancient TB genomic DNA. Hence, in this study, we focused on extending the collection of available ancient TB genomes and successfully identified a mixed TB infection with two distinct strains of L4 sublineages L4.1.2.1/Haarlem and L4.10/PGG3 by single-nucleotide polymorphism (SNP) based and phylogenetic analysis. Furthermore, we compared our findings to historical records and social factors to identify how they might have contributed to the local dispersal of TB at that time.

2. Methods

2.1. Sample preparation and shotgun sequencing

The remains of body210 were first macromorphologically examined before a rib bone sample (Fig. 1A) was collected at the Department of Anthropology of the Hungarian National History Museum in Budapest, Hungary. All preparation steps including sampling, DNA extraction, and library preparation were performed in a dedicated clean room laboratory located at the Eurac Research, Institute for Mummy Studies in Bolzano, Italy. The surface of the sample was UV irradiated for 5 min before processing a part of the bone into powder with a vibrating mill. For DNA extraction, 100 mg of bone powder were incubated at 56 °C for 48 h with 0.5 M EDTA (pH 8) and 20 mg/μL Proteinase K and then subjected to three times snap freezing with liquid nitrogen. For DNA purification and precipitation, a modified protocol by Maixner et al. was applied [22]. The extracted genomic DNA was converted into two double-indexed and double-stranded Illumina libraries (body210A, body210B) according to the published protocol of Meyer and Kircher [23] and sequenced on the Illumina HiSeqX platform using the 151 bp paired-end sequencing kit.

2.2. Post-sequencing processing and taxonomic profiling

After quality evaluation with FastQC v.0.11.8 [24], paired-end reads of library body210A and body210B were subjected to adapter removal and quality filtering using fastp v.0.20.1 with standard parameters [25]. All quality passed reads of both libraries were merged into one dataset (body210). All reads under 30 bp were removed using seqkit v.0.8.2 [26]. For general taxonomic profiling DIAMOND v.2.0.4 blastx was run against the non-redundant protein database [27]. The output was converted with MEGAN v.6.19.9 [28] and visualized using Krona v.2.7.1 [29].

2.3. Human and Treponema pallidum DNA analysis

Paired-end reads of body210 were mapped against the full human genome (build hg19) [30], human mitochondrial genome (rCRS) [31], and the *Treponema pallidum* reference genome (GenBank accession no. GCA_000604125.1) using Bowtie2 v.2.3.5 with standard parameters [32]. After filtering out alignments with mapping qualities lower than 30 with samtools v.1.12 [33], remaining reads were deduplicated with DeDup v.0.11.3 [34]. To verify the authenticity of ancient DNA, the damage pattern was calculated with DamageProfiler v.1.1 [35].

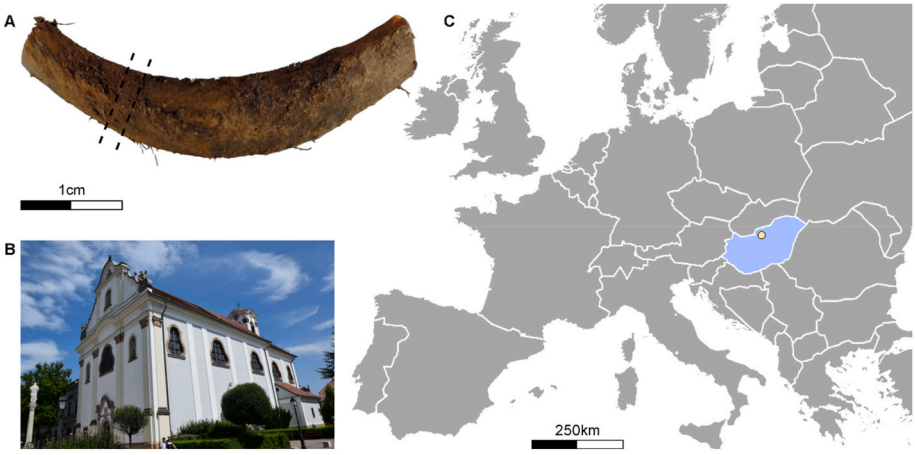

Fig. 1. Material used in this study and site description. A, partial rib bone of body210; dotted lines indicate the bone section used for ancient DNA extraction B, the Dominican Church of Whites in Vác, Hungary C, map of Europe with Hungary highlighted in blue and city of Vác with yellow dot. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.4. Mycobacterium tuberculosis DNA analysis

Paired-end reads of body210 were mapped against the reconstructed ancestral MTB genome [36] with stringent Bowtie2 parameters not allowing more than three mismatches per 100 bases. The resulting alignment file was quality evaluated using QualiMap2 v.2.2.1 [37]. Variant calling for MTB reads was performed using samtools mpileup with a minimum base and mapping quality of 30. All variants were subsequently filtered with VarScan v.2.4.2 mpileup2snp calling SNPs with a minimum read and SNP coverage of 5X [38].

2.5. Mixed infection detection

The datasets of three Vác individuals published by Kay et al. were included for comparative analysis [21]. Body68 and body92 are positive controls for MIs with two strains and body80 serves as a negative control for MIs as this individual was infected with only a single TB strain. We considered homozygous SNPs with a frequency of 100% as fixed for all strains and discarded heterozygous SNPs with frequencies lower than 5% or higher than 95%. The number of heterozygous SNPs was visualized in a frequency histogram and was compared with the SNP distribution pattern of the comparative datasets. Additionally, to investigate whether the observed allele frequency range is reliable, lineage-specific SNPs were identified using a reference database [39] and were highlighted in a scatter plot representing the allele frequency of each SNP against their respective position on the reference genome.

2.6. Phylogenetic reconstruction

We compiled two datasets for phylogenetic tree reconstruction including 252 previously published modern MTBC genomes and 142 previously published modern L4 genomes of the MTBC, respectively [40–42]. Both datasets include the genome data presented in this publication as well as four previously published ancient genome datasets [21,43]. *Mycobacterium canetti* was used as an outgroup for the MTBC dataset and L2_N0020 for the L4 dataset. All reference datasets were subjected to the alignment and variant calling pipeline as described for the body210 dataset using standard mapping parameters in Bowtie2. Overrepresented regions such as proline-glutamic acid (PE) and proline-proline-glutamic acid (PPE) protein families were removed (Data S1 and S2), and multi-FASTA alignment files were generated using bcftools v.1.12. After selecting the optimal substitution model with the ModelFinder [44] program implemented in IQ-TREE 2 v.2.1.2 [45], two maximum likelihood trees with 1,000 bootstrap samples were generated. The resulting tree files in Newick format were visualized and annotated with iTOL v.6.3 [46].

3. Results

3.1. Palaeopathological examination

The partially mummified state of body210 exposed gross pathologies on the bones. Morphological studies of the non-mummified bony elements of the lower limbs revealed advanced-stage osteo-periosteal lesions and typical osteitis (Fig. S1C), corresponding to Hackett's criteria in treponematosis [47]. The thickened and deformed tibiae were involved most severely (Fig. S1A). Less developed alterations were detected in the long bones of the upper limbs. Clustered pits - as earliest phase of *caries sicca* – are visible on the outer surface of the skull vault (Fig. 2D) [48–50] and advanced stage of osteoarthritis of the first metatarsophalangeal joint and well-developed marginal osteophytes were probably caused by the chronic overuse of this articulation in dorsiflexion (Fig. S1B). This degenerative articular process correlates with the long-term kneeling or sitting in a low chair in the everyday working life of a midwife. Three right-side ribs have evident traces of periosteal appositions and remodelling on the visceral surfaces indicating a probable chronic pulmonary infection consistent with tuberculosis (Fig. 2A–C). Palaeopathological studies have shown that inflammatory rib lesions are an important indicator in differential diagnosis of tuberculosis in skeletal remains [51].

3.2. Ancient DNA authentication and taxonomic profiling

We assessed the overall quantity and quality of sequenced reads and summarized the obtained data in Table S1. According to the generated taxonomic profile, over 77% of the reads were assigned to the MTBC, 9% to Firmicutes, 12% to other bacteria, leaving 2% of the reads unassigned (Fig. S2). We could not detect any *Treponema pallidum* specific reads. The human endogenous content was very low and therefore insufficient to perform ancient DNA authentication (Table S2). However, we could confirm a damage pattern and short reads typical for ancient DNA in MTB specific reads (Figs. S3A–C). In total, 2,559,128 unique reads mapped to the MTB ancestral genome with a mean coverage of 39.5-folds and a standard deviation of 52.9-folds.

3.3. Verification of mixed infection

A total of 513 SNPs were called for body210 out of which 245 SNPs

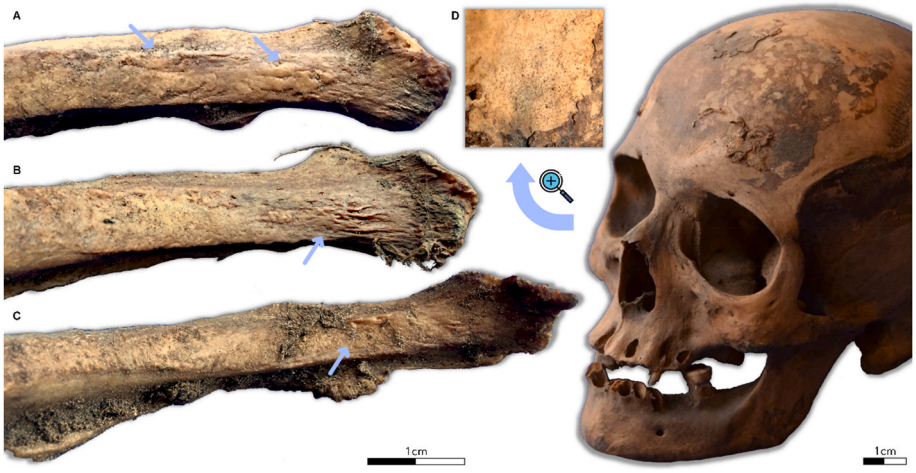

Fig. 2. Skeletal remains of body210 with palaeopathological evidence. A-C, three right-side rib bones with evidence for longer term remodelling and lamellar bone on the visceral surfaces (blue arrows). D, the skull of body210 with mummified tissue remains visible on the frontal lobe and close-up image of clustered pits indicating a syphilis infection. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

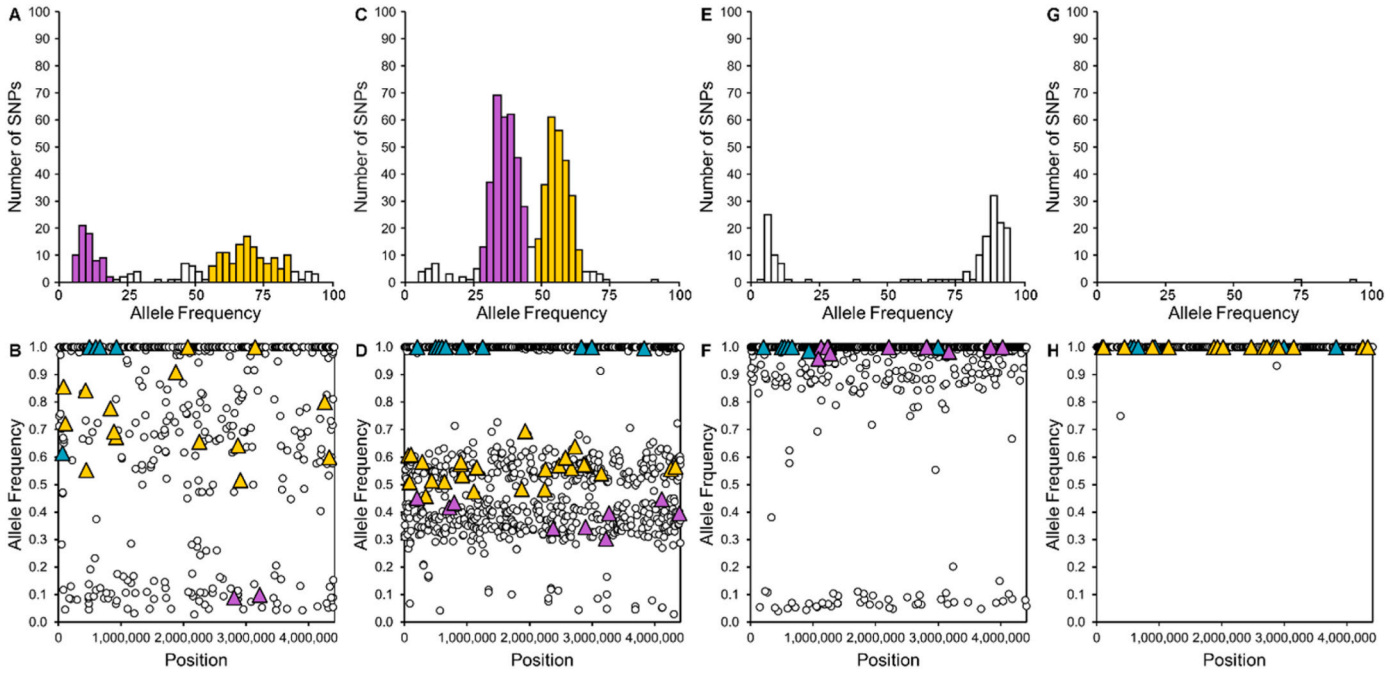

Fig. 3. Heterozygous and lineage-specific SNPs analysis for MI detection. Top row: Histograms with number of heterozygous SNPs for each allele frequency (1 bar = 2.5%). Only SNPs with allele frequency of 5–95% are included in analysis. Highlighted bars indicate allele frequencies of strains. Bottom row: Scatterplots with allele frequency of SNPs against their respective position. Triangles indicate lineage specific SNPs. A-B, body210, C-D, body68, and E-F, body92 are positive controls for MIs and G-H, body80 for single strain infection. Colour key: turquoise: L4, yellow: L4.1.2/Haarlem, purple: L4.10/PGG3. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are homozygous. The SNP frequency distribution pattern of body210 with two prominent peaks in Fig. 3A is in line with the similar bimodal SNP distribution observed in body68 (Fig. 3C) and body92 (Fig. 3E), indicating a MI with two strains. The mean and standard deviation of the majority strain body210-1 and minority strain body210-2 allele frequencies are 68.20 ± 12.14% and 12.91 ± 6.08%, respectively. Except for a few outliers, no heterozygous SNPs were found in body80 (Fig. 3G), confirming with a single strain infection.

Scatterplots are used to analyse strain frequencies as they are visualized by SNP clusters at a certain allele frequency along the reference genome (Fig. 3, bottom row). Most likely due to variable coverage, SNPs in body210 were too widely scattered for a confident estimation of the strain frequencies. Therefore, lineage specific SNPs were manually inspected (Fig. 3B). For body210, a total of 21 L4 specific SNPs were detected, out of which 14 are specific for L4.1 and L4.1.2/Haarlem and two for L4.7 and L4.8, also known as L4.10/PGG3 (Table S3). The mean and standard deviation for heterozygous SNPs is 70.53 ± 11.91% for L4.1 and L4.1.2/Haarlem and 9.47 ± 0.45% for L4.10/PGG3, which is accordance with our earlier observations. Further, there is a weak negative correlation between the allele frequency of a SNP and its respective total coverage, pointing to an overestimation of the body210-1 majority strain frequency with lower coverage SNPs (Fig. S4). Contrary to expectations, the strain frequencies of body210 are not as proportional as observed in body68 and body92 (Fig. 3D and F). We exclude that low frequency SNPs derived from non-tuberculous reads by identifying L4.10/PGG3 specific SNPs and argue that the discrepancy of strain frequency is caused by uneven genome coverage (Fig. S5). For further phylogenetical analysis, we included heterozygous SNPs between 55% and 85% and all homozygous SNPs for body210-1. Body210-2 was excluded from further analyses due to insufficient data.

Body68 was divided into majority strain body68-1 including heterozygous SNPs between 30% and 46%, and minority strain body68-2 including heterozygous SNPs between 48% and 64%, as previously stated by Kay et al. [21]. Homozygous SNPs were included in both body68 datasets (Fig. 3D). We excluded the minority strain of body92 as we could not identify lineage specific SNPs within the low frequency SNPs (Fig. 3F). Hence, only heterozygous SNPs with a frequency over 95% and homozygous SNPs were included for this dataset.

3.4. Phylogeny

All genome datasets used in this analysis are summarized in the Data S3 and S4. Body68-1, body68-2, body80, body92 and LUND1 were included as ancient reference datasets as they had sufficient genome coverage for further phylogenetical analysis (Table S4).

The MTBC dataset included 16,172 parsimony informative sites and the L3PU + F + R19 model was estimated as the best fit model by ModelFinder. The TVM + F + R4 model was applied for the L4 dataset with 5,112 parsimony informative sites. We had high confidence values for all branches of both trees (Fig. 4A and B). Already published ancient genome datasets were assigned correctly to the same sublineages as previously reported and served as reference points [21,43]. As expected body210-1 was grouped with other Vác individuals in L4, more specifically L4.1.2.1/Haarlem. Body210-2 was highlighted on a basal level according to L4.10/PGG3 specific SNPs in Fig. 4B.

4. Discussion

In this study, we successfully recovered high coverage ancient MTB genome data from a bone sample of a thus far undescribed individual of the VMC. The molecular analyses were initiated based on evidence of palaeopathological skeletal changes across the partially mummified remains of body210, indicating a possible TB-syphilis co-infection. However, the taxonomic profile of the metagenomic data acquired by shotgun sequencing displayed a high proportion of ancient MTB reads but no reads for *Treponema pallidum*. The pathogen load of *Treponema pallidum* is initially relatively high in primary lesions but becomes extremely low after it spreads throughout the body of the host and enters the tertiary or latency stage. Skeletal changes in the host most frequently occurs in this final phase of infection, therefore making it challenging to

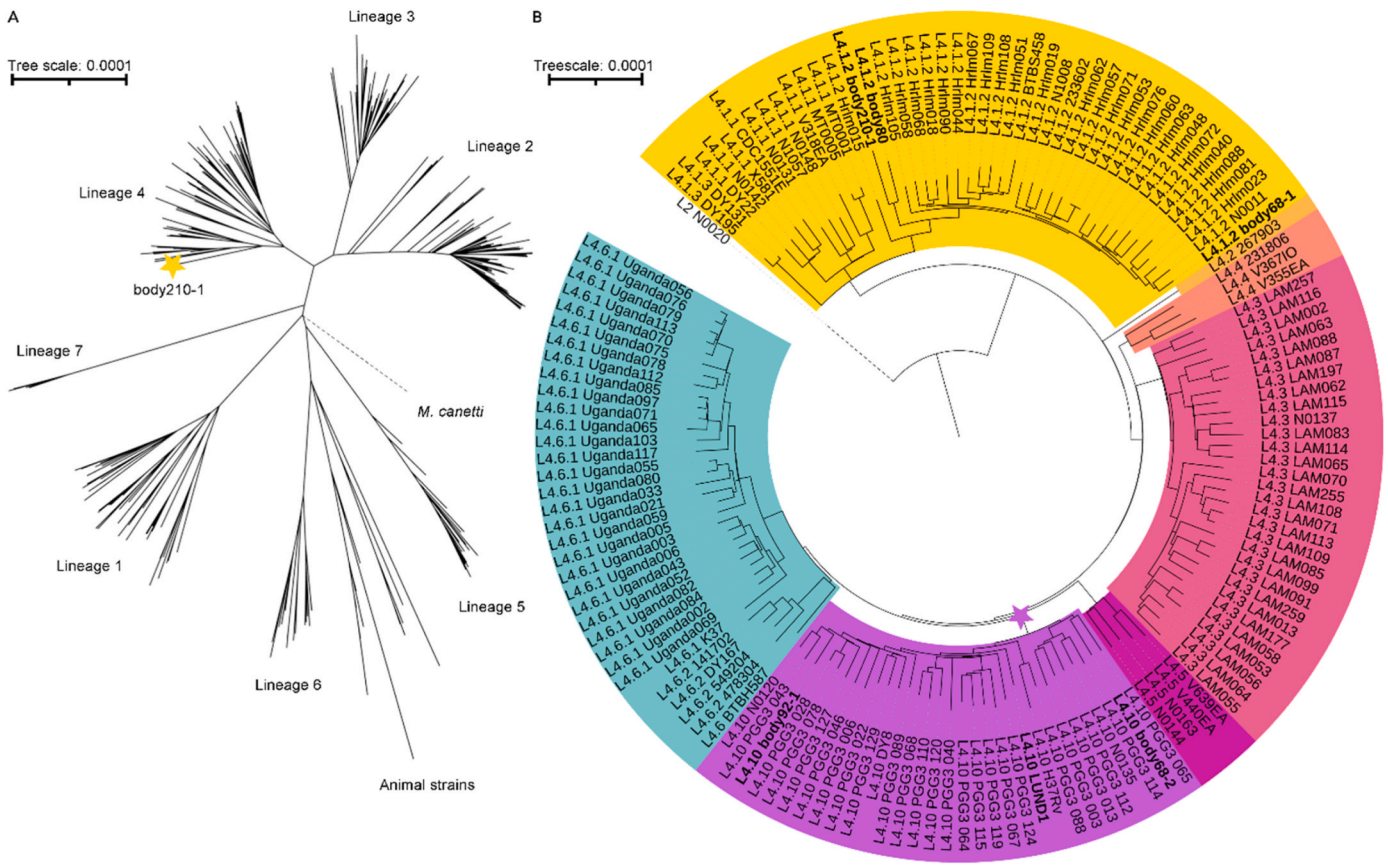

Fig. 4. Maximum-likelihood trees with 1,000 bootstrap samples. Generated with IQ-TREE 2 visualized with iTOL. A, unrooted maximum likelihood tree of MTBC dataset, M. canetti served as an outgroup; yellow star highlighting body210-1. B, circular maximum likelihood tree of L4 dataset, L2 N0020 as outgroup; purple star highlighting body210-2. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

detect treponemal DNA from such individuals. In addition, the pathogen load in a sample can vary depending on the tissue and specific sampling point. Lastly, *post-mortem* growth of other bacteria and the associated primary putrefaction processes causes degradation of originally existing endogenous DNA, including that of the pathogen, consequently making it no longer detectable [52,53].

We were able to confirm a mixed TB infection with two strains of the MTBC L4 sublineages L4.1.2.1/Haarlem and L4.10/PGG3, though the data for the latter strain was not sufficient to perform a phylogenetic analysis and we could only confirm it based on lineage-specific SNPs. These genotypes as well as sublineage L4.3/LAAM were previously identified in other individuals of the VMC. L4 belongs to one of the most globally widespread and virulent strains of the human adapted MTBC and comprises ten sublineages. The geographical distribution of these sublineages is likely linked to host-susceptibility to TB and host genetic diversity. L4 is divided into specialist sublineages which are restricted to specific regions and therefore might be adapted to local host populations and generalists, which are genetically more diverse than specialists. This means that generalists have higher proportions of variable human T cell epitopes and indicates past interactions with diverse immune escape mechanisms and/or host populations. Generalists and specialists acquired their ability to adapt to different host population independently from each other and were probably dispersed through different levels of mobility as they are entirely dependent on humans to migrate. Interestingly, all thus far identified Vác L4 genotypes L4.1.2/Haarlem, L4.3/LAM, and L4.10/PGG3 belong to generalists, which are the most frequently found L4 sublineages today in Europe [54].

Moreover, MIs with more than one strain of MTB have become an increasingly frequent phenomenon in high burden environments and often cause heteroresistance, compromising the treatment success of patients. Nevertheless, the detection of MIs remains complex due to the lack of distinct intraspecies markers [55]. Compared to other strain typing methods that are limited to specific nucleotide sequences, whole genome sequencing can provide high resolution data to decipher the genetic diversity of the MTBC, detect drug resistances, and transmission dynamics [56–59]. SNPs are informative indicators for genotyping and distinction between true MIs and other types of infections such as microevolution, reinfections, or relapses [60–62]. A standardized approach to analyse MIs has not yet been proposed and potentially informative SNPs are often dismissed in the analysis as either sequencing or alignment errors [1,43,63]. Here, we were able to verify a MI with two distinct strains by applying stringent alignment parameters and analysing heterozygous and lineage-specific SNPs. The verification of this mixed TB infection places the introduction of L4.1.2.1/Haarlem and L4.10/PGG3 to the local community to an earlier time point than previously suspected (Fig. 5), demonstrating a long-time prevalence of over 100 years in pre-antibiotic Hungary between the 17th and 19th century, almost 100 years before TB reached its peak in Hungary between 1896 and 1905 [64].

The community of Vác was greatly affected by a series of terrible events starting from the late 17th century (Fig. 5). In 1686, after the end of Turkish rule for over 150 years, the city was left in ruins, with only a few families surviving. An era of recovery followed with great efforts made by the Catholic Church to repopulate and rebuild the city, starting with the building of the Church of Whites in 1699. Families who had once fled as refugees resettled, and many new artisan families from German, Czech, and Moravian regions who were promised economic benefits arrived. The city's development was halted when it fell under the control of the Habsburg Empire and was severely devastated by Serbian mercenaries in 1709. Two-thirds of Vác burned down in 1731,

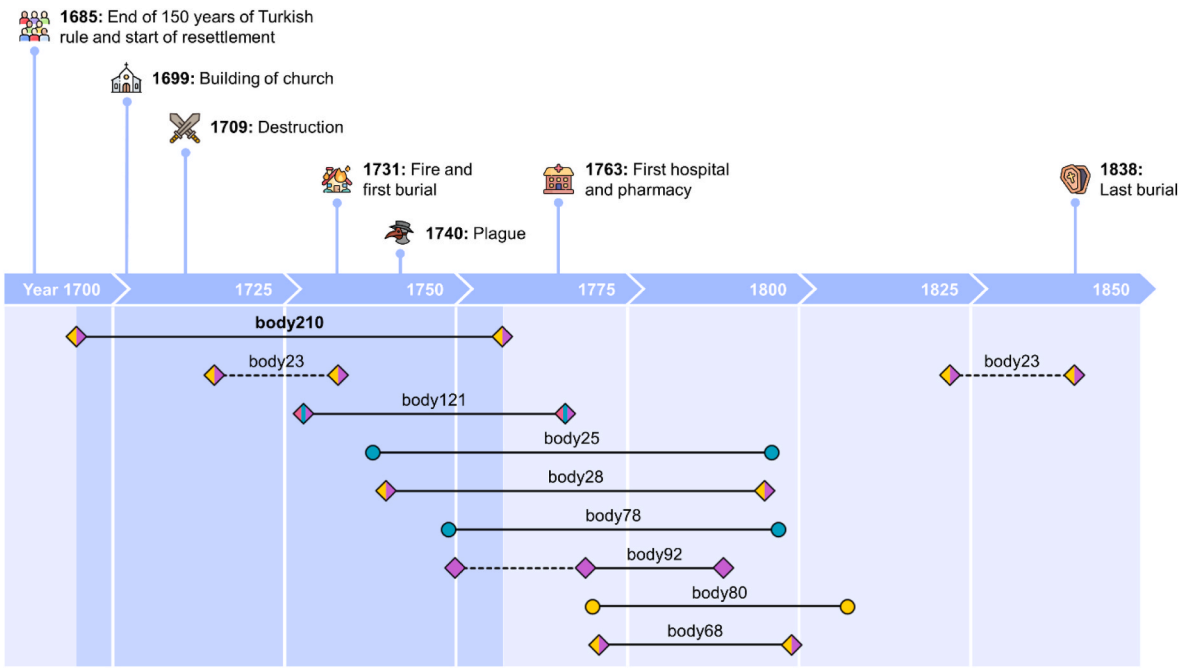

Fig. 5. Local historical events of Vác, Hungary between year 1685 and 1838 CE in relation to the lifetime of eight additional Vác individuals with available TB genome data from Kay et al. 2015. Shape indicates type of infection (circles: single strain infection; diamonds: mixed strain infection) and colours indicate strain type (turquoise: L4, yellow: L4.1.2.1/Haarlem; pink: L4.3/LAAM; purple: L4.10/PGG3). Dotted lines indicate ambiguous dates. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the same year as the first burial in the crypt. In 1740, soldiers returned from war carrying the plague, killing ten percent of the population. The first hospital and pharmacy were not built until 1763 [65]. We suggest that the high rates of population mobility and the establishment of strong social ties might have led to the introduction of new TB strains to the community and that body210, who was born in 1693, was most likely exposed to TB from early age. Upon infection, it is common for MTB to enter a latent state that can last for decades and during which no transmissions occur. This evolutionary strategy of MTB is considered to have developed through co-evolution with their human host to sustain in small populations over time [66,67]. The urbanization of the city and the substandard living conditions caused by the repeated destruction were both advantageous factors for the persistence of TB within the community, which is in accordance with the diversity of L4 strains and high prevalence of MIs as previously shown by Kay et al. and is now further supported by our findings [21]. Furthermore, growing host population density and distinct demographic histories are associated with phenotypic differences such as increased pathogen virulence, reduced latency periods or higher transmission potential [58,66,68]. While the concrete infection scenario with multiple TB strains of body210 remains inconclusive, we believe that the advanced stage bone remodelling on the ribs of body210 indicate an active and long-term tuberculosis infection until the time of death. Additionally, due to the individual's profession as a midwife and the lack of medical facilities during lifetime of body210, we argue that the individual had frequent contact with different households and consequently with potential actively infectious TB hosts. This, considered together with the acute inflammation of joints and a possible co-infection with syphilis, points to a continuous weakening of the immune system, ultimately leading to an active TB infection. For comparison to body210, the lifetime and the type of TB infection of the eight individuals with available TB genomic data published by Kay et al. are shown in Fig. 5. This highlights the possibility that L4.1.2.1/Haarlem and L4.10/PGG3 were likely introduced to the community at an even earlier timepoint at the turn of the century.

5. Conclusions

Our study provides further evidence for the long-time prevalence of L4 generalists and high occurrence of MIs within the community of Vác. In addition, we highlight the unique potential of the VMC to offer highly valuable information about genetical and historical epidemiological links within the community to reconstruct transmission events. This is useful for better estimation of mutation rates and how different TB strains have spread and persisted on a community and global level in the presence of other virulent pathogens at the dawn of the Industrial Revolution.

Funding

This work was supported by the European Regional Development Fund 2014–2020 CALL FESR2017 Research and Innovation Autonomous Province of Bolzano South Tyrol Project: FESR1078 MummyLabs, the University of Szeged Open Access Fund (I.P., I.S., G.P.), and the National Research Development and Innovation Office (Hungary) (grant number: NKTIH K 125561) (G.P., I.P., I.S.). H.Y.J. was supported by the German Academic Exchange Service (57437987) and the Eurac Research PhD Programme.

Data accessibility statement

Raw sequencing data generated within this study are uploaded to the NCBI Sequence Read Archive under BioProject PRJNA795622 (SRA accessions: SRR17499018, SRR17499019).

CRediT authorship contribution statement

Heidi Y. Jäger: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **Frank Maixner:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing.

Ildikó Pap: Investigation, Resources, Writing – review & editing. **Ildikó Szikossy:** Investigation, Resources. **György Pálfi:** Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing. **Albert R. Zink:** Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

We declare we have no competing interests.

Acknowledgements

Céline Jacqueroud is acknowledged for the documentation of the sample material.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tube.2022.102181.

References

[1] Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 2014;514:494–7. https://doi.org/10.1038/nature13591.
[2] Baker O, Lee OY-C, Wu HHT, Besra GS, Minnikin DE, Llewellyn G, et al. Human tuberculosis predates domestication in ancient Syria. Tuberculosis 2015;95:S4–12. https://doi.org/10.1016/j.tube.2015.02.001.
[3] World Health Organization. Global tuberculosis report 2020: executive summary. 2020. Geneva.
[4] Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, et al. The 5300-year-old Helicobacter pylori genome of the Iceman. Science 2016;351:162–5. https://doi.org/10.1126/science.aad2545.
[5] Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. Elife 2018;7. https://doi.org/10.7554/eLife.36666.
[6] 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. https://doi.org/10.1371/journal.ppat.1006997.
[7] Correa-Macedo W, Cambri G, Schurr E. The interplay of human and *Mycobacterium tuberculosis* genomic variability. Front Genet 2019;10. https://doi.org/10.3389/fgene.2019.00865.
[8] Spyrou MA, Keller M, Tukhbatova RI, Scheib CL, Nelson EA, Andrades Valtueña A, et al. Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. Nat Commun 2019;10. https://doi.org/10.1038/s41467-019-12154-0.
[9] Mühlemann B, Vinner L, Margaryan A, Wilhelmson H, de la Fuente Castro C, Allentoft ME, et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science 2020;369. https://doi.org/10.1126/science.aaw8977.
[10] Kerner G, Laval G, Patin E, Boisson-Dupuis S, Abel L, Casanova J-L, et al. Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years. Am J Hum Genet 2021;108:517–24. https://doi.org/10.1016/j.ajhg.2021.02.009.
[11] Müller R, Roberts CA, Brown TA. Complications in the study of ancient tuberculosis: presence of environmental bacteria in human archaeological remains. J Archaeol Sci 2016;68:5–11. https://doi.org/10.1016/j.jas.2016.03.002.
[12] Szikossy I, Bernert Z, Pap I. Anthropological investigation of the 18th-19th century ossuary of the Dominican Church at Vác, Hungary. Acta Biol Szeged 1997;42:145–50.
[13] Pap I, Józsa L, Repa I, Bajzik G, Lakhani S, Donoghue H, et al. 18-19th century tuberculosis in naturally mummified individuals (Vác, Hungary). Tuberculosis: past and Present. Budapest - Szeged, Hungary: Golden Book Publisher Ltd.; 1999. p. 421–8.
[14] Fletcher HA, Donoghue HD, Holton J, Pap I, Spigelman M. Widespread occurrence of Mycobacterium tuberculosis DNA from 18th-19th century Hungarians. Am J Phys Anthropol 2003;120:144–52. https://doi.org/10.1002/ajpa.10114.
[15] Donoghue HD, Pap I, Szikossy I, Spigelman M. Detection and characterization of *Mycobacterium tuberculosis* DNA in 18th century Hungarians with pulmonary and extra-pulmonary tuberculosis. Yearbook of Mummy Studies 2011;1:51–6.
[16] Váradi OA, Szikossy I, Spekker O, Rakk D, Terhes G, Urbán E, et al. Lipid biomarker-based verification of TB infection in mother's and daughter's mummified human remains (Vác Mummy Collection, 18th century, CE, Hungary). Acta Biol Szeged 2021;64:99–109. https://doi.org/10.14232/abs.2020.2.99-109.
[17] Váradi OA, Rakk D, Spekker O, Terhes G, Urbán E, Berthon W, et al. Verification of tuberculosis infection among Vác mummies (18th century CE, Hungary) based on lipid biomarker profiling with a new HPLC-HESI-MS approach. Tuberculosis 2021;126. https://doi.org/10.1016/j.tube.2020.102037.
[18] Murray JF. The Industrial Revolution and the decline in death rates from tuberculosis. Int J Tubercul Lung Dis 2015;19:502–3. https://doi.org/10.5588/ijtld.14.0856.
[19] Spekker O, Schultz M, Paja L, Váradi OA, Molnár E, Pálfi G, et al. Tracking down the White Plague. Chapter two: the role of endocranial abnormal blood vessel impressions and periosteal appositions in the paleopathological diagnosis of tuberculous meningitis. PLoS One 2020;15. https://doi.org/10.1371/journal.pone.0238444.
[20] Fletcher HA, Donoghue HD, Taylor GM, van der Zanden AG, Spiegelman M. Molecular analysis of *Mycobacterium tuberculosis* DNA from a family of 18th century Hungarians. Microbiology 2003;149(Pt 1):143–51. https://doi.org/10.1099/mic.0.25961-0.
[21] Kay GL, Sergeant MJ, Zhou Z, Chan JZ-M, Millard A, Quick J, et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. Nat Commun 2015;6. https://doi.org/10.1038/ncomms7717.
[22] Maixner F, Mitterer C, Jäger HY, Sarhan MS, Valverde G, Lücker S, et al. Linear polyacrylamide is highly efficient in precipitating and purifying environmental and ancient DNA. Meth Ecol Evol 2021. https://doi.org/10.1111/2041-210X.13772.
[23] Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010;2010. https://doi.org/10.1101/pdb.prot5448.
[24] Andrews S. FastQC A quality control tool for high throughput sequence data. 2014 [Online].
[25] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
[26] Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0163962.
[27] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12:59–60. https://doi.org/10.1038/nmeth.3176.
[28] Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol 2016;12. https://doi.org/10.1371/journal.pcbi.1004957.
[29] Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinf 2011;12. https://doi.org/10.1186/1471-2105-12-385.
[30] Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 2015;43:D670–81. https://doi.org/10.1093/nar/gku1177.
[31] Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;23. https://doi.org/10.1038/13779.
[32] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
[33] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
[34] Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome Biol 2016;17. https://doi.org/10.1186/s13059-016-0918-z.
[35] Neukamm J, Peltzer A, Nieselt K. DamageProfiler: fast damage pattern calculation for ancient DNA. Bioinformatics 2021;37:3652–3. https://doi.org/10.1093/bioinformatics/btab190.
[36] Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet 2010;42:498–503. https://doi.org/10.1038/ng.590.
[37] Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 2015;32:292–4. https://doi.org/10.1093/bioinformatics/btv566.
[38] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76. https://doi.org/10.1101/gr.129684.111.
[39] Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. Genome Med 2020;12. https://doi.org/10.1186/s13073-020-00817-3.
[40] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet 2013;45. https://doi.org/10.1038/ng.2744.
[41] Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, et al. Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. Emerg Infect Dis 2013;19:969–76. https://doi.org/10.3201/eid1906.121012.
[42] Domogalla J, Prodinger WM, Blum H, Krebs S, Gellert S, Muller M, et al. Region of difference 4 in alpine *Mycobacterium caprae* isolates indicates three variants. J Clin Microbiol 2013;51:1381–8. https://doi.org/10.1128/JCM.02966-12.
[43] Sabin S, Herbig A, Vågene ÅJ, Ahlström T, Bozovic G, Arcini C, et al. A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex. Genome Biol 2020;21. https://doi.org/10.1186/s13059-020-02112-1.
[44] Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.

[45] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
[46] Letunic I, Bork P. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49:W293–6. https://doi.org/10.1093/nar/gkab301.
[47] Hackett CJ. An introduction to diagnostic criteria of syphilis, treponarid and Yaws (treponematoses) in dry bones, and some implications. Virchows Arch A Pathol Anat Histol 1975;368:229–41. https://doi.org/10.1007/BF00432525.
[48] Ortner D. Identification of pathological conditions in human skeletal remains. second ed. London: Academic Press; 2003.
[49] Powell ML, Cook DC. The myth of syphilis: the natural history of treponematosis in North America. first ed. University Press of Florida; 2005.
[50] Pap I, Szikossy I, Váradi O, Szekeres A, Karlinger K, Pölöskei G, et al. Paleopathological study of an 18th century midwife's mummy (Vác, Hungary). A probable syphilis – tuberculosis co-infection. Athanatos. Extraordinary World Congress on Mummy Studies, Canarian Institute of Bioanthropology and Tenerife's Archaeological Museum, Santa Cruz de Tenerife; 2018. p. 152.
[51] Roberts C, Lucy D, Manchester K. Inflammatory lesions of ribs: an analysis of the Terry Collection. Am J Phys Anthropol 1994;95:169–82. https://doi.org/10.1002/ajpa.1330950205.
[52] Theel ES, Katz SS, Pillay A. Molecular and direct detection tests for Treponema pallidum subspecies pallidum: a review of the literature, 1964–2017. Clin Infect Dis 2020;71:S4–12. https://doi.org/10.1093/cid/ciaa176.
[53] Majander K, Pfrengle S, Kocher A, Neukamm J, du Plessis L, Pla-Díaz M, et al. Ancient bacterial genomes reveal a high diversity of Treponema pallidum strains in early modern Europe. Curr Biol 2020;30:3788–803. https://doi.org/10.1016/j.cub.2020.07.058. e10.
[54] Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet 2016;48:1535–43. https://doi.org/10.1038/ng.3704.
[55] Moreno-Molina M, Shubladze N, Khurtsilava I, Avaliani Z, Bablishvili N, Torres-Puente M, et al. Genomic analyses of *Mycobacterium tuberculosis* from human lung resections reveal a high frequency of polyclonal infections. Nat Commun 2021;12. https://doi.org/10.1038/s41467-021-22705-z.
[56] Lee RS, Radomski N, Proulx J-F, Manry J, McIntosh F, Desjardins F, et al. Reemergence and amplification of tuberculosis in the Canadian arctic. JID (J Infect Dis) 2015;211:1905–14. https://doi.org/10.1093/infdis/jiv011.
[57] Bjorn-Mortensen K, Soborg B, Koch A, Ladefoged K, Merker M, Lillebaek T, et al. Tracing Mycobacterium tuberculosis transmission by whole genome sequencing in a high incidence setting: a retrospective population-based study in East Greenland. Sci Rep 2016;6. https://doi.org/10.1038/srep33180.
[58] Mulholland Cv, Shockey AC, Aung HL, Cursons RT, O'Toole RF, Gautam SS, et al. Dispersal of *Mycobacterium tuberculosis* driven by historical European trade in the South pacific. Front Microbiol 2019;10:2778. https://doi.org/10.3389/fmicb.2019.02778.
[59] Byrne AS, Goudreau A, Bissonnette N, Shamputa IC, Tahlan K. Methods for detecting mycobacterial mixed strain infections–A systematic review. Front Genet 2020;11. https://doi.org/10.3389/fgene.2020.600692.
[60] Sobkowiak B, Glynn JR, Houben RMGJ, Mallard K, Phelan JE, Guerra-Assunção JA, et al. Identifying mixed Mycobacterium tuberculosis infections from whole genome sequence data. BMC Genom 2018;19. https://doi.org/10.1186/s12864-018-4988-z.
[61] Kargarpour Kamakoli M, Farmanfarmaei G, Masoumi M, Khanipour S, Gharibzadeh S, Sola C, et al. Prediction of the hidden genotype of mixed infection strains in Iranian tuberculosis patients. Int J Infect Dis 2020;95:22–7. https://doi.org/10.1016/j.ijid.2020.03.056.
[62] Gabbassov E, Moreno-Molina M, Comas I, Libbrecht M, Chindelevitch L. SplitStrains, a tool to identify and separate mixed *Mycobacterium tuberculosis* infections from WGS data. Microb Genom 2021;7. https://doi.org/10.1099/mgen.0.000607.
[63] Cohen T, van Helden PD, Wilson D, Colijn C, McLaughlin MM, Abubakar I, et al. Mixed-strain *Mycobacterium tuberculosis* infections and the implications for tuberculosis treatment and control. Clin Microbiol Rev 2012;25:708–19. https://doi.org/10.1128/CMR.00021-12.
[64] Hutás I. A tuberkulózis járványtana. Budapest: Pulmonális És Extrapulmonális Tuberkulózis Medicina Könyvkiadó; 2007. p. 33–46.
[65] Donoghue HD, Pap I, Szikossy I, Spigelman M. The Vác mummy project: investigation of 265 eighteenth-century mummified remains from the TB pandemic era. The handbook of mummy studies. Singapore: Springer Singapore; 2021. p. 777–805. https://doi.org/10.1007/978-981-15-3354-9_21.
[66] Ebert D, Bull JJ. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? Trends Microbiol 2003;11:15–20. https://doi.org/10.1016/S0966-842X(02)00003-3.
[67] Zheng N, Whalen CC, Handel A. Modeling the potential impact of host population survival on the evolution of M. tuberculosis latency. PLoS One 2014;9. https://doi.org/10.1371/journal.pone.0105721.
[68] Barnes I, Duda A, Pybus OG, Thomas MG. Ancient urbanization predicts genetic resistance to tuberculosis. Evolution 2011;65:842–8. https://doi.org/10.1111/j.1558-5646.2010.01132.x.