The Genome Sequence of a SNP Type 3K strain of *Mycobacterium leprae* isolated from a 7th Century Hungarian case of Lepromatous Leprosy.

Tom A Mendum^a, G Michael Taylor^a, Helen D Donoghue^b, Huihai Wu^a, Csaba Szalontai^c,
Antónia Marcsik^d, Erika Molnár^d, György Pálfi^d, Graham R Stewart^a.

^a = Department of Microbial Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, GU2 7XH, UK.

^b = Centre for Clinical Microbiology, Division of Infection and Immunity, Royal Free Campus, University College London, London NW3 2PF, UK.

^c = Salisbury Ltd., 2040 Budaörs, Hungary.

^d = Department of Biological Anthropology, University of Szeged, H-6726 Szeged, Hungary.

Short running title: A strain of leprosy from 7th century Hungary.

Keywords. Ancient DNA, *Mycobacterium leprae*, phylogeny, Branch 0, whole genome sequencing.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/oa.2673

Abstract.

We report on a *Mycobacterium leprae* genome isolated from the remains of an individual with lepromatous leprosy that were excavated from a 7th century Hungarian cemetery. We determined that the genome was from a SNP type 3K0 *M. leprae* strain, a lineage that diverged early from other *M. leprae* lineages. This is one of the earliest 3K0 *M. leprae* genomes to be sequenced to date. A number of novel SNPs as well as SNPs characteristic of the 3K0 lineage were confirmed by conventional PCR and Sanger sequencing. Recovery of accompanying human DNA from the burial was poor, particularly when compared to that of the pathogen. Modern 3K0 *M. leprae* strains have only been isolated from East Asia and the Pacific and so these findings require new scenarios to describe the origins and routes of dissemination of leprosy during antiquity that have resulted in the modern phylogeographical distribution of *M. leprae*.

Introduction.

The spread of leprosy around the world in antiquity has previously been inferred by characterising the infrequent single nucleotide polymorphisms (SNPs) of modern strains of *M. leprae*, collected from across the world (Monot *et al*, 2005). More recently, the genotyping of isolates recovered from archaeological human remains displaying skeletal evidence of lepromatous leprosy (LL) has provided unique opportunities to refine our understanding of the distribution of *M. leprae* strains seen in extant populations, as well as allowing analysis of strains from regions where the disease is no longer present. In this respect, we have previously studied archaeological cases from various parts of Britain and Europe (Taylor *et al*, 2013; Inskip *et al*, 2015; Donoghue *et al*, 2015) as well as modern cases from regions where the disease is still common (Monot *et al*, 2009).

One of the cases identified previously was burial KD271, a seventh century male individual whose remains were excavated from Kiskundorozsma (Szeged) in Hungary (Pálfi and Molnár, 2009). Study of a limited number of informative SNPs indicated that the strain of *M. leprae* present in KD271 was a type 3K on the original scheme of 16 genotypes (ranging from 1A – 4P) proposed by Monot and colleagues (2009).

With the application of whole genome sequencing (WGS) to both modern and ancient cases of leprosy (Schuenemann *et al*, 2013; Mendum *et al*, 2014), it has become clear that SNP type 3K isolates belong on a separate lineage of the *M. leprae* phylogenetic tree, now designated Branch 0. Although a number of modern examples of this lineage have been studied, all from East Asian locations (Schuenemann *et al*, 2013; Benjak *et al*, 2018), no ancient isolates have previously been available for such genomic analysis. The current study presents WGS analysis of the strain from burial KD271 and provides a rare opportunity to study what is one of the oldest cases of *M. leprae* to be sequenced. This is likely to represent a strain close to the most recent common ancestor (MRCA) of extant strains that is predicted to have existed only 3-4,000 years ago (Schuenemann *et al*, 2013). As previous biomolecular tests applied to KD271 have appeared in diverse sources, a summary of the conventional genotyping is brought together here, along with some novel observations. These serve both to provide context and as a useful comparator for the WGS study.

Materials and Methods.

Sampling.

The subject of the current study, burial KD271, was a mature male individual judged to be between 50 and 60 years of age at the time of death. The burial was excavated from the cemetery of Kiskundorozsma Szeged in Hungary and the diagnosis of leprosy made on osteological grounds (Pálfi, 1991; Pálfi and Molnár, 2009). For the present genomic study, a 50 mg sample of bone from the palate of KD271 was ground into a fine powder in a sterilized pestle and mortar.

A second burial, Sk12, an adult male, without osteological signs of leprosy, acted as a control for the WGS protocol. Bone was sampled from around the vomer region (Schuenemann *et al*, 2013).

DNA extraction.

DNA was extracted using an in-house version of the Boom method (Boom *et al*, 1990). In this procedure, bone powder (50 mg) was incubated in 1 ml of 1 × Tris-EDTA buffer, containing 40 mAU/ml of proteinase K at 37 °C for 48 h with occasional mixing. The sample was then centrifuged at 3,000 rpm for 3 mins and the supernatant transferred into 5 volumes of 6 M guanidinium thiocyanate (GUSCN, product G9020 from US Biologicals, Salem, MA.) containing 1 % Triton X-100 and buffered in 1 × Tris-EDTA buffer adjusted to pH 6.5 with 3M sodium acetate, pH 5.5. Bone powder was mixed with the GUSCN buffer on a mixing wheel for 1 h at 4 °C. The samples were then subjected to 3 freeze-thaw cycles. Bone powder was removed by centrifugation at 12,000 rpm for 5 mins and the supernatant transferred to a sterile 1.5 ml Eppendorf tube. Pre-washed silica suspension (40 µl of 0.5-10 µm, Sigma-Aldrich, S5631) was added and kept in contact for 3 h to maximize recovery of fragmented DNA. After centrifugation, silica was further washed twice with 1 ml aliquots of GUSCN extraction buffer, followed by 3 washes with 75 % ethanol and finally with 1 ml of acetone. After thorough drying of the silica pellet, DNA residues were eluted in 60 µl HPLC grade water at 55 °C. This was then sub-divided into 2 × 30 µl aliquots and stored in low retention plastic tubes to minimize loss of DNA through repeated freeze-thawing events.

M. leprae screening.

Before undertaking WGS, we screened for evidence of *M. leprae* DNA in the new extract using a PCR for the RLEP multi-copy element. This method amplifies a 78 bp amplicon with product monitored with a specific FAM-labeled hybridization probe. Details of this and the primer sequences and conditions have been previously reported (Inskip *et al*, 2015).

VNTR and SNP genotyping.

The PCR methods for the VNTR and individual SNP loci amplification and sequencing have been previously reported (Taylor *et al*, 2009; Taylor and Donoghue, 2011). The *rpoT* locus (ML1022) was PCR amplified and sequenced for the present study using the primers and conditions described in Taylor *et al*, 2009.

M. lepromatosis screening.

The opportunity was taken to screen the new extract for any evidence of the second leprosy agent, *M. lepromatosis*. The primers were modifications of those described by Singh *et al*, 2015, 5'-CTGTTCGTGAGGTACCGGTGAAA and 5'-GTTCGGCCGGAGTGTAGGTGTTA. These amplify a 135 bp fragment from the *hemN* gene, present in *M. lepromatosis* but absent in *M. leprae*. The PCR reagents and conditions

were as described previously for *M. leprae* specific primers (Inskip *et al*, 2015), except that an annealing temperature of 56 °C was used.

Screening of extract for *Mycobacterium tuberculosis* complex DNA.

The KD271 extract was also tested for the presence of *Mycobacterium tuberculosis* (MTB) complex organisms using a real-time PCR method for the IS1081 repetitive element (Taylor *et al*, 2007).

Conventional PCR amplification and sequencing of newly identified SNPs in the KD271 isolate of *M. leprae*.

Scrutiny of the WGS data from KD271 (Table 3) revealed sixteen novel polymorphic SNPs not found in other sequenced strains. Conventional PCR amplification and sequencing was undertaken to confirm the WGS findings for three randomly selected SNPs. Sequences of primers and amplicon sizes are given in Supplementary Table S1.

Human DNA.

An attempt was made to confirm the sex of individual KD271 using a PCR method based on a polymorphism in the amelogenin gene (Mannucci *et al*, 1994). This generates two bands from males of 106 bp and 112 bp (AMELX and AMELY products respectively), and a single AMELX product of 106 bp from females.

Whole genome sequencing of *M. leprae* genome from skeleton KD271.

DNA from skeleton KD271, was enriched for *M. leprae* sequences using microarrays and sequenced as described previously for cases Sk2, Sk8 and the control Sk12 from medieval Winchester (Schuenemann *et al*, 2013), but with only a single round of micro-array based enrichment. The raw sequence files were deposited in the Sequence Read Archive (SRA) database with Submission ID SAMN08093649. After initial quality controls and alignment to the *M. leprae* TN reference genome (Schuenemann *et al*, 2013), SNPs were identified and accepted if they had a read depth of 3 or greater; more than 75 % of reads agreed; an alignment quality score (MQ) ≥ 30 and an absence of reads in the control sample Sk12.

Phylogenies were generated by aligning all SNPs from selected leprosy genomes (Supplementary Excel data file 1) using both the Maximum Likelihood and Neighbour-Joining functions of MEGA7 (Kumar *et al*, 2016) and the Tamura 3-Parameter model (as determined to be the best fit by the Model Selection function of MEGA7). Equivalent loci in *M. lepromatosis* were determined by aligning *M. leprae* TN, *M. leprae* Br423 and *M. lepromatosis* FJ924 (Han *et al*, 2015) with Mauve v2.4.0 (Darling *et al*, 2004).

Results.

Paleopathological lesions in KD271.

The diagnosis of LL in individual KD271 was originally made on palaeopathological grounds. The lesions have been described in detail previously (Donoghue *et al*, 2005; Molnár *et al*, 2006), so only a summary is given here. The skull shows signs of the rhino-maxillary syndrome, with erosion and widening of the nasal margins and resorption of the anterior nasal spine (Figure 1 and inset). There is loss of bone from around the alveolar region on the maxillary process and the upper incisor teeth have been lost *ante mortem* with remodelling of the tooth sockets (Figure 1). There is also pitting on the nasal surface of the palate due to the disease. Pitting due to periostitis is also present on some tarsal bones and on the surfaces of the tibiae and fibulae. Subperiosteal exostoses seen on the fibulae are more evident on the distal third of the bone shafts (Figure 2).

Burial Sk12. This individual was found to be free of any macroscopic signs of the rhino-maxillary syndrome, or other indications of leprosy on the hands, feet or distal lower limbs (Taylor *et al*, 2013).

Screening of KD271 extract for *M. leprae* DNA.

The extract prepared from the palatal region of the skull tested positive using the RLEP PCR probe method (Figure 3). The control case was negative for *M. leprae* DNA (not shown).

SNP genotyping and multiple loci variable number tandem repeat (VNTR) analysis.

The results of SNP and VNTR typing are summarised in Table 1. Supplementary Figure S1 shows conventional Sanger sequencing of 6 informative SNP loci used in the Monot typing scheme (2009).

Screening for *M. lepromatosis* DNA.

No PCR products were amplified from the extracts using the *hemN* primers specific for *M. lepromatosis*, therefore there is no evidence that this individual was co-infected with this pathogen. To show that this PCR method can amplify *M. lepromatosis*, an appropriately sized amplicon (135 bp) was PCR amplified from partially-purified *M. lepromatosis* DNA harvested from infected tissue of a Scottish red squirrel (*Sciurus vulgaris*) (Supplementary Data, Figure. S2).

Screening for MTB complex DNA.

The extract prepared from the palate of KD271 was negative for evidence of *M. tuberculosis* complex DNA.

Human DNA.

No PCR products were obtained using the amelogenin PCR, which probably reflects the extremely fragmented nature of DNA in this skeleton.

Genome sequencing of KD271.

DNA extracts from skeletons KD217 and Sk12 were enriched for *M. leprae* sequences, PCR amplified and sequenced. Details of the read depths, their percentage alignment to the *M. leprae* TN reference genome, percentage genome coverage and average read length are given in Table 2. Samples from the KD271 burial gave an average read depth of 11.94, while the Sk12 control sample's average read depth was 0.39 % and less than 1 % of reads aligned to the *M. leprae* genome. As observed previously (Schuenemann *et al*, 2013; Mendum *et al*, 2014), the alignments were punctuated with regions with pan-genus similarity that had large numbers of highly heterogeneous reads that presumably derive from environmental mycobacteria. The KD271 sequences had short read lengths (Sawyer *et al*, 2012) typical of ancient DNA. These data confirm that there was little or no cross-contamination between samples, or from modern *M. leprae* DNA during the sample preparation.

The WGS sequencing was in broad agreement with Sanger sequencing of targeted SNPs (Table 1), so validating both conventional typing and genome sequencing methods. However, the WGS had insufficient coverage for 3 of the 9 Sanger determined SNP loci. Analysis of VNTR genotypes in KD271 was limited to the 21-3 locus (*ML0058*) and to the *rpoT* locus (*ML1022*), as insufficient numbers of reads spanned the necessary regions for the other two VNTR loci, GTA(9) and AGA(20). Sixteen novel SNPs were identified that have not been found in other published strains, including the 3K0 strains, S9, S10, CM-1 or Kyoto-1. None

of the SNPs in coding regions are likely to have phenotypic effects. To validate the WGS data, three randomly chosen, newly identified SNPs were additionally confirmed by conventional PCR and Sanger sequencing (Table 3).

Phylogenetic analysis of KD271

Phylogenetic analyses were consistent between both Maximum Likelihood and Neighbour-joining methods (Figure. 4 and Supplementary Figure S3). Both placed KD271 within the 3K0 lineage, branching from the common lineage before the S10/Kyoto-1 branch but after the S9/CM-1 branch (Schuenemann *et al*, 2013; Mendum *et al*, 2014; Avanzi *et al*, 2015, Honap *et al*, 2018). The branch lengths of KD271 were shorter than for the modern strains as would be expected for an ancient genome. KD271 was found to exhibit 26 SNPs, 24 of which were unique to the 3K0 group (S9, S10, CM-1 and Kyoto-1), six of these SNPs were uniquely shared with S10 and Kyoto-1. In contrast, no SNPs were uniquely shared with S9 or CM-1.

rpoT locus.

The presence of only three copies of GACATC in the *rpoT* tandem repeat distinguishes KD271, as well as S9 and S10, all of which have earlier branch points, from the modern isolates from Japan, Korea and parts of China that result from an apparently recent radiation of 3K0 strains and all contain 4 copies of the hexanucleotide (Kai *et al*, 2013; Weng *et al*, 2013).

ML0411 locus.

Polymorphism was noted in *ML0411*. Locus *ML0411* is a serine rich, 45kDa antigen (408 aa) recognised by B cells of the immune system. It is a member of the PPE protein gene families of pathogenic mycobacteria, having the characteristic Pro-Pro-Glu motif at the N termini of the proteins. These proteins, along with the related PE family, are likely to give rise to antigenic variation and may modulate the immune response. *ML0411* is the single most variable gene in the genome of *M. leprae*. Comparison of the KD271 polymorphic SNPs in *ML0411* (Table 4) with 3K0 strains from NE Asia (Kai *et al*, 2013) are compatible with its phylogenetic position as an early member of the 3K0 lineage.

Subdivision of the Branch 0 lineage.

Recent work shows the type 3K strains of Branch 0 may be further sub-divided into 3K0 and 3K1 lineages, depending on specific SNP subsets (Avanzi *et al*, 2015). These workers report that the 3K0 lineage displays a total of 20 specific SNPs and the newly described branch 3K1 demonstrates 23 specific SNPs. The 3K common branch is characterised by C at position 711,197 and G at SNPs 563,796 and 57,633. Inspection of the data from KD271 shows the isolate is consistent with 3K0 rather than 3K1 lineage of Branch 0 (Table 5).

Discussion.

The subject of the current report, KD271, was a mature male individual showing skeletal evidence of lepromatous leprosy. The remains were excavated from the cemetery of Szeged-Kiskundorozsma-Daruhalom dűlő II, Hungary in the summer of 2003. Based on the associated grave goods (Paluch and Szalontai, 2004; Mészáros *et al*, 2005; Szalontai, 2012), the burial ground was in use in the early/middle Avar transition period, in the late 7th century. The cemetery contained 94 individuals and skeletal signs of leprosy were noted in eight of these. This presumptive diagnosis was previously confirmed by ancient DNA (aDNA) analysis in three cases, including the subject of the current work, KD271 (Molnár *et al*,

2006). Subsequently, this case was partially genotyped for a later study of leprosy using conventional PCR and Sanger sequencing targeting phylogenetically informative SNP loci (Monot *et al*, 2009). This showed the isolate to be a type 3K strain of *M. leprae*.

Since those initial investigation, the understanding of the phylogeography of leprosy has improved considerably, with the recognition of an additional lineage (Branch 0) into which the 3K strains may be placed (Schuenemann *et al*, 2013) and which branches earlier from the common lineage than the other groups. To extend our understanding of how KD271 and the other leprosy remains at Szeged-Kiskundorozsma-Daruhalom dűlő II relate to other ancient and modern *M. leprae* strains, the genome of KD271 was sequenced. This genome is, to date, one of the oldest Branch 0 *M. leprae* genomes to have been studied and so is likely to represent a strain more similar to the notional MRCA (predicted to have existed ~3,000-4,000 years ago). From recent and ongoing studies on modern Branch 0 isolates, it is evident that there is a greater heterogeneity within this lineage than first realized, so that the type 3K strains may be further sub-divided into 3K0 and 3K1, depending on specific SNP subsets (Avanzi *et al*, 2015). The genome sequencing, phylogenetic analysis and SNP specific sequencing all confirm that KD271 falls within the currently known monophyletic *M. leprae* radiation and does indeed represent a 3K strain, being placed on an early branch of the 3K0 lineage.

This position is reflected in the SNPs of the hypervariable gene, *ML0411*, which, although not identical to modern strains, was found to be consistent with its place on the 3K0 lineage.

The isolate affecting KD271, along with S9 and S10, the two other strains with early branch points in the 3K0 lineage, possessed only 3 copies of the *rpoT* hexanucleotide GACATC, rather than the 4 copies typical of modern strains as described by Avanzi and colleagues (2015) and Kai *et al*, 2013. When first reported, this VNTR polymorphism was suggested as one, albeit limited, way of examining differences in the *M. leprae* genome (Matsuoka *et al*, 2000). The locus may therefore be helpful in defining those strains mentioned above with a more recent radiation. However, the relevance of the *rpoT* locus is difficult to assess without further investigation of strains isolated from various regions of India (Lavana *et al*, 2007; Lavana *et al*, 2009) where strains with 4 copies have been described. This is because the recently described second agent of leprosy, *Mycobacterium lepromatosis* (Han *et al*, 2008), also contains 4 copies of this sequence, as reported for cases in Western and Central Mexico (Han *et al*, 2012). So in studies where the *rpoT* primers recognize both pathogens, and where it was the only typing method used (Lavana *et al*, 2007; Lavana *et al*, 2009), further testing will be needed to distinguish between *M. lepromatosis* and the 3K lineage of *M. leprae* with the 4 copies of the *rpoT* hexanucleotide.

In the earlier study by Monot *et al*, (2009), two further LL cases from another Hungarian burial ground were also found to be positive for *M. leprae*. These were burials 222 and 503, a 45-50 year old male and a 30-35 year old female respectively. These were both excavated from the 10-11th century Eastern Hungarian cemetery at Püspökladány-Eperjesvölgy. They were found to be SNP types 3K and 3M respectively. Interestingly, they were both subsequently shown to be co-infected with *M. tuberculosis* (Donoghue *et al*, 2005) hence we checked burial KD271 for any signs of MTB complex DNA, but found none. The lack of either MTB complex or *M. lepromatosis* mycobacterial DNA is important for the accurate interpretation of WGS data.

The presence of 3K0 strains in 7th century Hungarian remains and of a 3K strain from the 10th century, is consistent with two contrasting scenarios for the origins of geographical

distribution of 3K *M. leprae* strains. The global distribution of 3K0 and 3K1 strains is today restricted to regions of the Western Pacific such as Japan (except Okinawa), Korea, China, The Philippines, New Caledonia and Indonesia amongst others (Kai *et al*, 2013; Avanzi *et al*, 2015; Monot *et al*, 2009; Weng *et al*, 2013; Honap *et al*, 2018). This could indicate that the 3K lineage originated in Northern or Eastern Asia. The presence of two type 3K cases (KD271 and 222) in early medieval Hungary would then suggest a route of dissemination from Asia to central Europe, perhaps via trade links or migrations. This would be consistent with what is known of the origins of the Panonian Avars, who are believed to have reached the Hungarian plain from the Eurasian steppe in the late 6th to early 9th centuries (Curta, 2006). The other possibility is that Europe was a centre of dissemination of the ancestral 3K0 and related strains, some of which later became less common or even absent from Europe but persisted in East Asia and the Pacific. Determining the likelihood of each of these scenarios will require more sampling and characterisation of both ancient and modern strains.

Conclusions.

Ancient DNA analysis is a powerful approach for understanding past human diseases like leprosy. In particular, it allows us to obtain strain typing data from geographical locations where the disease may no longer be found, to compare ancient with modern strain distributions and to assist with understanding earlier human migrations. Recognition of greater diversity within 3K lineage strains (Branch 0) and its recognition as the deepest lineage, has come from WGS studies applied to both ancient and modern cases. Burial KD271 represents one of the earliest examples of this archaic lineage to be studied to date.

Acknowledgements.

We thank Dr. Charlotte Avanzi of The Global Health Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland for the generous gift of *M. lepromatosis* genomic DNA, partially purified from red squirrel tissue. We also thank Dr. Tanvi Honap of the Department of Anthropology, University of Oklahoma for making available at short notice her sequence variant data. This study was supported in part by a Small Research Grant, reference SG150114, from the British Academy.

Supporting Information.

Supporting Figure S1. Sanger sequencing of informative SNP loci in KD271

Sanger sequence reads used to determine 3K subgenotype in burial KD271. Panel a) SNP 4, nt 413,903. Panel b) SNP 5, nt 591,858. Panel c) SNP 6, nt 1,133,495. Panel d) SNP 7, nt 2,312,066. Panel e) SNP 8, nt 7,614. Panel f) SNP 9, nt 1,113,926. Nt = nucleotide positions in the re-annotated version of the Tamil Nadu (TN) genome. Polymorphic loci are marked on each panel with a yellow highlight.

Supporting Figure S2. Q-PCRs showing amplification of *M. lepromatosis* DNA with *hemN* primers.

M. lepromatosis specific *hemN* primers were used to PCR amplify a 135 bp region of the *hemN* gene from *M. lepromatosis* genomic DNA. a) Amplification profile. b) Dissociation melt analysis. Reactions with template are in pink and with no template in black.

Supporting Figure S3. Phylogenetic relationships between modern and ancient (bold text) *M. leprae* strains.

The phylogeny was inferred by the Neighbour-Joining method of MEGA7 (Kumar et al, 2016) and the Tamura 3-parameter model. Bootstrap percentages from 1000 replicates are shown next to the branches. The scale indicates the number of substitutions per site. All positions with less than 90% site coverage were eliminated. *M. lepromatosis* was used as an outgroup (not shown).

Supporting Table S1. Primers used to amplify and confirm novel SNPs from the *M. leprae* isolate in KD271.

Supporting Data File S1 (Excel). An alignment of *M. leprae* SNPs used for phylogenetic analysis. Loci are numbered with reference to the genome of strain TN. Base calls that cannot be determined are listed as 'N'. Invariant sites and sites with reads in control samples are not included.

References.

- Avanzi C, Benjak A, Kai M, Busso P, Singh P, Matsuoka M, Cole ST. 2015. Analysis of *Mycobacterium leprae* strains from Japan: new trends in phylogeography. Eposter EP009 at 25th ECCMID in Copenhagen, Denmark, 25-28th April.
- Benjak A, Avanzi C, Singh P, Loiseau C, Girma S, Busso P, Fontes ANB, Miyamoto Y, Namisato M, Bobosha K, Salgado CG, da Silva M, Bouth RC, Frade MAC, Filho FB, Barreto JG, Nery JAC, Bühner-Sékula S, Lupien A, Al-Samie AR, Al-Qubati Y, Alkubati AS, Bretzel G, Vera-Cabrera L, Sakho F, Johnson CR, Kodio M Fomba A, Sow SO, Gado M, Konaté O, Stefani MMA, Penna GO, Suffys PN, Sarno EN, Moraes MO, Rosa PS22, Baptista IMFD, Spencer JS, Aseffa A, Matsuoka M, Kai M, Cole ST. 2018. Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. *Nature Communication* **9(1)**:352. doi: 10.1038/s41467-017-02576-z
- Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, van der Noordaa J. 1990. Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**: 495-503.
- Curta F. 2006. *Southeastern Europe in the Middle Ages, 500–1250*. Cambridge medieval textbooks. Cambridge: Cambridge University Press. ISBN 978-0-521-81539-0.
- Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* **14**:1394-1403. Doi: 10.1101/gr.2289704.
- Donoghue HD, Marcsik A, Molnár E, Paluch T, Szalontai C. 2005. Lepra nyomai a kiskundorozsmai avar temetőből Előzetes beszámoló. In Hadak útján” – Népeség és iparok a népvándorlás korában, 171-186. [Osteological signs of leprosy from the Avar period series of Kiskundorozsma. Preliminary report. In: In the way of the armies– Population and industries in the Migration Period] pp 171-186.
- Donoghue HD, Marcsik A, Matheson C, Vernon K, Nuorala E, Molto JE, Greenblatt CE, Spigelman M. 2005. Co-infection of *Mycobacterium tuberculosis* and *Mycobacterium leprae*

in human archaeological samples: a possible explanation for the historical decline of leprosy. *Proceedings of the Royal Society B* **272**: 389-394. Doi: 10.1098/rspb.2004.2

Donoghue HD, Taylor GM, Marcsik A, Molnár E, Palfi G, *et al.* 2015. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. *Infection, Genetics and Evolution* **31**:250-256. Doi: 10.1016/j.meegid.2015.02.001.

Han XY, Seo Y-H, Sizer KC, Schoberle T, May GS, *et al.* 2008. A new Mycobacterium species causing diffuse lepromatous leprosy. *American Journal of Clinical Pathology*. **130**: 856–864. Doi: 10.1309/AJCPP72FJZZRRVMM.

Han XY, Sizer KC, Velarde-Félix JS, Frias-Castro LO, Vargas-Ocampo F. 2012. The Leprosy Agents *Mycobacterium lepromatosis* and *Mycobacterium leprae* in Mexico. *International Journal of Dermatology* 51: 952–959. Doi: 10.1111/j.1365-4632.2011.05414.x.

Han XY, Mistry NA, Thompson EJ, Tang H-L, Khanna K, Zhang L. 2015. Draft genome sequence of new leprosy agent “*Mycobacterium lepromatosis*”. *Genome Announcements* **3**: e00513-15. Doi: 10.1128/genomeA.00513-15

Honap TP, Pfister L-A, Housman G, Mills S, Tarara RP, Suzuki K, Cuozzo FP, Sauter ML, Rosenberg MS, Stone AC. 2018. *Mycobacterium leprae* genomes from naturally infected nonhuman primates. *PLoS Neglected Tropical Diseases* 12(1): e0006190. doi.org/10.1371/journal.pntd.0006190.

Inskip SA, Taylor GM, Zakrewski S, Mays SA, Pike AW, Llewellyn G, Williams CM, Lee OY, Wu HH, Minnikin DE, Besra G. Stewart GR. 2015. Osteological, biomolecular and geochemical analysis of an early Anglo-Saxon case of lepromatous leprosy. *PLoS ONE* **13**:10 e0124282. Doi: 10.1371/journal.pone.0124282.

Kai M, Nakata N, Matsuoka M, Sekizuka T, Kuroda M, Makino M. 2013. Characteristic mutations found in the *ML0411* gene of *Mycobacterium leprae* isolated in Northeast Asian countries. *Infection, Genetics and Evolution* **19**: 200-204. Doi: 10.1016/j.meegid.2013.07.014.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution* **33**: 1870-1874. Doi: 10.1093/molbev/msw054. Epub 2016 Mar 22.

Lavania M, Katoch K, Singh H, Das R, Gupta AK, Sharma R, Chauhan DS, Sharma VD, Sachan P, Sachan SP, Katoch VM. 2007. Predominance of three copies of tandem repeats in *rpoT* gene of *Mycobacterium leprae* from Northern India. *Infection, Genetics and Evolution* **7**: 627–631. Doi: 10.1016/j.meegid.2007.05.011.

Lavania M, Lal R, Joseph G, Darlong J, Abraham S, Nanda NK, Jadhav RS. 2009. Genotypic analysis of *Mycobacterium leprae* strains from different regions of India on the basis of *rpoT*. *Indian Journal of Leprosy* **81**: 119-124.

Mannucci A, Sullivan KM, Ivanov PL, Gill P. 1994. Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. *International Journal of Legal Medicine* **106**:190–193. Doi.org/10.1007/BF01371335

Matsuoka M, Maeda S, Kai M, Nakata N, Chae GT, Gillis TP, Kobayashi K, Izumi S, Kashiwabara Y. 2000. *Mycobacterium leprae* typing by genomic diversity and global distribution of genotypes. *International Journal of Leprosy* **68**: 121–128.

Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, Tucker K, Hinds J, Cole ST, Kierzek AM, Nieselt K, Krause J, Stewart GR. 2014. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. *BMC Genomics* **15**: 270. Doi: 10.1186/1471-2164-15-270.

Mészáros P, Paluch T, Szalontai CS. 2005. Avar kori temetők Kiskundorozsma határában. Előzetes beszámoló az M5 autópályán feltárt lelőhelyről. – MKCsM (2004) Szeged, pp. 145–162.

Molnár E, Marcsik A, Bereczki Z, Donoghue HD. 2006. Pathological cases from the 7th century in Hungary. 16th European Meeting of the Paleopathology Association, Fira, Santorini, Greece.

Monot M, Honoré N, Garnier T, Araoz R, Coppe'e J -Y, Lacroix C, Sow S, Spencer JS, Truman RW, Williams DL, Gelber R, Virmond M, Flageul B, Cho SN, Ji B, Paniz-Mondolfi A, Convit J, Young S, Fine PE, Rasolofo V, Brennan PJ, Cole ST. 2005. On the origin of leprosy. *Science* **308**: 1040–1042. Doi: 10.1038/ng.477.

Monot M, Honoré N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, Matsuoka M, Taylor GM, Donoghue HD, Bouwman A, Mays S, *et al.* 2009. Phylogeography of Leprosy. *Nature Genetics* **41**: 1282-1289. Doi: 10.1126/science/1109759.

Pálfi G, Molnár E. 2009. The paleopathology of specific infectious diseases from Southeastern Hungary: a brief overview. *Acta Biologica Szegediensis* **53**: 111-116. (<http://www.sci.u-szeged.hu/ABS>)

Paluch T, Szalontai CS. 2004. Kiskundorozsma-Daruhalom dűlő II. In Kisfaludy I (ed) Régészeti kutatások Magyarországon 2003. [Archaeological Investigations in Hungary, 2003], 293-329.

Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS ONE* **7**(3): e34131. Doi: 10.1371/journal.pone.0034131.

Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, Herbig A, Economou C, Benjak A, Busso P, Nebel A, Boldsen JL, Kjellström A, Wu H, Stewart GR, Taylor GM, Bauer P, Lee OY, Wu HH, Minnikin DE, Besra GS, Tucker K, Roffey S, Sow SO, Cole ST, Nieselt K, Krause J. 2013. Genome-wide comparison of Medieval and modern *Mycobacterium leprae*. *Science* **341**: 179-183. Doi: 10.1126/science.1238286.

Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, Nieselt K, Krause J, Vera-Cabrera L, Cole ST. 2015. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proceedings of the National Academy of Science U S A*. **112**: 4459-4464. Doi: 10.1073/pnas.1421504112.

Szalontai C 2012. Ismét az avar kori lepráról. — Again, about the leprosy in the Avar period. In Pető Zs. (ed) „Hadak Útján”. A Népvándorláskor Fiatal kutatóinak XX. konferenciája. Budapest pp. 149–161.

Taylor GM, Murphy E, Hopkins R, Rutland PC, Chistov Y. 2007. First report of *Mycobacterium bovis* DNA in archaeological human remains. *Microbiology* **153**: 1243-1249. Doi: 10.1099/mic.0.2006/002154-0.

Taylor GM, Blau S, Mays SA, Monot M, Lee OY-C, Minnikin DE, Besra GS, Cole ST, Rutland PC. 2009. *Mycobacterium leprae* Genotype Amplified from an Archaeological Case of Lepromatous Leprosy in Central Asia. *Journal of Archaeological Science* **36**: 2408-2414. Doi: 10.1016/j.jas.2009.06.026.

Taylor GM, Donoghue HD. 2011. Variable nucleotide tandem repeat (VNTR) typing of *Mycobacterium leprae* isolates amplified from European archaeological human remains with lepromatous leprosy. *Microbes and Infection* **13**: 923-929. Doi: 10.1016/j.micinf.2011.05.003.

Taylor GM, Tucker K, Butler R, Pike AWG, Lewis J, Roffey S, Marder P, Lee O-Y, Wu HH, Minnikin DE, Besra GS, Singh P, Cole ST, Stewart GR. 2013. Detection and Strain Typing of Ancient *Mycobacterium leprae* from a Medieval Leprosy Hospital. *PLoS ONE* **8**: e62406. Doi: 10.1371/journal.pone.0062406.

Weng X, Xing Y, Liu J, Wang Y, Ning Y, Li M, Wu W, Zhang L, Li W, Heiden JV. 2013. Molecular, ethno-spatial epidemiology of leprosy in China: Novel insights for tracing leprosy in endemic and non-endemic provinces. *Infection, Genetics and Evolution* **14**: 361–368. Doi.org/10.1016/j.meegid.2012.12.009.

Figure 1. The skull of KD271.

Frontal view of skull from the burial (less mandible) showing evidence of rhino-maxillary syndrome, notably loss of several incisors, resorption of the anterior nasal spine and widening of the piriform aperture. Inset highlights loss of anterior nasal spine.

Figure 2. The fibulae from KD271.

KD271 fibulae showing subperiosteal exostoses on the distal third of the bone shafts.

Figure 3. Confirmation of *M. leprae* DNA in extracts from KD271.

Real-time amplification of RLEP PCR product (78 bp) monitored with a dual-labelled fluorescent hybridization probe. KD271 is shown in blue and negative controls in grey.

Figure 4. Phylogenetic relationships between selected modern (regular text) and ancient (bold text) *M. leprae* strains.

The phylogeny was inferred by the Maximum Likelihood method of MEGA7 (Kumar *et al*, 2016) and the Tamura-3-Parameter model. The tree with the highest log likelihood value is shown. Bootstrap percentages from 1000 replicates are shown next to the branches. The scale indicates the number of substitutions per site. All positions with less than 90% site coverage were eliminated. *M. lepromatosis* was used as an outgroup (not shown). CM1 and Br15-1 are derived from a cynomolgus macaque and a red squirrel respectively.

Table 1.

SNP Locus ¹	Amplicon Size (bp)	Nucleotide base.	SNP typing inference.	Nucleotide base by WGS.
14,676	136	C	Type 3	nd
1,642,879	122	T		T
2,935,693	107	C		nd
413,903	120	G	3I-3K	G
591,858	107	C	3I-3L	C
1,133,495	121	G	3J-3M	G
2,312,066	120	G	3K-3M	G
7,614	109	C	Not 3I	nd
1,113,926	117	A		A
Overall			3K	
VNTR loci		Copies		Copies by WGS
AGA(20) 2,785,364- 2,785,494.	Variable	16		nd
GTA(9) 2,583,816- 2,583,839.	Variable	24		nd
21-3, <i>ML0058</i> 72,683-73,686	96	2		2
<i>rpoT</i> , (<i>sigA</i>) <i>ML1022</i>	91	3		3

Table 2.

Total number of reads	Percentage of reads aligning to the <i>M. leprae</i> TN genome	Percentage of reads aligning to the human genome	Average <i>M. leprae</i> read depth	Average length of aligned reads	Percentage Genome coverage ≥ 3 reads, MQ score ≥ 30
52,430,302	17.1%	0.9%	11.9	88.6 bp	83.8 %

Table 3.

Position (TN)	Loci	Base (TN/KD271)	Amino acid (TN/KD271)	Function
18136	ML0014	C/T	A/A	Probable conserved membrane protein (pseudogene) ¹
19223	ML0016	C/T	L/L	<i>pknB</i>
250837	ML0185	C/T	T/I	Probable conserved transmembrane protein
2127288	Intergenic	C/T		
2223039	Intergenic	C/T		
2235810	ML1852	C/T	V/M	Probable conserved transmembrane protein (pseudogene) ¹
2297562	ML1914	T/A	I/N	<i>lpqN</i>
2564872	Intergenic	T/C		
2731405	Intergenic	G/A		
2881087	ML2410	G/A	A/A	Probable conserved transmembrane protein
3021965	ML2536	C/T	V/V	ESX conserved component <i>eccB3</i> , possible membrane protein.
3108100	ML2603	G/T	P/P	Possible lysophospholipase ¹
3158822	ML2639	G/A	L/L	Probable aldehyde dehydrogenase, (NAD ⁺) dependent
3256354	ML2700	G/A	A/T	Probable conserved transmembrane protein
3261877	Intergenic	G/A		
3268026	Intergenic	G/A		

Table 4.

<i>ML0411</i> position (Kai <i>et al</i> , 2013)	TN loci	TN	S9	S10	3K0 strains eg Kyoto-1 (Kai <i>et al</i> , 2013)	KD271 base
-41st	508714	G	G	G	T	No coverage
276th	509030	C	C	A	A	A
424th	509178	G	A	A	A	A
571st	509325	C	G	G	G	G
657th	509411	C	C	C	C	C
671st	509425	C	C	T	T	C
1097th	509851	G	T	T	T	T
1153rd	509907	G	G	G	G	No coverage

Table 5.

Genome	ML0585c	ML0466	ML0046c
	Pseudogene <i>gor</i>	Hypothetical protein	Pseudogene <i>espJ</i>
	SNP 711,197	SNP 563,796	SNP 57,633
3K1 branch			
Ryukyu-2	C	G	G
3K0 branch			
S9	C	C	T
S10	C	C	T
CM-1	C	C	T
KD271	C	C	T
Other strains			
TN	T	G	T
Br 4923	T	G	T