



OPEN

# A treponemal genome from an historic plague victim supports a recent emergence of yaws and its presence in 15<sup>th</sup> century Europe

Karen Giffin<sup>1,5</sup>, Aditya Kumar Lankapalli<sup>1,5</sup>, Susanna Sabin<sup>1</sup>, Maria A. Spyrou<sup>1</sup>, Cosimo Posth<sup>1,4</sup>, Justina Kozakaitė<sup>2</sup>, Ronny Friedrich<sup>3</sup>, Žydrūnė Miliauskienė<sup>2</sup>, Rimantas Jankauskas<sup>2</sup>, Alexander Herbig<sup>1</sup>✉ & Kirsten I. Bos<sup>1</sup>✉

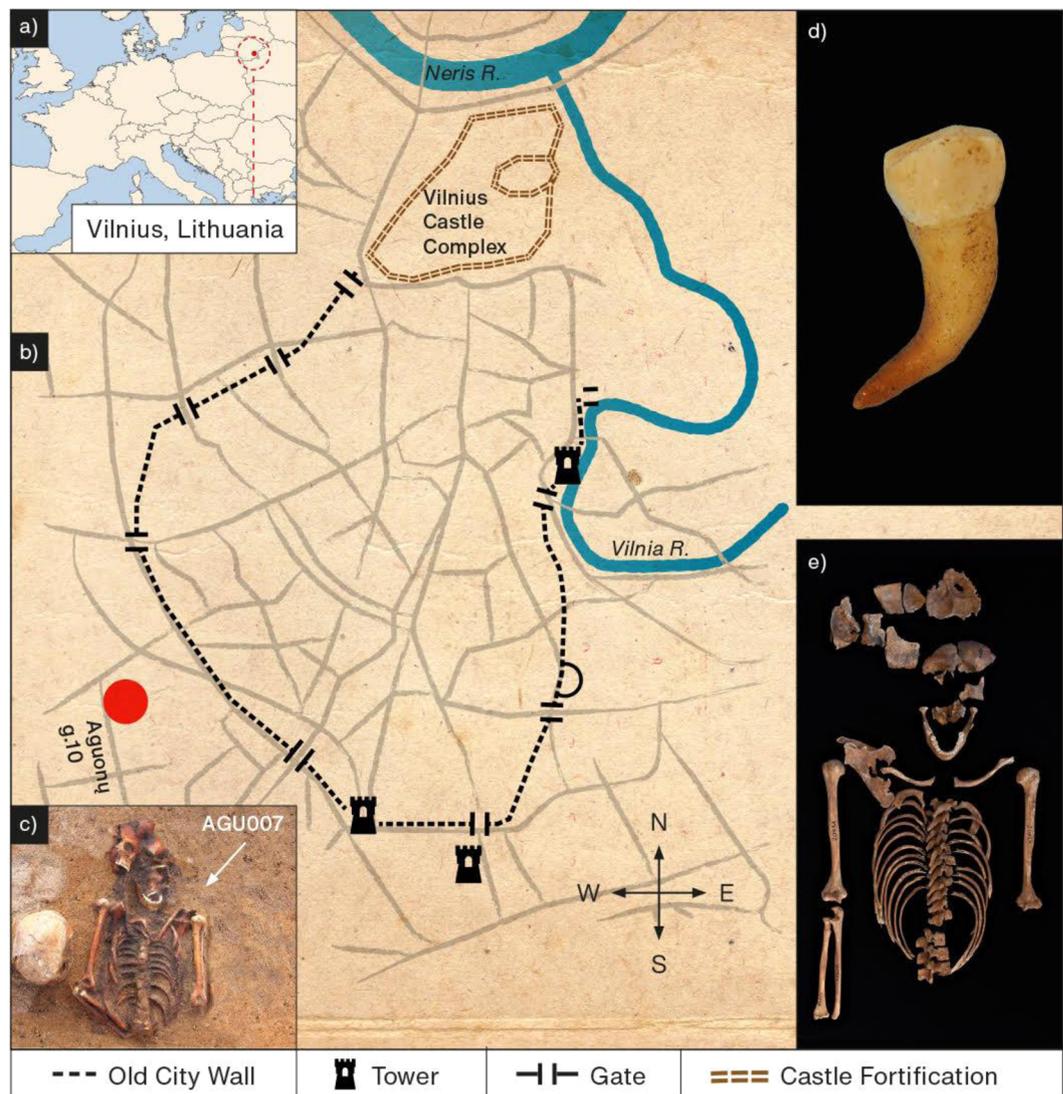
Developments in techniques for identification of pathogen DNA in archaeological samples can expand our resolution of disease detection. Our application of a non-targeted molecular screening tool for the parallel detection of pathogens in historical plague victims from post-medieval Lithuania revealed the presence of more than one active disease in one individual. In addition to *Yersinia pestis*, we detected and genomically characterized a septic infection of *Treponema pallidum pertenu*, a subtype of the treponemal disease family recognised as the cause of the tropical disease yaws. Our finding in northern Europe of a disease that is currently restricted to equatorial regions is interpreted within an historical framework of intercontinental trade and potential disease movements. Through this we offer an alternative hypothesis for the history and evolution of the treponemal diseases, and posit that yaws be considered an important contributor to the sudden epidemic of late 15<sup>th</sup> century Europe that is widely ascribed to syphilis.

Ancient DNA analyses have the potential to reveal history that is hidden within archaeological samples. Increased sequencing capacity made available by technological innovations has translated into concomitant increases in temporal depth for human genomes<sup>1</sup> as well as those from commensal<sup>2</sup> and invasive<sup>3</sup> microbes. Genome-level analyses of ancient bacteria and viruses have revealed unexpected patterns of pathogen dispersal in the past, and have prompted discussions to revise long standing theories about the evolution of some of our most well-studied diseases such as plague<sup>4</sup>, tuberculosis<sup>5</sup> and smallpox<sup>6</sup> to name but a few.

Genomic investigations of ancient microbes have largely been made possible via selective enrichment for DNA of a target organism. By design, this process requires *a priori* knowledge of a taxon of interest, which is typically acquired through either initial molecular screening<sup>7</sup>, inference from historical context<sup>8</sup>, diagnostic gross pathology<sup>9</sup>, or combinations thereof<sup>5</sup>. More recent approaches have explored the analytical resolution of broad multi-species enrichments<sup>10</sup> or fully non-targeted approaches<sup>11,12</sup> for pathogen detection. Though demonstration of these techniques is currently restricted to single microbial organisms, they indeed offer the flexibility to detect co-morbidities on historical timescales, thus giving a glimpse into past disease ecology.

Here we present the application of an ancient pathogen detection method to identify co-morbidities in archaeological specimens associated with an historical epidemic. The influence of co-infections as exacerbating the progression of certain diseases is an established phenomenon<sup>13,14</sup> and has been proposed to partially account for plague's unusually high mortality in the past<sup>15,16</sup>, its potential influence on selective mortality<sup>16</sup>, as well as its persistence and subsequent rapid extinction in post-medieval Europe<sup>8</sup>. As a first investigation into the existence of co-morbidities in post-medieval plague victims we applied a non-targeted pathogen screening approach offered through the HOPS platform<sup>17</sup> to skeletal material from a suspected 15<sup>th</sup>–16<sup>th</sup> century plague burial from Vilnius, Lithuania. HOPS couples high sensitivity in taxon assignment with an evaluation of molecular qualities common to ancient DNA: together these features make it an ideal tool for detection of concurrent infections in

<sup>1</sup>Max Planck Institute for the Science of Human History, Jena, Germany. <sup>2</sup>Vilnius University, Vilnius, Lithuania. <sup>3</sup>Curt-Engelhorn-Zentrum Archäometrie, Mannheim, Germany. <sup>4</sup>Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Rümelinstraße 23, 72070, Tübingen, Germany. <sup>5</sup>These authors contributed equally: Karen Giffin and Aditya Kumar Lankapalli. ✉e-mail: [herbig@shh.mpg.de](mailto:herbig@shh.mpg.de); [bos@shh.mpg.de](mailto:bos@shh.mpg.de)



**Figure 1.** (a,b) Map showing the city of Vilnius, Lithuania, and the relative positions of the castle fortification, the ancient city wall, and the archaeological site, Agunu g. 10 (red circle). (c) Multiple burial containing individual AGU007, indicated by the white arrow (photograph courtesy of Robertas Žukovskis). (d) Premolar from individual AGU007 used for molecular analyses. (e) exhumed skeleton of individual AGU007 (photograph courtesy of Justina Kozakaitė). Figure created by Hans Sell.

metagenomic data from archaeological tissues. In addition to *Yersinia pestis*, our analyses succeeded in identifying a disease of the treponemal family, *Treponema pallidum pertenue* (yaws), in one of the four confirmed plague victims. While a physiological relationship between the two infections is not explored here, genomic data from both pathogens are considered independently in terms of their unique evolutionary histories. Analyses of the newly reconstructed treponemal genome led us to propose a novel hypothesis that a recent introduction of yaws may have contributed to the devastating post-medieval outbreaks that are commonly attributed to the onset of syphilis in Europe.

### Site Description

In 2006 and 2007, a post-medieval cemetery containing 119 burials and 216 individuals was excavated in Vilnius, Lithuania. As historical or cartographic descriptions of the site are lacking, its existence was unknown prior to excavation. The cemetery is located outside the boundary of the medieval city walls (Fig. 1), and contained a number of multiple burials – 33 of these contained up to 15 individuals in a single grave, and are thus interpreted to each represent single mortality events<sup>18</sup>. These contexts, as well as the demographic composition that differs from other attritional contexts (here with young adults and adolescents prevailing), led to the site's recognition as a potential burial ground for victims of epidemics<sup>18</sup>.

Individual	Library Type	Raw reads	Mapping Reads post duplicate removal	Target DNA (%) post enrichment	Mean Fold Genomic Coverage	Percentage of Genome Covered at 3X(%)
<b>a) <i>Y. pestis</i> mapping:</b>						
AGU007	non-UDG	20901682	1704427	36.22	26.81	94.10
	UDG	23998414	2344695	37.33	38.01	94.32
AGU010	non-UDG	24499218	970324	30.18	12.52	92.57
	UDG	24725162	1480281	25.45	20.76	93.51
AGU020	*merged	14397991	132259	2.55	1.68	23.66
	non-UDG	19136670	3978	0.19	0.04	0.06
	UDG	19925684	8235	0.17	0.09	0.09
AGU025	non-UDG	19964532	266287	23.19	4.17	74.30
	UDG	23621324	849774	24.94	14.36	93.65
<b>b) <i>T. pallidum</i> mapping:</b>						
AGU007	non-UDG	46614028	146398	7.64	9.31	96.68
	UDG	45771712	330505	8.10	21.36	97.69

**Table 1.** Mapping statistics for *Y. pestis* and *T. pallidum* capture products. Results are reported for the second processed tooth only unless otherwise specified. \*Merged data from 2 aliquots of the UDG library of the first tooth.

## Results

**Pathogen screening and genomic reconstruction.** Powdered material from the pulp chamber of one tooth from each of 26 individuals and their associated negative controls were processed following methods described elsewhere (Supplementary Information) in the dedicated ancient DNA facility of the Max Planck Institute for the Science of Human History. To identify plague victims in the skeletal series, presence of *Yersinia pestis* DNA in the extract was assessed via a targeted amplification-based quantitative PCR (qPCR) approach for the *pestis*-specific *pla* gene of the PCP1 plasmid<sup>19</sup>. This assay revealed six putative plague victims: individuals AGU003, AGU009, AGU010, AGU013, AGU020 and AGU025 (Table S1; Figures S1; S2). Negative controls were free of amplification products.

In a separate strategy, double-stranded Illumina screening libraries were prepared for all 26 individuals from 10  $\mu$ L of DNA extract following established protocols<sup>20</sup>. Double-stranded libraries from 40  $\mu$ L of DNA extract were prepared alongside for the six putative plague victims, with enzymatic treatment to remove taphonomic chemical damage in the form of cytosine deamination<sup>21</sup>. All libraries were subsequently processed following established protocols<sup>22</sup> and were sequenced on an Illumina HiSeq 4000 to a depth of ca. 10 million reads each. Amplified UDG-treated and untreated libraries of the putative plague victims, as well as their associated negative controls, were prepared for two consecutive rounds of in-solution whole genomic *Y. pestis* capture performed as previously described<sup>23</sup>. Samples were captured in individual wells, with negative controls pooled in a single capture reaction. Capture products were sequenced to a depth of ca. 10 million reads each on an Illumina HiSeq 4000. All negative controls were sequenced on a NextSeq 500 to a depth of ca. 2 million reads each.

Mapping of the sample screening (i.e. shotgun) libraries to the *Y. pestis* CO92 reference genome (NC\_003143.1)<sup>24</sup> yielded between 2 and 87 unique *Y. pestis* chromosomal fragments for each screening library after duplicate removal (Table S3). At most 30 chromosomal fragments were detected in any one of our negative controls. Unexpectedly, the greatest number of mapping reads was identified in individual AGU007: this sample performed poorly in our initial *Y. pestis* qPCR screen despite meticulous evaluation of PCR inhibition prior to *pla* amplification (Figure S1, Table S1). For authentication, a non-UDG library from this individual was prepared for two rounds of in-solution *Y. pestis* capture and sequenced on a NextSeq 500. A UDG-treated library was also prepared, shotgun sequenced, and captured for *Y. pestis*, all carried out as described above with sequencing done on an Illumina HiSeq 4000.

Enriched products yielded convincing signals of ancient *Y. pestis* preservation in individuals AGU007, AGU010, AGU020, and AGU025 (Table S4, Table S5, Figure S4), though at levels too low to permit genomic reconstruction. This motivated us to process a second tooth from each of these individuals with the goal of increasing genomic coverage. DNA from these additional teeth was extracted and shotgun sequenced, as well as captured and sequenced to a depth of approximately 20 million reads following methods described above. Due to the low coverage of individual AGU020 (Table S4), an additional UDG library aliquot for the first tooth was captured along with this second set. Enriched reads from both AGU020 UDG-treated fractions from the first tooth were merged for analysis. Subsequent mapping to the *Y. pestis* reference after duplicate removal yielded mean coverages of 1.68- to 38.01-fold for all UDG-treated fractions, with the highest amounts identified in individual AGU007, which also showed excellent preservation of human DNA (Table 1, Tables S6-S10). Signals of ancient damage were detected in all untreated libraries (Figure S4). At most 24 reads mapped to the *Y. pestis* chromosome in our negative controls (Table S8).

To investigate whole microbial content, shotgun libraries from the initial screening dataset for the 26 teeth were evaluated using the MEGAN Alignment Tool (MALT) and queried for the presence of pathogen DNA via the screening pipeline HOPS<sup>17,25</sup>. The MALT database was constructed from a custom RefSeq Genome set in November 2017 that contained bacteria, viruses, and eukaryotes (<ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/>).

Predictably, *Y. pestis* signals were detected in both the UDG-treated and untreated libraries for individuals AGU007, AGU010, AGU020 and AGU025, all of which showed evidence of DNA damage (Figure S4). In further contrast to the qPCR screening results, no *Y. pestis* signals were detected in either the UDG-treated or untreated libraries for individuals AGU003, AGU009, and AGU013, indicating that these individuals may have yielded false positive identifications in the initial qPCR screen.

Individual AGU007, however, showed a weak signal for *Treponema pallidum*, a spirochete bacterium associated with the human diseases pinta, bejel, yaws, and most famously syphilis. Though its identification in HOPS was limited to only 12 assigned reads in the UDG treated library (Figure S3), a close examination of data from the untreated library revealed 7 reads assigned to *T. pallidum*, with additional reads assigned to *Treponema* at the genus level for both libraries. Mapping against the Nichols reference for *T. pallidum pallidum* yielded 36 and 27 reads in the treated and untreated libraries, respectively (Table S11, Figure S3). Intrigued by this, we captured both the UDG-treated and untreated libraries for individual AGU007 for *T. pallidum* DNA as previously described<sup>9</sup>. *T. pallidum*-enriched libraries were mapped to the Nichols reference yielding 31,882 and 126,552 unique reads and 1.78 and 6.67 mean fold coverage for the untreated and treated fractions, respectively, following duplicate removal (Table S12). The untreated library showed a C to T transition pattern comparable to that from the *Y. pestis* mapping, consistent with chemical damage of ancient molecules (Figures S4 and S5). *T. pallidum* capture of the second tooth from individual AGU007 yielded 330,505 mapping reads (following duplicate removal) that assembled to build a high-quality genome of 21-fold mean coverage (Table S14). Ancient damage patterns observed in the untreated fraction support its authenticity (Table S14; Figure S5). By contrast, the 8 negative controls associated with this capture together yielded only one read in each of three blank libraries that mapped to the reference (Table S14) indicating exceptionally clean working conditions. A BLASTn<sup>26</sup> evaluation of each read revealed them to come from regions that are conserved in many bacterial taxa, and are hence likely to derive from non-target sources. The two UDG-treated enrichment datasets were subsequently merged to yield a *T. pallidum pertenuis* genome of 26-fold mean coverage (Fig. 2).

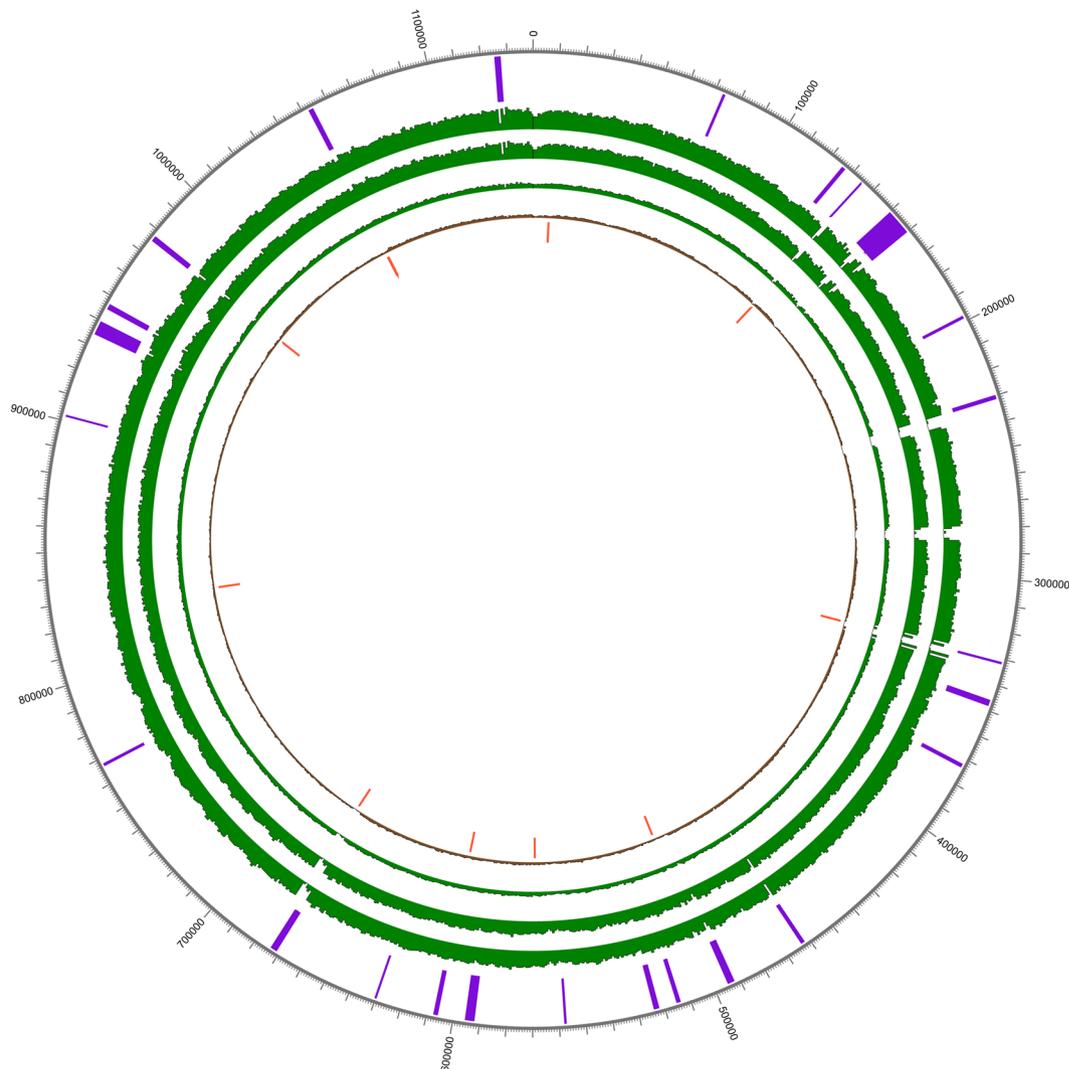
**Morphological assessment of individual AGU007.** The postcranial skeleton of individual AGU007 is extremely fragmented. Morphological analysis was therefore limited to the skull, clavulae, the right scapula, the upper limbs, ribs, and thoracic vertebrae. Periosteal lesions were identified as active when the affected portion of the bone appeared porous with unremodeled edges. Lesions were identified as healed if the affected portion of bone demonstrated rounded and remodeled edges. Individual AGU007 had mixed (active and healed) periosteal lesions on the distal portion of the right humerus and the proximal portion of the right ulna. A slight active periosteal reaction was observed on small fragments of the internal surface of the parietal bones. Images of pathological bone are shown in Figure S6. Pathological surveys are not reported for the remaining individuals of the skeletal series since they yielded no detectable infections other than *Y. pestis*, which does not cause skeletal involvement.

**Human DNA analysis of individual AGU007.** *T. pallidum pertenuis* is currently restricted to equatorial regions. To determine if individual AGU007 had ancestry associated with areas of modern yaws epidemiology, we undertook human genetic analyses. Shotgun sequencing data from both untreated and UDG-treated libraries for the second tooth of individual AGU007 were mapped against the human reference genome (hg19) with BWA<sup>27</sup> as implemented in EAGER<sup>28</sup>, and the damage pattern was estimated with *mapDamage2.0*<sup>29</sup> (Table S6). Based on coverage of the sex chromosomes relative to the autosomes the individual was determined to be female (Table 2a). Untreated sequencing data were also mapped against the mtDNA reference sequence (rCRS) (Table 2b). The resulting alignment was provided as input to *schmutzi*<sup>30</sup> and the reconstructed mtDNA consensus sequence was assigned to haplogroup H1ap1 using *Haplofind*<sup>31</sup>. MtDNA contamination level was estimated to be ca. 1% (0–2%), which can also be considered a proxy for nuclear contamination due to the low mitochondrial to nuclear DNA ratio (Table 2a)<sup>32</sup>. Sequencing data from the UDG-treated library was subsequently used for pseudo-haploid genotyping with *pileupCaller* (<https://github.com/stschiff/sequenceTools>) and then merged to the *Human Origins* dataset<sup>33</sup> resulting in 96,953 overlapping SNPs. A principal component analysis (Fig. 3) was calculated based on modern-day West Eurasian populations<sup>34</sup> onto which the ancient individual was projected using *smartpca*<sup>35</sup>. This analysis places her genome in close proximity to present-day Baltic populations such as Estonians and Lithuanians.

**Phylogenetic analyses.** *Yersinia pestis*. Phylogenetic analyses for *Y. pestis* were carried out using UDG treated datasets obtained from the second tooth only, with the exception of individual AGU020: inclusion of the ca. 8000 mapping reads from the second tooth did not increase its coverage beyond 1.68-fold, and its persistent low recovery led to its exclusion from phylogenetic tree construction. Its tentative position in the tree was rather determined based on manual evaluation of SNPs (Fig. 4, Table S16).

*Y. pestis* SNPs were called using UnifiedGenotyper in the Genome Analysis Toolkit (GATK)<sup>36</sup> implemented in EAGER, and analyzed within a published set of 275 genomes, of which 41 were ancient (Table S15). Excluded genomic regions were consistent with those selected in other published works<sup>22</sup>. SNPs were identified in all positions that had at least 3-fold coverage, a minimum genotyping quality of 30, and consensus calling at 90% support. This produced a final SNP alignment of 6949 positions (Table S16). From this, a maximum likelihood phylogenetic tree was constructed with RAxML<sup>37</sup> (Figs. 4 and S7).

All Lithuanian *Y. pestis* genomes share a number of positions that place them within the diversity of the second pandemic descendants of the Black Death (London ES, NAB003, and Barcelona 3031) that are hypothesised to represent plague's post-Black Death persistence in Western Europe. Based on our current phylogeny, two of these genomes, AGU007 and AGU010, are most closely related to the strain from Ellwangen, Germany<sup>8</sup>, here identified to have one unique derived position based on the analysis of a higher quality genome of this isolate that is

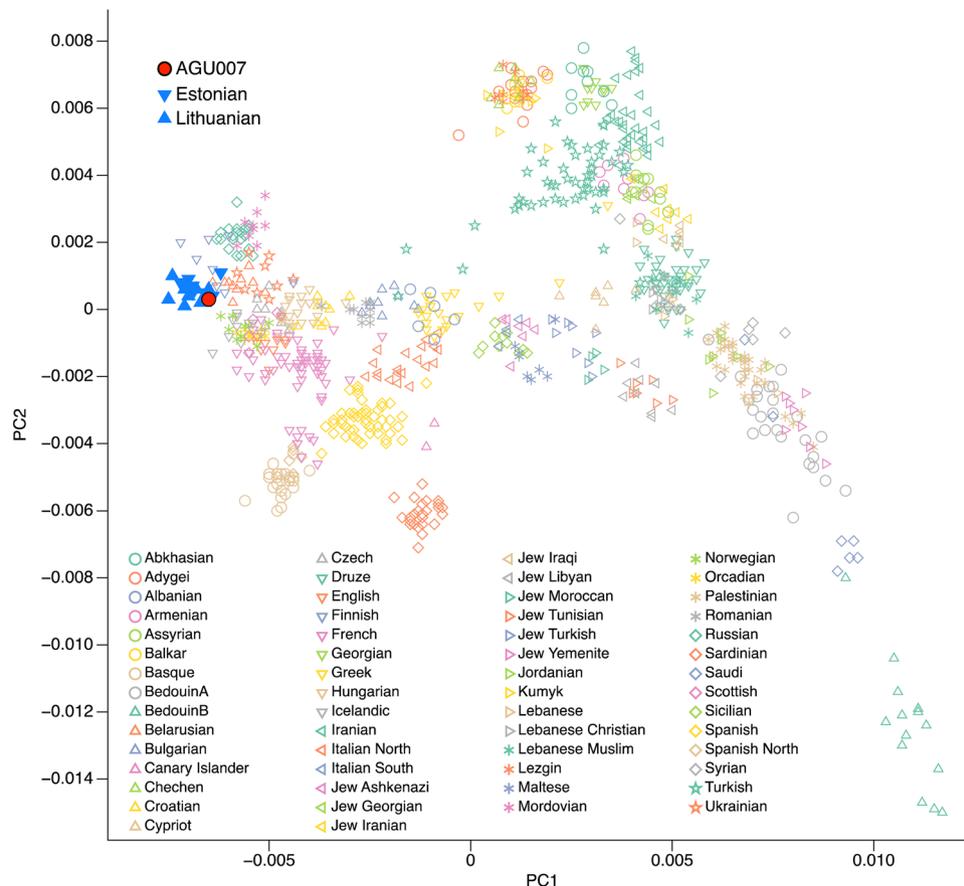


**Figure 2.** Circos coverage plot for ancient *Treponema pallidum pertenu*. The innermost circle represents 13 homoplastic sites (orange) detected between *pallidum*, *pertenu* and *endemicum* strains through SNP evaluation. Histogram showing the coverage of ancient *pertenu* genome 133 (Schuenemann *et al.*, 2018) is shown in brown. Coverages of enriched data for AGU007 are shown in green for first enrichment (inner), second enrichment (middle) and merged set (outer). The outermost positions in purple highlight the recombining regions identified by ClonalFrameML.

Library type	Raw Reads	Mapped Reads	Target DNA (%)	Mt/nuc Ratio	C to T 5'end (%)	Average fragment length (bp)	X/Y rate
<b>a) nuclear genome</b>							
non-UDG	12,853,927	7,492,473	69.99	147.81	16	62.02	0.84/0.04
UDG	14,325,984	9,752,652	68.98	127.29	1	61.29	0.83/0.05
<b>b) mitochondrial genome</b>							
Mean Coverage	C to T 5'end (%)	MtDNA Haplogroup	Contamination (%)				
26.80	21	H1ap1	1 (0–2)				

**Table 2.** Human DNA mapping statistics.

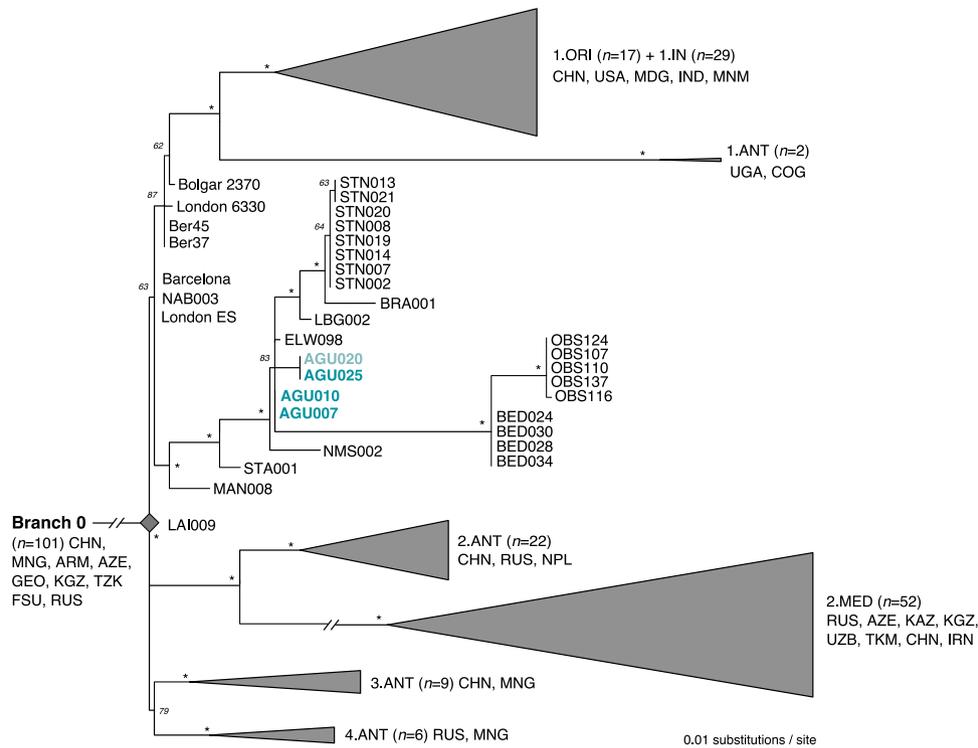
now available (ELW098). Our analysis identified an additional 4 positions that are unique to genome AGU025. While the coverage of genome AGU020 was deemed too low for inclusion in tree construction, visual inspection of the data revealed it to have at least one read spanning three of the four positions that are unique to AGU025, with identical SNP calls (Table S17). One of these shared positions causes a non-synonymous change in the flavin reductase gene *fre* of the FMN-NADPH reduction pathway (Table S18). An additional position present in only AGU025 affects the *mrcB* gene, which is a penicillin binding protein.



**Figure 3.** Principal component analysis based on 96,953 SNPs queried against the Human Origins dataset.

Though all burials stem from a common cemetery, it cannot be determined archaeologically whether or not they are from a single mortality event or several events that took place over a series of decades. Radiocarbon dates were thus obtained for all four individuals in whom *Y. pestis* infection was identified (Fig. 5). A conservative estimate based on common overlap of the sigma-2 (95% probability) radiocarbon distributions places all four Lithuanian strains within 12 years of each other, from 1463 CE (oldest age estimate from AGU007) to 1475 CE (youngest age estimate from AGU020). Of note, genome STA001 from Starnberg, Germany has an archaeological date of 1433–1523 CE<sup>22</sup>, which is consistent with our identification of a late 15<sup>th</sup> century plague outbreak in the Baltics from a more derived lineage. While a contemporaneous co-circulation of the four Lithuanian genomes presented here, which would also include the Ellwangen strain, remains possible based on their overlapping radiocarbon intervals, strain diversity on this scale (4 SNPs) has never been reliably observed in genomic data obtained from any single historical plague outbreak: all published Black Death genomes (1347–1351) are identical<sup>8</sup>, as are four genomes sequenced from a presumed single mortality event in 17<sup>th</sup> century London<sup>22</sup>. Only a single SNP distinguishes one genotype from the others in the catastrophe cemetery of l'Observance from 1722 Marseille, France<sup>38</sup> and the 15<sup>th</sup>–17<sup>th</sup>-century Stans mass burial from Switzerland<sup>22</sup>. The identical *Y. pestis* genome from individuals AGU007 and AGU010 thus parsimoniously implies they are contemporaneous. The similarity in radiocarbon distribution between individuals AGU010 and AGU020, the latter of which has at least three additional derived positions, provides support for constraining their age in the mid-15<sup>th</sup> century. Our phylogeny places genomes AGU007 and AGU010 at the base of a polytomy from which several descendent European lineages arise. As such events in *Y. pestis* evolution have been correlated with large-scale outbreaks<sup>39</sup>, our data provide support for a prolific late 15<sup>th</sup> century mortality event. As plague outbreaks were, however, frequent in Europe at this time, it is difficult to draw support from the historical literature to add resolution to its time interval. Regardless, it appears this event gave rise to two parallel tracks of persistent European plague: one thus far identified in continental Europe (Germany and Switzerland: LBG, BRA, STN) and another that has thus far been found in the port cities of 17<sup>th</sup> century London (BED) and 18<sup>th</sup> century Marseille.

As both London and Marseille were connected to a vast system of trade networks in the Mediterranean and beyond, many candidate plague source locations exist in areas underexplored from the vantage point of ancient DNA. The expansive range of the Ottoman Empire figures naturally into discussions of plague movements and source locations that could have fuelled these later outbreaks: its frequent interactions in the form of either war or trade with Lithuania or its allies in the 15<sup>th</sup> and 16<sup>th</sup> centuries<sup>40</sup> would have provided opportunities for plague introductions from the north such as the AGU007/AGU010 strain identified in this work. Further support for a northern origin comes from the base calls observed at position 4,208,536 (Table S16): although excluded in our

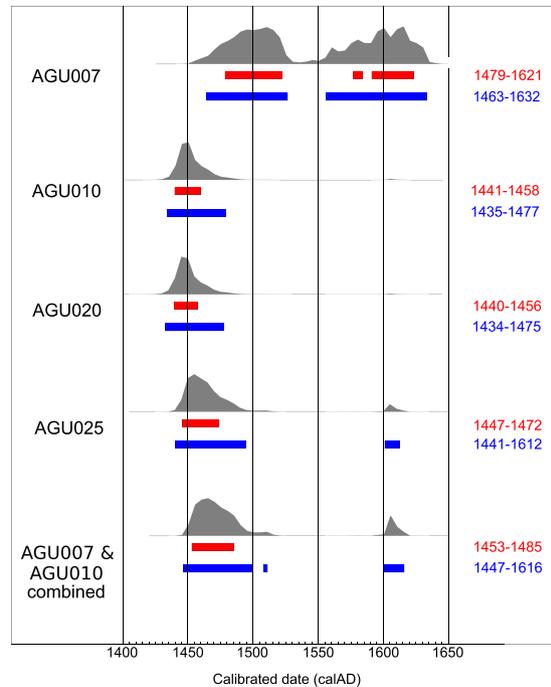


**Figure 4.** Maximum likelihood tree of post-Black Death genomes of *Y. pestis*. Constructed from 275 genomes with the Generalised Time Reversible (GTR) model, SNPs identified at 3-fold coverage, 1000 bootstrap replicates, and a 98% partial deletion filter (considering 5801 SNPs). Bootstraps with a value of 95 or greater are indicated with an asterisk (\*), and all those lower than 95 are presented with their numeric value. A phylogeny showing all genomes considered in this analysis is presented in Figure S7. The extrapolated position for the low coverage genome from individual AGU020 is shown in faded text.

phylogeny through application of a 98% partial deletion filter (Fig. 4, Table S16), this position shows the derived state uniquely in genomes AGU007/AGU010, Bedlam (London), and l'Observance (Marseille), thus indicating shared ancestry. Regardless, westerly Mediterranean sources for the inland movement of plague through the Balkans during these periods has received more attention in the historical literature<sup>41</sup>. Of note, the frequency of recorded plague outbreaks in regions under Ottoman control increased abruptly in the late 15<sup>th</sup> century and persisted thereafter<sup>41</sup>, which is consistent with a regional establishment of the disease, as has been suggested for post-Black Death Europe<sup>8</sup>. A connection between plague and the Ottoman Empire is also consistent with the historical narrative for the Great Plague of Marseille having been introduced via the *Grand-Saint-Antoine* merchant ship that collected a variety of trade goods, including supposedly plague-ridden silks, from several Levantine trading destinations days before the disease erupted amongst its crew<sup>42</sup>.

*Treponema pallidum*. Phylogenetic relationships of *T. pallidum pallidum* (n = 2), *endemicum* (n = 2), and *pertenue* (n = 27) (Table S19) were determined by calling all variant positions against the Nichols syphilis reference at a minimum coverage threshold of 5-fold and mapping quality of 30 using MultiVCFAnalyzer v0.85 (<https://github.com/alexherbig/MultiVCFAnalyzer>). Both genomes A12, sampled from a non-human primate<sup>43</sup>, and the 17<sup>th</sup> century genome from Colonial Mexico<sup>9</sup> (133) were removed due to low coverage (Table S20). Yaws genomes isolated from non-human primates were included because their phylogenetic placement may inform on directions of host switching. As treponemal genomes are known to recombine<sup>9,44,45</sup> we sought to identify recombinant regions in our dataset with ClonalFrameML<sup>46</sup> using a maximum likelihood tree reconstructed in RAXML<sup>37</sup> with a GTR + GAMMA substitution model for 1000 bootstrap replicates as input. This yielded 54 recombinant regions ranging in size from 3 to 8825 bp, which contained 737 positions that overlapped with, and were subsequently removed from, our list of variants (Table S20). Recombination events specific to AGU007 were not observed. Presence of additional homoplastic positions between yaws and either syphilis or bejel were identified through automated filtering on the SNP table (see supplementary methods). This process revealed an additional 13 positions that were removed for phylogeny construction and molecular dating (Table S21).

The SNP alignment generated after removal of detectable recombinant and homoplastic sites was used as input for phylogeny construction in RAXML<sup>37</sup> with parameters as described above. All trees consistently yielded high statistical support for discrete grouping of the three treponemal sub-clades of *pallidum* (syphilis, n = 2), *pertenue* (yaws, n = 25), and *endemicum* (bejel, n = 2). Figure 6 shows an estimate of topology based on removal of all sites with missing data. The genomes considered here derive from an extensive polytomy indicative of an expansive event coincident with the emergence of all currently available yaws genotypes. Our ancient genome



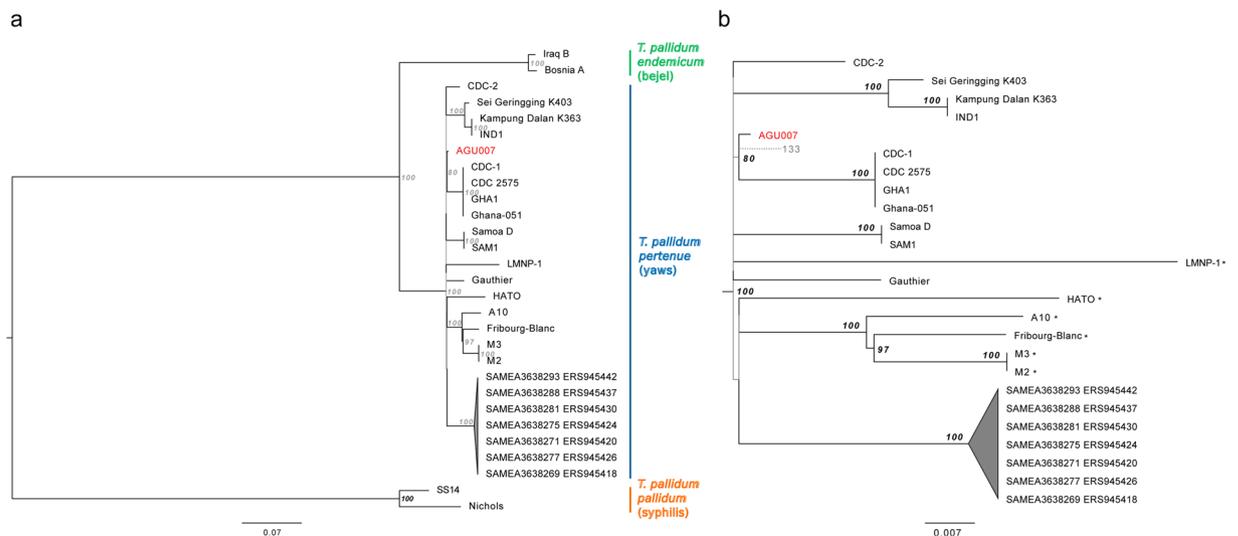
**Figure 5.** Comparative radiocarbon distribution plots for the individuals with confirmed plague infection. AGU007 (MAMS40889):  $^{14}\text{C}$  age =  $353 \pm 18$  yrBP, AGU010 (MAMS40326):  $^{14}\text{C}$  age =  $426 \pm 19$  yrBP, AGU020 (MAMS40328):  $^{14}\text{C}$  age =  $428 \pm 19$  yrBP, AGU025 (MAMS40329):  $^{14}\text{C}$  age =  $405 \pm 18$  yrBP. Sigma-1 ranges are shown in red, sigma-2 in blue. The curve for the combined plots of individuals AGU007 and AGU010 is also shown. (created with Oxcal v4.3.2<sup>82</sup>).

AGU007 groups consistently with genomes recently typed from human cases in Ghana, West Africa (CDC-1, CDC 2575<sup>47</sup>, and Ghana-051<sup>47</sup>). In addition, AGU007 has a much shorter branch length than its modern counterparts, as expected of ancient genotypes<sup>48</sup>. When all sites are considered, AGU007 differs from the extrapolated yaws MRCA by only five positions (Table S22), two of which are unique to this genome. Both of these private substitutions cause non-synonymous changes, one in the GroES chaperone protein (position 39,267) and another in a methyl-accepting chemotaxis protein (position 523,975) (Table S23), hence their functional significance is worthy of future investigation. While assessment of genome 133 revealed it to have a coverage considered too low for inclusion in tree construction, manual evaluation of its shared SNPs reveals it to possess the derived state for the position shared between AGU007 and the Ghana cluster, thus placing it within this clade. With regard to the positions unique to AGU007, 133 shows the ancestral state for one and no coverage for the other, hence resolution to infer further details on their relationship is currently lacking (Fig. 6; Table S22).

Our identification of a yaws genome in individual AGU007 could provide important clues on disease dissemination and ecology in post-medieval Europe. Given that this individual is suspected to be contemporaneous with AGU010 based on their identical *Y. pestis* genomes, additional resolution for their extrapolated age was sought through merging their radiocarbon data. This resulted in a calibrated radiocarbon date of 1447–1616 CE at a 95% confidence level (sigma-2), though with 83% of the probability density contained in the interval 1448–1498 CE (Fig. 5). This encompasses a time period consistent with the onset of an unknown acute and physically disfiguring disease in Europe.

**Molecular dating of the yaws cluster.** Given the small genetic distance between the genome reconstructed from AGU007 and the ancestral node of all yaws considered here, we employed a Bayesian molecular dating approach to formally estimate the time to the most recent common ancestor (tMRCA) of the *Treponema pallidum pertenuis* genomes in our dataset. Molecular dating was performed on the same alignment used for phylogenetic inference. Clock-like structure was evaluated using TempEst<sup>49</sup> and the MEGA maximum likelihood clock test<sup>50</sup> prior to molecular dating using BEAST (v2.5.2)<sup>51</sup>. The null hypothesis of equal evolutionary rates throughout the tree was rejected.

Due to the range of possible dates for AGU007 based on radiocarbon intervals, we prepared two models in which we sampled the AGU007 tip date from (1) the sigma-2 date range for the radiocarbon date of AGU007 alone (most conservative estimate, 1464–1633 CE; sig2 model) and (2) given their identical *Y. pestis* genomes, the combined sigma-1 date range for AGU007 and AGU010 (least conservative estimate, 1453–1485 CE; sig1 model). The clock rate was not meaningfully impacted by differences in tip dating ranges for AGU007, as both yielded mean values well within the 95% highest posterior density (HPD) intervals for each model (Figure S8). The mean dates for the tMRCA for the yaws cluster included here differ by only 169 years between the Sig1 and Sig2 models (Table 3). The 95% HPD intervals for this parameter also overlapped substantially and contained the mean value



**Figure 6.** Maximum likelihood phylogenetic tree of *T. pallidum* with 1000 bootstrap replicates generated after removal of 13 homoplastic positions and 54 regions of recombination, followed by deletion of all sites with ambiguous bases. SNP calling was based on a coverage of 5-fold, and the tree was constructed from 929,012 positions of which 1223 were polymorphic. Figure 6a shows a phylogeny of the full treponemal dataset considered here and 6b shows a zoom of the monophyletic clade of *pertenuae* strains, where all those isolated from non-human primates are distinguished with an asterisk (\*). Branches with statistical support less than 80 are depicted in grey. The scale represents mean substitutions per site given a GTR + GAMMA substitution model. Genome AGU007 is shown in red. The dashed branch and faded text depicts the extrapolated position of ancient genome 133.

Model	Clock Rate (95% HPD)	Yaws tMRCA (95% HPD; YBP)	AGU Tip Date (95% HPD)
Sig1	4.6E-8 (2.4E-8, 6.9E-8)	865 (571, 1235)	1469 CE (1455, 1485)
Sig2	5.8E-8 (2.7E-8, 8.8E-8)	696 (424, 1043)	1561 CE (1477, 1633)

**Table 3.** Selected parameter estimates from BEAST dating.

of the alternative model (Figure S8). Together these results suggest an emergence of this cluster within the last millennium.

## Discussion

Our recovery of *T. pallidum pertenuae* in archaeological human remains from post-medieval Lithuania invites an exploration of the historical context surrounding treponemal infections during this period. In the late 15<sup>th</sup> century, historical accounts begin to accumulate in Europe<sup>52</sup> and the Middle East<sup>53</sup> that describe a physically disfiguring disease seemingly unknown to medicine at the time. Early documents surfaced in Spain with Ruy Díaz de Isla's description of "bubas", though more widely accepted is Gonzalo Fernández Oviedo y Valdés's description of a malady that affected the mercenary armies of Charles VIII of France that laid siege to Naples in early 1495<sup>52</sup>. By mid-spring of that year the armies began retreating back to their homelands, though a new, highly contagious, disfiguring and rapidly progressing disease accompanied their military victory. In subsequent decades the illness took on many names such as the "French pox" or the "Pox of Naples", all of which were more reflective of political and social tensions of the era than biological truths<sup>54</sup>. Today these accounts are widely assumed to be the first documented descriptions of the treponemal disease known as syphilis.

Theories on its origin began to surface shortly after its first classification as a distinct condition, and fervent controversy over the varying hypotheses persists to this day. Most discussants favour the Columbian theory that assumes acquisition from New World populations by Columbus' crew on their first trans-Atlantic voyage<sup>52</sup>, with its rapid transmission in Europe facilitated through cultures of sexual activity<sup>55</sup>. Genetic data from both human and non-human primate treponemal infections are interpreted as providing support for the theory, and have led to the proposal of a prehistoric emergence of the treponemal disease cluster in the form of yaws in Africa<sup>56</sup>. The spirochete is then thought to have followed modern humans as an "heirloom" disease in their Pleistocene dispersal across the globe, where it adapted locally into the endemic form bejel (*Treponema pallidum endemicum*) in the Middle East and several hypothetical sub-types of yaws (*Treponema pallidum pertenuae*) in both the New and Old World. Introduction of New World yaws to 15<sup>th</sup> century Europe is then thought to have provided the ecological context for the emergence of the sexually transmitted variety, syphilis (*Treponema pallidum pallidum*)<sup>52,57</sup>. Phylogenetic relationships revealed from genome-level analyses<sup>9,43,44</sup> show cladistic separation of the

three treponemal pathogens, and while they share a common origin, there is no support for yaws as the ancestor. Dispersal of yaws in both humans and non-human primates, however, is still thought compatible with it being an heirloom disease<sup>58</sup>: this would push emergence of the pathogenic treponemal cluster far back into the Pleistocene, leaving open multiple models compatible with an early evolution in the Americas<sup>58</sup>.

Though the theory of a New World origin is almost as old as historical documentation of the disease itself<sup>59</sup>, it is not without critics. A growing body of evidence from the meticulous recording of skeletal lesions considered pathognomonic for syphilis identified in pre-Columbian Europe challenges popular assumptions on its origin. Under this model, syphilis may well have been present in Europe for centuries before New World contact, though may have been obscured by alternative nomenclature such as “leprosy”, a singular term thought to have described a variety of disfiguring conditions in antiquity. Syphilis could then have been among the many diseases brought to the New World via European contact. Both theories are fuelled by analyses of skeletal material from New<sup>60</sup> and Old World contexts<sup>61–63</sup>, with the latter questioned by the Columbianists based on scepticism about diagnostic resolution for syphilis pathology in archaeological bone or the influence of marine reservoir effects on radiocarbon dates<sup>57,64</sup>. Others contend that these pre-Columbian forms could result from related treponemal infections, which were later replaced by syphilis in the late 15<sup>th</sup> century<sup>65</sup>. In contrast, sparser attention is given to the “alternative” theory<sup>66</sup>, which also assumes a treponemal emergence in sub-Saharan Africa at some unspecified time in the past, though here either syphilis or its immediate predecessor became introduced to Europe via increased trade contacts with West Africa in the late 15<sup>th</sup> century. This less popular scenario is thus entirely independent of New World involvement, and emergence of the disease in Europe would have been merely correlative with Columbus’ travels<sup>67</sup>.

Importantly, all three hypotheses operate on the assumption that the new disease of late 15<sup>th</sup> century Europe was unequivocally an intense epidemic of the sexually transmitted *Treponema pallidum pallidum*. Its rapid onslaught and perceived decrease in virulence within half a century is frequently cited as a hallmark example of an emergent disease within a naïve host population<sup>68</sup>, wherein either pathogen or host rapidly adapt to achieve a closer approximation of a symbiotic relationship<sup>52,56,69</sup>. Here we propose a hypothesis that is to our knowledge novel and requires fewer assumptions to explain the epidemiological trends – that yaws be considered an important contributor to the late 15<sup>th</sup> century European pandemic. Yaws, and for that matter the other currently recognised treponemal diseases bejel and pinta, have not figured into discussions and theories of treponemal ecology in medieval Europe because their modern distribution reveals them to be restricted to warmer climates. Regions of sub-Saharan Africa, Southeast Asia, and Oceania are current targets for yaws eradication efforts, where the disease is invariably linked to rural areas lacking adequate resources to sustain global hygienic standards<sup>70–72</sup>. Our unexpected finding of yaws in a European individual from the Baltics, where skeletal evidence of pathognomonic treponemal pathology is purportedly absent prior to the 16<sup>th</sup> century<sup>67</sup>, reveals that at least the past form of the disease was indeed infectious under climatic conditions that differ from its current niche. Its phylogenetic position in close proximity to the radial divergence of all known members of the yaws cluster permits the establishment of a temporal dimension for its evolution, further supported by our dating estimates, and demonstrates the disease, as it is currently defined, to be far younger than previously assumed<sup>56</sup> – with an inferred emergence in the 12<sup>th</sup>–14<sup>th</sup> century. This recent emergence would imply that any older material presenting treponemal pathology results from strains that are not defined by the current yaws cluster.

Although the lack of a clear ancestor-descendent relationship with any typed lineage prevents resolution on directions of transmission between humans and non-human primates, the close relationship of our ancient genome to extant West African lineages, coupled with the diversity observed in other modern African yaws genomes, is consistent with its emergence in this region. As noted by the alternative theory, increased European presence in West Africa in the mid- to late-15<sup>th</sup> century would have provided a means for intercontinental disease movement<sup>66</sup>. The lure of West African gold motivated the Portuguese establishment of Elmina (modern Ghana) in 1482, though its activities quickly expanded beyond trade of minerals to include importation of African people to Europe as slaves. By the mid-sixteenth century, an estimated 10% of Lisbon’s population was of African descent<sup>73</sup>. Introduction of a highly contagious skin infection such as yaws during this period could have easily led to its presence in the 20,000 strong 1494 mercenary army of Charles VIII, whose contingent was assembled from a wide recruitment of men across Europe<sup>74</sup>.

Cutaneous lesions from yaws infections can occur anywhere in the body, and usually take the form of raised chancres similar in shape to a wild raspberry. Such pathology is consistent with reports of “welts the size of acorns” that described early symptoms of the 15<sup>th</sup> century disease<sup>75</sup>. Skeletal involvement in yaws is common, and can be manifested as dactylitis and periostitis of long bones. Later stages of the disease, if left untreated, can result in tibial bowing, caries sicca on the skull and gangosa of the nasal bones and maxilla, hence comprising a spectrum of pathology that can be indistinguishable from advanced syphilis in archaeological<sup>54,76</sup> and even modern clinical contexts<sup>71</sup>. Recovery with subsequent immunity is common today, and contributes to the disease’s reputation as an illness of children in remote and underserved areas. Immunity amongst survivors of the historical condition could have been a driving mechanism behind the perceived reduction in virulence of the causative agent for the post-Columbian European outbreak, and would also be consistent with its low-level persistence into the 19<sup>th</sup> century in Europe, potentially in a variety of rare forms such as sibbens of Scotland, button scurvy of Ireland, and radesyge of Norway<sup>77</sup>, all of which receded with changing standards of personal hygiene. Though the skeletal manifestation in the Lithuanian woman investigated here is consistent with, though not specific to, endemic (or non-venereal) treponemal pathology, our identification of ancient yaws DNA in one of her teeth is confirmatory. As yaws infection undergoes a short period of haematogenous involvement in advance of prolonged infection, we regard the presence of treponemal DNA in the dental pulp chamber as stemming from increased vascular permeability associated with the acute septic progression of plague, coincident with an underlying and pre-existing active infection of yaws. This proposed causal relationship implies merit in deeper non-targeted pathogen surveys of skeletal collections from known periods of high mortality associated with the many examples of plague or other septic diseases in our history. The further support we demonstrate for plague’s

endemic status in post-medieval Europe and its continued potential for large-scale outbreaks beyond the Black Death would have provided a wealth of opportunity for co-morbidity and sepsis for a variety of infections.

Earlier pessimism regarding the inaccessibility of treponemal diseases in archaeological material were made on the assumption of the spirochete's fragile structure, its limited *in vivo* detection in animal bone under controlled laboratory infection trials, and simple lack of success in PCR amplifications from ancient bone of presumed treponemal sufferers<sup>78,79</sup>. As the roster of archaeological samples showing molecular preservation of these diseases increases<sup>9,80,81</sup>, and their detection is further enhanced by analytical advances, the resolution of enigmas related to the origins of the full treponemal cluster, inclusive of its other members syphilis, bejel, and pinta, are within reach. Such feats are likely to be realised in the coming years as the burgeoning field of palaeopathogenomics continues to mature and move beyond previously established boundaries.

## Methods

Additional methods can be found in our supplementary file.

## Data availability

Raw data from the above analyses can be found under ENA accession number PRJEB37508.

Received: 23 March 2020; Accepted: 30 March 2020;

Published online: 11 June 2020

## References

1. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207, <https://doi.org/10.1038/nature14317>, <https://www.nature.com/articles/nature14317-supplementary-information> (2015).
2. Weyrich, L. S. *et al.* Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**, 357, <https://doi.org/10.1038/nature21674>, <https://www.nature.com/articles/nature21674-supplementary-information> (2017).
3. Krause-Kyora, B. *et al.* Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife* **7**, e36666, <https://doi.org/10.7554/eLife.36666> (2018).
4. Rasmussen, S. *et al.* Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* **163**, 571–582, <https://doi.org/10.1016/j.cell.2015.10.009> (2015).
5. Bos, K. I. *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494, <https://doi.org/10.1038/nature13591>, <https://www.nature.com/articles/nature13591-supplementary-information> (2014).
6. Duggan, A. T. *et al.* 17th Century Variola Virus Reveals the Recent History of Smallpox. *Current Biology* **26**, 3407–3412, <https://doi.org/10.1016/j.cub.2016.10.061> (2016).
7. Mühlemann, B. *et al.* Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proceedings of the National Academy of Sciences* **115**, 7557, <https://doi.org/10.1073/pnas.1804921115> (2018).
8. Spyrou, M. A. *et al.* Historical *Y. pestis* Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host & Microbe* **19**, 874–881, <https://doi.org/10.1016/j.chom.2016.05.012> (2016).
9. Schuenemann, V. J. *et al.* Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains. *PLoS Neglected Tropical Diseases* **12**, e0006447, <https://doi.org/10.1371/journal.pntd.0006447> (2018).
10. Mühlemann, B. *et al.* Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423, <https://doi.org/10.1038/s41586-018-0097-z> (2018).
11. Devault, A. M. *et al.* Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array. *Scientific Reports* **4**, 4245, <https://doi.org/10.1038/srep04245>, <https://www.nature.com/articles/srep04245-supplementary-information> (2014).
12. Vågene, Å. J. *et al.* *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution* **2**, 520–528, <https://doi.org/10.1038/s41559-017-0446-6> (2018).
13. Singer, M., Bulled, N., Ostrach, B. & Mendenhall, E. Syndemics and the biosocial conception of health. *The Lancet* **389**, 941–950, [https://doi.org/10.1016/S0140-6736\(17\)30003-X](https://doi.org/10.1016/S0140-6736(17)30003-X) (2017).
14. Bell, L. C. K. & Noursadeghi, M. Pathogenesis of HIV-1 and *Mycobacterium tuberculosis* co-infection. *Nature Reviews Microbiology* **16**, 80, <https://doi.org/10.1038/nrmicro.2017.128> (2017).
15. Bos, K. I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506, <https://doi.org/10.1038/nature10549>, <https://www.nature.com/articles/nature10549-supplementary-information> (2011).
16. DeWitte, S. N. & Wood, J. W. Selectivity of Black Death mortality with respect to preexisting health. *Proceedings of the National Academy of Sciences* **105**, 1436, <https://doi.org/10.1073/pnas.0705460105> (2008).
17. Huebler, R. *et al.* HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol* **20**, 280, <https://doi.org/10.1186/s13059-019-1903-0> (2019).
18. Žukovskis, R. Agonuų g. 10, Vilniuje 2007 metų archeologinių tyrinėjimų ataskaita (Vilnius, Lithuania, 2007).
19. Schuenemann, V. J. *et al.* Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proceedings of the National Academy of Sciences* **108**, E746, <https://doi.org/10.1073/pnas.1105107108> (2011).
20. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb.prot5448, <https://doi.org/10.1101/pdb.prot5448> (2010).
21. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Research* **38**, e87–e87, <https://doi.org/10.1093/nar/gkp1163> (2009).
22. Spyrou, M. A. *et al.* Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nature Communications* **10**, 4470, <https://doi.org/10.1038/s41467-019-12154-0> (2019).
23. Andrades Valtueña, A. *et al.* The Stone Age Plague and Its Persistence in Eurasia. *Current Biology* **27**, 3683–3691.e3688, <https://doi.org/10.1016/j.cub.2017.10.025> (2017).
24. Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527, <https://doi.org/10.1038/35097083> (2001).
25. Herbig, A. *et al.* MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv*, 050559, <https://doi.org/10.1101/050559> (2016).
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410, [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324> (2009).
28. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, <https://doi.org/10.1186/s13059-016-0918-z> (2016).
29. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684, <https://doi.org/10.1093/bioinformatics/btt193> (2013).

30. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* **16**, 224, <https://doi.org/10.1186/s13059-015-0776-0> (2015).
31. Vianello, D. *et al.* HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum Mutat* **34**, 1189–1194, <https://doi.org/10.1002/humu.22356> (2013).
32. Furtwangler, A. *et al.* Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Sci Rep* **8**, 14075, <https://doi.org/10.1038/s41598-018-32083-0> (2018).
33. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093, <https://doi.org/10.1534/genetics.112.145037> (2012).
34. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413, <https://doi.org/10.1038/nature13673> (2014).
35. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909, <https://doi.org/10.1038/ng1847> (2006).
36. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491–498, <https://doi.org/10.1038/ng.806> (2011).
37. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, <https://doi.org/10.1093/bioinformatics/btu033> (2014).
38. Bos, K. I. *et al.* Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994, <https://doi.org/10.7554/eLife.12994> (2016).
39. Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proceedings of the National Academy of Sciences* **110**, 577, <https://doi.org/10.1073/pnas.1205750110> (2013).
40. Davies, B. *Warfare, State and Society on the Black Sea Steppe, 1500–1700*. (Routledge, 2014).
41. Varlik, N. *Plague and Empire in the Early Modern Mediterranean World*. (Cambridge University Press, 2015).
42. Signoli, M., Séguy, I., Biraben, J.-N., Dutour, O. & Belle, P. Paleodemography and Historical Demography in the Context of an Epidemic: Plague in Provence in the Eighteenth Century. *Population (English Edition, 2002-)* **57**, 829–854, <https://doi.org/10.2307/3246618> (2002).
43. Knauf, S. *et al.* Nonhuman primates across sub-Saharan Africa are infected with the yaws bacterium *Treponema pallidum* subsp. *pertenue*. *Emerging Microbes & Infections* **7**, 1–4, <https://doi.org/10.1038/s41426-018-0156-4> (2018).
44. Arora, N. *et al.* Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nature Microbiology* **2**, 16245, <https://doi.org/10.1038/nmicrobiol.2016.245> (2016).
45. Pětrošová, H. *et al.* Whole Genome Sequence of *Treponema pallidum* ssp. *pallidum*, Strain Mexico A, Suggests Recombination between Yaws and Syphilis Strains. *PLOS Neglected Tropical Diseases* **6**, e1832, <https://doi.org/10.1371/journal.pntd.0001832> (2012).
46. Didelot, X. & Wilson, D. J. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology* **11**, e1004041, <https://doi.org/10.1371/journal.pcbi.1004041> (2015).
47. Strouhal, M. *et al.* Complete genome sequences of two strains of *Treponema pallidum* subsp. *pertenue* from Ghana, Africa: Identical genome sequences in samples isolated more than 7 years apart. *PLOS Neglected Tropical Diseases* **11**, e0005894, <https://doi.org/10.1371/journal.pntd.0005894> (2017).
48. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature Reviews Genetics* **20**, 323–340, <https://doi.org/10.1038/s41576-019-0119-1> (2019).
49. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2**, <https://doi.org/10.1093/ve/vev007> (2016).
50. Kumar, S., Stecher, G., Peterson, D. & Tamura, K. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* **28**, 2685–2686, <https://doi.org/10.1093/bioinformatics/bts507> (2012).
51. Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* **10**, e1003537, <https://doi.org/10.1371/journal.pcbi.1003537> (2014).
52. Baker, B. J. *et al.* The Origin and Antiquity of Syphilis: Paleopathological Diagnosis and Interpretation [and Comments and Reply]. *Current Anthropology* **29**, 703–737, <https://doi.org/10.1086/203691> (1988).
53. Patel, S., Kasiri, M., Barnes, W. & Mirjalili, S. A. Syphilis: A fresh Persian perspective. *New Zealand Medical Student Journal* (2018).
54. Aufderheide, A. C. & Rodriguez-Martin, C. *The Cambridge Encyclopedia of Human Paleopathology*. (Cambridge University Press, 1998).
55. Meyer, C. *et al.* Syphilis 2001 a palaeopathological reappraisal. *HOMO* **53**, 39–58, <https://doi.org/10.1078/0018-442X-00037> (2002).
56. Harper, K. N. *et al.* On the Origin of the Treponematoses: A Phylogenetic Approach. *PLOS Neglected Tropical Diseases* **2**, e148, <https://doi.org/10.1371/journal.pntd.0000148> (2008).
57. Harper, K. N., Zuckerman, M. K., Harper, M. L., Kingston, J. D. & Armelagos, G. J. The origin and antiquity of syphilis revisited: An Appraisal of Old World pre-Columbian evidence for treponemal infection. *American Journal of Physical Anthropology* **146**, 99–133, <https://doi.org/10.1002/ajpa.21613> (2011).
58. Baker, B. J. *et al.* Advancing the understanding of treponemal disease in the past and present. *American Journal of Physical Anthropology* **n/a**, <https://doi.org/10.1002/ajpa.23988> (2020).
59. Quétel, C. *History of syphilis*. (Johns Hopkins University Press, 1990).
60. Powell, M. L. & Cook, D. C. *The myth of syphilis: the natural history of treponematoses in North America*. (University Press of Florida, 2005).
61. Stirland, A. Pre-Columbian treponematoses in medieval Britain. *International Journal of Osteoarchaeology* **1**, 39–47, <https://doi.org/10.1002/oa.1390010106> (1991).
62. Mays, S., Vincent, S. & Meadows, J. A possible case of treponemal disease from England dating to the 11th–12th century AD. *International Journal of Osteoarchaeology* **22**, 366–372, <https://doi.org/10.1002/oa.1210> (2012).
63. von Hunnius, T. E., Roberts, C. A., Boylston, A. & Saunders, S. R. Histological identification of syphilis in pre-Columbian England. *American Journal of Physical Anthropology* **129**, 559–566, <https://doi.org/10.1002/ajpa.20335> (2006).
64. Armelagos, G. J., Zuckerman, M. K. & Harper, K. N. The Science Behind Pre-Columbian Evidence of Syphilis in Europe: Research by Documentary. *Evolutionary Anthropology: Issues, News, and Reviews* **21**, 50–57, <https://doi.org/10.1002/evan.20340> (2012).
65. Brothwell, D. In *The Myth of Syphilis* (eds Mary Lucas Powell & Della Collins Cook) Ch. 21, 480–496 (University Press of Florida, 2005).
66. Livingstone, F. B. On the Origin of Syphilis: An Alternative Hypothesis. *Current Anthropology* **32**, 587–590, <https://doi.org/10.1086/204004> (1991).
67. Jankauskas, R. In *The Origin of Syphilis in Europe, Before or After 1493?* (eds Olivier Dutour, György Pálfi, Jacques Bérato, & Jean-Pierre Brun) (Centre Archéologique du Var, Toulon, FR, 1994).
68. Knell, R. J. Syphilis in Renaissance Europe: rapid evolution of an introduced sexually transmitted disease? *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**, S174–S176, <https://doi.org/10.1098/rsbl.2003.0131> (2004).
69. Brothwell, D. R. North American treponematoses against the bigger world picture. *The myth of syphilis: the natural history of treponematoses in North America*. Gainesville: University Press of Florida, 480–496 (2005).
70. Organization, W. H. Summary report of a consultation on the eradication of yaws, 5–7 March 2012, Morges, Switzerland. (Geneva, CH, 2012).

71. Giacani, L. & Lukehart, S. A. The Endemic Treponematoses. *Clinical Microbiology Reviews* **27**, 89, <https://doi.org/10.1128/CMR.00070-13> (2014).
72. Mitjà, O. *et al.* Global epidemiology of yaws: a systematic review. *The Lancet Global Health* **3**, e324–e331, [https://doi.org/10.1016/S2214-109X\(15\)00011-X](https://doi.org/10.1016/S2214-109X(15)00011-X) (2015).
73. Saunders, A. C. d. C. M. *A Social History of Black Slaves and Freedmen in Portugal, 1441-1555*. 59 (Cambridge University Press, 1982).
74. Frith, J. Syphilis-its early history and treatment until penicillin, and the debate on its origins. *Journal of Military and Veterans Health* **20**, 49 (2012).
75. von Hutten, U. In *Classic descriptions of disease* (ed Ralph Hermon Major) (Charles C. Thomas, 1978).
76. Ortner, D. J. *Identification of Pathological Conditions in Human Skeletal Remains*. (Academic Press, 2003).
77. Morton, R. S. The Sibbens of Scotland. *Medical History* **11**, 374–380, <https://doi.org/10.1017/S0025727300012515> (1967).
78. von Hunnius, T. E., Yang, D., Eng, B., Wayne, J. S. & Saunders, S. R. Digging deeper into the limits of ancient DNA research on syphilis. *Journal of Archaeological Science* **34**, 2091–2100, <https://doi.org/10.1016/j.jas.2007.02.007> (2007).
79. Bouwman, A. S. & Brown, T. A. The limits of biomolecular palaeopathology: ancient DNA cannot be used to study venereal syphilis. *Journal of Archaeological Science* **32**, 703–713, <https://doi.org/10.1016/j.jas.2004.11.014> (2005).
80. Kolman, C. J., Centurion-Lara, A., Lukehart, S. A., Owsley, D. W. & Tuross, N. Identification of *Treponema pallidum* Subspecies *pallidum* in a 200-Year-Old Skeletal Specimen. *The Journal of Infectious Diseases* **180**, 2060–2063, <https://doi.org/10.1086/315151> (1999).
81. Montiel, R. *et al.* Neonate Human Remains: A Window of Opportunity to the Molecular Study of Ancient Syphilis. *Plos One* **7**, e36371, <https://doi.org/10.1371/journal.pone.0036371> (2012).
82. Bronk Ramsey, C. Methods for Summarizing Radiocarbon Datasets. *Radiocarbon* **59**, 1809–1833, <https://doi.org/10.1017/RDC.2017.108> (2017).

## Acknowledgements

This work was funded by the Max Planck Society, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program Starting Grant agreement number 805268 CoDisEASe (to KIB), Social Sciences and Humanities Research Council of Canada Doctoral Scholarship 752-2016-0442 (to KG) and DAAD Doctoral Scholarship 57299294 (to KG). We thank Aida Andrades-Valtueña and Ron Hübler for technical assistance, and the laboratory technicians of the Max Planck Institute for the Science of Human History for assistance with data generation. We also thank Philip Slavin for informative discussions regarding European plague outbreaks in the 15<sup>th</sup> century.

## Author contributions

K.I.B. and J.K. conceived of the investigation. J.K., Z.M. and R.J. provided archaeological material and context analysis. K.G. performed laboratory work. A.K.L., A.H., K.I.B., K.G., and M.A.S. performed phylogenetic analysis of pathogen data. C.P. performed analysis of human DNA. R.F. performed radiocarbon dating analysis. S.S. performed molecular dating simulations. K.I.B. and K.G. drafted the manuscript with input from all coauthors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-66012-x>.

**Correspondence** and requests for materials should be addressed to A.H. or K.I.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020