

ARTICLE

<https://doi.org/10.1038/s41467-019-12154-0>

OPEN

# Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes

Maria A. Spyrou <sup>1,2,19</sup>, Marcel Keller <sup>1,3,19</sup>, Rezeda I. Tuxhatova<sup>1,4</sup>, Christiana L. Scheib <sup>5</sup>, Elizabeth A. Nelson<sup>1,2</sup>, Aida Andrades Valtueña <sup>1</sup>, Gunnar U. Neumann <sup>1</sup>, Don Walker <sup>6</sup>, Amelie Alterauge<sup>7</sup>, Niamh Carty<sup>6</sup>, Craig Cessford <sup>8</sup>, Hermann Fetz<sup>9</sup>, Michaël Gourvenec <sup>10</sup>, Robert Hartle<sup>6</sup>, Michael Henderson<sup>6</sup>, Kristin von Heyking<sup>3</sup>, Sarah A. Inskip <sup>11</sup>, Sacha Kacki<sup>12,13</sup>, Felix M. Key<sup>14</sup>, Elizabeth L. Knox<sup>6</sup>, Christian Later<sup>15</sup>, Prishita Maheshwari-Aplin<sup>8</sup>, Joris Peters<sup>3,16</sup>, John E. Robb<sup>8</sup>, Jürgen Schreiber<sup>17</sup>, Toomas Kivisild <sup>5,18</sup>, Dominique Castex<sup>12</sup>, Sandra Lösch <sup>7</sup>, Michaela Harbeck<sup>3</sup>, Alexander Herbig <sup>1</sup>, Kirsten I. Bos <sup>1</sup> & Johannes Krause <sup>1,2</sup>

The second plague pandemic, caused by *Yersinia pestis*, devastated Europe and the nearby regions between the 14<sup>th</sup> and 18<sup>th</sup> centuries AD. Here we analyse human remains from ten European archaeological sites spanning this period and reconstruct 34 ancient *Y. pestis* genomes. Our data support an initial entry of the bacterium through eastern Europe, the absence of genetic diversity during the Black Death, and low within-outbreak diversity thereafter. Analysis of post-Black Death genomes shows the diversification of a *Y. pestis* lineage into multiple genetically distinct clades that may have given rise to more than one disease reservoir in, or close to, Europe. In addition, we show the loss of a genomic region that includes virulence-related genes in strains associated with late stages of the pandemic. The deletion was also identified in genomes connected with the first plague pandemic (541–750 AD), suggesting a comparable evolutionary trajectory of *Y. pestis* during both events.

<sup>1</sup>Max Planck Institute for the Science of Human History, 07745 Jena, Germany. <sup>2</sup>Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany. <sup>3</sup>SNSB, State Collection for Anthropology and Palaeoanatomy Munich, 80333 Munich, Germany. <sup>4</sup>Laboratory of Structural Biology, Kazan Federal University, Kazan, Russian Federation 420008. <sup>5</sup>Institute of Genomics, University of Tartu, Riia 23b, 51010 Tartu, Estonia. <sup>6</sup>MOLA (Museum of London Archaeology), London N1 7ED, UK. <sup>7</sup>Department of Physical Anthropology, Institute for Forensic Medicine, University of Bern, 3007 Bern, Switzerland. <sup>8</sup>Department of Archaeology, University of Cambridge, Downing St, Cambridge CB2 3ER, UK. <sup>9</sup>Archaeological Service, State Archive Nidwalden, 6371 Nidwalden, Switzerland. <sup>10</sup>Archeodunum SAS, Agency Toulouse, 8 allée Michel de Montaigne, 31770 Colomiers, France. <sup>11</sup>McDonald Institute for Archaeological Research, University of Cambridge, Downing St, Cambridge CB2 3ER, UK. <sup>12</sup>PACEA, CNRS Institute, Université de Bordeaux, 33615 Pessac, France. <sup>13</sup>Department of Archaeology, Durham University, South Rd, Durham DH1 3LE, UK. <sup>14</sup>Institute for Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>15</sup>Bavarian State Department of Monuments and Sites, 80539 Munich, Germany. <sup>16</sup>ArchaeoBioCenter and Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig Maximilian University Munich, Kaulbachstr. 37/III, 80539 Munich, Germany. <sup>17</sup>Dig it! Company GbR, 86971 Peiting, Germany. <sup>18</sup>Department of Human Genetics, Katholieke Universiteit Leuven, 3000 Leuven, Belgium. <sup>19</sup>These authors contributed equally: Maria A. Spyrou, Marcel Keller. Correspondence and requests for materials should be addressed to M.A.S. (email: [spyrou@shh.mpg.de](mailto:spyrou@shh.mpg.de)) or to K.I.B. (email: [bos@shh.mpg.de](mailto:bos@shh.mpg.de)) or to J.K. (email: [krause@shh.mpg.de](mailto:krause@shh.mpg.de))

One of the most devastating pandemics of human history was the second plague pandemic, which began with the infamous Black Death (BD, 1346–1353 AD) and continued with recurrent outbreaks in Europe, the Near East and North Africa until the 18th century AD<sup>1,2</sup>. Its causative agent, *Yersinia pestis*<sup>3</sup>, is a highly virulent bacterium that causes bubonic, pneumonic, and septicaemic plague and today is maintained among wild rodent populations in eastern Europe, Asia, Africa and the Americas<sup>4–6</sup>.

The first historically documented outbreaks of the second pandemic seem to have occurred in 1346 in the Lower Volga and Black Sea regions<sup>1,7</sup>. Subsequently, the bacterium dispersed through the rest of Europe over the next seven years, causing reductions in the human population estimated to be as high as 60%<sup>1</sup>. Recent studies on ancient *Y. pestis* DNA from medieval plague victims have contributed insights into these initial stages of the pandemic. Specifically, mid-14th-century *Y. pestis* genomes reconstructed from Saint-Laurent-de-la-Cabrerisse (southern France)<sup>8</sup>, Barcelona (Spain)<sup>9</sup>, London (England)<sup>10</sup> and Oslo (Norway)<sup>8</sup> were shown to be identical, suggesting the rapid dispersal of a single strain across Europe during the BD. Recently, the analysis of an additional low-coverage genome from Siena, Italy (BSS31)<sup>8</sup>, revealed the purported existence of *Y. pestis* strain diversity during the BD, a possibility that should be further explored.

After the BD, plague was a common scourge in Europe as evidenced by the thousands of recorded epidemics it supposedly caused between 1353 and the late 18th century<sup>2,11</sup>. Whether these were caused by multiple introductions of the disease from an Asian source or by its local persistence in Europe is currently a topic of debate<sup>9,12–14</sup>. While data from climatic proxies are considered as supportive of the former hypothesis<sup>12</sup>, genetic evidence is interpreted in both directions<sup>8,9,13</sup>. To date, ancient *Y. pestis* genomes from epidemics closely succeeding the BD in Europe have been sequenced from late-14th-century individuals in Bergen op Zoom (Netherlands), London (England) and the Middle Volga region of Russia. They cluster on a phylogenetic lineage that is a precursor to strains associated with the 19th-century third plague pandemic<sup>9,15,16</sup>, and thus provide a link between medieval and modern plague. Moreover, *Y. pestis* genomes recovered from Ellwangen, Germany (1485–1627 calAD), and the Great Plague of Marseille in France (L'Observance, 1720–1722 AD) cluster on an independent lineage, here termed the “post-BD” lineage, that is to date unidentified among modern *Y. pestis* diversity. Both lineages descended from the strain associated with the BD and, hence, likely represent plague's legacy in or around Europe after 1353.

At present, the source of the second pandemic and the route that the bacterium followed during its course of entry into Europe remain hypothetical since genomic data from early outbreaks in western Russia have thus far been elusive. In addition, the limited number of published ancient *Y. pestis* genomes<sup>9,10,14</sup> challenges our ability to construct hypotheses regarding the number of lineages responsible for the numerous post-BD outbreaks in Europe<sup>2,11</sup> and whether they derived from a single or multiple disease reservoirs. Here, we take steps to overcome these limitations by expanding the number of available *Y. pestis* genomes from multiple time periods and locations in order to gain additional knowledge on the early stages of the second pandemic, and to study the genetic diversity of the bacterium present in Europe after the BD. Additionally, we present a reanalysis of recently published data from the same time period<sup>8</sup>. Our results support the entrance of *Y. pestis* into Europe through the east during the initial wave of the pandemic and consistently demonstrate an absence of genetic diversity in the bacterium during the BD. Moreover, our genomic analysis of post-BD outbreaks from

central and western Europe suggests the local diversification of an extinct *Y. pestis* lineage between the late-14th and 18th centuries that may have resided in more than one disease reservoir.

## Results

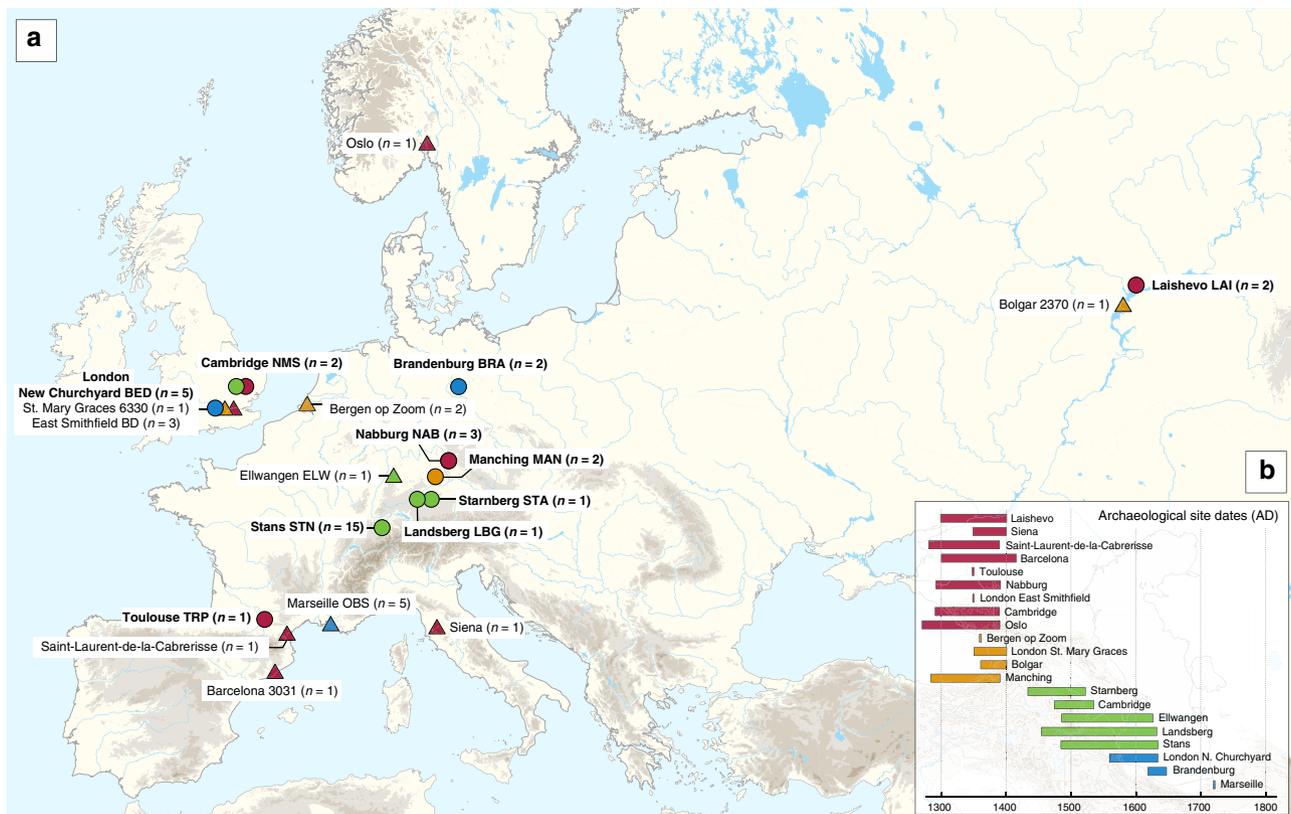
**Sample screening for signatures of *Y. pestis* DNA.** Two approaches were used for the assessment of *Y. pestis* DNA in tooth specimens ( $n = 206$ ) from ten archaeological sites spanning the 14th–17th centuries AD in Europe (Fig. 1, Supplementary Figs. 1–10 and Supplementary Note 1). First, a qPCR screening approach was employed for detection of the *Y. pestis*-specific gene, *pla*, located on the pPCP1 plasmid<sup>17</sup> in 180 specimens from the cities of London ( $n = 40$ ) in England, Toulouse ( $n = 42$ ) in France, Brandenburg an der Havel<sup>13</sup> ( $n = 3$ ), Landsberg am Lech ( $n = 10$ ), Manching-Pichl<sup>13</sup> ( $n = 28$ ), Nabburg ( $n = 12$ ) and Starnberg ( $n = 3$ ) in Germany, Laishevo ( $n = 10$ ) in the Volga region of Russia, and Stans ( $n = 32$ ) in Switzerland. Extracts from 41 teeth across these sites tested positive for *pla* (Supplementary Table 1). All extraction negative controls were free of amplification products. Amplification products from putatively positive individuals were not sequenced, as the presence of *Y. pestis* was subsequently assessed through whole-genome capture and high-throughput Illumina sequencing.

In addition, shotgun Next Generation Sequencing (NGS) data from individuals unearthed at the New Museums site (Augustinian Friary) in Cambridge ( $n = 26$ ) were screened for *Y. pestis* with the MEGAN alignment tool (MALT)<sup>18</sup> (see Methods). The output was post-processed within the pathogen screening pipeline HOPS<sup>19</sup>. The assessment of shotgun NGS reads produced from non-uracil-DNA-glycosylase (non-UDG) libraries revealed the potential presence of *Y. pestis* DNA in four individuals (Supplementary Table 2, Supplementary Fig. 11).

## *Y. pestis* in-solution capture and whole-genome reconstruction.

We prepared UDG-treated libraries<sup>20,21</sup> from all putatively positive samples and used a *Y. pestis* whole-genome in-solution capture approach<sup>22</sup> combined with high-throughput sequencing for the retrieval of 1,299,105–79,055,317 raw reads per sequenced library. All data were mapped against the *Y. pestis* CO92 reference genome (NC\_003143.1)<sup>3</sup>. This resulted in 86,278–3,822,030 unique mapping reads yielding 1.1–80.1-fold coverage across 34 individuals that span the time transect between the 14th and 17th centuries in Europe (Supplementary Table 3). More specifically, we could retrieve two *Y. pestis* genomes from Cambridge (England), five from London (England), one from Toulouse (France), three from Nabburg, two from Manching-Pichl<sup>13</sup>, one from Starnberg, one from Landsberg am Lech, two from Brandenburg an der Havel<sup>13</sup> (all from Germany), two from Laishevo (Russia) and 15 from Stans (Switzerland). Of those, 24 isolates showed at least 50% of the reference genome covered at 5-fold (Table 1), which allowed for their confident inclusion in phylogenetic analysis. In addition, we nearly tripled the genomic coverage of the published “549\_O” isolate from Ellwangen, Germany (now reaching 14.1-fold), which was previously processed by array-based capture using a different probe design<sup>9</sup> (Supplementary Table 3).

***Y. pestis* phylogenetic reconstruction.** To infer genetic relationships between the new and previously published *Y. pestis* isolates, we constructed phylogenies using the maximum likelihood (ML) method, allowing for up to 3% missing data (97% partial deletion) to accommodate lower coverage genomes. As a reference dataset, we used 233 modern isolates<sup>23–27</sup> (as listed in ref. 28), which represent most of the published *Y. pestis* genetic diversity. In addition, we included previously published second



**Fig. 1** Archaeological site locations and chronologies. **a** Map showing the geographic locations of archaeological sites from which second pandemic (14<sup>th</sup>- to 18<sup>th</sup>-century AD) *Y. pestis* genomes have been reconstructed ( $\geq 1$ -fold). The number (*n*) of genomes obtained from each site is shown in brackets. Locations of previously published genomes appear in triangles, whereas genomes that are newly described in this study appear in circles (labels in bold). Base map purchased from [vectormaps.de]. **b** Specimen chronologies combining archaeological and radiocarbon dates of previously published and new second plague pandemic isolates (see Supplementary Note 1 and Supplementary Table 3)

pandemic isolates ( $n = 15$ )<sup>8-10,14</sup>, a 6th-century AD isolate from Germany<sup>29</sup>, a 2nd- to 3rd-century AD isolate from the Tian Shan mountains in Kyrgyzstan<sup>30</sup>, as well as three Bronze Age isolates from the Altai and Volga regions<sup>31,32</sup> (Supplementary Fig. 12).

All newly reconstructed genomes appear on Branch 1 and are closely related to the previously published second pandemic isolates from Europe (Fig. 2), thus confirming their authenticity. In addition, they seem to represent a diverse group of strains that were present across Europe between the 14th and 18th century AD (Fig. 2, Supplementary Data 1). A number of genomes (NAB005, BRA003, STN011 and STN004) were excluded from further analyses as they showed evidence of excess heterozygosity, which is atypical of bacterial genomes (Supplementary Fig. 13). This phenomenon likely arises from enrichment of non-target DNA stemming from closely related organisms, an issue frequently encountered in ancient metagenomic datasets<sup>18,29,33</sup>. Moreover, these genomes had notably longer branch lengths in comparison to other contemporaneous isolates from the same archaeological contexts (Supplementary Fig. 14). Their assessment using the recently developed SNPEvaluation tool<sup>28</sup> (see Methods) classified their private SNP calls as false-positive, suggesting that the observed branch lengths are erroneous (Supplementary Data 2). Similarly, the previously published SLC1006 and BSS31 genomes<sup>8</sup> were also excluded from further analyses as they also showed high heterozygosity (Supplementary Fig. 15) and exceedingly longer branch lengths compared to other 14th-century *Y. pestis* genomes (Supplementary Figs. 14 and 16).

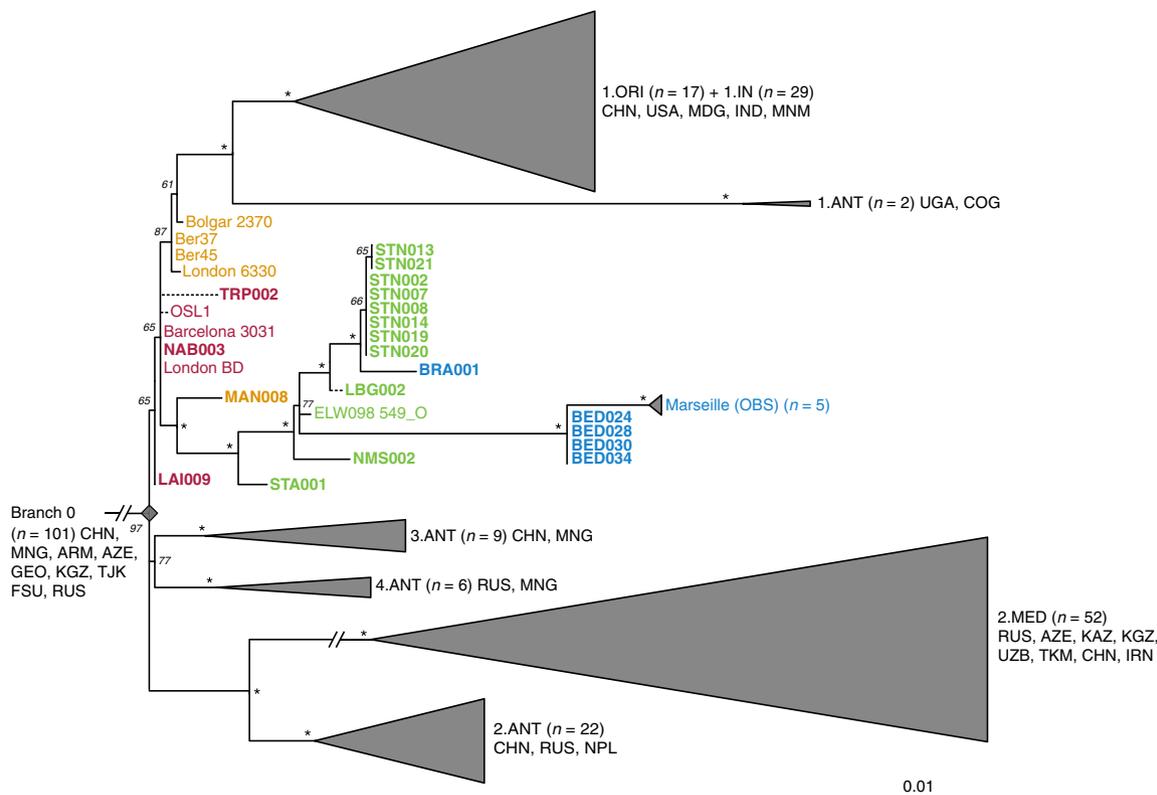
Our phylogenetic reconstruction shows that the LAI009 isolate from Laishevo is ancestral to the BD isolates from southern, central, western and northern Europe, as well as to the previously

published late 14th-century isolates from London (6330)<sup>10</sup> and Bolgar City<sup>9</sup> (Fig. 2). This genome possesses only one derived SNP distinguishing it from the N07 polytomy that gave rise to Branches 1–4 (Fig. 2; Supplementary Data 1)<sup>23</sup>. Since all other second pandemic genomes share an additional derived SNP on Branch 1, we interpret LAI009 as the most ancestral form of the strain that entered Europe during the initial wave of the second pandemic that has been identified to date. Regarding the central and western European genomes, NAB003 from Nabburg does not show differences compared to previously published BD genomes from London and Barcelona<sup>9,10</sup>. In addition, NMS003 from Cambridge was genotyped based on inspection of its SNP profile, despite it not fulfilling the genomic coverage criteria for inclusion in our phylogenetic analysis (Supplementary Table 3), as its archaeological context makes it distinct from other *Y. pestis*-positive individuals from the same site (see Supplementary Note 1). As a result, SNP inspection classified it as potentially identical to other BD genomes (Supplementary Data 3). By contrast, certain isolates associated with the BD period are seemingly distinct. For example, TRP002 from Toulouse, which dates to 1347–1350 based on archaeological evidence, forms its own unique branch (Fig. 2; Supplementary Data 1). Qualitative assessment of eight unique SNPs in TRP002 with SNPEvaluation<sup>28</sup> classified them as potential false-positives (see Methods, Supplementary Data 2). In addition, after visual inspection, all such variants appear in regions of the genome where reads from diverse sources seem to be mapping (Supplementary Fig. 17) and, therefore, were considered to be of exogenous origin. Similarly, we assessed one unique SNP identified in our re-analysis of the recently published OSL-1 genome from Oslo, Norway<sup>8</sup> (Fig. 2).

**Table 1** Post-capture sequencing statistics of all new *Yersinia pestis* genomes that passed quality criteria for inclusion in phylogenetic analysis

Sample name	Site name	Date (AD)	Uniquely mapping reads	Endogenous DNA post enrichment (%)	Mean fold coverage	Genome covered $\geq 5$ -fold (%)	Average fragment length (bp)	GC content (%)
BED030.A0102	New Churchyard, London	1560-1635 <sup>a</sup>	3,624,482	36.2	80.1	93.6	102.9	48.5
BED028.A0102	New Churchyard, London	1560-1635 <sup>a</sup>	2,665,238	22.2	37.2	91.4	65.0	49.0
BED034.A0102	New Churchyard, London	1560-1635 <sup>a</sup>	1,371,698	10.5	18.3	89.1	62.2	49.2
BED024.A0102	New Churchyard, London	1560-1635 <sup>a</sup>	1,000,524	18.1	12.6	84.7	58.5	49.1
BRA001.A0101	Domlinden 12, Brandenburg an der Havel	1618-1648 <sup>b</sup>	2,387,557	23.2	23.8	92.0	46.4	47.5
LA1009.A0101	Laishevo III, Laishevo	1300-1400 <sup>b</sup>	2,549,926	23.9	28.4	92.1	51.8	48.0
LBG002.A0101	Kirchhof St. Johannes, Landsberg	1455-1632 <sup>a</sup>	621,713	27.9	7.2	66.4	54.2	49.9
MAN008.B0101	St. Leonhardi, Manching-Pfchl	1283-1390 <sup>a</sup>	1,974,399	44.9	25.8	88.7	60.8	50.8
NAB003.B0101	Sankt Johans Freidhof, Nabburg	1292-1392 <sup>a</sup>	684,029	33.5	8.1	70.2	54.8	49.7
NIMS002.A0101	Augustinian Friary (New Museums site), Cambridge	1475-1536 <sup>a,b</sup>	855,185	2.3	12.5	94.8	67.9	47.3
STA001.A0101	Possenhofer Str. 3, Starnberg	1433-1523 <sup>a,b</sup>	1,110,049	9.9	11.7	84.3	49.0	44.7
STN014.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	3,822,030	48.2	55.3	93.0	67.3	48.9
STN020.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	2,020,769	44.3	28.2	90.3	64.8	48.5
STN021.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	1,588,442	35.1	21.7	88.6	63.7	48.5
STN019.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	1,325,076	35.3	18.7	87.1	65.8	49.1
STN007.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	1,293,507	32.8	18.0	86.7	64.8	49.4
STN002.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	935,795	27.6	12.7	83.3	63.0	48.4
STN008.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	875,153	30.2	11.7	77.7	62.5	50.1
STN013.A0101	Nägelgasse, Stans	1485-1635 <sup>a</sup>	714,482	24.5	9.2	73.8	59.9	49.0
TRP002.A0101	Trente-Six Ponts 16, Toulouse	1347-1350 <sup>b</sup>	632,303	19.8	5.9	50.9	43.2	48.7

<sup>a</sup>Dates based on radiocarbon dating of collagen<sup>b</sup>Dates based on archaeological context information

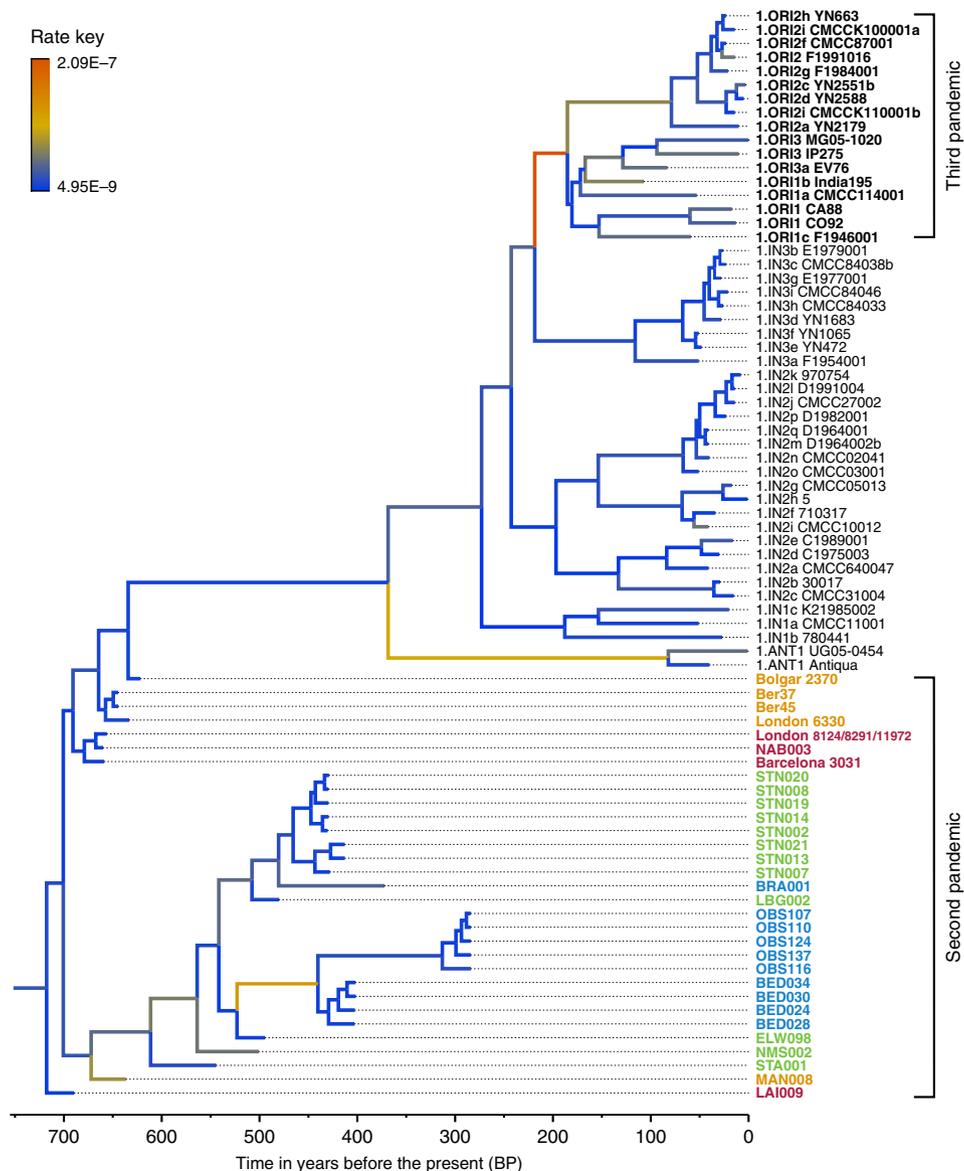


**Fig. 2** Phylogenetic positioning of second pandemic strains. A maximum likelihood phylogeny was generated allowing for up to 3% missing data (97% partial deletion) and considering a total of 6,058 single nucleotide polymorphisms (SNPs). The image shows a graphical representation of Branches 1–4 (see Supplementary Fig. 12 for a complete phylogeny), to emphasise the phylogenetic positioning of the new and previously published second pandemic strains (labels of new 14th- to 17th-century strains appear in bold). Dashed branches denote uncertainty in the private SNP calls of the respective genomes. Sub-clades of published genomes are collapsed to enhance tree visibility. Numbers (*n*) in brackets indicate the number of strains represented in each collapsed branch. Node support was estimated using 1,000 bootstrap iterations. Nodes that have bootstrap values of  $\geq 95$  are indicated by asterisks (\*). Scale denotes substitutions per site. Geographic abbreviations of modern strain isolation locations are as follows: China (CHN), United States of America (USA), Madagascar (MDG), India (IND), Myanmar (MNM), Congo (COG), Uganda (UGA), Mongolia (MNG), Nepal (NPL), Iran (IRN), Kazakhstan (KAZ), Kyrgyzstan (KGZ), Tajikistan (TJK), Armenia (ARM), Georgia (GEO), Azerbaijan (AZE), Uzbekistan (UZB), Turkmenistan (TKM), Russia (RUS) and unspecified regions of the Former Soviet Union (FSU)

Visual inspection revealed it as a low-quality C-to-T transition that could be confined by aDNA damage (Supplementary Fig. 18). Finally, despite exclusion of BSS31 (Siena, Italy) from phylogenetic analysis, two previously identified unique SNPs in this genome were manually inspected, since they were presented as evidence for *Y. pestis* genetic diversity in Europe during the BD<sup>8</sup>. Importantly, BLASTn analysis of reads overlapping those regions (Supplementary Fig. 18, Supplementary Data 4 and 5) showed a 100% identity to environmental or other enteric bacterial species, but not to *Y. pestis*. We, hence, conclude that apart from LAI009 all reconstructed genomes associated with the initial pandemic wave have identical genotypes. In addition, we note that structural rearrangements could provide alternative means of genetic diversity. Although architectural differences are vastly abundant among modern *Y. pestis* genomes<sup>34</sup>, their assessment in ancient *Y. pestis* is limited by the short read aDNA data produced here.

We find a number of genomes grouping with the previously described “post-BD” lineage together with published strains from Ellwangen (ELW098/549\_O), Germany (1486–1630)<sup>9</sup>, and Marseille, France (1720–1722)<sup>14</sup>, which are descended from the European BD isolates (Fig. 2; Supplementary Data 1). Here, we identify the earliest evidence of this lineage in a 14th-century isolate from Manching-Pichl (MAN)<sup>13</sup> (see Supplementary

Note 1), which is followed by the more derived 15th- to 17th-century isolates from Starnberg (STA), Landsberg (LBG), Stans (STN) and Cambridge (NMS), as well as the 17th-century Brandenburg an der Havel (BRA)<sup>13</sup> and London (BED), all of which provide further evidence for plague’s continuous presence in Europe after the BD. Of note, we retrieved eight nearly identical genomes from Stans (STN, maximum one SNP difference in two of eight genomes; mean SNP distance  $d = 0$ ), and together with the four identical genomes from 17th-century London (BED) ( $d = 0$ ), the five previously published nearly identical genomes from Marseille (OBS, maximum one SNP difference in one of five genomes,  $d = 0$ ), and the seven identical BD isolates from various regions in Europe ( $d = 0$ ), our results demonstrate low genetic diversity of the bacterium within local outbreaks and/or major epidemics of the second pandemic. In addition, we find that this “post-BD lineage” gave rise to (at least) two distinct clades within Europe, with the Ellwangen isolate being positioned closest to an apparent population split (Fig. 2). From this divergence, one clade gave rise to the strains associated with outbreaks in Germany and Switzerland (15th–17th century AD), and the second encompassed strains from 17th-century London (BED) and 18th-century Marseille (OBS). Notably, these two clades show dissimilar rates of substitution accumulation. For example, the mean SNP distance between the Ellwangen genome



**Fig. 3** Substitution rate variation across the *Y. pestis* Branch 1. The figure presents a maximum clade credibility (MCC) phylogenetic tree generated using BEAST v1.8<sup>85</sup> (rooted with 2.MED KIM10—outgroup not shown). The tree was viewed in FigTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>), and modified so that branch colours represent mean substitution rates (substitutions per site per year). The tree depicts the substitution rate variation across Branch 1 of the *Y. pestis* phylogeny, which ranges from 2.09E-7 (highest-red) to 4.95E-9 (lowest-blue) substitutions per site per year (see rate key). The isolates used for this analysis overlap with the ones used for the SNP and maximum likelihood phylogenetic analysis (see Supplementary Fig. 12), with the exception of the TRP002 and OSL1 genomes since their private SNP calls are likely affected by environmental contamination and other representative genomes exist in our dataset from the BD time period (1346–1353 AD). Labels of genomes associated with the second and third plague pandemics appear in bold. The mean substitution rate across the tree (including 2.MED KIM10) was calculated to 2.85E-8 substitutions per site per year. Lengths of branches are scaled to represent sample ages, and the depicted Branch 1 sequences are estimated to represent 731 years (95% HPD: 672–823) of *Y. pestis* evolution. The time scale is shown in years before the present (BP), where present denotes the most recently isolated modern *Y. pestis* strain (year 2005)

(ELW098/549\_O) and the London (BED) genomes ( $d = 45$ ) is double that observed between Ellwangen and Brandenburg (BRA,  $d = 22$ ), despite an assumption of them being contemporaneous (early 17th century AD) based on archaeological dating (Fig. 2; Supplementary Table 1; Supplementary Note 1).

**Analysis of substitution rate variation in *Y. pestis*.** We used the Bayesian framework BEAST v1.8 in order to make an assessment of substitution rate variations across the genealogy of Branch 1 ( $n = 80$ ), retaining high-quality second pandemic *Y. pestis* genomes and using available calibration points in our modern and

ancient datasets (Supplementary Data 6). Previous studies have demonstrated that overdispersion among *Y. pestis* branch lengths is unlikely a result of natural selection, and have rather suggested a link between rate acceleration and geographic expansion of certain lineages during epidemic spread<sup>16,23</sup>. Our analysis based on the coalescent skyline model (Fig. 3, Supplementary Fig. 19) suggests an over 40-fold difference between the fastest and slowest substitution rates identified on Branch 1 (Fig. 3). In particular, we observe the fastest rates in three internal branches (Fig. 3). The first spans the genetic distance between the strains from Ellwangen (549\_O) and London (BED), and supports the conflicting branch lengths of BED and BRA strains described



previously characterised historical strains are influenced by the capture design used for their retrieval. Specifically, the second pandemic genomes “Bolgar 2370” and “Barcelona 3031” (ref. 9) and the first pandemic genome “Altenerding 2148” (ref. 29) seem to lack coverage in certain *Y. pestis*-specific regions, since *Yersinia pseudotuberculosis* was previously used as a probe-design template for their enrichment<sup>9,29</sup> (Fig. 4a). Regarding the newly reconstructed strains, we find that most possess all analysed genes with the exception of the New Churchyard (BED) and Marseille (OBS) strains that lack the magnesium transporter genes *mgtB* and *mgtC*, as well as the Cambridge (NMS002) strain that is lacking the *inv* gene (Fig. 4). While invasins are associated with epithelial colonisation of *Y. pseudotuberculosis* and *Yersinia enterocolitica*, it is known to have been inactivated in *Y. pestis*<sup>38</sup>. By contrast, magnesium transporters are considered vital for *Y. pestis* intracellular survival under low Mg<sup>2+</sup> conditions<sup>39</sup>, such as those encountered within macrophage phagosomes. Specifically for *Y. pestis*, *mgtB* disruption has been associated with a decreased ability for macrophage invasion resulting in its attenuated virulence in mice<sup>40</sup>. Both *mgtB* and *mgtC* are present in all 233 modern *Y. pestis* genomes used in our comparative dataset. We explored these gene deletions in greater detail using BWA-MEM and identified them as part of a 49-kb missing region within the BED and OBS genomes (1,879,467–1,928,869 on CO92) (Fig. 4b, Supplementary Fig. 21) flanked by an *IS100* element immediately following its downstream end, which is consistent with previously characterised disruptions or losses of *Y. pestis* genomic regions via insertion elements<sup>41</sup>. Apart from *mgtB* and *mgtC*, this region encompasses a set of 34 additional genes that code for both characterised and hypothetical proteins, most of which seem to be associated with phenotypic characteristics that appear inactivated in *Y. pestis* such as motility and chemotaxis as well as few genes associated with metabolism, structure synthesis and environmental stress response (Supplementary Fig. 21, Supplementary Table 4). In addition, the clade encompassing this deletion is associated with some of the late outbreaks of the second plague pandemic, i.e. during the 17th century in London, England (BED) (see Supplementary Note 1), and during the 18th-century Plague of Marseille, in France (OBS 1720–1722 AD)<sup>14</sup>, which was one of the last major epidemics that occurred in continental Europe<sup>42</sup>. Intriguingly, a nearly identical genomic deletion (45 kb), also including the *mgtB* and *mgtC* virulence-associated genes, was recently identified in ancient isolates from France (LVC, LSD)<sup>28</sup> sequenced from victims of the first plague pandemic (6th–8th centuries AD)<sup>28</sup>. These genomes are described elsewhere and date within a wide temporal interval (550–650 AD), though based on existing data they appear to be the youngest first pandemic isolates sequenced to date<sup>28</sup>.

## Discussion

A series of studies have sufficiently demonstrated the preservation of *Y. pestis* in ancient human remains from a wide temporal transect<sup>8–10,14,22,29,31,32,43</sup>. This study presents an extensive sampling of multiple European epidemic burials from the period between the 14th and 17th centuries in order to gain a more complete picture of *Y. pestis*' genetic history during the second plague pandemic. Here, we nearly triple the amount of genomic data available from that time period (Fig. 1, Table 1 and Supplementary Table 3) and integration with existing datasets reveals key aspects regarding the initiation and progression of the second plague pandemic in Europe.

Based on historical sources alone, it has been difficult to determine the time at which *Y. pestis* first reached different parts of western Russia<sup>7</sup>. A commonly accepted view dates its arrival in the southwest, particularly in cities of Astrakhan and Sarai, in

1346<sup>1,44</sup> with subsequent spread into southern Europe from the Crimean peninsula. On the other hand, the dispersal of plague into northwestern Russia (i.e. in the cities of Pskov and Novgorod<sup>7,44</sup>) may have followed an alternative route via the Baltic Sea, occurring at the end of the BD between 1351 and 1353<sup>1,7,44</sup>. Such a notion of plague's expansion from northern Europe eastwards is also supported by published ancient genomic data from the late 14th-century Middle Volga region of Russia<sup>9</sup>, though other scenarios may come to light with incorporation of additional genomic and historical data. Importantly, through analysis of our new strain from Laishevo (LAI009), which is phylogenetically ancestral to all second pandemic strains sequenced to date (Fig. 2), we provide evidence for the bacterium's presence in the same region, ~2000 km northeast of the Crimean peninsula, prior to reaching southern Europe in 1347–1348<sup>1</sup> (currently represented by strains from Siena, Saint-Laurent-de-la-Cabrerisse, Barcelona and Toulouse<sup>8,9</sup>). These results suggest that the N07-derived SNP previously termed “p1”<sup>9</sup> (Fig. 2, Supplementary Fig. 12), that is common to all other second pandemic strains, was likely acquired within Europe during the onset of the BD. In addition, given the proximity of the LAI009 genome to the N07 node often associated with the initiation of the BD (Fig. 2, Supplementary Fig. 12)<sup>23</sup>, further data will be necessary to accurately re-evaluate the geographic origin of Branch 1. Previous analyses have proposed East Asia as the mostly likely candidate for the N07 polytomy<sup>10,23</sup> (Fig. 2). Such claims, however, cannot yet be verified given; (1) the apparent East Asian sampling bias of modern isolates<sup>23,45</sup>, (2) the lack of molecular evidence from East Asia dating to the early 14th century and (3) the scarcity of historical documentary sources from this region describing precise disease symptoms<sup>46</sup>. In addition, recently published modern *Y. pestis* genomes from Central Asia show a rich diversity in the local plague foci<sup>26,27</sup>, and further sampling from these regions has the potential to inform hypotheses on plague movement and evolution.

The identification of low genomic diversity during the initial wave of the second pandemic becomes particularly informative when attempting to reconstruct the spread of plague after 1353. Previous research based on climatic proxies<sup>12</sup> as well as PCR<sup>47</sup> and genomic<sup>8</sup> data have proposed multiple introductory waves of *Y. pestis* into Europe as the main source for the post-BD outbreaks recorded until the 18th century. Here, using previously published<sup>8–10,14</sup> and new whole-genome data from 20 archaeological sites, we identify that all genomes associated with post-BD outbreaks in Europe derived from a single ancestral strain that was present in southern, central, western and northern Europe during the BD. We, therefore, interpret the current data as supporting a single entry of *Y. pestis* during the BD, though additional interpretations may arise through the discovery of unsampled diversity in western Eurasia. Subsequent to its entry, we observe the formation of two sister lineages (Fig. 2). The first lineage is responsible for the bacterium's possible eastward expansion after the BD. It contains strains from late-14th-century Bergen op Zoom, London (6330)<sup>10</sup> and the city of Bolgar (2370)<sup>9</sup>, as well as extant strains from Africa (1.ANT)<sup>48</sup>, and, most importantly, a worldwide set of isolates associated with the third pandemic (1.ORI, 19th–20th centuries)<sup>15,16,23</sup> (Fig. 2). The second, here termed the “post-BD lineage”, is characterised by a profound genomic diversity identified within Europe that seems to have been restricted to the second pandemic, as no modern descendants have been identified for this lineage to date. It is represented by historical genomes isolated from 14th- to 18th-century Germany (MAN, STA, ELW, LBG and BRA), Switzerland (STN), England (NMS, BED) and France (OBS) (Fig. 2), suggesting that it persisted in Europe or its vicinity and caused infections over a wide geographic range. The fact that this lineage

has no identified modern descendants is likely related to the disappearance of plague from Europe in the 18th century, possibly due to extinction of local reservoirs, as previously suggested<sup>9</sup>.

We find that the “post-BD lineage” gave rise to (at least) two distinct clades that separate the strains identified in Central Europe during the 15th–17th centuries, and those identified in 17th- to 18th-century England and France. Their distinction is corroborated not only by their genetic and geographic separation (Fig. 2), but also by potential differences in their genomic profiles (Fig. 4) and substitution rates (Fig. 3). The clade that exhibits a slower substitution rate is mainly represented by temporally and genetically closely related isolates from Germany and Switzerland (Fig. 2), which could indicate endemic circulation of the bacterium in that region. Such an observation may be compatible with the hypothesis of an Alpine rodent reservoir facilitating the spread of plague in Central Europe after the BD<sup>49</sup>, although a possible sampling bias should be noted since the majority of our data derive from this region. On the other hand, the clade that exhibits a faster substitution rate (Fig. 3) appears to have had a wider geographic distribution. Given that both Marseille and London were among the main maritime trade centres in Europe during that time, it is plausible that introduction of the disease in these areas occurred via ships<sup>50</sup>, although sources favouring local epidemic eruptions also exist<sup>51</sup>. Previous studies have demonstrated that transmission of *Y. pestis* via steamships during the 19th century played a significant role in initial introduction of the bacterium to several regions worldwide, such as in Madagascar where it persists until today<sup>15,16,52,53</sup>. As such, the possibility of maritime introductions of plague into London and Marseille during the second pandemic vastly expands the breadth of potential geographic source(s) for these strains. Nevertheless, the phylogenetic positioning of the BED and OBS genomes within the “post-BD lineage” and in relation to other second pandemic isolates suggests they arose within Europe or its vicinity.

We identified a 49-kb deletion within both BED and OBS genomes (Fig. 4b), which caused the loss of two virulence-associated genes, *mgtB* and *mgtC* (Fig. 4a). This deletion could not be identified in other second pandemic or modern strains in our dataset (Supplementary Fig. 21). The inferred virulence potential of *mgtB* and *mgtC* genes is associated with intracellular survival of *Y. pestis* within macrophages<sup>40,54</sup>. Their co-expression has been shown to affect the virulence exerted by other pathogenic enterobacteria under laboratory conditions<sup>55,56</sup> and both genes have been proposed as potential drug targets<sup>40,57</sup>. Moreover, the function of *mgtB* was shown to be temperature-dependent, being active at 37 °C but not at 20 °C<sup>58</sup>, suggesting its loss affects the bacterium in warm-blooded hosts. Intriguingly, a 45-kb deletion in the same region was identified in genomes associated with late outbreaks of the first plague pandemic (6th–8th century AD)<sup>28</sup>, which sets it as a candidate for convergent evolution and raises questions regarding its functional importance. Given that all genomes displaying this deletion were obtained from plague victims, including the Great Plague of Marseille (1720–1722 AD) that is known to have caused high mortality, its occurrence may not have reduced the pathogen’s virulence, particularly since genome decay is a well-established characteristic of *Y. pestis* evolution<sup>59,60</sup>. Nevertheless, since both lineages that show this deletion are likely extinct, its functional characterisation will be of importance to evaluate potential effects on maintenance in mammalian and arthropod hosts, in Europe, during the first and second pandemics.

The second plague pandemic has arguably caused the highest levels of mortality of the three recorded plague pandemics<sup>1,61</sup>. It serves as a classic historical example of rapid infectious disease emergence, long-term local persistence and eventual extinction

for reasons that are currently not understood. We have shown that extensive sampling of ancient *Y. pestis* genomic data can provide direct molecular evidence on the genetic relationships of strains present in Europe during that time. In addition, we provide relevant information regarding the initiation and progression of the second pandemic and suggest that a single source reservoir may be insufficient to explain the breadth of epidemics and *Y. pestis*’ genetic diversity in Europe during the 400-year course of the pandemic. Although certain key regions in western Eurasia remain under-sampled for ancient *Y. pestis* DNA, namely the eastern Mediterranean, Scandinavia and the Baltics, vast amounts of high-quality genomic data are becoming increasingly available. Their integration into disease modelling efforts, which consider vector transmission dynamics<sup>62,63</sup>, climatic<sup>12,64,65</sup> and epidemiological data<sup>66</sup>, as well as a critical re-evaluation of historical records<sup>67</sup>, will become increasingly important for better understanding the second plague pandemic.

## Methods

**Tooth sampling, DNA extraction and *Y. pestis* qPCR screening.** Laboratory work was primarily performed in the dedicated aDNA facilities of the Max Planck Institute for the Science of Human History in Jena. Part of the sampling and DNA extractions were performed at aDNA facilities of the ArchaeoBioCenter of the Ludwig Maximilian University of Munich and aDNA facilities of the University of Cambridge, Department of Archaeology.

One-hundred and eighty teeth from nine sites located in England (BED), France (TRP), Germany (NAB, MAN, STA, LBG, BRA), Russia (LAI) and Switzerland (STN) spanning the 14th–17th centuries (see Supplementary Note 1) were sectioned in the cemento-enamel junction, and 30–70 mg of powder was removed from the surface of the pulp chamber using a dentist drill. This powder was then used for DNA extraction, using a protocol optimised for the retrieval of short fragments that are characteristic of ancient DNA<sup>68</sup>. Tooth powder was incubated in 1 ml of lysis buffer (0.45 M EDTA, pH 8.0, and 0.25 mg/ml proteinase K) overnight (12–16 h) at 37 °C. Then, DNA was bound to the silica membrane of spin columns using 10 ml of GuHCl-based binding buffer as described before<sup>68</sup>, followed by a purification that was performed using either the MinElute purification kit (Qiagen) or the High Pure Viral Nucleic Acid Large Volume Kit (Roche). DNA was eluted in 100 µl of TET (10 mM Tris-HCl, 1 mM EDTA pH 8.0, 0.05% Tween 20). Extraction blanks and a positive extraction control (cave bear specimen) were taken along for every extraction batch. All extracts were then evaluated for PCR inhibition, by spiking 2 µl of each extract in a qPCR reaction containing a standard of known concentration<sup>17</sup>. None of the extracts showed signs of PCR inhibitions and, therefore, all were tested by qPCR for the presence of the plasminogen activator gene (*pla*), located on the *Y. pestis*-specific pPCP1 plasmid using a published protocol<sup>17</sup>. PCR products were not sequenced as all putatively positive samples were subsequently evaluated through whole-genome enrichment and next-generation sequencing. All extraction and PCR blanks were free of amplification products.

In addition, 26 specimens from the Augustinian Friary in Cambridge (NMS) were sampled and DNA was extracted at the University of Cambridge. Roots were sawed from teeth using a sterile dremel cutting wheel and a UV-irradiated toothbrush was then used to briefly brush the roots with 5% w/v NaOCl. Subsequently, roots were soaked in 6% w/v bleach for 5 min, then rinsed twice with ddH<sub>2</sub>O, and finally soaked in 70% ethanol for 2 min. The roots were then transferred to a sterile paper towel and UV irradiated for 50 min on each side. After irradiation, teeth were weighed and subsequently transferred in 5-ml or 15-ml tubes for DNA extraction. DNA extraction was carried out as follows: 2 ml of EDTA (0.5 M, pH 8.0) and 50 µl of Proteinase K (10 mg/ml) were used for every 100 mg of sample. Extractions were then incubated at room temperature for 72 h. Extracted DNA was concentrated using the Amicon Ultra-15 concentrators with a 30-kDa filter, down to 250 µl. DNA was then purified using the MinElute PCR purification kit (Qiagen) according to manufacturer’s instruction. For the elution step, column-bound DNA was incubated with 100 µl of Elution buffer for 10 min at 37 °C.

**Non-UDG library preparation and metagenomic screening with HOPS.** The following protocol was carried out in the ancient DNA facility of the University of Cambridge, Department of Archaeology.

Non-UDG libraries were prepared for the NMS samples (Augustinian Friary, Cambridge; Supplementary Table 2) with the NEBNext® Library Preparation Kit for 454 (E6070S, New England Biolabs, Ipswich, MA) using a modified version of the manufacturer’s protocol<sup>69</sup>. Adaptors were constructed as previously described<sup>21</sup>. Indexing PCR reactions were set up as follows: 50 µl of DNA library, 1× PCR buffer, 2.5 mM MgCl<sub>2</sub>, 1 mg/ml BSA, 0.2 µM in PE 1.0, 0.2 mM dNTP each, 0.1 U/µl HGS Taq Diamond and 0.2 µM indexing primer, with the following cycling conditions: 5 min at 94 °C, followed by 18 cycles of 30 s each at 94 °C, 60 °C

and 68 °C, with a final extension of 7 min at 72 °C. Amplified products were purified using the MinElute kit (Qiagen) and DNA was eluted in 35 µl EB. The indexed libraries were then quantified using the Quant-iT™ PicoGreen® dsDNA kit (P7589, Invitrogen™ Life Technologies) on the Synergy™ HT Multi-Mode Microplate Reader with Gen5™ software. Subsequent shotgun sequencing of these libraries was carried out on an Illumina NextSeq500 platform (using the High-Output kit 1 × 75 cycle chemistry) at the University of Cambridge Biochemistry DNA Sequencing Facility.

The program MALT (version 0.4.0)<sup>18</sup>, integrated in the pathogen screening pipeline HOPS<sup>19</sup>, was used to assess the presence of *Y. pestis* DNA in the NMS specimens. A custom NCBI RefSeq (November 2017) database was used for running MALT, including all bacterial and viral assemblies marked as complete, a selection of eukaryotic pathogen genomes, as well as the human reference sequence (GRCh38). Genomes with keywords such as “unknown” were removed. A total of 15,361 genomes were retained in the database. Pre-processed shotgun NGS reads (.fastq) were used as input and the parameters were set as follows: 85 for the minimum percentage identity (-minPercentIdentity), 1 for the minimum support (-minSupport), using a top percentage value of 1 (-topPercent), a semi-global alignment mode, and with all remaining parameters set to default. The resulting “.rma6” output files were automatically post-processed with MALTEExtract (in HOPS) against a list of 100 target bacterial pathogen species, and the resulting profiles were qualitatively assessed within HOPS for the number of aligning reads, the read edit distance against different taxa and the presence of aDNA damage patterns<sup>19</sup>.

**UDG library preparation and *Y. pestis* whole-genome capture.** All putative *Y. pestis*-positive samples were subsequently converted into Illumina double-stranded DNA libraries as described before<sup>21</sup>, using a starting volume of 50–60 µl, with an initial USER (New England Biolabs) treatment step, where UDG was used in combination with endonuclease VIII to excise uracil nucleotides that result from post-mortem DNA damage<sup>20,70</sup>. Subsequently, full UDG-treated and partially UDG-treated libraries were quantified on a qPCR using the IS7/IS8 primer combination. Following, a double-indexing step was performed where libraries were split into multiple PCR reactions based on their initial quantification<sup>71</sup>, in order to ensure maximal amplification efficiency. Every reaction was assigned a maximum input of  $2 \times 10^{10}$  DNA molecules. A unique index combination (index primer containing a unique 8-bp identifier) was assigned to every library, and a 10-cycle amplification reaction was used to attach index combinations to DNA library molecules using *Pfu Turbo Cx Hotstart DNA Polymerase* (Agilent). PCR products were purified using the MinElute DNA purification kit (Qiagen), and eluted in TET (10 mM Tris-HCl, 1 mM EDTA pH 8.0, 0.05% Tween 20). After indexing, all libraries were amplified using *Herculase II Fusion DNA Polymerase* (Agilent) to a concentration of 200–300 ng/µl, in order to achieve 1–2 µg of DNA in a total of 7 µl. Products were again purified using the MinElute DNA purification kit (Qiagen), and eluted in TET (10 mM Tris-HCl, 1 mM EDTA pH 8.0, 0.05% Tween 20). In-solution whole-genome *Y. pestis* capture was then performed as described previously<sup>22</sup>, where the following genomes were used as templates for probe design: CO92 chromosome (NC\_003143.1), CO92 plasmid pMT1 (NC\_003134.1), CO92 plasmid pCD1 (NC\_003131.1), KIM10 chromosome (NC\_004088.1), Pestoides F chromosome (NC\_009381.1) and *Y. pseudotuberculosis* IP 32953 chromosome (NC\_006155.1). DNA captures were carried out on 96-well plates. Each sample was either captured in its individual well, or pooled with maximum one more sample from the same site. Capture enrichment was carried out for two rounds, except for sample NMS002 that was captured for one round. Blanks with non-overlapping index combinations were captured together.

**Sequencing and read processing.** After capture, all products were sequenced on an Illumina NextSeq500 platform using (1 × 151 + 8 + 8 cycles or 1 × 76 + 8 + 8 cycles) or on a HiSeq4000 (using 1 × 76 + 8 + 8 cycles or 2 × 76 + 8 + 8 cycles). Preprocessing of de-multiplexed reads was performed on the automated pipeline EAGER (v1.92.55)<sup>72</sup> and involved the removal of Illumina adaptors and read merging using AdapterRemoval v2 (ref. 73), as well as filtering all reads for sequencing quality (minimum base quality of 20) and length (to retrieve only reads ≥ 30 bp). Subsequently, reads were mapped with BWA (version 0.7.12)<sup>74</sup>, implemented in EAGER, against the CO92 reference genome (NC\_003143.1)<sup>3</sup> using stringent parameters (-n 0.1, -l 32) for genome reconstruction and mean coverage calculation and more lenient parameters (-n 0.01, -l 32) for inspection of regions surrounding potential variants. Reads with mapping quality lower than 37 (-q) were removed using SAMtools (<http://samtools.sourceforge.net/>), and PCR duplicates were removed using the MarkDuplicates tool (<http://broadinstitute.github.io/picard/>). Prior to SNP identification, raw pre-processed reads from partially-UDG-treated libraries were trimmed for 2-bp at both ends to remove sites that could be affected by aDNA damage and, subsequently, were re-filtered for length and re-mapped using stringent parameters.

**SNP calling and phylogenetic analysis.** SNP calling was performed using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK v3.5)<sup>75</sup>. Our newly reconstructed genomes were analysed alongside previously published *Y. pestis* genomes, which included a modern-day dataset of 233 genomes<sup>23–27,48</sup> (as listed in

ref. 28), three Bronze Age strains<sup>31</sup>, a 2nd- to 3rd-century AD isolate from the Tian Shan mountains in Kyrgyzstan<sup>30</sup>, one Justinianic strain (Altnerding 2148)<sup>29</sup>, 15 previously published historical genomes from the second plague pandemic<sup>8–10,14</sup> and a *Y. pseudotuberculosis* strain (IP32953)<sup>60</sup> that was used as outgroup for rooting the phylogeny. A vcf file was produced for every genome using the “EMIT\_ALL\_SITES” option, which generated a call for every position present in the reference genome. Furthermore, we used the custom java tool MultiVCFAnalyzer v0.85 (ref. 33) (<https://github.com/alexherbig/MultiVCFAnalyzer>) to produce a SNP table of variant positions across all genomes analysed, using the following parameters: for homozygous alleles, a SNP would be called when covered at least 3-fold with a minimum genotyping quality of 30, and for heterozygous alleles, a variant would be called when 90% of reads would support it. In cases where none of the parameters would be met, an “N” would be inserted in the respective genomic position. In addition, we omitted previously defined noncore regions, as well as annotated repetitive elements, homoplasies, tRNAs, rRNAs and tmRNAs from our SNP analysis<sup>16,23</sup>. In the present dataset, a total of 7,510 variant positions were identified. Subsequently, the annotation as well as the effect of each SNP was determined through the program SnpEff v3.11 (ref. 76).

We used a SNP alignment produced by MultiVCFAnalyzer v0.85 to construct phylogenetic trees using the ML and maximum parsimony (MP) methods. Up to 3% missing data were included in the analysis (97% partial deletion), resulting in a total number of 6,058 SNPs used for phylogenetic reconstruction. The MP phylogeny was produced in MEGA7 (ref. 77) in order to make a first assessment of genome topologies. The ML phylogenies were constructed with the program RAxML (version 8.2.9)<sup>78</sup> using the Generalised Time Reversible (GTR)<sup>79</sup> model with four gamma rate categories and 1000 bootstrap replicates to assess tree topology support.

**Reanalysis of recently published non-UDG *Y. pestis* genomes.** A recent study described the phylogenetic positioning and SNP analysis of five 14th century *Y. pestis* genomes<sup>8</sup>. As these genomes were non-UDG treated, they were reanalysed here using different criteria compared to other second pandemic and modern genomes in our dataset. Read pre-processing and merging was done as described in the above section “Sequencing and read processing”. In addition, read mapping against the CO92 reference genome (NC\_003143.1) was performed using more lenient parameters in BWA<sup>80</sup> (-n 0.01, -l 16) than the ones previously reported<sup>8</sup>, to account for ancient DNA deamination within mapping reads. In our view, the usage of strict BWA mapping parameters for non-UDG data (i.e. -n 0.1) could potentially introduce a reference bias to the analysis, which could in turn affect SNP discovery and phylogenetic inferences. PCR duplicates were removed from all five datasets using MarkDuplicates (<http://broadinstitute.github.io/picard/>) and reads were filtered for mapping quality (q 37) using SAMtools (<http://samtools.sourceforge.net/>). The obtained mean coverage for each of the five re-analysed genomes was: 3.4-fold for BSS31 (27.8% covered 5-fold), 6.7-fold for SLC1006 (59.1% covered 5-fold), 30.5-fold for OSL-1 (91.7% covered 5-fold), 38.1-fold for Ber37 (95.2% covered 5-fold) and 46.1-fold for Ber45 (94.1% covered 5-fold). In addition, the obtained average fragment lengths for the five re-analysed genomes were as follows: 52.2 bp for BSS31, 71.5 bp for SLC1006, 108 bp for OSL-1, 61.9 bp for Ber37 and 69.7 bp for Ber45. Before SNP calling, MapDamage2.0 (ref. 81) was used to rescale base qualities, primarily on the extremities of mapped reads, to account for sites that could potentially be affected by aDNA deamination. Subsequently, SNPs were called using GATK and the resulting vcf files were comparatively assessed in MultiVCFAnalyzer v0.85 (ref. 33) to compile a SNP table including all genomes in the dataset as described in the above section “SNP calling and phylogenetic analysis”.

**Qualitative SNP assessment in UDG-treated data using SNPEvaluation.** A frequent challenge faced when using ancient metagenomic datasets to reconstruct bacterial genomes is a strong environmental signal that can interfere with SNP assignments, especially in low-coverage data<sup>29</sup>. Such an effect can interfere with phylogenetic analyses by creating artificial branch lengths, which can in turn affect evolutionary inferences. As such, in order to avoid erroneous SNP assignments, we qualitatively evaluated all private SNP calls for the newly reconstructed genomes that were used for phylogenetic analysis in this study (minimum 50% of the genome covered 5-fold (Table 1)). We used the recently developed SNPEvaluation tool ([https://github.com/andreasKroepelin/SNP\\_Evaluation](https://github.com/andreasKroepelin/SNP_Evaluation)) to compare the SNP profiles that arise for each dataset under both stringent (BWA parameters -n 0.1, -l 32) and more lenient (BWA parameters: -n 0.1, -l 16) mapping parameters. Subsequently, the region around each SNP was evaluated within a 50-bp window, and was accepted as true when fulfilling the following criteria: (i) the ratio of coverage between the lenient and stringent mapping was not higher than 1.00, (ii) there were no heterozygous positions within this window when considering a high stringency mapping and (iii) no missing regions/bases were observed within close proximity to the identified SNP (see Supplementary Data 2). Note that the above criteria in SNPEvaluation have been determined and evaluated in UDG-treated metagenomic data<sup>28</sup> and, therefore, would need to be re-adapted for non-UDG-treated data that are heavily affected by aDNA deamination.

**Heterozygosity estimates.** Heterozygous variant calls were investigated given the disparity of branch lengths observed in certain newly reconstructed and previously published genomes (see Supplementary Figs. 14 and 16). Our approach takes into account the “haploid” nature of prokaryotic genomes, suggesting that “heterozygous” SNPs could either arise as a result of mixed infections or from erroneous mapping of DNA reads that belong to closely related bacterial contaminants. We performed SNP calling with the UnifiedGenotyper in GATK<sup>75</sup>, using the “EMI-T\_ALL\_SITES” option that generated a call for all positions in the reference genome. We then used MultiVCFAnalyzer v0.85 (ref. <sup>33</sup>) to compile a SNP table of variant positions with allele frequencies 10–90% across our dataset, hence accounting for all ambiguous heterozygous positions. Histograms of allele frequencies for all SNPs with <100% read support were constructed with R v3.4.1 (ref. <sup>82</sup>) using representative genomes from all sites.

**Estimates of substitution rate variation in *Y. pestis*.** In order to calculate the substitution rate variation across *Y. pestis* isolates associated with the second pandemic, we first assessed the temporal signal across Branch 1 that includes all genomes from both the second and third plague pandemics. For this, we computed an ML phylogeny in RaxML<sup>78</sup> using all Branch 1 genomes<sup>3,8–10,14,16,23,48,83,84</sup> (modern + ancient  $n = 79$ ), with the exception of the BD genomes TRP002 and OSL-1 that showed possible environmental contamination to be affecting their SNP calls. In addition, we used the strain 2.MED KIM10 (Branch 2) as outgroup for rooting the tree. Variant positions across this set of genomes were used for the analysis, allowing for up to 3% missing data (550 SNPs). We used TempEst v1.5 (<http://tree.bio.ed.ac.uk/software/tempest/>) for calculating the root-to-tip regression in relation to specimen or sampling ages. The calculated correlation coefficient ( $R$ ) and  $R^2$  values were 0.57 and 0.33, respectively, which permitted the proceeding with molecular dating analysis.

The Bayesian framework BEASTv1.8 (ref. <sup>85</sup>) was used to assess the substitution rate variation across the *Y. pestis* Branch 1 using the 2.MED KIM10 as outgroup. Our BEAUti setup took into consideration all archaeological, radiocarbon and sampling dates of both ancient and modern genomes (Supplementary Data 6) that were used as calibration points for the Bayesian phylogeny. Divergence dates for each node in the tree were estimated as years before the present, where the year 2005 was set as the present since it represents the most recently isolated modern *Y. pestis* strain on Branch 1 (1.ORI MG05). Monophyletic clades were defined based on the ML phylogeny (Supplementary Fig. 12). The GTR<sup>79</sup> substitution model (4 gamma rate categories) and a lognormal relaxed clock (clock rate tested and strict clock rejected in MEGA<sup>77</sup>) were used to set up two separate analyses using the coalescent constant size<sup>86</sup> and coalescent Bayesian skyline<sup>87</sup> demographic models. For each analysis, three independent chains of 50,000,000 states each were carried out and then combined using LogCombiner to ensure run convergence, with 10% burn-in. In addition, we estimated marginal likelihoods to determine the best demographic model for our dataset using path sampling and stepping stone sampling (PS/SS) implemented in BEAST v1.8 (ref. <sup>85</sup>). For this, each of the described runs was carried out for an additional 50,000,000 states (500,000 states divided into 100 steps) using an alpha parameter of 0.3, which determined the coalescent Bayesian skyline model as better fit for the current dataset. The results produced by the run considering this demographic model were then viewed in Tracer v1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>) to ensure all relevant effective sample sizes (ESS) were >200. We used TreeAnnotator<sup>85</sup>, to produce a maximum clade credibility (MCC) phylogeny using the best-fit model with 10% burn-in, which resulted in the processing of 13,503 trees. The MCC tree was viewed and modified in FigTree v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) where branch lengths were represented as a function of age and mean rates were used to colour individual branches. Finally, the skyline plot was produced and viewed using Tracer v1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>) after resampling states at a lower frequency (every 100,000) using LogCombiner<sup>85</sup>.

**Gene presence/absence and deletion analysis.** In order to investigate the virulence-associated gene profiles of the newly reconstructed second pandemic genomes, the highest quality (coverage) genome from every site (LAI009, NAB003, TRP002, MAN008, STA001, NMS002, LBG002, STN014, BRA001, BED030) was used for comparison with each other and with previously published representatives of ancient (London BD 8124/8291/11972, OSL-1 Ber45, London 6330, Bolgar 2370, Barcelona 3031, Ellwangen 549\_O, OBS137, RISE509, RT5, Altenerding 2148) and modern (1.ORI-CO92, 0.PE2-PESTF, 0.PE4-Microtus 91001) *Y. pestis* isolates as well as a *Y. pseudotuberculosis* strain (IP32953). All listed genomes were re-mapped against the CO92 chromosomal reference genome (NC\_003143.1) without the use of a mapping quality filter of ( $-q$  0). The coverage across 80 chromosomal and 43 plasmid virulence-associated and evolutionary determinant genes that were previously defined<sup>36</sup> was calculated using bedtools<sup>88</sup>. The results are plotted in the form of a heatmap using the ggplot2 (ref. <sup>89</sup>) package in R version 3.4.1 (ref. <sup>82</sup>) and can be viewed in Fig. 4. In addition, we used BWA-MEM<sup>80</sup> to explore the precise coordinates of observed gene or region losses in all affected genomes using default parameters. For the visualisation of an identified deletion across BED and OBS isolates, we computed the average coverage across 3,000-bp windows in representative *Y. pestis* genomes from all analysed periods of the second pandemic, and subsequently used the program Circos<sup>90</sup> to produce coverage plots of a 20-fold maximum coverage. The coverage plots were arranged in chronological order as

follows: LAI009, London BD 8124/8291/11972, Ber45, Bolgar 2370, MAN008, STA001, NMS002, ELW098, LBG002, STN014, BRA001, BED030, OBS137 and the reference genome CO92.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Raw sequencing data of the deep-sequenced genomes are available on the European Nucleotide Archive under project accession number PRJEB29990. Other data supporting the findings of the study are available in this article and its Supplementary Information files, or from the corresponding authors upon request.

Received: 18 December 2018 Accepted: 15 August 2019

Published online: 02 October 2019

## References

- Benedictow, O. J. *The Black Death, 1346-1353: The Complete History* (Boydell and Brewer, Woodbridge, UK, and Rochester, N.Y., 2004).
- Biraben, J.-N. *Les Hommes et la peste en France et dans les pays européens et méditerranéens. t. 2, les hommes face à la peste* (Mouton, Paris, 1976).
- Parkhill, J. et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
- Gage, K. L. & Kosoy, M. Y. Natural history of plague: perspectives from more than a century of research. *Annu. Rev. Entomol.* **50**, 505–528 (2005).
- Prentice, M. B. & Rahalison, L. Plague. *Lancet* **369**, 1196–1207 (2007).
- Tikhomirov, E. Epidemiology and Distribution of Plague. in *Plague manual: epidemiology, distribution, surveillance and control*. (eds. Dennis, D. T. et al.) 11–41 (World Health Organisation, Geneva, 1999).
- Alexander, J. T. *Bubonic Plague in Early Modern Russia: Public Health and Urban Disaster* (Johns Hopkins University Press, Baltimore, Maryland, USA, 1980).
- Namouchi, A. et al. Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proc. Natl Acad. Sci. USA* **115**, E11790–E11797 (2018).
- 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 (2016).
- Bos, K. I. et al. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
- Büntgen, U., Ginzler, C., Esper, J., Tegel, W. & McMichael, A. J. Digitizing historical plague. *Clin. Infect. Dis.* **55**, 1586–1588 (2012).
- Schmid, B. V. et al. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc. Natl Acad. Sci. USA* **112**, 3020–3025 (2015).
- Seifert, L. et al. Genotyping *Yersinia pestis* in historical plague: evidence for long-term persistence of *Y. pestis* in Europe from the 14th to the 17th Century. *PLoS ONE* **11**, e0145194 (2016).
- Bos, K. I. et al. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).
- Pollitzer, R. *The Plague No. 26* (World Health Organization, Geneva, 1954).
- Morelli, G. et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).
- Schuenemann, V. J. et al. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* **108**, E746–E752 (2011).
- Vägene, A. J. et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
- Huebler, R. et al. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. Preprint at <https://www.biorxiv.org/content/10.1101/534198v2> (2019).
- Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
- Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
- Andrades Valtueña, A. et al. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691 (2017). e3688.
- Cui, Y. et al. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl Acad. Sci. USA* **110**, 577–582 (2013).
- Kislichkina, A. A. et al. Nineteen whole-genome assemblies of *Yersinia pestis* subsp. *microtus*, including representatives of *Biovvars caucasica*, *talassica*, *hissarica*, *altaica*, *xilingolensis*, and *ulegeica*. *Genome Announc.* **3**, e01342–01315 (2015).

25. Zhgenti, E. et al. Genome assemblies for 11 *Yersinia pestis* strains isolated in the Caucasus region. *Genome Announc.* **3**, e01030–01015 (2015).
26. Kutnyrev, V. V. et al. Phylogeny and classification of *Yersinia pestis* through the lens of strains from the plague foci of Commonwealth of Independent States. *Front. Microbiol.* **9**, 1106 (2018).
27. Eroshenko, G. A. et al. *Yersinia pestis* strains of ancient phylogenetic branch O. ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. *PLoS ONE* **12**, e0187230 (2017).
28. Keller, M. et al. Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750 CE). *Proc. Natl. Acad. Sci. USA* **116**, 12363–12372 (2019).
29. Feldman, M. et al. A high-coverage *Yersinia pestis* genome from a sixth-century Justinianic plague victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).
30. de Barros Damgaard, P. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369 (2018).
31. Rasmussen, S. et al. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
32. Spyrou, M. A. et al. Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234 (2018).
33. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
34. Darling, A. E., Miklós, I. & Ragan, M. A. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet.* **4**, e1000128 (2008).
35. Green, M. H. Putting Africa on the Black Death map: Narratives from genetics and history. *Afriques* (2018). Available at: <http://journals.openedition.org/afriques/2125> (Accessed: 3rd September 2019).
36. Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect. Immun.* **77**, 2242–2250 (2009).
37. Zhou, D. et al. Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, microtus. *J. Bacteriol.* **186**, 5147–5152 (2004).
38. Simonet, M., Riot, B., Fortineau, N. & Berche, P. Invasin production by *Yersinia pestis* is abolished by insertion of an IS200-like element within the *inv* gene. *Infect. Immun.* **64**, 375–379 (1996).
39. Groisman, E. A. et al. Bacterial Mg<sup>2+</sup> homeostasis, transport, and virulence. *Annu. Rev. Genet.* **47**, 625–646 (2013).
40. Ford, D. C., Joshua, G. W., Wren, B. W. & Oyston, P. C. The importance of the magnesium transporter MgtB for virulence of *Yersinia pseudotuberculosis* and *Yersinia pestis*. *Microbiology* **160**, 2710–2717 (2014).
41. Fetherston, J. D. & Perry, R. D. The pigmentation locus of *Yersinia pestis* KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. *Mol. Microbiol.* **13**, 697–708 (1994).
42. Signoli, M., Bello, S. & Dutour, O. [Epidemic recrudescence of the Great Plague in Marseille (May–July 1722): excavation of a mass grave]. *Med. Trop. (Mars)* **58**, 7–13 (1998).
43. Wagner, D. M. et al. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
44. Benedictow, O. J. *The Black Death and Later Plague Epidemics in the Scandinavian Countries: Perspectives and Controversies* (Walter de Gruyter GmbH & Co KG, Warsaw/Berlin, 2016).
45. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
46. Hymes, R. Epilogue: a hypothesis on the East Asian beginnings of the *Yersinia pestis* polytomy. in *Pandemic Disease in the Medieval World: Rethinking the Black Death* Vol. 1 (ed. Green, M. H.) 285–308 (Arc Medieval Press, Kalamazoo and Bradford, 2014).
47. Haensch, S. et al. Distinct clones of *Yersinia pestis* caused the black death. *PLoS Pathog.* **6**, e1001134 (2010).
48. Chain, P. S. et al. Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J. Bacteriol.* **188**, 4453–4463 (2006).
49. Carmichael, A. G. Plague persistence in Western Europe: a hypothesis. in *Pandemic Disease in the Medieval World: Rethinking the Black Death* Vol. 1 (ed. Green, M. H.) 157–191 (Arc Medieval Press, Kalamazoo and Bradford, 2014).
50. Signoli, M., Séguin, I., Biraben, J.-N., Dutour, O. & Belle, P. Paleodemography and historical demography in the context of an epidemic. *Population* **57**, 829–854 (2002).
51. Cummins, N., Kelly, M. & Ó Gráda, C. Living standards and plague in London, 1560–1665. *Econ. Hist. Rev.* **69**, 3–34 (2016).
52. Brygoo, E.-R. Épidémiologie de la peste à Madagascar. *Arch. Inst. Pasteur Madagascar* **35**, 9–147 (1966).
53. Vogler, A. J. et al. Temporal phylogeography of *Yersinia pestis* in Madagascar: insights into the long-term maintenance of plague. *PLOS Negl. Trop. Dis.* **11**, e0005887 (2017).
54. Grabenstein, J. P., Fukuto, H. S., Palmer, L. E. & Bliska, J. B. Characterization of phagosomal trafficking and identification of PhoP-regulated genes important for survival of *Yersinia pestis* in macrophages. *Infect. Immun.* **74**, 3727–3741 (2006).
55. Blanc-Potard, A. B. & Groisman, E. A. The *Salmonella* selC locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J.* **16**, 5376–5385 (1997).
56. Snively, M., Miller, C. & Maguire, M. The *mgtB* Mg<sup>2+</sup> transport locus of *Salmonella typhimurium* encodes a P-type ATPase. *J. Biol. Chem.* **266**, 815–823 (1991).
57. Belon, C. et al. A macrophage subversion factor is shared by intracellular and extracellular pathogens. *PLoS Pathog.* **11**, e1004969 (2015).
58. Snively, M., Florer, J., Miller, C. & Maguire, M. Magnesium transport in *Salmonella typhimurium*: 28Mg<sup>2+</sup> transport by the CorA, MgtA, and MgtB systems. *J. Bacteriol.* **171**, 4761–4766 (1989).
59. Achtman, M. et al. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* **96**, 14043–14048 (1999).
60. Chain, P. S. et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* **101**, 13826–13831 (2004).
61. Perry, R. D. & Fetherston, J. D. *Yersinia pestis*-etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).
62. Dean, K. R. et al. Human ectoparasites and the spread of plague in Europe during the Second Pandemic. *Proc. Natl. Acad. Sci. USA* **115**, 1304–1309 (2018).
63. Keeling, M. K. & Gilligan, C. A. Metapopulation dynamics of bubonic plague. *Nature* **407**, 903–906 (2000).
64. Xu, L. et al. The trophic responses of two different rodent–vector–plague systems to climate change. *Proc. Biol. Sci.* **282**, 20141846 (2015).
65. Xu, L. et al. Wet climate and transportation routes accelerate spread of human plague. *Proc. Biol. Sci.* **281**, 20133159 (2014).
66. Whittles, L. K. & Didelot, X. Epidemiological analysis of the Eyam plague outbreak of 1665–1666. *Proc. Biol. Sci.* **283**, 20160618 (2016).
67. Roosen, J. & Curtis, D. R. Dangers of noncritical use of historical plague data. *Emerg. Infect. Dis.* **24**, 103 (2018).
68. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763 (2013).
69. Rasmussen, M. et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225 (2014).
70. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
71. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
72. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
73. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
74. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
75. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
76. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
77. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
78. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
79. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. in *Lectures on Mathematics in the Life Sciences* Vol. 17 (ed. Miura, R. M.) 57–86 (American Mathematical Society, Providence, Rhode Island, 1986).
80. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
81. Jónsson, 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 (2013).
82. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2015).
83. Auerbach, R. K. et al. *Yersinia pestis* evolution on a small timescale: comparison of whole genome sequences from North America. *PLoS ONE* **2**, e770 (2007).
84. Eppinger, M. et al. Draft genome sequences of *Yersinia pestis* isolates from natural foci of endemic plague in China. *J. Bacteriol.* **191**, 7628–7629 (2009).
85. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).

86. Kingman, J. F. C. The coalescent. *Stoch. Process. Appl.* **13**, 235–248 (1982).
87. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
88. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
89. Wickham, H. *ggplot2: Elegant Graphics For Data Analysis* (Springer International Publishing AG, Switzerland, 2016).
90. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

## Acknowledgements

We thank Aditya K. Lankapalli and Stephen Clayton for computational analysis support. We thank Guido Brandt, Antje Wissgott, Cécilia Freund and Marta Burri for laboratory support. We are grateful to Monica Green for critical comments on the manuscript. We thank Hans Sell and Michelle O'Reilly for graphics support. We thank Rafail' M. Fat-tahov for facilitating excavations of the Laishevo III archaeological site, Ayrat Sitdikov for providing access to the Laishevo III skeletal assemblage and Elizaveta V. Volkova for assisting with sampling of skeletal material. In addition, we would like to thank Joke Somers for the anthropological analysis and sampling of the Stans individuals. We thank Bettina Jungklaus for providing the samples from Brandenburg an der Havel, Bernd Trautmann for morphological analyses, Jochen Haberstroh and Mathias Hensch for providing archaeological information, and the staff of the SAPM for support during sample collection. We also thank Benoît Kirschenbilder, for his initial involvement in this project in association with the Toulouse archaeological site (16 rue des Trente Six Ponts). The fieldwork at the New Churchyard was led by Alison Telfer, and radiocarbon dating was carried out by 14CHRONO Centre, The Queen's University, Belfast, Northern Ireland. Analysis of radiocarbon dates from New Churchyard was performed by Derek Hamilton of the Scottish Universities Environmental Research Centre (SUERC), East Kilbride, Scotland, and Peter Marshall of Historic England. Radiocarbon dating for the Stans collection was performed at the LARA laboratory of the Department of Chemistry and Biochemistry at the University of Bern. Radiocarbon dating for all other material was performed in the Curt-Engelhorn-Zentrum Archäometrie gGmbH in Mannheim, Germany. The Cambridge work is supported by the Wellcome Trust (Award no. 2000368/Z/15/Z) and St. John's College, Cambridge (J.E.R., T.K., C.C., C.L.S.); the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.16-0030) (C.L.S.); and the Estonian Research Council personal research grant (PRG243) (C.L.S.). M.A.S., M.K., K.I.B. and J.K. were supported by the Max Planck Society and the ERC starting grant APGREID (to J.K.). R.T., A.H. and K.I.B. were supported by the Max Planck Society.

## Author contributions

M.A.S., M.K., R.I.T., M.Ha., K.I.B. and J.K. designed the study. M.A.S., M.K., R.T., E.A.N., C.L.S., G.U.N. and P.M.-A. performed laboratory work. M.A.S., M.K., A.A.V., F.M.K. and A.H. performed data analysis. D.W., A.A., N.C., H.F., M.G., R.H., M.He., K.v.H., S.A.I., S.K., E.L.K., J.P., J.E.R., D.C., S.L. and M.Ha. performed anthropological analysis, as well as identified and provided access to appropriate archaeological material. A.A., J.S., K.v.H., C.L. and C.C. facilitated excavations and provided access to unpublished archaeological information. T.K., M.Ha., A.H., K.I.B. and J.K. supervised different aspects of the study. M.S., M.K. and K.I.B. wrote the paper with contribution from all co-authors.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-12154-0>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Peer review information** *Nature Communications* thanks Francois Balloux and Ludovic Orlando for their contribution to the peer review of this work. Peer reviewer reports are available.

**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) 2019