A phylogeography of the second plague pandemic revealed through the analysis of historical *Y. pestis* genomes

Maria A. Spyrou^{1,2,#,*}, Marcel Keller^{1,3,#}, Rezeda I. Tukhbatova^{1,4}, Elizabeth A. Nelson^{1,2}, Aida Andrades Valtueña¹, Don Walker⁵, Amelie Alterauge⁶, Niamh Carty⁵, Hermann Fetz⁷, Michaël Gourvennec⁸, Robert Hartle⁵, Michael Henderson⁵, Kristin von Heyking³, Sacha Kacki^{9,10}, Elizabeth L. Knox⁵, Christian Later¹¹, Joris Peters^{3,12}, Jürgen Schreiber¹³, Dominique Castex¹⁰, Sandra Lösch⁶, Michaela Harbeck³, Alexander Herbig¹, Kirsten I. Bos^{1*} and Johannes Krause^{1,2*}

[#]Maria A. Spyrou and Marcel Keller contributed equally to this work

*Address correspondence to M.A.S: spyrou@shh.mpg.de; K.I.B: bos@shh.mpg.de; J.K: krause@shh.mpg.de

Affiliations:

¹Max Planck Institute for the Science of Human History, Jena 07745, Germany

²Institute for Archaeological Sciences, University of Tübingen, Tübingen 72070, Germany

³SNSB, State Collection for Anthropology and Paleoanatomy Munich, Munich 80333, Germany

⁴Center of Excellence “Archaeometry”, Kazan Federal University, Kazan 420008, Russian Federation

⁵MOLA (Museum of London Archaeology), London N1 7ED, United Kingdom

⁶Department of Physical Anthropology, Institute for Forensic Medicine, University of Bern, Bern 3007, Switzerland

⁷Archaeological Service, State Archive Nidwalden, Nidwalden 6371, Switzerland

⁸Archeodunum SAS, Agency Toulouse, 8 allée Michel de Montaigne, 31770 Colomiers

⁹Department of Archaeology, Durham University, South Rd, Durham DH1 3LE, United Kingdom

¹⁰PACEA, CNRS Institute, Université de Bordeaux, Pessac 33615, France

¹¹Bavarian State Department of Monuments and Sites, 80539 Munich, Germany

¹²ArchaeoBioCenter 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

¹³Dig it! Company GbR, Peiting, 86971, Germany

Abstract

The second plague pandemic (14th - 18th century AD), caused by the bacterium *Yersinia pestis*, is infamous for its initial wave, the Black Death (1346-1353 AD), and its repeated scourges in Europe and the vicinity until the Early Modern Era. Here, we report 32 ancient *Y. pestis* genomes spanning the 14th to 17th century AD through the analysis of human remains from nine European archaeological sites. Our data support an initial entry of the bacterium from Eastern Europe and the absence of genetic diversity during the Black Death as well as low diversity during local outbreaks thereafter. Moreover, 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. Finally, we show the loss of a genomic region that includes virulence-associated genes in strains associated with late stages of the second plague pandemic (17th - 18th century AD). This deletion could not be detected in extant strains within our modern dataset, though it was identified in a today-extinct lineage associated with the first plague pandemic (6th - 8th century AD), suggesting convergent evolution during both pandemic events.

Introduction

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 subsequently continued with recurrent outbreaks in Europe until the 18th century AD^{1,2}. Its causative agent, *Yersinia pestis*³, is a highly virulent bacterium that causes bubonic, pneumonic, and septicemic plague and today is maintained among wild rodent populations in Eastern Europe, Asia, Africa and the Americas⁴⁻⁶. The source of the second pandemic, as well as the route that plague followed prior to its entry into Europe, remains hypothetical. Genetic characterisation of BD genomes analysed alongside modern diversity has associated its initiation with a star-like diversification of *Y. pestis* lineages⁷, and has led to the proposal of an East Asian origin, as modern strains that are directly ancestral to this event have been identified in China near historical trade routes^{7,8}. However, the fact that until today the majority of *Y. pestis* strains have been isolated from China⁷ and to a lesser extent from Central Asia—which is also known to harbour a variety of disease foci^{9,10}—creates a sampling bias in the data and questions the accuracy of such conclusions. The view of an East Asian origin for the BD is also supported by certain historical interpretations suggesting that the bacterium travelled westward either via the Mongol army or via Silk Road traders during the early 14th century¹¹. Such claims, however, can not yet be verified given the

scarcity of documentary sources from China describing precise disease symptoms¹², as well as the lack of molecular evidence from Central and East Asia dating to the early 14th century. The first clearly documented outbreaks of the second pandemic occurred in 1346 in the Lower Volga and Black Sea regions^{1,13}. From there, the bacterium dispersed through the rest of Europe within the next five years, causing reductions in the human population estimated to be as high as 60%¹. At present, the only ancient genetic evidence from Eastern Europe is from the city of Bolgar in the Middle Volga region and stems from the time period shortly after the BD (1362-1400 AD)¹⁴, during which *Y. pestis* seems to have expanded eastward to later become a precursor of the third plague pandemic that spread worldwide from China during the 19th century¹⁴⁻¹⁶. Thus, it has been difficult to genetically assess whether this region also contributed to the disease's introduction to Europe, since supporting genomic evidence from the earliest phases of the second pandemic have thus far been elusive.

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^{2,17}. Whether these were caused by multiple introductions of the disease from an Asian source or rather by its local persistence in Europe is an on-going subject of scientific debate^{14,18-20}. While data from climatic proxies are interpreted to support the former hypothesis¹⁸, genetic evidence currently favours the latter^{14,19}. Analysis of *Y. pestis* genomes from 16th-century Ellwangen, Germany¹⁴ and the Great Plague of Marseille in France (L'Observance, 1720-22 AD)²⁰ revealed the existence of a previously unidentified *Y. pestis* lineage that was responsible for both of these post-Black Death (post-BD) outbreaks. This lineage descends from the strain associated with the BD thus far identified in both London and Barcelona, and, therefore, likely represents plagues' legacy and persistence in or around Europe after 1353. The limited number of published ancient *Y. pestis* genomes^{8,14,20}, however, challenges our ability to construct hypotheses of whether this lineage was solely responsible for the numerous post-BD outbreaks in Europe and whether it derived from a single or multiple reservoirs.

Here, we take steps to overcome these limitations by greatly 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. We analyse material recovered from nine archaeological sites located in Russia, Germany, Switzerland, England and France from which we reconstruct 32 *Y. pestis* genomes. Our data supports the entrance of *Y. pestis* into Europe through the East during the initial wave of the pandemic and consistently demonstrates an absence of genetic diversity in the bacterium during the BD. In addition, our genomic analysis of

post-BD outbreaks from Central and Western Europe suggests the local diversification of a single plague lineage between the 14th and 18th centuries that may have resided in more than one reservoir.

Results

Y. pestis qPCR screening

Human archaeological material was screened for the presence of the *Y. pestis*-specific gene, *pla*, located on the pPCP1 plasmid, using a previously described qPCR approach²¹. A total of 181 teeth were retrieved and tested from archaeological sites in the cities of Nabburg (*n*=12), Manching-Pichl¹⁹ (*n*=28), Starnberg (*n*=3), Landsberg am Lech (*n*=10) and Brandenburg an der Havel¹⁹ (*n*=3) in Germany; Stans (*n*=32) in Switzerland, London (*n*=40) in England, Toulouse (*n*=39) in France and the city of Laishevo (*n*=10) in the Volga region of Russia (Figure 1 and Supplementary Information). Extracts from 41 teeth across all 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.

Y. pestis capture and whole genome reconstruction

We prepared UDG-treated libraries^{22,23} from putatively positive extracts and used a *Y. pestis* in-solution capture approach²⁴ combined with high-throughput sequencing for the retrieval of 1,299,105 to 79,055,317 raw reads per sequenced library. All data were mapped against the *Y. pestis* CO92 reference genome (NC_003143.1)³. This resulted in 86,278 to 3,822,030 unique mapping reads yielding 1.1 to 80.1-fold coverage across 32 individuals that span the time transect between the 14th and 17th century in Europe (Supplementary table 2). More specifically, we could retrieve three *Y. pestis* genomes from Nabburg, two from Manching-Pichl¹⁹, one from Starnberg, one from Landsberg am Lech, two from Brandenburg an der Havel¹⁹ (all from Germany), 15 from Stans (Switzerland), five from London (England), one from Toulouse (France) and two from Laishevo (Russia). Of those, 23 isolates showed at least 50% of the reference genome covered at 5-fold (Supplementary table 2), 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 immobilized probes)¹⁴ (Supplementary table 2).

Y. pestis phylogenetic reconstruction

To infer genetic relationships between the new and previously published *Y. pestis* isolates we constructed phylogenies using the maximum likelihood 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^{7,9,10,25,26}, which represent most of the published *Y. pestis* genetic diversity. In addition, we included all previously published second pandemic isolates ($n=12$)^{8,14,20}, a 6th-century AD isolate from Germany²⁷, a 2nd- to 3rd-century AD isolate from the Tian Shan mountains in Kyrgyzstan²⁸, as well as three Bronze Age isolates from the Altai and Volga regions^{29,30} (Supplementary fig. 1).

All newly reconstructed genomes appear on Branch 1 and are closely related to the previously published second pandemic isolates from Europe (Figure 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 (Supplementary data 1). A number of genomes were excluded from the phylogeny (NAB005, BRA003, STN011 and STN004) as they showed an excess amount of heterozygous SNP calls in comparison to other contemporaneous isolates from the same archaeological contexts (Supplementary fig. 2). Such calls potentially arise from enrichment of non-target DNA stemming from closely related organisms, an effect frequently identified in ancient metagenomic datasets^{27,31,32} that can result in false branch elongation (Supplementary fig. 3) and misinterpretation of phylogeographic inferences.

Our phylogenetic reconstruction shows that the LAI009 isolate from Laishevo is ancestral to the BD isolates from Central, Western and Southern Europe, as well as to the previously published late 14th-century isolates from London (6330)⁸ and Bolgar City¹⁴ (Figure 2). This genome possesses only one derived SNP distinguishing it from the N07 polytomy that gives rise to Branches 1 to 4 (Figure 2; Supplementary data 1)⁷. Since all other second pandemic genomes share an additional derived SNP on Branch 1, we interpret LAI009 as the most ancestral thus far identified form of the strain that entered Europe during the initial wave of the second pandemic. Regarding the Central and Western European genomes, NAB003 from Nabburg does not show differences compared to previously published BD genomes from London and Barcelona^{8,14}, whereas TRP002 from Toulouse, which dates to 1347-1350 based on archaeological evidence, is seemingly distinct from other isolates and forms its own unique branch (Figure 2; Supplementary data 1). Although this isolate shows indications of high heterozygosity (Supplemental fig. 2) and suffers from comparatively lower coverage, we nevertheless considered it further in our analysis, as it is the sole genome available from this site. To

investigate whether its branch arose from true unique variants or contaminant reads that resulted in erroneous SNP calling, the genome alignment was evaluated in greater detail. After visual inspection, all eight SNPs unique to TRP002 appear in regions of the genome where reads from diverse sources seem to be mapping (Supplementary fig. 4). In addition, these variants exhibit all previously defined characteristics of problematic SNP assignments in low coverage genomes²⁷ and, therefore, were considered to be of exogenous origin. We, hence, conclude that apart from LAI009 all other reconstructed genomes that date to the Black Death period have identical genotypes.

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)¹⁴, and Marseille, France (1720-1722)²⁰, which are descended from the European BD isolates (Figure 2; Supplementary data 1). Here, we identify the earliest evidence of this lineage in a 14th-century isolate from Manching-Pichl (MAN)¹⁹ (see Supplementary Information), which is followed by the more derived 15th- to 17th-century isolates from Landsberg (LBG), Stans (STN) and Starnberg (STA), as well as the 17th-century Brandenburg an der Havel (BRA)¹⁹ 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 (Figure 2). From this divergence, one clade gives rise to the strains associated with outbreaks in Germany and Switzerland, and the second encompasses 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 (ELW098/549_O) and the London (BED) genomes ($d=45$) is double than that observed between Ellwangen and Brandenburg (BRA, $d=22$), despite an assumption of them being contemporaneous (early 17th century AD) based on their archaeological dating (Figure 2; Supplementary Information; Supplementary table 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 ($n=76$), which contains all second pandemic *Y. pestis* genomes, using available calibration points in our modern and ancient datasets (Supplementary data 2). 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^{7,16}. Our analysis suggests an over 40-fold difference between the fastest and slowest mutation rates observed in Branch 1 (Figure 3). In particular, we observe the fastest rates in three internal branches (Figure 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 earlier (Figure 3 and Supplementary data 3). The second is the branch leading to the 1.ANT strains isolated from Africa (Congo and Uganda) (Figure 3 and Supplementary data 4). The broad history of 1.ANT and the time period associated with its establishment in Africa are unknown, though an introduction from Eurasia has been hypothesized¹⁴. The third, which displays the fastest rate within the entire Branch 1, is the branch leading to 1.ORI isolates (Figure 3 and Supplementary data 5), which is associated with the global spread of *Y. pestis* via maritime routes during the third plague pandemic (1894 - 1950s)^{15,16}. Our results, therefore, support the idea of faster substitution rates during epidemic spread, here particularly noticeable for lineages known to have expanded over wide geographic areas.

Virulence factor analysis

To investigate the virulence profiles of all newly reconstructed genomes, we analysed the presence or absence of virulence-associated genes located on the *Y. pestis* chromosome (Figure 4a) and plasmids (Supplementary fig. 5)^{33,34}, in comparison to representatives of published ancient and modern strains. We find that the genetic profiles of some of the previously characterised historical strains are influenced by the capture design used for their retrieval. Specifically, the second pandemic genomes ‘Bolgar 2370’ and ‘Barcelona 3031’¹⁴, and the first pandemic genome ‘Altenerding 2148’²⁷ seem to lack coverage in certain *Y. pestis*-specific regions, since *Y. pseudotuberculosis* was used as a probe-design template for enrichment of these genomes (Figure 4a). Regarding the newly reconstructed strains, we find that most possess all analysed virulence determinants with the exception of the New Churchyard (BED) and Marseille (OBS) strains that lack the magnesium transporter genes *mgtB* and *mgtC* (Figure 4). Magnesium transporters are considered vital for bacterial intracellular survival under low Mg^{2+}

conditions³⁵, 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 mice³⁶. 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) (Figure 4b) 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 elements³⁷. 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 table 3). 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 Information) and during the 18th-century Plague of Marseille, in France, (OBS - 1720-1722)²⁰, which was one of the last major epidemics that occurred in continental Europe³⁸. Intriguingly, a nearly identical genomic deletion (45 kb), also including the *mgtB* and *mgtC* virulence-associated genes, was recently identified in an ancient isolate from France (LVC)³⁹ sequenced from a victim of the first plague pandemic (6th – 8th centuries AD)³⁹. This genome is described elsewhere and dates within a wide temporal interval (550-650 AD), though based on the existing genomic data it seems to be the most derived isolate on the first pandemic lineage sequenced to date³⁹. Therefore, the identified deletion, which includes potential virulence factors, seems to be specific to derived isolates of two lineages within the *Y. pestis* phylogeny that seem to be today extinct.

Discussion

A series of studies have sufficiently demonstrated the preservation of *Y. pestis* in ancient human remains from a wide temporal transect^{8,14,20,24,27,29,40}. In this study, this has given a great opportunity for the extensive sampling of multiple European epidemic burials from the time period between the 14th and 17th century, in order to gain a more complete picture of *Y. pestis*' genetic history during the second plague pandemic. As a result, we nearly quadruple the amount of genomic data available from that time period (Figure 1 and Supplementary tables 1, 2) and their integration in 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 Russia¹³. A commonly accepted view dates its arrival in the southwest, particularly in cities of Astrakhan and Sarai, in 1346^{1,41} 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^{13,41}) may have followed an alternative route, occurring at the end of the BD, between 1351-1353, via the Baltic Sea^{1,13,41}. Such a notion of plague's expansion from northern Europe towards the East is also supported by published ancient genomic data from the late 14th-century Middle Volga region of Russia¹⁴, though other employable 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 (Figure 2), we provide evidence for the bacterium's presence in the same region, ~2,000 km northeast of the Crimean peninsula, prior to reaching Southern Europe in 1347 – 1348 (here represented by strains from Barcelona and Toulouse)¹. These results lead us to propose that the N07 derived SNP previously termed “p1”¹⁴ (Figure 2, Supplementary fig. 1) 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 (Figure 2, Supplementary fig. 1)⁷ and the apparent East Asian sampling bias of modern isolates⁷, additional data will be necessary for an accurate re-evaluation of the geographic origin of Branch 1.

Such insights of low genomic diversity during the initial pandemic wave (BD) become valuable when attempting to reconstruct the spread of plague after 1353. Previous research based on climatic proxies¹⁸ and PCR data⁴² 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^{8,14,20} and new whole genome data from 15 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 and Western Europe during the BD. We, therefore, interpret our data as supporting a single entry of *Y. pestis* during the BD, though additional interpretations may be revealed through the discovery of un-sampled diversity in Western Eurasia. Subsequent to its entry, we observe the formation of two sister lineages (Figure 2). The first lineage is responsible for the bacterium's possible eastward expansion after the BD. It contains strains from late-14th-century London (6330)⁸ and Bolgar City (2370)¹⁴, as well as extant strains from Africa (1.ANT)⁴³, and, most importantly, a worldwide set of isolates

associated with the third pandemic (1.ORI, 19th – 20th centuries)^{7,15,16} (Figure 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 (BED) and France (OBS) (Figure 2), suggesting that it persisted in Europe 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 suggested¹⁴.

We find that the “post-BD lineage” gave rise to (at least) two distinct clades within Europe 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 (Figure 2), but also by potential differences in their virulence profiles (Figure 4a) and substitution rates (Figure 3). The clade that exhibits a slower substitution rate is mainly represented by temporally and genetically closely related isolates from Germany and Switzerland (Figure 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 BD⁴⁴, although a possible sampling bias should be noted since the majority of our data points derive primarily from this region. On the other hand, the clade that exhibits a faster substitution rate (Figure 3) seems 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 ships⁴⁵, although sources favouring local epidemic eruptions also exist⁴⁶. Published research has demonstrated that transmission of *Y. pestis* via steamships during the 19th-century third pandemic played a significant role in initial introduction of the bacterium to several regions worldwide, such as in Madagascar where it persists until today^{15,16,47,48}. Although the possibility of maritime introductions of plague into London and Marseille during the second pandemic vastly expand the location of its potential geographic source(s), the phylogenetic positioning of the BED and OBS genomes within the “post-BD lineage” and in relation to other second pandemic isolates suggests it likely arose within Europe or the vicinity.

We identified a 49 kb deletion within both BED and OBS genomes (Figure 4b), which results in the loss of two virulence-associated genes, *mgtB* and *mgtC* (Figure 4a). This deletion could not

be identified in any other second pandemic or modern strain within our dataset. The inferred virulence importance of *mgtB* and *mgtC* genes is associated with intracellular survival of *Y. pestis* within macrophages^{36,49}. Their co-expression has been shown to affect the virulence exerted by other pathogenic enterobacteria under laboratory conditions^{50,51} and both genes have been proposed as potential drug targets^{36,52}. Moreover, the function of *mgtB* was shown to be temperature-dependent, being active at 37°C but not at 20°C⁵³, which suggests that its loss may primarily affect warm-blooded hosts. Intriguingly, a 45 kb deletion within the same region was identified in a genome likely associated with a late outbreak of the first plague pandemic (6th - 8th century AD)³⁹, which sets it as a candidate for convergent evolution and raises a question regarding its functional importance. Given that all the genomes that contain this deletion were isolated from historical plague victims, it is plausible it may not have affected the bacterium's virulence, particularly since genome decay is a well known mode of evolution in *Y. pestis*^{54,55}. Nevertheless, since both lineages that harbour it are thought to be today extinct, a full functional characterisation will be of importance to evaluate its potential significance in survival across different mammalian hosts and arthropod vectors, which may have affected the bacterium's maintenance 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^{1,56}. It serves as a classic historical example of rapid infectious disease emergence in Europe, 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 genealogical 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 Europe remain under-sampled for ancient 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^{57,58}, climatic^{18,59,60} and epidemiological data⁶¹, as well as a critical re-evaluation of historical records⁶² 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 and part of the sampling and DNA extractions were performed at aDNA facilities of the ArchaeoBioCenter of the Ludwig Maximilian University of Munich.

181 teeth from nine sites located in Germany, Switzerland, England, France and Russia spanning the 14th – 17th centuries (see Supplementary Information) were sectioned in the cemento-enamel junction, and 30-70 mg of powder were 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 DNA⁶³. Tooth powder was incubated in 1 ml lysis buffer (0.45M EDTA, pH 8.0, and 0.25 mg/ml proteinase K) overnight (12-16 h) at 37°C. After, DNA was bound to the silica membrane of spin columns using 10 ml of GuHCl-based binding buffer as described before⁶³, followed by a purification that was performed using either the MinElute purification kit (Qiagen) or the Viral Nucleic Acid Kit (Roche). DNA was eluted in 100 µl TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20). 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 concentration²¹. 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 protocol²¹. 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.

UDG library preparation and Y. pestis whole genome capture

After *pla* screening, DNA extracts yielding an amplification product were considered as putatively positive for *Y. pestis* and were subsequently converted into Illumina double-stranded DNA libraries as described before²³, using a starting volume of 50-60 µl, with an initial USER (New England Biolabs) treatment step, where uracil-DNA-glycosylase (UDG) was used in combination with endonuclease VIII to excise uracil nucleotides that result from post-mortem DNA damage^{22,64}. 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 quantification⁶⁵, in order to ensure maximal amplification efficiency. Every reaction was assigned a maximum input of 2×10^{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 (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20). After indexing all libraries were amplified using *Herculase II Fusion DNA Polymerase* (Agilent) to a concentration of 200-300 ng/ μ l, in order to achieve 1-2 μ g of DNA in a total of 7 μ l. Products were again purified using the MinElute DNA purification kit (Qiagen), and eluted in TET (10mM Tris-HCl, 1 mM EDTA pH 8.0, 0.05% Tween20). In-solution *Y. pestis* capture was then performed as described previously²⁴, 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), KIM 10 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. 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 (1x151+8+8 cycles or 1x76+8+8 cycles) or on a HiSeq4000 (using 1x76+8+8 cycles or 2x76+8+8 cycles). Preprocessing of de-multiplexed reads was performed on the automated pipeline EAGER⁶⁶ and involved removal of Illumina adapters and read merging by AdapterRemoval v2⁶⁷, as well as filtering for sequencing quality (minimum base quality of 20) and read length (to retrieve only reads ≥ 30 bp). Subsequently, reads were mapped with BWA (version 0.7.12)⁶⁸, implemented in EAGER, against the CO92 reference genomes (NC_003143.1)³ using stringent parameters (-n 0.1, -l 32) for genome reconstruction and more lenient parameters (-n 0.01, -l 32) for SNP inspection. 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/>).

SNP calling and phylogenetic analysis

SNP calling was performed using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK)⁶⁹. Our newly reconstructed genomes were analysed alongside previously published *Y. pestis* genomes, which included a modern-day dataset of 233 genomes^{7,9,10,25,26,43}, one Justinianic strain (Altenerding)²⁷, three Bronze Age strains²⁹, 12 previously published historical genomes from the second plague pandemic and a *Y. pseudotuberculosis* strain (IP32953) which was used as outgroup. 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 v0.85³² (<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^{7,16}. In the present dataset, a total of 7,251 variant positions were identified. Subsequently, the annotation as well as the effect of each SNP was determined through the program SnpEff⁷⁰. For the phylogenetic analysis, we used a SNP alignment produced by MultiVCFAnalyzer v0.85 to construct phylogenetic trees using the maximum likelihood (ML) and maximum parsimony (MP) methods. Up to 3% missing data was included in the analysis (97% partial deletion), resulting in a total amount of 6,353 SNPs used for phylogenetic reconstruction. The MP phylogeny was produced in MEGA7⁷¹ in order to make a first assessment of genome topologies. The ML phylogenies were reconstructed with the program RAxML (version 8.2.9)⁷² using the Generalised Time Reversible (GTR)⁷³ model with four gamma rate categories and 1000 bootstrap replicates to assess tree topology support.

Heterozygosity estimates

Heterozygous variant calls in all genomes were investigated given the disparity of branch lengths observed in isolates from the same archaeological sites (see Supplementary figure 2). 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 subsampled all newly reconstructed genomes to ~5-fold coverage, and performed SNP calling using GATK⁶⁹,

as described above. We then compiled a SNP table of all variant positions across our dataset using MultiVCFAnalyzer v0.85³² accounting for all heterozygous positions. Histograms of allele frequencies for all SNPs with <100% read support were constructed with R v3.4.1⁷⁴ 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 a maximum likelihood phylogeny in RaxML⁷² using all Branch 1 genomes (modern + ancient $n=76$), with the exception of the low coverage TRP002 genome that showed possible environmental contamination in its 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 (545 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 0.53 and 0.28 respectively, which permitted the proceeding with molecular dating analysis.

The Bayesian framework BEASTv1.8⁷⁵ was used to assess the substitution rate variation across the *Y. pestis* Branch 1 using the 2.MED KIM10 as outgroup. Our BEAUti set-up took into consideration all archaeological, radiocarbon, and sampling dates of both ancient and modern genomes (Supplementary data 2) 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). All genomes within this Branch 1 were constrained as a monophyletic clade. The Generalised Time Reversible (GTR)⁷³ substitution model (6 gamma rate categories) and a lognormal relaxed clock (clock rate tested and strict clock rejected in MEGA⁷¹) were used to set up two separate analyses using the Coalescent Constant Size⁷⁶ and Coalescent Bayesian Skyline⁷⁷ 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-fitted demographic model for our dataset using path sampling and stepping stone sampling (PS/SS) implemented in BEAST v1.8⁷⁵. 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 expected sample sizes (ESS) were > 200. We used TreeAnnotator⁷⁵, 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.

Virulence factor presence/absence and deletion analysis

In order to investigate the virulence profiles of the newly reconstructed second pandemic genomes, the highest quality (coverage) genome from every site (LAI009, NAB003, TRP002, MAN008, STA001, LBG002, STN014, BRA001, BED030) was used for comparison with each other and with previously published representatives of ancient (London ES, 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. The newly reconstructed TRP002 and the published London 6330 genomes were excluded from this analysis due to their low coverage. All other listed genomes were re-mapped against the CO92 chromosomal reference genome (NC_003143.1) using a mapping quality (-q) filter of 0. The coverage across 80 chromosomal and 43 plasmid virulence-associated genes that were previously defined³³, was calculated using bedtools⁷⁸. The results are plotted in the form of a heatmap of using the ggplot2⁷⁹ package in R version 3.4.1⁸⁰ and can be viewed in Figure 4. In addition, we used BWA-MEM⁸¹ to explore the precise coordinates of observed gene or region losses in all affected genomes using default parameters. For a visualisation of the identified deletion we computed average coverages across 2,000-bp windows in representative *Y. pestis* genomes from all analysed time periods of the second pandemic, and subsequently used the program Circos⁸² to produce coverage plots of a 20-fold maximum coverage for all genomes (LAI009, London BD 8124/8291/11972, Bolgar 2370, MAN008, STA001, ELW098, LBG002, STN014, BRA002, BED030, OBS137 and the reference genomes CO92).

Data availability

Raw sequencing data of the deep-sequenced genomes will become available on the European Nucleotide Archive under project accession number PRJEB29990 upon publication. 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.

Figures & Tables

Figure 1 – Archaeological site locations

(a) Archaeological/radiocarbon dates of previously published and new second plague pandemic isolates (see Supplementary Information).

(b) Map showing geographic locations of the archaeological sites from which second pandemic (14th- to 18th-century AD) *Y. pestis* genomes have been reconstructed. The number (*n*) of whole 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.

A maximum likelihood phylogeny was generated allowing for up to 3% missing data (97% partial deletion) and considering a total of 6,353 SNP positions. The image shows a graphical representation of Branches 1 - 4 (see Supplementary fig. 1 for a complete phylogeny), to emphasize the phylogenetic positioning of the new and previously published second pandemic strains (14th -18th centuries). Subclades of published genomes are collapsed to enhance tree visibility. Numbers (*n*) in brackets indicate the amount of strains represented in each collapsed branch. Nodes that have bootstrap values of ≥ 95 are indicated by asterisks (*). 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), Armenia (ARM), Georgia (GEO), Azerbaijan (AZE), Uzbekistan (UZB), Turkmenistan (TKM), Russia (RUS), and unspecified regions of the Former Soviet Union (FSU).

The figure presents a maximum clade credibility (MCC) phylogenetic tree generated using BEAST v1.8⁷⁵ (rooted with 2.MED KIM10 - outgroup not shown). The tree was viewed in FigTree v1.4, and modified so that branch colours represent mean substitution rates (substitution/site/year). The tree depicts the substitution rate variation across Branch 1 of the *Y. pestis* phylogeny, which ranges from 2.12E-7 (highest) to 5.10E-9 (lowest) substitutions/site/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 figure 1), with the exception of the TRP002 genome since its SNP calls were likely affected by a possible environmental contamination. The mean substitution rate across Branch 1 (including 2.MED KIM10) was calculated to 2.67E-8 substitutions/site/year. Lengths of branches are scaled to represent sample ages, and the depicted Branch 1 sequences are estimated to represent 756 years (95% HPD: 671 - 887) 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).

Table 1 – Post-capture sequencing statistics of all new *Yersinia pestis* genomes that passed quality tests 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 ^a	3,624,482	36.2	80.1	93.6	102.9	48.5
BED028.A0102	New Churchyard, London	1560–1635 ^a	2,665,238	22.2	37.2	91.4	65.0	49.0
BED034.A0102	New Churchyard, London	1560–1635 ^a	1,371,698	10.5	18.3	89.1	62.2	49.2
BED024.A0102	New Churchyard, London	1560–1635 ^a	1,000,524	18.1	12.6	84.7	58.5	49.1
BRA001.A0101	Domlinden 12, Brandenburg	1618-1648 ^b	2,387,557	23.2	23.8	92.0	46.4	47.5
LAI009.A0101	Laishevo III, Laishevo	1300-1400 ^b	2,549,926	23.9	28.4	92.1	51.8	48.0
LBG002.A0101	Kirchhof St. Johannis, Landsberg	1455-1632 ^a	621,713	27.9	7.2	66.4	54.2	49.9
MAN008.B0101	St. Leonhardi, Manching-Pichl	1283-1390 ^a	1,974,399	44.9	25.8	88.7	60.8	50.8
NAB003.B0101	Sankt Johans Friedhof, Nabburg	1292-1392 ^a	684,029	33.5	8.1	70.2	54.8	49.7
STA001.A0101	Possenhofener str. 3, Starnberg	1433-1523 ^{a,b}	1,110,049	9.9	11.7	84.3	49.0	44.7
STN014.A0101	Nägeligasse, Stans	1485-1635 ^a	3,822,030	48.2	55.3	93.0	67.3	48.9
STN020.A0101	Nägeligasse, Stans	1485-1635 ^a	2,020,769	44.3	28.2	90.3	64.8	48.5
STN021.A0101	Nägeligasse, Stans	1485-1635 ^a	1,588,442	35.1	21.7	88.6	63.7	48.5
STN019.A0101	Nägeligasse, Stans	1485-1635 ^a	1,325,076	35.3	18.7	87.1	65.8	49.1
STN007.A0101	Nägeligasse, Stans	1485-1635 ^a	1,293,507	32.8	18.0	86.7	64.8	49.4
STN002.A0101	Nägeligasse, Stans	1485-1635 ^a	935,795	27.6	12.7	83.3	63.0	48.4
STN008.A0101	Nägeligasse, Stans	1485-1635 ^a	875,153	30.2	11.7	77.7	62.5	50.1
STN013.A0101	Nägeligasse, Stans	1485-1635 ^a	714,482	24.5	9.2	73.8	59.9	49.0
TRP002.A0101	Trente-Six Ponts 16, Toulouse	1347-1350 ^b	632,303	19.8	5.9	50.9	43.2	48.7

^aDates based on radiocarbon dating of collagen^bDates based on archaeological context information

References

- 1 Benedictow, O. J. *The Black Death, 1346-1353: The complete history*. (Boydell & Brewer, 2004).
- 2 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. (1976).
- 3 Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523-527, doi:10.1038/35097083 (2001).
- 4 Gage, K. L. & Kosoy, M. Y. Natural history of plague: perspectives from more than a century of research. *Annu. Rev. Entomol.* **50**, 505-528, doi:10.1146/annurev.ento.50.071803.130337 (2005).
- 5 Prentice, M. B. & Rahalison, L. Plague. *Lancet* **369**, 1196-1207, doi:10.1016/S0140-6736(07)60566-2 (2007).
- 6 Tikhomirov, E. Epidemiology and distribution of plague. *World Health Organisation* (1999).
- 7 Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci U S A* **110**, 577-582, doi:10.1073/pnas.1205750110 (2013).
- 8 Bos, K. I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506-510, doi:10.1038/nature10549 (2011).
- 9 Kutyrev, 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** (2018).
- 10 Eroshenko, G. A. *et al.* *Yersinia pestis* strains of ancient phylogenetic branch 0. ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. *PloS one* **12**, e0187230 (2017).
- 11 McNeill, W. H. *Plagues and peoples*. 1976. Garden City, NY: Anchor P (1998).
- 12 Hymes, R. Epilogue: a hypothesis on the East Asian beginnings of the *Yersinia pestis* polytomy. *The Medieval Globe* **1**, 12 (2014).
- 13 Alexander, J. T. *Bubonic plague in early modern Russia: public health and urban disaster*. (Johns Hopkins University Press, Baltimore, Maryland 21218, USA, 1980).
- 14 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).
- 15 Pollitzer, R. The Plague. Geneva. *World Health Organization*, 26 (1954).
- 16 Morelli, G. *et al.* *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140-1143, doi:10.1038/ng.705 (2010).
- 17 Büntgen, U., Ginzler, C., Esper, J., Tegel, W. & McMichael, A. J. Digitizing historical plague. *Clin. Infect. Dis.* **55**, 1586-1588, doi:10.1093/cid/cis723 (2012).
- 18 Schmid, B. V. *et al.* Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci U S A* **112**, 3020-3025, doi:10.1073/pnas.1412887112 (2015).
- 19 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, doi:10.1371/journal.pone.0145194 (2016).
- 20 Bos, K. I. *et al.* Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* **5**, e12994, doi:10.7554/eLife.12994 (2016).
- 21 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 U S A* **108**, E746-752, doi:10.1073/pnas.1105107108 (2011).

- 22 Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87, doi:10.1093/nar/gkp1163 (2010).
- 23 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).
- 24 Andrades Valtueña, A. A. *et al.* The Stone Age Plague and Its Persistence in Eurasia. *Curr. Biol.* **27**, 3683-3691. e3688 (2017).
- 25 Kislichkina, A. A. *et al.* Nineteen Whole-Genome Assemblies of *Yersinia pestis* subsp. *microtus*, Including Representatives of Biovars *caucasica*, *talassica*, *hissarica*, *altaica*, *xilingolensis*, and *ulegeica*. *Genome Announc* **3**, e01342-01315, doi:10.1128/genomeA.01342-15 (2015).
- 26 Zhgenti, E. *et al.* Genome Assemblies for 11 *Yersinia pestis* Strains Isolated in the Caucasus Region. *Genome Announc* **3**, e01030-01015, doi:10.1128/genomeA.01030-15 (2015).
- 27 Feldman, M. *et al.* A High-Coverage *Yersinia pestis* Genome from a Sixth-Century Justinianic Plague Victim. *Mol. Biol. Evol.* **33**, 2911-2923, doi:10.1093/molbev/msw170 (2016).
- 28 de Barros Damgaard, P. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369 (2018).
- 29 Rasmussen, S. *et al.* Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* **163**, 571-582, doi:10.1016/j.cell.2015.10.009 (2015).
- 30 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).
- 31 Vågane, A. J. *et al.* *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol* **2**, 520-528, doi:10.1038/s41559-017-0446-6 (2018).
- 32 Bos, K. I. *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494-497, doi:10.1038/nature13591 (2014).
- 33 Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect. Immun.* **77**, 2242-2250, doi:10.1128/IAI.01477-08 (2009).
- 34 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).
- 35 Groisman, E. A. *et al.* Bacterial Mg²⁺ homeostasis, transport, and virulence. *Annu. Rev. Genet.* **47**, 625-646 (2013).
- 36 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).
- 37 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).
- 38 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).
- 39 Keller, M., Spyrou, M. A., Scheib, C. I., Kröpelin, A., Haas-Gebhard, B., Pääffeng, B., Haberstroh, J., Ribera i Lacomba, A., Raynaud, C., Cessford, C., Stadler, P., Nägele, K., Neumann, G.U., Bates, J. S., Trautmann, B., Inskip, S., Peters, J., Robb, J.E., Kivisild, T., McCormick, M., Bos, K.I., Harbeck, M., Herbig, A., Krause, J. Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541-750 CE). *bioRxiv* (2018).
- 40 Wagner, D. M. *et al.* *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319-326, doi:10.1016/s1473-3099(13)70323-2 (2014).

- 41 Benedictow, O. J. *The Black Death and Later Plague Epidemics in the Scandinavian Countries: Perspectives and Controversies*. (Walter de Gruyter GmbH & Co KG, 2016).
- 42 Haensch, S. *et al.* Distinct clones of *Yersinia pestis* caused the black death. *PLoS Pathog* **6**, e1001134, doi:10.1371/journal.ppat.1001134 (2010).
- 43 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).
- 44 Carmicheal, A. G. *Plague persistence in western Europe: A hypothesis*. Vol. 1 157-192 (ARC Medieval Press, 2014).
- 45 Signoli, M., Séguy, I., Biraben, J.-N., Dutour, O. & Belle, P. Paleodemography and Historical Demography in the Context of an Epidemic. *Population* **57**, 829-854 (2002).
- 46 Cummins, N., Kelly, M. & Ó Gráda, C. Living standards and plague in London, 1560–1665. *The Economic History Review* **69**, 3-34 (2016).
- 47 Brygoo, E.-R. Epidémiologie de la peste à Madagascar. *Les Archives de l'Institut Pasteur de Madagascar* (1966).
- 48 Vogler, A. J. *et al.* Temporal phylogeography of *Yersinia pestis* in Madagascar: Insights into the long-term maintenance of plague. *PLoS neglected tropical diseases* **11**, e0005887 (2017).
- 49 Grabenstein, J. P., Fukuto, H. S., Palmer, L. E. & Bliska, J. B. Characterization of phagosome trafficking and identification of PhoP-regulated genes important for survival of *Yersinia pestis* in macrophages. *Infect. Immun.* **74**, 3727-3741 (2006).
- 50 Blanc - Potard, A. B. & Groisman, E. A. The *Salmonella* selC locus contains a pathogenicity island mediating intramacrophage survival. *The EMBO journal* **16**, 5376-5385 (1997).
- 51 Snively, M., Miller, C. & Maguire, M. The mgtB Mg²⁺ transport locus of *Salmonella typhimurium* encodes a P-type ATPase. *J. Biol. Chem.* **266**, 815-823 (1991).
- 52 Belon, C. *et al.* A macrophage subversion factor is shared by intracellular and extracellular pathogens. *PLoS Path.* **11**, e1004969 (2015).
- 53 Snively, M., Florer, J., Miller, C. & Maguire, M. Magnesium transport in *Salmonella typhimurium*: 28Mg²⁺ transport by the CorA, MgtA, and MgtB systems. *J. Bacteriol.* **171**, 4761-4766 (1989).
- 54 Achtman, M. *et al.* *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **96**, 14043-14048 (1999).
- 55 Chain, P. S. *et al.* Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A* **101**, 13826-13831, doi:10.1073/pnas.0404012101 (2004).
- 56 Perry, R. D. & Fetherston, J. D. *Yersinia pestis*--etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35-66 (1997).
- 57 Dean, K. R. *et al.* Human ectoparasites and the spread of plague in Europe during the Second Pandemic. *Proceedings of the National Academy of Sciences*, 201715640 (2018).
- 58 Keeling, M. J. & Gilligan, C. A. Bubonic plague: a metapopulation model of a zoonosis. *Proceedings of the Royal Society B: Biological Sciences* **267**, 2219-2230, doi:10.1098/rspb.2000.1272 (2000).
- 59 Xu, L. *et al.* The trophic responses of two different rodent–vector–plague systems to climate change. *Proceedings of the Royal Society of London B: Biological Sciences* **282**, 20141846 (2015).
- 60 Xu, L. *et al.* Wet climate and transportation routes accelerate spread of human plague. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, 20133159 (2014).

- 61 Whittles, L. K. & Didelot, X. Epidemiological analysis of the Eyam plague outbreak of 1665–1666. *Proc. R. Soc. B* **283**, 20160618 (2016).
- 62 Roosen, J. & Curtis, D. R. Dangers of Noncritical Use of Historical Plague Data. *Emerging Infect. Dis.* **24**, 103 (2018).
- 63 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* **110**, 15758–15763, doi:10.1073/pnas.1314445110 (2013).
- 64 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, doi:10.1098/rstb.2013.0624 (2015).
- 65 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3, doi:10.1093/nar/gkr771 (2012).
- 66 Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).
- 67 Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* **5**, 337, doi:10.1186/1756-0500-5-337 (2012).
- 68 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
- 69 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498, doi:10.1038/ng.806 (2011).
- 70 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* **6**, 80–92 (2012).
- 71 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874, doi:10.1093/molbev/msw054 (2016).
- 72 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, doi:10.1093/bioinformatics/btu033 (2014).
- 73 Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences* **17**, 57–86 (1986).
- 74 Team, R. D. C. R: A Language and Environment for Statistical Computing: Vienna, Austria. (2015).
- 75 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- 76 Kingman, J. F. C. The coalescent. *Stoch Process Their Appl* **13**, 235–248 (1982).
- 77 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, doi:10.1093/molbev/msi103 (2005).
- 78 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- 79 Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).
- 80 R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2015).
- 81 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
- 82 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645, doi:10.1101/gr.092759.109 (2009).
- 83 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* Springer-Verlag. *New York* (2009).

Acknowledgements

We thank Aditya K. Lankapalli, Felix M. Key and Stephen Clayton for computational analysis support. We thank Guido Brandt, Antje Wissgott, Cäcilia Freund and Marta Burri for laboratory support. We are grateful to Monica Green for critical comments on the manuscript. We thank Rafail' M. Fattahov for facilitating excavations of the Laishevo III archaeological site, to 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 specimens. 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. 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.Harbeck, K.I.B. and J.K. conceived the investigation. M.A.S., M.K., R.T. and E.A.N. performed laboratory work. M.S., M.K., A.A.V and A.H. performed data analysis. D.W., A.A., N.C., H.F., M.G., R.H., M.Henderson, K.v.H., S.K., E.L.K., J.P., D.C., S.L., M.Harbeck performed anthropological analysis as well as identified and provided access to appropriate archaeological material, A.A., J.S, K.v.H., C.L., facilitated excavations and provided access to unpublished archaeological information, M.S., M.K. and K.I.B wrote the paper with contribution from all co-authors.

Competing interests

The authors declare no competing financial interests.