Yersinia pestis genomes of the first and second plague pandemic recovered from ancient DNA

Dissertation

to fufill the requirements for the degree of
"doctor rerum naturalium" (Dr. rer. nat.)

Submitted to the council of the
Faculty of Biological Sciences
of the Friedrich Schiller University Jena

by Marcel Keller, B.Sc. M.Sc.
born on the 3rd of June 1989 in Tübingen, Germany

Reviewers:

1. Prof. Dr. Johannes Krause (Max Planck Institute for the Science of Human History, Jena)
2. Prof. Dr. Thilo M. Fuchs (Institute of Molecular Pathogenesis, Friedrich Loeffler Institute, Jena)
3. Prof. Dr. Joachim Burger (Institute of Anthropology, Johannes Gutenberg University Mainz)

Date of defense: 3rd of July, 2020

Tout ce que l'homme pouvait gagner au jeu de la peste et de la vie,
c'était la connaissance et la mémoire.

So all a man could win in the conflict between plague and life
was knowledge and memories.

Alles, was der Mensch beim Spiel der Pest und des Lebens gewinnen konnte,
waren Erkenntnis und Erinnerung.

Albert Camus, La Peste (1947)

Table of Contents

Summary 7

Zusammenfassung 8

1 Introduction 9

1.1 Diseases in History – History of Diseases 9

1.2 Ancient DNA 10

1.3 Ancient pathogen genomics 11

1.4 Plague and its causative agent, *Yersinia pestis* 13

1.4.1 Plague in Antiquity and the First Pandemic 13

1.4.2 The Second Pandemic 14

1.4.3 The Third Pandemic and its legacy 15

1.4.4 The biology of plague and the genetic makeup of *Yersinia pestis* 15

1.4.5 Archaeogenetics and the evolutionary history of plague 18

2 Objectives 22

3 Overview of Manuscripts 23

4 Manuscript A: Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750) 27

5 Manuscript B: Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes 39

6 Manuscript C: Ancient *Yersinia pestis* genomes provide no evidence for the origins or spread of the Justinianic Plague 55

7 Manuscript D: Von der Seuchengeschichte der Pest zu einer Naturgeschichte ihres Erregers – Neue Einblicke durch alte DNA 71

8 Discussion 90

8.1 Archaeogenetic research on *Yersinia pestis* 90

8.2 The origins of the First and Second Pandemic 93

8.3 Progression and persistence of the First and Second Pandemic 98

9 Concluding Remarks and Outlook 105

References 107

Ehrenwörtliche Erklärung 125

Acknowledgements 126

Appendix 128

Supplementary Material for Manuscript A 128

Supplementary Material for Manuscript B 210

Supplementary Material for Manuscript C 262

Summary

Infectious diseases and epidemic events have shaped populations, societies, minds and environments throughout human history. However, studies on the history of diseases were in the past limited to indirect evidence due to the elusiveness of the causative agents. This has fundamentally changed through the recent emergence of ancient pathogen genomics from the field of palaeogenetics. This methodology allows us to reconstruct genomes of bacteria, viruses and protozoans from ancient DNA, retrieved from human remains.

This dissertation focuses on *Yersinia pestis*, the causative agent of bubonic plague, responsible for at least two historical pandemics: the First Pandemic (541–750) starting with the Justinianic Plague in the Mediterranean basin; and the Second Pandemic, starting with the infamous Black Death (1346–1353) and ravaging through Europe for more than four centuries. Previous studies could already identify *Y. pestis* as causative agent of both pandemics but left many questions unanswered that shall be addressed here in three papers, an essay and the final discussion.

The first paper presents eight *Y. pestis* genomes of the First Pandemic, retrieved from sites in England, France, Germany and Spain, covering at least the first century of this pandemic. The results suggest that the Justinianic Plague (541–544) already reached the British Isles, show that the causative lineage diversified early during the pandemic in multiple strains, and give indications for the persistence of plague in Europe or close-by. A deletion discovered in the youngest strains might hold clues for ecological adaptations.

The second paper offers new insights into the onset and progression of the Second Pandemic through 34 ancient genomes, recovered from ten sites in England, France, Germany, Russia and Switzerland dating to the 14th–17th centuries. A genome from Russia testifies the initial entry through East Europe, the low diversity during the Black Death shows a rapid spread. The close relationship of all Second Pandemic strains suggests a local persistence and diversification. A deletion in one clade similar to the one detected in the first paper, coinciding with an accelerated substitution rate, could be interpreted as convergent evolution.

The third paper challenges previous claims in a recently published paper about the origin of the Justinianic Plague through a reanalysis of the two presented genomes. The phylogenetic analysis of one sample suggests rather an identification as a strain potentially basal to the Black Death.

The essay gives a short introduction on the history of plague research and a comprehensive overview of the recent discoveries of archaeogenetic studies, including insights into the evolution of the bacterium enabled by prehistoric plague genomes.

In the last part, the results of the papers are discussed in a more comparative approach and broader framework, focusing on the three topics – methodological challenges, origins of the pandemics, and progression and persistence – critically assessing the limitations of the archaeogenetic approach and offering an outlook of future research directions.

Zusammenfassung

Durch die Geschichte hindurch hatten Infektionskrankheiten und Epidemien einen prägenden Einfluss auf Populationen, Gesellschaften, Mentalitäten und unsere Umwelt. Die Seuchengeschichte war in der Vergangenheit jedoch weitgehend auf indirekte Zeugnisse beschränkt, da die Erreger schwer greifbar sind. Dies hat sich jüngst durch die Entwicklung der ‚Paläopathogenomik' aus der Paläogenetik fundamental verändert. Diese Methodik erlaubt die Rekonstruktion von bakteriellen, viralen und Protozoen-Genomen aus alter DNA, gewonnen aus menschlichen Überresten.

Diese Dissertation widmet sich *Yersinia pestis*, dem Erreger der Beulenpest, verantwortlich für mindestens zwei historische Pandemien: Die Erste Pandemie (541–750), die mit der Justinianischen Pest im Mittelmeerraum begann, und der Zweiten Pandemie, die nach dem berüchtigten Schwarzen Tod (1346–1353) für mehr als 400 Jahre in Europa wütete. Frühere Studien konnten bereits *Y. pestis* als Erreger beider Pandemien identifizieren, ließen jedoch viele Fragen offen, die hier in drei Studien, einem Essay und der abschließenden Diskussion angesprochen werden sollen.

Die erste Studie stellt acht *Y.-pestis*-Genome der Ersten Pandemie von Fundorten in Deutschland, England, Frankreich und Spanien vor, die zumindest das erste Jahrhundert dieser Pandemie abdecken. Die Ergebnisse legen nahe, dass bereits die Justinianische Pest (541-544) die Britischen Inseln erreicht hat, dass der verantwortliche Stamm bereits im frühen Stadium der Pandemie diversifizierte, und dass die Pest möglicherweise in Europa oder benachbarten Regionen überdauerte. Eine Deletion in den jüngsten Stämmen könnte auf eine ökologische Anpassung hindeuten.

Die zweite Studie bietet neue Einblicke in den Beginn und Verlauf der zweiten Pandemie durch 34 alte Genome aus Deutschland, England, Frankreich, Russland und der Schweiz, datiert auf das 14.–17. Jh. Ein Genom aus Russland bezeugt den Eintrag der Pest über Osteuropa, die geringe Diversität während des Schwarzen Todes eine rasante Ausbreitung. Die enge Verwandtschaft aller Stämme der Zweiten Pandemie deutet an, dass der Erreger lokal überdauerte und diversifizierte. In einer Klade wurde neben einer erhöhten Substitutionsrate eine Deletion ähnlich jener der ersten Studie beobachtet, was möglicherweise als konvergente Evolution interpretiert werden kann.

Die dritte Studie ficht die Behauptungen einer kürzlich publizierten Studie zum Ursprung der Justinianischen Pest durch eine Neuanalyse der vorgestellten Genome an. Die phylogenetische Analyse eines der Genome zeigt vielmehr, dass es basal zum Schwarzen Tod fallen könnte.

Der Essay bietet einen kurzen Abriss zur Geschichte der Pestforschung und eine umfassende Übersicht über die neuesten Entdeckungen der Archäogenetik, darunter auch die Einblicke in die Evolution des Erregers, die durch prähistorische Pestgenome gewonnen wurden.

Im letzten Teil werden die Ergebnisse der Studien vergleichend und in einem weiteren Rahmen unter drei Schwerpunkten – methodische Herausforderungen, Ursprünge der beiden Pandemien sowie Verlauf und Persistenz – diskutiert, mit einem kritischen Blick auf die Grenzen des archäogenetischen Ansatzes und einem Ausblick auf zukünftige Forschungsfelder.

1 Introduction

1.1 Diseases in History – History of Diseases

Written record of epidemic diseases goes back to the very first sources known to us, such as the Akkadian Epic of Atra-Hasis from 18th century BCE (Lambert et al., 1969), and back-references on past epidemics can be found already in Late Antiquity, e.g., in Procopius imitating Thucydides' description of the Plague of Athens in 540 BCE, when writing about the Justinianic Plague almost one thousand years later (Little, 2007). The existential threat of infectious diseases in contrast to the invisibility and inexplicability of their causes provoked contemporaries to record their experience and is held responsible for significant shifts in the history of mentalities (Bergdolt, 1994; Meier, 2016).

The emerging scientific interest in the history of diseases in the early 19th century can be seen as an outgrowth of medical history, but also the context of the second cholera pandemic, hitting Central Europe in 1831, is apparent: Justus Hecker directly refers to this contemporaneous modern pandemic (*Weltseuche*) in the preface of his book *"Der schwarze Tod im vierzehnten Jahrhundert"* (Hecker, 1832), which can be seen as the founding document of what he calls "historical pathology". The rise of germ theory with early pioneers like Ignaz Semmelweis, John Snow and finally Louis Pasteur further intensified the scientific engagement with the history of diseases (Lederberg, 2000). While the new knowledge about the cause and transmission of diseases opened new perspectives on historical pandemics, studying them in turn seemed promising for understanding and coping with contemporaneous threats.

Similar to the microbiological revolution of the mid-19th century, attempts to recover ancient DNA of pathogens from human remains in the last 20 years fueled again the history of disease, offering completely new and independent perspectives on the past. The establishment of ancient pathogen genomics led on the one hand to a specialization into the evolutionary history of pathogens; on the other hand, there is a growing interest to integrate it into consilient approaches on human history. The concept of consilience, popularized by Wilson, 1998, is to align multiple lines of evidence of different disciplines to explain and understand complex phenomena and is increasingly utilized to bridge humanities and natural sciences. Although notoriously in suspicion of determinism, consilience was particularly successful for studying the past, shifting from a mere history of events or classical archaeology to a more holistic view on the history on human and environmental history by integrating the natural scientific methods, from radiocarbon dating over isotope analyses to complex climate reconstructions (Izdebski et al., 2016).

Strikingly, from Hecker's publication in 1832 onwards, plague *sensu stricto* became a focal point for the history of disease, which can be explained by the fulminant impact and long-lasting

impression of the Black Death in Europe (cf. Bergdolt, 1994), but also by the pathognomonic buboes which allowed the retrospective diagnosis from historical documents – later verified by ancient DNA as shown in the following chapters. Archaeogenetic approaches were of course facilitated by preceding historical and archaeological research, e.g., by identifying plague burials (Hawkins, 1990); and by coincidence, *Yersinia pestis* DNA turned out to be comparatively easy to retrieve from skeletal material due to common fulminant and lethal bacteremia after infection. Not least because of that, *Y. pestis* became a 'model organism' to study pathogens and diseases in the past.

Referring to the epigraph of this dissertation, infectious diseases are major agents in human history, and keeping the *memory* about the devastating consequences of pandemics alive gives us motivation for our research. However, the *knowledge* that we gain by studying the history of diseases expands constantly beyond mere historical questions, from scholarly studies on human reaction and resilience in the face of disasters to insights in the evolutionary history of their agents.

1.2 Ancient DNA

The field of palaeogenetics was established by two pioneering works by Higuchi et al., 1984, who sequenced DNA of a museum specimen of the extinct quagga; and by Pääbo, 1985, who was the first to sequence allegedly ancient human DNA of an Egyptian mummy (now thought to be contamination; see Pääbo et al., 2004; Zagorski, 2006).

As realized already back then, ancient DNA features three main characteristics and challenges, since the DNA of an organism is neither replicated nor repaired after death. These are low concentration, fragmentation and damage due to molecular degradation (Pääbo et al., 2004; Willerslev and Cooper, 2005). Fragmentation is caused either by hydrolytic cleavage of phosphodiester bonds in the backbone or depurination of glycosidic bonds causing abasic sites, susceptible to strand breaks by β-elimination, both leading ultimately to the disintegration of DNA. Damage includes the described abasic sites as well as hydrolysis of bases, such as cytosine to uracil, leading to misincorporation of thymine, when replicated *in vitro* (Brotherton et al. 2007). Maillard reactions are responsible for crosslinks between DNA and proteins or with itself, making the DNA inaccessible for molecular biological methods.

At least the low concentration could be overcome to some degree by the parallel development of the polymerase chain reaction (PCR) which allowed the amplification of DNA fragments (Pääbo et al., 1989). However, in the following decades, the prospect of finding 'molecular fossils' – evolutionary snapshots of past millennia, led to a gold rush with studies going back thousands and thousands of years, ultimately presenting alleged DNA of million years old dinosaurs (Woodward et al., 1994) or amber-embedded insects (DeSalle et al., 1992). In the next two decades, the excitement was gradually replaced by disillusion and several published aDNA results were identified as

artefacts, mainly due to modern contamination or amplification of non-specific sequences (cf. Austin et al., 1997; Hedges and Schweitzer, 1995; Zischler et al., 1995). This led to the establishment of strict quality standards for the performance and publication of ancient DNA studies; and two of the challenging characteristics of ancient DNA – the short fragment size and the specific damage pattern – were eventually utilized as sanity checks (Willerslev and Cooper, 2005). The quick adoption of high throughput sequencing techniques and DNA hybridization enrichment finally caused the recent transformation of palaeo*genetics* into palaeo*genomics*, allowing for the complete reconstruction of ancient genomes (Hofreiter et al., 2015), at the same time also facilitating the authentication of ancient DNA with the described criteria – fragment length and damage pattern – and allowing contamination estimation (Briggs et al., 2007; Green et al., 2009).

A breakthrough was the sequencing of the Neanderthal genome (Green et al., 2010) and the previously unknown Denisovan (Krause et al., 2010), offering new insights into the evolution of archaic, but also modern humans by tracing ancient introgressions of both hominins (Green et al., 2010; Reich et al., 2010). Studies on Neolithic and Bronze Age human remains were able to track migrations in prehistory that were previously opaque or controversial to archaeologists (Allentoft et al., 2015; Brandt et al., 2013; Haak et al., 2015). Studies on animal and plant remains were not only able to offer new findings about extinct species (Heintzman et al., 2017; Meyer et al., 2017; Noonan et al., 2005), but also allowed us to comprehend the processes of domestication in space and time (Mascher et al., 2016; Ottoni et al., 2013).

A whole new subject arose from the study of ancient microbes; not only pathogens, as discussed in the next chapter, but also commensal microbial communities reconstructed from dental calculus or palaeofeces (Warinner et al., 2015). In parallel, the field experienced a diversification of source materials such as parchment (Teasdale et al., 2017), archaeobotanical remains (Sønstebø et al., 2010), collection specimens (Guschanski et al., 2013) or sediments (Slon et al., 2017).

1.3 Ancient pathogen genomics

Already in the first two decades of palaeogenetics, researchers attempted to recover not only DNA of hominins, other animals and plants, but also their microbial pathogens. For physical anthropologists, this approach was of particular interest, since most infectious diseases do not leave macroscopic traces on the skeleton, or only lead to nonspecific reactions such as *cribra orbitalia* (Waldron, 2008). Counterintuitively, the absence of skeletal lesions or osteological stress markers in a population can even be an indicator of higher frailty, since fulminant progression of diseases would lead to death before leaving traces on the bone, a phenomenon described as 'osteological paradox' (Woods et al., 1992). Therefore, early studies followed the paradigm of a molecular approach in palaeopathology, providing an independent diagnostic tool especially for morphologically ambiguous

or invisible diseases. However, early PCR studies implicitly covered already broader research questions, such as the Columbian Exchange of infectious diseases between the Old and New World (Salo et al., 1994), although such studies might not withstand modern authentication requirements. The first publication of a fully reconstructed ancient pathogen genome, a *Yersinia pestis* genome recovered from a plague pit associated with the Black Death in 1348–1350 (Bos et al., 2011), ultimately opened the field towards the molecular evolutionary history of pathogens.

Today, the full reconstruction of ancient pathogen genomes allows not only for the precise phylogenetic placements, but also refined molecular dating of MRCAs and determination of substitution rates by calibrating with centuries- or millennia-old strains. One example is *Mycobacterium lepra*e, the causative agent of leprosy, where aDNA studies showed a remarkable genomic conservation (Schuenemann et al., 2013). The reconstruction of a pathogen's diversity in the past furthermore enables temporal-phylogeographic studies: Although tuberculosis lesions are known from pre-Columbian skeletal remains (Roberts and Buikstra, 2008), the modern strains in the Americas descend from European lineages (Hershberg et al., 2008). *Mycobacterium tuberculosis* genomes reconstructed from such remains were found to be closely related to strains found in pinnipeds of the Southern Hemisphere with a MRCA around 5000 BP (Bos et al., 2014), suggesting that tuberculosis might have reached the Americas in the Holocene by zoonotic transmission, before replacement through European strains. Moreover, the recovered genomes can be examined for their genetic makeup (like presence of virulence factors or genome decay) or recombination events, as shown for closely related substrains of *Treponema pallidum,* including the causative agents of venereal syphilis, yaws and bejel (Schuenemann et al., 2018).

With the exception of plague, where retrospective diagnosis was facilitated by contemporary descriptions of buboes and massive mortality spikes during the Second Pandemic (14^{th}–18^{th} centuries) allowed for the identification of known or presumed plague pits, archaeogenetic research on acute infectious diseases was very limited until powerful metagenomic pipelines became available. With tools such as MALT (Vågene et al., 2018), MetaPhlAn (Segata et al., 2012), or Kraken (Wood and Salzberg, 2014), it is possible to screen shotgun sequencing datasets for the presence of hundreds or thousands of taxa in parallel. By this approach, Vågene et al., 2018 were able to detect *Salmonella enterica* in an early-contact epidemic cemetery in southern Mexico, identifying Paratyphus C as a possible cause for *Cocoliztli*, an epidemic that diminished the indigenous populations of post-Columbian Mesoamerica. Other studies were able to identify previously unknown lineages of *Yersinia pestis* in Neolithic and Bronze Age burials initially sampled for human DNA (Andrades Valtueña et al., 2017; Rascovan et al., 2019; Rasmussen et al., 2015), or recover a *Borrelia recurrentis* genome from a 15^{th}-century double burial from Oslo with an untargeted approach (Guellil et al., 2018).

Recent approaches focused also on DNA viruses, presenting ancient genomes for the variola (Duggan et al., 2016) and the Hepatitis B virus (Krause-Kyora et al., 2018; Mühlemann et al., 2018),

or eukaryotic pathogens such as *Plasmodium falciparum* (Marciniak et al, 2016). Still underexplored remain RNA viruses. The additional hydroxyl group of ribose makes the ribose-phosphate backbone much more susceptible to hydrolysis, therefore RNA is not expected to preserve well in archaeological remains. Exceptions are the HIV and Influenza virus genomes retrieved from archival specimens (Reid et al., 1999; Worobey et al., 2016) and a permafrost burial (Taubenberger et al., 2005; Tumpey et al., 2005;).

1.4 Plague and its causative agent, *Yersinia pestis*

1.4.1 Plague in Antiquity and the First Pandemic

The history of plague is traditionally divided into three pandemics: the First Pandemic (6th–8th c.), the Second Pandemic (14th–18th c.) and the Third Pandemic (19th–21st c.). Both the First and Second Pandemic started with fulminant outbreaks spreading over the whole Mediterranean basin and Europe within a few years, called the Justinianic Plague (541–544) and Black Death (1346–1353).

Speculations about previous plague epidemics go back to scholars such as Heinrich Haeser (Haeser, 1875) and are based primarily on historical records and medical texts of Antiquity. Especially the Plague of Athens (430–428 BCE), the Antonine Plague (165–180 CE) and the Plague of Cyprian (249–270 CE) were subject to attempts in retrospective diagnosis but could not convincingly be identified as plague *sensu stricto* (Harper, 2015; Littman and Littman, 1973; Morens and Littman, 1992). The only known written record that withstands modern philological critique has been identified in a text by the physician Rufus of Ephesus (fl. 100 CE) citing works of Dionysius Kurtos, Posidonius, and Dioscorides on bubonic plague in "Libya, Egypt and Syria" (Mulhall, 2019).

The first commonly accepted record of the Justinianic Plague is for summer 541 in Pelusium, Egypt, as documented by Procopius, although there are accounts for plague to have started in "Ethiopia" (Sarris, 2002). John of Ephesus left us vivid descriptions of how the pandemic spread further along the Mediterranean coast, along his travel route from Alexandria to Constantinople via Palestine and Syria (Little, 2007). The pandemic spread further West, affecting Illyricum, Italy, Gaul and Spain, but also North Africa within the next two years (Sallares, 2007). Even records of unspecified mass mortalities in 544 in Ireland and Britain were read as testimonies of the Justinianic Plague (Dooley, 2007; Maddicott, 2007). For the following decades, reports of numerous outbreaks both in the Byzantine Empire and the West have come down to us (Harper, 2017) prompting historians repeatedly to assign them to "waves" (Stathakopoulos, 2004) or "amplification events" (Harper, 2017), till the pandemic ended in a final blaze in 743–750. The identification of epidemic events in literary sources as well as the demographic, political and cultural impact of the First Pandemic are matter of ongoing scholarly dispute. While the raising interest in the First Pandemic and therefore

emerging evidence (Stathakopoulos, 2004) lead to far-reaching claims as paraphrased in the volume title "Plague and the End of Antiquity" (Little et al., 2007), others argued for a more critical assessment and virtually dismissed much of the adjudged consequences (Durliat, 1989; Meier, 2016). The preliminary peak of this discussion was reached with Harper, 2017, branded as 'maximalist' and extensively criticized by 'minimalist' colleagues (Haldon et al., 2018; Mordechai and Eisenberg, 2019).

1.4.2 The Second Pandemic

For almost 600 years, between 750 and 1346, we lack any records testifying bubonic plague in Europe or its neighboring regions, prompting the conclusion of a complete vanish of *Y. pestis* from the continent. This is even more striking with regard to the well-documented onset of the Second Pandemic, the Black Death in 1346–1353. In difference to the Justinianic Plague, the Black Death entered Europe via the Black Sea, this time however with scarce evidence for its origin and itinerary: Around lake Issyk Kul (today the borderland between Kazakhstan and Kyrgyzstan), Nestorian cemeteries were discovered in the late 19th century with gravestone inscriptions pointing towards plague as a cause of death, dating to 1337–1339, which was often referred to as a prelude of the Black Death (Slavin, 2019). More securely linked to the European pandemic are the outbreaks in Sarai and Astrakhan, by then part of the Khanate of the Golden Horde, in 1346 (Benedictow, 2004). The Genuese notary Gabriele de Mussis reports the famous story about the siege of Caffa (modern Feodosia, Crimea) by Tartars, who, falling sick of plague themselves, catapulted corpses into the city, often cited as an early example of biological warfare (Wheelis, 2002). Merchants then spread the disease on maritime trade routes first to Mediterranean port cities, reaching Alexandria, Constantinople and Messina in 1347 and from there to Genoa, Venice, Marseille and Barcelona. By 1353, plague had spread across the whole of Europe including the British Isles, Scandinavia and Western Russia, leaving only a few regions unaffected (Benedictow, 2004). The high mortality rate, estimated up to 60 %, had a strong impact on the demography and economy of Medieval Europe, which is even mirrored in lead pollution records (More et al., 2017).

The Black Death was only the onset of a pandemic that lasted at least until the 18th century, affecting large areas, e.g., in the *pestis secunda* in 1356–1366, during the Thirty Years' War (1618-1648) or in Russia in 1770–1772 (Alexander, 1986; Biraben, 1975). Other notable outbreaks, like the Great Plague of London in 1665, or the Great Plague of Marseille in 1720–1722 appear more focalized. Lately, more than 7000 plague occurrences have been collected for the Second Pandemic (Schmid et al., 2015), certainly still underestimating the true number due to biased data acquisition (Roosen and Curtis, 2018).

1.4.3 The Third Pandemic and its legacy

The beginning of the Third Pandemic is matter of definition: Traditionally it is set in context of a Muslim revolt in Yunnan in 1855, when the suppression by military troops and forced migrations enabled the transmission of plague beyond the local focus. However, plague was present in Yunnan at least since 1772 (Ben-Ari et al., 2012) causing outbreaks of epidemic character. In the second half of the 19th century, plague spread to the Qinghai and Guangdong provinces, notably affecting Hong Kong in 1894. With the outbreak in Hong Kong, plague finally became subject to international attention. Kitasato Shibasaburō, a former student of Robert Koch, was sent by the Japanese government to Hong Kong in 1894 to identify the causative agent. Alexandre Yersin, a Swiss physician working for the *Institut Pasteur* and sent by the French government, arrived only a few days later. Although Kitasato published his discovery first, his characterization of the bacterium was inaccurate in several morphological aspects, so credit was later given to Yersin who published his results shortly after Kitasato (Kupferschmidt, 1993). Notably, already Yersin made the connection of the Hong Kong plague with the European plague outbreaks of the 14th–18th centuries (Yersin, 1894).

In the following years, national and international plague commissions intensified research on plague facing the worldwide spread of the Third Pandemic. The accelerated and intensified maritime transport via steamships allowed the bacterium for the first time in its history to reach every inhabited continent. Only Europe and Australia were able to contain local outbreaks through rigid sanctions and are considered to be plague-free today. However, both North and South America, Africa and Asia are since then home to countless natural reservoirs of *Y. pestis*. The worldwide dissemination from a common origin and recurrent local epidemic outbreaks are reasons to call the Third Pandemic still ongoing, although the worldwide yearly death toll has been constantly in the lower hundreds in recent years (WHO, 2016, 2012). However, the 2017 plague outbreak in Madagascar with suspected 2417 cases and 209 deaths received worldwide attention and was classified by the WHO as a class 2 emergency (WHO, 2017).

1.4.4 The biology of plague and the genetic makeup of *Yersinia pestis*

The microbiological research in plague started in the last years of the 19th century in China, linked to the so-called Third Pandemic, with Alexandre Yersin. It was only four years after Yersin's discovery that Paul-Louis Simond described the flea-borne transmission from black rats (*Rattus rattus*) to humans (Simond, 1898). Bacot and Martin then established in 1914 the classical paradigm of the 'blocked flea': The oriental rat flea, *Xenopsylla cheopis*, sucks blood from a plague-infected rat. Mediated by the abrupt temperature change from 37°C in the mammalian host to less than 26°C in the arthropod, *Y. pestis* builds up a biofilm in the flea's foregut while proliferating for the next days, causing a full blockage of the flea's ability to ingest and digest further blood meals. When trying to

suck blood from another host, the blockage forces the flea to regurgitate contaminated blood back into the host. The inability to ingest blood leads the starving flea to repeated attempts to suck blood, promoting the dissemination of the bacterium, until the flea eventually dies. However, as iconic as this transmission model became, it is overly simplistic and does not account for multiple factors: The natural reservoir of *Y. pestis* is in most regions not the well-studied black rat but wild rodent species such as gerbils, marmots and voles, carrying a variety of different, host-specific fleas (Pollitzer, 1952a). *Y. pestis*' capacity of infection, biofilm formation and transmission vary significantly between different flea species and the mammalian blood donor (Bland et al., 2018). Furthermore, the 'early phase transmission' was proposed as an alternative model which does not require biofilm formation and is thought to account for transmission without blockage or observed rapid spread of the disease in rodent populations (Eisen et al., 2015). The complex relationship of different transmission modes, ecology and immunology of different vectors and hosts remains a subject for future research.

Human infection with modern *Y. pestis* presents in three main forms: bubonic, septicaemic and pneumonic plague. The clinical manifestation is however not a consequence of genetically differing strains, but primarily related to the mode of transmission. Infection via flea bite typically leads to the bubonic form. The subcutaneously injected bacteria travel to the nearest lymph node either by lymphatic flow or by intracellular carriage in phagocytic cells while actively suppressing a host innate immune response. In the lymph node, the bacteria replicate, causing a palpable or even visible swelling, the eponymic *bubo*, and disseminate further via the lymphatic system. After this first pre-inflammatory stage, *Y. pestis* enters the inflammatory stage characterized by spread into the bloodstream, leading to a secondary septicaemia. What follows is a fulminant progression by a massive immune response and cell death induced by bacterial endotoxins, leading ultimately to multi-organ failure and death (Demeure et al., 2019). Cutaneous infection can also result in the less common primary septicaemic plague without detectable lymphadenopathy and rapid progression (Dennis et al., 1999).

Both septicaemic and bubonic plague can lead to secondary pneumonic plague by invasion of and multiplication in the pulmonary tissue. This enables droplet and thus human-to-human transmission, leading to primary pneumonic plague. This disease form is the most fulminant and leads to death within 24 h without treatment and an incubation time of only 1–3 days. Similar to bubonic plague, primary pneumonic plague progresses in two phases with massive neutrophil invasion, intra-alveolar hemorrhage and edema in the inflammatory stage (Pollitzer, 1952b; Zimbler et al., 2015).

Y. pestis is one of three *Yersiniae* species that are considered pathogenic to humans. However, the other two – *Y. pseudotuberculosis* and *Y. enterocolitica* – cause comparatively mild and most often self-limiting gastrointestinal infections. The genus *Yersinia* currently consists of 18 species; other species of this genus are pathogenic to fish (*Y. ruckeri*) or insects (*Y. entomophaga*) or are non-

pathogenic environmental species (McNally et al., 2016). The evolutionary history of *Y. pestis* is remarkable since it forms a clonal lineage within the diversity of *Y. pseudotuberculosis* that evolved rather recently (Achtman et al., 1999). Within this period, *Y. pestis* became highly adapted to its peculiar ecology while losing characteristic features of its evolutionary past as environmental or enteric bacterium, both by acquisition of virulence factors and deletion and pseudogenization of genes (Hinnebusch et al., 2016).

From its ancestor *Y. pseudotuberculosis*, *Y. pestis* inherited the pYV/pCD1 virulence plasmid (70 kb) which was also acquired two times independently by *Y. enterocolitica* (Reuter et al., 2014). pCD1 encodes a type 3 secretion system (T3SS) that forms a needle-like structure on the cell surface and allows the bacteria to inject effector proteins into host cells (Tardy et al., 1999). These effector proteins, here called Yops (*Yersinia* outer proteins) are crucial for the prevention of phagocytosis and suppression of the innate immune response by inhibition of cytokine response. Similar T3SS are shared between multiple pathogenic species such as *Salmonella* or *Shigella* and acquired by horizontal gene transfer via plasmids or pathogenicity islands (Gophna et al., 2003).

The most significant genetic change in the evolution of *Y. pestis* is the acquisition of two additional virulence plasmids, pMT1 (100 kb) and pPCP1 (9.6 kb). pMT1 encodes for the Yersinia murine toxin (Ymt) and the F1 capsular protein (Caf1). Ymt, a phospholipase D, protects the bacterium in a flea's midgut against cytotoxic byproducts of blood plasma digestion (Hinnebusch et al., 2002). Caf1 is a polymeric protein that forms a capsule on the cell membrane in the mammalian host and seems to enhance virulence in flea-mediated transmission (Sebbane et al., 2009), potentially by antiphagocytic or immunomodulatory functions (Sha et al., 2011). The smallest plasmid, pPCP1, encodes the plasminogen activator (Pla), a surface protease that is necessary to cause pneumonic plague, as shown for atypical pPCP1-deficient strains of *Y. pestis* such as Pestoides F (Zimbler et al., 2015). By cleavage of host Plasminogen to Plasmin it induces fibrinolysis and allows the bacteria to escape fibrin clots during infection and thus rapid proliferation in lung tissue (Sebbane et al., 2006). However, a second function of this protein seems to be associated with subcutaneous dissemination in case of flea-borne transmission. This is apparent from a single nucleotide mutation, I259T, that is shared among all derived non-*Pestoides Y. pestis* strains. *Pestoides* strains show a significant impairment in invasive infection during bubonic plague (Zimbler et al., 2015). Since the pPCP1 plasmid occurs in high copy number at least in the reference strain CO92 (Parkhill et al., 2001), *pla* is commonly used for plague detection in PCR or qPCR assays in medical or archaeogenetic contexts (Demeure et al., 2019; Schuenemann et al., 2011).

Besides these events of gene gain, *Y. pestis* shows a characteristic profile of gene loss or pseudogenization in adaptation to its lifestyle. An attempt to detect orthologs of functional genes through sequence alignment in the reference *Y. pestis* strain CO92 revealed a number of 337 pseudogenes (Lerat and Ochman, 2005), suggesting that pseudogenization is an important mecha-

nism in the evolution of *Y. pestis.* Pseudogenization events are thereby often caused by insertion elements (such as IS100), fundamental drivers of genomic rearrangements (Chain et al., 2006) and overly abundant in *Y. pestis* compared to *Y. pseudotuberculosis* (Parkhill et al., 2001).

Several of these pseudogenization events have been identified for virulence factors of *Y. pseudotuberculosis* that were detrimental or useless for *Y. pestis'* specific ecology. Examples are Invasin (Inv) and the pCD1-encoded adhesin YadA which enable enteric *Yersiniae* to invade cells of mucosal tissues (Simonet et al., 1996; Skurnik and Wolf-Watz, 1989; Wren, 2003). Another detrimental gene that was pseudogenized by a frameshift mutation early in *Y. pestis'* evolution is the urease *ureD*, killing up to 40 % of infected fleas infected with *Y. pseudotuberculosis* by metabolizing urea to cytotoxic ammonia (Chouikha and Hinnebusch, 2014). Other affected genes suppress biofilm formation encoded in the hmsHFRS locus (including *rcsA, pde2* and *pde3*), present but only active under environmental conditions in *Y. pseudotuberculosis* (Hinnebusch et al., 2016; Sun et al., 2008). Loss-of-function mutations are also found for flagellar or chemotaxis-related genes, both unnecessary for the non-motile bacterium (McNally et al., 2016; Parkhill et al., 2001).

Although these examples illustrate well how microbiological research on modern bacterial strains combined with genomic sequencing can elucidate the evolution of highly virulent pathogens, the chronology of the relevant mutations stays broadly unexplored. Comparisons between phylogenetically extant 'archaic' and 'modern' strains allow limited insights (e.g. on the I259T mutation of *pla*) but are flawed by a literal 'survivorship bias', meaning that less adapted strains might have existed but went extinct. This led to predicted models of *Y. pestis'* evolution that have been shown to be wrong by reconstructed ancient genomes (cf. Sun et al., 2014 and Hinnebusch et al., 2016).

1.4.5 Archaeogenetics and the evolutionary history of plague

Speculations about the phylogeographic history of plague can be traced back to the pre-genetic era. Although Wu Lien-Teh speculated on 'ancient' foci of plague (Wu et al., 1936), Devignat's study was the first systematic attempt: based on their ability to ferment glycerol and reduce nitrate, he divided modern *Y. pestis* strains in three biovars that he linked to historical pandemics and natural foci (Devignat, 1951). The strains causing the third pandemic were characterized as unable to ferment glycerol, originating in East Asia and therefore named *Orientalis*. The biovar capable of both metabolic functions was identified as the oldest and therefore named *Antiqua*. The peculiar geographic distribution of this biovar, today found in Central Asia and Africa, was explained by "Aryan" migrations in prehistory, causing the Philistine Plague and establishing foci in Central Africa, from where it allegedly spread during the Justinianic Plague. The biovar unable to reduce nitrate was supposed to originate in the Russian steppe and therefore associated with the Second Pandemic and named *Mediaevalis*.

With the rise of palaeogenetics in the 1980 and 90s, researchers got particularly interested in the identification of *Y. pestis* in skeletal remains. Drancourt et al. published a pioneering study in

1998, presenting allegedly ancient *Y. pestis* DNA amplificates from plague burials from 16th–18th-century France. This was followed by a number of PCR studies of the same research team (Drancourt et al., 2007, 2004; Raoult et al., 2000; Tran et al., 2011a, 2011b), covering sites dating from the Justinianic Plague till the 17th century. Surprisingly, they repeatedly claimed to have identified *Orientalis* strains in samples of both the First and Second Pandemic, contradicting the ideas of Devignat. Other research groups published PCR studies on plague detection in German sites of the First and Second Pandemic (Garrelt and Wiechmann, 2003; Pusch et al., 2004; Wiechmann et al., 2010; Wiechmann and Grupe, 2005). The findings of the First Pandemic in Aschheim were particularly remarkable, regarding the lack of historical records for that time and region.

However, the increasing doubt in ancient DNA studies with regard to possible contaminations or unspecific detections affected also these early plague studies: a re-examination of previously published and additional potential plague pits was unable to confirm previous or to present new results (Gilbert et al., 2004). Moreover, later PCR studies were not able to verify the identification of *Orientalis* strains in human remains of the First and Second Pandemic (Haensch et al., 2010; Harbeck et al., 2013; Seifert et al., 2016). Meanwhile, genotyping of modern *Y. pestis* strains drew a more complex picture of the bacterium's evolutionary history (Achtman et al., 2004): The biovars could be shown to be polyphyletic groups and therefore unusable as phylogenetic classifiers or even historical interpretations. Instead, Achtman et al. introduced a new system of numbered branches, with Branch 0 forming the 'stem' leading from *Y. pseudotuberculosis* to the node with Branch 1 and 2, but also included the biovar information in their new labelling system (e.g., 1.ANT or 2.MED) which is still used today.

The sequencing of the first complete *Y. pestis* genome (Parkhill et al., 2001) and the rapid development of novel sequencing techniques were a major breakthrough, allowing for the first whole genome comparisons and phylogenies (Chain et al., 2004; Morelli et al., 2010). Morelli et al. were already able to falsify the alleged association of the First Pandemic with Central African *Antiqua* strains based on molecular dating.

The availability of a reference genome also opened the field of archaeogenetics to genomic approaches, enabling both the design of probe sets for target enrichment and short read alignment. This was first accomplished with the reconstruction of a complete *Y. pestis* genome of the Black Death (1348–1350), recovered from the known plague cemetery East Smithfield, London (Bos et al., 2011). The authors showed that the Black Death genome falls in an ancestral position to all Branch 1 genomes including the ones associated with the Third Pandemic, only two SNPs derived from the node with Branch 0 and 2. With the analysis of 133 new modern *Y. pestis* genomes, (Cui et al., 2013) were able to show that two more branches, Branch 3 and 4, emerged from the same node, indicating that such polytomies might be associated with epidemics. With the Black Death as an additional calibration point, they could also show drastic variations in the molecular clock rate, again strikingly coinciding with the Second and Third Pandemic. Based on the geographic origin of the

most basal modern strain, they identified the Qinghai-Tibet plateau in China as potential 'homeland' of plague and the trade routes crossing there as facilitating its spread.

After the clear identification of *Y. pestis* as causative agent of the Black Death, research focused on post-Black Death outbreaks in Europe: early archaeogenetic (Haensch et al., 2010) and climatic data (Schmid et al., 2015) supported a model of multiple reintroductions of plague into Europe, contradicting common ideas of plague persistence in a local reservoir (Biraben 1975; Biraben and Le Goff 1969). However, in 2016 several studies on post-Black Death *Y. pestis* genomes from France and Germany found evidence for plague persistence in Europe, since they formed a phylogenetic clade emerging from the Black Death with no known extant strains (Bos et al., 2016; Seifert et al., 2016; Spyrou et al., 2016). The conundrum about the strains found in Bergen op Zoom (Haensch et al., 2010), London St Mary Graces (Bos et al., 2011; cf. Green and Schmid, 2016), and Bolgar (Spyrou et al., 2016), all further derived on modern Branch 1, was finally explained by a subbranch travelling back to Asia after the Black Death (Spyrou et al., 2016). The *pestis secunda*, to which these genomes were assigned, though remains a matter of dispute: Namouchi et al., 2018 interpreted the data as supportive for reintroductions from the Caspian Sea region.

A whole genome reconstruction for the Justinianic Plague in Aschheim (Wagner et al., 2014) could verify not only previous PCR studies on this site, but also the proposed phylogenetic position (Harbeck et al., 2013) between two modern lineages, 0.ANT1 and 0.ANT2, on the tip of a comparatively long branch. In addition, Wagner et al. suggested a phylogeographic history of plague with two independent introductions of plague into Europe during the First and Second Pandemic. Feldman et al., 2016 were able to reconstruct another genome of the First Pandemic from the close-by site of Altenerding, presenting an identical genome after exclusion of false positive SNPs in the Aschheim genome.

Whereas all previous studies had a targeted approach in examining potential or documented plague burials, new metagenomic tools allowed for the screening of shotgun sequencing data. This led to the discovery of an ancient plague lineage in Late Neolithic and Bronze Age (LNBA) burials spanning from Poland to the Altai region (Rasmussen et al., 2015). These genomes revealed not only the existence of a today extinct, basal lineage of *Y. pestis*, but also provided a valuable snapshot of *Y. pestis*' evolution. The bacterium, arguably able to cause severe bacteremia in humans, already carried all three virulence plasmids. However, it was missing the *ymt* gene on plasmid pMT1 and two loss-of-function mutations on *pde2* and *rcsA*, and therefore not efficiently transmittable via fleas, contradicting models of *Y. pestis'* early emergence as flea-transmittable pathogen and acquisition of *ymt* (Sun et al., 2014). Intriguingly, the authors also identified a distinct strain from an Iron Age site in Armenia, which – although too low in genomic coverage for phylogenetic analyses – already carried *ymt*. Spyrou et al., 2018 were able to present a full genome of such a Bronze Age *Y. pestis* strain with a 'modern' genetic set-up, dating the flea-transmissible and therefore bubonic plague back to at least 1800 BCE. Andrades Valtueña et al., 2017 presented not only six additional LNBA genomes on the

same branch as the Rasmussen et al., 2015 genomes, but also linked the geographic distribution through time to the Yamnaya expansion in the 3rd century BCE and the arrival of a European Neolithic component in Siberia 2200–1500 BCE. This was however challenged later by Rascovan et al., 2019, instead arguing for emergence of plague in the context of Neolithic decline and spread along trade routes, although their major finding, an even more basal Neolithic *Y. pestis* genome from Sweden, does not directly support this hypothesis.

Another approach to link the phylogeographic history of plague to human mobility was presented by Damgaard et al., 2018. They recovered a plague genome from a 2nd–3rd century burial in the Tian Shan region that branches basal on the First Pandemic lineage and report on a second, low coverage genome from the Caucasus region dated to the 6th–9th centuries. Based on that, they conclude that the First Pandemic might have spread along the steppe in the context of the Hunnic expansions, questioning the common narrative of an introduction via Egypt.

The First Pandemic remains vastly unexplored by archaeogeneticists. Neither the genetic diversity, nor the progression of the pandemic and the potential persistence can be investigated on the basis of the two identical Bavarian genomes published so far (see above). Furthermore, common ideas about the geographic extent and the identification of later epidemic events as bubonic plague (Little et al., 2007) were recently challenged (Mordechai and Eisenberg, 2019), underlining the research potential of archaeogenetic studies. For the Second Pandemic, the open questions concern primarily the progression after the Black Death and the ongoing dispute regarding persistence versus reintroductions. However, recent claims about the Black Death and the *pestis secunda* (Namouchi et al., 2018) should be critically assessed. This includes the reanalysis of previously published data, which has already been shown to be inevitable (Feldman et al., 2016). The determination of the origins of both pandemics is still challenging and requires a careful differentiation between the place of emergence of the causative lineage and the starting point of the human-associated phenomenon. Nevertheless, the combination of modern and ancient phylogeography with traditional historical approaches can lead to profound, evidence-based hypotheses. Finally, the increasing number of available ancient genomes will allow for comparative studies, elucidating common patterns and distinct features of both the First and Second Pandemic.

2 Objectives

The aim of this thesis is to elucidate the microevolution of *Yersinia pestis* in the context of human history, primarily during the First Pandemic (541–750) and Second Pandemic (1346–18th century), through the analysis of ancient *Y. pestis* genomes recovered from human remains. This archaeogenetic approach offers not only invaluable snapshots of the evolutionary history of one of the most disastrous human pathogens but can also complement historical and archaeological research on the two pandemics.

Manuscript A is focused on the First Pandemic. By recovering genomic data of *Y. pestis* from nine archaeological sites in England, France, Germany and Spain, the aims of this study are to investigate the geographic and temporal extent of the pandemic in comparative assessment with the historical data; to analyze the phylogenetic relationship of the reconstructed strains and possible implications for the historical phylogeography, such as local reservoirs versus reintroductions; and to analyze the discovered diversification and evolution of First Pandemic lineages in reference to other modern and ancient *Y. pestis* clades.

Manuscript B addresses the Black Death (1346–1353) and the consecutive outbreaks of the Second Pandemic. With the reconstruction of *Y. pestis* genomes from ten archaeological sites in England, France, Germany, Russia and Switzerland, this study aims to elucidate the onset and rapid spread of the Black Death; to engage in ongoing discussions about persistence and progression of the pandemic by analyzing the underlying phylogenetic diversity; and to investigate evolutionary trajectories over the course of this pandemic.

Manuscript C provides a reanalysis of two previously published *Y. pestis* genomes from the Tian Shan region and Northern Ossetia; challenges the historical interpretations presented in the original paper; and provides a previously omitted phylogenetic assessment of the second, low coverage genome, potentially representing a pre-Black Death strain.

Manuscript D locates the recent archaeogenetic approaches on plague in the broader history of plague research and gives a synopsis on the respective ancient DNA studies, including the recent discoveries on Neolithic and Bronze Age plague, but also previous versions of Manuscript A and B. Written as a review for a broad audience yet on a scientific level, this manuscript aims to highlight the major achievements of ancient plague genomics, but also to critically assess its boundaries and summarize the open questions.

Finally, the findings of all manuscripts are discussed in a more comparative approach, focusing on three overarching topics: (1) the methodological challenges of ancient pathogen genomics; (2) the question of the geographic origins of the First and Second Pandemic; and (3) controversial aspects about the progression and persistence of plague after the Justinianic Plague and the Black Death. The latter two are discussed in a broader framework as feasible in the individual manuscripts, with emphasis on the contributions and limitations of the archaeogenetic approach.

3 Overview of Manuscripts

Manuscript A: Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750)

Marcel Keller*, Maria A. Spyrou*, Christiana L. Scheib, Gunnar U. Neumann, Andreas Kröpelin, Brigitte Haas-Gebhard, Bernd Päffgen, Jochen Haberstroh, Albert Ribera i Lacomba, Claude Raynaud, Craig Cessford, Raphaël Durand, Peter Stadler, Kathrin Nägele, Jessica S. Bates, Bernd Trautmann, Sarah A. Inskip, Joris Peters, John E. Robb, Toomas Kivisild, Dominique Castex, Michael McCormick, Kirsten I. Bos, Michaela Harbeck, Alexander Herbig & Johannes Krause
*contributed equally

Published in *PNAS* 116 (25), 12363–12372 (2019), doi:10.1073/pnas.1820447116

Summary: In this paper, we present eight ancient *Y. pestis* genomes of the First Pandemic from the 6th–8th centuries. We provide evidence that the Justinianic Plague (541–544) reached already the British Isles. The genomes from France, Germany and Spain show a rapid diversification and persistence of plague in Europe or its vicinity over the course of this pandemic. In addition, we identified a genomic deletion found in the youngest genomes including two virulence-associated genes.

Author contributions: M.M., K.I.B., M.H., A.H., and J.K. designed the study; **M.K.**, M.A.S., C.L.S., G.U.N., K.N., and J.S.B. performed laboratory work; A.K. developed the new analytical tool; **M.K.**, M.A.S., and G.U.N. performed data analyses; B.T. and S.A.I. performed anthropological examination; C.L.S. and T.K. provided genomic data; B.H.-G., B.P., J.H., A.R.i.L., C.R., P.S., J.P., J.E.R., D.C., and M.H. identified and provided access to archaeological material; B.H.-G., B.P., J.H., A.R.i.L., C.R., C.C., R.D., P.S., and M.M. provided archaeological and historical information; and **M.K.**, M.A.S., M.M., and A.H. wrote the paper with contributions from all authors.

In total, Marcel Keller contributed 40 % to the laboratory work, data analysis and manuscript writing.

Manuscript B: Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes

Maria A. Spyrou*, **Marcel Keller***, Rezeda I. Tukhbatova, Christiana L. Scheib, Elizabeth A. Nelson, Aida Andrades Valtueña, Gunnar U. Neumann, Don Walker, Amelie Alterauge, Niamh Carty, Craig Cessford, Hermann Fetz, Michaël Gourvennec, Robert Hartle, Michael Henderson, Kristin von Heyking, Sarah A. Inskip, Sacha Kacki, Felix M. Key, Elizabeth L. Knox, Christian Later, Prishita Maheshwari-Aplin, Joris Peters, John E. Robb, Jürgen Schreiber, Toomas Kivisild, Dominique Castex, Sandra Lösch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos & Johannes Krause

*contributed equally

Published in *Nature Communications* 10, 4470 (2019) doi:10.1038/s41467-019-12154-0

Summary: This paper presents 34 ancient *Y. pestis* genomes of the Second Pandemic, spanning a broad geographic area from the 14th–17th centuries. We show evidence for the introduction of the Black Death via East Europe, low genetic diversity of *Y. pestis* during the Black Death and local epidemics as well as persistence and diversification of plague lineages during the Second Pandemic. Furthermore, we identify an accelerated substitution rate in one lineage, coinciding with a genomic deletion similar to the one described in Manuscript A.

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 the 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 the 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 contributions from all co-authors.

In total, Marcel Keller contributed 40 % to the laboratory work, data analysis and manuscript writing.

Manuscript C: Ancient *Yersinia pestis* genomes provide no evidence for the origins or spread of the Justinianic Plague

Authors: **Marcel Keller**, Maria A. Spyrou, Michael McCormick, Kirsten I. Bos, Alexander Herbig, Johannes Krause

Submitted to *bioRxiv* on 31st of October 2019, revised on 12th of November 2019, doi:10.1101/819698

Summary: In their paper "137 ancient human genomes from across the Eurasian steppes", Damgaard et al., 2018 published two *Y. pestis* genomes from the Tian Shan region and the Caucasus, from which they draw conclusions about the spread of the Justinianic Plague from Central Asia to Europe in the context of the Hunnic migrations. In this re-analysis of the published data, we show that the age and phylogenetic position of the first genome does not substantiate their claims, and the second low-coverage genome even falls in a completely different phylogenetic position, potentially holding clues about the onset of the Black Death.

Author contributions:
M.K., A.H., K.I.B and J.K. planned and designed the study. **M.K.** performed data processing and phylogenetic analyses; M.A.S. and **M.K.** performed dating analyses. M.M. provided and reviewed historical context information. **M.K.** wrote the manuscript with contributions from M.A.S., K.I.B. and A.H. and edits from all co-authors.

In total, Marcel Keller contributed 80 % to the data analysis and manuscript writing.

Manuscript D: Von der Seuchengeschichte der Pest zu einer Naturgeschichte ihres Erregers – Neue Einblicke durch alte DNA

Marcel Keller

Published in *Pest! Eine Spurensuche.* Edited by LWL-Museum für Archäologie, Westfälisches Landesmuseum, Herne; Stefan Leenen; Alexander Berner; Sandra Maus. wbg Theiss (Darmstadt, 2019)

Summary: This book chapter is a contribution of the exhibition catalogue accompanying the special exhibition *Pest!* of the LWL-Museum für Archäologie Herne, 20th of September 2019 – 10th of May 2020. With its diachronic approach – following plague from its emergence in the Neolithic period until today – the exhibition explicitly integrates the results of recent archaeogenetic studies. The essay is conceived as both an introduction for lay readers and a literature review on the evolutionary history of plague as revealed through ancient DNA. With a historical approach, the archaeogenetic studies are located in a broader scientific and scholarly framework, discussing their contributions and limitations in answering key questions on the history of plague.

Marcel Keller conceived the outline, researched literature and wrote the manuscript to 100 %.

4 Manuscript A: Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750)

Marcel Keller*, Maria A. Spyrou*, Christiana L. Scheib, Gunnar U. Neumann, Andreas Kröpelin, Brigitte Haas-Gebhard, Bernd Päffgen, Jochen Haberstroh, Albert Ribera i Lacomba, Claude Raynaud, Craig Cessford, Raphaël Durand, Peter Stadler, Kathrin Nägele, Jessica S. Bates, Bernd Trautmann, Sarah A. Inskip, Joris Peters, John E. Robb, Toomas Kivisild, Dominique Castex, Michael McCormick, Kirsten I. Bos, Michaela Harbeck, Alexander Herbig & Johannes Krause
*contributed equally

Published in *PNAS* 116 (25), 12363–12372 (2019), doi:10.1073/pnas.1820447116

This page is intentionally left blank.

Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750)

Marcel Keller[a,b,1,2,3], Maria A. Spyrou[a,1], Christiana L. Scheib[c,d], Gunnar U. Neumann[a], Andreas Kröpelin[a,e], Brigitte Haas-Gebhard[f], Bernd Päffgen[g], Jochen Haberstroh[h], Albert Ribera i Lacomba[i], Claude Raynaud[j], Craig Cessford[c], Raphaël Durand[k], Peter Stadler[l], Kathrin Nägele[a], Jessica S. Bates[c], Bernd Trautmann[b], Sarah A. Inskip[m], Joris Peters[b,n,o], John E. Robb[c], Toomas Kivisild[c,p], Dominique Castex[q], Michael McCormick[r,s], Kirsten I. Bos[a], Michaela Harbeck[b,2], Alexander Herbig[a,2], and Johannes Krause[a,s,2]

[a]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; [b]State Collection of Anthropology and Palaeoanatomy Munich, Staatliche Naturwissenschaftliche Sammlungen Bayerns, 80333 Munich, Germany; [c]Department of Archaeology, University of Cambridge, Cambridge CB2 3ER, United Kingdom; [d]Institute of Genomics, University of Tartu, 51010 Tartu, Estonia; [e]Friedrich Schiller University Jena, 07743 Jena, Germany; [f]Archaeological Collection of the Bavarian State, 80538 Munich, Germany; [g]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilian University Munich, 80799 Munich, Germany; [h]Bavarian State Department of Monuments and Sites, 80539 Munich, Germany; [i]Department for Municipal Archaeology, Valencia City Council, 46014 Valencia, Spain; [j]CNRS, UMR5140, Archéologie des Sociétés Méditerranéennes, 34199 Montpellier, France; [k]Service d'Archéologie Préventive de l'Agglomération de Bourges Plus, 18023 Bourges Cedex, France; [l]Department of Pre- and Protohistory, University of Vienna, 1190 Vienna, Austria; [m]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge CB2 3ER, United Kingdom; [n]ArchaeoBioCenter, Ludwig Maximilian University Munich, 80539 Munich, Germany; [o]Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig Maximilian University Munich, 80539 Munich, Germany; [p]Department of Human Genetics, Katholieke Universiteit Leuven, 3000 Leuven, Belgium; [q]UMR 5199, PACEA, CNRS Institute, 33615 Pessac Cedex, France; [r]Initiative for the Science of the Human Past, Department of History, Harvard University, Cambridge, MA 02138; and [s]Max Planck–Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, 07745 Jena, Germany

Edited by Nils Chr. Stenseth, University of Oslo, Oslo, Norway, and approved May 9, 2019 (received for review November 30, 2018)

The first historically documented pandemic caused by *Yersinia pestis* began as the Justinianic Plague in 541 within the Roman Empire and continued as the so-called First Pandemic until 750. Although paleogenomic studies have previously identified the causative agent as *Y. pestis*, little is known about the bacterium's spread, diversity, and genetic history over the course of the pandemic. To elucidate the microevolution of the bacterium during this time period, we screened human remains from 21 sites in Austria, Britain, Germany, France, and Spain for *Y. pestis* DNA and reconstructed eight genomes. We present a methodological approach assessing single-nucleotide polymorphisms (SNPs) in ancient bacterial genomes, facilitating qualitative analyses of low coverage genomes from a metagenomic background. Phylogenetic analysis on the eight reconstructed genomes reveals the existence of previously undocumented *Y. pestis* diversity during the sixth to eighth centuries, and provides evidence for the presence of multiple distinct *Y. pestis* strains in Europe. We offer genetic evidence for the presence of the Justinianic Plague in the British Isles, previously only hypothesized from ambiguous documentary accounts, as well as the parallel occurrence of multiple derived strains in central and southern France, Spain, and southern Germany. Four of the reported strains form a polytomy similar to others seen across the *Y. pestis* phylogeny, associated with the Second and Third Pandemics. We identified a deletion of a 45-kb genomic region in the most recent First Pandemic strains affecting two virulence factors, intriguingly overlapping with a deletion found in 17th- to 18th-century genomes of the Second Pandemic.

Justinianic Plague | ancient DNA | bacterial evolution | Anglo-Saxons | Merovingians

Yersinia pestis, the causative agent of plague, is a Gram-negative bacterium that predominantly infects rodents and is transmitted by their ectoparasites such as fleas. As a zoonosis, it is also able to infect humans with a mortality rate of 50–100% without antibiotic treatment (1), manifesting as bubonic, septicemic, or pneumonic plague. In addition to the ancient foci that exist in Central and East Asia, the pathogen spread worldwide at the end of the 19th century in the so-called Third Pandemic that started in 1855 in Yunnan, China, establishing new local foci in Africa and the Americas. Today, *Y. pestis* causes sporadic infections annually and occasional local recurrent epidemics such as that documented in 2017 in Madagascar (2).

Although recent paleogenetic analyses have reconstructed an ancient form of *Y. pestis* that infected humans as early as in the prehistoric period [2,900–1,700 BCE (3–6)], the First Pandemic

Significance

The first historically reported pandemic attributed to *Yersinia pestis* started with the Justinianic Plague (541–544) and continued for around 200 y as the so-called First Pandemic. To date, only one *Y. pestis* strain from this pandemic has been reconstructed using ancient DNA. In this study, we present eight genomes from Britain, France, Germany, and Spain, demonstrating the geographic range of plague during the First Pandemic and showing microdiversity in the Early Medieval Period. Moreover, we detect similar genome decay during the First and Second Pandemics (14th to 18th century) that includes the same two virulence factors, thus providing an example of potential convergent evolution of *Y. pestis* during large-scale epidemics.

Author contributions: M.M., K.I.B., M.H., A.H., and J.K. designed the study; M.K., M.A.S., C.L.S., G.U.N., K.N., and J.S.B. performed laboratory work; A.K. developed the new analytical tool; M.K., M.A.S., and G.U.N. performed data analyses; B.T. and S.A.I. performed anthropological examination; C.L.S. and T.K. provided genomic data; B.H.-G., B.P., J.H., A.R.i.L., C.R., P.S., J.P., J.E.R., D.C., and M.H. identified and provided access to archaeological material; B.H.-G., B.P., J.H., A.R.i.L., C.R., C.C., R.D., P.S., and M.M. provided archaeological and historical information; and M.K., M.A.S., M.M., and A.H. wrote the paper with contributions from all authors.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The raw sequencing data of plague-positive samples have been deposited in the European Nucleotide Archive (accession no. PRJEB29991).

[1]M.K. and M.A.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: marcel.keller@ut.ee, harbeck@snsb.de, herbig@shh.mpg.de, or krause@shh.mpg.de.

[3]Present address: Institute of Genomics, University of Tartu, 51010 Tartu, Estonia.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820447116/-/DCSupplemental.

Published online June 4, 2019.

(541–750) is the earliest historically recorded pandemic clearly attributed to *Y. pestis* (7, 8), starting with the fulminant Justinianic Plague (541–544). It was later followed by the Second Pandemic, which started with the Black Death of 1346–1353 (9, 10) and persisted in Europe until the 18th century (11–13).

The 2000s saw first attempts to amplify *Y. pestis*-specific DNA fragments from burials of the sixth century (14–16). Although early studies on two French sites (15, 16) are controversial due to methodological limitations (17) and proved inconsistent with a later genotyping study (18), more recent studies have been successful in authenticating the latter and reconstructing whole *Y. pestis* genomes from two early medieval burial sites in modern-day Bavaria, Germany (7, 8).

These genomic investigations identified a previously unknown lineage associated with the First Pandemic that was found to be genetically identical at both sites and falls within the modern diversity of *Y. pestis*. Moreover, this lineage is distinct from those associated with the Second Pandemic that started ∼800 y later, indicating two independent emergence events.

Although these studies have unequivocally demonstrated the involvement of *Y. pestis* in the First Pandemic, the published genomes represent a single outbreak, leaving the genetic diversity of that time entirely unexplored. Here, we assess the diversity and microevolution of *Y. pestis* during that time by analyzing multiple and mass burials from a broader temporal and spatial scope than previously attempted. After screening 183 samples from 21 archaeological sites, we were able to reconstruct eight genomes with higher than 4.5-fold mean coverage from Britain, France, Germany, and Spain. Furthermore, we identified a large deletion in the most recent First Pandemic strains that affects the same region as a deletion observed in late Second Pandemic strains, suggesting similar mechanism of pathogen adaptation in the waning period of the two separate pandemics.

Results

Screening and Capture. We used a previously described qPCR assay (19) that targets the *Y. pestis*-specific *pla* gene on the pPCP1 plasmid to test 171 teeth from a minimum of 122 individuals from 20 sites, spanning from ∼300 to 900 CE (*SI Appendix*, Table S1). For the remaining site, Edix Hill, Britain, the 22 samples only had shotgun sequencing data available, and therefore, pathogen DNA screening was performed using the metagenomic tool MALT (20). This analysis revealed six putatively *Y. pestis*-positive samples after visual inspection of aligned reads in MEGAN (21) (*SI Appendix*, Table S4). All 30 PCR-positive extracts and 5 of the Edix Hill samples were subsequently turned into double-stranded, double-indexed, and UDG-treated DNA libraries and were enriched for the *Y. pestis* genome following an in-solution capture approach (20).

Whereas some samples reached up to 38.1-fold chromosomal mean coverage after whole-genome capture, nine of the PCR-positive samples yielded a coverage of lower than 0.1-fold. Since the qPCR assay can amplify nonspecific products and subsequent capture can enrich for environmental DNA that sporadically maps to the *Y. pestis* reference, it is crucial to differentiate between samples that show low DNA preservation and those that are false positives.

False-positive samples are unlikely to show similar mapping success on all genetic elements compared with true-positive samples. Therefore, mapping to all three plasmids was used in combination with a statistical outlier detection for the verification of low coverage genomes. Ratios of reads mapping to the *Y. pestis* chromosome and the three individual plasmids were determined to normalize for the variable coverage between samples. Since the samples were captured with the same probe set and assuming no vast differences in plasmid copy number, the ratios should be consistent over all positive samples independent of their genomic coverage. This is, however, not expected for false-positive samples. Therefore, we calculated the Mahalanobis distance (22), a standard method for outlier detection in multivariate datasets, to find false-positive and authenticate low coverage true-positive samples ($\chi^2 = 5.991$, df = 2, $P = 0.05$; *SI Appendix*, Table S2). Five samples, EDI002.A, DIR002.A, LVC001.B, LVC001.C, and PEI001.A, were classified as outliers. Despite having chromosomal coverage, DIR002.A, EDI002.A, and PEI001.A had no or only a few reads mapping to the plasmids and were therefore considered as *Y. pestis* negative. LVC001.B and LVC001.C had an exceptionally high ratio of reads mapping to pPCP1 and are still considered positive. The remaining 33 samples come from four sites in Germany (Dittenheim [DIT], $n = 3$; Petting [PET], $n = 3$; Waging [WAG], $n = 1$; Unterthürheim [UNT], $n = 5$), two in France (Lunel-Viel [LVC], $n = 6$; Saint-Doulchard [LSD], $n = 11$), one in Britain (Edix Hill [EDI], $n = 4$), and one in Spain (Valencia [VAL], $n = 1$; Table 1 and Fig. 1).

After mapping to the chromosome, 10 genomes showed a higher than 4.5-fold mean coverage and were used for downstream analyses. These were DIT003.B (9.4-fold), EDI001.A (38.1-fold), EDI003.A (5.2-fold), EDI004.A (7.5-fold), LSD001.A (4.8-fold), LSD023.A (7.2-fold), PET004.A (5.6-fold), VAL001.B (9.6-fold), as well as UNT003.A and UNT004.A (7.6-fold and 5.2-fold, respectively) (*SI Appendix*, Table S3). The raw reads of six positive samples of the individuals LVC001, LVC005, and LVC006 were combined to yield a single genome with a mean coverage of 6.7-fold for the site of Lunel-Viel after assuring that they represent an identical strain. From each site, only the genome with the highest coverage was used for phylogenetic analyses when multiple genomes were available but shown to be identical in the evaluation of their single-nucleotide polymorphism (SNP) profiles. As such, the genomes of EDI003.A, EDI004.A, and UNT004.A were omitted.

SNP Evaluation. Phylogenetic analyses based on SNP alignments are prone to wrong topologies and artificial branches introduced by false-positive SNPs. This is especially true for low-coverage genomes that derive from metagenomic specimens. In the context of ancient pathogen DNA, there are three main sources for false-positive SNPs: First, DNA damage such as deamination of cytosine to uracil can lead to misincorporation of nucleotides during sample processing (23). Second, the mapping of closely related environmental species to the reference sequence of the target organism is likely, especially for conserved regions of the genome (24). Third, mapping of short reads is more prone to mismapping and calling of false-positive SNPs generated at sites of genome rearrangement. Whereas the first source can be circumvented via in vitro protocols like UDG treatment (25), the latter two can be reduced but not eliminated with strict mapping parameters and exclusion of problematic regions (26) as applied here. A fourth source for false SNP assignments could result from multiple genetically distinct strains that would lead to a chimeric sequence. The latter was not observed in our data (*SI Appendix*, Fig. S1) and this phenomenon might be limited to chronic infections with pathogens such as *Mycobacterium tuberculosis*, where mixed infections have been previously documented (27). The introduction of false-positive SNPs by sequencing errors is stochastically negligible as shown in simulated datasets (*SI Appendix*).

The retrieval of genomes that span a wide geographic area gives us the opportunity to assess *Y. pestis* microdiversity present in Europe during the First Pandemic. Given that our genomes are of relatively low genomic coverage, we critically evaluated uniquely called and shared SNPs among the First Pandemic genomes to accurately determine their phylogenetic position. This analysis was performed for all genomes retrieved from UDG-treated libraries with higher than 4.5-fold mean coverage, including the previously published high-quality Altenerding genome (17.2-fold mean coverage).

For this, we developed the tool "SNPEvaluation" and defined three different criteria, all applying for a 50-bp window surrounding the SNP: (A) Comparing the mean coverage after BWA mapping with high and low stringency and excluding all SNPs that showed a higher coverage under low stringent mapping than in high stringent mapping. In metagenomic datasets, reads of related species map

Table 1. List of all sites that were tested with country in brackets (AUS = Austria, DEU = Germany, ESP = Spain, FRA = France, GBR = Great Britain)

Site	Lab ID	Context	Graves in total	Multiple burials	Time frame	Positive/total samples
Alladorf (DEU)	ALL	Separate burial area (*Hofgrablege*)	163	5×2	630–720	0/6
Dirlewang (DEU)	DIR	Early medieval cemetery	40	2×2	650–700	0/2
Dittenheim (DEU)	DIT	Early medieval cemetery	238, 10 crem.	4×2	550–700	3/9
Edix Hill (GBR)	EDI	Early medieval cemetery	115	1×4, 9×2	500–650	4/22
Forchheim (DEU)	FOR	Special burial	1	1×4	650–700	0/3
Grafendobrach (DEU)	GRA	Settlement burials (*Hofgrablege*)	85	1×3, 1×2+1	850–930	0/3
Kleinlangheim (DEU)	KLH	Early medieval cemetery	244, 56 crem.	8×2, 1×3	470–720	0/5
Leobersdorf (AUS)	LEO	Early medieval cemetery	154	16×2, 4×3, 2×4, 1×5	640–800	0/3
Lunel-Viel Les Horts (FRA)	LVH	Early medieval cemetery	140	1×2	475–700	0/5
Lunel-Viel Quartier centrale (FRA)	LVC	Demolition trench inhumations	—	6+2 individuals in 2 trenches	400–600	6/16
München-Aubing (DEU)	AUB	Early medieval cemetery	896	4×2	400–700	0/8
Neuburg an der Donau (DEU)	NEU	Late Roman cemetery	130	3×2, 1×3	300–400	0/2
Peigen (DEU)	PEI	Early medieval cemetery	274	3×2	450–700	0/5
Petting (DEU)	PET	Early medieval cemetery	721	min. 1×3, 2×2, 1×2+1	530–730	3/7
Regensburg Fritz-Fend-Str. (DEU)	RFF	Late Roman cemetery	115, 48 crem.	2×2	350–450	0/3
Saint-Doulchard Le Pressoir (FRA)	LSD	Early medieval cemetery	175	12×2, 2×3	530–1200	11/26
Sindelsdorf (DEU)	SIN	Early medieval cemetery	331	3×2, 1×3+1	500–720	0/5
Straubing Azlburg I/II (DEU)	SAZ	Late Roman cemetery	541, 1 crem.	2×2, 1×3	300–450	0/3
Unterthürheim (DEU)	UNT	Early medieval cemetery	256	14×2, 2×3, 1×4	525–680	5/7
Valencia, Plaça de Almoina (ESP)	VAL	Visigothic intramural cemetery	67	3×2, 3×3, 4×4, 2×5, 15×5+	500–700	1/36
Waging (DEU)	WAG	Early medieval cemetery	239	min. 2×2, 1×2+1	530–700	1/12
Westheim (DEU)	WES	Early medieval cemetery	228	5×2, 1×3	500–650	0/3

The number of graves is counting multiple burials as single graves; cremations are counted separately. Multiple burials are listed as number of graves times number of individuals (5x2 translates to five double burials, and 1x2+1 to one double burial associated with a single burial). Detailed site descriptions are given in *SI Appendix*, and a table of all screened samples in *SI Appendix*, Table S1.

frequently to conserved regions in the reference genome. When the position is not covered by reads from the target organism (*Y. pestis*) but the genomic region is similar enough in other environmental organisms so that their reads can map, they might mimic a SNP in *Y. pestis* when the contaminant species carries a different allele in that position. (B) Excluding all SNPs for which heterozygous calls were identified in the surrounding regions. Heterozygous calls accumulate in conserved regions due to the above-described effect. (C) Excluding all SNPs within regions that include positions that lack genomic coverage. Variants in genome architecture often appear as gaps in mapped data and are likely to cause mapping errors, potentially resulting in false-positive SNPs.

The performance of the presented tool and the validity of the selected criteria were assessed using simulated datasets, where each dataset consisted of both reads of a known *Y. pestis* genotype (CO92) with different coverages and reads from a captured sample previously determined as *Y. pestis*-negative (DIR002.A) for representation of an environmental background that is typical for ancient DNA datasets. The SNP evaluation on the artificial datasets—applied with the same criteria as for the First Pandemic genomes presented in this study—showed a maximum sensitivity (all false-positive CO92 SNPs detected) and a high specificity (3.49–8.57% of true-positive CO92 positions erroneously filtered out; *SI Appendix*, Figs. S2 and S3 and Tables S5, S6, and S7).

This evaluation was applied to all nonshared SNPs within the First Pandemic lineage, totaling between 1 and 87 chromosomal SNPs per genome (*SI Appendix*, Table S8). Forty-four chromosomal SNPs, three SNPs on the pCD1 plasmids, and two on the pMT1 plasmid were classified as true positive across all 11 genomes (*SI Appendix*, Table S10). The following 39 chromosomal SNPs appear unambiguous and phylogenetically informative (see "Tree" in *SI Appendix*, Table S10 and Fig. 2*B*): The Edix Hill genomes (EDI001.A, EDI003.A, and EDI004.A) share one unique SNP, and they are missing one SNP previously identified in Altenerding and shared in all other genomes. The Altenerding genome (AE1175) as well as the genomes of Unterthürheim (UNT003.A, UNT004.A) and Dittenheim (DIT003.B) appear identical after SNP evaluation at all positions. The genomes from Petting (PET004.A) and Valencia (VAL001.B) appear distinct, each occupying a unique branch composed of two and three unique SNPs, respectively (Fig. 2*B* and *SI Appendix*, Table S10). The genomes of Lunel-Viel (LVC_merged) and Saint-Doulchard (LSD001.A, LSD023.A) share in total 12 SNPs and Lunel-Viel has one more unique derived SNP. The two genomes of Saint-Doulchard share two additional SNPs but are separated by branches of nine (LSD001.A) and seven SNPs (LSD023.A). Three SNPs, two of which were previously called in Altenerding (*SI Appendix*, Table S5), were identified as potentially shared among all First Pandemic genomes and therefore classified as uninformative for the microdiversity. One SNP might be shared among all genomes except Edix Hill, but cannot be reliably classified due to low coverage in multiple samples. Another SNP appears to be homoplastic since it appears in the Altenerding cluster and Saint-Doulchard, but not in Lunel-Viel. Whereas the low-coverage genomes of LSD007.A, LSD019.A, LSD020.A, LSD021.A, LSD022.A, and LSD024.A appear identical with LSD023.A by full or partial coverage of its unique SNPs, the genomes of LSD002.A, LSD013.A, and LSD026.A cannot be assigned to the genotype of LSD023.A or LSD001.A, since none of their unique SNPs is covered in these genomes.

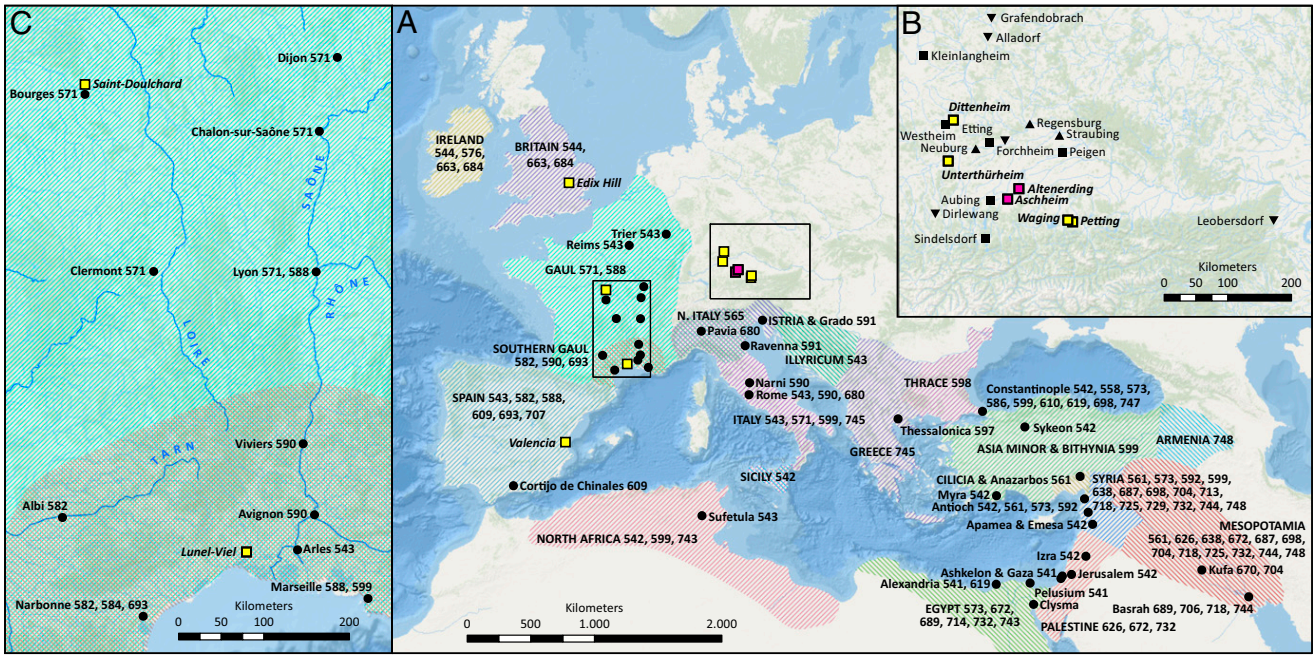


Fig. 1. Geographic extent of the First Pandemic and sampled sites. (*A*) Map of historically documented occurrences of plague (regions shaded, cities depicted by circles, both with respective years of occurrence) between 541 and 750 in Europe and the Mediterranean basin. All sources are given in *SI Appendix*. Sites with genomic evidence for *Y. pestis* are shown as pink (previously published) and yellow squares (presented here). (*B*) Enlarged rectangular space of *A* (*Right*) showing all sites in Germany and Austria that were included in this study. Sites tested negative are depicted in black upward-pointing triangles (burials dating before 541), squares (dating around 541–544), and downward-pointing triangles (dating after 544). (*C*) Enlarged *Inset* of *A* (*Left*) shows reported occurrences in France and main rivers.

On the pCD1 plasmid, one SNP was identified as missing in the Edix Hill genomes (EDI001.A, EDI003.A, EDI004.A), one as shared between both Saint-Doulchard genomes (LSD001.A, LSD023.A), and one as unique to the genome LSD001.A. One additional SNP was found on the pMT1 plasmid in the Valencia genome (VAL001.B). An analysis of the Aschheim genome as well as a SNP effect analysis is presented in *SI Appendix*, Tables S7 and S14.

SNPs shared by at least two genomes without a conflicting call in any other genome were evaluated as potentially shared SNPs among the First Pandemic lineage. We applied the exact same parameters as for the nonshared SNPs, but also considered positions with less than threefold coverage (*SI Appendix*, Table S11). Only SNPs that pass all three criteria of our SNP evaluation in at least half of the analyzed genomes (i.e., 6 out of 12) were accepted as true shared SNPs, reducing the number from 50 SNPs identified in a previous study (7)—after removal of nonshared and ambiguous SNPs—to 45.

The Waging sample (WAG001.A) had a genomic coverage too low for inclusion in our phylogenetic analysis. Since it was the only sample giving evidence for *Y. pestis* presence at this site, it was assessed for all SNPs that were either shared or unique in the other First Pandemic genomes. Visual inspection revealed 7 of the 43 shared SNPs to be present in the WAG001.A genome at low coverage (less than threefold) and one SNP absent in Edix Hill but potentially present in all other genomes. For both shared and unique SNPs, no conflicting positions were found. This strain could, therefore, be attributed to the First Pandemic lineage without, however, resolving its exact phylogenetic position (*SI Appendix*, Table S11).

Phylogenetic Analysis. A set of 233 modern *Y. pestis* genomes (*SI Appendix*, Table S12) as well as 7 Second Pandemic genomes, including a representative of the Black Death strain (London) and 7 post-Black Death genomes [14th-century Bolgar, 16th-century Ellwangen (12); 18th-century Marseille (13)], and an ancient genome from Tian Shan [DA101, second to third century (28)] were used for phylogenetic analyses alongside our First Pandemic genomes presented here (*SI Appendix*, Table S3) and the previously published genome of Altenerding. The *Y. pseudotuberculosis* isolate IP32953 (29) was used as an outgroup.

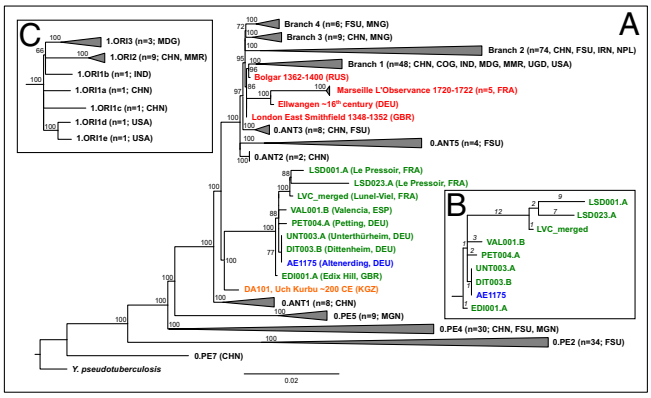


Fig. 2. Phylogenetic tree. (*A*) Maximum-likelihood tree with full SNP alignment (6,580 positions) of 233 modern *Y. pestis* and one *Y. pseudotuberculosis* genome, 10 published (second- to third-century Tian Shan in orange; Altenerding in blue; Second Pandemic in red) and eight genomes presented here (green) with country given in brackets (DEU = Germany, ESP = Spain, FRA = France, GBR = Great Britain, RUS = Russia). Numbers and origins of modern genomes are given in brackets (CHN = China, COG = Congo, FSU = Former Soviet Union, IND = India, IRN = Iran, MDG = Madagascar, MMR = Myanmar, MNG = Mongolia, NPL = Nepal, UGA = Uganda). Numbers on nodes are showing bootstrap values (1,000 iterations). (*B*) Detailed, manually drawn tree of the First Pandemic genomes showing all remaining SNP positions after SNP evaluation (number of SNPs given in italics). (*C*) Detailed tree of the 1.ORI clade within branch 1, showing the polytomy.

Our maximum-likelihood tree (30) constructed from the full SNP alignment reveals that all of the genomes presented here occupy positions on the same lineage (Fig. 2*A* and *SI Appendix*, Fig. S4). This confirms their authenticity and is congruent with previous association of this lineage to the First Pandemic (541–750). In addition, the previously reported genome from Altenerding (2148) is identical to the genomes from Dittenheim (DIT003.B) and Unterthürheim (UNT003.A) presented here. Moreover, the genomes of Petting (PET004.A), Valencia (VAL001.B), and the clade giving rise to the French genomes of Lunel-Viel (LVC_merged) and Saint-Doulchard (LSD001.A, LSD023.A) seem to diverge from the Altenerding cluster through a polytomy (Fig. 2*B*; 88% bootstrap support). The French clade further diversifies into two branches, one giving rise to Lunel-Viel (LVC_merged, 100% bootstrap support), and a second one splitting into the two genomes from Saint-Doulchard (LSD001.A, LSD023.A; 88% bootstrap support). The British genome of EDI001, however, branches off one SNP ancestral to this polytomy (100% bootstrap support) and possesses one unique SNP. This is remarkable, since the British Isles are one of the most remote places where the First Pandemic was suspected of reaching in relation to its presumed starting point in Egypt.

Virulence Factor and Deletion Analysis. We screened for the presence/absence of 80 chromosomal and 42 plasmid-associated virulence genes (31, 32) in all First Pandemic genomes with higher than 4.5-fold coverage (Fig. 3 and *SI Appendix*, Fig. S5). Only the filamentous prophage was consistently found absent in all presented genomes. This is expected, since it has integrated into the genome of only a number of modern branch 1 genomes (33). Reduced coverages for a set of virulence factors can be seen in the Altenerding (AE1175), Bolgar, and Ellwangen genomes due to a capture bias, since the capture probe set in the respective studies was designed on the basis of *Y. pseudotuberculosis* rather than of *Y. pestis* (7, 12).

Intriguingly, the most derived First Pandemic genomes from Lunel-Viel (LVC_merged) and Saint-Doulchard (LSD001.A, LSD023.A) show a deletion of two chromosomal virulence-associated genes, *mgtB* and *mgtC* (Fig. 3). These magnesium transporters are part of the PhoPQ regulon, which is important for survival of *Y. pestis* in the magnesium-deficient environment of macrophages. However, functional studies on *mgtB* hint at an important role during macrophage invasion rather than intracellular survival (34).

A second deletion was observed for the gene YPO2258, categorized as a potential virulence factor based on the presence of a frameshift mutation in the avirulent 0.PE2_Microtus91001 strain (32). Its inactivation in the 2.ANT1_Nepal516 strain, isolated from a human patient, nevertheless indicates that this gene is not essential for virulence in humans (35).

Further exploration of the deletion of the two neighboring genes *mgtB* and *mgtC* revealed that they are part of a ∼45-kb deletion (positions 1,883,402–1,928,869 in the CO92 reference), affecting 34 genes including multiple motility (*motA*, *motB*) and chemotaxis genes (*cheA*, *cheB*, *cheD*, *cheR*, *cheW*, *cheY*, *cheZ*) (*SI Appendix*, Fig. S6). On the downstream end, the deletion is flanked by an IS100 insertion element. A potential upstream insertion element might be undetectable at our current resolution due to a genome rearrangement in the reference genome CO92. This is in agreement with previous findings concerning the highly abundant IS100 element in *Y. pestis*, responsible not only for disruptions of multiple genes caused by homologous recombination (29), but also for the loss of the 102-kb-long *pgm* locus containing a high-pathogenicity island in several strains (36). To address the specificity of this deletion to the sixth- to eighth-century strains from France, we also investigated the presence of the two virulence factors in all other modern and ancient strains in this study. Intriguingly, a similar deletion affecting the same region including *mgtB* and *mgtC* was observed in the late Second Pandemic genomes from London New Churchyard [1560–1635 (37)] and Marseille L'Observance [1720–1722 (13)]. However, a full deletion of this 45-kb region was not found in any of the other ancient or modern genomes. Therefore, the deletion appeared independently in the later course of both the First and Second Pandemics. A second but smaller chromosomal deletion and a homoplastic deletion on the pMT1 plasmid are presented in *SI Appendix*, Fig. S7.

Discussion

Identifying *Y. pestis* DNA in Low-Complexity Specimens. In total, we screened 171 samples from 20 sites in France, Germany, and Spain for *Y. pestis* with a qPCR assay (19) and 22 additional samples from Edix Hill, Britain, with the metagenomic tool MALT (20). All putatively positive samples were turned into UDG libraries and subsequently enriched for *Y. pestis*, resulting in mean coverages ranging from 0.01- to 38.1-fold.

The validation of genomic data with relatively low amounts of mapping reads as presented here is challenging; DNA extracted from archaeological remains results in metagenomic data, and differentiating between target organism DNA and environmental background can be difficult.

The identification of *Y. pestis* DNA based on PCR targeting the *pla* locus on the pPCP1 plasmid has theoretically been shown to be problematic (38), leading to discussions about false-positive results (17). However, assignment to *Y. pestis* based on reads retrieved from shotgun sequencing and mapping to a reference genome also can be challenging in case of extremely low genomic coverage (3, 4). Since all of the presented genomes are derived from DNA libraries specifically enriched for *Y. pestis* DNA and are thus biased toward the target organism, a previously suggested competitive mapping approach (3) would not be suitable. Instead, we considered the relative number of mapping reads to the plasmids and chromosome to identify false-positive samples from captured data. We were able to verify that 30 out of 33 samples were positive for *Y. pestis* with as few as 2,000 reads mapping to the chromosome. Since the three plasmids pCD1,

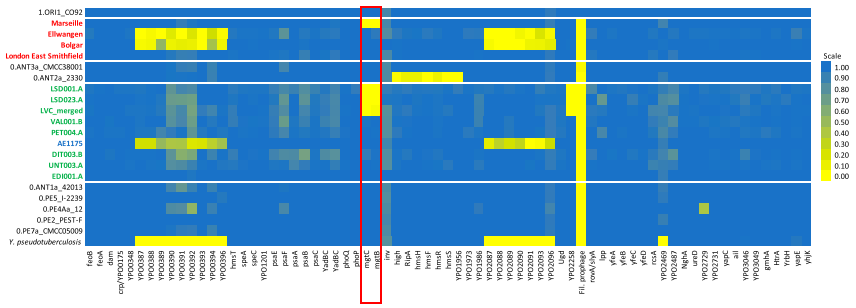


Fig. 3. Heatmap showing the percentage of coverage of chromosomal virulence factors. First Pandemic genomes (blue and green) and Second Pandemic genomes (red) are shown in combination with selected strains of main clades of modern *Y. pestis* diversity on branch 0 as well as the reference genomes of *Y. pseudotuberculosis* and *Y. pestis* (CO92).

pMT1, and pPCP1 were already present in the early divergent Neolithic and Bronze Age strains (3, 4) and loss of plasmids has only been observed sporadically in attenuated strains (39), this method could be reliably applied to data stemming from other branches in the *Y. pestis* phylogeny.

Analyzing Microdiversity with Low-Coverage Genomes. Reliable SNP calling is crucial for the phylogenetic analysis of verified low-coverage genomes and can be challenging when dealing with ancient pathogen DNA stemming from metagenomic contexts. This has been demonstrated on *Y. pestis* genomes (7), but previously applied visual inspections are time-consuming and not easily reproducible.

Here, we present an approach for SNP authentication using a semiautomated SNP evaluation. We selected three criteria for our evaluation to assess the likelihood of mismapping. We excluded all SNPs that (A) had higher coverage when mapped with less strict parameters, (B) had "heterozygous" positions in close proximity, or (C) were flanked by gaps. With these filters, we tolerate a loss of specificity (3.59–8.57% true positive erroneously filtered) to reach a maximum sensitivity (100% false positives filtered), as shown with simulated data. Our method is therefore tailored for the reliable characterization of microdiversity. Moreover, the tool "SNPEvaluation" that was newly developed for this analysis offers a highly flexible framework for the assessment of VCF files and can be utilized also for a variety of analyses on different organisms.

Phylogenetic Analysis. We were able to confidently reconstruct eight genomes associated with the First Pandemic from Britain, France, Germany, and Spain, providing insights into the microdiversity and persistence of *Y. pestis* in Europe between the sixth and eighth centuries.

Our presented genomes add diversity to a phylogenetic lineage that was previously shown to contain two identical sixth-century genomes from southern Germany [Aschheim and Altenerding (7, 8)]. It diverges between the 0.ANT1, 0.ANT2, and 0.ANT5 clades in the main *Y. pestis* phylogeny and shares a short branch with a second- to third-century genome from the Tian Shan mountains (28). Intriguingly, a single diversification event gave rise to the published as well as three of the presented additional branches, two composed of single genomes with two to three derived SNPs and one branch diversifying into three distinct strains. Similar polytomies can be detected in other parts of the phylogeny of *Y. pestis* that have been related to human epidemics (40): one gave rise to branches 1–4 (including ancient Second Pandemic genomes, Fig. 2*A*) and is dated to 1142–1339 (40), shortly before the European Black Death. To date, it is unknown whether this event was restricted to a rodent reservoir, or if it was already associated with a human epidemic. A second polytomy gave rise to the 1.ORI clade, which includes strains related to the worldwide spread of plague during the Third Pandemic in the 19th century (Fig. 2*C*).

Within the First Pandemic lineage, the genomes that derive from this polytomy display variable terminal branch lengths (1–23 SNPs), which are likely concurrent with their different ages (see below). Given that *Y. pestis* is a pathogen that can cover large geographic distances without accumulating genetic diversity (12), it is challenging to elucidate the geographic origin for this diversification event. A first hypothesis suggests an origin of this diversification event within the historically recorded geographic range of the First Pandemic, i.e., either in Europe, the Mediterranean basin, or the Middle East. Our current data may lend some credibility to this scenario for two reasons: First, we identify four European strains with short genetic distances from this polytomy, the shortest of which is identified in three locations in rural Bavaria, and second, we identify an almost direct ancestor of this polytomy to be present in Europe during the sixth century, represented by a genome from Britain. Alternatively, the bacterium may have been recurrently introduced to the affected regions from a single remote reservoir.

The hypothesis of a single introduction would require the establishment of a local reservoir, since the genomes recovered from Lunel-Viel and Saint-Doulchard are clearly not associated with the initial outbreak in 541–544 but rather with subsequent ones (see below). The establishment of a local reservoir is further substantiated by two diversification events in the French clade, one giving rise to the genome of Lunel-Viel with only one unique SNP and a second event only two SNPs derived, giving rise to both Saint-Doulchard genomes. Possible locations for reservoirs during the First Pandemic have been suggested in the Iberian Peninsula and the Levant (41). There is also a growing body of evidence for the presence of black rats (*Rattus rattus)* in Europe in late Antiquity and the Early Medieval Period (42, 43), suspected to represent the main reservoir species during the Second Pandemic (42).

Such a scenario would be congruent with the Second Pandemic, where the phylogeny of ancient genomes is in line with a single introduction and subsequent persistence in a local host species (12, 37, 44), although this hypothesis was challenged by an alternative scenario claiming multiple introductions on the basis of climatic data (45). Similar to the European Second Pandemic lineage (12, 13), strains emerging from the First Pandemic lineage have so far been recovered solely from ancient DNA of European plague burials, suggesting that the lineage either went extinct or persists in a yet-unsampled reservoir.

Origin of the Justinianic Plague. Based on available data, it has been suggested that the most parsimonious location for the divergence event that gave rise to the First Pandemic lineage is Central Asia (28). All published genomes of the branches 0.ANT1, 0.ANT2, and 0.ANT5 that frame the First Pandemic lineage in the phylogenetic tree were sampled in the autonomous Xingjiang region in northwestern China or in Kyrgyzstan (40, 46). In addition, an ancient second- to third-century *Y. pestis* genome from the Tian Shan mountains in Central Asia (28) branches off basal to all the First Pandemic genomes. The resulting claim that the Huns might have brought plague to Europe is, however, unsubstantiated due to the gap of more than three centuries before the onset of the First Pandemic.

Since the long shared branch of the First Pandemic genomes (45 SNPs) does not have any known extant descendants, this strain might have been maintained in a now extinct reservoir after its emergence in Central Asia. The first outbreak is reported in Pelusium, Egypt; an introduction from either Africa or Asia was presumed, given the sudden and dramatic onset of the pandemic. Previous assumptions of an African origin were mainly based on a single deeply diverging 0.PE strain "Angola" (47) and the reports of the Byzantine historian Evagrius Scholasticus, who wrote in his *Ecclesiastical History* that the plague began in "Ethiopia." However, there are legitimate doubts about the characterization of the "Angola" genome as a genuine African strain (26, 48) and the account of Evagrius has been assessed critically with historical and philological methods (49, 50). For an Asian origin, the sea route via the Red Sea and the Indian Ocean is a plausible scenario since India was well connected by marine traffic with the early Byzantine Empire (41). A suggested alternative scenario would require overland transport from the Eurasian Steppe via Iran to the Red Sea that is, so far, not supported by any data (51). In conclusion, we interpret the current data as insufficient to resolve the origin of the Justinianic Plague as a human epidemic.

Archaeological and Historical Context. Here, we present genomic evidence for the First Pandemic reaching the British Isles in the sixth century. This genome was recovered from a burial on the site of Edix Hill, close to Cambridge (Roman *Duroliponte*) and near a Roman road running north from London (*Londinium*) toward Lincoln (*Lindum Colonia*) via Braughing, all of which were Roman settlements. Based on archaeological dating in combination with its rather basal position within the clade, this genome is likely related to the very first occurrence of plague in

Britain suggested for 544 (*SI Appendix*), potentially introduced via sea communications with Brittany following the outbreak in central Gaul in 543 (Fig. 1, ref. 52). Interestingly, the genome was recovered from a single burial, underlining that, in small settlements, plague-induced mortality crises need not always involve a radical change in mortuary practice toward multiple or mass burials. The fact that two of the four additional Edix Hill individuals that appeared positive for plague in the MALT screening were buried in two simultaneous double burials nevertheless suggests that otherwise broadly typical cemeteries where multiple burials are relatively frequent are indeed a good indicator for epidemic events (18).

In addition, we were able to reconstruct four genomes from the Mediterranean basin and central France, where the historical records are more explicit about the presence of plague during the First Pandemic. Regarding Spain, the radiocarbon dating of the *Y. pestis*-positive individual from Valencia (432–610) would include the first outbreak reported for Spain in 543 in a contemporary chronicle (*SI Appendix*). The three unique SNPs identified in this genome, which separate it from the identified polytomy, however, may suggest its association with a later outbreak. Intriguingly, a canon of a church council held in 546 in Valencia dealing with burial practices for bishops in case of sudden death was recently connected with plague by philological and contextual analysis (53). Later outbreaks within the relevant time frame are documented in Spain's Visigothic kingdom, e.g., in 584 and 588 by Gregory of Tours, and by a funerary inscription dated 609 at Cortijo de Chinales 35 km northeast of Malaga (*SI Appendix*).

The second Mediterranean genome from Lunel-Viel in southern France represents another and significantly younger outbreak, since it belongs to an independent clade that derives from the same polytomy as the Spanish and German genomes, sharing 12 SNPs with the genomes of Saint-Doulchard and possessing 1 unique SNP. The radiocarbon dates for the inhumations give an interval of at least 567–618 (youngest lower and oldest upper boundary; *SI Appendix*, Table S13) overlapping with documented outbreaks in 571, 582, 588, 590, and possibly 599–600 in southern France (Fig. 1 *A* and *C*). Lunel-Viel's broader vicinity includes Arles, the seaport city of Marseille, and the Rhône mouth. Close to important coastal and fluvial shipping routes as well as Roman roads that facilitated the spread of plague (41), Lunel-Viel could have been affected by all five recorded epidemics. The initial outbreak, documented for Arles ca. 543, falls outside of some of the radiocarbon intervals. This is consistent with the phylogenetic analysis that shows a higher accumulation of SNPs in this genome. Thus, the victims at Lunel-Viel can most likely be attributed to an outbreak in the last quarter of the sixth century.

Within the site of Saint-Doulchard, two distinct genomes were found, one of which is represented by only one sample (LSD001.A), and the other is present in seven, including the sample LSD023.A with the highest coverage. The presence of two independent genomes in the same site, i.e., without one of them being directly ancestral to the other, has so far not been reported for the First or the Second Pandemic. Furthermore, the similar branch lengths of seven and nine SNPs derived from a common node do not allow for a clear temporal distinction. Also based on the stratigraphy of the site, the temporal structure cannot be fully resolved: all 11 plague-positive burials are dug into a trench that must have been open over the whole course of these inhumations. However, since the individuals were buried in distinct graves, they cannot be clearly assigned to a single event. Therefore, this finding can be explained by two hypotheses: First, the two strains might have struck the local population at the same time in a single outbreak; therefore, the victims were buried indiscriminately. This, however, would have implications for the epidemiology of *Y. pestis*, showing the parallel presence of different strains in a single outbreak. Second, the two strains could belong to two independent outbreaks within a shorter period of time, so the local community returned to the same structure, i.e., the trench, for emergency burials. Regarding the radiocarbon dating of adjacent burials within the trench, ca. 650–880, the closest historically reported outbreak is in Narbonne in 693. This would correspond with the relatively derived state of the two Saint-Doulchard strains compared with all other First Pandemic genomes. Other outbreaks in the West such as 663–666 and 684–687 in the British Isles, 707–709 in Spain, or 680 and 745–746 in Italy, might have been spatially limited and might not have spread to central France. Finally, an anecdote by Gregory of Tours in his sixth *Liber Historiarum* reports how the city of Narbonne was struck by plague repeatedly between 582 and 584, claiming the lives of those who fled the city when they returned to it. Although this episode is too early to account for the two strains in Saint-Doulchard, it showcases how a city was struck by plague multiple times over a short interval, as proposed in our second hypothesis.

In Bavaria, Germany, we detected *Y. pestis* in four sites (Dittenheim, Petting, Unterthürheim, Waging) in addition to the two previously published sites [Altenerding (7), Aschheim (8)]. Two of the reconstructed genomes were identical to Altenerding and Aschheim, suggesting that these four can be attributed to the same epidemic event. Some of the radiocarbon intervals of these sites fall even slightly before the onset of the First Pandemic, suggesting an association of this outbreak directly with the Justinianic Plague. Regarding the Edix Hill genome, this would in turn necessitate the accumulation of one (Edix Hill) to two (Altenerding cluster) SNPs within the onset of the First Pandemic between 541 and 544.

Intriguingly, the genome of Petting, Bavaria, falls not with the Altenerding cluster but in a distinct phylogenetic position. Since this strain also branches off from the common node with the other Bavarian strain as well as the French and Spanish genomes, this shows the presence of two independent strains and, therefore, presumably two independent epidemic events in early medieval Bavaria. This is striking, since we lack any historical records of the First Pandemic affecting southern Germany. The radiocarbon dates for the Bavarian sites are inconclusive and do not allow for a clear temporal separation of the two events. The higher number of accumulated SNPs nevertheless suggests a younger date for the epidemic represented by Petting. Further phylogeographic analyses are presented in *SI Appendix*.

Deletion Analysis. The analysis of virulence factors revealed a deletion of a ∼45-kb region in the most derived and most recent genomes thus far identified for the First Pandemic. This deletion contained two previously described virulence factors involved in host cell invasion and intracellular growth (*mgtB* and *mgtC*). Intriguingly, a similar deletion covering the same genomic region was detected in the most derived available Second Pandemic genomes from London New Churchyard (1560–1635) and Marseille (1720–1722). Genome decay by deletion or pseudogenization is a well-known trait of *Y. pestis* and has contributed to its distinct ecology and pathogenicity (54). Both deletions from the First and Second Pandemics are observed in genomes recovered from human victims. Therefore, it is reasonable to assume that the deletion may not have reduced the bacterium's virulence. Moreover, it affects a number of cell surface proteins—remnants of the motile lifestyle of nonpestis *Yersiniae* (55)—so the deletion might have even facilitated immune evasion.

Because none of the investigated modern strains harbored this specific deletion, this possible case of convergent evolution might be an adaptation to a distinct ecological niche in Europe or the Mediterranean basin since an ancient local reservoir is the most parsimonious hypothesis for both historical pandemics (13, 44).

Concluding Remarks. Our study offers insights into the first historically documented plague pandemic, complementing the limited power of conventional historical, archaeological, or paleoepidemiological research. Moreover, we show the potential of paleogenomic research for understanding historical and modern pandemics by a comparative approach on genomic features across millennia. Facing the problem of low-coverage genomic data with

a high environmental background—a notorious challenge in ancient DNA research—we have developed approaches to facilitate the authentication and confident phylogenetic placement of such genomes.

In the future, more extensive sampling of putative plague burials will help to draw a more comprehensive picture of the onset and persistence of the First Pandemic, especially on sites in the eastern Mediterranean basin, where not only is the Justinianic Plague reported to have started, but where also the eighth century outbreaks clustered according to the written records presently available. This will contribute to the comparative exploration of *Y. pestis*' microevolution and human impact in the course of past and present pandemics.

Materials and Methods

Sites and Samples. The acquisition and selection of samples followed two approaches: Focusing on Bavaria, we concentrated on one region, where the two previously reconstructed *Y. pestis* genomes attributed to the Justinianic Plague had been found (7, 8). Additionally, given the absence of robust genetic evidence from Gaul and the Mediterranean basin, which the surviving historical records depict as the epicenter of the pandemic, and the controversial presence of plague on the British Isles during the Justinianic Plague, we extended our screening to four sites with multiple burials in a broader geographical scope on the Mediterranean coast in France and Spain, central France, and inland Britain. Table 1 gives an overview of all tested sites.

For the first focus, we collected samples of 79 individuals from 46 burials belonging to 16 archaeological sites in Bavaria, Germany, and one site in Austria (Fig. 1*B*). Importantly, the dating of the burials spans the 4th to 10th century, including also burials dating before (8 individuals on three sites) and after (17 individuals on five sites) the Justinianic Plague (541–544). Since mass graves that could be indicative of an epidemic are unsurprisingly rare for the small settlements associated with early medieval cemeteries in Bavaria, we followed the approach of the previous successful studies (7, 8, 18): we systematically screened multiple burials, i.e., where two or more individuals were found in a context indicating a simultaneous burial, such as a common grave pit and articulated remains on the same level. Single burials were sporadically tested, if the context suggested a close connection to a multiple burial. Burials with indications of a violent death of the interred were excluded, since a coincidental acute infection with *Y. pestis* seems unlikely.

Within the Mediterranean basin, we tested inhumations from Valencia, Spain, and Lunel-Viel (Hérault), France. A contemporary chronicler records that bubonic infection devastated Spain during the first phase of the Justinianic Plague (541–544), and a recently published interpretation of a contemporary record argues that it reached Valencia presumably before 546 (53). Further textual references, including an epitaph dating to 609, document later Iberian outbreaks (56) (Fig. 1). In the Visigothic levels of the *Plaça de l'Almoina* in Valencia, several collective burials in an intramural cemetery were interpreted as possible plague burials (56, 57).

The historical evidence for the First Pandemic in France is more substantial, mainly based on the contemporary bishop and historian Gregory of Tours (58). He reports several plague outbreaks spanning from ca. 543 in the province of Arles through 588 in Marseille to 590 in Avignon (Fig. 1*C*). The site of Lunel-Viel, around 30 km southwest of the ancient Roman city of Nîmes and less than 100 km from the mentioned cities, revealed eight exceptional inhumations in demolition trenches unrelated to the nearby contemporary cemeteries (59).

In central France, we screened material from the site Le Pressoir in Saint-Doulchard, close to Bourges. Gregory of Tours (d. 594), explicitly mentions an outbreak at Bourges only in 571. Surviving written records are scarce leaving it undocumented whether other outbreaks in southern Gaul such as those just mentioned or the 693 outbreak in Narbonne reached Bourges (Fig. 1*C* and *SI Appendix*). The use of an existing ditch, most likely intended as an enclosure for the cemetery, as funerary space, gave however a first indication of a local mortality crisis, which was further substantiated by the presence of multiple burials and the demographic profile (60). From the 48 burials within the trench, 26 samples were selected mainly based on preservation, including 9 samples of multiple burials.

For the British Isles, the historical evidence for plague presence in the sixth century is controversial. Unlike later outbreaks in seventh-century Britain that are reported, e.g., by Bede, the identification of a disease occurring in the 540s and called *blefed* in Irish chronicles as bubonic plague, is mainly based on the coincidence with the Continental European outbreaks and thus uncertain. The same is true for Britain, where a great mortality (*mortalitas magna*) is reported in the *Annales Cambriae* (*SI Appendix*). For this study, we screened 22 individuals from the Anglo-Saxon cemetery of Edix Hill, well-connected to the Roman road network and Roman towns, and characterized by a number of multiple burials.

For the screening, one tooth (preferentially molar) per individual was used for every individual of a multiple burial, if available. For a number of individuals, additional teeth were tested, if sequencing the first gave a weak positive. For the collective burials from Valencia, a clear attribution to individuals was not assured, so multiple teeth were sampled per feature number, where possible. Detailed site descriptions can be found in *SI Appendix*, including a table with all screened samples (*SI Appendix*, Table S1).

Details on the radiocarbon dating and the cartography of the presented maps are described in separate sections of the *SI Appendix*.

Sample Preparation, DNA Extraction, qPCR, and MALT Screening. The sample preparation and DNA extraction for samples from Austria, France, Germany, and Spain were done in the ancient DNA facilities of the ArchaeoBioCenter of the Ludwig Maximilian University Munich, Germany, and the Max Planck Institute for the Science of Human History in Jena, Germany.

All teeth were cut along the cementoenamel junction, and the surface of the pulp chamber was drilled out with a dental drill from the crown and in some cases the root, aiming for 30–50 mg of bone powder. DNA was extracted based on the protocol published in ref. 61: The powder was suspended in 1 mL of extraction buffer (0.45 M EDTA pH 8.0, and 0.25 mg/mL proteinase K in UV-irradiated HPLC water) and incubated at 37 °C overnight on a rotor. After centrifugation, the supernatant was mixed with 10 mL of binding buffer (5 M guanidinium hydrochlorid, 40% isopropanol, and 90 mM sodium acetate) to bind the DNA on a silica column of either the MinElute purification kit (Qiagen) or the High Pure Viral Nucleic Acid Kit (Roche). After purification with washing buffer of the respective kit, the DNA was eluted in 100 μL of TET buffer (10 mM Tris·HCl, 1 mM EDTA, pH 8.0, 0.05% Tween 20).

All extracts were tested with the qPCR assay targeting a 52-bp region on the pPCP1 plasmid published in ref. 19 with minor changes (0.75 mg/mL BSA, additional 5% DMSO, EVA green instead of SYBR green, annealing for 30 s, elongation for 30 s, gradient from 60 to 90 °C). All samples showing an amplification with a melting peak between 74 and 80 °C were captured for *Y. pestis*.

The samples of Edix Hill, Britain, were prepared in the ancient DNA facility of the University of Cambridge, Department of Archaeology. Root portions of teeth were removed with a sterile drill wheel. These root portions were briefly brushed with 5% (wt/vol) NaOCl using a UV-irradiated toothbrush that was soaked in 5% (wt/vol) NaOCl for at least 1 min between samples. Roots were then soaked in 6% (wt/vol) bleach for 5 min. Samples were rinsed twice with ddH_2O and soaked in 70% ethanol for 2 min, transferred to a clean paper towel on a rack inside the glove box, UV irradiated for 50 min on each side, and then allowed to dry. They were weighed and transferred to clean, UV-irradiated 5-mL or 15-mL tubes for chemical extraction. Per 100 mg of each sample, 2 mL of EDTA Buffer (0.5 M, pH 8.0) and 50 μL of proteinase K (10 mg/mL) were added. Tubes were rocked in an incubator for 72 h at room temperature. Extracts were concentrated to 250 μL using Amplicon Ultra-15 concentrators with a 30-kDa filter. Samples were purified according to manufacturer's instructions using the Minelute PCR Purification Kit with the only change that samples were incubated with 100 μL of Elution Buffer at 37 °C for 10 min before elution.

Library Preparation. Of putatively positive extracts in the qPCR or MALT screening, 50 μL were turned into Illumina double-stranded DNA libraries with initial USER treatment (New England Biolabs) to remove postmortem damage in form of deaminated cytosines by consecutive incubation with uracil-DNA-glycosylase (UDG) and endonuclease VIII (25). To enhance the efficiency of subsequent double indexing, UDG-treated libraries were quantified by qPCR using IS7/IS8 primer and split for a maximum of 2×10^{10} DNA molecules. Every library was indexed with a unique index combination in a 10-cycle amplification reaction using Pfu Turbo Cx Hotstart DNA Polymerase (Agilent) (62, 63). The amplification 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). For the capture, the indexed libraries were amplified to 200–300 ng/μL using Herculase II Fusion DNA Polymerase (Agilent) and purified a second time as described.

The non-UDG library preparation for all Edix Hill samples was conducted using a protocol modified from the manufacturer's instructions included in the NEBNext Library Preparation Kit for 454 (E6070S; New England Biolabs) as detailed in ref. 64. DNA was not fragmented and reactions were scaled to half volume; adaptors were made as described in ref. 62 and used in a final

concentration of 2.5 μM each. DNA was purified on MinElute columns (Qiagen). Libraries were amplified using the following PCR setup: 50-μL DNA library, 1× PCR buffer, 2.5 mM $MgCl_2$, 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. Cycling conditions were as follows: 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 MinElute columns and eluted in 35 μL of EB. Samples were quantified using Quant-iT PicoGreen dsDNA kit (P7589; Invitrogen Life Technologies) on the Synergy HT Multi-Mode Microplate Reader with Gen5 software.

In-Solution Capture. For the in-solution capture, a probe set was generated using a fragment size of 52 bp and a tiling of 1 bp with the following genomes as templates: 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). The capture was performed as previously described (65) on 96-well plates with a maximum of two samples pooled per well and all blanks with unique index combinations in one well.

Sequencing and Data Processing. All captured products were sequenced either on an Illumina NextSeq500 or HiSeq4000 platform at the Max Planck Institute for the Science of Human History in Jena, Germany. The non-UDG libraries of Edix Hill samples were sequenced on Illumina NextSeq500 at the University of Cambridge Biochemistry DNA Sequencing Facility, and the FastQ files were processed on the Estonian Biocenter server and screened with MALT (20) using a reference set including full bacterial and viral genomes with 85% identity.

For all sequenced UDG libraries, de-multiplexed reads were processed with the EAGER pipeline (66) starting with Illumina adapter removal, sequencing quality filtering (minimum base quality of 20) and length filtering (minimum length of 30 bp). Sequencing data of paired-end and single-end sequencing were concatenated after adapter removal and merging. The same was done for samples from the same individual (DIT004) and all data from Lunel-Viel (LVC) due to low genomic coverage after ensuring an identical genotype. The sequencing results are shown in *SI Appendix*, Table S3.

Mapping against reference genomes of CO92 (chromosome NC_003143.1, plasmid pMT1 NC_003134.1, plasmid pCD1 NC_003131.1, plasmid pPCP1 NC_003132.1) was done with BWA using stringent parameters (–n 0.1, –l 32). Reads with low mapping quality were removed with Samtools (–q 37), and duplicates were removed with MarkDuplicates. For the plasmids, a merged reference was used, consisting of the CO92 reference of pCD1 (NC_003131.1), pMT1 (NC_003134.1), and pPCP1 [NC_003132.1, with base pairs 3,000–4,200 masked (19)], to avoid overestimation of coverage due to homologous regions. For the verification of positive qPCR results, we normalized the number of reads mapping to each plasmid with reads mapping to the chromosome and calculated the Mahalanobis distance for each sample to detect outliers. Based on this, we excluded the samples PEI001.A and DIR002.A as false positives (*SI Appendix*, Table S2).

The raw data of the Aschheim and Altenerding genomes were processed identically, however considering only the A120 sample for Aschheim instead of the combined A120+A76 data (7, 8).

SNP Calling and Evaluation. All genomes recovered from UDG-libraries with higher than 4.5-fold mean coverage including the Altenerding genome were assessed in the SNP analysis. Additionally, the sample WAG001.A was evaluated to explore its phylogenetic position, since it was the only positive sample of the relevant site.

The UnifiedGenotyper within the Genome Analysis Toolkit was used for SNP calling and creating VCF files for all genomes, using "EMIT_ALL_SITES" to generate calls for all positions in the reference genome. For the subsequent analyses, 233 previously published modern *Y. pestis* genomes (*SI Appendix*, Table S12), one genome from second- to third-century Tian-Shan mountains [DA101 (28)], one genome representing the Black Death from London East Smithfield [8291-11972-8124 (13)], and seven Second Pandemic genomes [Ellwangen; Bolgar; Marseille L'Observance OBS107, OBS110, OBS116, OBS124, OBS137 (12, 13)] were taken along together with *Y. pseudotuberculosis* (IP32953) as an outgroup. Previously identified problematic regions (26, 40) as well as regions annotated as repeat regions, rRNAs, tRNAs, and tmRNAs were excluded for all following analyses. MultiVCFAnalyzer, version 0.85 (67), was used for generating a SNP table with the following settings: Minimal coverage for base call of 3 with a minimum genotyping quality of 30 for homozygous positions, minimum support of 90% for calling the dominant nucleotide in a "heterozygous" position. All positions failing these criteria would be called "N" in the SNP table. For the SNP evaluation, all N positions of unique SNPs within the First Pandemic lineage were reevaluated, replacing N by "0" for not covered and lowercase letters for homozygous positions with maximum twofold coverage. To test for possible mixed infections or elevated contamination, all SNPs not passing the 90% threshold were plotted (*SI Appendix*, Fig. S1).

For the evaluation of unique and shared SNPs of First Pandemic genomes retrieved from UDG-treated libraries, we used the newly developed tool "SNPEvaluation" (https://github.com/andreasKroepelin/SNP_Evaluation) and a comparative mapping, using BWA with high stringent (–n 0.1, –l 32) and low stringent (–n 0.01, –l 32) mapping parameters, allowing for more mismatches in the latter. SNPs were called true positive when meeting the following criteria within a 50-bp window: (A) the ratio of mean coverage of low-stringent to high-stringent mapping is not higher than 1.00, (B) no "heterozygous" positions, and (C) no noncovered positions (*SI Appendix*, Tables S8 and S9). An assessment of this method is presented in *SI Appendix*, showing a maximal sensitivity (100% false positives detected) while accepting a high specificity (up to 3.49–8.57% of true positions filtered out).

SNP evaluation on the plasmids was done using the same criteria after mapping to the individual references as described above. For the SNP effect analysis, the remaining unique true SNPs were compared with the genome annotations of the CO92 *Y. pestis* reference genome (*SI Appendix*, Table S10).

Shared SNPs (*SI Appendix*, Table S11) were evaluated with the same criteria with minor modifications: The minimum threshold for calling a position was set to one read covering and SNPs were called true positive, if the SNP passed the criteria in more than half of the genomes under examination.

The Aschheim genome was evaluated separately (*SI Appendix*, Table S14) but with the same criteria. As previously addressed (7), the enormously high number of potential false-positive SNPs might not be explained solely by contamination by soil bacteria or sequencing errors but additionally by PCR or capture artifacts.

Phylogenetic Analyses. For the phylogenetic analyses, we aimed for one high coverage genome per site to minimize missing data in the SNP alignment, excluding the genome of EDI003.A, EDI004.A, and UNT004.A after assuring no conflicting positions with EDI001.A and UNT003.A, respectively, in the SNP evaluation. A maximum-likelihood tree [RAxML 8 (30) using the GTR substitution model, Fig. 2*A*; for full tree, see *SI Appendix*, Fig. S4] was generated without exclusion of missing and ambiguous data (full SNP alignment), resulting in a total number of 6,580 SNPs. Robustness of all tree nodes was tested by the bootstrap methods using 1,000 pseudoreplicates.

A detailed tree of the First Pandemic lineage was drawn manually based on the performed SNP evaluation, excluding all potential false-positive SNPs (Fig. 2*B*).

Analysis of Virulence Factors and Genome Decay. The presence/absence analysis for genes was performed with BEDTools (68) by calculating the percentage across each gene using the function "coverage," which calculates the percentage of bases in a given window being covered by at least one read (3). Since gene duplications can affect the mapping quality, the mapping quality filter of BWA was set to 0 (–q 0) to generate a bam-file as input. For the heatmap of virulence factors (Fig. 3), a collection of proven and putative virulence genes (31, 32) was evaluated. With this method, only deletions and pseudogenization by large gene truncations can be detected; a test for pseudogenization by frameshift mutations was not attempted. The more extensive analysis on genome decay was based on the annotation file for the reference genome CO92 (55) by extracting all regions annotated as "gene." For the exact determination of the start and end positions of deletions, mapping with BWA-MEM was performed (69).

ACKNOWLEDGMENTS. We are grateful to Aditya K. Lankapalli, Aida Andrades Valtueña, and all members of the Department of Archaeogenetics, Max Planck Institute for the Science of Human History for support and fruitful discussions, and Raphaela Stahl, Marta Burri, Cäcilia Freund, Franziska Aron, Antje Wissgott, and Guido Brandt for their assistance in the laboratory. We thank the staff of the State Collection for Anthropology and Palaeoanatomy Munich for support during sample collection and Ronny Friedrich at the Curt-Engelhorn-Centre Archaeometry gGmbH, Mannheim, for providing additional information on radiocarbon dates. Furthermore, we thank Kyle Harper and Henry Gruber for their correspondence. The excavation of Saint-Doulchard Le Pressoir was led by Philippe Maçon, Service d'Archéologie Préventive de l'Agglomération de Bourges Plus, who initiated the archaeogenetic investigation together with D.C. and R.D. This study was supported by the European Research Council starting grant APGREID (to J.K.), by the Wellcome Trust (Award 2000368/Z/15/Z to J.E.R.), and by the Justinianic Pandemic Working Group at Max Planck–Harvard Research Center for the Archaeoscience of the Ancient Mediterranean/Initiative for the Science of the Human Past at Harvard.

1. R. Yang, A. Anisimov, *Yersinia pestis: Retrospective and Perspective* (Springer, Dordrecht, 2016).
2. T. Burki, Plague in Madagascar. *Lancet Infect. Dis.* **17**, 1241 (2017).
3. A. Andrades Valtueña *et al*., The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691.e8 (2017).
4. S. Rasmussen *et al*., Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
5. M. A. Spyrou *et al*., Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234 (2018).
6. N. Rascovan *et al*., Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic decline. *Cell* **176**, 295–305.e10 (2019).
7. M. Feldman *et al*., A high-coverage *Yersinia pestis* genome from a sixth-century Justinianic Plague victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).
8. D. M. Wagner *et al*., *Yersinia pestis* and the Plague of Justinian 541–543 AD: A genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
9. Benedictow OJ (2004) *The Black Death 1346–1353: The Complete History* (Boydell & Brewer, Woodbridge, UK).
10. K. I. Bos *et al*., A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
11. J.-N. Biraben, *Les Hommes et la Peste en France et dans les Pays Européens et Méditerranéens* (Mouton, Paris, 1975).
12. M. A. Spyrou *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).
13. K. I. Bos *et al*., Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).
14. I. Wiechmann, G. Grupe, Detection of *Yersinia pestis* DNA in two early medieval skeletal finds from Aschheim (Upper Bavaria, 6th century A.D.). *Am. J. Phys. Anthropol.* **126**, 48–55 (2005).
15. M. Drancourt *et al*., Genotyping, Orientalis-like *Yersinia pestis*, and plague pandemics. *Emerg. Infect. Dis.* **10**, 1585–1592 (2004).
16. M. Drancourt *et al*., *Yersinia pestis* Orientalis in remains of ancient plague patients. *Emerg. Infect. Dis.* **13**, 332–333 (2007).
17. M. T. P. Gilbert *et al*., Absence of *Yersinia pestis*-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology* **150**, 341–354 (2004).
18. M. Harbeck *et al*., *Yersinia pestis* DNA from skeletal remains from the 6th century AD reveals insights into Justinianic Plague. *PLoS Pathog.* **9**, e1003349 (2013).
19. V. J. Schuenemann *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–E752 (2011).
20. Å. J. Vågene *et al*., *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
21. D. H. Huson *et al*., MEGAN community Edition–Interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Comput. Biol.* **12**, e1004957 (2016).
22. P. C. Mahalanobis, On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* **2**, 49–55 (1936).
23. A. W. Briggs *et al*., Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).
24. C. Warinner *et al*., A robust framework for microbial archaeology. *Annu. Rev. Genomics Hum. Genet.* **18**, 321–356 (2017).
25. A. W. Briggs *et al*., Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
26. G. Morelli *et al*., *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).
27. G. L. Kay *et al*., Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat. Commun.* **6**, 6717 (2015).
28. P. B. Damgaard *et al*., 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
29. P. S. G. Chain *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 (2004).
30. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
31. D. Zhou, R. Yang, Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect. Immun.* **77**, 2242–2250 (2009).
32. D. Zhou *et al*., Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, microtus. *J. Bacteriol.* **186**, 5147–5152 (2004).
33. A. Derbise, E. Carniel, YpfΦ: A filamentous phage acquired by *Yersinia pestis*. *Front. Microbiol.* **5**, 701 (2014).
34. D. C. Ford, G. W. P. Joshua, B. W. Wren, P. C. F. Oyston, The importance of the magnesium transporter MgtB for virulence of *Yersinia pseudotuberculosis* and *Yersinia pestis*. *Microbiology* **160**, 2710–2717 (2014).
35. P. S. G. Chain *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).
36. J. D. Fetherston, R. D. Perry, 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).
37. M. A. Spyrou *et al*., A phylogeography of the second plague pandemic revealed through the analysis of historical *Y. pestis* genomes. bioRxiv:https://doi.org/10.1101/481242 (30 November 2018).
38. S. Hänsch *et al*., The *pla* gene, encoding plasminogen activator, is not specific to *Yersinia pestis*. *BMC Res. Notes* **8**, 535 (2015).
39. E. Garcia *et al*., Pestoides F, an atypical *Yersinia pestis* strain from the former Soviet Union. *Adv. Exp. Med. Biol.* **603**, 17–22 (2007).
40. Y. Cui *et al*., Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 577–582 (2013).
41. K. Harper, *The Fate of Rome: Climate, Disease, and the End of an Empire* (Princeton University Press, Princeton, 2017).
42. M. McCormick, Rats, communications, and plague: Toward an ecological history. *J. Interdiscip. Hist.* **34**, 1–25 (2003).
43. F. Audoin-Rouzeau, J.-D. Vigne, Le rat noir (*Rattus rattus*) en Europe antique et mediévale: Les voies du commerce et l'expansion de la peste. *Anthropozoologica* **25**, 399–404 (1997).
44. L. Seifert *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).
45. B. V. Schmid *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 (2015).
46. G. A. Eroshenko *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).
47. M. Eppinger *et al*., Genome sequence of the deep-rooted *Yersinia pestis* strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. *J. Bacteriol.* **192**, 1685–1699 (2010).
48. M. H. Green, "Taking "pandemic" seriously: Making the Black Death global" in *Pandemic Disease in the Medieval World: Rethinking the Black Death*, M. H. Green, Ed. (ARC Humanities Press, Kalamazoo, MI, 2014).
49. P. Sarris, The Justinianic Plague: Origins and effects. *Contin. Chang.* **17**, 169–182 (2002).
50. P. Allen, The "Justinianic" Plague. *Byzantion* **49**, 5–20 (1979).
51. U. Schamiloglu, "The plague in the time of justinian and central Eurasian history: An agenda for research" in *Central Eurasia in the Middle Ages. Studies in Honour of Peter B. Golden*, I. Zimonyi, O. Karatay, Eds. (Harrasowitz Verlag, Wiesbaden, 2016).
52. L. K. Little *et al*., *Plague and the End of Antiquity: The Pandemic of 541–750*, L. K. Little, Ed. (Cambridge University Press, 2007).
53. H. Gruber, Indirect evidence for the social impact of the Justinianic pandemic: Episcopal burial and conciliar legislation in Visigothic Hispania. *J. Late Antiq.* **11**, 193–215 (2018).
54. A. McNally, N. R. Thomson, S. Reuter, B. W. Wren, 'Add, stir and reduce': *Yersinia* spp. as model bacteria for pathogen evolution. *Nat. Rev. Microbiol.* **14**, 177–190 (2016).
55. J. Parkhill *et al*., Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
56. M. Kulikowski, "Plague in Spanish Late Antiquity" in *Plague and the End of Antiquity: The Pandemic of 541–750*, L. K. Little, Ed. (Cambridge University Press, Cambridge, 2007), pp. 150–170.
57. L. Alapont Martín, A. V. Ribera i Lacomba, (2009) "Topografía y jerarquía funeraria en la Valencia tardo-antigua" in *Morir en el Mediterráneo Medieval: Actas del III Congreso Internacional de Arqueología, Arte e Historia de La Antigüedad Tardía y Alta Edad Media Peninsular*, J. López Quiroga, A. M. Martínez Tejera, Ed. (Oxford), pp 59–88.
58. A. J. Stoclet, "Consilia humana, ops divina, superstitio: Seeking succor and solace in times of plague, with particular reference to Gaul in the Early Middle Ages" in *Plague and the End of Antiquity: The Pandemic of 541–750*, L. K. Little, Ed. (Cambridge University Press, Cambridge, 2007), pp. 135–149.
59. C. Raynaud, "Les nécropoles de Lunel-Viel (Hérault) de l'Antiquité au Moyen Âge." *Revue Archéologique de Narbonnaise*, Supplément 40 (Presses Universitaires de la Méditerranée, Montpellier, France, 2010).
60. P. Maçon, R. Durand, M. Salin, F. Dominin, *Le Pressoir–Rapport Final d'Opération de Fouille Préventive* (Service d'Archéologie Préventive de l'Agglomération de Bourges Plus, Bourges, 2011).
61. J. Dabney *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 (2013).
62. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
63. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
64. M. Rasmussen *et al*., The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
65. Q. Fu *et al*., DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013).
66. A. Peltzer *et al*., EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
67. K. I. Bos *et al*., Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
68. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
69. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*:1303.3997 (16 March 2013).

5 Manuscript B: Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes

Maria A. Spyrou*, Marcel Keller*, Rezeda I. Tukhbatova, Christiana L. Scheib, Elizabeth A. Nelson, Aida Andrades Valtueña, Gunnar U. Neumann, Don Walker, Amelie Alterauge, Niamh Carty, Craig Cessford, Hermann Fetz, Michaël Gourvennec, Robert Hartle, Michael Henderson, Kristin von Heyking, Sarah A. Inskip, Sacha Kacki, Felix M. Key, Elizabeth L. Knox, Christian Later, Prishita Maheshwari-Aplin, Joris Peters, John E. Robb, Jürgen Schreiber, Toomas Kivisild, Dominique Castex, Sandra Lösch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos & Johannes Krause
*contributed equally

Published in *Nature Communications* 10, 4470 (2019) doi:10.1038/s41467-019-12154-0

This page is intentionally left blank.

nature COMMUNICATIONS

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

The second plague pandemic, caused by *Yersinia pestis*, devastated Europe and the nearby regions between the 14th and 18th 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.

[1] Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [2] Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany. [3] SNSB, State Collection for Anthropology and Palaeoanatomy Munich, 80333 Munich, Germany. [4] Laboratory of Structural Biology, Kazan Federal University, Kazan, Russian Federation 420008. [5] Institute of Genomics, University of Tartu, Riia 23b, 51010 Tartu, Estonia. [6] MOLA (Museum of London Archaeology), London N1 7ED, UK. [7] Department of Physical Anthropology, Institute for Forensic Medicine, University of Bern, 3007 Bern, Switzerland. [8] Department of Archaeology, University of Cambridge, Downing St, Cambridge CB2 3ER, UK. [9] Archaeological Service, State Archive Nidwalden, 6371 Nidwalden, Switzerland. [10] Archeodunum SAS, Agency Toulouse, 8 allée Michel de Montaigne, 31770 Colomiers, France. [11] McDonald Institute for Archaeological Research, University of Cambridge, Downing St, Cambridge CB2 3ER, UK. [12] PACEA, CNRS Institute, Université de Bordeaux, 33615 Pessac, France. [13] Department of Archaeology, Durham University, South Rd, Durham DH1 3LE, UK. [14] Institute for Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [15] Bavarian State Department of Monuments and Sites, 80539 Munich, Germany. [16] 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. [17] Dig it! Company GbR, 86971 Peiting, Germany. [18] Department of Human Genetics, Katholieke Universiteit Leuven, 3000 Leuven, Belgium. [19] 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) or to K.I.B. (email: bos@shh.mpg.de) or to J.K. (email: 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[1,2]. Its causative agent, *Yersinia pestis*[3], 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[4–6].

The first historically documented outbreaks of the second pandemic seem to have occurred in 1346 in the Lower Volga and Black Sea regions[1,7]. 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%[1]. 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)[8], Barcelona (Spain)[9], London (England)[10] and Oslo (Norway)[8] 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)[8], 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[2,11]. 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[9,12–14]. While data from climatic proxies are considered as supportive of the former hypothesis[12], genetic evidence is interpreted in both directions[8,9,13]. 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[9,15,16], 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[9,10,14] challenges our ability to construct hypotheses regarding the number of lineages responsible for the numerous post-BD outbreaks in Europe[2,11] 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[8]. 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[17] in 180 specimens from the cities of London ($n = 40$) in England, Toulouse ($n = 42$) in France, Brandenburg an der Havel[13] ($n = 3$), Landsberg am Lech ($n = 10$), Manching-Pichl[13] ($n = 28$), 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)[18] (see Methods). The output was post-processed within the pathogen screening pipeline HOPS[19]. 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[20,21] from all putatively positive samples and used a *Y. pestis* whole-genome in-solution capture approach[22] 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)[3]. 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[13], one from Starnberg, one from Landsberg am Lech, two from Brandenburg an der Havel[13] (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[9] (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[23–27] (as listed in ref. [28]), which represent most of the published *Y. pestis* genetic diversity. In addition, we included previously published second

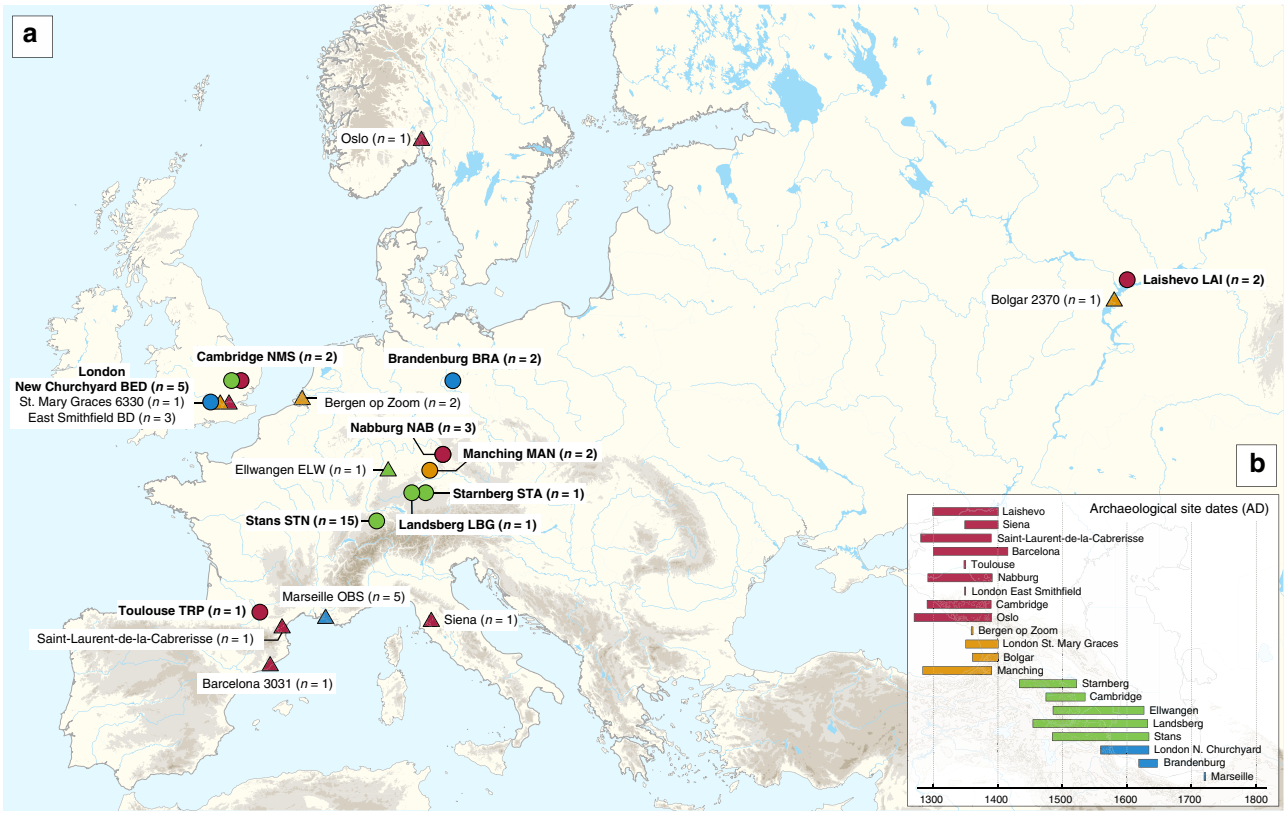


Fig. 1 Archaeological site locations and chronologies. **a** Map showing the geographic locations of archaeological sites from which second pandemic (14th- to 18th-century AD) *Y. pestis* genomes have been reconstructed (≥1-fold). The number (*n*) of genomes obtained from each site is shown in brackets. 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 ($n = 15$)[8–10,14], a 6th-century AD isolate from Germany[29], a 2nd- to 3rd-century AD isolate from the Tian Shan mountains in Kyrgyzstan[30], as well as three Bronze Age isolates from the Altai and Volga regions[31,32] (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[18,29,33]. 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[28] (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[8] 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)[10] and Bolgar City[9] (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)[23]. 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[9,10]. 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[28] 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[8] (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 ≥ 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 an der Havel	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 Freidhof, Nabburg	1292–1392[a]	684,029	33.5	8.1	70.2	54.8	49.7
NMS002.A0101	Augustinian Friary (New Museums site), Cambridge	1475–1536[a,b]	855,185	2.3	12.5	94.8	67.9	47.3
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

[a]Dates based on radiocarbon dating of collagen
[b]Dates based on archaeological context information

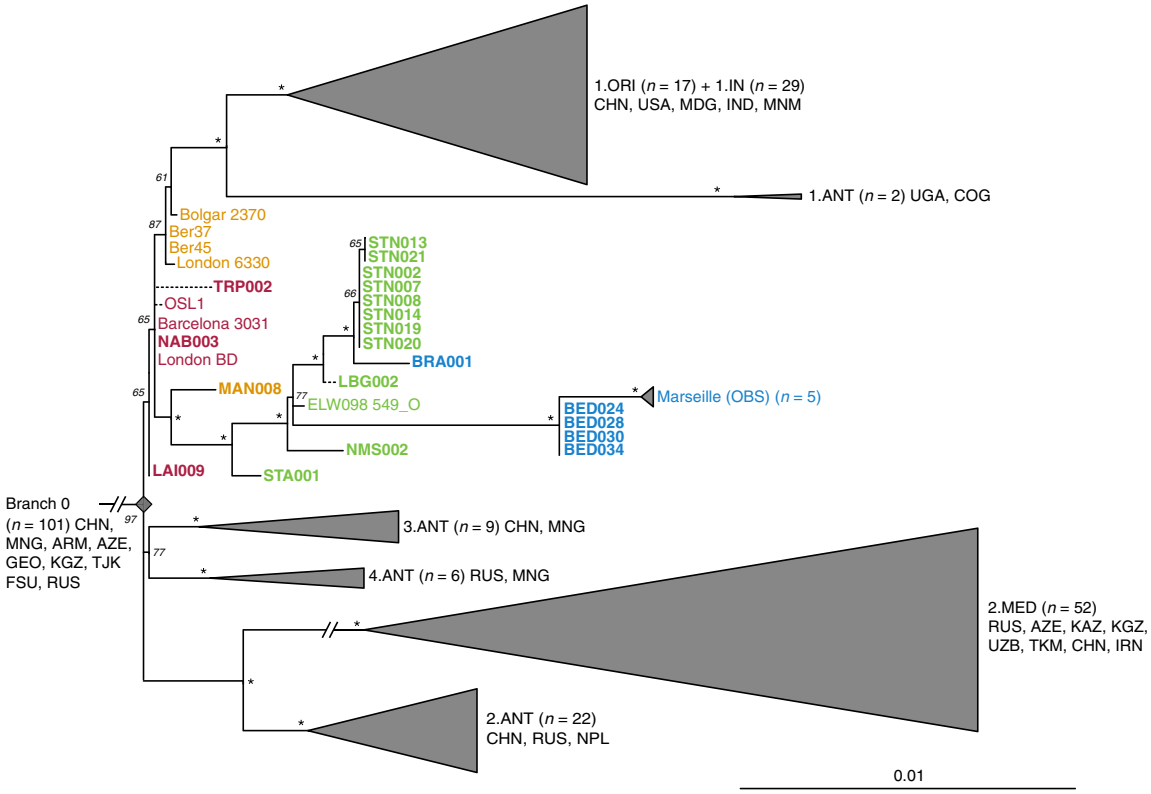


Fig. 2 Phylogenetic positioning of second pandemic strains. A maximum likelihood phylogeny was generated allowing for up to 3% missing data (97% partial deletion) and considering a total of 6,058 single nucleotide polymorphisms (SNPs). The image shows a graphical representation of Branches 1–4 (see Supplementary 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 ≥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[8]. 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[34], 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)[9], and Marseille, France (1720–1722)[14], 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)[13] (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)[13] 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

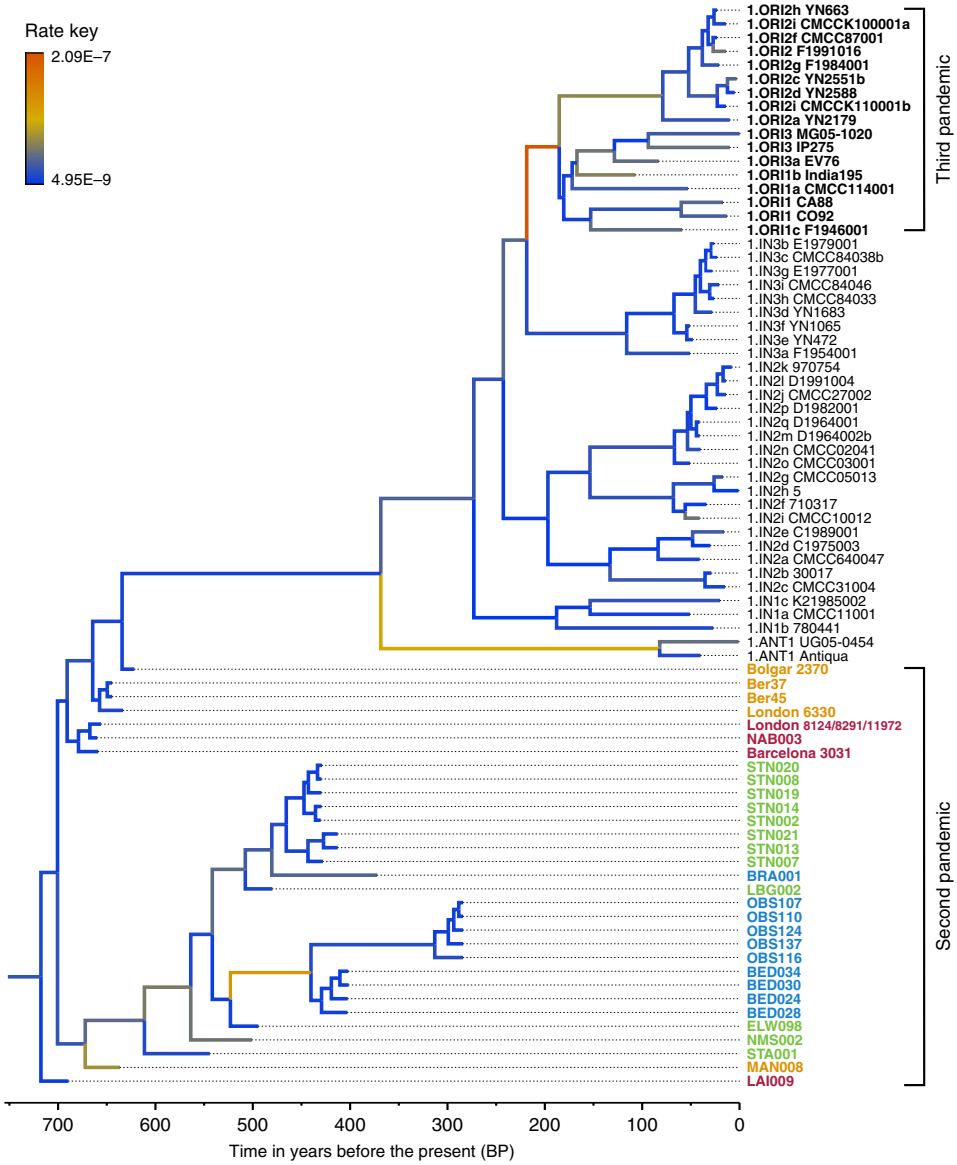


Fig. 3 Substitution rate variation across the *Y. pestis* Branch 1. The figure presents a maximum clade credibility (MCC) phylogenetic tree generated using BEAST v1.8[85] (rooted with 2.MED KIM10—outgroup not shown). The tree was viewed in FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/), and 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[16,23]. 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

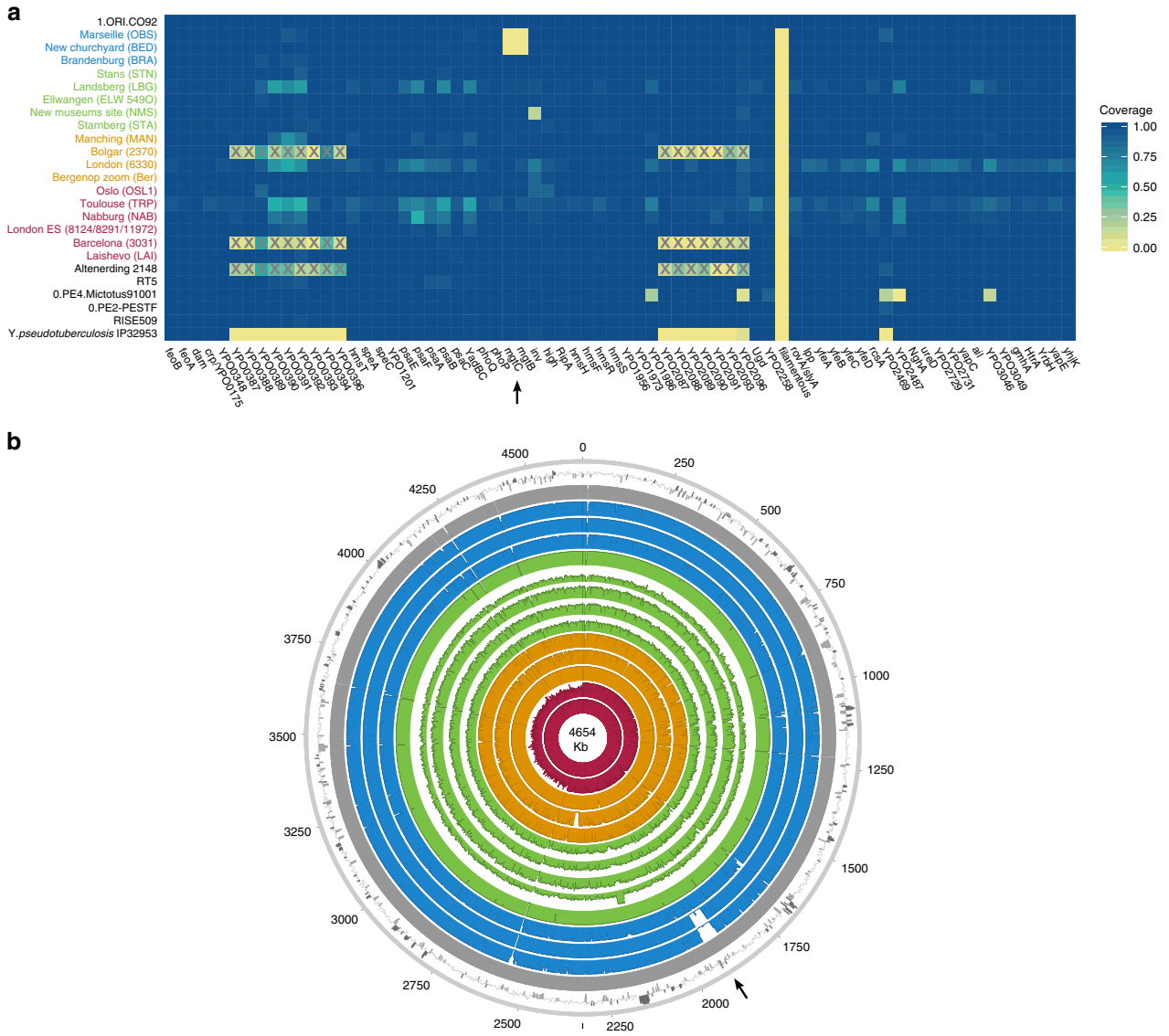


Fig. 4 Assessment of chromosomal and gene-specific coverage in *Y. pestis*. **a** A comparison of genetic profiles was performed across newly reconstructed and previously published second pandemic genomes (in red, orange, green and blue). Here, we show an assessment of the presence or absence of 80 previously defined[36] potential virulence and evolutionary determinants across the *Y. pestis* chromosome. Published genomes from the Bronze Age period[31,32] (RISE509 and RT5), from the first pandemic[29] (6th-century Altenerding 2148), from modern-day isolates (0.PE2, 0.PE4 and 1.ORI)[23], as well as *Y. pseudotuberculosis* IP32953[60], are also shown for comparative purposes. The colour scale ranges from 0 (not covered—yellow) to 1 (entirely covered—blue) according to the relative proportion of gene/locus covered. The heatmap was plotted in R version 3.4.1[82] using the ggplot2 package[89]. Boxes marked with "X" indicate genomic loci that were not part of the *Y. pestis* probe design when the respective isolates were captured[9,29]. Refer to Supplementary Fig. 20 for presence/absence of virulence-associated genes across the pMT1, pPCP1 and pCD1 plasmids. **b** Chromosomal coverage plots made with the Circos[90] software. The plots were constructed to a maximum coverage of 20-fold, and the average coverage was calculated over 3,000-bp windows. Genomes are shown in chronological order from oldest (innermost circle) to youngest (outermost circle) as follows: LAI009, London BD 8124/8291/11972 (BD representative), Ber45, Bolgar 2370, MAN008, STA001, NMS002, ELW098/549_O, LBG002, STN014, BRA002, BED030, OBS137 and the reference genome CO92. The outermost ring represents fluctuations in GC content (%) across CO92, where dark and light grey bars show deviations from the genomic mean (47.6%) by at least one standard deviation

earlier (Fig. 3 and Supplementary Data 7). The second is the branch leading to the 1.ANT strains isolated from Africa (Congo and Uganda) (Fig. 3 and Supplementary Data 8). 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 hypothesised[9,35]. The third, which displays the fastest rate within the entire Branch 1, is the branch leading to 1.ORI isolates (Fig. 3 and Supplementary Data 9), 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.

Analysis of virulence-associated genomic profiles. To investigate the genomic profiles of all newly reconstructed genomes, we analysed the presence or absence of potential virulence-associated and evolutionary determinant genes located on the *Y. pestis* chromosome (Fig. 4a) and plasmids (Supplementary Fig. 20)[36,37], in comparison to published representatives of 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" (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[9,29] (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 invasin is associated with epithelial colonisation of *Y. pseudotuberculosis* and *Yersinia enterocolitica*, it is known to have been inactivated in *Y. pestis*[38]. By contrast, magnesium transporters are considered vital for *Y. pestis* intracellular survival under low Mg^{2+} conditions[39], 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[40]. 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[41]. 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)[14], which was one of the last major epidemics that occurred in continental Europe[42]. 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)[28] sequenced from victims of the first plague pandemic (6th–8th centuries AD)[28]. 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[28].

Discussion

A series of studies have sufficiently demonstrated the preservation of *Y. pestis* in ancient human remains from a wide temporal transect[8–10,14,22,29,31,32,43]. 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[7]. A commonly accepted view dates its arrival in the southwest, particularly in cities of Astrakhan and Sarai, in 1346[1,44] 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[7,44]) may have followed an alternative route via the Baltic Sea, occurring at the end of the BD between 1351 and 1353[1,7,44]. 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[9], 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[1] (currently represented by strains from Siena, Saint-Laurent-de-la-Cabrerisse, Barcelona and Toulouse[8,9]). These results suggest that the N07-derived SNP previously termed "p1"[9] (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)[23], 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[10,23] (Fig. 2). Such claims, however, cannot yet be verified given; (1) the apparent East Asian sampling bias of modern isolates[23,45], (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[46]. In addition, recently published modern *Y. pestis* genomes from Central Asia show a rich diversity in the local plague foci[26,27], 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[12] as well as PCR[47] and genomic[8] 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[8–10,14] 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)[10] and the city of Bolgar (2370)[9], as well as extant strains from Africa (1.ANT)[48], and, most importantly, a worldwide set of isolates associated with the third pandemic (1.ORI, 19th–20th centuries)[15,16,23] (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[9].

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[49], 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[50], although sources favouring local epidemic eruptions also exist[51]. 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[15,16,52,53]. 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[40,54]. Their co-expression has been shown to affect the virulence exerted by other pathogenic enterobacteria under laboratory conditions[55,56] and both genes have been proposed as potential drug targets[40,57]. Moreover, the function of *mgtB* was shown to be temperature-dependent, being active at 37 °C but not at 20 °C[58], 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)[28], 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[59,60]. 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[1,61]. 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[62,63], climatic[12,64,65] and epidemiological data[66], as well as a critical re-evaluation of historical records[67], 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 cementoenamel 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[68]. 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[68], 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[17]. 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[17]. 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_2O, 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[69]. Adaptors were constructed as previously described[21]. Indexing PCR reactions were set up as follows: 50 µl of DNA library, 1× PCR buffer, 2.5 mM $MgCl_2$, 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)[18], integrated in the pathogen screening pipeline HOPS[19], 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 MALTExtract (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[19].

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[21], 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[20,70]. 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[71], 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 (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[22], 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)[72] 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)[74], implemented in EAGER, against the CO92 reference genome (NC_003143.1)[3] 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)[75]. Our newly reconstructed genomes were analysed alongside previously published *Y. pestis* genomes, which included a modern-day dataset of 233 genomes[23–27,48] (as listed in ref. [28]), three Bronze Age strains[31], a 2nd- to 3rd-century AD isolate from the Tian Shan mountains in Kyrgyzstan[30], one Justinianic strain (Altenerding 2148)[29], 15 previously published historical genomes from the second plague pandemic[8–10,14] and a Y. pseudotuberculosis strain (IP32953)[60] 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[16,23]. 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 v3.1i (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)[78] using the Generalised Time Reversible (GTR)[79] 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[8]. 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[80] (-n 0.01, -l 16) than the ones previously reported[8], 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[29]. 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) 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[28] 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[75], using the "EMIT_ALL_SITES" option that generated a call for all positions in the reference genome. We then used MultiVCFAnalyzer v0.85 (ref. [33]) 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. [82]) 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[78] using all Branch 1 genomes[3,8–10,14,16,23,48,83,84] (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. [85]) 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[79] substitution model (4 gamma rate categories) and a lognormal relaxed clock (clock rate tested and strict clock rejected in MEGA7[77]) were used to set up two separate analyses using the coalescent constant size[86] and coalescent Bayesian skyline[87] 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. [85]). 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[85], 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[85].

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[36] was calculated using bedtools[88]. The results are plotted in the form of a heatmap using the ggplot2 (ref. [89]) package in R version 3.4.1 (ref. [82]) and can be viewed in Fig. 4. In addition, we used BWA-MEM[80] 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[90] 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

1. Benedictow, O. J. *The Black Death, 1346-1353: The Complete History* (Boydell and Brewer, Woodbridge, UK, and Rochester, N.Y., 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* (Mouton, Paris, 1976).
3. Parkhill, J. et al. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413**, 523–527 (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 (2005).
5. Prentice, M. B. & Rahalison, L. Plague. *Lancet* **369**, 1196–1207 (2007).
6. 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).
7. Alexander, J. T. *Bubonic Plague in Early Modern Russia: Public Health and Urban Disaster* (Johns Hopkins University Press, Baltimore, Maryland, USA, 1980).
8. 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).
9. 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).
10. Bos, K. I. et al. A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506–510 (2011).
11. Büntgen, U., Ginzler, C., Esper, J., Tegel, W. & McMichael, A. J. Digitizing historical plague. *Clin. Infect. Dis.* **55**, 1586–1588 (2012).
12. 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).
13. 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).
14. Bos, K. I. et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).
15. Pollitzer, R. *The Plague* No. 26 (World Health Organization, Geneva, 1954).
16. Morelli, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).
17. 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).
18. Vågene, A. J. et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
19. 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).
20. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
21. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5448 (2010).
22. Andrades Valtueña, A. et al. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691 (2017). e3688.
23. Cui, Y. et al. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc. Natl Acad. Sci. USA* **110**, 577–582 (2013).
24. 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 (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. 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**, 1106 (2018).
27. 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).
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 Mg2+ 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., Séguy, 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. Epidémiologie 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 phagosome 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. Snavely, 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).
57. Belon, C. et al. A macrophage subversion factor is shared by intracellular and extracellular pathogens. *PLoS Pathog.* **11**, e1004969 (2015).
58. Snavely, 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).
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, Cäcilia 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. Fattahov 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.

This page is intentionally left blank.

6 Manuscript C: Ancient *Yersinia pestis* genomes provide no evidence for the origins or spread of the Justinianic Plague

Marcel Keller, Maria A. Spyrou, Michael McCormick, Kirsten I. Bos, Alexander Herbig, Johannes Krause

Submitted to *bioRxiv* on 31st of October 2019, revised on 12th of November 2019, doi:10.1101/819698

Ancient *Yersinia pestis* genomes provide no evidence for the origins or spread of the Justinianic Plague

Marcel Keller[1,2]*, Maria A. Spyrou[1], Michael McCormick[3,4], Kirsten I. Bos[1], Alexander Herbig[1]*, Johannes Krause[1,4]*

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

[2]Institute of Genomics, University of Tartu, Tartu, Estonia

[3]Initiative for the Science of the Human Past, Department of History, Harvard University, Cambridge, USA

[4]Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean

*For correspondence: marcel.keller@ut.ee (MK); herbig@shh.mpg.de (AH); krause@shh.mpg.de (JK)

Competing interests: The authors declare that no competing interests exist.

Abstract

Along with the publication of 137 ancient human genomes retrieved from archaeological remains of the Eurasian steppe, Damgaard et al., 2018 identified two individuals infected with *Yersinia pestis*, yielding one genome with 0.24x average coverage (DA147, 6th–9th c. AD) and another with 8.7x (DA101, 2nd–3rd c. AD). A phylogenetic analysis performed on the latter placed it in a position ancestral to a 6th-century Justinianic genome from Aschheim, Germany. These results are used to fuel an argument that the Justinianic Plague (541–544 AD) "was brought to Europe towards the end of the Hunnic period through the Silk Road along the southern fringes of the steppes" in contrast to the leading hypothesis of introduction via the Red Sea that is supported by historical accounts. In our reanalysis, we question the contested historical context of the presented genomes with the Justinianic Plague and show that the lower coverage genome might be rather related to the Black Death (1346–1353 AD).

Introduction

The recent sequencing of dozens of pathogen genomes reconstructed from ancient DNA enabled increased-resolution phylogeographic studies on the spread of infectious diseases in prehistoric and historic times, especially in the context of human migration, mobility and trade (Andrades Valtueña et al., 2017; Bos et al., 2016; Keller et al., 2019; Namouchi et al., 2018; Rascovan et al., 2019; Rasmussen et al., 2015; Spyrou et al., 2019, 2016; Vågene et al., 2018). This is especially true for plague with its long and richly documented history and the abundance of published ancient genomes of its causative agent, *Yersinia pestis*. Interpretation of phylogenetic data in the context of human history requires careful assessment of tree topologies, branch lengths and mutation rates as well as thoughtful consilient approaches in integrating historical and archaeological data to prevent overly simplistic, deterministic or even erroneous interpretations.

For the Second Pandemic, the geographic origin and the possible persistence within Europe after the Black Death (1346–1352 AD) are the subject of ongoing scientific and scholarly discussion (Bos et al., 2016; Namouchi et al., 2018; Schmid et al., 2015; Spyrou et al., 2016; Spyrou et al., 2019). Whereas the first comprehensive phylogenetic study on *Y. pestis* favoured an East Asian origin (Cui et al., 2013), other scenarios assume an origin in Central Asia or the Caucasus (Benedictow, 2004; Namouchi et al., 2018; Sussman, 2011). Similarly, the origin of the Justinianic Plague (541–544 AD) has long been hypothesized to have originated in Africa (Achtman et al., 1999; Cui et al., 2008; Sarris, 2002). More recent studies however agree that the strains causing the First Pandemic (541–750 AD) likely emerged in Central Asia (Eroshenko et al., 2017; Harper, 2017; Wagner et al., 2014). The fact that the first outbreak of the Justinianic Plague is reported for Pelusium, Egypt nevertheless raises questions about the history and itinerary of the causative *Y. pestis* strain prior to this outbreak. The currently favoured scenario is an introduction via the Red Sea from India (Harper, 2017; Tsiamis et al., 2009) since there are no historical sources supporting a land route via the Levant or the Arabian peninsula (Schamiloglu, 2016). As such, discrepancies that arise from different analytical approaches raise questions about the history of *Y. pestis* prior to the first documented Justinianic Plague outbreak.

In a recent publication, Damgaard et al., 2018 presented two ancient *Y. pestis* genomes: one from the Tian Shan region (DA101, 2nd–3rd c. AD, 8.7-fold average coverage), branching ancestral to the First Pandemic lineage; and one from the Caucasus (DA147, 6th–9th c. AD, 0.24-fold average coverage) which was not further investigated. Damgaard et al.'s phylogenetic analysis places DA101 ancestral to the published genome from Aschheim. This

positioning is supported by a single SNP shared between the two genomes. An additional five SNPs are unique to DA101 compared to 95 in Aschheim, which is provisionally consistent with Aschheim's younger age (reported in the SI and in Extended Data Fig. 9, though the latter does not present the tree at full resolution). The identified shared ancestry was interpreted as setting DA101 within the context of the "Justinian plague" (*sic*). The longer branch in the Aschheim genome is explained by its younger age and a seemingly accelerated substitution rate, which is supposedly indicative of an epidemic context.

Although the Justinianic Plague was previously thought to represent the first major onslaught of plague in humans (i.e., the First Pandemic), plentiful examples of human infections of *Y. pestis* are surfacing as far back as the Neolithic (Andrades Valtueña et al., 2017; Rascovan et al., 2019; Spyrou et al., 2018). Here, we present a reanalysis of both DA101 and DA147 genomes which does not seem to support the arguments made by Damgaard et al., 2018. Instead, the analysis of DA101 suggests it to be yet another example of a pre-Justinianic human infection. Furthermore, in contrast to its suggested archaeological dating, we show DA147 to occupy a phylogenetic position much closer to the Black Death (1346–1353 AD) than the Justinianic Plague (Fig. 1C). As such, neither genome can address the origin of the First Pandemic.

Results and Discussion

We reanalysed both presented genomes with a more extensive dataset of published modern and ancient *Y. pestis* genomes (Fig. 1A, Table S1). We opted to include the genome from Altenerding (Feldman et al., 2016) as a representative for the Justinianic Plague: though genetically identical to Aschheim (Wagner et al., 2014), its higher coverage makes it less prone to false positive SNPs that are common in metagenomic data with high environmental backgrounds. Of note, the Aschheim genome has been shown to carry a high number of false positive SNPs (Feldman et al., 2016), which might in part account for its longer branch and accelerated substitution rate observed by Damgaard et al. (see SI).

Analysis of the DA101 genome revealed a minimum of 3 SNPs shared with Altenerding and a minimum of 9 that are unique. By contrast, Altenerding has 51 unique SNPs (Table S2). This sets both nodes, i.e., that giving rise to the shared Justinianic/DA101 branch and the one separating them, deeper in time compared to what is presented in the original publication (Damgaard et al., 2018).

Further to this, we attempted a molecular dating analysis, though the age of individual DA101 proved difficult to determine given discrepancies in the text and SI, ranging from

"approximately 180 AD" (main text) to 214–261 calAD/1701 BP (Damgaard et al., 2018 Supplementary Table 2). Ultimately, we opted to use the calibrated radiocarbon interval, which yielded a mean age of 154 BC (95% HPD: 527 BC to 153 AD) for the emergence of the shared lineage and 9 BC (95% HPD: 318 BC to 221 AD) for their divergence time (Table S4). For comparison, dating results without the recently published RT5 genome (Spyrou et al., 2018) are shown in Table S4. This strongly supports a pre-Justinianic provenience for the DA101 genome. A number of shared or unique SNPs might be undetected for DA101 due to low coverage, hence the estimated divergence dates are conservative and might be even older.

Regarding the substitution rate, we do not observe a notable acceleration on the Altenerding branch (mean 2.67E-08) compared to the overall mean (1.48E-08) across the tested dataset (Fig. S2), particularly since both estimates show overlapping 95 % HPD intervals (Table S5). Nevertheless, we do observe an overdispersion of substitution rates across different *Y. pestis* lineages (described previously in Cui et al., 2013 and Spyrou et al., 2019) with the highest estimate here yielding an 17-fold deviation from the mean (2.46E-07).

Damgaard et al. do not discuss the fact that DA101 predates the onset of the Justinianic Plague by three centuries according to its radiocarbon date. This fact, however, is incompatible with their hypothesis of a 6th-century pandemic disease introduction to Europe through Hunnic expansion based on this genome alone, as argued suggestively multiple times in their work: in the abstract ("Scythians […] moved westward in about the second or third century BC, forming the Hun traditions in the fourth–fifth century AD, and carrying with them plague that was basal to the Justinian plague."), the subheader ("Origins and spread of the Justinian plague", p. 372) and the concluding sentence ("[…], we find provisional support for the hypothesis that the pandemic was brought to Europe towards the end of the Hunnic period through the Silk Road along the southern fringes of the steppes.", p. 373).

Previously published data demonstrating the absence of detectable genetic changes in *Y. pestis* and its extremely rapid movement during the Black Death in Europe (1347–1353 AD; (Namouchi et al., 2018; Spyrou et al., 2019, 2016) clearly indicate that this pathogen is able to travel vast geographic expanses quickly, without accumulating genetic diversity in the process. As such, the depth of the time interval for the coalescence of DA101 and the Justinianic genomes offers little to no evidence on the temporal or geographic origin of the Justinianic Plague (beginning in 541 AD, Fig. 1B). Since individual DA101 comes from a geographical location that today houses multiple plague foci including modern lineages 0.ANT1, 0.ANT2 and the newly described 0.ANT5 (*sensu* Eroshenko et al., 2017; Fig. 2), it may even be

surmised whether the sampled individual fell victim to an epidemic event or a to a sporadic individual infection.

The second *Y. pestis* genome from individual DA147 from North Ossetia, supposedly 6th–9th centuries, could substantiate a spread of plague along the "southern fringes of the steppe" (p. 373), although its phylogenetic placement was not investigated by Damgaard et al. Even though the coverage is low, our re-analysis of the raw sequence data from this individual and an assessment of phylogenetically informative positions reveals that it does not share any derived SNPs with Altenerding or DA101 (Table S2). None of the positions shared between Altenerding and DA101 are covered in DA147, but 2 out of the 9 unique SNPs of DA101 are covered and show the ancestral state. Of the unique Altenerding SNPs, 9 are covered in DA147 with 8 showing the ancestral state. The only SNP possibly shared with DA147 is a C>T change that is potentially caused by DNA damage, as it appears only in a single read.

Such initial results motivated a further exploration of DA147's possible phylogenetic position. For this, we used MultiVCFAnalyzer v0.85 for a comparative SNP analysis against our dataset of ancient and modern *Y. pestis* genomes (Table S1), while omitting all private calls in DA147 since their vast majority will represent DNA damage and sequencing errors due to the genome's low coverage. The remaining SNPs forming the branch of DA147 in Fig. S1 (red) are an artefact caused by homoplastic or triallelic sites. We computed a maximum likelihood phylogenetic tree that, unexpectedly, placed DA147 closest to the previously described polytomy of Branches 1–4 (Fig. S1). The genomes's placement was further investigated by visual inspection of all diagnostic SNPs separating Branches 1, 2, 3&4 and Branch 0 (see Table S3). Our analysis reveals several potential placements for DA147: (1) it is one SNP ancestral to the polytomy but derived with respect to the 0.ANT3 node, (2) it is directly on the polytomy, (3) it is one SNP ancestral to the Black Death strain (Bos et al., 2011) on Branch 1, or (4) it is one to 16 SNPs basal on Branch 2 (Fig. 1C; Table S3). The third scenario is of particular interest in the context of a recently discovered genome from Laishevo, Russia (Spyrou et al., 2019) which could be identical to DA147. Therefore, DA147 might instead offer currently unexplored insights into the origin of the Black Death.

Furthermore, this finding raises doubts about the precision in the archaeological dating of this specimen (6th–9th centuries; Damgaard et al., 2018). Unfortunately, the provenience of this genome cannot be further investigated since metadata from this individual are absent in Table S2 in Damgaard et al., 2018. Based on our molecular dating analysis, the node giving rise to 0.ANT3, which is basal to all possible placements of DA147, is dated to a mean age of 1030

AD (95% HPD: 732 AD to 1274 AD), thus placing this low coverage genome within the diversity that has accumulated within the last millennium.

Finally, we would like to correct two inaccuracies in nomenclature in the study: First, the label "0.ANT5" has already been given to a modern clade of *Y. pestis* strains reported by Eroshenko et al., 2017. In general, we recommend against applying nomenclature combining phylogenetic and metabolic features to ancient genomes (Achtman, 2016), since their metabolic profile has not yet been characterized. Second, the "Justinianic Plague" is named after the Roman emperor Justinian I (c. 482–565 AD) who reigned during the onset of this pandemic (Little et al., 2007). The term "Justinian Plague" as used by the authors is misleading, since it suggests a connection to either Justin I or Justin II of the Justinianic dynasty.

Overall, we argue that the two presented *Y. pestis* genomes cannot contribute to our understanding of the Justinianic Plague that began in 541 AD in the southeast Mediterranean basin due to their phylogenetic, temporal and geographical distance. Moreover, these genomes offer no support for a connection between the Justinianic Plague and the Hunnic expansion, or for a spread through the southern steppe, both of which are also in conflict with the leading, document-based hypothesis of a plague introduction via trade routes linking India to the Red Sea (Harper, 2017; Fig. 2). The low coverage genome might rather hold clues for the onset of the Black Death or on the origins of Branch 2. We suggest a redirected focus here, especially if higher coverage data from this or a similar archaeological sample becomes available in the future.

Materials and Methods

Sequencing data for the samples DA101 and DA147 were retrieved from ENA with the provided accession numbers (Damgaard et al., 2018) and processed with the EAGER pipeline (Peltzer et al., 2016), including Illumina adapter removal, sequencing quality filtering (minimum base quality of 20) and length filtering (minimum length of 30 bp).

For the DA101 sample with higher coverage, reads were clipped on both ends by 3 bases to remove the majority of damaged sites and subsequently filtered again for length using the same parameter. Mapping against the CO92 reference genome (chromosome NC_003143.1) was done with BWA (-l 32, -n 0.1, -q 37), reads with low mapping quality (-q 37) were removed with Samtools and duplicates were removed with MarkDuplicates.

SNP calling was performed with the UnifiedGenotyper within the Genome Analysis Toolkit (GATK) using the 'EMIT_ALL_SITES' option to generate a call for every position in the reference genome.

For the DA147 sample with low coverage, mapping was performed without prior damage clipping and with less stringent parameters in BWA (-l 16, -n 0.01, -q 37) to retrieve a maximum of coverage. Reads with low mapping quality were removed with Samtools (-q 37) and duplicates were removed with MarkDuplicates. For a phylogenetic analysis of the low coverage DA147 genome (0.24-fold), the bam-file was converted into a fastq-file using bedtools, multiplied by 5 and mapped again with identical parameters but without duplicate removal to reach the necessary coverage of positions for SNP calling. SNP calling was performed with the UnifiedGenotyper within the Genome Analysis Toolkit using 'EMIT_ALL_SITES' to generate calls for all positions in the reference genome.

For the phylogenetic analyses, we used 166 previously published modern *Y. pestis* genomes (Cui et al., 2013; Eroshenko et al., 2017; Kislichkina et al., 2015; Zhgenti et al., 2015), a *Y. pseudotuberculosis* reference genome (IP32953; Chain et al., 2004) as an outgroup and the following ancient genomes: nine genomes from Neolithic/Bronze Age contexts (Andrades Valtueña et al., 2017; Rasmussen et al., 2015; Spyrou et al., 2018), one genome of the Justinianic Plague (Altenerding; Feldman et al., 2016), one genome representing Black Death (8291-11972-8124; (Bos et al., 2011), and six genomes of the subsequent second plague pandemic (Observance OBS116, OBS137, OBS110, OBS107, OBS124; Bos et al., 2016); Bolgar, (Spyrou et al., 2016)). A complete list of all *Y. pestis* genomes used is given in Table S1. Previously identified problematic regions as well as regions annotated as repeat regions, rRNAs, tRNAs and tmRNAs were excluded for all subsequent analyses (Cui et al., 2013; Morelli et al., 2010). MultiVCFAnalyzer v0.85 (Bos et al., 2014) was used for generating a SNP table with the following settings: Minimal coverage for base call of 5 with a minimum genotyping quality of 30 for homozygous positions, minimum support of 90% for calling the dominant nucleotide in a 'heterozygous' position. The sample DA147 was processed in the outgroup mode in MultiVCFAnalyzer to remove all singletons, for the most part representing damaged sites called due to the prior multiplication of reads. The unfiltered SNP alignment produced by MultiVCFAnalyzer was used for all following analyses. Additionally, all phylogenetically informative positions were visually inspected in IGV v2.4 (Thorvaldsdóttir et al., 2013). Maximum likelihood trees were generated with RAxML v8 (Stamatakis, 2014) using the GTR substitution model based on a partial deletion (95 %) SNP alignment for DA101 (3673 SNPs) and a full SNP alignment for DA147 (3885 SNPs). Robustness of all trees was tested by the bootstrap methods using 1000 pseudo-replicates.

For the estimation of divergence times and substitution rates with BEAST 1.10 (Drummond and Rambaut, 2007), we used the coalescent Bayesian skyline model with a setup identical to that published in Spyrou et al., 2018 with the following modifications: Integration of the DA101 sample (Damgaard et al., 2018), 95% partial deletion SNP alignment and 800,000,000 states as chain length. A second run was performed without the recently published Bronze Age genome RT5 (Spyrou et al., 2018) to investigate whether the results are affected by previously unavailable data. The recently published genomes by Eroshenko et al., 2017 were not included in the dating analysis due to exceptionally long branches and unavailability of raw data to address potential mismapping. An MCC tree was produced using TreeAnnotator of BEAST v1.10, showing the relative mean substitution rates (Fig. S2). All trees were visualized in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

For Fig. 2, we used the coordinates of DA146 and DA160, since they are identical and frame the sample DA147 which is not part of the Table S2 of Damgaard et al., 2018. The first outbreaks of the Justinianic Plague come from Harper, 2017 and (Stathakopoulos, 2004). For the trade routes, we used the "Indian and Persian trade routes with the West 50 BCE - 300 CE" and "Silk Road routes 1–1400 CE" from OWTRAD (http://www.ciolek.com/owtrad.html). Important trade centres are adopted from Harper, 2017.

Author Contributions

M.K., A.H., K.I.B and J.K. planned and designed the study. M.K. performed data processing and phylogenetic analyses; M.A.S. and M.K. performed dating analyses. M.M. provided and reviewed historical context information. M.K. wrote the manuscript with contributions from M.A.S., K.I.B. and A.H. and edits from all co-authors.

253 **References**

254 Achtman M. 2016. How old are bacterial pathogens? *Proc Biol Sci* **283**:e1003471–e1003471.

255 Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. 1999. Yersinia pestis, the
256 cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. *Proc Natl*
257 *Acad Sci U S A* **96**:14043–14048.

258 Andrades Valtueña A, Mittnik A, Key FM, Haak W, Allmäe R, Belinskij A, Daubaras M,
259 Feldman M, Jankauskas R, Janković I, Massy K, Novak M, Pfrengle S, Reinhold S,
260 Šlaus M, Spyrou MA, Szécsényi-Nagy A, Tõrv M, Hansen S, Bos KI, Stockhammer
261 PW, Herbig A, Krause J. 2017. The Stone Age Plague and Its Persistence in Eurasia.
262 *Curr Biol* **27**:3683–3691.e8.

263 Benedictow OJ. 2004. The Black Death 1346–1353: The Complete History. Woodsbridge:
264 Boydell & Brewer.

265 Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM,
266 Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA,
267 Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson
268 D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J. 2014.
269 Pre-Columbian mycobacterial genomes reveal seals as a source of New World human
270 tuberculosis. *Nature* **514**:494–497.

271 Bos KI, Herbig A, Sahl J, Waglechner N, Fourment M, Forrest SA, Klunk J, Schuenemann
272 VJ, Poinar D, Kuch M, Golding GB, Dutour O, Keim P, Wagner DM, Holmes EC,
273 Krause J, Poinar HN. 2016. Eighteenth century Yersinia pestis genomes reveal the long-
274 term persistence of an historical plague focus. *Elife* **5**:e12994.

275 Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee
276 JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJD, Herring DA, Bauer P,
277 Poinar HN, Krause J. 2011. A draft genome of Yersinia pestis from victims of the Black
278 Death. *Nature* **478**:506–510.

279 Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM,
280 Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M,
281 Medigue C, Simonet M, Chenal-Francisque V, Souza B, Dacheux D, Elliott JM, Derbise
282 A, Hauser LJ, Garcia E. 2004. Insights into the evolution of Yersinia pestis through
283 whole-genome comparison with Yersinia pseudotuberculosis. *Proceedings of the*
284 *National Academy of Sciences* **101**:13826–13831.

285 Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV,

Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R. 2008. Insight into microevolution of Yersinia pestis by clustered regularly interspaced short palindromic repeats. *PLoS One* **3**:e2652.

Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, Weinert L a., Wang Z, Guo Z, Xu L, Zhang Y, Zheng H, Qin N, Xiao X, Wu M, Wang X, Zhou D, Qi Z, Du Z, Wu H, Yang X, Cao H, Wang H, Wang J, Yao S, Rakin A, Li Y, Falush D, Balloux F, Achtman M, Song Y, Wang J, Yang R. 2013. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc Natl Acad Sci U S A* **110**:577–582.

Damgaard P de B, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E, Baimukhanov N, Loman V, Hedeager L, Pedersen AG, Nielsen K, Afanasiev G, Akmatov K, Aldashev A, Alpaslan A, Baimbetov G, Bazaliiskii VI, Beisenov A, Boldbaatar B, Boldgiv B, Dorzhu C, Ellingvag S, Erdenebaatar D, Dajani R, Dmitriev E, Evdokimov V, Frei KM, Gromov A, Goryachev A, Hakonarson H, Hegay T, Khachatryan Z, Khaskhanov R, Kitov E, Kolbina A, Kubatbek T, Kukushkin A, Kukushkin I, Lau N, Margaryan A, Merkyte I, Mertz IV, Mertz VK, Mijiddorj E, Moiyesev V, Mukhtarova G, Nurmukhanbetov B, Orozbekova Z, Panyushkina I, Pieta K, Smrčka V, Shevnina I, Logvin A, Sjögren K-G, Štolcová T, Taravella AM, Tashbaeva K, Tkachev A, Tulegenov T, Voyakin D, Yepiskoposyan L, Undrakhbold S, Varfolomeev V, Weber A, Wilson Sayres MA, Kradin N, Allentoft ME, Orlando L, Nielsen R, Sikora M, Heyer E, Kristiansen K, Willerslev E. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**:369–374.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **8**. doi:10.1186/1471-2148-7-214

Eroshenko GA, Nosov NY, Krasnov YM, Oglodin YG, Kukleva LM, Guseva NP, Kuznetsov AA, Abdikarimov T, Dzhaparova AK, Kutyrev VV. 2017. Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high- mountain plague foci of Kyrgyzstan. *PLoS One* **12**:e0187230–e0187230.

Feldman M, Harbeck M, Keller M, Spyrou MA, Rott A, Trautmann B, Scholz HC, Päffgen B, Peters J, McCormick M, Bos K, Herbig A, Krause J. 2016. A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol Biol Evol* **33**:2911–2923.

Harper K. 2017. The Fate of Rome: Climate, Disease, and the End of an Empire. Princeton: Princeton University Press.

Keller M, Spyrou MA, Scheib CL, Neumann GU, Kröpelin A, Haas-Gebhard B, Päffgen B, Haberstroh J, Ribera I Lacomba A, Raynaud C, Cessford C, Durand R, Stadler P, Nägele K, Bates JS, Trautmann B, Inskip SA, Peters J, Robb JE, Kivisild T, Castex D, McCormick M, Bos KI, Harbeck M, Herbig A, Krause J. 2019. Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). *Proc Natl Acad Sci U S A* **116**:12363–12372.

Kislichkina AA, Bogun AG, Kadnikova LA, Maiskaya NV, Platonov ME, Anisimov NV, Galkina EV, Dentovskaya SV, Anisimov AP. 2015. Nineteen Whole-Genome Assemblies of Yersinia pestis subsp. microtus, Including Representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica: TABLE 1. *Genome Announc* **3**:e01342–15.

Little LK, Hays JN, Morony MG, Kennedy HN, Stathakopoulos D, Sarris P, Stoclet AJ, Kulikowski M, Maddicott J, Dooley A, Sallares R, McCormick M. 2007. Plague and the End of Antiquity: The Pandemic of 541–750. Cambridge University Press.

Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, Cui Y, Thomson NR, Jombart T, Leblois R, Lichtner P, Rahalison L, Petersen JM, Balloux F, Keim P, Wirth T, Ravel J, Yang R, Carniel E, Achtman M. 2010. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* **42**:1140–1143.

Namouchi A, Guellil M, Kersten O, Hänsch S, Ottoni C, Schmid BV, Pacciani E, Quaglia L, Vermunt M, Bauer EL, Derrick M, Jensen AØ, Kacki S, Cohn SK, Stenseth NC, Bramanti B. 2018. Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proceedings of the National Academy of Sciences* **115**:E11790-E11797.

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. 2016. EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**:60–60.

Rascovan N, Sjögren K-G, Kristiansen K, Nielsen R, Willerslev E, Desnues C, Rasmussen S. 2019. Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline Article Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. *Cell* **176**:295–305.e10.

Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren K-G, Pedersen AG, Schubert M, Van Dam A, Kapel CMO, Nielsen HB, Brunak S, Avetisyan P, Epimakhov A, Khalyapin MV, Gnuni A, Kriiska A, Lasak I, Metspalu M, Moiseyev V, Gromov A, Pokutta D, Saag L, Varul L, Yepiskoposyan L, Sicheritz-Pontén T, Foley RA, Lahr MM,

Nielsen R, Kristiansen K, Willerslev E. 2015. Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* **163**:571–582.

Sarris P. 2002. The Justinianic plague: origins and effects. *Contin Chang* **17**:169–182.

Schamiloglu U. 2016. The Plague in the Time of Justinian and Central Eurasian History: An agenda for Research In: Zimonyi I, Karatay O, editors. Wiesbaden: Harrasowitz Verlag.

Schmid BV, Büntgen U, Easterday WR, Ginzler C, Walløe L, Bramanti B, Stenseth NC. 2015. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci U S A* **112**:3020–3025.

Spyrou MA, Keller M, Tukhbatova RI, Scheib CL, Nelson EA, Andrades Valtueña A, Neumann GU, Walker D, Alterauge A, Carty N, Cessford C, Fetz H, Gourvennec M, Hartle R, Henderson M, von Heyking K, Inskip SA, Kacki S, Key FM, Knox EL, Later C, Maheshwari-Aplin P, Peters J, Robb JE, Schreiber J, Kivisild T, Castex D, Lösch S, Harbeck M, Herbig A, Bos KI, Krause J. 2019. Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. *Nat Commun* **10**:4470.

Spyrou MA, Tukhbatova RI, Feldman M, Drath J, Kacki S, Beltrán de Heredia J, Arnold S, Sitdikov AG, Castex D, Wahl J, Gazimzyanov IR, Nurgaliev DK, Herbig A, Bos KI, Krause J. 2016. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host Microbe* **19**:874–881.

Spyrou MA, Tukhbatova RI, Wang C-C, Andrades Valtueña A, Lankapalli AK, Kondrashin VV, Tsybin VA, Khokhlov A, Kühnert D, Herbig A, Bos KI, Krause J. 2018. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nat Commun* **9**:2234.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313.

Stathakopoulos D. 2004. Famine and Pestilence in the Late Roman and Early Byzantine Empire: A Systematic Survey of Subsistence Crises and Epidemics. Aldershot: Ashgate.

Sussman GD. 2011. Was the Black Death in India and China? *Bull Hist Med* **85**:319–355.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**:178–192.

Tsiamis C, Poulakou-Rebelakou E, Petridou E. 2009. The Red Sea and the Port of Clysma. A Possible Gate of Justinian's Plague. *Gesnerus* **66**:209–217.

Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, Spyrou MA,

Andrades Valtueña A, Huson D, Tuross N, Bos KI, Krause J. 2018. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology and Evolution* **2**:520–528.

Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN, Kuch M, Lumibao C, Poinar D, Pearson T, Fourment M, Golding B, Riehm JM, Earn DJD, Dewitte S, Rouillard J-M, Grupe G, Wiechmann I, Bliska JB, Keim PS, Scholz HC, Holmes EC, Poinar H. 2014. Yersinia pestis and the plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect Dis* **14**:319–326.

Zhgenti E, Johnson SL, Davenport KW, Chanturia G, Daligault HE, Chain PS, Nikolich MP. 2015. Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus Region. *Genome Announc* **3**:1030–1015.

Fig. 1: A. Maximum likelihood tree based on 3673 SNPs of 167 modern and 18 ancient genomes. Main branches are collapsed for clarity. Numbers on nodes indicate bootstrap support. Highlighted are the Justinianic genome from Altenerding (green), the investigated Tian Shan genome DA101 (blue) and the recently characterized modern strains of clade 0.ANT5 (brown). Relevant parts of the tree are highlighted in pink (Fig. 1B) and light blue (Fig. 1C). **B.** Schematic tree of the common branch of Altenerding and DA101 showing the minimum number of SNPs (italics) and divergence dates (bold, mean and 95 % HPD). **C.** Schematic tree of the highlighted part of A (light blue) for the positioning of DA147 with numbers of SNPs (italics, # of diverged SNPs in DA147/# of covered SNPs in DA147/# of total SNPs on branch) and estimated divergence date of clade 0.ANT3 (bold, mean and 95 % HPD) based on the dating analysis with BEAST 1.10 (see Materials and Methods and SI). Possible placements of DA147 in the phylogeny are highlighted in red.

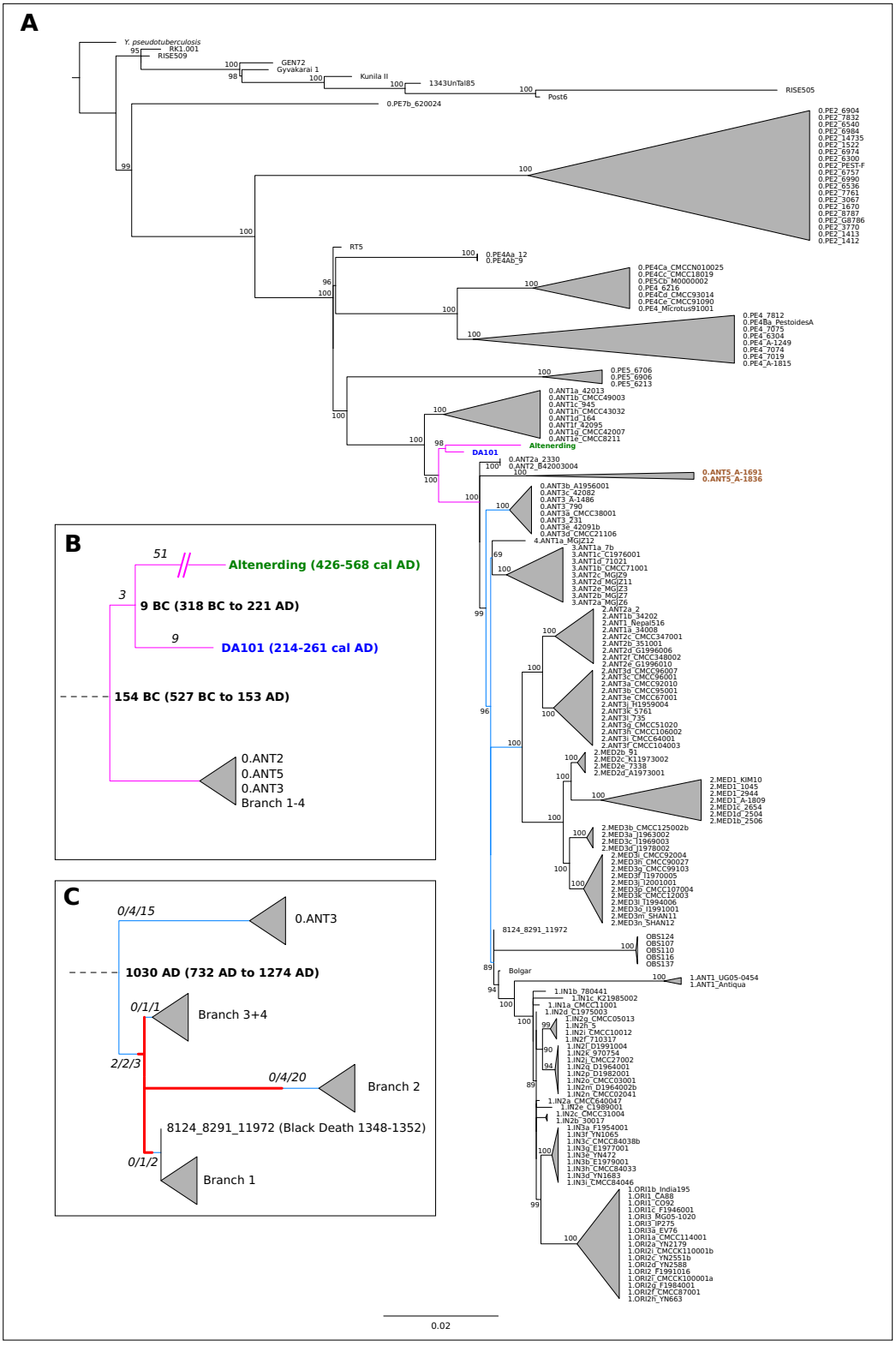
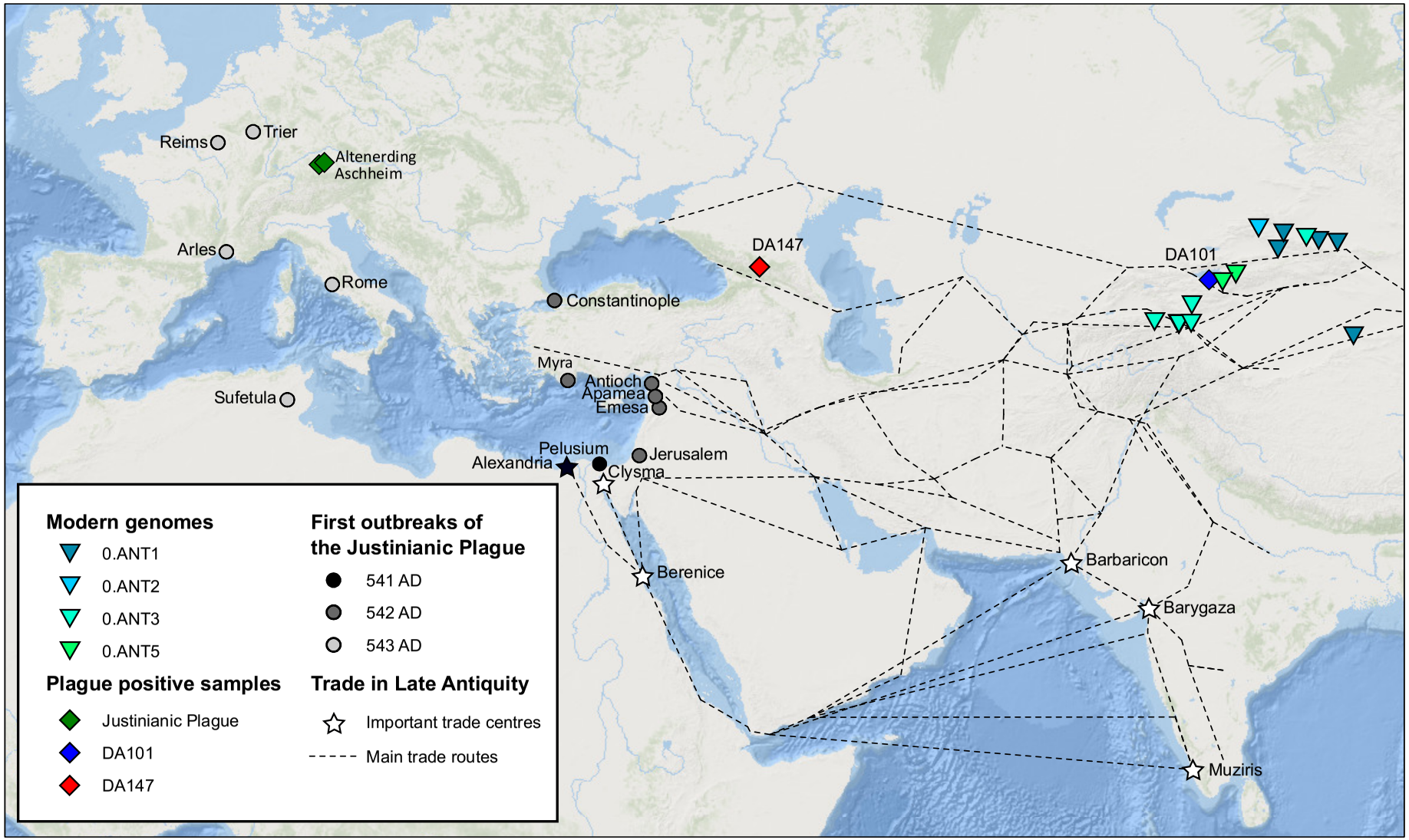


Fig. 2: Map of the origin of selected modern and ancient *Y. pestis* genomes in relation to the first outbreaks of the Justinianic Plague as well as important trade centres and routes. An introduction of plague via the Indian Ocean and the Red Sea is supported by well-established sea communications connecting Egyptian and Indian trade centres. The hypothesis of an overland transport through the steppe is in conflict with the directionality of the first outbreaks of the Justinianic Plague and lacks historical support. Although DA147 presumably dates later than 750 AD, it is shown for comparison.



7 Manuscript D: Von der Seuchengeschichte der Pest zu einer Naturgeschichte ihres Erregers – Neue Einblicke durch alte DNA

Marcel Keller

Published in *Pest! Eine Spurensuche*. Edited by LWL-Museum für Archäologie, Westfälisches Landesmuseum, Herne; Stefan Leenen; Alexander Berner; Sandra Maus. wbg Theiss (Darmstadt, 2019)

Von der Seuchengeschichte der Pest zu einer Naturgeschichte ihres Erregers

Neue Einblicke durch alte DNA

Dieser Beitrag wechselt die Pespektive: Nicht der Mensch, sondern das Bakterium steht im Zentrum der Geschichte. Über mehrere tausend Jahre mutierte ein vergleichsweise harmloses Bakterium zum Pesterreger *Yersinia pestis* wie wir ihn heute kennen. Was musste eigentlich geschehen, um aus einem Keim einen Killer zu machen, der die Welt in Atem hält? Und was verrät der genetische Stammbaum des Bakteriums über die Verbreitungswege der Seuche? Die Antwort auf diese Fragen liefert die Archäogenetik.

Marcel Keller

Anfang Oktober 2017 verbreitete sich die Meldung eines Pestausbruchs in Madagaskar wie ein Lauffeuer durch deutschsprachige wie internationale Medien.[1] Zunächst war von zwei Dutzend Toten die Rede, jedoch wuchs die Zahl der Opfer in den nächsten Wochen auf über 200. In Schlagzeilen wie »Die Pest ist zurück« stützten zahlreiche Zeitungen den Eindruck der breiten Öffentlichkeit, es handle sich um die Wiederkehr einer »mittelalterlichen« Seuche.[2] Und tatsächlich hat sich die Pestpandemie von 1346 bis 1353 als Schwarzer Tod tief in das kollektive Gedächtnis Europas eingeprägt und gilt als der Inbegriff einer Seuche. Dies verdeckt allerdings den Blick auf die Tatsache, dass die Pest auch in heutiger Zeit jedes Jahr weltweit für einige Todesfälle verantwortlich ist, dass Ausbrüche wie in Madagaskar gar Ausläufer einer Pestpandemie sind, die im ausgehenden 19. Jahrhundert und dem, beginnenden 20. Jahrhundert um den gesamten Globus wanderte.

Dass wir historische Epidemien mit einem modernen Erreger in Verbindung bringen können, ist allerdings alles andere als selbstverständlich, sondern eher die Ausnahme: Von den biblischen Plagen bis hin zu neuzeitlichen Epidemien wie dem »Englischen Schweiß« kann oft nur spekuliert werden, was sich hinter den überlieferten Berichten verbirgt. Einen fundamentalen Beitrag zur Identifizierung des Erregers historischer Pestausbrüche leistete dabei die Analyse »alter DNA«.[3] Mit der Anwendung modernster molekularbiologischer Methoden auf archäologisches Material ermöglicht die Paläogenetik aber viel mehr als nur die Identifizierung von (prä-)historischen Krankheitserregern. Durch sie können wir heute eine Naturgeschichte des Pestbakteriums von der Jungsteinzeit bis in unsere Zeit erzählen und daran exemplarisch die Evolution eines Pathogens von einem relativ harmlosen Darmkeim zu einem gefährlichen Killer nachvollziehen.

Kurzer Abriss zur Geschichte der Pestforschung

Eine nach heutigen Maßstäben wissenschaftliche Auseinandersetzung mit der Pest in Geschichte und Gegenwart reicht bis ins frühe 19. Jahrhundert zurück, im 20. Jahrhundert schließlich explodiert die biologisch-medizinische Literatur im Angesicht der Dritten Pandemie geradezu. Eine auch nur ansatzweise vollständige Forschungsgeschichte kann und soll deshalb gar nicht angestrebt werden. Jedoch sollen wesentliche Kontinuitätslinien, aber auch manche Sackgassen der Pestforschung aufgezeigt werden, die schließlich zu den paläogenetischen Studien der letzten zwei Jahrzehnte führten und diese somit in den Diskursen der Seuchengeschichte verorten.

Es ist wohl das Verdienst des Arztes Justus Hecker, neben der Medizingeschichte mit der *historischen Pathologie* eine Geschichte der Krankheiten als eigenständige Disziplin zu etablieren, etwa mit seiner Publikation *Der schwarze Tod im vierzehnten Jahrhundert* aus dem Jahr 1832.[4] Während schon im ausgehenden 18. Jahrhundert Pathologen medizinische Abhandlungen zur Geschichte einzelner Krankheiten veröffentlicht hatten, formuliert Hecker im Angesicht der Cholera-Epidemie von 1831 einen neuen Anspruch dieser Disziplin, durch den Blick in die Geschichte auch die zeitgenössischen Ausbrüche besser verstehen zu wollen. Dies ist insofern bemerkenswert, als die medizinische Mikrobiologie und damit der Nachweis von Mikroorganismen als Ursache für Infektionskrankheiten erst in der zweiten Hälfte des 19. Jahrhunderts durch Louis Pasteur und Robert Koch begründet wurde. Dass Hecker neben anderen Krankheiten ausgerechnet die Pest aufgegriffen hat, ist dabei sicherlich mehreren Faktoren geschuldet. Zum einen lag die letzte große europäische Pestepidemie nur etwa 60 Jahre zurück (1771 in Moskau[5]), zum anderen erlaubten die charakteristischen Symptome der Beulenpest auch eine vergleichsweise sichere

Identifizierung in historischen Quellen. Nicht zuletzt hat sich die Pest auch tief in die europäische Kulturgeschichte eingeschrieben.

Mit dem dritten Band seines *Lehrbuch*[s] *der Geschichte der Medizin und der epidemischen Krankheiten* legte Heinrich Haeser im Jahr 1845 erstmals eine umfassende Seuchengeschichte von der attischen Seuche im 4. Jahrhundert v. Chr. bis zu den zeitgenössischen Influenzaepidemien vor.[6] Auch wenn Haeser den Ausdruck »Pest« hier noch allgemein für Infektionskrankheiten benutzt, versucht er sich an einer Differentialdiagnose, um die »Bubonenpest« von anderen Infektionskrankheiten abzugrenzen. Haeser identifiziert die erste eindeutige Beschreibung der Beulenpest bei Rufus von Ephesos im 1. Jahrhundert n. Chr.,[7], die auch moderner philologischer Textkritik standhält, und ordnet auch die Pest des Justinian im 6. Jahrhundert und den Schwarzen Tod des 14. Jahrhunderts mit der »Bubonenpest« einer gemeinsamen Ursache zu.[8] Auch für die folgenden Jahrhunderte identifiziert er zahlreiche Pestausbrüche bis ins Jahr 1874 im Nahen Osten.[9] Hingegen hat ihn die Nachricht über eine neue Pestepidemie in der Provinz Yunnan/China 1855 wohl nicht erreicht.

Nachdem sich die Pest in den folgenden Jahren bis an die südchinesische Küste ausbreitete und im Jahr 1894 schließlich die britische Kolonie Hongkong erreichte, wurde auch Europa auf die Epidemie aufmerksam. In der Zwischenzeit feierten Louis Pasteur und Robert Koch große Erfolge durch die Entdeckung von Bakterien als Krankheitserreger, und so lieferten sich deren Labore ein Wettrennen um die Beschreibung neuer Bakterien. Es gelang schließlich dem Pasteur-Schüler Alexandre Yersin, 1894 den Erreger der sogenannten Hongkong-Pest zu isolieren,[10] womit er die eigentliche biologisch-medizinische Erforschung der Pest begründete. Aber bereits in seiner Erstbeschreibung des Erregers *Yersinia pestis*, welcher später nach ihm benannt werden sollte, stellte Yersin aufgrund des Symptombildes die Verbindung zu historischen Ausbrüchen her, ausdrücklich zur Pest von Marseille 1720–1722.[11] Außerdem beobachtete er, dass den Pestausbrüchen häufig ein Massensterben von Hausratten (*Rattus rattus*) vorausging. Nur vier Jahre später gelang es Yersins Kollegen Paul-Louis Simond, den Infektionsweg über den Rattenfloh (*Xenopsylla cheopis*) nachzuweisen.[12] 1914 beschrieben schließlich Bacot und Martin das seither klassische Paradigma der Pest-Übertragung: Nachdem ein Floh das Blut eines infizierten Nagetiers aufgesogen hat, bildet *Yersinia pestis* im Vormagen des Flohs einen Biofilm aus, der zur Blockade weiterer Nahrungsaufnahme führt.[13] Der so ausgehungerte Floh verlässt das eventuell bereits verendete Tier auf der Suche nach einem neuen Wirt. Hat er diesen gefunden, kommt es mehrfach zu Stichen, wobei der Floh aufgrund der Blockade das aufgenommene und mit *Yersinia pestis* kontaminierte Blut wieder in den Wirt erbricht.

Während der Dritten Pandemie (in Abfolge mit der Ersten Pandemie, beginnend mit der Justinianischen Pest, und der Zweiten Pandemie), die in Yunnan ihren Anfang nahm, verbreitete sich die Pest schließlich weltweit und erreichte vermutlich erstmals Nord- und Südamerika. In der Folge gründeten sich mehrere Pest-Kommissionen zur Erforschung und Bekämpfung der Pest, welche nach der Gründung der World Health Organization auch von dieser fortgeführt wurde.[14]

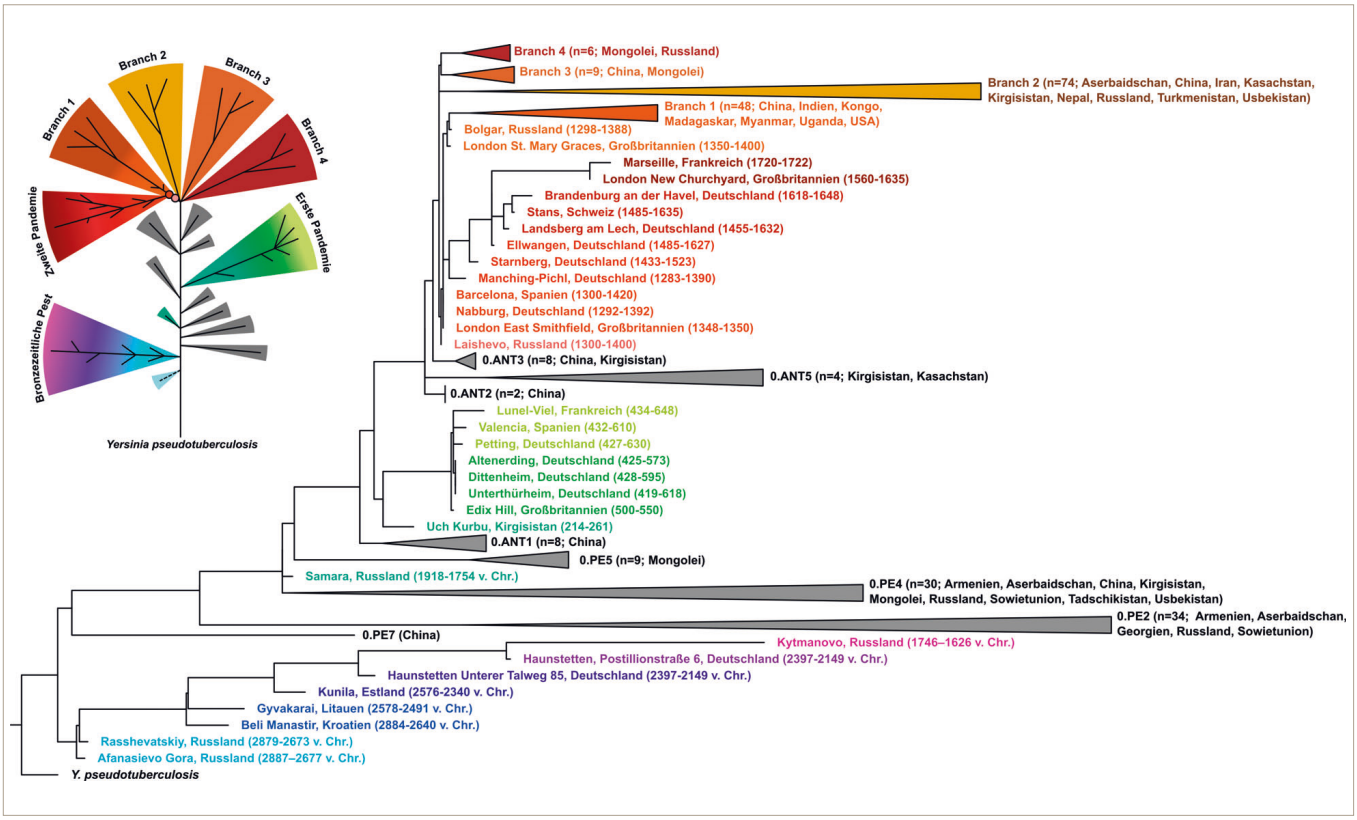
Dabei wurden zahlreiche Erkenntnisse sowohl zur Ökologie als auch der Pathogenese der Pest erlangt, so etwa dass nicht Ratten, als »Kulturfolger« eng mit dem Menschen assoziiert, sondern wilde Nager wie Murmeltiere oder Erdhörnchen die natürlichen Reservoire der Pest darstellen.[15] Infolge der Dritten Pandemie etablierten sich solche Reservoire auch etwa in Präriehund-Populationen in den USA, wo es seither zu sporadischen Fällen kommt.[16]

Neben der Beulenpest, bei der das Bakterium den dem Eintrittspunkt nächstgelegenen Lymphknoten befällt und die charakteristische Schwellung (lateinisch: *bubo*) verursacht, wurden auch weitere Krankheitsverläufe näher untersucht: Bei der septikämischen Pest gelangt das Bakterium in die Blutbahn und vermehrt sich dort, was einen fulminanten Verlauf mit einer Sterblichkeit bis zu 100 % führt. Ähnlich verhält es sich mit der Lungenpest, wobei durch die Infektion der Lunge die gefürchtete direkte Übertragung von Mensch zu Mensch möglich wird.[17]

Ein anhaltendes Interesse an der Pestforschung auch in Industrienationen ohne epidemische Pestfälle begründet sich schließlich im Potenzial von *Yersinia pestis* als Biowaffe, was durch die japanische Armee während des Zweiten Weltkriegs an Zivilisten und Kriegsgefangenen experimentell erprobt wurde.[18]

Durch die moderne mikrobiologische Forschung boten sich im 20. Jahrhundert auch für die Seuchengeschichte neue Ansätze. So versuchte Devignat 1951 aufgrund biochemischer Eigenschaften und geographischer Verteilung verschiedene Varietäten von *Yersinia pestis* zu klassifizieren, die er mit historischen Pestwellen in Verbindung brachte: Demnach waren *Antiqua*-Stämme für die Justinianische Pest verantwortlich, *Medievalis*-Stämme für den Schwarzen Tod, und *Orientalis* ordnete er der Dritten Pandemie zu.[19] Auch wenn sich diese Einteilung später als unzureichend herausstellen sollte, hat sich die Nomenklatur bis heute gehalten.[20]

Einen Durchbruch stellte die Veröffentlichung des ersten *Yersinia pestis*-Genoms durch Parkhill u. a. 2001 dar.[21] Waren davor schon erste »Stammbäume« der



1 Phylogenetischer Stammbaum moderner und alter Genome des Pesterregers *Yersinia pestis* (Nr. 114). Eine schematische Skizze des Baums ist links oben abgebildet, der Schwarze Tod und der in Laishevo gefundene direkte Vorgänger sind als Kreise hervorgehoben. Kladen mit modernen Genomen sind als Dreiecke zusammengefasst, die Anzahl und Herkunft der Genome ist in den Klammern angegeben. Die modernen Branches 1–4 sind in Brauntönen hervorgehoben, moderne Kladen entlang von Branch 0 sind grau dargestellt. Mit Hilfe alter DNA rekonstruierte Genome sind farblich hervorgehoben und mit Fundort und archäologischer bzw. Radiokarbon-Datierung beschriftet. Nicht dargestellt sind das älteste Pestgenom aus Gökhem (Schweden, gestrichelte Linie im schematischen Baum) sowie die Genome des Schwarzen Todes aus Abbadia San Salvatore (Italien), Oslo (Norwegen) und Saint-Laurent-de-la-Cabrerisse (Frankreich) und das *Pestis-secunda*-Genom aus Bergen op Zoom (Niederlande). Der Baum wurde nach der Maximum-Parsimony-Methode basierend auf 5403 SNPs erstellt, die in 98 % der Genome abgedeckt waren.

Pest auf der Basis kurzer DNA-Fragmente erstellt worden, erlaubte der Vergleich kompletter Genome in den folgenden Jahren ungleich detailliertere Analysen. Innerhalb der nächsten Jahre führte die technische Weiterentwicklung im Bereich der DNA-Sequenzierung zu einem rasanten Anstieg der Zahl publizierter moderner Genome, sodass uns heute bereits über 200 zur Verfügung stehen.[22]

Aber auch von Seiten der »klassischen« Seuchengeschichte gab es im 20. Jahrhundert wichtige Impulse, beispielsweise durch Birabens *L'homme et la peste en moyen age* 1976.[23] Interessanterweise hatte die moderne Pestforschung aber Zweifel über die Ursachen der historischen Pestwellen provoziert, insbesondere auf Basis der Epidemiologie. 1971 veröffentlichte Shrewsbury eine Monographie, in der er *Yersinia pestis* nur eine untergeordnete Rolle zuweist.[24] Während Cohn 2002 nur nachgewiesen haben wollte, dass der Schwarze Tod mit Sicherheit nicht durch *Yersinia pestis* verursacht wurde[25], machten andere Biologen und Historiker Alternativvorschläge: Twigg schlug 1984 den Anthraxerreger *Bacillus anthracis* vor,[26] Scott und

Duncan propagierten 2001 ein bisher unbeschriebenes Virus ähnlich dem Ebola- und Marburgvirus.[27]
In einer ungeheuren Fleißarbeit hat der Historiker Ole J. Benedictow, der 2004 bereits eine umfassende Geschichte des Schwarzen Todes vorgelegt hatte[28], schließlich 2010 alle Alternativtheorien systematisch geprüft und zu widerlegen versucht.[29] Dabei widmete Benedictow auch ein kurzes Kapitel den paläogenetischen Studien, die in den Jahren zuvor behaupteten, DNA-Spuren des Erregers in Skeletten aus historischen Pestgräbern nachgewiesen zu haben.[30] 2011 wurde von Bos u. a. schließlich das erste vollständige *Yersinia pestis*-Genom aus alter DNA rekonstruiert, welche aus Skeletten eines Londoner Pestgrabs des Schwarzen Todes gewonnen wurde.[31] Dies räumte nicht nur endgültig letzte Zweifel an *Yersinia pestis* als Verursacher historischer Epidemien aus, es öffnete auch völlig neue Perspektiven auf die Geschichte des Erregers.

Die Biologie und Genetik des Pesterregers *Yersinia pestis*

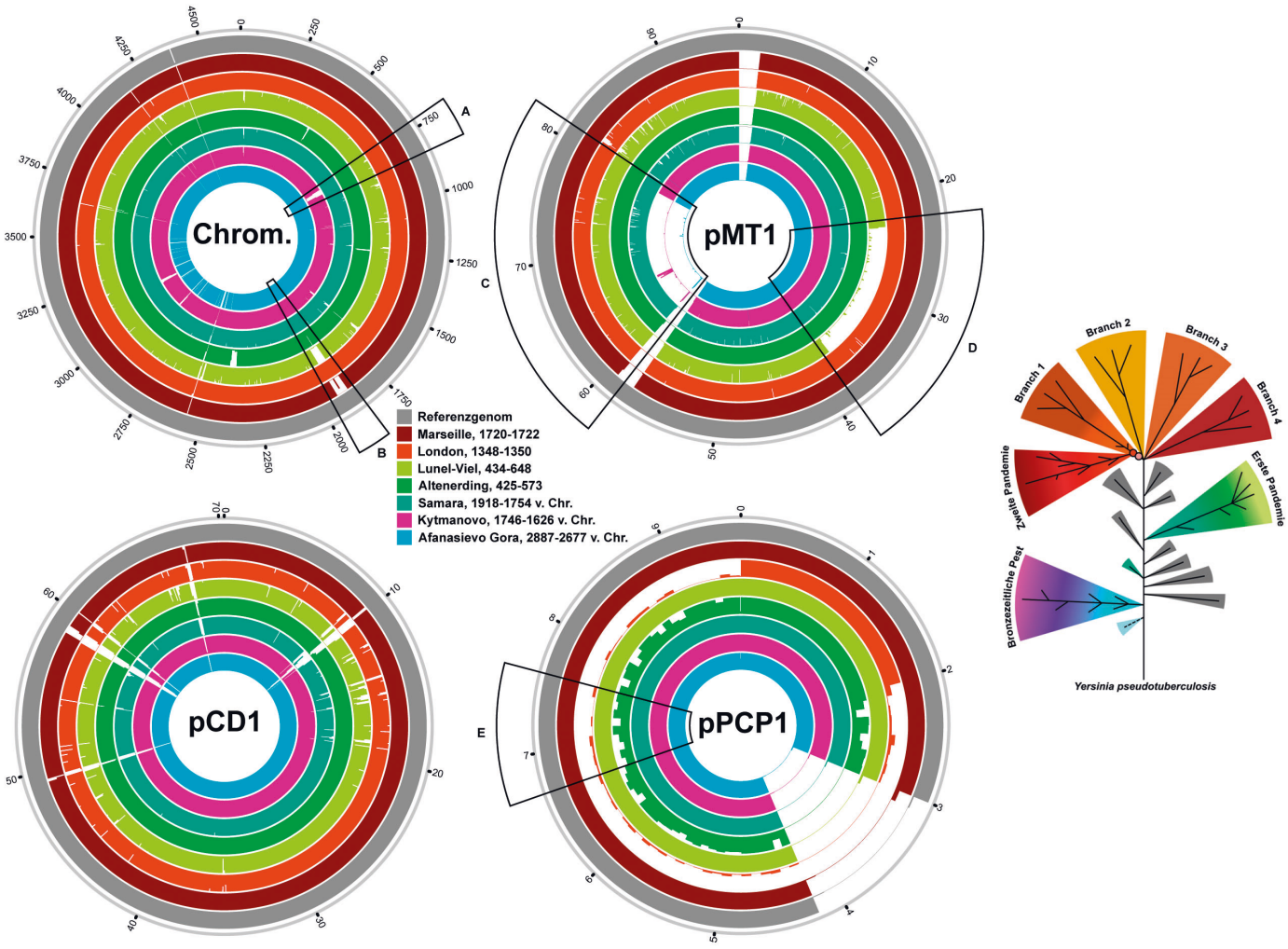
Die Gattung *Yersinia* in der Familie der Enterobakterien umfasst derzeit über ein Dutzend Arten, wovon die meisten Arten frei in der Umwelt vorkommen und nur gelegentlich aus erkrankten Tieren und Menschen isoliert werden. Als Krankheitserreger sind dabei drei Arten für den Menschen relevant: *Yersinia enterocolitica*, *Yersinia pseudotuberculosis* und *Yersinia pestis*.[32] Die ersten beiden verursachen beim Menschen v. a. Darmerkrankungen mit Durchfall und Fieber, jedoch sind diese meist selbstlimitierend, sodass keine Behandlung mit Antibiotika notwendig ist. Interessanterweise stellt die Spezies *Yersinia pestis* eigentlich eine Untergruppe von *Yersinia pseudotuberculosis* dar, vermutlich geht die gesamte Diversität von *Yersinia pestis* sogar auf einen einzigen »Klon« zurück.[33] Um zu verstehen, wie aus einem vergleichsweise harmlosen Darmbakterium ein höchst gefährlicher Erreger mit einer so besonderen Ökologie entstanden ist, lohnt sich ein Blick in das Erbgut des Pesterregers.
Das Genom von *Yersinia pestis* besteht aus einem ringförmigen Chromosom und drei deutlich kleineren, ebenfalls ringförmigen Plasmiden. Die über 4,6 Millionen Basenpaare (bp) des Chromosoms enthalten die essenziellen »Haushaltsgene« des Bakteriums, die Plasmide dagegen sind im Wesentlichen für die Pathogenität verantwortlich. Nur das Plasmid pCD1 mit etwa 70.000 bp ist auch Teil der Genome von *Yersinia enterocolitica* und *Yersinia pseudotuberculosis* und kodiert etwa für ein Sekretionssystem, das es den Bakterien erlaubt, Proteine in Zellen des Wirts zu injizieren um so eine Immunreaktion zu verhindern.
Die zwei anderen Plasmide, pMT1 (ca. 102.000 bp) und pPCP1 (ca. 9.600 bp) sind spezifisch für *Yersinia pestis* und wurden vermutlich von anderen Bakterienarten (etwa Salmonellen) im Laufe der Evolution aufgenommen. Sie kodieren für einige Gene, die für die besondere Ökologie nötig sind: Das Gen *ymt* auf dem pMT1-Plasmid ist etwa für das Überleben des Bakteriums im Vormagen des Flohs essenziell,[34] das Gen *pla* des pPCP1-Plasmids ist für den subkutanen Infektionsweg durch den Flohbiss wichtig.[35] Während das Chromosom in einer Bakterienzelle nur in einfacher Version vorliegt, können sich Plasmide selbst replizieren, für das kleinste Plasmid, pPCP1, wird etwa eine Kopienanzahl von ca. 200 pro Zelle angenommen.[36]
Die Diversität heutiger Yersinia *pestis*-Stämme wird klassisch in fünf Branches (englisch: Äste) geteilt, wobei der Branch 0 den Stamm bzw. die Wurzel zu *Yersinia pseudotuberculosis* darstellt. Bereits die basalen Stämme, die von diesem Branch abzweigen, wurden teilweise von an Pest erkrankten Menschen isoliert, nur einzelne, aus wilden Nagetieren isolierte Stämme scheinen später ihre Virulenz verloren zu haben.[37] Alle diese Stämme wurden, soweit nachvollziehbar, in Zentral- oder Ostasien isoliert, weshalb Asien als »Heimat« des Pesterregers vermutet wird. Der Branch 0 führt schließlich zu einem Knoten, von dem die Branches 1–4 abzweigen, wobei sich die Branches 3 und 4 einen kurzen Ast teilen. Die Stämme der Branches 2–4 stammen ebenfalls alle aus Asien, einige der Stämme von Branch 1 hingegen wurden auch in Afrika und in den USA isoliert, wohin diese Stämme im Zuge der Dritten Pandemie Ende des 19./Anfang des 20. Jahrhunderts gelangten.[38]

Die Molekulare Paläopathologie

Die Paläopathologie als Teildisziplin der biologischen Anthropologie beschäftigt sich mit der Krankheitsbelastung des Menschen in der Vergangenheit, hauptsächlich durch Analyse von menschlichen Überresten wie Skeletten oder Mumien.[39] Da sich das Skelett in ständigem Umbau befindet und auf äußere Einflüsse reagiert, lassen sich beispielsweise Mangelerkrankungen wie Skorbut, degenerative Erkrankungen wie Osteoporose, aber auch einige chronisch verlaufende Infektionskrankheiten wie Tuberkulose oder Lepra diagnostizieren. Einige Skelettveränderungen sind allerdings unspezifisch und lassen damit keine genauere Diagnose zu, ebenso entziehen sich rapide verlaufende Infektionskrankheiten dem makroskopischen Nachweis.
Als 1984 Higuchi u. a. erstmals DNA-Fragmente aus einer Gewebeprobe einer ausgestorbenen Zebra-Art sequenzierten und 1985 Pääbo kurz darauf gleiches

für eine ägyptische Mumie gelang, war mit der Paläogenetik eine neue Disziplin geboren.[40] Innerhalb weniger Jahre feierte die Paläogenetik spektakuläre Erfolge, und auch der Paläopathologie eröffneten die neuen Methoden ganz neue Möglichkeiten. 1998

2 Genomische Abdeckung des Chromosoms sowie der drei Plasmide pMT1, pCD1 und pPCP1 von *Yersinia pestis* durch ausgewählte alte Genome der Bronzezeit, der Ersten und der Zweiten Pandemie. Der äußerste Ring gibt die Position in 1000 bp (Basenpaaren) an. Die Abdeckung ist als ringförmiges Balkendiagramm dargestellt, jeder Balken gibt die durchschnittliche Abdeckung von 0 (nicht abgedeckt) bis ≥ 5 (voll abgedeckt) für einen Abschnitt von 5000 bp (Chromosom) bzw. 100 bp (Plasmide) an. A: Region mit Flagellin-Genen, die im jüngsten bronzezeitlichen Genom verloren gingen (violett), da sie für den nichtbeweglichen Lebensstil von Yersinia *pestis* nicht mehr benötigt werden. B: Region mit zwei Virulenzgenen, die zweimal unabhängig im Laufe der Ersten Pandemie (grün, hellgrün) und Zweiten Pandemie (rot, dunkelrot) verloren ging. C: Region des pMT1-Plasmids, die das *ymt*-Gen enthält. Während das für die Flohübertragung wichtige Gen in den beiden verwandten bronzezeitlichen Genomen fehlt (blau, violett), besitzt das spätbronzezeitliche Genom aus Samara das Gen bereits (türkis, siehe Abb. 1). D: Region mit Genen unbekannter Funktion, die im jüngsten Genom der ersten Pandemie fehlt. E: Position des *Yersinia pestis*-spezifischen *pla*-Gens, wonach bei dem PCR-Screening gesucht wird. Die geringe Abdeckung des pPCP1-Plasmids durch die Probe aus London (rot) ist ein technisches Artefakt, da die DNA dieser Probe anders angereichert wurde. Ebenso wurde die Region 3000–4200 bp aus technischen Gründen entfernt.



schließlich publizierten Drancourt u. a. den ersten Nachweis sogenannter »alter DNA« des Pesterregers aus der Pulpahöhle von Zähnen aus Massengräbern des 16. und 18. Jahrhunderts.[41] Da die Zahnpulpa neben dem Zahnnerv auch zahlreiche Blutgefäße enthält, gelangen während einer Sepsis Bakterien in hoher Zahl in die Pulpahöhle. Gleichzeitig schützt der Zahnschmelz das Innere vor äußeren Einflüssen.

Durch die seinerzeit neu entwickelte Polymerasen-Kettenreaktion (PCR) – welche Enzyme verwendet, die auch während der Zellteilung das Erbgut kopieren – war es nun möglich, kurze DNA-Fragmente zu vervielfältigen, was sowohl den Nachweis geringster Spuren als auch das anschließende Sequenzieren vereinfachte. Dabei wird sich zunutze gemacht, dass DNA aus zwei komplementären Basensträngen besteht, welche durch die Basenpaarung A–T und G–C eine Doppelhelix formt. Für die PCR werden kleine DNA-Fragmente designt, die komplementär zu Abschnitten rechts und links der gesuchten Region sind. Diese können an diese Region binden und werden entlang des vorliegenden Strangs in komplementärer Basenabfolge verlängert.[42] Durch zyklische Wiederholung dieser Reaktion können DNA-Fragmente dann nahezu exponentiell vervielfältigt werden.

Innerhalb weniger Jahre geriet die noch junge Disziplin der Paläogenetik jedoch in eine schwere Krise: Einige der veröffentlichten Ergebnisse stellten sich als Kontaminationen heraus, andere Resultate konnten nicht repliziert werden.[43] Und auch die Pestnachweise der frühen Jahre, die mit Hilfe der PCR versuchten, eine Region des *pla*-Gens zu vervielfältigen, ließen sich teilweise nicht reproduzieren.[44] Dies führte schließlich zur Aufstellung ganzer Kataloge von Maßnahmen und Kriterien, um die Authentizität der paläogenetischen Ergebnisse sicherzustellen.[45] Heute arbeiten Paläogenetiker in speziellen Reinräumen, die durch Überdruck, Luftfilter, Schutzkleidung und eine Reihe von Dekontaminationsmaßnahmen den Eintrag moderner DNA verhindern sollen. Auch findet eine strikte Trennung der Arbeitsbereiche statt, in denen alte DNA gewonnen und vervielfältigt wird, um eine Kontamination der Proben untereinander zu verhindern. Schließlich haben wir heute auch eine realistischere Einschätzung der DNA-Erhaltung in archäologischen Proben. Unmittelbar nach dem Tod eines Lebewesens setzt nämlich ein chemischer Zerfall der DNA ein, der durch mikrobiellen Befall und physikalische Faktoren noch verstärkt wird. Nach einigen Jahrhunderten beträgt schließlich die durchschnittliche Fragmentlänge unter 100 bp. Zusätzlich kommt es zu chemischen Modifikationen bevorzugt an den Enden der Fragmente, die ein charakteristisches Schadensmuster ergeben, welches nach der Sequenzierung sichtbar wird. So sehr diese DNA-Degradation auch die Möglichkeiten der Paläogenetik limitieren, bietet sie andererseits auch eine Chance zur Authentifizierung alter DNA.[46]

Der nächste große Durchbruch für die Paläogenetik stellte die Anwendung der Hochdurchsatz-Sequenzierung (auch *Next-Generation-Sequencing*, NGS) dar.[47] Zuvor mussten in relativ aufwendiger Laborarbeit einzelne kurze DNA-Fragmente gezielt vervielfältigt und sequenziert werden. Durch die neue Methode ist es heute aber möglich, sämtliche DNA-Fragmente einer Probe gleichzeitig zu sequenzieren. Dies stellt nicht nur eine enorme Zeitersparnis dar, die Methode beschränkt sich auch nicht mehr auf einzelne kleine Regionen, sondern macht die gesamte gewonnene DNA zugänglich. Dadurch erlaubt das NGS heute das Sequenzieren ganzer bakterieller Genome in einigen Stunden. Die Paläogenetik stellte die neue Methode aber auch vor neue Herausforderungen. Durch die Lagerung der archäologischen Skelettreste im Erdreich sind diese reich an DNA aus der Umwelt, insbesondere von Bodenbakterien. Die über die Jahrhunderte stark fragmentierte DNA des Menschen bzw. des ihn befallenden Krankheitserregers hingegen stellt oft nur einen geringen Bruchteil der gesamten DNA dar. Deshalb wurden sogenannte Capture-Methoden entwickelt, womit die gesuchte DNA angereichert wird, um anschließend Sequenzierkosten zu sparen.[48] Dabei macht man sich wie bei der PCR zunutze, dass einzelne DNA-Stränge mit komplementärer Basenfolge eine starke Doppelhelix-Struktur bilden. Für ein Capture erstellt man nun synthetische DNA-Fragmente des gesuchten Organismus, die fest an eine Glasplatte oder kleine Kügelchen gebunden werden. Diese können die komplementären DNA-Fragmente aus der Probe binden. Die nicht-gebundene DNA wird herausgewaschen und die gebundene DNA gelöst, welche schließlich in erheblich höherer Konzentration vorliegt.

Mithilfe der Bioinformatik können anschließend die DNA-Fragmente des gesuchten Organismus herausgefiltert und an eine Referenzsequenz angelegt werden. Auf diese Weise lassen sich vollständige Genome rekonstruieren, die durch Abgleich mit modernen Genomen in einen Stammbaum gesetzt und für weitere Analysen verwendet werden können.

Allerdings wird für das Screening größerer Probenmengen, etwa eines Massengrabes mit dutzenden Individuen, häufig noch eine quantitative PCR für ein Fragment des *pla*-Gens eingesetzt.[49] Bei dieser Methode wird durch einen Fluoreszenzfarbstoff die DNA-Menge eines Reaktionsansatzes bestimmt, welche bei erfolgreicher Vervielfältigung eines Fragments beinahe exponentiell zunimmt. Dadurch kann nicht nur die Anwesenheit der gesuchten DNA, sondern auch deren ungefähre Menge nachgewiesen werden. Die Methode

ist ungleich schneller und günstiger als das direkte Sequenzieren und hat eine relativ geringe Fehlerquote, somit eignet sie sich sehr gut, um aus einer größeren Menge an Proben die besten Kandidaten zu bestimmen. Seit 2011 wurden zahlreiche weitere Pest-Genome veröffentlicht, die vom 18. Jahrhundert bis in die Jungsteinzeit zurückdatieren und von Fundorten vom zentralasiatischen Altai-Gebirge bis nach Spanien reichen.[50] Es ist damit zu rechnen, dass von der Verfassung dieses Beitrags bis zur Veröffentlichung dieses Katalogs noch weitere Pestgenome veröffentlicht werden, die in den folgenden Betrachtungen natürlich nicht berücksichtigt werden können.

Wie aber wurden diese Pestgräber entdeckt? Zwar sind in Europa zahlreiche Pestfriedhöfe aus der frühen Neuzeit bekannt, insbesondere wenn sie außerhalb der Städte angelegt und möglicherweise noch durch Kapellen markiert wurden. Allerdings werden diese aus dem gleichen Grund in der Regel nicht von Archäologen ausgegraben, sondern als Bodendenkmäler erhalten. Hingegen werden bei Bauarbeiten, besonders im urbanen Kontext, regelmäßig Teile von Friedhöfen entdeckt, die im Rahmen einer Rettungsgrabung gesichert werden müssen. Grundsätzlich ist jede Bestattung aus einer relevanten Periode ein potenzielles Pestgrab, sofern nicht etwa schwere Traumata am Skelett auf eine andere Todesursache hindeuten. Die Wahrscheinlichkeit eines positiven Tests auf Pest-DNA ist aber natürlich bei Mehrfachbestattungen und Massengräbern höher, die bereits auf eine höhere Sterblichkeitsrate hindeuten. Selbst hierbei ist die Rate der Negativergebnisse relativ hoch, sodass hinter jedem Pestgenom oft Dutzende von Negativproben stehen.

Der geographische Fokus auf einzelne Regionen kommt dabei in der Regel durch besondere Fragestellungen oder durch Kooperationen zustande, wie etwa zwischen dem Max-Planck-Institut für Menschheitsgeschichte in Jena und der Staatssammlung für Anthropologie und Paläoanatomie in München, weshalb Bayern im Moment die Region mit den meisten alten Pestgenomen sein dürfte.[51] Dies birgt natürlich aber auch Gefahren für die Interpretation der Daten, wenn diese Verzerrung, also der willkürliche Fokus auf einzelne Regionen, nicht berücksichtigt wird. Die Diskrepanz zwischen der fundamentalen Rolle des Mittelmeers bei der Ausbreitung der ersten beiden historischen Pandemien und den vergleichsweise wenigen Genomen aus dieser Region lässt sich schließlich auch durch die schlechtere DNA-Erhaltung erklären, die mit klimatischen Parametern korreliert.

Andererseits handelt es sich etwa bei den vorgeschichtlichen Pestgenomen durchgehend um Zufallsfunde.[52] Bei der Sequenzierung von Hunderten von menschlichen Proben für populationsgenetische Fragestellungen, etwa um Migrationsbewegungen in der Bronzezeit nachzuvollziehen, wurden große Datenmengen durch die Sequenzierung des »Hintergrunds« produziert, also Umwelt-DNA und gelegentlich auch DNA von Krankheitserregern. Mithilfe der Bioinformatik wurden diese Sequenzdaten auf *Yersinia pestis* durchforstet, was überraschenderweise einige positive Proben zutage förderte.

Der Schwarze Tod und die Zweite Pandemie

Nachdem 2010 eine PCR-Studie schon glaubhaft machen konnte, dass sich authentische DNA des Pesterregers aus archäologischem Skelettmaterial gewinnen lässt,[53] wurden schließlich 2011 durch Bos u. a. die ersten alten Genome des Pesterregers aus Londoner Massengräbern gewonnen.[54] Bei dem einen handelte es sich um den Fundort »East Smithfield«, welcher aufgrund historischer Aufzeichnungen dem Schwarzen Tod zugeordnet wurde und zuvor bereits anthropologisch bearbeitet worden war.[55] Erst Jahre später konnte ein irrtümlich diesem Fundort zugeschriebenes Genom einem anderen, etwas jünger datierenden Fundort in London (St. Mary Graces) zugeordnet werden, was auch die etwas andere Position des Genoms im phylogenetischen Baum erklärt.[56]

Inzwischen ist die Zahl der verfügbaren Genome der Zweiten Pandemie auf über 30 gestiegen und verteilt sich auf Fundorte von Spanien bis Russland und vom 14. bis 18. Jahrhundert.[57]

Dabei zeigte sich, dass einige der Genome des 14. Jahrhunderts mit dem Londoner Genom des Schwarzen Todes identisch sind, nämlich Barcelona, Toulouse, Saint-Laurent-de-la-Cabrerisse, Abbadia San Salvatore, Nabburg und Oslo.[58] Dies ist insofern überraschend, als das Bakterium über weite geographische Distanzen gewandert sein muss, ohne dabei nachweisbare Mutationen im Genom fixiert zu haben. Für London und Barcelona konnte das noch durch schnelle maritime Transportrouten erklärt werden, die Fälle von Toulouse und Nabburg beweisen jedoch, dass sich die Pest auch über Land schnell ausgebreitet haben muss.

Aufgrund früherer PCR-Studien wurde bereits angenommen, dass das Genom des Schwarzen Todes (also der Pandemie von 1346–1353) im Stammbaum in die Nähe des Knotens fällt, von dem die Branches 1–4 abzweigen. Mithilfe der Genomanalysen wurde nachgewiesen, dass das der Schwarze Tod nur zwei SNPs von diesem Knoten entfernt auf Branch 1 liegt. Als SNP, kurz für *short nucleotide polymorphism*, werden einfache Mutationen der Basenfolge genannt, bei der eine Base durch eine andere ersetzt wird, etwa in der Basenfolge ATGCTG zu ATGTTG. Da die meisten modernen Stämme der vier Branches in Asien gefunden

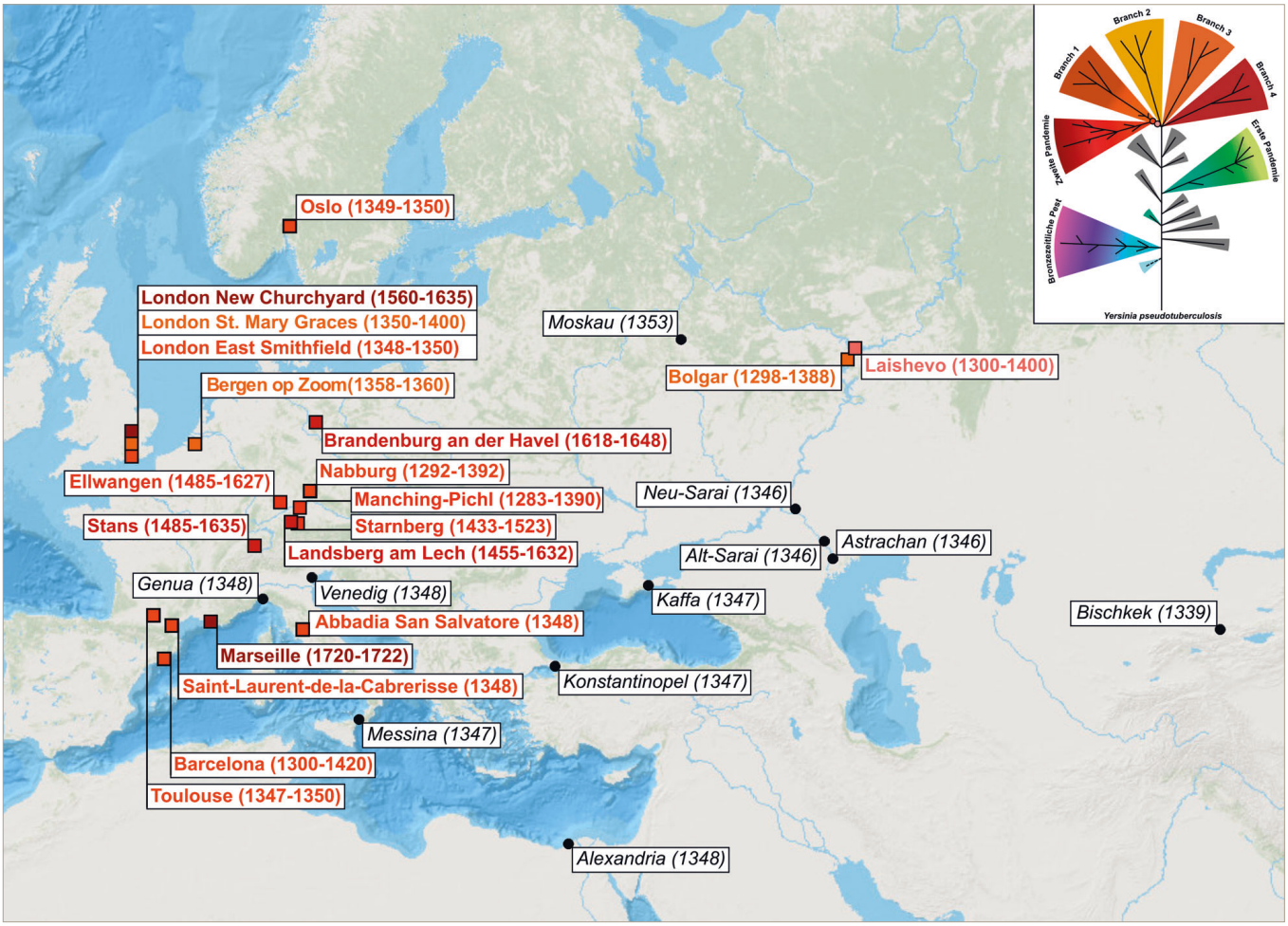
wurden, lässt sich darauf schließen, dass die Diversifizierung in die Branches jedoch in Asien stattfand.[59] Üblicherweise wird versucht, phylogenetische Stammbäume so weit aufzulösen, dass von einem Knoten immer nur zwei Äste abgehen. Neben dem besagten Knoten basal zu Branches 1–4 gibt es aber in der Phylogenie von *Yersinia pestis* noch weitere Knoten mit mehr als zwei Ästen (etwa ein Knoten, von dem die Stämme der Dritten Pandemie abzweigen), die sich nicht weiter auflösen lassen. Es wurde daher postuliert, dass solche sternförmigen Knoten in Zusammenhang mit einer raschen Ausbreitung stehen. Für die Geburt der Branches 1–4 muss vorerst offenbleiben, ob dies auch mit einer Pandemie in Asien in Verbindung steht. Spannend ist allerdings, dass sämtliche Genome von Branch 1 direkt von dem Genom des Schwarzen Todes abstammen. Da die basaleren dieser Genome in Asien zu finden sind, lässt sich daraus ableiten, dass ein *Yersinia pestis*-Stamm aus Asien nach Europa gelangt sein und den Schwarzen Tod verursacht haben muss, anschließend aber wieder nach Asien zurückgewandert sein muss, um sich dort in Wildnagetier-Reservoiren bis heute zu halten.[60] Da sich auch die Stämme der Dritten Pandemie auf Branch 1 befinden, kann sogar gesagt werden, dass etwa die Stämme des Ausbruchs auf Madagaskar 2017 direkt vom europäischen Schwarzen Tod abstammen.

Inzwischen gibt es auch ein Genom aus der Wolga-Region in Russland, die den Eintrag der Pest im 14. Jahrhundert aus Asien dokumentiert: Das Genom fällt im Stammbaum zwischen den Knoten der Branches 1–4 und den Schwarzen Tod (London East Smithfield etc.).[61] Der Fundort, Laishevo, fiel damals in das Reich der Goldenen Horde, und für die politischen Zentren des Reichs, Astrachan und Sarai, sind Pestausbrüche für das Jahr 1346 überliefert.[62] Die Städte liegen nahe der Mündung der Wolga ins Kaspische Meer, sodass angenommen werden kann, dass die Pest über die Wolga flussaufwärts nach Laishevo gelangte.

Von den jüngeren Genomen der Zweiten Pandemie liegen drei auf dem Ast, welcher zu den modernen Stämmen von Branch 1 führt. Das erste aus Bergen op Zoom (Niederlande) ist nur wenige SNPs vom Schwarzen Tod entfernt, fällt basal zu allen modernen Branch-1-Genomen, dem Genom aus London St. Mary Graces und dem Genom aus Bolgar, Russland.[63] Die Genome aus Bergen op Zoom und London sind dabei vermutlich der *Pestis Secunda* zuzuordnen, die Europa zwischen 1357 und 1366, nur wenige Jahre nach dem Schwarzen Tod heimgesucht hat. Bolgar wiederum liegt sehr nahe bei Laishevo, sodass diese Region innerhalb weniger Jahre sowohl Zeuge des Einzugs der Pest in Europa als auch deren Rückkehr nach Asien wurde.

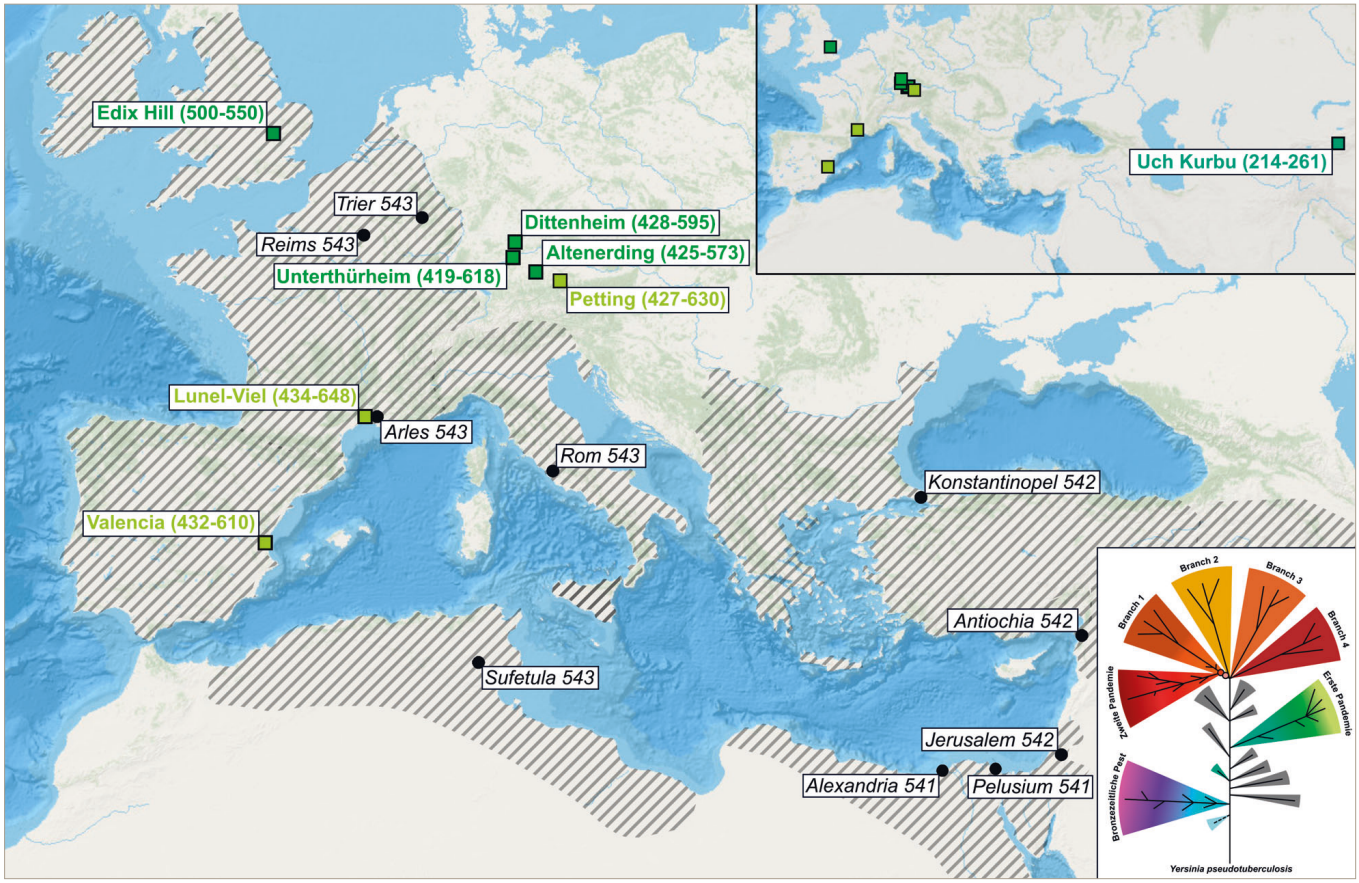
Alle restlichen Genome der Zweiten Pandemie hingegen fallen auf einen Ast, der ebenfalls vom Genom des Schwarzen Todes abstammt, dem aber keine modernen Stämme zugeordnet werden können. Es muss daher angenommen werden, dass diese Klade (geschlossene Abstammungsgemeinschaft, die auf einen gemeinsamen Vorfahren zurückgeht), die nur aus alten Genomen der Zweiten Pandemie in Europa besteht, heute ausgestorben ist. Das jüngste Genom dieser Klade kann der Großen Pest von Marseille 1720–1722 zugeordnet werden, eine der letzten großen Pestepidemien auf dem europäischen Kontinent.[64] Für alle weiteren alten Genome ist so eine präzise Zuordnung schwierig bis unmöglich. Mit Aufgabe der Sitte der Grabbeigaben im Frühmittelalter im Zuge der Christianisierung finden sich in Gräbern des Spätmittelalters und der frühen Neuzeit nur selten gut datierbare Artefakte wie Schmuck oder Münzen. Auch lassen sich viele Massengräber, die bei archäologischen Grabungen zutage gefördert werden, nicht historisch dokumentierten Pestausbrüchen zuordnen, dies gilt erst recht für kleinere Mehrfachbestattungen. Die Methode der Radiokarbondatierung kann hier zu einem gewissen Grad von Nutzen sein, allerdings erstrecken sich die damit bestimmten Zeitspannen oft über ein bis zwei Jahrhunderte.[65] Eine Ausnahme stellt hier etwa eine Dreifachbestattung aus Brandenburg an der Havel dar, die im Hinterhof eines Wohnhauses auf der Dominsel entdeckt wurde.[66] Bei einem der drei Individuen, bei denen es sich um junge Männer handelte, wurde ein Tonpfeifenkopf mit den Initialen eines Pfeifenbäckers gefunden, welcher in den 1630er-Jahren produzierte. Durch weitere Indizien ließ sich rekonstruieren, dass es sich möglicherweise um eine Bestattung von schwedischen Söldnern des Dreißigährigen Krieges handelt, welche 1631 in der Stadt stationiert waren.

Da sich die weiteren Genome sowohl grob zeitlich als auch im Baum als eine Abfolge zwischen dem Schwarzen Tod und der Großen Pest von Marseille einordnen lassen, muss darauf geschlossen werden, dass die Pest sich über die Jahrhunderte in Europa bzw. dem mediterranen Raum gehalten hat und es sich bei den Ausbrüchen nicht um wiederholte Einführungen des Erregers aus Asien handelte.[67] Letzteres wurde aufgrund klimatischer und epidemiologischer Modellierungen postuliert,[68] allerdings würde man dann eine höhere Diversität der Genome der Zweiten Pandemie erwarten. Da der Mensch nicht den natürlichen Wirt von *Yersinia pestis* darstellt, bedeutet dies, dass zu dieser Zeit ein Reservoir außerhalb Zentral- und Ostasiens bestanden haben muss. Ob es sich dabei um die Hausratte handelte oder um ein Wildtier wie das Europäische Murmeltier, muss vorerst unbeantwortet bleiben.[69]



3 Karte mit sämtlichen publizierten Genomen der Zweiten Pandemie. Die Beschriftung gibt den Fundort sowie die archäologische bzw. Radiokarbon-Datierung an. Die Farbkodierung folgt dabei dem phylogenetischen Baum (siehe Abb. 1) bzw. dem schematischen Baum oben rechts: Der Schwarze Tod ist in rot dargestellt, der unmittelbare Vorgänger aus Laishevo in rosa, die jüngeren Genome des europäischen Asts als dunkelroter Gradient und die Genome der *Pestis secunda* in orange. Wichtige historisch dokumentierte Pestausbrüche zum Beginn des Schwarzen Todes sind als schwarze Punkte mit kursiver Beschriftung kartiert.

Etwa in die Zeit des 16. Jahrhunderts kann auf der europäischen Linie ein weiterer Knoten mit mehr als zwei Ästen datiert werden. Aus diesem Knoten gehen nicht nur das Genom von Ellwangen und der gemeinsame Zweig von London New Churchyard und Marseille hervor, sondern auch ein zweiter Ast, der zu den Genomen von Landsberg, Stans und Brandenburg führt.[70] Interessanterweise datieren die Pestgräber von London New Churchyard und Brandenburg beide etwa in die erste Hälfte des 17. Jahrhunderts, die Genome zeigen jedoch unterschiedlich lange Äste von ihrem letzten gemeinsamen Vorfahren. Die Astlänge spiegelt in etwa die Anzahl der fixierten SNPs wider, und in der Tat konnte durch eine molekulare Datierung nachgewiesen werden, dass die Substitutionsrate, also die Anzahl fixierter Mutationen pro Zeiteinheit, auf dem Teilzweig von Ellwangen zu London New Churchyard deutlich erhöht ist. Die Ursache dafür muss vorerst offenbleiben, allerdings zeichnet sich auf diesem Ast eine weitere Besonderheit ab: Sowohl die Genome von London New Churchyard, als auch jene aus Marseille weisen eine große Deletion (also eine Mutation, bei dem ein Teil des Genoms verloren geht) von etwa 45.000 bp auf,



4 Karte mit sämtlichen publizierten Genomen der Ersten Pandemie. Die Beschriftung gibt den Fundort sowie die archäologische bzw. Radiokarbon-Datierung an. Die Farbkodierung folgt dabei dem phylogenetischen Baum (siehe Abb. 1) bzw. dem schematischen Baum unten rechts: Das Justinianische Genom aus Edix Hill in hellgrün, die identischen Genome aus Bayern in mittelgrün, die jüngeren Genome aus Petting, Lunel-Viel und Valencia in gelbgrün. Der Fundort eines deutlich älteren aber verwandten Genoms aus dem Tian-Shan-Gebirge ist im Kartenausschnitt oben rechts dargestellt. Wichtige historisch dokumentierte Pestausbrüche zum Beginn der Justinianischen Pest sind als schwarze Punkte mit kursiver Beschriftung angegeben. Regionen, für welche zeitgenössische Berichte zur Ersten Pandemie vorliegen, sind grau schraffiert.

was etwa 1% des Chromosoms ausmacht. Der Verlust einer so großen Genom-Region kann dadurch erklärt werden, dass das Genom von *Yersinia pestis* in seiner Architektur relativ instabil ist und durch besondere Sequenzabschnitte, sogenannte Insertionselemente, regelmäßig umgebaut wird. Eine solches Insertionselement findet sich etwa an einem Ende dieser Deletion. Mit dieser Region gingen etwa 35 Gene verloren, darunter auch zwei, welche in modernen Peststämmen als mögliche Virulenzfaktoren identifiziert wurden, also zur Gefährlichkeit des Erregers beitragen. Einige der anderen Gene hingegen sind Relikte aus der evolutiven Vergangenheit von *Yersinia pestis* als freilebendes oder darmpathogenes Bakterium, das sich mithilfe von Rezeptoren und Flagellen selbstständig fortbewegen konnte. Als Gene für Oberflächenproteine, die vom Immunsystem des Wirts erkannt werden können, könnte der Verlust dieser Region somit auch die Virulenz des Erregers erhöhen. Bis die Auswirkungen dieser Mutation an modernen Stämmen erforscht werden, bleibt dies jedoch Spekulation.[71]

Die Justinianische Pest und die Erste Pandemie

Im Gegensatz zum Schwarzen Tod sind die Justinianische Pest und die folgenden Ausbrüche des Frühmittelalters heute außerhalb der Fachkreise praktisch unbekannt. So hat sich die Seuchengeschichte auch erst in den letzten Jahrzehnten intensiver mit der Ersten Pandemie beschäftigt.[72] Dies mag neben dem größeren zeitlichen Abstand daran liegen, dass etwa aus dem deutschsprachigen Raum keine schriftlichen Quellen überliefert sind. Dagegen sind die Ausbrüche im Oströmischen Reich sehr gut dokumentiert, auch für das ehemalige weströmische Reich haben sich einige Aufzeichnungen von Klerikern erhalten. Der gesamte Schatz der islamischen Quellen zu späteren Epidemien im Nahen Osten ist vermutlich noch gar nicht gehoben, was sicher auch mit den modernen Fachgrenzen und Sprachbarrieren zusammenhängt. Die Ausmaße der Justinianischen Pest und die Auswirkungen dieser und folgender Epidemien am Übergang von der Spätantike zum Frühmittelalter sind schwer einzuschätzen, aber zumindest für das Oströmische Reich muss mit erheblichen Bevölkerungsverlusten gerechnet werden.[73]

Aufgrund des vorgeblichen Mangels an archäologisch belegten Massengräbern wurde mitunter die Bedeutung der Justinianischen Pest heruntergespielt. Dies wurde allerdings durch Arbeiten von McCormick infrage gestellt, welche etwa auch ein umfangreiches »Inventar« von Massengräbern der Spätantike enthalten.[74] Wie im Folgenden gezeigt werden wird, sind aber auch manche ansonsten relativ unauffällige Mehrfachbestattungen mit zwei bis fünf Individuen in den Kontext der Ersten Pandemie zu stellen. Aufgrund der niedrigeren Populationsdichte fielen der Pest abseits der urbanen Zentren vermutlich nicht genug Menschen zum Opfer, um ein Massengrab zu füllen.

Die erste genetische Spur der Justinianischen Pest kam für Historiker aus völlig unerwarteter Richtung: 2003 erschien eine Publikation von Wiechmann und Garrelt, die DNA-Fragmente von *Yersinia pestis* mittels PCR in einer frühmittelalterlichen Doppelbestattung aus Aschheim in Bayern nachwiesen.[75] Aus historischen Quellen war es zuvor nicht klar, ob die Pest in dieser Zeit bis ins heutige Bayern vordrang, mit Zusammenbruch der weströmischen Herrschaft riss auch die schriftliche Überlieferung ab.

Auch für französische Fundorte wurden PCR-Studien veröffentlicht, deren Ergebnisse aber teilweise im Widerspruch mit neueren Erkenntnissen stehen.[76] 2014 gelang es schließlich Wagner u. a. aus Proben desselben Gräberfelds in Aschheim ein komplettes Genom für die Justinianische Pest zu rekonstruieren, womit zumindest eine Beteiligung von *Yersinia pestis* an der Ersten Pandemie unzweifelhaft war.[77] Anders als der Stamm des Schwarzen Todes hat der Stamm der Justinianischen Pest keine überlebenden Nachfahren. Vielmehr bildet die Justinianische Pest einen eigenen, relativ langen Ast, der von Branch 0 abzweigt, bevor dieser in den Knoten mit den Branches 1–4 mündet. Die nächstverwandten modernen Stämme wurden alle in Zentralasien isoliert, womit diese Region als Ursprung des »Justinianischen Asts« angenommen werden kann.[78] Ein altes Genom aus dem 2.–3. Jahrhundert aus dem zentralasiatischen Tian-Shan-Gebirge (Uch Kurbu, Kirgisistan), welches sehr basal auf diesen Ast fällt, stützt diese These.[79] Ob es sich bei dem Individuum jedoch um das Opfer einer Epidemie handelt oder um einen sporadischen Fall – es handelt sich offenbar um eine Doppelbestattung in einem Grabhügel, und vermutlich zirkulierten in dieser Zeit bereits verschiedene Peststämme in den lokalen Nagetierpopulationen –, bleibt offen.

Inzwischen liegen uns auch für die Erste Pandemie mehr Genome aus Bayern, von der spanischen wie französischen Mittelmeerküste und aus Großbritannien vor, die eine gemeinsame Klade bilden.[80] Mit den Pestnachweisen in Valencia und in Lunel-Viel nahe des Rhône-Deltas war zu rechnen, zumindest für Südfrankreich sind durch den Bischof und Geschichtsschreiber Gregor von Tours zahlreiche Ausbrüche überliefert.[81] Eine größere Überraschung hingegen war das britische Genom, welches von einem angelsächsischen Gräberfeld in der Nähe von Cambridge stammt.[82] Über Grabbeigaben können die Pestgräber in etwa auf die Mitte des 6. Jahrhunderts datiert werden. Für diese Zeit ist uns lediglich eine Epidemie auf den Britischen Inseln überliefert, deren Ursache sich aufgrund fehlender Symptombeschreibungen und einer rätselhaften Bezeichnung nicht genauer bestimmen ließ. Nur aufgrund der zeitlichen Nähe zum Beginn der Justinianischen Pest wurde von einigen Historikern postuliert, es handle sich um die Pest.[83] Erst für die Mitte des 7. Jahrhunderts sind uns Quellen erhalten, die eine relativ sichere Zuordnung von Epidemien in Großbritannien zur Pest erlauben.

Auch die Position des Genoms im Stammbaum spricht für eine frühe Datierung. Es zweigt mit einem fehlenden SNP basal zu einem sternförmigen Knoten ab, aus dem sämtliche anderen bekannten Genome in vier Ästen entspringen, und trägt nur einen zusätzlichen eigenen SNP. Mit besagtem Knoten sehen wir wieder ein Phänomen, welches uns bereits weiter oben begegnet ist und das möglicherweise auf eine schnelle Ausbreitung hindeutet.[84]

Aus Bayern liegen uns inzwischen identische Genome von vier Reihengräberfeldern vor: Zu den bereits länger publizierten aus Aschheim und Altenerding traten noch Unterthürheim und Dittenheim hinzu, was den epidemischen Charakter nur noch unterstreicht.[85] Al-

lerdings weicht ein weiteres Genom aus Petting ab und fällt auf einen der drei anderen beschriebenen Äste. Somit muss also davon ausgegangen werden, dass die Erste Pandemie Bayern mindestens zweimal erreicht hat. Die geographische Verteilung gibt uns hier möglicherweise weiter Aufschluss: Die mit Aschheim identischen Genome fallen alle in oder nahe der ehemaligen Provinz *Raetia secunda*, wohingegen Petting in der ehemaligen Provinz *Noricum ripense* lag, die Grenze verlief dabei entlang des Inn. Ob wir es hier aber tatsächlich mit zwei geographisch separierten Epidemien zu tun haben, ist mit der geringen Anzahl an Proben noch nicht abschließend zu klären.

Für Valencia, das zu dieser Zeit unter westgotischer Herrschaft stand, konnte kürzlich durch Gruber aus Textquellen ein Pestausbruch für das Jahr 548 rekonstruiert werden.[86] Ob es sich bei dem Genom um ein Zeugnis der gleichen oder einer anderen Epidemie handelt, kann aufgrund der unpräzisen Datierung nicht bestimmt werden. Es stammt aus einem intramuralen Bestattungsplatz nahe einer Kirche, welcher mehrere Kollektivgräber enthält. Da diese relativ leicht wieder zu öffnen waren, boten sie sich vermutlich als Notbestattungsplatz in Krisenzeiten wie einer Pestwelle an.[87]

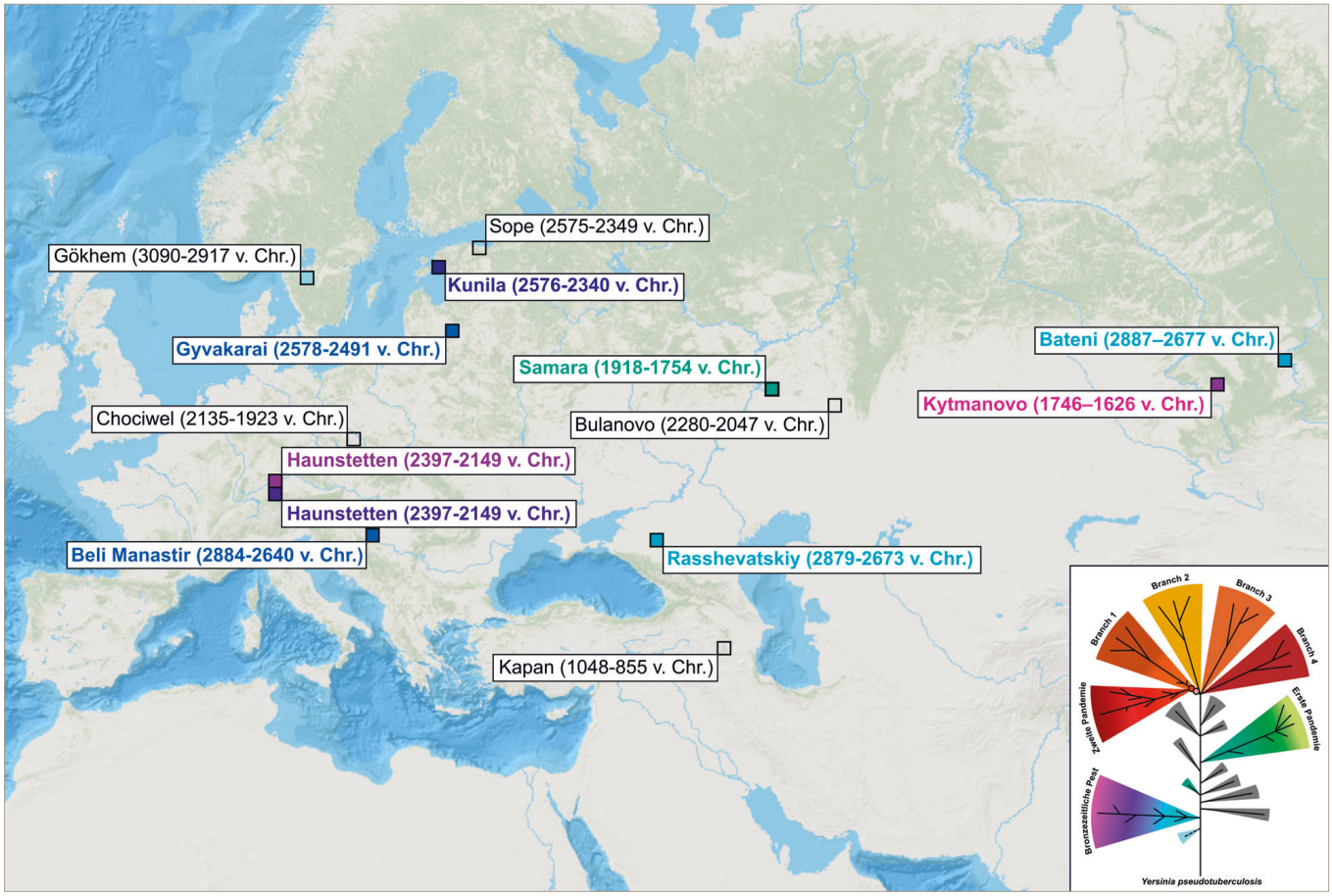
Das Genom aus Lunel-Viel (Südfrankreich) hat schließlich von dem gemeinsamen Knoten ausgehend den längsten Ast, was auch mit der jüngeren Radiokarbondatierung korreliert. Somit kann auch für die Erste Pandemie angenommen werden, dass die wiederkehrenden Ausbrüche nicht durch unabhängige Wiedereintragungen der Pest, sondern durch einen lokal persistierenden Stamm verursacht wurden. Im Gegensatz zu den relativ regelkonformen Pestgräbern aus Bayern und Großbritannien deutet der Kontext dieses Genoms tatsächlich auf eine Mortalitätskrise hin: Acht Individuen unterschiedlichen Alters und Geschlechts wurden in Gräben geworfen, welche durch den Aushub von Fundamentsteinen einer römischen Villa zur Wiederverwendung entstanden sind.[88] Völlig unerwartet zeigte sich bei der Analyse dieses Genoms, dass beinahe exakt die gleiche Genomregion fehlt, wie dies für die Genome der Zweiten Pandemie aus London New Churchyard und Marseille beschrieben wurde.[89] Da das Genom aus Lunel-Viel jedoch im Baum auf einen anderen Ast fällt und die jeweils nächstverwandten Stämme der Ersten und Zweiten Pandemie diese Region besitzen, geschah diese Mutation mindestens zwei Mal unabhängig voneinander, zwischen dem 6. und 7. Jahrhundert und zwischen dem 16. und 17. Jahrhundert. Überraschenderweise zeigte keines der über 200 modernen Genome an dieser Position eine ähnlich große Lücke im Genom, sodass für die Mutation ein gemeinsamer Hintergrund zu bestehen scheint. Eine Besonderheit der Stämme der Ersten und Zweiten Pandemie ist, dass sie außerhalb der heute bekannten Reservoire zirkuliert haben müssen. Demnach könnte es sich bei der Mutation um eine Anpassung an ein neues Wirtstier handeln, welcher das Reservoir während der historischen Pandemien stellte, wie es etwa für die Hausratte angenommen wurde.

Die Pest in der Vor- und Frühgeschichte

Berichte über Epidemien sind in etwa so alt wie die ersten Schriftquellen, die wir besitzen, und spätestens mit der Sesshaftwerdung in größeren Siedlungen und dem engen Kontakt zu Tieren durch die Domestizierung in der Jungsteinzeit wurde den Seuchen ein idealer Nährboden gelegt. Allerdings ließ sich von den zahlreich überlieferten Epidemien der Frühgeschichte und Antike, vom Atrahasis-Epos und der Bibel bis zur »Cyprianischen Pest« im 3. Jahrhundert, bisher keine eindeutig dem Pesterreger *Yersinia pestis* zuordnen. In den Schriften des antiken Arztes Rufus von Ephesos aus dem 1. Jahrhundert jedoch fanden sich Zitate älterer Werke zurück bis ins 3. Jahrhundert v. Chr., die die charakteristischen Symptome der Beulenpest beschreiben.[90] Bis vor wenigen Jahren war dies der erste vage Beleg der Pest vor der Justinianischen Pest.

Dabei stand auch das Alter von *Yersinia pestis* selbst zur Diskussion. Über die Substitutionsrate kann mittels molekularer Datierung errechnet werden, wann beispielsweise der letzte gemeinsame Vorfahre zweier Arten gelebt haben muss. Da hierbei zahlreiche Parameter berücksichtigt werden müssen und in der Regel auch keine völlig konstante Substitutionsrate angenommen werden kann, werden dafür bioinformatische Programme benutzt, die stochastische Algorithmen mit Bayes'scher Statistik kombinierten.[91] Alten Genomen kommt bei der molekularen Datierung eine besondere Bedeutung zu, da diese gleichsam Kalibrierungspunkte darstellen, sofern das Alter zuverlässig archäologisch oder über Radiokarbondatierung bestimmt wurde. So wurde dann auch in der Publikation des ersten alten Pestgenoms von Bos u. a. 2011 der letzte gemeinsame Vorfahre aller *Yersinia pestis*-Stämme auf das 4. bis 2. Jahrhundert v. Chr. datiert, wobei damals jedoch nur 17 moderne *Yersinia pestis*-Sequenzen zur Verfügung standen.[92]

2015 erschien schließlich eine Publikation von Rasmussen u. a., die zeigen konnte, dass bereits in der Bronzezeit Menschen der Pest zum Opfer fielen.[93] Die insgesamt sieben positiven Individuen verteilen sich auf sechs Fundorte vom Altaigebirge bis nach Estland und Armenien, wobei aber nur für zwei Proben, datierend in das 3. bis 2. Jahrtausend v. Chr., komplette Genome rekonstruiert werden konnten. Überraschend



5 Karte mit sämtlichen publizierten Genomen der Jungsteinzeit und Bronzezeit. Die Beschriftung gibt den Fundort sowie die Radiokarbon-Datierung an. Die Farbkodierung folgt dabei dem phylogenetischen Baum (siehe Abb. 1) bzw. dem schematischen Baum unten rechts: Das älteste Genom aus Gökhem in hellblau, die verwandten Bronzezeitlichen Genome als blau-violett-Gradient und das jüngste und phylogenetisch eigenständige Genom aus Samara in türkis. Für die Fundorte mit leeren Kästchen konnte der Pesterreger zwar nachgewiesen werden, allerdings reichte die genomische Abdeckung nicht für eine phylogenetische Analyse aus.

war hierbei, dass diese Genome auf einen eigenen Ast fielen, der basal zur gesamten modernen Diversität abzweigt. Es handelt sich also um eine Frühform des Bakteriums, die in der Folge ausstarb und von Stämmen verdrängt wurde, die wir auch heute noch in Zentralasien finden.

Wie sich herausstellte, zeichneten sich diese Genome durch weitere Besonderheiten aus: Sie zeigten nur einen Teil der genetischen Veränderungen, die *Yersinia pestis* so besonders machen. Zwar besaßen die bronzezeitlichen Stämme bereits alle drei Virulenzplasmide, allerdings fehlte ihnen beispielsweise das Gen *ymt*, wohingegen andere Gene wie etwa *ureD*, die in allen modernen Peststämmen ihre Funktion verloren haben, noch intakt waren. Diese Mutationen zeichnen sich dadurch aus, dass sie eine Anpassung an den Übertragungsweg mittels Floh darstellen, da sie einerseits das Überleben von *Yersinia pestis* im Vormagen des Flohs erlaubten (*ymt*),[94] andererseits aber auch das Überleben des Flohs als Vektor garantierten (*ureD*).[95] Sowohl die bisher bekannten modernen Pestgenome als auch die Genome der Ersten und Zweiten Pandemie waren genetisch relativ uniform, mit den bronzezeitlichen Genomen wurde ein gänzlich anderer Erregertypus identifiziert, der ohne die Paläogenetik nie entdeckt worden wäre. Dies stellt gleichsam ein Fenster in die Evolution des Erregers dar, da anhand

der alten Genome festgestellt werden kann, welche Anpassungen an die neue Ökologie zu diesem Zeitpunkt bereits vorlagen und welche noch ausstanden. So wurde zuvor vermutet, dass Mutationen wie die Aufnahme des *ymt*-Gens sehr früh geschehen sein musste, um die Flohübertragung zu gewährleisten.[96] Wie sich nun aber zeigte, war diese Frühform der Pest, der bereits Menschen zum Opfer fielen, noch gar nicht voll an den Flohvektor angepasst. Somit bleibt auch offen, wie genau der Infektionsweg verlief. Zwar gibt es auch einen Übertragungsweg mittels Flöhen und Läusen, der dem Prinzip der »dreckigen Nadel« folgend eine schnelle Übertragung ermöglicht und keine genetische Anpassung des Erregers verlangt.[97] Allerdings ist dieser Infektionsweg relativ ineffizient,[98] sodass es fraglich ist, ob damit allein das Bakterium in einer Nagetierpopulation persistieren kann.

Inzwischen haben wir auch für die bronzezeitliche Pest mehr Daten: Andrades Valtueña u. a. publizierten 2017 sechs Genome aus Russland, dem Baltikum, Kroatien und Süddeutschland.[99] Diese fallen alle auf den gleichen Ast wie die zuvor publizierten Genome aus Russland, genauer gesagt reihen sie sich quasi chronologisch auf. Überraschend war hierbei, dass sowohl die jüngsten als auch ältesten Genome aus Russland stammen, die mitteleuropäischen jedoch sowohl zeitlich als auch phylogenetisch dazwischen liegen. Damit zeichnet sich also eine Bewegung des Erregers aus Zentralasien nach Europa in der frühen Bronzezeit und eine spätere Rückkehr ab. Interessanterweise korreliert dies zeitlich mit menschlichen Migrationsbewegungen aus der pontischen Steppe (Yamnaya-Kultur) nach Europa und Südsibirien um 2800 v. Chr., die nicht nur zu einen erheblichen Eintrag neuer genetischer Komponenten in die europäische Bevölkerung führte, sondern auch mit der Verbreitung des domestizierten Pferdes und womöglich auch der Ausbreitung der indoeuropäischen Sprachen in Zusammenhang steht.[100] Demnach hätten die höhere Mobilität und die großen Migrationsbewegungen auch die Ausbreitung der Pest begünstigt.

Ein kürzlich publiziertes Genom aus Schweden beweist allerdings, dass auch bereits während des Spätneolithikums etwa 3000 v. Chr. ein Individuum der Trichterbecherkultur an Pest verstarb.[101] Allerdings handelt es sich hierbei um einen noch basaleren, unabhängigen Seitenast im Stammbaum der Pest. Dass dieser Stamm, wie von den Autoren postuliert, tatsächlich mit dem Zusammenbruch der jungsteinzeitlichen Kulturen in Europa zusammenhängt, lässt sich bisher aber noch nicht beweisen.

Die jüngste Probe des Datensatzes von Rasmussen u. a. 2015 stammte von einem eisenzeitlichen Individuum aus Armenien (ca. 1000 v. Chr.), allerdings war die Abdeckung des *Yersinia pestis*-Genoms zu gering, um es in einem phylogenetischen Baum zu platzieren. Jedoch schien dieses Genom bereits das *ymt*-Gen zu besitzen, welches einen der wesentlichen Unterschiede zwischen der bronzezeitlichen und der modernen Pest ausmacht. Spyrou u. a. gelang es schließlich 2018, einen dem beschriebenen Genom vermutlich nahe verwandten Stamm zu beschreiben, der aus einer spätbronzezeitlichen Doppelbestattung aus Samara, Russland, rekonstruiert wurde.[102] Neben dem *ymt*-Gen trägt das Genom auch jene Mutationen, die für moderne Peststämme charakteristisch sind, wie etwa den Funktionsverlust von *ureD*. Auch phylogenetisch fällt dieser Stamm zwischen Äste moderner Peststämme, die teilweise aus erkrankten Patienten isoliert wurden. Somit stellt dieser Stamm den ältesten Nachweis der »modernen« Pest dar, der bereits voll an die Übertragung mittels Floh angepasst war und die Beulenpest verursachen konnte. Damit wäre also auch die Beulenpest als Ursache für Epidemien der Antike wieder im Spiel, und auch wenn etwa die »Pest des Thukydides« (430–426 v. Chr.), die »Antoninische« (165–180 n. Chr.) oder die »Cyprianische Pest« (250–271 n. Chr.) vermutlich anderen Erregern zugeordnet werden können,[103] bleibt fraglich, woher die von Rufus von Ephesos zitierten antiken Ärzte die Symptome der Beulenpest kannten. Schließlich gibt es aber mit der »Pest der Philister« und der »Pest der Hethiter« noch Epidemien in der Bronzezeit, die anhand der literarischen Quellen quasi nicht zu identifizieren sind.

Die offenen Fragen

Einige der offenen Fragen wurden oben bereits angerissen, beispielsweise jene nach einem Pestreservoir in Europa. In der Gesamtschau drängt sich allerdings nicht nur die Frage auf, welches Tier dabei als Reservoir diente, sondern auch, warum die Pest nach der bronzezeitlichen Pest und den historischen Pandemien jeweils wieder aus Europa verschwand. Solange es keine Nachweise von *Yersinia pestis* in Tierknochen aus archäologischen Kontexten der entsprechenden Epochen gibt, werden sich diese Fragen nur schwer mit Gewissheit beantworten lassen. Dies stellt eine besondere Herausforderung dar, da Knochen von Ratten oder wilden Nagetieren aus den entsprechenden Perioden dafür zuerst in Kontexte wie Gräber oder Kloaken gelangen müssen, um eine Überlieferung zu gewährleisten. Schließlich müssen diese aber die Aufmerksamkeit des archäologischen Ausgrabungsteams erlangen. Dies ist nicht nur aufgrund der Größe ein Problem, manche Nagetiere legen auch metertiefe Gangsysteme an, weshalb solche Funde oft mit störender »Bioturbation« in Zusammenhang gebracht werden. Noch komplexer ist

die Frage nach den Übertragungswegen. Neben dem klassischen Modell des »blockierten Flohs« wurden weitere Szenarien vorgeschlagen, etwa Menschenflöhe und -läuse, was die Anwesenheit von Nagetieren für Epidemien obsolet machen würde.

Weiterhin ist auch unklar, inwieweit andere Krankheitserreger an den Pestepidemien im Sinne einer »Syndemie« beteiligt waren. Mit den methodischen Weiterentwicklungen der letzten Jahre, die das Screening für Pathogene ohne konkreten Verdacht deutlich erleichtert haben, ist diese Frage theoretisch schon bis zu einem gewissen Grade beantwortbar. Allerdings gibt es eine Reihe von tödlich verlaufenden Infektionskrankheiten, die mit den herkömmlichen Methoden der Paläogenetik nicht nachweisbar sind: Zum einen sind dies gastrointestinale Erkrankungen, bei denen der Erreger in der Regel nicht in die Blutbahn gelangt, sondern etwa durch Dehydration zum Tod führt. Solche Pathogene wie etwa der Cholera-Erreger lassen sich aller Wahrscheinlichkeit nach auch in Zukunft nicht in den Zähnen der Verstorbenen nachweisen. Die zweite große Lücke stellen derzeit noch RNA-Viren dar. Dazu zählen etwa die Erreger von Ebola, Polio, Influenza oder Tollwut. Bei diesen Viren besteht das Genom nicht aus DNA (Desoxyribonukleinsäure), wie bei allen Bakterien oder auch dem Pockenvirus, sondern aus RNA (Ribonukleinsäure), bei der die Ribose, welches das Rückgrat des Stranges bildet, eine zusätzliche Hydroxygruppe besitzt. Dies macht das Molekül instabil, da dadurch das Rückgrat chemisch leicht gebrochen werden kann. Es ist somit anzunehmen, dass sich in archäologischem Material keine alte RNA mehr nachweisen lässt. Es bleibt allerdings die Hoffnung, dass sich in Zukunft mit den Methoden der Proteomik, der qualitativen Analyse von Proteinen mittels Massenspektrometrie, die spezifischen Kapselproteine dieser Viren nachweisen lassen.

Alle diese Fragen hängen zumindest teilweise mit dem großen Rätsel zusammen, wie *Yersinia pestis* in der Geschichte so verheerende Pandemien verursachen konnte und ob wir solche auch in Zukunft wieder zu befürchten haben. Heute sind wir hauptsächlich mit sporadischen zoonotischen Fällen und vergleichsweise kleinen Ausbrüchen konfrontiert, eine verringerte Virulenz des Bakteriums lässt sich aber nicht nachweisen. Eine mögliche Erklärung wäre eine höhere Immunität der Menschen durch Selektion, zumindest für Regionen, die auch in der Vergangenheit von der Pest heimgesucht wurden. Ein solches Szenario wurde für die Mutation CCR5-Δ32 vorgeschlagen, welche zu einer Resistenz gegen Infektionen mit HIV-1 führt und heute mit 4–16 % in der europäischen Bevölkerung in höherer Frequenz auftritt als anderswo. Aufgrund der Häufigkeit und einem vermeintlichen Alter der Mutation von 700 Jahren wurde spekuliert, dass diese Mutation auch eine Resistenz gegen *Yersinia pestis* bewirkt und deshalb positiv selektiert wurde.[104] Allerdings wurde diese Hypothese später aus mehreren Gründen abgelehnt, u. a. weil auch Mäuse mit dieser Mutation mit *Yersinia pestis* infiziert werden können[105] und sich keine positive Selektion nachweisen lässt.[106]

Offen bleibt die Frage, ob andere Mutationen zu einer erhöhten Immunität beigetragen haben könnten. Durch die neuesten paläogenetischen Erkenntnisse sind solche Studien auch konzeptionell anspruchsvoll, da für Eurasien wiederholte Pestausbrüche seit der Bronzezeit, für Nordafrika spätestens seit der Justinianischen Pest angenommen werden müssen. Dies macht es schwierig, eine geeignete Vergleichspopulation zu finden, die diesem Pathogen nicht ausgesetzt war.

Schließlich liegt der Schlüssel zum Verständnis der historischen Pandemien im Vergleich zur modernen Pest aber auch gar nicht im Erreger *Yersinia pestis* selbst oder dem menschlichen Immunsystem, sondern bei ökologischen Faktoren. Für Pestausbrüche in Nagetieren konnte der Zusammenhang mit klimatischen Schwankungen bereits gut untersucht werden. Da es sich um eine Zoonose handelt, spielen auch komplexe Mensch-Tier-Interaktionen eine wichtige Rolle, sei es die Nutzung von Nagetieren etwa als Pelzlieferanten und der Kontakt von Menschen mit Ratten und deren Flöhen, welcher wiederum auch von sozialen Faktoren abhängig ist. Um diesen Fragen nachzugehen, bedarf es anderer Ansätze der Paläogenetik (etwa der Populationsgenetik der Ratten) oder gänzlich anderer Disziplinen wie der Klimageschichte oder Umweltarchäologie. Die molekulare Paläopathologie stellt also ein exzellentes Werkzeug dar, um den Erreger nachzuweisen und deren Evolutionsgeschichte nachzuvollziehen, greift damit aber nur einen von etlichen Faktoren heraus, die wir für ein Verständnis historischer Pandemien berücksichtigen müssen.

Anmerkungen

1 Uhlmann 2017; McNeil 2017.

2 Gerhard 2017.

3 Als »alte DNA« (englisch: ancient DNA) bezeichnet man DNA, die aus paläontologischem oder archäologischem Material gewonnen wird, dementsprechend wird die Disziplin Paläo- oder Archäogenetik genannt. Allerdings dient auch zunehmend organisches Material aus Archiven, Bibliotheken und Sammlungen als Untersuchungsgegenstand. Die Besonderheit der alten DNA liegt nicht in ihrer Quelle (sie kann menschlichen, bakteriellen, tierischen oder pflanzlichen Ursprungs sein) sondern den gemeinsamen Eigenschaften: Durch biologische, chemische und physikalische Prozesse nach dem Zelltod ist die DNA stark fragmen-

tiert und weist charakteristische chemische Modifikationen auf, weshalb die Analyse alter DNA besondere Methoden verlangt.

4 Hecker 1832.
5 Alexander 2003.
6 Haeser 1875.
7 Haeser 1875, 16–18.
8 Haeser 1875, 37–56 (Justinianische Pest), 97–181.
9 Haeser 1875, 727–731.
10 Yersin 1894.
11 Yersin 1894, 663.
12 Simond 1898.
13 Bacot/Martin 1914a; Bacot/Martin 1914b.
14 Eine exzellente, knappe Zusammenfassung der modernen Pestforschung findet sich in Kupferschmidt 1993.
15 Gage/Kosoy 2005.
16 Dennis u. a. 1999, 33–34. 80–81.
17 Siehe Dennis u. a. 1999, 43–54 zu den klinischen Formen.
18 Harris 2002.
19 Devignat 1951.
20 Zusätzlich wurden in diese Klassifizierung »Pestoides« (PE) und »Intermedium« (IN) als weitere Biovare integriert, wobei ersteres für einige basale Stämme mit abgeschwächter Virulenz postuliert wurde (Song u. a. 2004), auch wenn wir heute wissen, dass 0.PE-Stämme Menschen infizieren können.
21 Parkhill u. a. 2001.
22 Eroshenko u. a. 2017; Kutyrev u. a. 2018; Kislichkina u. a. 2015; Kislichkina u. a. 2017; Kislichkina u. a. 2018a; Kislichkina u. a. 2018b; Zhgenti u. a. 2015.
23 Biraben 1975/1976 (Bd. 1).
24 Shrewsbury 1970.
25 Cohn 2003.
26 Twigg 1984.
27 Scott/Duncan 2001.
28 Benedictow 2004.
29 Benedictow 2010.
30 Benedictow 2010, 381–395.
31 Bos u. a. 2011.
32 Wren 2003.
33 Achtman u. a. 1999. Bakterien vermehren sich asexuell durch einfache Zellteilung. Im Gegensatz zu anderen Bakterien tritt bei Yersinien zudem kaum Rekombination, also der Austausch genetischer Elementem auf. Deshalb kann *Yersinia pestis* als »klonal« bezeichnet werden, die Nachkommen sind also genetisch mit den Vorfahren identisch, lässt man sporadische Mutationen im Laufe der Evolution außen vor.
34 Hinnebusch u. a. 2002.
35 Sebbane u. a. 2006.
36 Parkhill u. a. 2001.
37 Siehe z. B. Song u. a. 2004.
38 Eine gute Übersicht zur modernen Diversität und Phylogeographie der Pest bietet Cui u. a. 2013.
39 Für eine deutschsprachige Einführung in die Paläopathologie, siehe Grupe/Harbeck/McGlynn 2015, 353–404.
40 Higuchi u. a. 1984; Pääbo 1985.
41 Drancourt u. a. 1998.
42 Eine gute Einführung in die PCR-Methode, insbesondere im Kontext alter DNA, bietet Grupe/Harbeck/McGlynn 2015, 479–527.
43 Pääbo u. a. 2004 bietet dazu einen guten Überblick.
44 Gilbert u. a. 2004.
45 Siehe etwa Willerslev/Cooper 2005.
46 Briggs u. a. 2007.
47 Knapp/Hofreiter 2010.
48 Maricic/Whitten/Pääbo 2010.
49 Schuenemann u. a. 2011.
50 Spyrou u. a. 2016; Bos u. a. 2016; Spyrou u. a. 2018a; Wagner u. a. 2014; Feldman u. a. 2016; Keller u. a. 2018; Rasmussen u. a. 2015; Andrades Valtueña u. a. 2017b; Spyrou u. a. 2018b; Damgaard u. a. 2018.
51 Feldman u. a. 2016; Keller u. a. 2018; Spyrou u. a. 2018a.
52 Andrades Valtueña u. a. 2017b; Rasmussen u. a. 2015.
53 Haensch u. a. 2010; Bos u. a. 2016.
54 Haensch u. a. 2010.
55 Siehe etwa Waldron 2001; DeWitte 2010.
56 Bos u. a. 2016.
57 Spyrou u. a. 2016; Bos u. a. 2016; Spyrou u. a. 2018a.
58 Spyrou u. a. 2016; Spyrou u. a. 2018a; Namouchi u. a. 2018.
59 Cui u. a. 2013.
60 Spyrou u. a. 2016.
61 Spyrou u. a. 2018a.
62 Dols 1977, 50.
63 Spyrou u. a. 2016; Namouchi u. a. 2018.
64 Bos u. a. 2016.
65 Siehe z. B. das Massengrab aus Ellwangen, publiziert in Spyrou u. a. 2016.
66 Dalitz/Grupe/Jungklaus 2012; Spyrou u. a. 2018a; Seifert u. a. 2016.
67 Spyrou u. a. 2018a; Spyrou u. a. 2016; Bos u. a. 2016; Seifert u. a. 2016.
68 Schmid u. a. 2013.
69 Letzteres wurde etwa von Carmichael 2014 vorgeschlagen.
70 Spyrou u. a. 2018a.
71 Funktionelle Studien am lebenden Bakterium bzw. im Wirtsorganismus betreibt etwa Joseph Hinnebusch am National Institute for Allergy and Infectious Diseases, National Institutes of Health, Hamilton, USA. Im Rahmen einer Kooperation mit dem Max-Planck-Institut für Menschheitsgeschichte sind auch Untersuchungen zu den Auswirkungen von Mutationen in alten Genomen wie der beschriebenen Deletion geplant.
72 Wichtige frühe Arbeiten stammen etwa von Russell 1968; Biraben/Le Goff 1975; Allen 1979; Sarris 2002. Mit seiner Dissertation stellte Stathakopoulos 2004 einen umfangreichen Katalog von Originalquellen und eine systematische Erfassung von Ausbrüchen der Ersten Pandemie für das Byzantinische Reich zusammen. Der Sammelband von Little 2007a schließlich versammelt mehrere Aufsätze, die sich auch intensiv mit der Ersten Pandemie in Westeuropa beschäftigen, beruhend auf einer Konferenz an der American Academy in Rom 2001. Eine neuere, knappe Zusammenfassung bietet Harper 2017, 199–245, im Appendix 304–315 findet sich außerdem eine Liste zu Ausbrüchen der Pest nach 544, ähnlich jener von Stathakopoulos 2004.

73 Die beiden Extreme stellen Russell 1968 und Durliat 1989 dar. Während Russell von einem Bevölkerungsverlust von bis zu 50 % ausgeht und die Erste Pandemie in den Zusammenhang mit der Ausbreitung des Islams stellt, bestreitet Durliat gravierende Auswirkungen, insbesondere außerhalb der größeren Städte. Neuere Studien wie Sarris 2002, Meier 2016 und Harper 2017 kommen, u. a. durch Hinzuziehung von nicht-schriftlichen Quellen, zu einem ausgewogeneren und detaillierteren Bild.
74 McCormick 2016; McCormick 2015.
75 Garrelt/Wiechmann 2003.
76 Drancourt u. a. 2004 propagierten, dass sowohl die Erste (541–750) als auch die Zweite Pandemie (1346 bis 18. Jahrhundert) dem Orientalis-Biovar zuzuordnen sind. Heute wissen wir, dass dieser Biovar vermutlich erst im 19. Jahrhundert entstand.
77 Wagner u. a. 2014.
78 Cui u. a. 2013; Eroshenko u. a. 2017.
79 Damgaard u. a. 2018.
80 Keller u. a. 2018.
81 Harper 2017.
82 Malim/Hines 1998.
83 Siehe Maddicott 2007 für England und Dooley 2007 für Irland.
84 Keller u. a. 2018; Cui u. a. 2013.
85 Keller u. a. 2018.
86 Gruber 2018.
87 McCormick 2016; McCormick 2015.
88 Raynaud 2010.
89 Spyrou u. a. 2018a; Keller u. a. 2018.
90 Sallares 2007, 251.
91 Drummond/Rambaut 2007.
92 Bos u. a. 2011.
93 Rasmussen u. a. 2015.
94 Hinnebusch u. a. 2002.
95 Chouikha/Hinnebusch 2014.
96 Hinnebusch 2005.
97 Johnson u. a. 2014.
98 Hinnebusch u. a. 2017.
99 Andrades Valtueña u. a. 2017b.
100 Haak u. a. 2015; Allentoft u. a. 2015; Gaunitz u. a. 2018.
101 Rascovan u. a. 2018.
102 Spyrou u. a. 2018b.
103 Littmann/Littmann 1973, Morens/Littmann 1992 und Harper 2015 etwa diagnostizieren die drei genannten als Pocken-Epidemien.
104 Stephens u. a. 1998.
105 Mecsas u. a. 2004.
106 Sabeti u. a. 2005.

8 Discussion

8.1 Archaeogenetic research on *Yersinia pestis*

The interest in the causative agents of past pandemics is arguably as old as the discipline of microbiology. However, in the last century, scientists were limited to two comparative approaches: retrospective diagnosis, based on contemporaneous symptom descriptions, and palaeoepidemiology, based primarily on quantitative data either from skeletal material or historical records (Sallares, 2007). A third option, palaeopathology, is only applicable for a small number of diseases that leave pathognomonic bone lesions, such as leprosy (Waldron, 2008) or soft tissue lesions in rare cases of mummification. These methodological approaches however depend on three fundamental presumptions: (1) that the historically described disease is indeed a (single) nosological entity in our modern sense, (2) that the disease still exists today, and (3) that it did not undergo pathomorphosis, i.e. the change of clinical manifestations over time (cf. Leven, 2013).

The research on historical plague experienced a series of refutations of the classical narrative of the Second Pandemic as caused by *Y. pestis* (Cohn, 2002; Duncan and Scott, 2005; Twigg, 1984), in fact primarily for the Black Death. Remarkably, the causative agent of the First Pandemic was never as controversial, which might be explained by the scarcity of historical data to fuel alternative hypotheses, or by an overall lower attention.

Archaeogenetic approaches on pathogens promised to end these discussions by offering an independent and supposedly more objective line of evidence for the nature of past pandemics (McCormick, 2007). However, not all historians were as enthusiastic, sometimes even questioning the legitimacy of a modern scientific approach on historical phenomena (Leven, 2013), and early PCR studies were justifiably criticized by peers on inadequate methodology, namely potential contamination and lack of specificity (Gilbert et al., 2004).

Since the publications of the first ancient *Y. pestis* genomes from the Black Death (Bos et al., 2011) and the First Pandemic (Wagner et al., 2014), the involvement of *Y. pestis* in both historical pandemics is virtually undoubted. However, the mere detection of plague in a specific context can still offer new insights into the history: the presence of the Justinianic Plague in Anglo-Saxon England, as an example, was little more than speculation prior to molecular proof presented in Manuscript A. Nevertheless, some caveats expressed on early PCR studies still maintain their justification partially: specificity remains a major challenge regarding low coverage data, and environmental contamination can interfere with phylogenetic analyses.

In both Manuscript A and B, the majority of samples were screened with a qPCR assay prior to target enrichment. As shown in Manuscript A, the qPCR assay has such a high sensitivity, that differentiation between false positive samples and samples close to the detection limit can be diffi-

cult. Also, metagenomic approaches, as used for the samples from Edix Hill (Manuscript A) occasionally produce ambiguous results. For questionable samples after enrichment, was solved with a novel approach comparing chromosomal and plasmidal coverage after capture (Manuscript A), since previously utilized methods, such as competitive mapping to multiple *Yersiniae* (Andrades Valtueña et al., 2017), are not feasible with regard to the capture bias.

The correct identification of true SNPs is particularly a problem of ancient pathogen genomics. DNA extracted from archaeological remains is *per se* metagenomic, often consisting predominantly of environmental DNA. Since bacterial pathogens such as *Yersiniae* often have closely related environmental taxa, misidentification of reads as stemming of pathogenic species is to be expected. One way to verify the authenticity of ancient DNA is the check for DNA damage patterns, i.e. higher C-to-T and G-to-A changes towards the end of the read. However, this supposes that the contaminations are of modern origin and did not accumulate DNA damage, which is commonly assumed for human DNA, where contamination prior to the excavation is not expected. However, the samples were constantly exposed to environmental DNA, so contaminant DNA with damage patterns is likely to be found. Furthermore, libraries for pathogen DNA studies are frequently UDG treated, removing all damaged sites, to facilitate genotyping. Another possibility is to check for 'heterozygosity', so positions were two or more different alleles appear both in a significant frequency (>10%). This was done in both Manuscript A and B (see Fig. S1 for A; Fig. S13 & 14 for B), which gave already first indications for strong contamination in several samples in Manuscript B. But checking for 'heterozygosity' gives only general estimates: 'heterozygous' SNPs are in general not considered for phylogenetic analyses, and especially for low coverage genomes, contamination can also introduce 'homozygous' SNPs as shown in Manuscript A (SI) with an artificial dataset.

In the presented studies, environmental contamination has not been observed to interfere with gross topologies of phylogenetic trees, since introduced SNPs are most often singletons and occasional shared SNPs appear either as homoplastic or are ignored due to missing data in regions with reduced 'mappability'. However, contaminants are frequently observed to introduce false branch lengths or false unique branches, as shown in Fig. S16, Manuscript B. This in turn can lead to wrong conclusions, either by mimicking an accelerated substitution rate (as shown in Manuscript C for Damgaard et al., 2018, caused by the use of a demonstrably contaminated genome) or by fabricating diversity where there is none (as shown in Manuscript B for the Black Death genomes TRP002 and BSS31, cf. Namouchi et al., 2018).

Previously, the classification of SNPs was either done by 'visual inspection' – checking for heterozygous SNPs, unusual peaks or drops in coverage (Feldman et al., 2016) – or with special tools, e.g. searching for clustering of SNPs (Namouchi et al., 2018). Whereas the first approach performs well but is difficult to replicate and can become challenging when handling dozens of ancient genomes, the second approach has only a limited power in detection of false positive SNPs, as shown in Manuscript B.

The challenge of false positive SNPs was accomplished by the development of a new methodology and implementation in a bioinformatic tool, as presented in Manuscript A. A key element of the methodology is the comparative mapping, i.e. mapping with higher and lower stringency. As expected, mapping with higher stringency filters contaminant reads with several mismatches that do not appear in the low stringent mapping, resulting in a lower mean coverage in genomic regions susceptible to contamination. This as well as the presence of coverage gaps and 'heterozygous' SNPs was systematically evaluated for a 50 bp window surrounding all unique and selected shared SNPs with the tool 'SNPEvaluation' (Manuscript A and B).

To validate this approach, the methodology was tested with an artificial dataset, showing an overall satisfying performance, with a maximum sensitivity and an acceptable specificity (Manuscript A, SI). Unfortunately, the comparative mapping is limited to UDG libraries, since DNA damage interferes with the stringency parameter. Therefore, this method was not applied for published non-UDG data (cf. Namouchi et al., 2018 in Manuscript B).

As shown in chapter 1.4.5, even modern *Y. pestis* phylogenies tempted researchers to historical interpretations, so the advancement of ancient pathogen genomics promised to give conclusive answers through evolutionary snapshots of *Y. pestis*. However, the ongoing dispute about plague persistence during the Second Pandemic is only one counterexample. The challenges and pitfalls of phylogeographic interpretations only begin with methodological problems as shown above and will be further discussed in the next chapters.

Coming back to the premises of palaeopathological studies as mentioned in the beginning, research on ancient plague genomes was able to show that concepts of diseases can indeed pursue through millennia and are sometimes attributable to single nosological entities. However, bubonic plague might be a special case in the sense that the clinical manifestations allow recognition from literary sources after centuries or even millennia – already Rufus of Ephesus (fl. 100 CE) observed pestilential *boubones*, arguably the first clear description of plague in history (Mulhall, 2019). This is remarkable with regard to major transformative shifts in medical history, from humorism and miasmatic theory in Antiquity to contagionism and modern microbiology.

The recent discoveries of the Bronze Age plague lineages (Andrades Valtueña et al., 2017; Rascovan et al., 2019; Rasmussen et al., 2015) though, as discussed in Manuscript D, challenge the other two premises: pathogens have indeed evolved and changed their clinical picture over the course of human history (pathomorphosis) and went extinct, so that we cannot always extrapolate from our modern knowledge on diseases to the past.

8.2 The origins of the First and Second Pandemic

From a historical perspective, the question about the origin of pandemics is often connected to political history, be it the Columbian Exchange in the case of syphilis or the siege of Caffa 1346 for the Black Death (see chapter 1.4.2), and therefore embedded in broader narratives. From a biological perspective, it is relevant with regard to the evolutionary history of the causative agents or the co-evolution with vectors and hosts, as shown for *Helicobacter pylori* and humans (Kodaman et al., 2014). In modern, integrative approaches, several lines of evidence are considered: the modern phylogeography of the pathogen; environmental factors such as climate or vector/host habitat; historical records, if available; and palaeopathological evidence, if applicable. These were recently supplemented by inferences through ancient genomes for a number of infectious diseases. Each of these lines has its challenges and pitfalls: Modern phylogeography is susceptible to sampling bias and blind to extinct diversity. Ancient phylogeography might mitigate the latter by recovering extinct lineages but is even more biased in sampling. Environmental factors might change over time (see, e.g., Sallares, 2006 for malaria), and for zoonotic diseases, the determination of 'original' host reservoirs can be difficult (see e.g. Bos et al., 2014 for tuberculosis and Honap et al., 2018 for leprosy). For historical records, there are at least three pitfalls: borrowing of historiographic *topoi*; misinterpretation of toponyms; and political agendas. A notorious problem that is common to all approaches are unavoidable *argumenta ex silentio*.

In an interdisciplinary context as set here, these challenges can even amplify when scientists and scholars of different disciplines are confronted with them, and occasionally can lead to circular arguments. Therefore, a thoughtful application of the concept of consilience, i.e. convergence of evidence, can help to find reasonable answers: by aligning multiple lines of evidence from different, independent sources, the most probable or parsimonious scenario can be determined.

The origin of the Justinianic Plague is a matter of ongoing investigations both by historians and biologists. Procopius reports the first outbreak for Pelusium, an Egyptian port city at the Red Sea. According to Evagrius and Zacharias of Mitylene, plague started in "Ethiopia", John of Ephesus specifies "in the inland peoples of the regions southeast of India, that is, of Kush, the Himyarites, and others" (McCormick, 2007). Historians warned for two reasons for a literal reading of these statements: The "Ethiopian" origin could be a *topos* or even a 'prejudice' that Evagrius borrowed from the influential description of the Plague of Athens by Thucydides (Allen, 1979; Dols, 1974); and the toponym "India" in Late Antique geography does merely translate to a land at the Indian Ocean or the Red Sea (Harper, 2017; Sarris, 2007). Nevertheless, both Africa and Asia were considered as the 'origin' of the First Pandemic.

The African hypothesis goes back at least to Devignat, 1951 and his idea of different biovars, each responsible for one of the three pandemics (see chapter 1.4.5). Although based on 'historical' interpretations, Devignat's 'biological' augmentation was in turn used by some historians (Biraben

and Le Goff, 1969). Devignat was right to assume the *Antiqua* biovar as the ancestral form, but the modern African *Antiqua* strains known today were not introduced before the 14th century, as we know today (Wagner et al., 2014), although this misconception survived long in microbiological literature (Achtman et al., 1999; Achtman et al., 2004). Sarris, 2002 argued for the African hypothesis and against India as origin based on an *argumentum ex silentio* since plague was not reported in China before 610 and Persia was hit by plague after Byzantium, although closer to India. He categorically excluded a third option, the Eurasian steppe, following McNeill's argumentation, who claimed that the steppe reservoir cannot have been established before the 14th century, regarding the plague-free centuries prior to the Black Death (McNeill, 1976). As a second 'biological' argument for an African origin, the existence of a basal lineage called 'Angola' was utilized and suggested as an extant descendant of the Justinianic Plague (Cui et al., 2013). However, this strain is an outlier by all means: it has accumulated an exceptional number of unique SNPs and is the only genome on Branch 0 that has ever been attributed to Africa. Appearing first in 1984 in the U.S. Army Medical Institute for Infectious Diseases (Eppinger et al., 2010), apparently all provenance information on this sample is lost (Green, 2014), casting doubt about any association of 'Angola' with the African state.

The Asian origin hypothesis goes also back to Wu Lien-Teh who identified the Central Asian Plateau as a "huge endemic area" and homeland of plague (Sussman, 2011). Allen, 1979 argued for an Asian origin and introduction through maritime routes from India or Sri Lanka – either through the Persian Gulf or the Red Sea with a potential intermediate station in Ethiopia, later adopted and adapted by other historians (Harper, 2017).

The first genome of the Justinianic Plague finally gave a solid substantiation for an Asian origin (Wagner et al., 2014) since the extant clades (including a recently discovered clade 0.ANT5, see Eroshenko et al., 2017 and Manuscript C), which frame the Justinianic branch in the phylogenetic tree, are found today in Central Asia. The retrieval of an ancient plague genome from a 2nd–3rd-century burial from the Tian Shan region, sharing a short branch with the genomes of the First Pandemic (cf. Damgaard et al., 2018 in Manuscript A and C) is in accordance with this hypothesis, revealing the borderland of modern China, Kazakhstan and Kyrgyzstan the potential birthplace of the lineage causative for the First Pandemic. As shown in Manuscript C, a second low coverage genome from Northern Ossetia, also published by Damgaard et al. 2018, is however not related to this lineage and therefore allows no phylogeographic inferences for the First Pandemic. Therefore, a potential overland route along the steppe in the context of Hunnic migrations, proposed by Damgaard et al., lacks any evidence. So far, the phylogeography does not allow for inferences about the itinerary, but the leading hypothesis of an introduction via the Indian Ocean and the Red Sea is supported by historical data: maritime trade connections to India were well established (Harper, 2017) and the directionality of the Justinianic Plague's spread as reported by John of Ephesus is striking (see chapter 1.4.1).

However, this only answers the question of the ultimate origin, or origin of the causative lineage for the First Pandemic. The estimated divergence date with the Tian Shan genome of 318 BCE to 221 CE (Manucsript C) indicates that the lineage survived for several hundred years in a reservoir prior to the Justinianic Plague, without leaving any modern descendants. Green, 2018 briefly discussed the possibility of an intermediate, secondary reservoir and therefore proximate origin in Northern Africa, more specifically in Ethiopia and the Axumite Kingdom. This scenario has even relevance for pre-Justinianic plague in Antiquity as concluded from Rufus of Ephesus' works for Libya, Egypt, and Syria (Mulhall, 2019 and Manuscript D). The estimated date for these observations of plague symptoms in the first century CE sets it in the realm of possibility that the First Pandemic lineage could be responsible for them, too. However, current phylogeographic data do not allow for any conclusions about the possibility and potential location of a secondary reservoir.

For the Black Death, an Asian origin was commonly agreed on, only the exact location of its emergence and the possible involvement of the Far East during the Black Death are a matter of ongoing discussions. The different hypotheses can basically be narrowed down to four geographic locations: India, China, Central Asia and the Caspian Sea/Caucasus region. The Indian origin is the least substantiated, and Sussman, 2011 is certainly right to assume that the severe outbreaks of the Third Pandemic contributed to this conviction popular in late 19th and early 20th century (cf. Norris, 1977). The first epidemic event in India clearly identifiable as plague dates not before the 17th century, potentially introduced from Europe. In addition, there are no indications of massive population losses in the 14th century (Sussman, 2011).

In his influential book, McNeill constructed a complex theory with an ultimate origin in the Himalayan foothills between India, Myanmar and China (McNeill, 1976). According to his theory, the Mongols caught plague there during their conquests in the 13th century and brought it back to Mongolia. Either from there or the original focus, plague spread to the Chinese province Hebei, causing the initial epidemic, and further westwards along caravan routes through the steppe, finally reaching Crimea 15 years later. A collection of epidemics compiled from the 18th-century Imperial Encyclopedia indeed reports epidemics for 1331–1334, 1344–1346 and the 1350s accompanied by massive population losses (McNeill, 1976), as acknowledged by Sussman, 2011. However, none of the alleged plague epidemic records provides symptom descriptions, which first appear for an outbreak in 1644 (Norris, 1977; Sussman, 2011). In summary, there is no convincing evidence of a Chinese origin of the Second Pandemic based on historical data.

But China was discussed as the origin of *Y. pestis* and therefore ultimate origin of the Black Death also based on modern genomic data (Cui et al., 2013). Due to the geographic origin of the most basal modern branch 0.PE7 and the highest diversity within China, Cui et al. suggested the Qinghai-Tibet Plateau as geographic origin of plague and a potential spread along trade routes. Notably, this study revealed that Branches 1–4 emerged in a "big bang" polytomy (Cui et al., 2013),

and that the Black Death genome (Bos et al., 2011) falls very close to this polytomy in an ancestral position on Branch 1. Hymes embraced this theory and collected supportive historical data for plague epidemics in 13th–14th-century China, but without arguing for a direct link to the European pandemic (Hymes, 2014).

However, Cui et al.'s argumentation was primarily built on the genetic diversity in China and therefore heavily biased. Later studies presented dozens of new genomes from Central Asia and the Caucasus region, namely Armenia, Azerbaijan, Georgia, Kyrgyzstan, Russia, Tadzhikistan (Eroshenko et al., 2017; Kislichkina et al., 2018a, 2018b, 2017, 2015; Kutyrev et al., 2018), and showed a previously unknown diversity in this region, falling on all major branches except Branch 3.

Regarding the Central Asian Plateau, Wu Lien-Teh's characterization as ancient or even original focus of plague set an influential foundation (Norris, 1977). However, one of the most frequently cited evidence for the Central Asian origin of the Black Death is archaeological: the Nestorian cemeteries near Issyk-Kul, excavated in the 1880s (Slavin, 2019). The cemeteries were established in the 13th century and contained around 620 tombstones with 439 dated Syriac inscriptions, revealing a sudden mortality peak between 1338 and 1339, and ten inscriptions dating to these years give "pestilence" as a cause of death. McNeill, 1976 interpreted the Nestorian cemeteries as intermittent station of plague spread westwards by the Mongols, others, such as Benedictow, 2013 and Norris, 1977 dismissed it due to the unclear attribution of the "pestilence" to *Y. pestis*. Slavin, 2019 showed mortality levels similar to those reconstructed of the European Black Death. Environmental shifts, abundance of rodent species and favorable precipitation levels further support the identification as plague (Slavin, 2019). Although the Issyk-Kul region was the heart on a widely spanned trade network, Slavin did however not exclude that the Issyk-Kul epidemic could have been caused by a strain unrelated to the Black Death, regarding the status as an enzootic region today (see also the discussion on the 2nd–3rd-century genome from the same region in Manuscript C).

On the other hand, the high diversity of Branch 0 strains in Central Asia, and more specifically the concentration of 0.ANT1, 0.ANT2, 0.ANT3 and 0.ANT5 strains in the Tian-Shan region (Eroshenko et al., 2017), all emerging prior to the "big bang" polytomy, are a strong argument for the region around Issyk-Kul as starting point of the Second Pandemic (Green, 2018).

Also, a contemporary source was interpreted as supportive of a Central Asian origin: the Arabic historian Ibn al-Wardi testified in 1349 that the pandemic began in the "land of darkness". Dols, 1978, translator of the original source, read it as "inner Asia". However, the presumed geographic location of this land is highly divergent. Slavin, 2019 locates the "land of darkness" in "western Siberia near the Arctic circle". Martin, 1978 in contrast, relying on the works of Ibn Battuta, identifies it as the region of the river Kama in Western Russia. For Norris, 1977 and Benedictow, 2013 it is congruent with the land of the Golden Horde, or southern Russia. The most recent example presented by Namouchi et al., 2018 illustrates how the ambiguity of topographic terms can tempt arbitrary interpretations: Although citing Martin, 1978 and his identification of the Kama

river region as the "land of darkness", they allege that Ibn al-Wardi meant the Caspian Sea region in lack of geographic knowledge – and in accordance with their hypothesis.

Arguments against a Central Asian origin were primarily in delineation and favor of the Caucasus/Caspian Sea hypothesis. Benedictow, 2013 cited a Russian chronicle reporting a plague outbreak in 1346 stretching from the northwestern shores of the Caspian Sea to Crimea, and Nikephoros Gregoras, a contemporary Greek historian, attesting that the plague came from the shores of the Sea of Azov and the river Don. Both Norris and Benedictow argue strongly against an origin east of the Caspian Sea (be it Central Asia, India or China) because of the vast distances and the aggravated trade through the land of the Golden Horde. After the death of Kublai Khan in 1259, the Mongol Empire split into the Il-Khanate (Persia), the Kipchak-Khanate (Southern Russia), the Chagatai-Khanate (Central Asia) and the Yuan Dynasty (China); and the first three gradually converted to Islam (1295–1313). This severely affected trade between Europe and China through Central Asia due to conflicts of the Mongolian successor states, but also because the Islamic Khanates became intolerant towards Christian and Persian merchants, which finally led to the siege of Caffa (Norris, 1977; Benedictow, 2013). Thus, both are contesting the narrative of a "ready-made pathway for propagation" by a "network of cavanserais" proposed by McNeill, 1976. Benedictow sees this strong influence of political borders reflected in the limited spread of the Black Death through Russia 1346–1348, stopping right before Nizhny Novgorod at the border of the Golden Horde (Benedictow, 2004).

Without providing archaeogenetic data from the relevant geographic region, Namouchi et al., 2018 published along with ancient genomes of the Black Death and the *pestis secunda* (1356–1363) their hypothesis of a secondary reservoir established prior to the Black Death, extending from the western shores of the Caspian Sea to Southern Russia. Furthermore, they propose this reservoir not only as proximate origin of the Black Death, but also of recurrent plague introductions into Europe facilitated by fur trade.

Based on the current state of archaeogenetic research alone, the question of the origin of the Black Death cannot be satisfiably answered. The genome from Laishevo (Manuscript B), one SNP ancestral to the Black Death and one derived from the "big bang" polytomy, is congruent with the presumed spread of plague along the lower Volga through the Kipchak-Khanate of the Golden Horde (Benedictow, 2013). Hence, the additional SNP, present in the Black Death genomes, would have been acquired somewhere on the way from the lower Volga to the European coasts of the Mediterranean Sea (see Abbadia San Salvatore, reanalyzed in Manuscript B; and Barcelona in Spyrou et al., 2016), therefore the Laishevo genome supports the uncontested Eastern European entry of the Black Death. The Bolgar genome (Spyrou et al., 2016), presumably dating to the late *pestis secunda* (1362–1366, cf. Namouchi et al., 2018), is more enlightening in the sense that it demonstrates the permeability of the border between the Kipchak-Khanate and the Russian principalities for plague.

The low coverage plague genome from Northern Ossetia (Damgaard et al., 2018), analyzed in Manuscript C, might hold clues for the involvement of the Caucasus region in the early Black Death, be it either as origin or passage. However, also the possibility that this ancient strain belongs to Branch 2 should not be dismissed.

More insightful about a possible origin is the distribution of modern diversity. So far, only 0.PE2 and 2.MED strains were isolated in the Caucasus region, so neither strains branching directly ancestral (e.g. 0.ANT3) nor directly derived (Branch 1) to the Black Death (Kutyrev et al., 2018). Hence, a supposed secondary reservoir must have been cleared from these strains in the last six centuries, premised that the modern diversity in the Caucasus region is fully explored. The constrained distribution of the modern 0.ANT strains, emerging basal to the "big bang" polytomy on Branch 0, in the Tian Shan and Alai mountains (Eroshenko et al., 2017) is indicative of this region as an ultimate origin of the Black Death strain. Moreover, it is a strong alternative hypothesis to the Qinghai-Tibet Plateau as location of the polytomy event: neglecting the Branch 1 distribution that was presumably established by a post-Black Death strain travelling from Europe to Asia (Spyrou et al., 2016) and ultimately to Africa (Green, 2018), and regarding the widespread distribution of Branch 2 strains (Cui et al., 2013; Kutyrev et al., 2018), the concentration of Branch 3 and 4 strains in Mongolia, Gansu and the Altai region (Cui et al., 2013; Kutyrev et al., 2018) is also compatible with an eastward spread from the Tian Shan region.

Ultimately, only future archaeogenetic analyses in the suspected regions might reveal where the Black Death originated as a pandemic. And only with ancient DNA will we finally be able to trace down potential historical plague epidemics in Asia, connected to or independent of the pandemics in Western Eurasia.

8.3 Progression and persistence of the First and Second Pandemic

Scientific and scholarly debates about the progression and persistence of the historical pandemics arose primarily from two circumstances: the disappearance of plague from Europe in the 18th century and presumably between 750 and 1346, and the zoonotic nature of plague. Therefore, the leading question is whether *Y. pestis* established local reservoirs in Europe or the Mediterranean basin – eventually becoming extinct twice – or if it was regularly reintroduced over several centuries.

The scenario of a local persistence of plague in Europe or the Mediterranean basin requires either a reservoir in commensal rats or wild rodents, or alternative models of maintenance and transmission. The presence of the black rat (*Rattus rattus*) in the temporal and geographic range of both pandemics was repeatedly questioned, since identification in contemporary literary sources is difficult; and in archaeological excavations, rat bones were regularly dismissed as signs for recent perturbations or simply not found due to their small size (Davis, 1986; McCormick, 2003).

Nevertheless, it is commonly accepted that the black rat, originally indigenous to Asia (Prendergast et al., 2017) reached the Mediterranean basin latest in Antiquity (Audoin-Rouzeau and Vigne, 1997; Davis, 1986; McCormick, 2003), probably introduced via the Indian Ocean (Prendergast et al., 2017) and the Red Sea (McCormick, 2003). Presumably with the expansion of the Roman Empire, the black rat spread across Western Europe along major commercial routes (Audoin-Rouzeau and Vigne, 1997). Also, for northwestern Russia – relevant for the Second Pandemic –, skeletal remains of black rats date back to the 12th century (Savinetsky and Krylovich, 2011). Moreover, as well known from Madagascar, *Rattus rattus* can serve as a long-term sylvatic reservoir, although with *Synopsyllus fonquerniei* as vector (Andrianaivoarimanana et al., 2013), and statistical modelling has shown that relatively small rat metapopulations allow for the persistence of plague (Keeling and Gilligan, 2000). The clear link of plague dissemination with ships, as reported both for the First and Second Pandemic (Benedictow, 2004; Harper, 2017) further supports hypothesis of the black rat as reservoir during epidemics.

Overall, however, the archaeological evidence for the black rat in premodern Europe is scarce. The most comprehensive database to date, "Archaeological Finds of Rats: AD 1–1500" (McCormick et al., 2013), lists in total 220 features, and neither for the First, nor the Second Pandemic contemporary records take note of rat epizootics, as observed during the Third Pandemic (Davis, 1986). Therefore, several researchers concluded that rats cannot have been the reservoir or intermediate host during the Second Pandemic (Hufthammer and Walløe, 2013; Karlsson, 1996).

Alternative models for persistence without a mammalian reservoir focus mainly on the human flea (*Pulex irritans*) or the human louse (*Pediculus humanus*). *Y. pestis* was sporadically isolated from *P. irritans* during modern plague outbreaks (Dennis et al., 1999), but it is in general considered as a poor vector (Hinnebusch et al., 2017). Nevertheless, studies tried to find evidence for the *P. irritans* hypothesis through historical and archaeological considerations (Hufthammer and Walløe, 2013) as well as epidemiological modelling (Dean et al., 2018). Drancourt et al., 2006 on the other hand argued for transmission via the human louse, although so far only observed in animal models (Ayyadurai et al., 2006). Finally, it remains uncertain how *Y. pestis* could persist in humans and their ectoparasites for several years, decades or even centuries.

As a third option, also sylvatic rodent reservoirs have been proposed. Carmichael, 2014 found supportive evidence for her hypothesis of the European marmot (*Marmota marmota*) as a reservoir in the Alps during the Second Pandemic; Pribyl, 2017 discussed the common vole (*Microtus arvalis*) as a possible host in Late Medieval England. For the Near East, where plague is prevalent today in the highlands of Iran and border regions, Varlık, 2014 suggested persistence in gerbils and other rodents, seeding the numerous outbreaks in the Ottoman Empire.

Although Harper, 2017 briefly mentions Varlık's hypothesis of sylvatic reservoirs in Ottoman Anatolia, alternatives to the conventional view of rats as central agents for persistence and transmission were never discussed extensively for the First Pandemic and most historians follow the conventional rat-flea transmission model (Harper, 2017; McCormick, 2003; Sallares, 2007).

Equally, discussion about persistence versus reintroductions focused primarily on the Second Pandemic. One notorious problem in this dispute though is the Eurocentrism in plague studies, as diagnosed by Varlık, 2017 and 2014, neglecting contemporaneous epidemic events in the Near East and Northern Africa. For the Second Pandemic, this Eurocentrism is to be found on both sides, arguing for reintroductions from Asia (Schmid et al., 2015) or persistence in Europe (Bos et al., 2016; Seifert et al., 2016; Spyrou et al., 2016), even in cases such as the Great Plague of Marseille 1720–1722, where the spread from the Near East via maritime trade is well documented and can be narrowed down to a single ship, the *Grand Saint Antoine* (Devaux, 2013). The problem becomes apparent with regard to a possible reservoir in the Caucasus region. Whereas arguing previously for reintroductions from Central Asia (Bramanti et al., 2016; Schmid et al., 2015), the same group of researchers recently proposed a reservoir in the Caspian Sea region (Namouchi et al., 2018). By redefining Europe as Western Europe, their results ostensibly seem to support their previous claims about reintroductions, although they are technically supporting the opposite scenario of persistence in (Eastern) Europe (Bos et al., 2016; Seifert et al., 2016; Spyrou et al., 2016).

The question of persistence versus reintroduction has been approached essentially by two methods: genetics and epidemiology. Schmid et al., 2015 used a modelling approach on epidemiological data of the Second Pandemic in combination with climatic data, as previously utilized for the Third Pandemic (Sun et al., 2019; Xu et al., 2014); however, the study suffers from its underlying dataset, primarily consisting of the digitized plague occurrences collected by Biraben, 1975. Roosen and Curtis, 2018 criticized the uncritical use of this dataset due to the geographic bias and the obscure body of original sources, and it is striking that all of the alleged reintroductions are found in the geographic periphery of Biraben's dataset. In the study by Namouchi et al., 2018, the phylogenetic structure of the previously (Bos et al., 2011; Spyrou et al., 2016) and newly sequenced ancient plague genomes of the *pestis secunda* was interpreted as congruent with reintroductions from a reservoir in the Caspian Sea region through fur trade, specifically via a northern route through the Volga region and Novgorod. However, this ignores evidence from the Biraben dataset used by the same group in Schmid et al., 2015: Biraben clearly identified Germany as a starting point of the *pestis secunda* in 1356 and a radial spread to Brabant in 1356, England in 1361 and Russia in 1359. This is a much more parsimonious explanation for the phylogenetic structure with the Bergen op Zoom genome in the ancestral position and the genomes from London St. Mary Graces and Bolgar as direct descendants on separate branches (Manuscript B). Remarkably, Schmid et al., 2015 identified the outbreak in Germany in 1356 as potentially associated with a wildlife reservoir but dismissed it due to missing links to local climate fluctuations. Finally, as discussed above, the modern diversity of

plague strains in the Caucasus gives no indication for its function as an origin or reservoir during the Second Pandemic.

Regarding the Archaeogenetic evidence for the Second Pandemic, the fact that the Black Death strain is directly ancestral to all post-Black Death strains sequenced so far (including the *pestis secunda*) strongly supports a local persistence, either in Europe or the nearby affected regions in North Africa and the Near East (Manuscript B). The later diversification into two clades (Manuscript B) might even indicate the existence of multiple reservoirs. Furthermore, the coincidence of the extinction of the post-Black Death clade and the disappearance of plague in Europe in the 18th century is well explained by clearance of local reservoirs from *Y. pestis*. Other scenarios would require selective reintroductions of strains of a single, today extinct clade from a region such as Central Asia, where strains of other clades (e.g. of Branch 0) were most likely circulating at the same time (Eroshenko et al., 2017). Due to the high mobility of *Y. pestis* while accumulating little to no genetic diversity, as showcased by the Black Death (Manuscript B), as well as the strong sampling bias, the potential location of one or multiple reservoirs in Europe or close-by remains for now inexplicable by archaeogenetic data.

For the First Pandemic, the newly discovered diversity, likely emerging already during the Justinianic Plague (541–544) and certainly in the later epidemics (Manuscript A), is a strong indication for local persistence. Similar to the Second Pandemic, all ancient genomes of the First Pandemic found so far are descendants of a single strain, and the clade is extinct today – again a coincidence with the supposed disappearance of plague from Europe and the Mediterranean basin in the 8th century, potentially explained by clearance of a local reservoir. Also, for the First Pandemic, the location of a possible reservoir cannot be concluded from archaeogenetic data. However, Harper, 2017 described focalizations in time and space as he outlined a Byzantine Phase (542–600) and an Iberian Phase (660–749) for the West while seeing continuous presence in Syria, perhaps indications of multiple local reservoirs.

A striking discrepancy in the phylogenies of both historical pandemics is the polytomy observed for early First Pandemic (Manuscript A) which has no equivalent in the Second Pandemic (Manuscript B). Neglecting the basal strain from Laishevo, no genetic diversity has so far been consistently described for the Black Death, and the emerging lineages can be completely resolved in bifurcations (except one trifurcation, Manuscript B). This could either be an artefact due to unsampled or undetected diversity during the early Second Pandemic, or indeed a difference in the evolutionary history of both pandemics. An early and wide geographic dispersal (e.g. across the Mediterranean Sea) is historically documented for both the Justinianic Plague and the Black Death. A slower pace of long-distance transmission for the Late Antiquity might explain why *Y. pestis* could acquire this microdiversity seen in the First Pandemic strains (Manuscript A), but remains speculative without futher investigation.

Two more observations made through ancient DNA analyses are remarkable regarding the First and Second Pandemic progression: the large genomic deletion observed in strains associated with later phases of both pandemics, and an accelerated substitution rate observed in the same Second Pandemic clade carrying the deletion (Manuscript A and B). Whether there is a causal relationship between these two observations for the Second Pandemic clade remains provisionally unclear – none of the affected genes could be identified as associated with a mutator gene. A determination of the substitution rate for the respective First Pandemic lineage was not attempted yet due to lack of precise dating of the later strains (Saint-Doulchard, Manuscript B), but could help answering this question.

The observed deletions of more than 45 kb are almost identical, the one observed in London/Marseille extends the Lunel-Viel/Saint-Doulchard deletion for 3.9 kb on one end. Pseudogenization and genome reduction are well known phenomena in *Y. pestis* (Chain et al., 2004), but it is striking that a deletion of this size and location was not observed in any modern strain of the comparative dataset (Manuscript A and B). This opens the question about a potential adaptive character of this mutation with regard to a specific ecological niche, since both lineages distinguish themselves from other strains as they stem from epidemic contexts in Europe.

A closer look into the deleted genomic region does – without functional testing – not solve the conundrum. Two of the affected genes, *mgtB* and *mgtC*, have previously been identified as virulence factors (Zhou and Yang, 2009). As Magnesium transporters they are crucial for invasion and maintenance in Mg^{2+}-deficient environments such as macrophages, relevant for immune evasion in early phases of infection of mammalians (Ford et al., 2014). As part of the PhoPQ regulon, they might also be involved in biofilm formation in the flea foregut (Rebeil et al., 2013). However, also the neighboring genes should be considered: compiling genes involved primarily in chemotaxis and motility – remnants of the enterobacterial lifestyle of *Y. pestis*' ancestors – as well as a number of hypothetical and pseudogenes, loss of such a region follows the general trend of genome reduction in *Yersiniae* and specifically *Y. pestis* (McNally et al., 2016).

The functional consequences of the deletion of both Magnesium transporters in *Y. pestis* as well as the neighboring regions still have to be explored, but the epidemiological data for the Great Plague of Marseille, in total around 100,000 victims (Devaux, 2013), do not suggest a loss of virulence (Slack, 1981). Considering all ancient genomes of the First and Second Pandemic, there are no clear indications of a change of transmissibility or lethality over the course of both pandemics that could potentially explain the eventual disappearance of plague from Europe in the 8th and 18th centuries.

Surprisingly, as Sallares, 2007 noted, "there is an almost complete absence of theories to explain the end of the first pandemic". This is remarkable, since several theories about the disappearance of plague from 18th-century Europe build on processes specific for this time and are therefore not transferable to the end of the First Pandemic (e.g., rat poisoning with arsenic, cf. Konkola, 1992).

Of course, the reasons or the interplay of multiple processes do not have to be identical, but regarding the striking similarities in the progression of both pandemics, a common explanatory model is more parsimonious than postulating two independent and causally unique events.

The major theories concentrate on humans, rats, or the causative agent *Y. pestis* itself. Regarding the pathogen itself, several researchers speculated about a loss of virulence. Such ideas as proposed by McNeill, 1976 can be retraced primarily to studies of the 1930s when bacteriologists claimed to have observed transformations of *Y. pestis* to *Y. pseudotuberculosis* (Pollitzer, 1952c), evidently in a time when their phylogenetic relationship was not clear yet. Even cross-immunization through *Y. pseudotuberculosis* (discussed, e.g., in McCormick, 2007 citing Devignat, 1953) was proposed, and promising efforts to develop live vaccines based on *Y. pseudotuberculosis* provisionally support this idea (Blisnick et al., 2008). However, this theory does not explain why this would happen selectively in Europe, regarding the Asian origin and early spread of *Y. pseudotuberculosis* (Seecharran et al., 2017), and as discussed above, the Great Plague of Marseille does not support a theory of loss-of-virulence or increased immunity.

Appleby, 1980 attributed the local extinction to immunity of rats. This is essentially built on observations in rat populations during epizootics/epidemics in India showing a clear increase of immunity (Pollitzer, 1952a). However, the (acquired) immunity is likely to decrease fast in plague-free years, especially regarding the short generation time of rats (Slack, 1981). The hypothesis of immunity of humans as sole factor is incompatible with a zoonotic disease, since no herd immunity leading to eradication can be reached (Keeling and Gilligan, 2000), and therefore was only seriously propagated in context of alternative etiological theories (Cohn, 2003). However, human immunity was occasionally discussed as a contributing factor or to explain epidemiological patterns (Ell, 1984).

Another theory emerged from the observed coincidence of a spread of the brown rat (*Rattus norvegicus)* and gradual replacement of the black rat *(Rattus rattus)* in the 18th century, first elaborated by Hirst, 1953. This theory builds on one hand on the behavioral differences of both species but also on their parasites. The black rat is known to stay closer to human settlements, to climb on trees as well as roofs and to go on ships (Davis, 1986; Hufthammer and Walløe, 2013), and while the black is thought to carry *Xenopsylla cheopis*, the flea of the brown rat, *Nosopsyllus fasciatus*, is a rather poor vector for *Y. pestis*. Although this 'replacement theory' is not applicable to the First Pandemic, Audoin-Rouzeau and Vigne, 1994 coincidentally observed a disappearance of rats in the 9th century for yet unclear reasons.

Finally, also human agency was suggested as a crucial factor, namely improvement of sanitation and establishment of quarantine measures and *cordons sanitaires* (Rothenberg, 1973; Slack, 1981). Documented first in Ragusa (modern Dubrovnik, Croatia) during a plague outbreak in 1377, the isolation of sick citizens and visitors in a place outside the city walls for thirty days prove to be effective and was increasingly adopted by other (primarily port) cities, eventually prolonged to 40 days (therefore called *quarantine*; Sehdev, 2002). *Cordons sanitaires*, i.e. militarily enforced isolation

from plague-infested regions along borders, often in combination with quarantine stations for controlled trade and mobility, were established in the 18th century, most prominently at the Austro-Hungarian border to the Balkan region. The effectiveness of this *cordon sanitaire* is controversial among historians, since its increasing success might merely coincide with a general retreat of plague from Europe due to other reasons (Eckert, 2000; Rothenberg, 1973). An important prerequisite for these theories though is that plague was introduced to the respective ports or regions from outside, and therefore are incompatible with Central European reservoirs. And indeed, Eckert, 2000 takes "the continued existence of endemic loci and periodic outbreaks in the Balkans and the Near East" for granted. Neither quarantine measures nor *cordon sanitaires* are known for the 8th century and therefore cannot explain the retreat of plague at the end of the First Pandemic.

The explanatory power of archaeogenetic approaches regarding the disappearance of plague has so far been limited. Significant changes in the genetic make-up, acquired during the pandemics, of the pathogen have only been found recently (see above and Manuscript A and B) and cannot yet be interpreted conclusively. Even for larger genetic differences concerning well-characterized virulence factors – as seen in Neolithic and Bronze Age plague lineages – conclusions on virulence or mode of transmission are challenging (Manuscript D). Remains of rats and other commensal or wild rodents have not been archaeogenetically investigated yet, not least because of their scarcity and small size.

In the past, attempts to find traces of an increased immunity in humans have focused on the CCR5-Δ32 mutation, providing protection against HIV, and its relatively high frequency in Europe. But both studies on a potential plague immunity as well as the evolutionary history of this mutation could not confirm this hypothesis, as discussed in Manuscript D. However, a recently discovered mutation in the immune-cell receptor FPR1 that provides protection against *Y. pestis* and occurs in higher frequency in Europeans and Asians (Osei-Owusu et al., 2019) opens new perspectives for research on plague immunity in rodents and humans.

9 Concluding Remarks and Outlook

The field of ancient pathogen genomics can contribute invaluable insights into the evolutionary history of pathogens, as shown here for *Yersinia pestis*. Recovery of ancient genomes can reveal today extinct diversity and gives a temporal dimension to phylogeographic studies, reshaping our ideas of origins and dispersals of pathogens.

Historical interpretations of these findings however extend beyond the expertise of pure evolutionary biology and require careful consilient approaches, considering and integrating both confirmatory and contradictory lines of evidence of other disciplines, such as history, archaeology or climate science.

Several caveats and limitations of current integrative, archaeogenetic approaches on the history of diseases were already discussed in the previous chapters: all disciplines struggle with missing data and are occasionally forced to mount *argumenta ex silentio,* especially regarding the origin and spread of pandemics; neither the question of reservoir and vector species in historical times, nor of potential locations of their sylvatic foci can yet be answered conclusively; and the reasons for the retreat of plague from Europe, at least two times in history, remains elusive.

Although conventional ancient pathogen genomics has the potential to uncover previously unknown or ambiguous outbreaks, to narrow down temporal and spatial windows for the dispersal of pandemics, and to reveal fine-scale evolutionary trajectories, many of the open questions are intangible by this discipline alone. Possible interactions of plague with other pathogens, e.g. as a syndemic, are currently understudied, and certain agents such as RNA viruses will require new methodological approaches. Questions about reservoirs and vectors can only be answered by studying their natural history in interplay with climate and anthropogenic activities. Epidemiological inferences require comprehensive anthropological examinations of human remains. Immunity and resistance in humans and animals can only be investigated through population genetic studies. Adaptive evolution can hardly be tracked without functional testing *in vivo*. Finally, the collaboration with historical disciplines has to be consolidated further, facilitated by more quantitative studies and digitization of existing datasets.

However, none of these outlined future directions will be able to write a comprehensive history of plague if they do not overcome the current 'blind spot', which is in fact more a 'peripheral blindness': the broad ignorance of the history of plague outside of Europe. Both the First and Second Pandemics are still primarily phenomena of 'Western' history and only peripherally of Northern Africa and the Middle East. Potential pre-Modern epidemic or pandemic events in Central and East Asia or (sub-Saharan) Africa are vastly unexplored and yet to be integrated into the classical narratives. This is an inherent bias of almost all disciplines and apparent even in basic concepts such as the 'three pandemics' which clearly emerged from a Western perspective. Although the attribution of

ancient and modern genomes of these pandemics to distinct clades provisionally supports this classification, it is hardly applicable for the continued presence of plague in the Near East after the Second Pandemic, not to mention potential epidemic and pandemic events in Central and East Asia. Remarkably, modern medical and biological plague research has rather the opposite bias: the strong sampling bias on sequenced modern genomes towards China has been overcome only gradually in recent years.

The objective of a 'global' history of plague holds new challenges for both the humanities and natural sciences, including archaeogenetics: linguistic and disciplinary boundaries impede scholarly exchange and comparative studies, intercontinental scientific collaborations are still more an exception than the rule. Furthermore, human remains as research objects of archaeogenetics are particularly sensitive, interfering with religious, cultural and political interests. Irrespective of that, the success of archaeogenetic research is highly dependent on the archaeological research and, regarding the DNA preservation, climatic conditions of target regions, further hampering an unbiased approach.

Nevertheless, *Yersinia pestis* has become a 'model organism' to study infectious diseases in the context of human history for good reasons; and also future research, aware of the pitfalls and caveats of inevitable biases, will provide meaningful insights into the evolution of this pathogen and the complex interplay with its natural environment and human agency.

References

Achtman M et al., 2004. Microevolution and history of the plague bacillus, Yersinia pestis. *Proc Natl Acad Sci USA* 101:17837–17842.

Achtman M et al., 1999. Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. *Proc Natl Acad Sci USA* 96:14043–14048.

Adjemian JZ et al., 2007. Initiation and spread of traveling waves of plague, Yersinia pestis, in the western United States. *Am J Trop Med Hyg* 76:365–375.

Alapont Martín L, Ribera i Lacomba AV, 2009. Topografía y jerarquía funeraria en la Valencia tardo-antigua. Oxford. pp. 59–88.

Alexander JT, 2003. Bubonic Plague in Early Modern Russia. Public Health and Urban Disaster. Oxford: Oxford University Press.

Alexander JT, 1986. Reconsiderations on plague in early modern Russia, 1500-1800. *Jahrb Gesch Osteur* 34:244–254.

Allen P, 1979. The "Justinianic" Plague. *Byzantion* 49:5–20.

Allentoft ME et al., 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.

Andrades Valtueña A et al., 2017. The Stone Age Plague and Its Persistence in Eurasia. *Curr Biol* 27:3683–3691.e8.

Andrianaivoarimanana V et al., 2013. Understanding the Persistence of Plague Foci in Madagascar. *PLoS Negl Trop Dis* 7:e2382–e2382.

Appleby AB, 1980. The Disappearance of Plague: A Continuing Puzzle. *Econ Hist Rev* 33:161–173.

Audoin-Rouzeau F, 2003. Les chemins de la peste. Rennes: Presses universitaires de Rennes.

Audoin-Rouzeau F, Vigne J-D, 1997. Le rat noir (Rattus rattus) en Europe antique et mediévale: Les voies du commerce et l'expansion de la peste. *Anthropozoologica* 25:399–404.

Auerbach RK et al., 2007. Yersinia pestis evolution on a small timescale: Comparison of whole genome sequences from North America. *PLoS One* 2:e770.

Austin JJ et al., 1997. Problems of reproducibility — does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society B: Biological Sciences* 264:467–474.

Ayyadurai S et al., 2006. Experimental model to evaluate the human body louse as a vector of plague. *Emerg Infect Dis* 194:1589–1596.

Bacot AW, Martin CJ, 1914. Reports on Plague Investigations in India. LXVII. Observations on the mechanism of the transmission of plague by fleas. *J Hyg* (Plague Suppl. 13):655–664.

Belon et al., 2015. A Macrophage Subversion Factor Is Shared by Intracellular and Extracellular Pathogens. *PLoS Pathog* 11:e1004969.

Ben-Ari T et al., 2012. Identification of Chinese plague foci from long-term epidemiological data. *Proc Natl Acad Sci USA* 109:8196–8201.

Benedictow OJ, 2016. The Black Death and Later Plague Epidemics in the Scandinavian Countries: Perspectives and Controversies. Warsaw, Berlin: Gruyter.

Benedictow OJ, 2010. What Disease was Plague? On the Controversy over the Microbiological Identity of Plague Epidemics in of the Past. Leiden: Koninglijke Brill NV.

Benedictow OJ, 2004. The Black Death 1346-1353: The Complete History. Woodsbridge: Boydell & Brewer.

Bergdolt K, 1994. Der Schwarze Tod in Europa: Die Große Pest und das Ende des Mittelalters. München: C. H. Beck.

Biraben JN, 1975. Les hommes et la peste en France et dans les pays européens et méditerranéens. Paris: Mouton.

Biraben JN, Le Goff J, 1975. The Plague in the Early Middle Ages. In: Forster R, Ranum O, editors. *Biology of man in history.* Baltimore, London: The Johns Hopkins University Press. pp. 48–80.

Biraben JN, Le Goff J, 1969. La Peste dans le Haut Moyen Age. *Annales Histoire, Sciences Sociales* 24:1484–1510.

Blanc-Potard AB, Groisman EA, 1997. The Salmonella selC locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J* 16:5376–5385.

Bland DM et al., 2018. Infectious blood source alters early foregut infection and regurgitative transmission of Yersinia pestis by rodent fleas. *PLoS Pathog* 14:e1006859.

Blisnick T et al., 2008. Oral vaccination against bubonic plague using a live avirulent Yersinia pseudotuberculosis strain. *Infect Immun* 76:3808–3816.

Bondioli L et al., 2016. Plasmodium falciparum malaria in 1st–2nd century CE southern Italy. *Curr Biol* 26:R1220–R1222.

Bos KI et al., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514:494–497.

Bos KI et al., 2016. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *Elife* 5:e12994.

Bos KI et al., 2011. A draft genome of Yersinia pestis from victims of the Black Death. *Nature* 478:506–510.

Bramanti B et al., 2016. Plague: A Disease Which Changed the Path of Human Civilization In: Yang R, Anisimov A, editors. *Yersinia Pestis: Retrospective and Perspective.* Dordrecht: Springer Netherlands. pp. 1–26.

Brandt G et al., 2013. Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity. *Science* 342:257–261.

Briggs AW et al., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA* 104:14616–14621.

Briggs AW et al., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38:e87–e87.

Brotherton P et al., 2007. Novel high-resolution characterization of ancient DNA reveals C>U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35: 5717–5728.

Brygoo ER, 1966. Epidémiologie de la peste à Madagascar. *Arch Inst Pasteur Madagascar* 35:9–147.

Büntgen U et al., 2012. Digitizing historical plague. *Clin Infect Dis* 55:1586–1588.

Burki T, 2017. Plague in Madagascar. *Lancet Infect Dis* 17:1241–1241.

Camus A, 1947. La Peste. Paris: Gallimard.

Carmichael AG, 2014. Plague Persistence in Western Europe: A Hypothesis. In: Green MH, editor. *The Medieval Globe, v. 1. Pandemic Disease in the Medieval World: Rethinking the Black Death.* Kalamazoo, MI: Bradford. pp. 157–192.

Chain PSG et al., 2004. Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. *Proc Natl Acad Sci USA* 101:13826–13831.

Chain PSG et al., 2006. Complete genome sequence of Yersinia pestis strains Antiqua and Nepal516: Evidence of gene reduction in an emerging pathogen. *J Bacteriol* 188:4453–4463.

Chouikha I, Hinnebusch BJ, 2014. Silencing urease: A key evolutionary step that facilitated the adaptation of *Yersinia pestis* to the flea-borne transmission route. *Proc Natl Acad Sci USA* 111:18709–18714.

Cingolani P et al., 2012. 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.

Claiborne Stephens J et al., 1998. Dating the Origin of the CCR5-D32 AIDS-Resistance Allele by the Coalescence of Haplotypes. *Am J Hum Genet* 62:1507–1515.

Cohn SK, 2003. The Black Death Transformed. London: Hodder Education.

Cohn SK, 2002. The Black Death: End of a Paradigm. *Am Hist Rev* 107:703–738.

Cui Y et al., 2013. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc Natl Acad Sci USA* 110:577–582.

Cummins N et al., 2016. Living standards and plague in London, 1560–1665. *Econ Hist Rev* 69:3–34.

Dabney J et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA* 110:15758–15763.

Dalitz S et al., 2012. Das kleinste Massengrab Brandenburgs. Drei Tote aus dem Dreißigjährigen Krieg auf der Dominsel der Stadt Brandenburg an der Havel. *Jahresbericht des Historischen Vereins Brandenburg (Havel)* 21.

Damgaard P de B et al., 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* 557:369–374.

Darling AE et al., 2008. Dynamics of genome rearrangement in bacterial populations. *PLoS Genet* 4:e1000128.

Davis DE, 1986. The Scarcity of Rats and the Black Death: An Ecological History. *J Interdiscip Hist* 16:455–470.

Dean KR et al., 2018. Human ectoparasites and the spread of plague in Europe during the Second Pandemic. *Proc Natl Acad Sci USA* 115:1304–1309.

Demeure CE et al., 2019. Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes & Immunity* 20:357–370.

Dennis DT et al., 1999. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: World Health Organization.

DePristo MA et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.

Derbise A, Carniel E, 2014. YpfΦ: a filamentous phage acquired by Yersinia pestis. *Front Microbiol* 5:701–701.

DeSalle R et al., 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* 257:1933–1936.

Devaux CA, 2013. Small oversights that led to the Great Plague of Marseille (1720-1723): lessons from the past. *Infect Genet Evol* 14:169–185.

Devignat R, 1953. La Peste antique du Congo belge dans le cadre de l'histoire et de la géographie, Mémoires. Collection in 8. Tome XXIII, fasc. 4. Brussels: Institut royal colonial belge.

Devignat R, 1951. Variétés de l'espéce Pasteurella pestis; nouvelle hypothése. *Bull World Health Organ* 4:247–263.

DeWitte SN, 2010. Age patterns of mortality during the Black Death in London, A.D. 1349-1350. *J Archaeol Sci* 37:3394–3400.

Dols MW, 1978. Geographical origin of the Black Death: comment. *Bull Hist Med* 52:112–120.

Dols MW, 1977. The Black Death in the Middle East. Princeton: Princeton University Press.

Dols MW, 1974. Plague in Early Islamic History. *J Am Orient Soc* 94:371–383.

Dooley A, 2007. The Plague and Its Consequences in Ireland. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 215–228.

Drancourt M et al., 1998. Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. *Proc Natl Acad Sci USA* 95:12637–12640.

Drancourt M et al., 2006. Yersinia pestis as a telluric, human ectoparasite-borne organism. *Lancet Infect Dis* 6:234–241.

Drancourt M et al., 2004. Genotyping, Orientalis-like Yersinia pestis, and plague pandemics. *Emerg Infect Dis* 10:1585–1592.

Drancourt M et al., 2007. Yersinia pestis Orientalis in remains of ancient plague patients. *Emerg Infect Dis* 13:332–333.

Drummond AJ, Rambaut A, 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 8. doi:10.1186/1471-2148-7-214

Drummond AJ et al., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–1192.

Duggan AT et al., 2016. 17th Century Variola Virus Reveals the Recent History of Smallpox. *Curr Biol* 26:15787–15792.

Duncan CJ, Scott S, 2005. What caused the Black Death? *Postgrad Med J* 81:315–320.

Durliat J, 1989. La peste du VIe siècle. Pour un nouvel examen des sources byzantines. In: Abadie-Reynal C, Kravari V, Lefort J, Morrisson C, editors. *Hommes et richesses dans l'empire byzantine.* Paris: Lethielleux. pp. 107–119.

Eckert EA, 2000. The Retreat of Plague from Central Europe, 1640–1720: A Geomedical Approach. *Bull Hist Med* 74:1–28.

Eisen RJ et al., 2015. The Role of Early-Phase Transmission in the Spread of Yersinia pestis. *J Med Entomol* 52:1183–1192.

Ell SR, 1984. Immunity as a factor in the epidemiology of medieval plague. *Rev Infect Dis* 6:866–879.

Eppinger M et al., 2009. Draft genome sequences of Yersinia pestis isolates from natural foci of endemic plague in China. *J Bacteriol* 191:7628–7629.

Eppinger M et al., 2010. Genome sequence of the deep-rooted Yersinia pestis strain angola reveals new insights into the evolution and pangenome of the plague bacterium. *J Bacteriol* 192:1685–1699.

Eroshenko GA et al., 2017. Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. *PLoS One* 12:e0187230.

Feldman M et al., 2016. A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol Biol Evol* 33:2911–2923.

Fetherston JD, Perry RD, 1994. 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.

Ford DC et al., 2014. The importance of the magnesium transporter MgtB for virulence of Yersinia pseudotuberculosis and Yersinia pestis. *Microbiology* 160:2710–2717.

Fu Q et al., 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci USA* 110:2223–2227.

Gage KL, Kosoy MY, 2005. Natural History of Plague: Perspectives from More than a Century of Research. *Annu Rev Entomol* 50:505–528.

Garcia E et al., 2007. Pestoides F, an atypical Yersinia pestis strain from the former Soviet Union. *Adv Exp Med Biol* 603:17–22.

Garrelt C, Wiechmann I, 2003. Detection of Yersinia pestis DNA in early and late Medieval Bavarian burials. *Documenta Archaeobiologiae* 1:247–254.

Gaunitz C et al., 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science* 360:111–114.

Gerhard S, 2017. Die Pest ist zurück. *Zeit online*. October 4, 2017.

Gilbert MTP et al., 2004. Absence of Yersinia pestis-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology* 150:341–354.

Gophna U et al., 2003. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* 312:151–163.

Grabenstein JP et al., 2006. Characterization of phagosome trafficking and identification of PhoP-regulated genes important for survival of Yersinia pestis in macrophages. *Infect Immun* 74:3727–3741.

Green MH, 2018. Putting Africa on the Black Death map: Narratives from genetics and history. *Afriques* 9. doi: 10.4000/afriques.2125.

Green MH, 2014. Taking "Pandemic" Seriously: Making the Black Death Global. In: Green MH, editor. *The Medieval Globe, v. 1. Pandemic Disease in the Medieval World: Rethinking the Black Death.* Kalamazoo, MI: Bradford. pp. 1:4.

Green MH, Schmid BV. 2016. Plague Dialogues: Monica Green and Boris Schmid on Plague Phylogeny (II). *Contagions*. https://contagions.wordpress.com/2016/06/29/plague-dialogues-monica-green-and-boris-schmid-on-plague-phylogeny-ii/

Green RE et al., 2009. The Neandertal genome and ancient DNA authenticity. *EMBO J* 28:2494–2502.

Groisman EA et al., 2013. Bacterial Mg^{2+} homeostasis, transport, and virulence. *Annu Rev Genet* 47:625–646.

Gruber H, 2018. Indirect Evidence for the Social Impact of the Justinianic Pandemic: Episcopal Burial and Conciliar Legislation in Visigothic Hispania. *Journal of Late Antiquity* 11:193–215.

Grupe G et al., 2015. Prähistorische Anthropologie. Berlin Heidelberg: Springer.

Guellil M et al., 2018. Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia. *Proc Natl Acad Sci USA* 115:10422–10427.

Guschanski K et al., 2013. Next-Generation Museomics Disentangles One of the Largest Primate Radiations. *Syst Biol* 62:539–554.

Haak W et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207–211.

Haensch S et al., 2010. Distinct clones of Yersinia pestis caused the Black Ceath. *PLoS Pathog* 6:e1001134–e1001134.

Haeser H, 1875. Lehrbuch der Geschichte der Medicin und der epidemischen Krankheiten. Dritte Bearbeitung. Dritter Band: Geschichte der epidemischen Krankheiten. Jena: Olms.

Haldon J et al., 2018. Plagues, climate change, and the end of an empire: A response to Kyle Harper's The Fate of Rome (3): Disease, agency, and collapse. *History Compass* 16:e12507.

Hänsch S et al., 2015. The pla gene, encoding plasminogen activator, is not specific to Yersinia pestis. *BMC Res Notes* 8:1–3.

Harbeck M et al., 2013. Yersinia pestis DNA from skeletal remains from the 6th century AD reveals insights into Justinianic Plague. *PLoS Pathog* 9:e1003349.

Harper K, 2017. The Fate of Rome: Climate, Disease, and the End of an Empire. Princeton: Princeton University Press.

Harper K, 2015. Pandemics and passages to late antiquity: rethinking the plague of c.249-270 described by Cyprian. *Journal of Roman Archaeology* 28:223–260.

Harper KN et al., 2011. The origin and antiquity of syphilis revisited: An Appraisal of Old World pre-Columbian evidence for treponemal infection. *Am J Phys Anthropol* 146 Suppl 53:99–133.

Harris SH, 2002. Factories of Death: Japanese Biological Warfare, 1932-45 and the American Cover-Up. New York, NY: Routledge.

Hawkins D, 1990. Black Death and the new London cemeteries of 1348. *Antiquity* 64:637–642.

Hecker JFC, 1832. Der schwarze Tod im vierzehnten Jahrhundert: Nach den Quellen für Ärzte und gebildete Nichtärzte bearbeitet. Berlin: F. A. Herbig.

Hedges SB, Schweitzer MH. 1995. Detecting Dinosaur DNA. *Science* 268:1191–1192.

Heintzman PD et al., 2017. A new genus of horse from Pleistocene North America. *Elife* 6:e29944.

Hershberg R et al., 2008. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biol* 6:e311.

Hinnebusch BJ et al., 2016. Ecological Opportunity, Evolution, and the Emergence of Flea-Borne Plague. *Infect Immun* 84:1932–1940.

Hinnebusch BJ et al., 2017. "Fleaing" the Plague: Adaptations of Yersinia pestis to its insect vector that lead to transmission. *Annu Rev Microbiol* 71:215–232.

Hinnebusch BJ et al., 2002. Role of Yersinia Murine Toxin in Survival of Yersinia pestis in the Midgut of the Flea Vector. *Science* 296:733–736.

Hirst LF, 1953. The Conquest of Plague: A Study of the Evolution of Epidemiology. Oxford: Clarendon Press.

Hofreiter M et al., 2015. The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays* 37:284–293.

Honap TP et al., 2018. Mycobacterium leprae genomes from naturally infected nonhuman primates. *PLoS Neglected Tropical Diseases* 12:e0006190.

Huebler R et al., 2019. HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *bioRxiv* doi:10.1101/534198.

Hufthammer AK, Walløe L, 2013. Rats cannot have been intermediate hosts for Yersinia pestis during medieval plague epidemics in Northern Europe. *J Archaeol Sci* 40:1752–1759.

Huson DH et al., 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol* 12:e1004957–e1004957.

Hymes R, 2014. Epilogue: A Hypothesis on the East Asian Beginnings of the Yersinia pestis Polytomy. In: Green MH, editor. *The Medieval Globe, v. 1. Pandemic Disease in the Medieval World: Rethinking the Black Death.* Kalamazoo, MI: Bradford. pp. 285–308.

Izdebski A et al., 2016. Realising consilience: How better communication between archaeologists, historians and natural scientists can transform the study of past climate change in the Mediterranean. *Quat Sci Rev* 136:5–22.

Johnson TL et al., 2014. Yersinia murine toxin is not required for early-phase transmission of Yersinia pestis by Oropsylla montana (Siphonaptera: Ceratophyllidae) or Xenopsylla cheopis (Siphonaptera: Pulicidae). *Microbiology* 160:2517–2525.

Jónsson H et al., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–1684.

Karlsson G, 1996. Plague without rats: The case of fifteenth-century Iceland. *Journal of Medieval History* 22:263–284.

Kay GL et al., 2015. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat Commun* 6:6717–6717.

Keeling MJ, Gilligan CA, 2000. Metapopulation dynamics of bubonic plague. *Nature* 407:903–906.

Keller M et al., 2018. Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541-750 CE). *bioRxiv* doi:10.1101/481226.

Keller M et al., 2019. Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541-750). *Proc Natl Acad Sci USA* 116:12363–12372.

Kingman JFC, 1982. The coalescent. *Stochastic Process Appl* 13:235–248.

Kircher M et al., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40:1–8.

Kislichkina AA et al., 2015. Nineteen Whole-Genome Assemblies of Yersinia pestis subsp. microtus, Including Representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica. *Genome Announc* 3:e01342–15.

Kislichkina AA et al., 2018a. Nine Whole-Genome Assemblies of Yersinia pestis subsp. microtus bv. Altaica Strains Isolated from the Altai Mountain Natural Plague Focus (No. 36) in Russia. *Genome Announc* 6:1–2.

Kislichkina AA et al., 2017. Eight Whole-Genome Assemblies of Yersinia pestis subsp. microtus bv. caucasica Isolated from the Common Vole (Microtus arvalis) Plague Focus in Dagestan, Russia. *Genome Announc* 5:5–6.

Kislichkina AA et al., 2018b. Six Whole-Genome Assemblies of Yersinia pestis subsp. microtus bv. ulegeica (Phylogroup 0.PE5) Strains Isolated from Mongolian Natural Plague Foci. *Genome Announc* 6:5–6.

Knapp M, Hofreiter M, 2010. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes* 1:227–243.

Kodaman N et al., 2014. Human and Helicobacter pylori coevolution shapes the risk of gastric disease. *Proc Natl Acad Sci USA* 111:1455–1460.

Konkola K, 1992. More than a coincidence? The arrival of arsenic and the disappearance of plaque in early modern Europe. *J Hist Med Allied Sci* 47:186–209.

Krause J et al., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464:894–897.

Krause-Kyora B et al., 2018. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *Elife* 7:e36666.

Krzywinski M et al., 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.

Kulikowski M, 2007. Plague in Spanish Late Antiquity. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 150–170.

Kumar S et al., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870–1874.

Kupferschmidt H, 1993. Die Epidemiologie der Pest. Der Konzeptwandel in der Erforschung der Infektionsketten seit der Entdeckung des Pesterregers in Jahre 1894. *Gesnerus Suppl* 43:1–222.

Kutyrev VV et al., 2018. Phylogeny and Classification of Yersinia pestis Through the Lens of Strains From the Plague Foci of Commonwealth of Independent States. *Front Microbiol* 9:1–11.

Lambert WG et al., 1969. Atra-Hasis: the Babylonian story of the Flood. Oxford: Clarendon Press.

Lederberg J, 2000. Infectious history. *Science* 288:287–293.

Lerat E, Ochman H, 2005. Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Res* 33:3125–3132.

Leven K-H, 2013. "Vandalisches Minimum" und molekularisierte Medizingeschichte: Neuere Entwicklungen in der Seuchengeschichte des Frühen Mittelalters In: Vögele J, Knöll S, Noack T, editors. *Epidemien Und Pandemien in Historischer Perspektive: Epidemics and Pandemics in Historical Perspective.* Wiesbaden: Springer. pp. 35–49.

Li H, 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997.

Li H, Durbin R, 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.

Little LK, 2007. Life and Afterlife of the First Plague Pandemic. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 3–32.

Little LK et al., 2007. Plague and the End of Antiquity: The Pandemic of 541-750. Cambridge University Press.

Littman RJ, Littman ML, 1973. Galen and the Antonine plague. *Am J Philol* 94:243–255.

Maçon P et al., 2011. Le Pressoir - Rapport final d'opération de fouille préventive. Bourges.

Maddicott J, 2007. Plague in Seventh-Century England. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 171–214.

Mahalanobis PC, 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India* 2:49–55.

Maixner F et al., 2016. The 5300-year-old Helicobacter pylori genome of the Iceman. *Science* 351:162–165.

Malim T, Hines J, 1998. The Anglo-Saxon cemetery at Edix Hill (Barrington A), Cambridgeshire. York: Council for British Archaeology.

Maricic T et al., 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004.

Martin J, 1978. The land of darkness and the Golden Horde. The fur trade under the Mongols, XIII-XIVth centuries. *Cahiers du monde russe et soviétique* 19:401–421.

Mascher M et al., 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat Genet* 48:1089–1093.

McCormick M, 2016. Tracking mass death during the fall of Rome's empire (II): a first inventory of mass graves. *Journal of Roman Archaeology* 29:1004–1046.

McCormick M, 2015. Tracking mass death during the fall of Rome's empire (I). *Journal of Roman Archaeology* 28:325–357.

McCormick M, 2007. Toward a Molecular History of the Justinianic Pandemic. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 290–312.

McCormick M, 2003. Rats, Communications, and Plague: Toward an Ecological History. *J Interdiscip Hist* 34:1–25.

McCormick M et al., 2013. Archaeological Finds of Rats: AD 1-1500. doi:10.7910/DVN/NARKTE

McNally A et al., 2016. "Add, stir and reduce": Yersinia spp. as model bacteria for pathogen evolution. *Nat Rev Microbiol* 14:177–190.

McNeil DG Jr, 2017. Fearsome Plague Epidemic Strikes Madagascar. *The New York Times*. October 6, 2017.

McNeill WH, 1976. Plagues and peoples. Garden City, N.Y.: Anchor Press.

Mecsas J et al., 2004. Evolutionary genetics: CCR5 mutation and plague protection. *Nature* 427:606–606.

Meier M, 2016. The "Justinianic Plague": the economic consequences of the pandemic in the eastern Roman empire and its cultural and religious effects. *Early Medieval Europe* 24:267–292.

Meyer M, Kircher M, 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010:db.prot5448.

Meyer M et al., 2017. Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. *Elife* 6:e25413.

Mordechai L, Eisenberg M, 2019. Rejecting Catastrophe: the case of the Justinianic Plague. *Past & Present* 244:3–50.

More AF et al., 2017. Next-generation ice core technology reveals true minimum natural levels of lead (Pb) in the atmosphere: Insights from the Black Death: Pandemic Reveals True Natural Pb Levels. *GeoHealth* 1:211–219.

Morelli G et al., 2010. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* 42:1140–1143.

Morens DM, Littman RJ. 1992. Epidemiology of the Plague of Athens. *Trans Proc Am Philol Assoc* 122:271–304.

Mühlemann B et al., 2018. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* 557:418–423.

Mulhall J, 2019. Plague before the Pandemics: The Greek Medical Evidence for Bubonic Plague before the Sixth Century. *Bull Hist Med* 93:151–179.

Namouchi A et al., 2018. Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proc Natl Acad Sci USA* 115:201812865–201812865.

Noonan JP et al., 2005. Genomic sequencing of Pleistocene cave bears. *Science* 309:597–599.
Norris J, 1978. Response. *Bull Hist Med* 52:114–120.
Norris J, 1977. East or west? The geographic origin of the Black Death. *Bull Hist Med* 51:1–24.
Osei-Owusu P et al., 2019. FPR1 is the plague receptor on host immune cells. *Nature* 1–6.
Ottoni C et al., 2013. Pig Domestication and Human-Mediated Dispersal in Western Eurasia Revealed through Ancient DNA and Geometric Morphometrics. *Mol Biol Evol* 30:824–832.
Pääbo S, 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314:644–645.
Pääbo S et al., 1989. Ancient DNA and the Polymerase Chain Reaction. *J Biol Chem* 264:1–4.
Pääbo S et al., 2004. Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645–679.
Parkhill J et al., 2001. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* 413:523–527.
Peltzer A et al., 2016. EAGER: efficient ancient genome reconstruction. *Genome Biol* 17:60–60.
Perry RD, Fetherston JD, 1997. Yersinia pestis--etiologic agent of plague. *Clin Microbiol Rev* 10:35–66.
Pollitzer R, 1954. The Plague. Geneva: World Health Organization.
Pollitzer R, 1952a. Plague studies: 6. Hosts of the infection. *Bull World Health Organ* 6:381–465.
Pollitzer R, 1952b. Plague studies: 4. Pathology. *Bull World Health Organ* 5:337–337.
Pollitzer R, 1952c. Plague studies: 2. The plague bacillus. *Bull World Health Organ* 5:73–108.
Pollitzer R, 1951. Plague studies: 1. A summary of the history and survey of the present distribution of the disease. *Bull World Health Organ* 4:475–533.
Prendergast ME et al., 2017. Reconstructing Asian faunal introductions to eastern Africa from multi-proxy biomolecular and archaeological datasets. *PLoS One* 12:e0182565.
Prentice MB, Rahalison L, 2007. Plague. *Lancet* 369:1196–1207.
Pribyl K, 2017. Farming, Famine and Plague: The Impact of Climate in Late Medieval England. Cham: Springer.
Pusch CM et al., 2004. Yersinial F1 antigen and the cause of Black Death. *Lancet Infect Dis*.
Quinlan AR, Hall IM, 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.
Raoult D et al., 2000. Molecular identification by "'suicide PCR'" of Yersinia pestis as the agent of Medieval Black Death. *Proc Natl Acad Sci USA* 97:12800–12803.
Rascovan N et al., 2019. Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline Article Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. *Cell* 176:295–305.e10.
Rasmussen M et al., 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Proc Natl Acad Sci USA* 506:6886–6891.

Rasmussen S et al., 2015. Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* 163:571–582.

Raynaud C, 2010. Les nécropoles de Lunel-Viel (Hérault) de l'Antiquité au Moyen Âge. Revue archéologique de Narbonnaise Supplément 40. Montpellier: Presses universitaires de la Méditerranée.

Rebeil R et al., 2013. Induction of the Yersinia pestis PhoP-PhoQ regulatory system in the flea and its role in producing a transmissible infection. *J Bacteriol* 195:1920–1930.

Reich D et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.

Reid AH et al., 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. *Proc Natl Acad Sci USA* 96:1651–1656.

Reuter S et al., 2014. Parallel independent evolution of pathogenicity within the genus Yersinia. *Proc Natl Acad Sci USA* 111:6768–6773.

Roberts CA, Buikstra J, 2008. The Bioarchaeology of Tuberculosis: A Global View on a Reemerging Disease.

Rohland N et al., 2015. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci* 370:20130624–20130624.

Roosen J, Curtis DR, 2018. Dangers of noncritical use of historical plague data. *Emerg Infect Dis* 24:103–110.

Rothenberg GE. 1973. The Austrian sanitary cordon and the control of the Bubonic plague: 1710-1871. *J Hist Med Allied Sci* 28:15–23.

Russell JC, 1968. That earlier plague. *Demography* 5:174–184.

Sabeti PC et al., 2005. The case for selection at CCR5-Delta32. *PLoS Biol* 3:1963–1969.

Sallares R, 2007. Ecology, Evolution, and Epidemiology of Plague. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 231–289.

Sallares R, 2006. Role of environmental changes in the spread of malaria in Europe during the Holocene. *Quat Int* 150:21–27.

Salo WL et al., 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci USA* 91:2091–2094.

Sarris P, 2007. Bubonic Plague in Byzantium. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 119–132.

Sarris P, 2002. The Justinianic plague: origins and effects. *Contin Chang* 17:169–182.

Savinetsky AB, Krylovich OA, 2011. On the history of the spread of the black rat (Rattus rattus L., 1758) in northwestern Russia. *Biol Bull Russ Acad Sci* 38:203–207.

Schamiloglu U, 2016. The Plague in the Time of Justinian and Central Eurasian History: An agenda for Research In: Zimonyi I, Karatay O, editors. *Central Eurasia in the Middle Ages. Studies in Honour of Peter B. Golden.* Wiesbaden: Harrasowitz.

Schmid BV et al., 2015. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci USA* 112:3020–3025.

Schubert M et al., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 9:88.

Schuenemann VJ et al., 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proc Natl Acad Sci USA* 108:746–752.

Schuenemann VJ et al., 2018. Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. *PLoS Negl Trop Dis* 12:e0006447.

Schuenemann VJ et al., 2013. Genome-Wide Comparison of Medieval and Modern Mycobacterium leprae. *Science* 341:179–183.

Scott S, Duncan CJ, 2001. Biology of Plagues: Evidence from Historical Populations. Cambridge: Cambridge University Press.

Sebbane F et al., 2009. The Yersinia pestis caf1M1A1 fimbrial capsule operon promotes transmission by flea bite in a mouse model of bubonic plague. *Infect Immun* 77:1222–1229.

Sebbane F et al., 2006. Role of the Yersinia pestis plasminogen activator in the incidence of distinct septicemic and bubonic forms of flea-borne plague. *Proc Natl Acad Sci USA* 103:5526–5530.

Seecharran T et al., 2017. Phylogeographic separation and formation of sexually discrete lineages in a global population of Yersinia pseudotuberculosis. *Microb Genom* 3:e000133.

Segata N et al., 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9:811–814.

Sehdev PS, 2002. The origin of quarantine. *Clin Infect Dis* 35:1071–1072.

Seifert L et al., 2016. 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–e0145194.

Sha J et al., 2011. Characterization of an F1 deletion mutant of Yersinia pestis CO92, pathogenic role of F1 antigen in bubonic and pneumonic plague, and evaluation of sensitivity and specificity of F1 antigen capture-based dipsticks. *J Clin Microbiol* 49:1708–1715.

Shrewsbury JFD, 1970. A history of bubonic plague in the British Isles. Cambridge: Cambridge University Press.

Signoli M et al., 1998. La rechute épidémique de la Grande Peste de Marseille (Mai–Juillet 1722): le charnier de l'Observance. *Med Trop* 58:7–13.

Signoli M et al., 2002. Paleodemography and Historical Demography in the Context of an Epidemic: Plague in Provence in the Eighteenth Century. *Population* 57:829–854.

Simond PL, 1898. La propagation de la peste. *Annales de l'Institut Pasteur* 12:625–687.

Simonet M et al., 1996. Invasin production by Yersinia pestis is abolished by insertion of an IS200-like element within the inv gene. *Infect Immun* 64:375–379.

Skurnik M, Wolf-Watz H. 1989. Analysis of the yopA gene encoding the Yop1 virulence determinants of Yersinia spp. *Mol Microbiol* 3:517–529.

Slack P, 1981. The disappearance of plague: an alternative view. *Econ Hist Rev* 34:469–476.

Slavin P, 2019. Death by the Lake: Mortality Crisis in Early Fourteenth-Century Central Asia. *J Interdiscip Hist* 50:59–90.

Slon V et al., 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* 608:605–608.

Smith O et al., 2014. A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci Rep* 4:4003.

Snavely MD et al., 1989. Magnesium transport in Salmonella typhimurium: 28Mg2+ transport by the CorA, MgtA, and MgtB systems. *J Bacteriol* 171:4761–4766.

Snavely MD et al., 1991. The mgtB Mg^{2+} transport locus of Salmonella typhimurium encodes a P-type ATPase. *J Biol Chem* 266:815–823.

Sønstebø JH et al., 2010. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol Ecol Resour* 10:1009–1018.

Spyrou MA et al., 2019. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat Rev Genet* 20:323–340.

Spyrou MA et al., Krause J. 2018a. A phylogeography of the second plague pandemic revealed through the analysis of historical Y. pestis genomes. *bioRxiv* doi:10.1101/481242.

Spyrou MA et al., 2016. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host Microbe* 19:874–881.

Spyrou MA et al., 2018b. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nat Commun* 9:2234.

Stamatakis A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Stathakopoulos D, 2004. Famine and Pestilence in the Late Roman and Early Byzantine Empire: A Systematic Survey of Subsistence Crises and Epidemics. Aldershot: Ashgate.

Stoclet AJ, 2007. Consilia humana, ops divina, superstitio: Seeking Succor and Solace in Times of Plague, with Particular Reference to Gaul in the Early Middle Ages. In: Little LK, editor. *Plague and the End of Antiquity: The Pandemic of 541–750.* Cambridge: Cambridge University Press. pp. 135–149.

Summers WC, 2012. The Great Manchurian Plague of 1910-1911: The Geopolitics of an Epidemic Disease. Yale: Yale University Press.

Sun YC et al., 2008. Experimental evidence for negative selection in the evolution of a Yersinia pestis pseudogene. *Proc Natl Acad Sci USA* 105:8097–8101.

Sun YC et al., 2014. Retracing the evolutionary path that led to flea-borne transmission of Yersinia pestis. *Cell Host Microbe* 15:578–586.

Sun Z et al., 2019. Human plague system associated with rodent diversity and other environmental factors. *R Soc Open Sci* 6:190216.

Sussman GD, 2011. Was the Black Death in India and China? *Bull Hist Med* 85:319–355.

Tardy F et al., 1999. Yersinia enterocolitica type III secretion-translocation system: channel formation by secreted Yops. *EMBO J* 18:6793–6799.

Taubenberger JK et al., 2005. Characterization of the 1918 influenza virus polymerase genes. *Nature* 437:889–893.

Tavare S, 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Some mathematical questions in biology / DNA sequence analysis edited by Robert M Miura*.

Teasdale MD et al., 2017. The York Gospels: a 1000-year biological palimpsest. *R Soc Open Sci* 4:170988.

Tran TNN et al., 2011. Beyond ancient microbial DNA: Nonnucleotidic biomolecules for paleomicrobiology. *Biotechniques* 50:370–380.

Tran TNN et al., 2011a. Brief communication: Co-detection of Bartonella quintana and Yersinia pestis in an 11th-15th burial site in Bondy, France. *Am J Phys Anthropol* 145:489–494.

Tran TNN et al., 2011b. High throughput, multiplexed pathogen detection authenticates plague waves in medieval Venice, Italy. *PLoS One* 6:e16735.

Tumpey TM et al., 2005. Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus. *Science* 310:77–81.

Twigg G, 1984. The Black Death: A Biological Reappraisal. London: Batsford Academic and Educational.

Uhlmann B, 2017. Pest auf Madagaskar beunruhigt Seuchenschützer. *Süddeutsche Zeitung*. October 7, 2017.

Vågene ÅJ et al., 2018. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology and Evolution* 2:520–528.

Varlık N, 2017. "Oriental plague" or epidemiological orientalism? Revisiting the plague episteme of the early modern Mediterranean. In: Varlık N, editor. *Plague and Contagion in the Islamic Mediterranean.* Kalamazoo, MI: Arc Humanities Press. pp. 57–88.

Varlık N, 2014. New Science and Old Sources: Why the Ottoman Experience of Plague Matters. In: Green MH, editor. *The Medieval Globe, v. 1. Pandemic Disease in the Medieval World: Rethinking the Black Death.* Kalamazoo, MI: Bradford. pp. 193–228.

Vogler AJ et al., 2017. Temporal phylogeography of Yersinia pestis in Madagascar: Insights into the long-term maintenance of plague. *PLoS Negl Trop Dis* 11:e0005887.

Wagner DM et al., 2014. Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. *Lancet Infect Dis* 14:319–326.

Waldron HA, 2001. Are plague pits of particular use to palaeoepidemiologists? *International Journal of Epidemiology* 30:104–108.

Waldron T, 2008. Palaeopathology. Cambridge: Cambridge University Press.

Walløe L, 2008. Medieval and modern bubonic plague: some clinical continuities. *Med Hist Suppl* 27:59–73.

Wang J et al., 2004. Complete Genome Sequence of Yersinia pestis Strain 91001, an Isolate Avirulent to Humans. *DNA Res* 11:179–197.

Warinner C et al., 2017. A Robust Framework for Microbial Archaeology. *Annu Rev Genomics Hum Genet* 18:13.1–13.36.

Warinner C et al., 2015. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos Trans R Soc Lond B Biol Sci* 370:20130376–20130376.

Wheelis M, 2002. Biological warfare at the 1346 siege of Caffa. *Emerg Infect Dis* 8:971–975.

Whittles LK, Didelot X, 2016. Epidemiological analysis of the Eyam plague outbreak of 1665-1666. *Proc Biol Sci* 283. doi:10.1098/rspb.2016.0618

WHO, 2017. Plague Outbreak Madagascar, External Situation Report 14.

WHO, 2016. Plague around the world, 2010–2015. *Weekly Epidemiological Record* 91:89–92.

WHO, 2012. Human plague: review of regional morbidity and mortality, 2004–2009. *Weekly Epidemiological Record* 85:40–45.

Wickham H, 2016. ggplot2: Elegant Graphics For Data Analysis. New York, NY: Springer.

Wiechmann I, Grupe G, 2005. Detection of Yersinia pestis DNA in two early medieval skeletal finds from Aschheim (Upper Bavaria, 6th century A.D.). *Am J Phys Anthropol* 126:48–55.

Wiechmann I et al., 2010. Yersinia pestis DNA Sequences in Late Medieval Skeletal Finds, Bavaria. *Emerg Infect Dis* 16:1806–1807.

Willerslev E, Cooper A, 2005. Ancient DNA. *Proc Biol Sci* 272:3–16.

Wilson EO, 1998. Consilience: The Unity of Knowledge. New York, NY: Knopf.

Wirth T et al., 2008. Origin, spread and demography of the Mycobacterium tuberculosis complex. *PLoS Pathog* 4:e1000160.

Wood DE, Salzberg SL, 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46.

Woods JW et al., 1992. The Osteological Paradox: Problems of Inferring Prehistoric Health from Skeletal Samples. *Curr Anthropol* 33:343–358.

Woodward SR et al., 1994. DNA Sequence from Cretaceous Period Bone Fragments. *Science* 266: 1229–1232.

Worobey M et al., 2016. 1970s and "Patient 0" HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* 539:98–101.

Wren BW, 2003. The Yersiniae — a model genus to study the rapid evolution of bacterial pathogens. *Nat Rev Microbiol* 1:55–64.

Wu L et al., 1936. Plague; a manual for medical and public health workers. Shanghai: Weishengshu National Quarantine Service, Shanghai Station.

Xiao YL et al., 2013. High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *J Pathol* 229:535–545.

Xu L et al., 2015. The trophic responses of two different rodent-vector-plague systems to climate change. *Proc Biol Sci* 282:20141846.

Xu L et al., 2014. Wet climate and transportation routes accelerate spread of human plague. *Proc Biol Sci* 281:20133159.

Yang R, Anisimov A, 2016. Yersinia pestis: Retrospective and Perspective. Dordrecht: Springer.

Yersin A, 1894. La Peste bubonique à Hong-Kong. *Annales de l'Institut Pasteur* 8:662–667.

Zagorski N, 2006. Profile of Svante Pääbo. *Proc Natl Acad Sci USA* 103:13575–13577.

Zhou D et al., 2004. Genetics of Metabolic Variations between Yersinia pestis Biovars and the Proposal of a New Biovar, microtus. *J Bacteriol* 186:5147–5152.

Zhou D, Yang R, 2009. Molecular Darwinian Evolution of Virulence in Yersinia pestis. *Infect Immun* 77:2242–2250.

Zimbler DL et al., 2015. Early emergence of Yersinia pestis as a severe respiratory pathogen. *Nat Commun* 6:7487–7487.

Zischler H et al., 1995. Detecting Dinosaur DNA. Comment on DNA sequence from Cretaceous period bone fragments. *Science* 268:1192–1193.

Ehrenwörtliche Erklärung

Die geltende Promotionsordnung der Fakultät für Biowissenschaften der Friedrich-Schiller-Universität Jena ist mir bekannt. Die vorliegende Dissertation habe ich selbständig angefertigt und dabei weder Textabschnitte aus einer eigenen Prüfungsarbeit oder von Dritten ohne Kennzeichnung übernommen. Ich habe alle verwendeten Quellen, persönliche Mitteilungen und Hilfsmittel angegeben. Alle Personen, die mich bei der Auswahl und Auswertung des Materials, bei der Erstellung der Publikationen und bei der Verfassung des Manuskripts unterstützt haben, sind in Kapitel 3 – *Overview of Manuscripts* und in den Danksagungen namentlich aufgeführt. Die Hilfe eines Promotionsberaters wurde nicht in Anspruch genommen. Dritte haben weder unmittelbar noch mittelbar geldwerte Leistung für Arbeiten, die mit dem Inhalt dieser Dissertation zusammenhängen, von mir erhalten. Ich versichere, dass ich die vorliegende Arbeit nicht für eine staatliche oder andere wissenschaftliche Prüfung als Prüfungsarbeit eingereicht habe. Ferner habe ich weder diese Arbeit noch eine in wesentlichen Teilen ähnliche, noch eine andere Abhandlung bei einer anderen Hochschule oder Fakultät als Dissertation eingereicht.

Jena, Juli 2020

_____

Marcel Keller

Acknowledgements

I would like to express my deep gratitude to all people who contributed in many ways to the success of this dissertation project:

first and foremost, to my doctorate supervisor Johannes Krause, for giving me the opportunity to conduct this research project in his group, for his support and trust as well as the freedom to become an independent researcher.

to Michaela Harbeck, for introducing me to the world of ancient DNA, for her enormous commitment to build the collaboration this thesis is substantially based on and for mentoring me over the last seven years.

to Alexander Herbig, who kindly agreed to supervise my work and guided me through the computational analyses as well as the publication process, and for his valuable scientific and personal advice.

to Maria A. Spyrou, who taught me most of the laboratory and analytical methods and who has been a reliable collaborator.

to Michael McCormick, for his impressive commitment and his tremendous knowledge shared on any given opportunity in endless conversations.

to Gunnar U. Neumann, who was the most diligent and independent Master student, lab assistant and dear colleague.

to all members of the pathogen group: foremost Kirsten Bos for her valuable input; Aditya Kumar Lankapalli for the inspiring discussion on false positive SNPs; Aida Andrades Valtueña for her generosity, producing countless coverage plots and sharing her scripts; likewise James Fellows Yates and Åshild J. Vågene for their generous sharing of their computational skills; and Andreas Kröpelin, who developed and customized his SNPEvaluation tool in an impressive speed.

to Guido Brandt, Franziska Aron, Marta Burri, Cäcilia Freund, Raphaela Stahl and Antje Wissgott for their support and help in the lab and for creating a pleasant work environment.

to Joris Peters, who kindly hosted me for a whole year in his group; as well as Michaela Zimmermann, Ptolemaios Paxinos, Ferdinand M. Neuberger, Simon Trixl, Ursula Mutze and all colleagues from the Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, LMU Munich, for the great time.

to all members and associates of the State Collection for Anthropology and Palaeoanatomy in Munich, for their generous support with organizational tasks, basic research and sampling, and for sharing their anthropological expertise.

to Brigitte Haas-Gebhard, Bernd Päffgen and Jochen Haberstroh for their commitment to our plague meetings and for sharing their archaeological expertise on the Bavarian sites.

to Christiana L. Scheib, for the collaboration on the Edix Hill samples, for proofreading the dissertation manuscript and for her support while finishing my thesis in Tartu.

to Stefan Leenen, Alexander Berner and Sandra Maus, for giving me the opportunity to write the essay for their exhibition catalogue, for valuable inspirations and insights, and for their warm welcome in Herne.

to Sandra Lösch, Dominique Castex, Craig Cessford, Raphaël Durand, Albert Ribera i Lacomba, Claude Raynaud, Amelie Alterauge, Sacha Kacki, Kristin von Heyking, Michaël Gourvennec, Christian Later, Mathias Hensch and Bettina Jungklaus for the excellent interdisciplinary collaboration.

to Beate Kerpen, Sabine Ziegler and Bettina Bock for their assistance dealing with administrative tasks, and for their professional and personal advice.

to Philip Slavin, Kyle Harper, Monica Green and members of her MEDMED-List, for insights in their ongoing research and fruitful discussions on the historical dimensions of both historical pandemics.

to past and present scientific and administrative members of the Max Planck Institute for the Science of Human History for creating such an inspiring and encouraging environment.

to my dear colleagues Elizabeth A. Nelson and Barbara Pavlek for their friendship and support, and for always having a sympathetic ear during and after work.

and last but not least to my family for their love and unconditional support.

Thank you so much!

Appendix

Supplementary Material for Manuscript A

Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750)

Marcel Keller*, Maria A. Spyrou*, Christiana L. Scheib, Gunnar U. Neumann, Andreas Kröpelin, Brigitte Haas-Gebhard, Bernd Päffgen, Jochen Haberstroh, Albert Ribera i Lacomba, Claude Raynaud, Craig Cessford, Raphaël Durand, Peter Stadler, Kathrin Nägele, Jessica S. Bates, Bernd Trautmann, Sarah A. Inskip, Joris Peters, John E. Robb, Toomas Kivisild, Dominique Castex, Michael McCormick, Kirsten I. Bos, Michaela Harbeck, Alexander Herbig & Johannes Krause
*contributed equally

Published in *PNAS* 116 (25), 12363-12372 (2019), doi:10.1073/pnas.1820447116

Supplementary Information Text

Assessment of SNPEvaluation

To test the validity of the tool SNPEvaluation and the applied criteria for detection of false positive SNPs, we generated a set of artificial datasets representing clean and contaminated *Y. pestis* genomes ranging from ~3x to ~30x coverage. To mimic the contamination with environmental bacteria (background), we used the sample DIR002.A, since enrichment with a *Y. pestis* probeset cannot be simulated reliably *in silico*. The sample DIR002.A was captured and sequenced after initial qPCR screening, but was classified as negative for *Y. pestis* after sequencing and mapping. Therefore, the 8470 reads mapping with strict parameters are derived from environmental sources. For the target *Y. pestis* reads (foreground), we generated artificial sequencing data based on the reference genome CO92 aiming for 3-fold, 5-fold, 10-fold and 30-fold coverage in triplicates using the tool gargammel (1). We simulated paired-end 75 bp sequencing reads using the coverage flag (-c) and the length distribution of the paired-end 75 bp sequenced sample DIT003.B. The achieved coverages are ~6 % lower (on average 2.83-fold, 4.72-fold, 9.43-fold, 28.20-fold, see SI Appendix, Table S5) due to mapping quality filtering. The artificial reads were combined with the background and mapped with stringent (-n 0.1) and lenient (-n 0.01) mapping parameters. Since the capture probe set was constructed based on CO92 among others, a possible capture bias was neglected for the artificial CO92 reads.

To test for the sensitivity of our method, we applied it to all false positive SNPs that were introduced by our background sample (SI Appendix, Fig. S2, Table S6). As expected, no SNPs were called in any of the simulated *Y. pestis* datasets without background. Using only the background (DIR002), a total number of 451 SNPs were called. One additional SNP was only called with spiked in *Y. pestis* reads (called as N in DIR002), summing up to a total of 452 possible false positive SNPs. With increasing coverage of spiked in *Y. pestis* reads, the number of false positive SNPs decreases from on average 139.67 for 3-fold, 40.33 for 5-fold, 5 for 10-fold to 0 for 30-fold, showing that the environmental contamination can introduce a significant number of false positive SNPs for low coverage samples but is negligible for high coverage samples. The remaining positions of putative false positive SNPs are called as heterozygous (N) or as reference call in high coverage samples (in average 0.67 for 10-fold, 63.67 for 30-fold).

Using the same criteria as for the First Pandemic samples in this study, SNPEvaluation was able to filter 100 % of false positive SNPs from every artificial dataset. Moreover, all chosen filtering criteria were necessary for the determination as false positive SNPs: Regarding the ratio of mean coverage with low/high stringent mapping, this criterion was able to filter out almost all SNPs. Only 2 of 451 SNPs from the pure background, 1 of 147 SNPs from a 3x-fold and 1 of 39 SNPs from a 5-fold coverage contaminated *Y. pestis* samples passed the criterion (ratio 1.00), but were filtered out either due to a heterozygous SNP or uncovered positions surrounding the SNP. The lowest ratio for the false positive SNPs without other criteria applying ranges between 1.05 and 1.21, justifying the very strict criterion of a ratio of 1.00 to be accepted as a true positive SNP.

To determine the specificity of our method, we generated a SNP table based only on modern *Y. pestis* strains and extracted the SNPs that are shared by at least 34 genomes (~15 %). This cutoff was chosen due to the local maximum in the distribution of the numbers of genomes sharing SNPs. This resulted in a list of 418 SNPs, which constitute the backbone phylogeny of *Y. pestis* and are a good proxy for true positive SNP positions (see SI Appendix, Fig. S3, Table S7). However, all positions appear as reference calls in our datasets, since the artificial reads are based on the same genome (CO92) that was arbitrarily chosen as reference genome.

In the simulated CO92 datasets without background, in average 79.33 positions were called as reference for the simulated 3-fold coverage samples, 230.33 for 5-fold, 403.33 for 10-fold and all 418 for 30-fold. All remaining positions were called as N due to low coverage. The ratio of mean coverage with low/high stringent mapping never exceeded 1.00 in any datasets, showing that sequencing errors simulated by gargammel are not interfering with this parameter. A

maximum of 2.63 % of the reference calls were filtered out due to uncovered bases surrounding the position in the simulated 3-fold or 5-fold coverage samples, but never for 10-fold or 30-fold coverage. Using solely the background sample, only 2 positions were called as reference but filtered out due to a >1.00 ratio, giving evidence that the DIR002 sample is indeed negative for *Y. pestis*.

For the simulated CO92 datasets contaminated with the background, in average 74.33 positions were called for the 3-fold samples, 227.67 for 5-fold, 401.33 for the 10-fold and all 418 for 30-fold. This means that 0 – 7.89 % of the positions are not called due to heterozygosity introduced by the background. After filtering with SNPEvaluation, in average 69.67 positions remained for the 3-fold samples, 216.67 for 5-fold, 385.00 for 10-fold and 401.67 for 30-fold, so 3.59 – 8.57 % of called positions are filtered out.

In summary, SNPEvaluation offers a maximum sensitivity (100 % of false positive SNPs filtered out), while retaining a high specificity (8.57 % or fewer of true shared positions filtered out). Although we used only one background sample in this analysis and were not able to model, i.e., the capture bias, the presented method is a powerful tool to filter low coverage datasets for false positive SNPs, a crucial step to build reliable bacterial phylogenies.

SNP Evaluation of the Aschheim Genome and SNP Effect Analysis

The Aschheim genome (2) was evaluated separately, given its peculiarly high number of potential false positive SNPs described previously (3). Our systematic evaluation verified previous classifications: all SNPs potentially unique to Aschheim that passed the criteria show a coverage lower than 5-fold, which was the threshold of their SNP calling. However, the high number of presumably shared SNPs that did not pass our stricter criteria underlines again the high 'heterozygosity' of the genome (see SI Appendix, Table S14) that might be explained not only by contamination by soil bacteria or sequencing errors but presumably also by PCR and capture artefacts, as previously discussed (3). Therefore, the Aschheim genome was excluded from subsequent analyses.

Of the 39 non-shared phylogenetically informative chromosomal SNPs that were detected among all new genomes, 20 are non-synonymous in coding regions of (hypothetical) proteins (SI Appendix, Table S10). The genome of VAL001.B shows non-synonymous SNPs in the genes *tyrP*, YPO1985 and YPO2588. TyrP is a transcriptional regulator for the metabolism of aromatic amino acids and was identified as a virulence factor crucial for the infection of mice (4). YPO1985 was identified as a glycosyl transferase gene inactivated in the avirulent strain 91001 and thus might be a virulence factor as well (5). The gene YPO2588 codes for an ABC transport protein. An additional non-synonymous SNP was detected on the pMT1 plasmid in the putative DNA-binding protein YPMT1.59C. For the genome PET004.A, only one non-synonymous SNP was identified, located on the hypothetical protein YPO3510. The genomes of LVC_merged, LSD001.A and LSD023.A show seven shared non-synonymous SNPs: in the genes *marC*, a multidrug resistance protein; *phrB,* coding a 3',5'-cyclic-nucleotide phosphodiesterase; *tyrA*, a bifunctional chorismite mutase/prephenate dehydrogenase; the phosphoenolpyruvate carboxylase *ppc,* the oligogalacturonate lyase ogl; the sugar transport ATP-binding protein YPO1554, identified as pseudogene in CO092 (6); and the hypothetical protein YPO4112. The gene *ppc* has been shown to be connected with the type III secretion system that is essential for pathogenicity in *Y. pestis* by injection of Yops (*Yersinia* outer membrane proteins) into host cells of the innate immune system (7). An additional non-synonymous SNP shared among this clade was found on the pCD1 plasmid, affecting *yopB*, a *Yersinia* outer membrane protein, identified as virulence factor since it is an important translocon for the injection of effector proteins via the type III secretion system (8). A SNP unique to LVC_merged was identified in the hypothetical protein YPO2238. Both LSD001.A and LSD023.A share a SNP in the *lacY*, a galactoside permease. In LSD001.A, four unique SNPs were identified as non-synonymous: in *hisP*, a histidine/lysine/arginine/ornithine transporter subunit; *yapB*, a putative autotransporter protein pseudogenized in KIM (9); YPO1999, a decarboxylase; and YPO1856, a hypothetical protein. In LSD023.A, three non-synonymous SNPs were found: in *cpxA*, a two-component sensor kinase; *lacY*, a galactoside permease already affected by a SNP shared with LSD001.A; and *znuC*, a high-affinity zinc transporter ATPase.

Extended Deletion Analysis

A deletion of ca. 12.9 kb was found on the chromosome (positions 2,533,444 to 2,546,401 in CO92) of all three genomes of the French clade (LVC_merged, LSD001.A, LSD023.A), affecting the genes *araF, araG, araH, araC, manA* and several hypothetical proteins (see SI Appendix, Fig. S7). Affecting the arabinose operon, strains with this deletion will be unable to metabolize arabinose. This is also a defining characteristic of the biovar *Microtus* (10), here caused by a 122-bp frameshift deletion in *araC*, a clade within the 0.PE4 branch including the genomes 0.PE4Ca_ CMCCN010025, 0.PE4Cc_CMCC18019, 0.PE4Cd_CMCC93014, 0.PE4Ce_ CMCC91090, 0.PE4m_I-3086, 0.PE4_I-3134, 0.PE4_Microtus91001 and 0.PE4_M0000002. The loss of the arabinose operon is however only one of multiple changes in genome architecture, gene loss and pseudogenization in this clade, causing these strains to be avirulent in humans. Therefore, the 12.9 kb deletion observed here in the First Pandemic genomes is presumably a similar case of genome decay.

Another deletion of 14.8 kb was found on the pMT1 plasmid (positions 23,133 to 37,975 in CO92) of the genomes of the French clade (LVC_merged, LSD001.A, LSD023.A) and both genomes of Unterthürheim (UNT003.A, UNT004.A). All 21 genes affected by this deletion are hypothetical proteins including a putative ABC transporter ATP-binding protein, therefore the functional consequences of the deletion cannot be determined. The fact that this deletion is found in the French clade as well as in the Unterthürheim genomes but not in the Altenerding (AE1175) and Dittenheim genomes (DIT003.B) that are otherwise identical, suggests that this deletion is more common and mediated by a process similar to the transposable elements on the chromosome.

Phylogeographic Analyses

The new genomes and radiocarbon dates combined suggest an association of the British genome as well as the polytomy giving rise to the four lineages with the early phase of the First Pandemic or even the Justinianic Plague itself (541–544). The accumulation of one (EDI001) or two (Altenerding cluster) SNPs from the basal node of all genomes could have happened on the way from Egypt to western Europe. The fact that the pandemic reportedly spread from Pelusium along the Mediterranean coastline in two independent waves, one heading west to Alexandria and the other east to Palestine, could explain the early branching event (11). Strikingly, such a diversification during the onset of a pandemic has not been found yet for the Black Death (1348–1352), where the two genomes from London East Smithfield (12) and Barcelona were found to be identical (13). Besides differing mutation rates, this might be due to differences in propagation speed between the 6th and 14th century, related to changes in human mobility by land and sea: a significantly slower or less direct transmission over large distances would allow the pathogen to acquire more substitutions.

The lineages found in Bavaria could have spread there by a 'western route' from Gaul, by a 'southern route' from Italy or by an 'eastern route' from Illyricum, which were all affected by plague in or around 543. The presence of plague in the British Isles even suggests a fourth 'northern route' upstream along the Rhine river. The 'western' and 'southern route' would have necessitated overland transport via the Roman road network that connected all of the relevant sites with the Mediterranean coastlines and was still functional in the 6th century (SI Appendix, Fig. S9). The 'southern route' would have required crossing the Alps via different passes that had been used since Antiquity (14). Navigation along the Danube could have facilitated the 'eastern route'. The importance of rivers for the spread of plague has already been shown exemplarily for the Rhône during the First Pandemic (15) and for the Black Death (16). However, attempts to prove the preferential spread via rivers during Second Pandemic have recently been criticized (17, 18).

The site of Petting is geographically situated only 100 km southeast of Aschheim and Altenerding (SI Appendix, Fig. S9). However, it was located in the Roman province *Noricum ripense* whereas the sites with the distinct uniform lineage were situated in *Raetia secunda* (Aschheim, Altenerding, Unterthürheim) or close by (Dittenheim). Although the administrative system of the Western Roman Empire had broken down by the mid-6th century, its political borders continued to be influential, not least because of the ecclesiastical system of dioceses that followed them. It is possible

that these ancient boundaries influenced the spread of the two epidemic outbreaks in modern-day Bavaria. Since the river Inn separated *Raetia secunda* and *Noricum ripense*, this might suggest that rivers could serve as physical barriers to the spread of plague where river transport was negligible. This in turn would rather suggest the 'eastern route' or the 'southern route' for Petting as described above.

Complementing the previous results from Aschheim and Altenerding, our new data from Unterthürheim and Dittenheim underline the epidemic extent of this plague outbreak in early medieval Bavaria, totalling 16 individuals with genomic evidence for *Y. pestis* and an additional five PCR-positive individuals in Aschheim (19). Far from the urban centres of the time and any recorded outbreak of plague, the new molecular evidence stands in strong contrast to Durliat's claim that the Justinianic Plague was merely an urban phenomenon. Instead, we view this data as being in line with ancient statements by Procopius, John of Ephesos and Paul the Deacon who reported that the countryside of the Levant and Italy were severely impacted.

Radiocarbon Dating

At least one individual per burial was sampled for radiocarbon dating for all burials that tested positive for *Y. pestis*, assuming simultaneity of interment for the multiple burials. Samples were dated at the CEZ Archaeometry gGmbH, Mannheim, Germany. For Saint-Doulchard, published radiocarbon dates of adjacent burials in the same trench are reported (20). These as well as the raw radiocarbon dates of Aschheim and Altenerding (2, 3) were recalibrated for consistency. The raw radiocarbon dates were calibrated with IntCal13 (21) in OxCal v4.3.2 (22). All raw and calibrated dates are given in SI Appendix, Table S12; Fig. S8 shows the respective probability distributions. Some of the intervals completely pre-date the onset of Justinianic Plague (541) which could be explained by a marine or freshwater reservoir effect (23, 24) or human bone collagen offset (25). In the absence of C/N isotope data and a well-established method for addressing the human bone collagen offset, we report calibrated dates without any correction. A combination of the probability distributions using the function "Combine" in Oxcal was attempted for the trench inhumations of Lunel-Viel (LVC001, LVC002, LVC003, LVC005, LVC006, LVC007), the double burial 1175/1176 of Altenerding (AE1175, AE1176) and the quadruple burial 131–134 of Unterthürheim (UNT004, UNT005, UNT008). For Lunel-Viel, the resulting 2-sigma interval is 562–604 calAD (A_{comb} 148.6 %; A 109–127.9 %), for Altenerding 428–552 calAD (A_{comb} 107.1 %; A 104–105.9 %) and for Unterthürheim 428–544 calAD (A_{comb} 29.6 %; 14.5–100.9 %). The low agreement index for Unterthürheim is caused by the comparatively recent dating sample of UNT005 with an individual agreement index of 14.5 %. Therefore, the combination of uncorrected dates is deemed problematic in general and was dismissed for interpretation.

Cartography

All maps were generated in ArcGIS 10.4.1 (ESRI) using the 'World Ocean Basemap' without references. The sources for all historical occurrences are given in the SI Appendix. The mapped regions in Fig. 1 are primarily based on the Digital Atlas of Roman and Medieval Civilizations (DARMC; https://darmc.harvard.edu) maps "Provinces AD303–324" for Western Europe and "Provinces ca. AD500" for Eastern Europe, Middle East and Africa. The provinces in SI Appendix, Fig. S9 are based on a georeferenced map by Rettner and Steidl (26), the Roman roads are combined from Rettner and Steidl and the DARMC map "Roman Roads". The main rivers in Figs. 1C and 3 were taken from Natural Earth (ne_10m_river_lake_centerlines, http://www.naturalearthdata.com), based on data provided by the European Commission, Joint Research Centre, Institute for Environment and Sustainability (JRC IES).

Archaeological Context Information

The following site descriptions present condensed information on all sites examined in this study. The classification of multiple burial types follows McCormick (2015). Sex and age determination is based only on morphological examination. For the age classification, the German system (27) *is used as follows: Infans I (0–6 years old), infans II (7–12 years old), juvenile (13–20 years old), adult (20–40 years old), mature (40–60 years old), senile (more than 60 years old). Sites that were tested positive for plague are set in bold.*

Alladorf (ALL; Markt Turnau, Landkreis Kulmbach, Germany):

The Carolingian cemetery of Alladorf, dating roughly between 630 to 720, revealed 163 graves with remains of 276 individuals. However, the total size of the cemetery is unknown, since the excavation did not reach the borders of the burial area. B. Leinthaler classified three burials as double burials (type 1): 179/180 (ALL001, early adult female; ALL002, infans I), 184/185 (ALL003, infans II; ALL004, late adult to early mature male) and 203/204 (ALL005, late adult male; ALL006, infans I). The two burials 188/189 (infans I; late adult to early mature male) and 208/209 (infans I; early adult female) were classified as double burials with unclear simultaneity. (28, 29)

Dirlewang (DIR; Landkreis Unterallgäu, Germany):

The Alemannic site of Dirlewang is a very small cemetery with 40 excavated burials and an expected number of 55 burials in total. It dates to the Late Merovingian period, from 650–700, based on the archaeological finds. Two double burials (type 1) were found on this site, 33/34 (DIR001, juvenile to early adult female; juvenile male) and 38/39 (adult male; DIR002, adult female). Graves 18 and 19 (mature male; early adult male) did not share the same grave pit but were buried very close to each other, indicating a connection. Burials 30/31, 36/37 and 2 were classified as non-simultaneous successive or additive double burials (type 2). (30)

Dittenheim (DIT; Landkreis Weißenburg-Gunzenhausen, Germany):

The early medieval cemetery of Dittenheim recruited from a settlement on the site of the modern village. Only 6.5 km south of the *limes*, the settlement was probably well connected to the remaining Roman infrastructure. 2.5 km Southeast of the cemetery, remains of a Germanic fortification dating to the Migration Period were found, known as *Gelbe Bürg*. This structure went out of use around 500, just before the region fell under the rule of Franks.

The cemetery was known already by 1937. The first excavation campaign in 1968 revealed the first 114 burials. Later campaigns took place in 1971 (burials 115–164) and 1972 (165–244). The excavations probably reached the borders of the cemetery. However, some burials might have been lost due to erosion from a nearby river and from plowing. Besides the early medieval burials, older settlement traces from the Linear Pottery culture, Pre-Roman Iron Age and the imperial Roman period as well as one Bronze Age burial were found on the site.

The cemetery of Dittenheim, dating from the middle of the 6th century to the end of the 7th century, revealed 238 graves containing 244 individuals and 10 cremations from the same period. Other remarkable finds were three horse burials and one circular pit. Four burials were classified as double burials type 1: 8A/B (DIT007, mature female; mature male; possibly an additive burial), 18A/B (DIT003, mature male; DIT004, adult female), 22A/B (DIT005, infans I; DIT006, adult female) and 188A/B (DIT001; DIT002, both infans II). There were also three exceptional single burials in prone position (180, 201, 204) deviating from the contemporary burial rites. (31)

Edix Hill (EDI; civil parish Barrington, Cambridgeshire, United Kingdom):

The Anglo-Saxon cemetery of Edix Hill, close to Barrington and Orwell was initially discovered in the 19th century. Excavations between 1989 and 1991 revealed part of an inhumation cemetery comprising 149 individuals in

115 graves, dating to between 500 and 625/650, although one burial was believed to be possibly Iron Age. It is estimated that there may originally have been around 300 burials in total with a complete cross-section of the population by age and sex suggesting that the burials relate to a community of 50–65 people spanning around 150 years. The initial dating of the cemetery was based primarily upon artefact typologies of the various grave goods and seriation by correspondence analysis, with burials broadly divided into earlier and later groups and some evidence for spatial patterning over time. Subsequent radiocarbon dating has broadly confirmed and refined the dating of the cemetery. Palaeopathological evidence suggested the presence of tuberculosis, leprosy and cancers. The human remains are currently held by Cambridgeshire County Council, who generously provided access to the material.

The cemetery contained a total of 18 multiple burials, of which 10 were classifiable as type 1: they comprise one quadruple burial with at least two of the individuals buried simultaneously and eight simultaneous double burials. Four double burials are of unclear simultaneity, another four and a triple burial are clearly understandable as type 2 (non-simultaneous). The site was initially not sampled for the purpose of plague screening, so samples from both multiple and single burials were screened. This includes the single graves 76 (Sk405, EDI001; juvenile), 69 (Sk359, EDI002, early adult female), 78 (Sk424, EDI003, juvenile), 46 (Sk146, EDI010, early mature male), 60 (Sk183, EDI011, adult female), 63 (Sk198, EDI012, early adult male), 83 (Sk436, EDI014, early adult female), 90 (Sk458, EDI017, senile female), 95 (Sk530, EDI018, late mature female), 97 (Sk551, EDI019, early adult male), 99 (Sk576, EDI020, mature male), 100 (Sk578, EDI021, early mature male), 105 (Sk592, EDI022, late mature female), the non-simultaneous double burial grave 66 (Sk322A, EDI013, at least adult male; SK322B, early adult female) and triple burial grave 18 (Sk42A1, infans I; Sk42A2, early adult with indet. sex); Sk42B, EDI009, early adult female), the simultaneous double burials grave 96 (Sk547A, adult female; Sk537B, EDI004, infans II), 106 (Sk626A, EDI005, early adult female; Sk626B, EDI006, early adult male), 9 (Sk13A, EDI007, mature to senile male; juvenile male) and 84 (Sk440A, EDI015, adult female; Sk440B, infans I) as well as the complex burial grave 2, with two individuals buried simultaneously (Sk3B, adult to mature male; Sk3C, EDI007, early adult male) and two more individuals buried later (Sk3A1 and Sk3A2, both adult and possibly female and male respectively). (32, 33)

Forchheim (FOR; Gemeinde Pförring, Landkreis Eichstätt, Germany):

Within the vestiges of a late medieval settlement, a quadruple burial was found without any context suggesting a larger burial ground. Two mature women (FOR002, individual 2; ind. 4), a mature man (FOR003, ind. 3) and a male juvenile (FOR001, ind. 1) were buried partially overlapping but all in East-West orientation. Remarkably, one of the women (2) was buried in the prone position. Grave goods such as a sax and a glass bead necklace date the burial to the second half of the 7th century. A belt buckle was indicative of a clothed inhumation. (34)

Grafendobrach (GRA; Gemeinde Kulmbach, Landkreis Kulmbach, Germany):

The Carolingian cemetery of Grafendobrach revealed 85 burials; the excavations 1975–1976 did not reach the edges of the burial ground. The various garment artefacts found in the graves point to the late 9th to early 10th century. Besides several secondary burials (type 2), there were two remarkable graves: complex 83/84/85 consisted of two adjacent stone cists holding two women (83, GRA002, early mature; 85, GRA003 early adult) and an infant (84, ca. 1 year old) buried at the feet of 83. A single burial of a man (42, GRA001, early senile) was covered with stones, on top of which two neonates (43, 44) were found without individual burial pits. (35)

Kleinlangheim (KLH, Gemeinde Großlangheim, Landkreis Kitzingen, Germany):

The Frankish cemetery of Kleinlangheim dates to the late 5th to early 8th century and was completely excavated from 1962 to 1969. It contained 244 burials and a remarkable number of 56 cremations; unlike other contemporary cemeteries, the graves showed signs of grouping, potentially indicating family groups. Nine multiple burials type 1 were

identified, eight of them double burials: 35/36 (KLE005, adult male; late mature male), 41/42 (late adult male with skull fracture; juvenile to early adult), 147/148 (min. adult; neonate), 171/172 (two infans I), 208/209 (infans I; early adult female), 272/273 (late adult male; KLE004, senile male), 211 (KLE001, mature male; second individual cremated) and 218/219 (KLE002, mid-adult to mature male, KLE003, late adult to early mature female). The triple burial 100–102 contained remains of two infants (1–2 years old; 4–6 years old) and one mature individual of indeterminable sex. (36)

Leobersdorf (LEO, Bezirk Baden, Austria):

The Avar cemetery of Leobersdorf was excavated between 1977 and 1983 and was dated to 640–800. In total, 154 burials containing remains of 171 individuals were excavated with an exceptionally high number of 25 multiple burials. 16 burials were identified as double burials, of which 9 contained each an adult and a subadult (16: adult female and infans I; 57: mature male and infans I; 74: adult female and infant; 93: mature male and infans I; 100: mature male and juvenile female; 104: adult female and infans I; 114: adult male and infans I; 119: late mature male and infans II; 140: adult female and infans II). Four consisted of two adults (35: mature male and female; 86: adult female and mature male; 114: adult female and male; 144: senile and adult male) and two held only subadults (23: male juvenile and infans II; 103: two juvenile females; 145: infans I and II). The remains of infant individuals often were identified only after the morphological examination following the excavation.

Three burials were identified as triple burials (21: senile male, mature female and infant; 67: adult female and two infants; 82: juvenile male, LEO001; adult female, LEO002; infans I, LEO003). A fourth (99) was interpreted as a double burial (mature male and infant) with a secondary burial (female).

Two burials contained four individuals (105: mature female, juvenile female, infans I and neonates; 134: adult female, mature male, infans I and II). Burial 79 contained the remains of five individuals, identified as the burial of a mature female with secondary burials of four additional individuals (senile female and male, juvenile female and infans II) that had probably been buried on the same spot previously, and were reburied after the mature female. (37, 38)

Lunel-Viel (LVH, LVC; Arrondissement Montpellier, Département Hérault, France):

The site of Lunel-Viel is equidistant (25 km) from the modern and Roman towns of Montpellier and Nîmes at the intersection of two paved roads that were active in late Antiquity: a later Roman secondary road running about 2 km southeast and parallel to the Roman *Via Domitia* that connected Iberia to Italy, and a road connecting the *Via Domitia* to the coastal lagoon *l'Étang d'or*, part of the complex of lagoons that included that of Lattes (anc. *Lattara*), famous already in Antiquity for its fishery (Pliny the Elder, Natural History, 9.29–32), and Maguelone, an island bishopric first mentioned in the 6th century when abundant emerging archaeological evidence documents far-flung Mediterranean shipping connections. A brief Roman-era occupation in the 2nd century BCE at Lunel-Viel was followed by continuous settlement beginning ca. 50–80 CE, although the ancient settlement area was abandoned in the 7th century, and the exact location of the new dwelling zones between the 7th and the 10th century remains unknown. Three inhumation cemeteries received the deceased of this community in succession from the 4th century. The earliest, at *Le Verdier,* functioned from the end of the 3rd century down to the beginning of the 6th century, and yielded 340 burials. The second, known as *Les Horts*, received burials of the late 5th to 7th centuries; 140 burials were excavated out of an estimated original total of ca. 200. The third is associated with the church of St. Vincent; burials began there ca. 520 and continued down to the 17th century; 97 have been excavated.

The cemetery *Les Horts* contained one clear case of a double burial, likely simultaneous (type 1), of two adult individuals (38A/B) placed head facing toes in a sarcophagus; the head and upper body of 38B were destroyed, along with the feet of 38A and part of the sarcophagus, during later leveling operations. Disarticulated remains of a third poorly preserved adult (38C, LVH001) were found pushed into the eastern end of the sarcophagus. Grave goods likely

stemming from the original burial (one buckle, one plated buckle, one fibula, one pin and one clasp) date the burial of 38C to 475–550.

Outside the cemeteries, anomalous (deviant) burials of the remains of eight individuals were found during the excavation of the Gallo-Roman structures, called the *Quartier central*. Robbing of foundation stones in Antiquity had left empty trenches which were subsequently used for the summary interment of these individuals. A pin buckle, a spindle whorl and a knife found with the remains of the presumably clothed individuals date the interment to the 5th to 6th century. The positions and postures of the individuals clearly deviate from the burial customs seen in the contemporaneous cemeteries of *Les Horts* and *Le Verdier*: two adult women (3A: LVC003, adult; 3B: LVC004, mature) were placed under limestone slabs, with one resting her head on the thighs of the other. Westward of this group, a woman (1: LVC001, early adult) and another adult individual (2: LVC002, late adult to early mature, sex indeterminable) were found; the latter however had been disturbed by later agricultural activities. Whereas these individuals appear to have been laid rather carefully on the ground, the two men adjacent to the west (4: LVC005, mature, 5: LVC006, early adult) seem to have been carelessly dropped into the trench, the latter with his arms stretched over his head. Two more individuals, a young adolescent woman (6, LVC007) and an infant (7, sex indeterminable), were found in a second trench to the east. However, all individuals were interred more or less in supine East-West oriented position, as expected for early medieval Christian burials. (39, 40)

München-Aubing (AUB, Stadt München, Germany):

The Baiuvarian cemetery of München-Aubing was excavated in two phases, 1938 and 1960–1963. It was used from the 5th to 7th century. A total of 896 graves was excavated. Four double burials (type 1) have been identified: 854/855 (AUB008, adult male; late mature male) 809/810 (AUB006; AUB007, two adult males), 724/725 (AUB004, adult male; AUB005, mature male; with possible later manipulation) and 676/677 (AUB002, mature female; AUB003, male individual). A third individual (675, infans I) was buried later on top of the double burial 676/677. (41)

Neuburg an der Donau (NEU; Landkreis Neuburg-Schrobenhausen, Germany):

The site of Neuburg an der Donau was occupied at least from the first half of the 2nd century by a Roman military base associated with an urn cemetery with around 130 cremations. The burial ground *Seminargarten* is related to a later occupation period during the 4th century. From an estimated number of 150 burials, this site revealed 130 burials with 133 individuals. One burial was identified as a double burial (32A/B, adult female and juvenile with indeterminable sex), one as a triple burial (34A/B/C: mature male; NEU001, mature male; NEU002, adult male). Both are part of zone 1, dating to 330–360. Two additional burials (14, 28) were suggested to contain "mother and child" (neonate/infant). (42, 43)

Peigen (PEI; Markt Pilsting, Landkreis Dingolfing-Landau, Germany):

The early medieval cemetery of Peigen dating to the mid-5th to the 7th century revealed around 274 burials. Among these, three could be identified as double burials. Double burial 18 contained an adult woman (18-1, PEI001) and an early adult woman (18-2, PEI002), buried in opposite orientation. Double burial 63 contained an adult woman (63-1, PEI004) and a senile man (63-2, PEI005). The burial of a late mature male (109-1) and a child (infans I, 109-2) form the third clear double burial. The burials 62-1 of a juvenile male and 62-2 of a late mature female (PEI003) were adjacent but at different depth and therefore of questionable simultaneity (type 2). (44)

Petting (PET; Landkreis Traunstein, Germany):

Although located in Germany today, Petting falls in a region that was in ancient times associated to the Roman city and later ecclesiastical center of *Iuvavum* (today Salzburg, Austria) and has therefore a different settlement history

than other parts of southern Bavaria. The cemetery of Petting was discovered in 1991 and excavated within the following two years. A comprehensive publication of the site of Petting is still pending, not least because the archaeological findings were not available for research until the ownership situation was clarified in 2010. Studies so far concentrated on individual burials or specific artefacts.

The cemetery was completely excavated, counting 721 burials. Only 24 burials were lost due to destruction prior to the excavation: around 50 % of the burials fell victim to ancient grave robbers. At least four multiple burials were observed in the cemetery, all presumably type 1. The double burial 172/173 contained a female and a male individual; the double burial 377/378 a late adult to mature and a late adult to early mature male (PET003/PET004), the triple burial 342/343/344 at least one early adult male (PET001) and an infans II (PET002). The double burial 630/631 of an early adult female (PET006) and an early mature female (PET007) was associated with a single burial of an infans II (PET005) on top. (45, 46)

Regensburg Fritz-Fend-Straße (RFF; Stadt Regensburg, Germany):

Regensburg, *Castra Regina,* was an important Roman military base with a civilian settlement of the province *Raetia*. The extensive adjacent cemetery *Kumpfmühl* was excavated in several campaigns following the first excavation in 1872 by J. Dahlem. He documented 827 burials and cremations but left the vast majority of features undocumented, which led S. v. Schnurbein to the assumption that the first excavation uncovered around 3000 cremations and 2000 burials, mainly dating to the late 3rd to mid 4th century. More features were lost in this area due to later construction works without archaeological survey. An excavation in 1999 revealed 100 additional cremations and 50 burials. The excavations of *Fritz-Fend-Straße* and *Im Güterbahnhof* took place in 2011 and revealed 48 cremations/115 burials and 161 cremations/116 burials, respectively. The site *Fritz-Fend-Straße* contained two exceptional double burials of type 1: In burial 30, the second individual (30-1: RFF001, late adult to early mature male) was laid in prone position and opposite orientation on top of the first (30-2), who was buried in supine position. In 53, the two individuals, a juvenile female (53-2: RFF002) and an early mature male (53-3, RFF003), were buried in crouched position facing in the same direction but in opposite orientation. An archaeological or radiocarbon date for *Fritz-Fend-Straße* has not yet been published, but a similar date as for *Kumpfmühl* can be assumed. (47)

Saint-Doulchard Le Pressoir (LSD; Arondissement Bourges, Département Cher, France):

In the former village of Saint-Doulchard, 2 km from Bourges, an archaeological survey in 2007 revealed a dense funerary space, dated from the 7th to 12th centuries. The town is first mentioned as *Sanctus Dulcardus* in the 7th century, named after the hermit Dulcardus (late 6th century). It is likely that the cemetery was established on the site of Dulcardus' hermitage. However, a tumulus (6th to 4th c. BCE) and a Roman villa found in the vicinity of the excavated funerary space attest earlier occupation of the site.

In 2009, the rescue excavation led by P. Maçon excavated a part of this cemetery measuring approximately 400 m^2, reaching one of its boundaries in the form of a ditch that borders the burial area and runs from the northeast to the southwest. This ditch was also used for burials: 57 individuals in 48 burials have been discovered within it whose organization and funerary practice may indicate a particular mortality crisis. However, some secondary burials and stratigraphically overlapping graves within the ditch could suggest that it was used on multiple occasions. Eleven multiple burials were found, including nine double burials F206-35/36 (mature; fetus), F206-68/69 (LSD008, mature to senile male; infans I), F206-71/72 (LSD009, senile male, infans I), F206-79a/79b (LSD011, mature to senile male; infans I), F206-93/145 (LSD012, early adult; early adult male), F206-152a/152b (infans I, infans II), F206-155a/155b (senile female; LSD019, mature to senile male), F206-156a/156b (LSD020, mature to senile male; infans I) and F206-204a/204b (LSD023, mature male; infans I-II), and two triple burials F206-91a/91b/91c (mature male, infans I, infans I), F 206-172a/172b/172c (infans I; LSD022, juvenile; LSD021, mature female). Three more double burials were found within

the cemetery area delimited by the ditch: F206-75a/-75b (infans I to juvenile; late adult to mature female), F206-229a/-229b (late adult to mature female, infans I) and F206-235 (at least adult male; subadult).

In addition to several multiple burials, a selection of single burials within the ditch were also sampled for this study: F206-31 (LSD001, early adult female), F206-38 (LSD002, mature female), F206-40 (LSD003, mature male), F206-45 (LSD004, mature to senile male), F206-51 (LSD005, mature male), F206-52 (LSD006, early adult), F206-65 (LSD007, mature to senile), F206-78 (LSD010, mature to senile male), F206-132 (LSD013, mature male), F206-136 (LSD014, mature female), F206-144 (LSD015, juvenile), F206-151 (LSD016, mature male), F206-153 (LSD017, infans II), F206-154 (LSD018, early adult), F206-208 (LSD024, juvenile), F206-215 (LSD025, mature to senile female), F206-216 (LSD026, infans II).

A palaeodemographic analysis highlighted similarities in the structure of the population buried in the ditch with mortality profiles frequently observed in epidemic contexts, including plague. This suspicion is reinforced by the taphonomic evidence arguing for simultaneous interments (type 1) in an easily available structure, separated from the rest of the community. Nevertheless, the presence of multiple burials in the space enclosed by the ditch might indicate a gradual shift in funerary practice at the beginning of an epidemic when mortality was still limited.

Radiocarbon dating was performed on five individuals buried in the ditch. The calibrated 2-sigma intervals cluster in two groups (ca. 650–880 and ca. 720–920), which might indicate the usage of the ditch as funerary space for multiple events. A combination of all five dates is rejected by Oxcal (df=4, T=11.2 [5% 9.5]), further substantiating the hypothesis of multiple events. (20)

Sindelsdorf (SIN; Landkreis Weilheim-Schongau, Germany):

The early medieval cemetery of Sindelsdorf consists of 331 burials with remains of 354 individuals dating from 500 to 720. It included three double burials type 1 (25/26 senile male and mature male; 51/52, senile and mature male (SIN006, SIN007); 154/155, infant and senile female) and a burial group of three individuals with one buried later on top (163: infant; 164: SIN002, late adult to early mature male; 165: SIN003, adult female; 162: SIN001, senile male). The double burial 25/26 with a later burial on top (24, mature female) may be identified as a probable homicide due to a perimortem skull fracture (26). (48)

Straubing Azlburg I/II (SAZ; Stadt Straubing, Germany):

Straubing, called *Sorviodurum* by the Romans, was an important *castrum* on the Danubian border of *Raetia*. The Late Roman cemeteries – around 200 m apart from each other – were excavated in 1981 (*Azlburg I*) and 1984 (*Azlburg II*). However, the borders of the cemeteries were not reached on all sides, so the original dimensions remain unclear. They were in use at roughly the same time, approximately between 300 and 450. *Azlburg I* contained 107 graves with 111 individuals and one cremation. The burial 54-1/-2/-3 was identified as a triple burial (SAZ002, male juvenile; SAZ003, male juvenile; third individual min. adult). The double burial 16/17 contained the remains of two children (SAZ001, infans I; neonate). *Azlburg II* contained additional 434 graves with 45 individuals. The double burial 5a/b of two adult men was interpreted as the burial of two soldiers. (49)

Unterthürheim (UNT; Gemeinde Buttenwiesen, Landkreis Dillingen an der Donau, Germany):

The town of Unterthürheim is probably the site of the early medieval settlement that used the burial ground *Buttenwiesen*. Traces of an older Imperial Roman settlement (1st to 5th century) were found on the *Thürlesberg* hill, around two kilometers from the cemetery; it was near the Roman roads *Via Iuxta Danuvium* (following the Danube) and the *Via Claudia Augusta* (connecting *Augusta Vindelicum*/Augsburg with Italy) which intersect at the fort *Submuntorium*, around 10 km Northeast of Unterthürheim.

The Alemannic cemetery has been known since 1889, when a local resident discovered the first six graves. Additional burials were found between 1943 and 1966. The first professional excavation took place in 1968 (graves 1–42), later excavations occurred between 1969 and 1972 (graves 43–178) and in 1979 (graves 180–238). All of them were rescue excavations, which explains the scattered trenches and limited dimensions: The excavations are thought to have reached the southern and northern border of the cemetery, but the extent to the west and east remain unknown. In total, 256 burials were excavated of which 230 were preserved well enough for more detailed examination. They are archaeologically dated to 525 to 680.

H. Lüdemann lists in total 14 double burials, two triple burials and one quadruple burial. The triple burials 120/122 and 124/125 as well as the double burials 167/168, 186/186a and 217/217a were classified as secondary burials (type 2). Burials 12 (mature with indeterminable sex, subadult) 63/64 (UNT006, adult female; UNT007, infans I), 79/80 (adult female, infans I), 116 (mature female; UNT001, infans II), 140/141 (mature male, early adult female), 145/146 (adult to early mature female, male) and 189 (adult females, neonates) presumably simultaneous double burials (type 1). 65/66 (infans I, infans II) and 190/190a (adult to early mature female, early adult male) were classified by C. Grünewald as possible secondary burials (type 2). For the burials 97/98 and 147/149, the classification is unclear.

The quadruple burial 131–134 (UNT004, infans I; UNT005, adult male; infans I; min. adult male) was associated with two single burials: 129 (UNT002, infans I) and 130 (UNT003, mature male). (38, 50)

Valencia Plaça de l'Almoina (VAL; Ciudad de Valencia, Spain):

The site of Valencia, *Plaça de l'Almoina* is an intramural cemetery that was founded in the 5th century and remained in use during the Visigothic period in the 6th and 7th centuries. It was excavated from 1985 to 1999. Contrary to the earlier Roman practice, the necropolis was not located outside the city walls but close to a shrine commemorating the martyrdom of St. Vincent, and adjacent to the cathedral. Many burials still follow the later Roman tradition that made use of *tegulae* and amphoras. Among the later burials, there are several monumentalizing collective tombs that were probably used as elite family tombs for successive burials.

The excavated areas of the necropolis contained a number of multiple burials of differing types and dates, including 15 or 16 collective slab tombs that appear to reflect successive burials (type 2: details in McCormick 2016, 1024n15). At the moment of sampling, the human remains were comingled and only sorted by find numbers (corresponding the numbers VAL001–009), so the assignment to individuals is not possible for this site. Four or five features seemed to be of type 1, i.e., reflecting at least in part simultaneous burials. The quadruple burial tomb 4 was covered with *tegulae* and contained four individuals who were piled one on top of another in an east-west orientation (two samples were taken, VAL008). Tomb 28 was stone-lined and just north of an apse identified as a memorial shrine to the martyrdom of St. Vincent; it contained a minimum of 21 individuals, from whom sixteen samples were taken (VAL005, VAL006, VAL007). Tomb 50, just southwest of Tomb 28 (and actually under the wall of that apse), may well have been a continuation of Tomb 28; it contained at least seven individuals, from whom seven samples were taken (VAL003, VAL004). Tombs 28 and 50 are lower than and earlier than the wall of the apse. The apse itself seems well dated to the late sixth or early seventh century, based on pottery excavated in a pit under the apse pavement. Multiple tomb 40 was a pit burial containing four supine individuals (five samples, VAL002) oriented east-west in the entry from the east into the apse of the memorial shrine. Tomb 41 is south of the apse structure which might indicate a privileged burial space. It is a pit that seems to have used pre-existing Roman walls to the south and east for two sides; bricks were simply placed without mortar to form the other two sides of the burial space, which might suggest a hasty or improvised burial. It contained the very disturbed remains of at least 15 individuals, including 4 subadults, from which we took seven samples (VAL001). (38, 51–53)

Waging (WAG; Landkreis Traunstein, Germany):

The early medieval cemetery of Waging, close to the site of Petting, was discovered and excavated in 1987/1988. The excavators deduced a discontinuity between the previous Roman settlement (until 300) and the early medieval colonization related to the burial ground that was in use between 530 and 700. Similarly to the site of Petting, the resettlement was most likely dominated by *Iuvavum* (modern Salzburg, Austria) in the southeast.

Although the 239 burials were subsequently analyzed and the artefacts went into a local exhibition after extensive conservation measures, the site is still not published in a comprehensive way. At least three burials were identified as multiple burials: the double burial 200/201 of an early mature male and infans II (WAG002; WAG003) were located on top of the burial of a potentially male infans II (WAG001). The double burial 37 contained remains of a potentially female infans II (WAG004) and a juvenile to early adult individual (WAG005). Burial 39 was identified as a double burial of a potentially female infans II (WAG006) and a late mature female (WAG007). (45, 54)

Westheim (WES; Landkreis Weißenburg-Gunzenhausen, Germany):

The Merovingian cemetery of *Westheim-Mehlbuck*, excavated between 1979 and 1985, revealed remains of 255 individuals in 228 graves and dates to the 6th to mid-7th century. Five burials were clearly identified as double burials: 17a/b, 26a/b, 36a/b, 52a/b, 106a/b and 128a/b. Six additional burials were interpreted as presumably double burials of a child with a parent: 172, 190, 202, 206, 208 and 210. The triple burial I, uncovered during a previous excavation in the 1910s, held the remains of a late adult to early mature male (1-1, WES001), a late mature female (1-3, WES003) and of a third individual (1-2, WES002, min. adult, probably female). Burial 13a/b was reported by H. Lüdemann as a quadruple. However, according to R. Reiss, the burial contained only disturbed remains of two males (adult, mature). (38, 55)

Sources for mapping plague outbreaks between 541 and 750 CE

Clysma as possible starting point

Tsiamis et al. (2009); Harper (2017) 215–218

541: Pelusium

Prokopios, *BP* 2.22, 6

See also: Stathakopoulos (2004) no. 102

541: Gaza, Ashkelon, Negev

John of Ephesos, *Fragment E* 77

Epigraphic material in Conrad (1996) 95

See also: Stathakopoulos (2004) no. 103

541: Alexandria

Prokopios, *BP* 2.22, 6

John of Ephesos, *Fragment E* 77, 80

Chron. Seert 185

Michael the Syrian 2 (235–238)

See also: Stathakopoulos (2004) no. 104

542: Jerusalem and countryside

Cyril of Scythopolis, *Vita* of Kyriakos 10

See also: Stathakopoulos (2004) no. 105

542: Izra

Epigraphic material in Koder (1995) 13–18

See also: Stathakopoulos (2004) no. 106

542: Antioch

Vita of Symeon Stylites Iunior 69

See also: Stathakopoulos (2004) no. 107

542: Apamea

Evagrios, *Hist. eccl.* 4.29

See also: Stathakopoulos (2004) no. 108

542: Emesa

Zacharias Rhetor, *Fragment ch. IX*

Leontios of Neapolis, *Vita* of Symeon Salos 151

See also: Stathakopoulos (2004) no. 109

542: Myra

Vita of Nicholas of Sion 52

See also: Stathakopoulos (2004) no. 110

542: Constantinople

Prokopios, *BP,* 2.22, 2.23

John of Ephesos, *Fragment E* 74–93

John Malalas, 482

Theophanes, *Chron.* AM 6034

See also: Stathakopoulos (2004) no. 111

542: Sykeon

Theodore of Sykeon, Vita 8 (I 7–8)

See also: Stathakopoulos (2004) no. 112

542: North Africa

Victor of Tunnuna, *Chronica* ad a. 542 (201)

See also: Stathakopoulos (2004) no. 114

542: Sicily

Byzantina Siciliae, 133

See also: Stathakopoulos (2004) no. 115

543: Sufetula

Epigraphic material from Sufetula nos. 1–4 (277–280)

See also: Stathakopoulos (2004) no. 117

543: Italy, Illyricum

Marcellinus Comes ad a. 543 (107)

See also: Stathakopoulos (2004) no. 116

543: Gaul, Arles, Reims, Trier

Gregory of Tours, *Lib. hist.* 4.5, 6.15, 6.33

Gregory of Tours, *Lib. vitae patrum* 6.6, 17.4

Gregory of Tours, *Liber in gloria confessorum* 78

543: Spain

Victoris Tunnunensis Chronicon, *Consularia Caesaraugustana*, ad a. p.c. Basili II

See also: Kulikowski (2007) 150–151

543–544: Rome

Inscriptiones Christianae Urbis Romae 1.1452, 2.4287, 7.17624, 2.5088, 2.4289, 8.20839, 2.5087, 2.5087, 2.5087.

See also: Stathakopoulos (2004) no. 118

544: Ireland, Britain

Annals of Tigernach 137, 198

Adomnán, *Vita Columbae* 348

See also: E. Phillimore (1888); A. Dooley (2007), 216; J. Maddicott (2007), 173–174

558: Constantinople

Agathias, *His.,* 5.10

John Malalas, *Chron.* 18.127 (489)

Theophanes, *Chron.* AM 6050

Agapios, *Kitab al-'Unwan*

See also: Stathakopoulos (2004) no. 134, Harper (2017) no. 1

561–562: Cilicia and Anazarbos, Syria, Mesopotamia, Antioch

Theophanes, *Chron.* AM 6053

Vita Symeon Stylites Iunior 126–129

Chron. ad a. 640

Barhadbšabba 388–389

Chron. Seert 185–186

Amr ibn Matta 42–43

See also: Stathakopoulos (2004) no. 136, Harper (2017) no. 2

565: Liguria, Northern Italy

Paul the Deacon, *Hist. Langobardorum* 2.4

See also: Stathakopoulos (2004) no. 139, Harper (2017) no. 3

571: Italy, Gaul, Bourges, Chalon-sur-Saône, Clermont, Dijon, Lyon

Marius of Avenches, a. 571

Gregory of Tours, *Lib. hist.* 4.31–32

See also: Stathakopoulos (2004) no. 144, Harper (2017) no. 4

573–574: Constantinople, Egypt, Syria and Antioch

John of Biclaro, a. 573

Agapios, *Kitab al-'Unwan*

John of Nikiu 94.18

Chron. ad a. 846

Michael the Syrian 10.8 (346)

See also: Stathakopoulos (2004) no. 145, Harper (2017) no. 5

576: Ireland

For secondary sources see:

Dooley (2007), 219

Woods (2003)

582–584: Southwestern Gaul, Narbonne, Spain

Gregory of Tours, *Lib. hist. 6*.14, 6.33

See also: Harper (2017) no. 6

586: Constantinople

Agapios, *Kitab al-'Unwan*

See also: Harper (2017) no. 7

588: Gaul, Lyon, Marseille, Spain

Gregory of Tours, *Lib. hist.* 9.21–22

See also: Harper (2017) no. 8

590–591: Rome, Narni, Ravenna, Istria and Grado, Rhône Valley, Avignon, Viviers

Gregory of Tours, *Lib. hist.* 10.1, 10.23

Gregory the Great, *Dial.* 4.18, 4.26, 4.37; *Ep.* 2.2

Paul the Deacon, *Hist. Langobardorum,* 3.24, 4.4

Liber pontificalis 65

See also: Stathakopoulos (2004) nos. 151, 154; Harper (2017) nos. 9, 10

592: Syria, Palestine, Antioch

Evagrius, *Hist. eccl.* 4.29

I. Palaestina Tertia nos. 68–70

Hassan ibn Thabit in Conrad (1981) 154

See also: Stathakopoulos (2004) no. 155, Harper (2017) no. 11

597: Thessalonica and countryside

Mir. Demetr., 3 & 14

See also: Stathakopoulos (2004) no. 156, Harper (2017) no. 12

598: Thrace

Theophylact Simocatta, *Hist.* 7.15.2

See also: Stathakopoulos (2004) no. 159, Harper (2017) no. 13

599–600: Constantinople, Asia Minor and Bithynia, Syria, North Africa, Italy, Marseille

Michael the Syrian, 10.23 (387)

Chron. ad a. 1234

Gregory the Great, *Ep.* 9.232, 10.20

Paul the Deacon, *Hist. Langobardorum,* 4.14

Elias of Nisibis, a. 911

Thomas of Margâ, *Book of Governors,* 11

Fredegar, *Chron.* 4.18

See also: Stathakopoulos (2004) no. 160, Harper (2017) no. 14, Biraben and LeGoff (1975) 75

609: Cortijo de Chinales, Spain

CIL II 7.677

See also: Harper (2017) no. 15, McCormick (2016) 327

619: Constantinople, Alexandria

Mirac. sanct. Artemii, 34

See also: Stathakopoulos (2004) no. 173, Harper (2017) no. 17

626–628: Palestine, Mesopotamia

Michael the Syrian, 11.3 (409)

Eutychius, *Annales*

al-Tabari, 1061

Arabic sources in Conrad (1981) 159–163

See also: Stathakopoulos (2004) nos. 177, 178; Harper (2017) no. 18

638–639: Palestine, Syria, Mesopotamia

Michael the Syrian, 11.8 (423)

Elias of Nisibis, (AH 18)

Chron. ad a. 1234, 76 (AH 18)

Arabic sources in Conrad (1981) 167ff.

See also: Stathakopoulos (2004) no. 180, Harper (2017) no. 20

663–666: England, Ireland

Adomnán, *Vita Columbae* 47

Bede, *Hist. eccl.* 3.23, 27, 30; 4.1, 7, 8

Bede, *Vit. Cuthb.* 8

See also: Harper (2017) no. 21, Maddicott (2007)

670–671: Kufa

Arabic sources in Conrad (1981) 250–253

See also Stathakopoulos (2004) no. 185, Harper (2017) no. 22

672–673: Egypt, Palestine, Mesopotamia, Kufa

Theophanes, *Chron.* AM 6164

Agapios, *Kitab al-'Unwan*

Arabic sources in Conrad (1981) 253ff.

See also: Stathakopoulos (2004) nos. 185, 186; Harper (2017) no. 23

680: Rome, Pavia

Paul the Deacon, *Hist. Langobardorum* 6.5

Liber pontificalis 81

See also Stathakopoulos (2004) no. 192, Harper (2017) no. 24

684–687: England, Ireland

Adomnán, *Vita Columbae* 47

Bede, *Hist. eccl.* 4.14

See also Harper (2017) no. 25, Maddicott (2007)

687–689: Syria, Mesopotamia, Basrah

John bar Penkaye, *Riš Mellē* XV 160–165 (68–71)

Arabic sources in Conrad (1981) 263ff.

See also: Stathakopoulos (2004) nos. 194, 195; Harper (2017) no. 26

689–690: Egypt

Arabic sources in Conrad (1981) 271ff.

See also: Stathakopoulos (2004) no. 196, Harper (2017) no. 27

693: Spain, southwestern Gaul

Mozarabic Chronicle of 754, 41

See also: Harper (2017) no. 28, Kulikowski (2007) 153–154

698–700: Constantinople, Syria, Mesopotamia

Elias of Nisibis (AH 79 and 80)

Chron. ad a. 819, AG 1011

Theophanes, *Chron.* AM 6190 & 6192

Nikephoros, *Brev.* 41

Leo Grammaticus, *Chron.* 167

Arabic sources in Conrad (1981) 274ff.

See also: Stathakopoulos (2004) nos. 198–200; Harper (2017) no. 29

704–706: Syria, Mesopotamia, Basrah, Kufa

Michael the Syrian, 11.17 (449)

Chron. Zuqnin a. 1016

Arabic sources in Conrad (1981) 278ff.

See also: Stathakopoulos (2004) nos. 201, 203; Harper (2017) no. 30

707–709: Spain

Akhbar majmu'a, 7.BkS

See also Harper (2017) no. 31, Kulikowski (2007)

713: Syria

Chronicle of Disasters a. 1024

Michael the Syrian, 11.17 (452)

Chron. ad a. 819 & ad a. 846 a. 1024

See also Stathakopoulos (2004) no. 205, Harper (2017) no. 32

714–715: Egypt

Severos, *History of the Patriarchs* 17

See also Stathakopoulos (2004) no. 207, Harper (2017) no. 33

718–719: Syria, Mesopotamia, Basrah

Arabic sources in Conrad (1981) 286ff.

See also: Stathakopoulos (2004) no. 209, Harper (2017) no. 34

725–726: Syria, Mesopotamia

Theophanes, *Chron.* AM 6218

Vita Willibaldi 4

Michael the Syrian, 11.19 (436)

Agapios, Kitab al-'Unwan

Elias of Nisibis (AD 107)

Chron. ad a. 819 a. 1036

See also Stathakopoulos (2004) no. 213, Harper (2017) no. 35

729: Syria

Michael the Syrian, 11.21 (463)

Harper (2017) no. 36

732–735: Egypt, Palestine, Syria, Mesopotamia

Theophanes, *Chron.* AM 6225

Agapios, *Kitab al-'Unwan*

Arabic sources in Conrad (1981) 291ff.

See also Stathakopoulos (2004) no. 214, Harper (2017) no. 37

743–750: Egypt, North Africa, Syria, Mesopotamia, Basrah, Sicily, Italy, Greece, Constantinople, Armenia

Severos, *History of the Patriarchs* 18

Michael the Syrian, 11.22 (465– 66)

Chron. Zuqnin a. 1055–1056, a. 1061– 62

Chron. ad a. 1234

Theophanes, *Chron.* AM 6238

Nikephoros, *Brev.* 67

Nikephoros, *Antirhetikos III* 496B

Theodore Studites 805B–D

Michael Glycas, *Annales* 527

John Zonaras, *Epit. hist.* 15.6

John of Naples, *Gesta episcoporum neapolitanorum* 42

Arabic sources in Conrad (1981) 293ff.

See also: Stathakopoulos (2004) nos. 218–222, Harper (2017) no. 38, McCormick (2007) 292

Primary sources:

Adomnán, *Vita Columbae = Adomnán's Life of St. Columba.* Ed. and tr. A.O. Anderson and M. O. Anderson (Oxford 1991)

Akhbar majmu'a = A History of Early Al-Andalus: The Akhbar Majmu'a. Tr. D. James (New York 2012)

al-Tabari = *The Sāsānids, the Byzantines, the Lakhmids, and Yemen.* Tr. C. E. Bosworth (New York 1999).

Amr ibn Matta = *Maris, Amri et Slibae de Patriarchis Netsorianorum commentaria.* Ed. H Gismondi (Rome 1896).

Annals of Tigernach = The Annals of Tigernach. Ed. W. Stokes. *Revue Celtique* 16–17 (1895–1896). Reprint, 2 vols. (Dyfed 1993).

Agapios, *Kitab al-'Unvan = Kitab al-'Unvan.* Ed. A. Vasiliev. In *PO* 5.4, 7.4; 8.3; 11.1 (1910, 1911, 1912, 1915).

Agathias = *Agathiae Myrinei Historiarum Libri Quinque.* Ed. R. Keydell. *CFHB* 2 (Berlin 1967).

Barhadbšabba = *Cause de la foundation des écoles.* Ed. A. Scher. In *PO 4*, 388–389.

Bede, *Hist. eccl. = Bede's Ecclesiastical History of the English People.* Eds. B. Colgrave, R. A. B. Mynors (Oxford 1991).

Bede, *Vit. Cuthb. = Two Lives of St. Cuthbert*. Ed. B. Colgrave (Cambridge 1940).

Chron. ad a. 640 = The Seventh Century in the West-Syrian Chronicles. Tr. A. Palmer. *Translated texts for historians 15* (Liverpool 1993).

Chron. ad a. 846 = Chronicon ad a. 846. Ed. I. B. Chabot, E. W. Brooks. CSCO 4, Syr. 4 (Louvain 1907)

Chron. ad a. 819 = Anonymi auctoris Chronicon ad annum Christi 1234 pertinens, I. praemissum est Chronicon Anonymum ad A. D. 819 pertinens. Ed. J.-B. Chabot. In CSCO 81; Syr. 36 (Paris 1916).

Chron. ad a. 1234 = Anonymi auctoris Chronicon ad annum Christi 1234 pertinens, I. praemissum est Chronicon Anonymum ad A. D. 819 pertinens. Ed. J.-B. Chabot. In CSCO 81; Syr. 36 (Paris 1916).

Chronicle of Disasters = A Chronicle of Disasters dated AD 716. In: *The Seventh Century in the West-Syrian Chronicles*. Tr. A. Palmer. *Translated texts for historians 15*, (Liverpool 1993)

Chron. Seert = Chronicle of Seert (Histoire Nestorienne). Ed. A. Scher. In *PO* 4, 213–313; 5, 217–344; 7, 93–203.

Chron. Zuqnin = Chronicon anonymum pseudo-dionysianum vulgo dictum. Ed. J.-B. Chabot. CSCO 91, 104; Syr. 43, 53 (Paris 1927, 1933)

Victoris Tunnunensis Chronicon = Victoris Tunnunensis Chronicon cum reliquiis ex consularibus Caesaraugustanis et Iohannis Biclarensis Chronicon. Ed. C. Cardelle de Hartmann. Corpus Christianorum, Series Latina 173A (Turnhout 2001).

Cyril of Scythopolis = *Kyrillos von Skytopolis.* Ed. E. Schwartz. TU 49.2 (Leipzig 1939).

Elias of Nisibis = *Opus Chronologicum.* Ed. E. W. Brooks. CSCO 63, Syr. 23 (Louvain 1964).

Eutychius, *Annales = Annales.* Ed. L. Cheikho, CSCO 50–51 (Paris 1904).

Evagrius *Hist. eccl. = Historia ecclesiastica.* Eds. J. Bidez, L. Parmentier. *The Ecclesiastical History of Evagrius* (London 1898).

Fredegar, *Chron*. = Fredegar, *Chronica*. Ed. B. Krusch, SRM, 2.128.5–6 (Hannover 1888).

Gregory of Tours, *Lib. hist. = Libri historiarum X.* Eds. B. Krusch, W. Levison. MGH SRM, 1.1 (Hannover 1937–1951).

Gregory of Tours, *Liber in Gloria confessorum = Liber in Gloria confessorum*. B. Krusch, MGH SRM, 1.2 (Hannover 1885).

Gregory of Tours, *Lib. vitae partum = Liber vitae partum*..Ed. B. Krusch, MGH SRM, 1.2 (Hannover 1885).

Gregory the Great, *Dial. = Dialogorum libri iv.* Ed. A: de Vogüé. SC 251, 260, 265 (Paris 1978–1980).

Gregory the Great, *Ep. = Registrum epistularum.* Ed. D. Norberg. CC 140–40A (Turnhout 1982).

John bar Penkaye, *Riš Mellē = North Mesopotamia in the Late Seventh Century: Book XV of John bar Penkaye's Riš Mellē*. S. Brock. Jerusalem Studies in Arabic and Islam 9 (Jerusalem 1989).

John of Biclaro = *Chronica.* Ed. T. Mommsen. MGH AA 9, 2, 213–214 (Berlin 1894).

John of Ephesos *Fragment E* = *Historiae Ecclesiasticae fragmenta.* Eds. W. J. van Douwen, J. P. N. Land. *Verhandelingen der koniglijke Akademie van Wetenschapen Afdeeling Letterkunde* 18, 197–264 (Amsterdam 1899).

John Malalas = *Ioannis Malalae chronographia*. Ed. I. Thurn (Berlin 2000).

John of Naples, *Gesta episcoporum neapolitanorum* = *Johannes Diaconus, Gesta episcoporum neapolitanorum*. Ed. G. Waitz. MGH SRL, 425.15–19 (Hannover 1878).

John of Nikiu = *The Chronicle of John, Bishop of Nikiu.* Ed. R. H. Charles (London 1916).

John Zonaras, *Epit. hist.* = *Ioannis Zonarae epitome historiarum*. Ed. L. Dindorf (Leipzig 1870).

Leo Grammaticus, *Chron.* = *Leonis Grammatici Chronographia*. Ed. I. Bekker. CSHB 42 (Bonn 1842).

Leontios of Neapolis, *Vita of Symeon Salos* = *Das Leben des heiligen Narren Symeon*. Ed. L. Rydén. *Acta Universitatis Upsalensis. Studia Graeca Upsaliensia* 4, 151 (Stockholm, Göteborg, Upsala 1963).

Liber pontificalis = *Liber pontificalis.* Ed. T. Mommsen. MGH Gesta Pont. Rom. 1 (Hannover 1898).

Marius of Avenches = *Chronica.* Ed. T. Mommsen. MGH AA 9, 2, 225–240 (Berlin, 1894).

Marcellinus Comes = *Chronicon ad annum DXVIII.* Ed. T. Mommsen, MGH AA 11, 60–108 (Hannover 1894).

Michael Glycas, *Annales* = *Michaelis Glycae Annales*. Ed. I. Bekker. CSHB 24 (Bonn 1836).

Michael the Syrian = *Chronique de Michel le Syrien, patriarche Jacobite d'Antioche (1166–1199).* Ed. J.-B. Chabot (Paris 1899–1924).

Mir. Demetr. = *Les plus anciens recueils de Miracles de Saint Démétrius et la pénétration des Slaves dans les Balkans.* P. Lemerle. 2 vols. (Paris 1979–1981).

Mirac. sanct. Artemii = *Miracula Sancti Artemii.* Ed. A. Papadopoulos-Kerameus. *Varia Graeca Sacra*. 34, 52 (Saint Petersburg 1909).

Mozarabic Chronicle of 754 = *Cronica Mozarabe de 754: Edicion critica y traduccion.* Ed. and tr. J. E. Lopez Pereira (Zaragoza, 1989).

Nikephoros, *Antirhetikos III* = *Nikephoros Patriarches*, *Antirhetikos* III. In PG 100.

Nikephoros, *Brev.* = *Nikephoros, Patriarch of Constantinople: Short History. Ed. C. Mango. CFHB 13 (Washington 1990).*

Paul the Deacon = *Historia Langobardorum*. Ed. G. Waitz. SS rer. Langobard. (Hannover 1878).

Prokopios *BP* = *Persian Wars*. Eds. J. Haury, G. Wirth, *Opera*. Vol. 1 (Leipzig 1962–1964).

Severos, *History of the Patriarchs* = *History of the Patriarchs of the Coptic Church of Alexandria.* Ed. B. Evetts, PO 1.2, 5.1, 10.5 (Paris 1906–1915).

Theodore of Sykeon, *Vita* = *La vie de Théodore de Sykéon.* Ed. A. J. Festugière. *Subsidia Hagiographica* 48 (Brussels 1970).

Theodore Studites = *Laudatio Platonis*. In PG99.

Theophanes = *Theophanis chronographia.* Ed. C. de Boor (Leipzig 1883).

Theophylact Simocatta, *Hist.* = *Theophylacti Simocattae historiae.* Ed. C. de Boor (Leipzig 1997).

Thomas of Margâ, Book of Governors = *The Book of Governors: the Historia monastica of Thomas, bishof of Margâ A.D. 840.* Ed. E. A. W. Budge (London 1893).

Victor of Tunnuna = *Chronica a. CCCCXLIV–DLXVII.* Ed. T. Mommsen, MGH AA 11, 184–206 (Berlin 1894).

Vita of Nicholas of Sion = *The Life of Saint Nicholas of Sion.* Eds. I. Ševčenko, N. P. Ševčenko (Brookline, 1984).

Vita of Symeon Stylites Iunior = *La vie ancienne de S. Syméon Stylite le jeune (521–592).* Ed. P. van den Ven (Brussels 1962).

Vita Willibaldi = *Vitae Willibaldi et Wynnebaldi auctore sanctimoniali Heidenheimensi.* Ed. O. Holder-Egger, MGH SS, 15/1 (Hannover 1887).

Zacharias Rhetor = *Historia Ecclesiastica.* Ed. E. W. Brooks, CSCO Syr. 3.6 (Louvain, 1924).

Inscriptions:

Byzantina Siciliae = *Byzantina Siciliae*. Ed. G. Manganaro. *Minima Epigraphica et Papyrologica* 4, 133 (2001).

Inscriptiones Christianae Urbis Romae = Inscriptiones Christianae Urbis Romae septimo seculo antiquiores, Nova Series (= ICUR NS) Eds. I.B. de Rossi et al. 10 vols. (Rome 1922–1992).

CIL = *Corpus Inscriptionum Latinarum* (Berlin 1863–).

Epigraphic material from Sufetula = Nouvelles recherches d'archéologie et d'épigraphie chrétienne à Sufetula (Byzacène). N. Duval. *Mélanges d'archéologie de l'École française de Rome* 68 (1956).

I. Palaestina Tertia = *Inscriptions from Palaestina Tertia*. Eds. Y. Meimaris, K. Kritikakou (Athens 2005).

Secondary sources:

L. I. Conrad: *The Plague in the Early Medieval Near East* (PhD, Princeton University 1981).

L. I. Conrad: Die Pest und ihr soziales Umfeld im Nahen Osten des frühen Mittelalters. *Der Islam* 73(1996): 81–112.

A. Dooley: The Plague and Its Consequences in Ireland. In: *Plague and the End of Antiquity: The Pandemic of 541–750.* Ed. L. K. Little (Cambridge 2007).

J.-N. Biraben, J. Le Goff: The Plague in the Early Middle Ages. In: *Biology of Man in History.* Ed. R. Forster and O. Ranum, trans. E. Forster and P. M. Ranum (Baltimore 1975).

K. Harper: *The Fate of Rome. Climate, Disease & the End of an Empire.* (Princeton, Oxford 2017).

J. Koder: Ein inschriftlicher Beleg zur Justinianischen Pest in Zora (Azra'a). Byzantinoslavica 56 (1995): 12–18.

M. Kulikowski: Plague in Spanish Late Antiquity. In: *Plague and the End of Antiquity: The Pandemic of 541–750.* Ed. L. K. Little (Cambridge 2007).

J. Maddicott: Plague in Seventh-Century England. In: *Plague and the End of Antiquity: The Pandemic of 541–750.* Ed. L. K. Little (Cambridge 2007).

M. McCormick: Tracking mass death during the fall of Rome's empire (II): a first inventory of mass graves. *Journal of Roman Archaeology* 29 (2016):1004–1046.

M. McCormick: Toward a Molecular History of the Justinianic Pandemic. In: *Plague and the End of Antiquity: The Pandemic of 541–750.* Ed. L. K. Little (Cambridge 2007).

E. Phillimore: The Annales Cambriae and Old-Welsh Genealogies from Harlein MS 3859. *Y Cymmrodor* 9 (1888):141–83.

D. C. Stathakopoulos: *Famine and Pestilence in the Late Roman and Early Byzantine Empire.* (Burlington 2004).

C. Tsiamis et al.: The Red Sea and the Port of Clysma. A Possible Gate of Justinian's Plague. *Gesnerus* 66 (2009):209–217.

D. Woods: *Acorns, the Plague and the 'Iona Chronicle.' Peritia* 17–18 (2003–4): 495–502.

Fig. S1: Heterozygosity plots for the twelve genomes with higher than 4.5-fold coverage. The y-axis displays percentage of allele frequency, the reference calls (0 %) and alternative calls (100 %) were excluded for scaling.

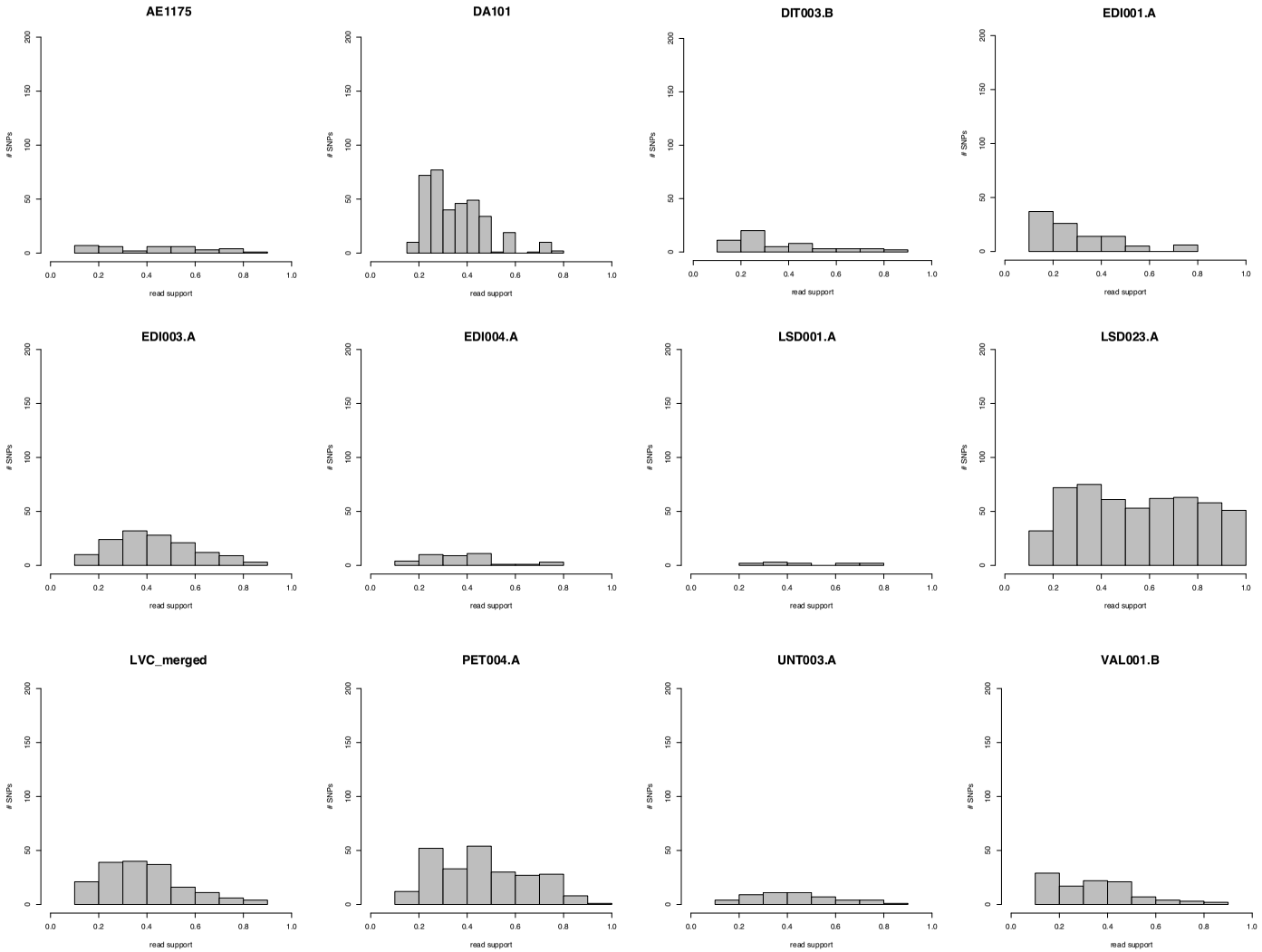


Fig. S2: Stacked bar chart for the false positive SNPs called in the background (DIR002.A) and the contaminated simulated datasets. The "Ns" are called due to heterozygosity (grey), all false positive SNPs were filtered out by our SNP evaluation (orange). With increasing coverage, the foreground has a high enough coverage to outshine the contamination and enable a reference call. See also SI Appendix, Table S6.

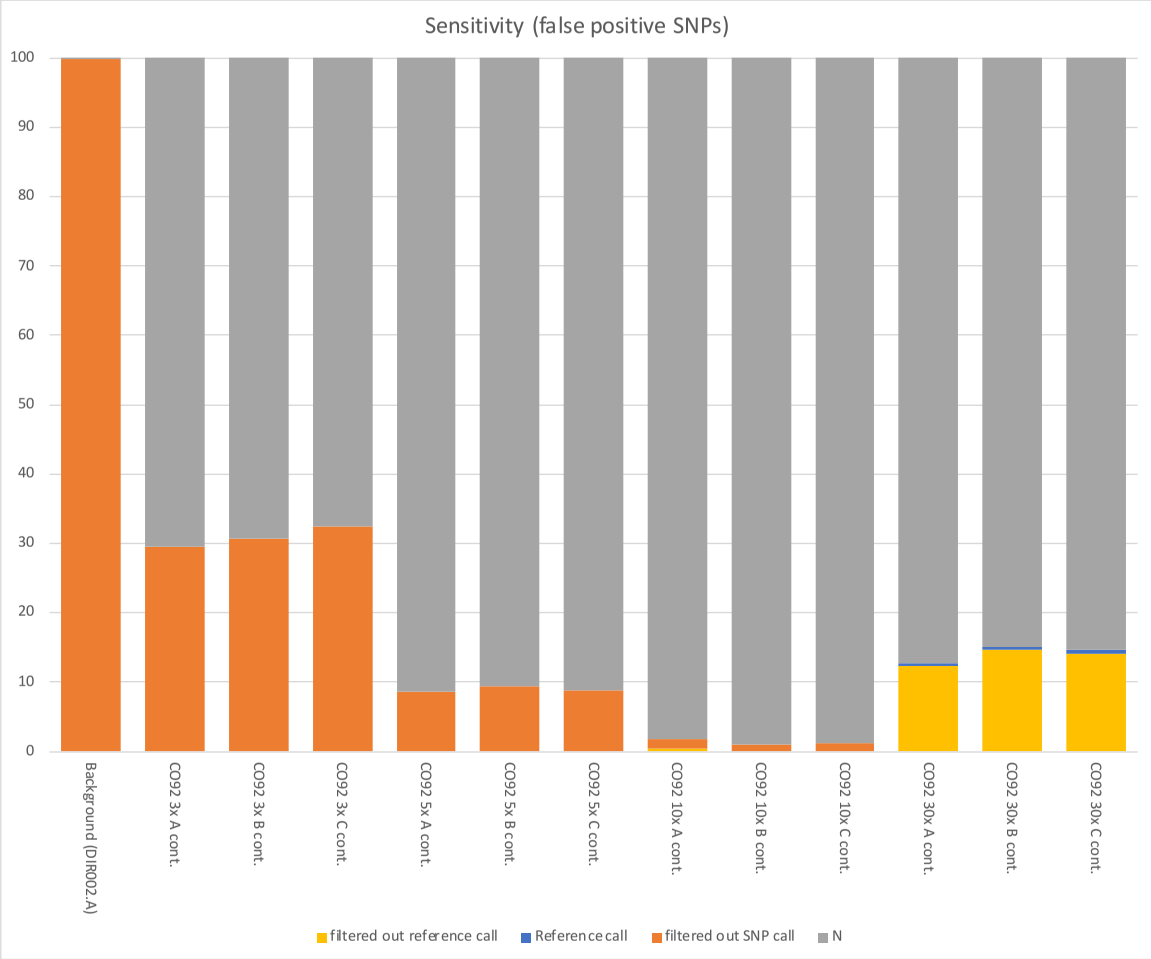


Fig. S3: Stacked bar chart for the 418 core positions for the backbone phylogeny of modern *Y. pestis* in the simulated datasets with ("Contaminated", right bar) and without ("Clean", left bar) background (DIR002.A). Whereas for the clean datasets, the number of "Ns" (grey) and filtered positions (orange) are only dependent on coverage and go to zero for the 30-fold coverage datasets, more positions are lost as "Ns" due to heterozygosity or filtered out due to contamination. See also SI Appendix, Table S7.

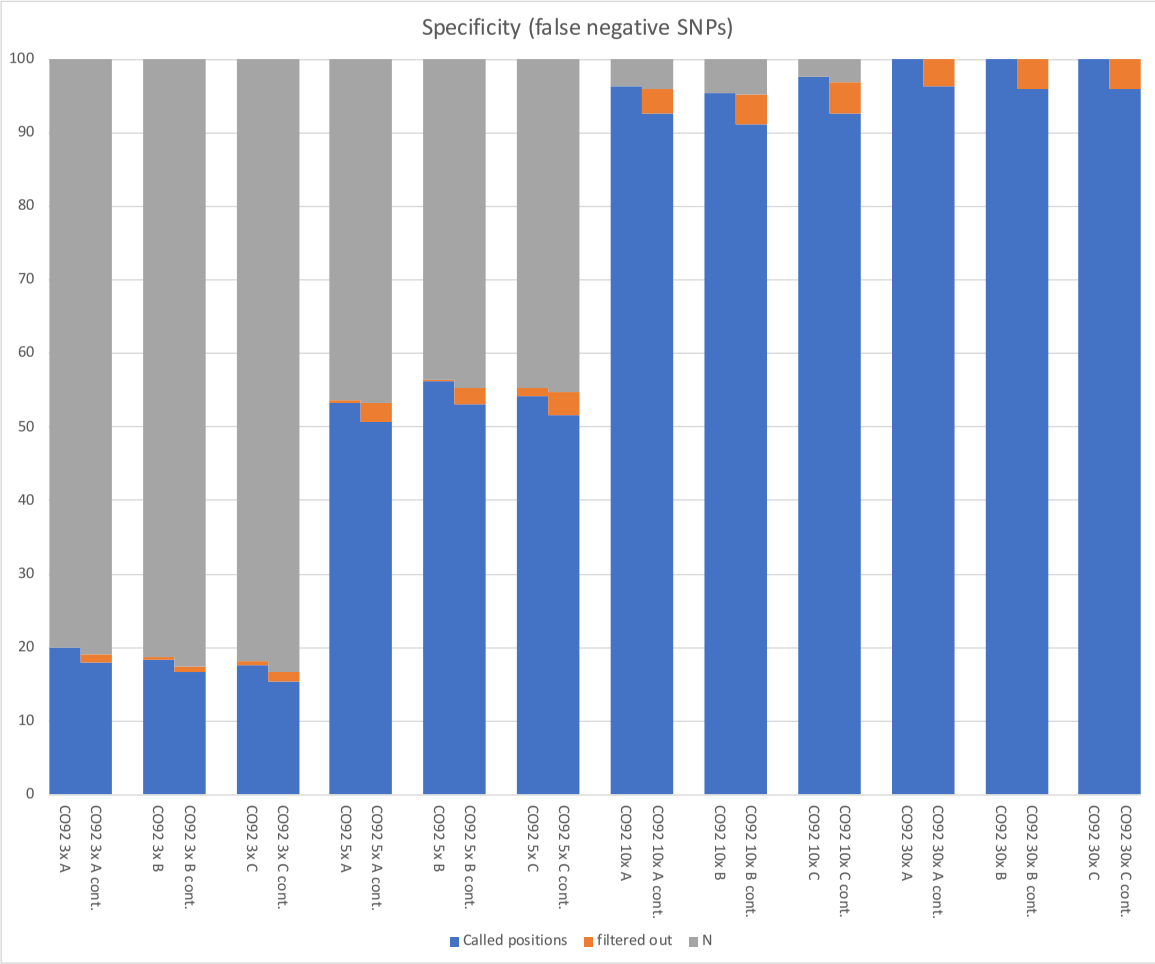


Fig. S4: Maximum Likelihood tree with full SNP alignment (6580 positions) of 233 modern *Y. pestis* and one *Y. pseudotuberculosis* genome, ten published (2nd–3rd century in orange; Altenerding in blue; Second Pandemic in red) and eight newly reconstructed genomes (green). Numbers on node are showing bootstrap values (1000 iterations).

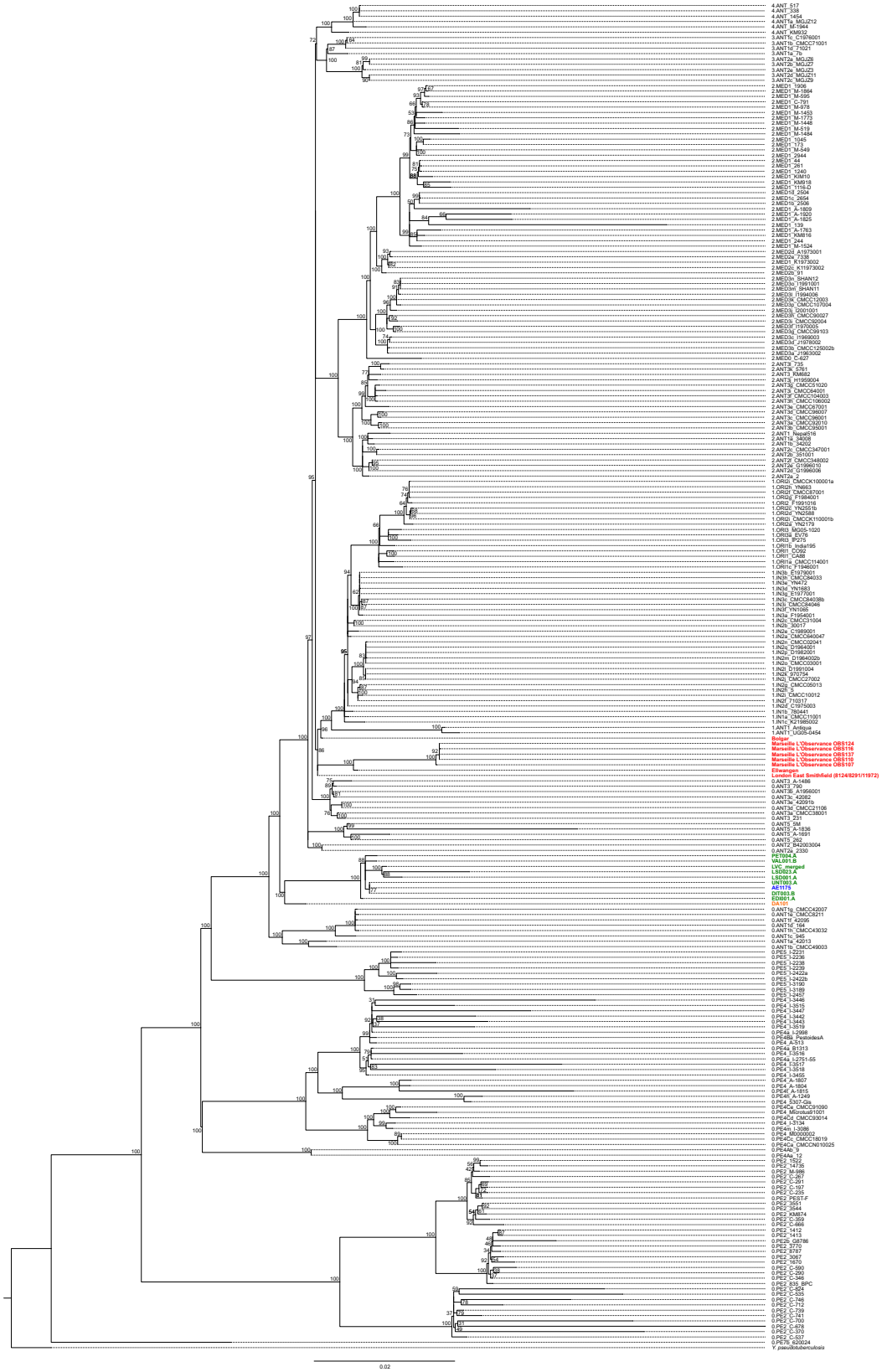


Fig. S5: Heatmap showing the percentage of coverage of plasmidal virulence factors in all analyzed First Pandemic genomes.

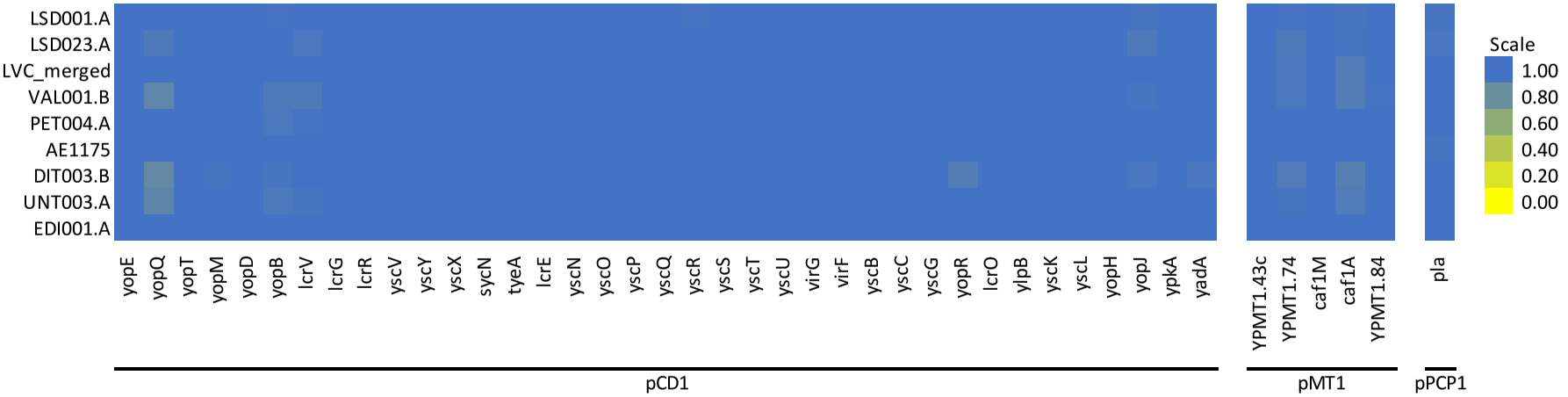


Fig. S6: Heatmap showing the percentage of coverage of the ca. 45 kb region of the chromosome deleted in LSD001.A, LSD023.A, LVC_merged and Marseille. First Pandemic genomes (blue and green) and Second Pandemic genomes (red) are shown in combination with selected strains of main clades of modern *Y. pestis* diversity on Branch 0 as well as the reference genomes of *Y. pseudotuberculosis* and *Y. pestis* (CO92).

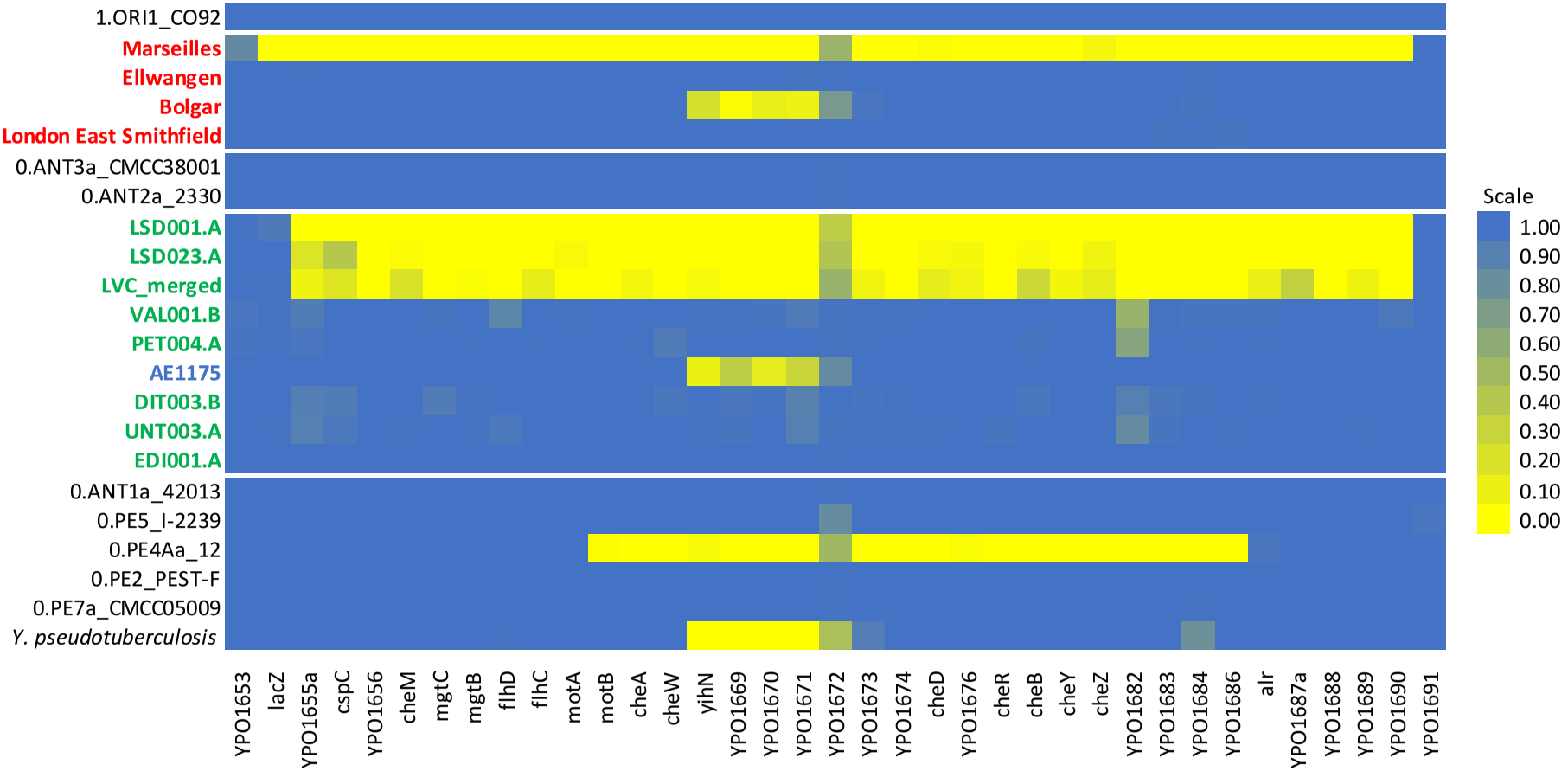


Fig. S7: Heatmap showing the percentage of coverage of a region on the chromosome (A) deleted in LSD001.A, LSD023.A and LVC_merged, and a region on the pMT1 plasmid (B) deleted in LSD001.A, LSD023.A, LVC_merged and UNT003.A.

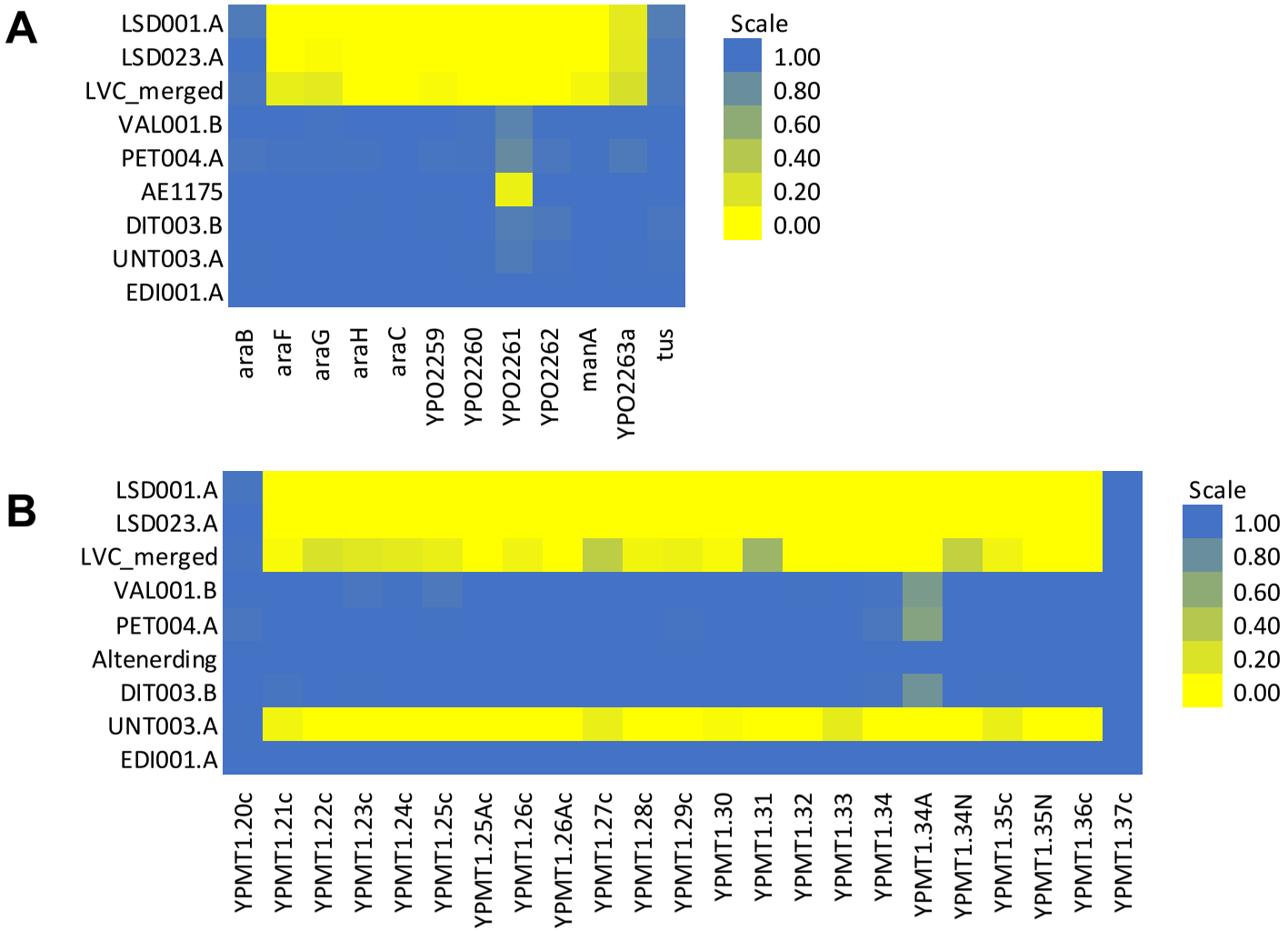


Fig. S8: Combined figure showing the probability plots for all radiocarbon samples (blue, 2-sigma ranges, see SI Appendix, Table S13 for details). The archaeological dating is shown as light blue box. Years with documented plague outbreaks are highlighted in red (Europe) and light orange (only Asia/Africa). The schematic tree of the presented genomes is shown on the left.

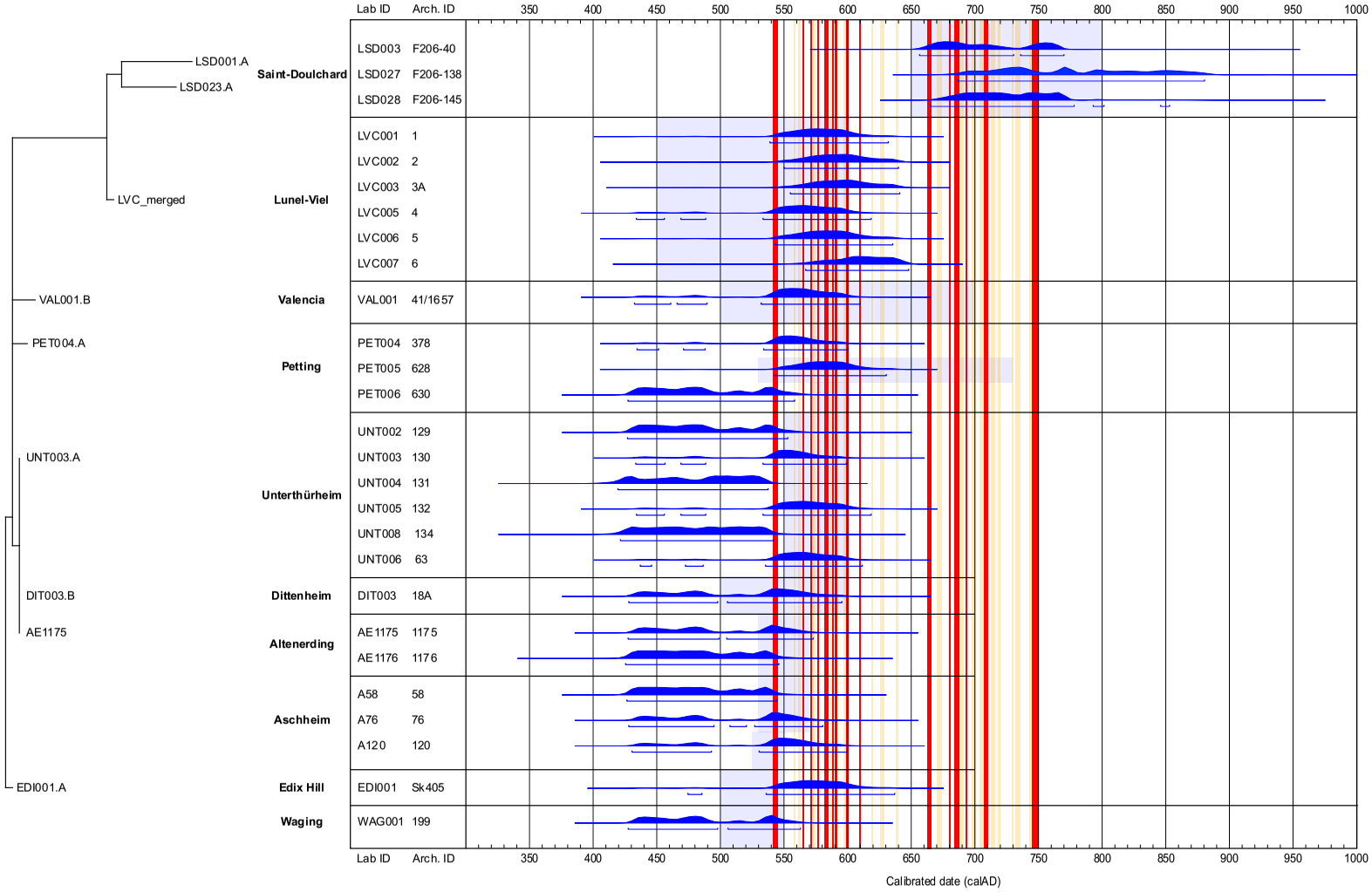


Fig. S9: Historical geography of the tested sites in Germany and Austria. A: Detailed map of all tested sites in Germany and Austria in relation to main rivers (blue), the Roman road network (orange) and the Roman provinces in their last state before the fall of the Western Roman Empire (shaded regions); all sources are given in the SI Appendix. The sites of published genomes are depicted by pink squares, the sites for the genomes presented here are represented by yellow squares. Sites tested negative are depicted in black upward-pointing triangles (burials dating before 541), squares (dating around 541–544), downward-pointing triangles (dating after 544). B: Relation of the detailed map (rectangular space) to the sites in Britain (Edix Hill), France (Lunel-Viel, Saint-Doulchard) and Spain (Valencia) presented in this study.

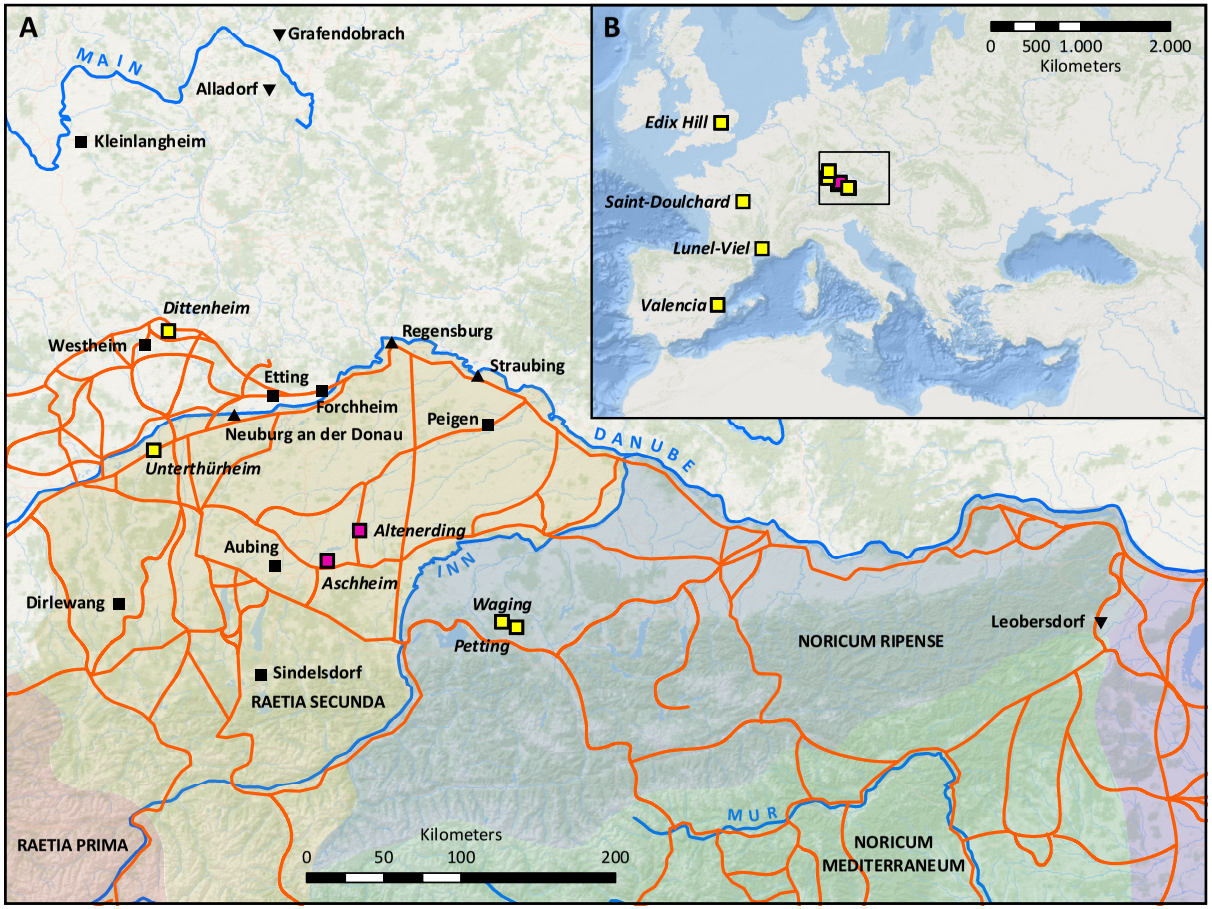


Fig. S10: Screenshot of IGV showing the positions 8528/8529 on the pPCP1 plasmid (175043/175044 in the merged reference) showing the 2-bp deletion causing mismapping.

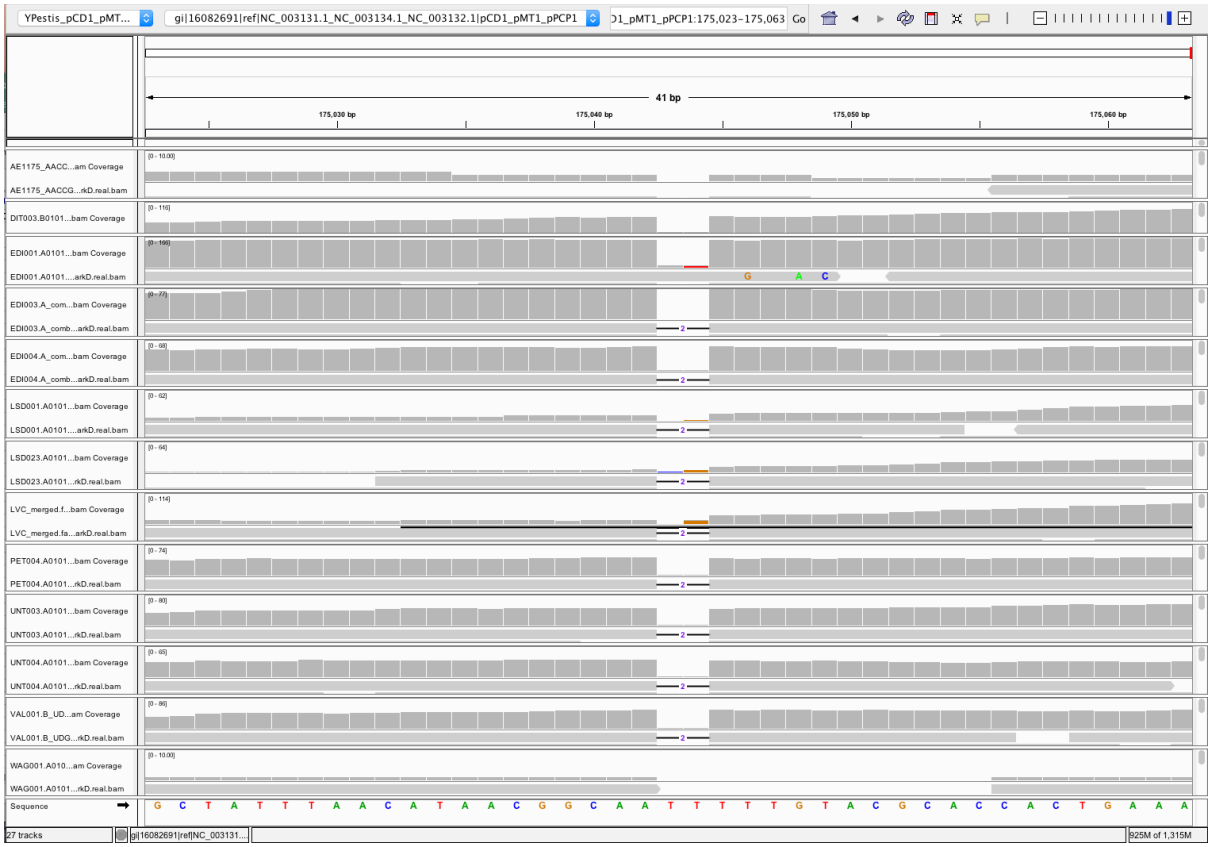


Table S1: Table of all tested individuals and samples. 'Group' gives the archaeological IDs of simultaneously buried individuals or individual from the same collective burial (Valencia: for details see site description, SI Apendix). 'Dating' gives the range of the archaeological dating CE for each burial. 'Total' list the total number of individuals, further broken down to adults (>20 ya) and subadults (<20 ya). 'Arch. ID' and 'Lab ID ind.' are showing the individual ID with the morphological 'sex'(f=female, m=male, ?=indeterminable) and 'age' (i=infans 1, ii=infans 2, j=juvenile, a=adult, m=mature, s=senile, (a)=at least adult, e=early, l=late). Teeth are noted in the FDI system or letters (I=Incisor, C=Canine, P=premolar, M=molar, lower case letters for deciduous teeth, indet=indeterminable). The qPCR screening results for *Y. pestis* are given in the 'Scr' column, samples screened with MALT are indicated by a star. The 'Seq' column shows if samples were considered as positive after sequencing.

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Alladorf	179/180	700–900	2	1	1	179	ALL001	f	ea		47	ALL001.A	neg	
						180	ALL002	?	i		64	ALL002.A	neg	
	184/185	700–900	2	1	1	184	ALL003	?	ii		85	ALL003.A	neg	
						185	ALL004	m	la–em		28	ALL004.A	neg	
	203/204	700–900	2	1	1	203	ALL005	m	la		18	ALL005.A	neg	
						204	ALL006	?	i	May be a later burial	74	ALL006.A	neg	
Dirlewang	33/34	650–700	2	1	1	33	DIR001	f	j–a		51	DIR001.A	neg	
	38/39	650–675	2	2	0	39	DIR002	f	a		17	DIR002.A	pos	neg
Dittenheim	188(A+B)	550–700	2	0	2	188A	DIT001	?	i		37	DIT001.A	neg	
						188B	DIT002	?	i		37	DIT002.A	neg	
	18(A+B)	**500–600**	**2**	**2**	**0**	**18A**	**DIT003**	**m**	**m**		18	DIT003.A	neg	
											M	**DIT003.B**	**pos**	**pos**
	18(A+B)	**500–600**	**2**	**2**	**0**	**18B**	**DIT004**	**f**	**a**		**17**	**DIT004.A**	**pos**	**pos**
											M	**DIT004.B**	**pos**	**pos**
	22(A+B)	600–700	2	1	1	22A	DIT005	?	i		85	DIT005.A	neg	
							DIT006	f	a		38	DIT006.A	neg	
	8(A+B)	500–700	2	2	0	8A	DIT007	f	m		18	DIT007.A	neg	

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Edix Hill	**Grave 76**	**500–550**	**1**	**0**	**1**	**Sk405**	**EDI001**	**?**	**j**		**M**	**EDI001.A**	**pos***	**pos**
	Grave 69	500–550	1	1	0	Sk359	EDI002	f	ea		M	EDI002.A	pos*	neg
	Grave 78	**500–550**	**1**	**0**	**1**	**Sk424**	**EDI003**	**?**	**j**		**M**	**EDI003.A**	**pos***	**pos**
	Grave 96	**500–550**	**2**	**1**	**1**	**Sk547B**	**EDI004**	**?**	**ii**	**Simultaneous double burial**	**M**	**EDI004.A**	**pos***	**pos**
	Grave 106	**500–550**	**2**	**2**	**0**	**Sk626A**	**EDI005**	**f**	**ea**	**Simultaneous double burial**	**M**	**EDI005.A**	**pos***	**pos**
						Sk626B	EDI008	m	ea	Simultaneous double burial	M	EDI008.A	neg	
	Grave 2	550–650	4	4	0	Sk3C	EDI009	m	ea	Sk3B and Sk3C simultaneous	M	EDI009.A	neg	
	Grave 9	500–550	2	1	1	Sk13A	EDI010	m	m-s	Simultaneous double burial	M	EDI010.A	neg	
	Grave 18	550–650	3	2	1	Sk42B	EDI006	f	ea	Nonsimultaneous double burial	M	EDI006.A	neg	
	Grave 46	500–550	1	1	0	Sk146	EDI011	m	em		M	EDI011.A	neg	
	Grave 60	550–650	1	1	0	Sk183	EDI012	f	a		M	EDI012.A	neg	
	Grave 63	500–650	1	1	0	Sk198	EDI013	m	ea		M	EDI013.A	neg	
	Grave 66	500–650	2	2	0	Sk322A	EDI014	m	(a)	Nonsimultaneous double burial	M	EDI014.A	neg	
	Grave 83	500–550	1	1	0	Sk436	EDI015	f	ea		M	EDI015.A	neg	
	Grave 84	500–550	2	1	1	Sk440A	EDI016	f	a	Simultaneous double burial	M	EDI016.A	neg	
	Grave 88	500–650	1	1	0	Sk453A	EDI017	m	ea		M	EDI017.A	neg	
	Grave 90	500–650	1	1	0	Sk458	EDI007	f	s		M	EDI007.A	neg	
	Grave 95	500–550	1	1	0	Sk530	EDI018	f	lm		M	EDI018.A	neg	
	Grave 97	550–650	1	1	0	Sk551	EDI019	m	ea		M	EDI019.A	neg	
	Grave 99	500–650	1	1	0	Sk576	EDI020	m	m		M	EDI020.A	neg	
	Grave 100	500–650	1	1	0	Sk578	EDI021	m	ea		M	EDI021.A	neg	
	Grave 105	500–650	1	1	0	Sk592	EDI022	f	lm		M	EDI022.A	neg	
Forchheim	1–4	650–700	4	2	1	1	FOR001	m	j		27	FOR001.A	neg	
						2	FOR002	f	m	Prone position	27	FOR002.A	neg	
						3	FOR003	m	m		47	FOR003.A	neg	
Grafendobrach	42/43/44	850–930	3	1	2	42	GRA001	m	es		46	GRA001.A	neg	
	83/84/85	850–930	3	2	1	83	GRA002	f?	em	May be a later burial	16	GRA002.A	neg	
						85	GRA003	f	ea		16	GRA003.A	neg	
Kleinlangheim	211/212	630–720	2	min 1	?	211	KLH001	m	m	Second burial: cremation	47	KLH001.A	neg	
	218/219	560–630	2	2	0	218	KLH002	m	ma–m		17	KLH002.A	neg	
						219	KLH003	f	la–em		23	KLH003.A	neg	
	272/273	670–720	2	2	0	273	KLH004	m	s		33	KLH004.A	neg	
	35/36	630–720	2	2	0	35	KLH005	m	a		37	KLH005.A	neg	
Leobersdorf	82(A–C)	770–800	3	1	2	82A	LEO001	m	j		36	LEO001.A	neg	
						82B	LEO002	f	a		27	LEO002.A	neg	
						82C	LEO003	?	i		55	LEO003.A	neg	

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Saint-Doul-chard	**F 206-31**	**750–800?**	**1**	**1**	**0**	**31**	**LSD001**	**f**	**ea**		**47**	**LSD001.A**	**pos**	**pos**
	F 206-38		**1**	**1**	**0**	**38**	**LSD002**	**f**	**m**		**47**	**LSD002.A**	**pos**	**pos**
	F 206-40		1	1	0	40	LSD003	m	m		17	LSD003.A	neg	
	F 206-45		1	1	0	45	LSD004	m	m–s		17	LSD004.A	neg	
	F 206-51		1	1	0	51	LSD005	m	m		47	LSD005.A	neg	
	F 206-52		1	1	0	52	LSD006	?	ea		47	LSD006.A	neg	
	F 206-65		**1**	**1**	**0**	**65**	**LSD007**	**?**	**m–s**		**47**	**LSD007.A**	**pos**	**pos**
	F 206-68/68		2	1	1	68	LSD008	m	m–s		28	LSD008.A	neg	
	F 206-71/72		2	1	1	71	LSD009	m	s		37	LSD009.A	neg	
	F 206-78		1	1	0	78	LSD010	m	m–s		47	LSD010.A	neg	
	F 206-79		2	1	1	79a	LSD011	m	m–s		47	LSD011.A	neg	
	F 206-93/145		2	2	0	93	LSD012	?	ea		47	LSD012.A	neg	
	F 206-132		**1**	**1**	**0**	**132**	**LSD013**	**m**	**m**		**47**	**LSD013.A**	**pos**	**pos**
	F 206-136		1	1	0	136	LSD014	f	m		37	LSD014.A	neg	
	F 206-144		1	0	1	144	LSD015	?	j		47	LSD015.A	neg	
	F 206-151		1	1	0	151	LSD016	m	m		47	LSD016.A	neg	
	F 206-153		1	0	1	153	LSD017	?	ii		46	LSD017.A	neg	
	F 206-154		1	1	0	154	LSD018	?	ea		38	LSD018.A	neg	
	F 206-155		**2**	**2**	**0**	**155b**	**LSD019**	**m**	**m–s**		**48**	**LSD019.A**	**pos**	**pos**
	F 206-156		**2**	**1**	**1**	**156a**	**LSD020**	**m**	**m–s**		**47**	**LSD020.A**	**pos**	**pos**
	F 206-172		**3**	**1**	**2**	**172c**	**LSD021**	**f**	**m**		**38**	**LSD021.A**	**pos**	**pos**
						172b	**LSD022**	**?**	**j**		**46**	**LSD022.A**	**pos**	**pos**
	F 206-204		**2**	**1**	**1**	**204a**	**LSD023**	**m**	**m**		**47**	**LSD023.A**	**pos**	**pos**
	F 206-208		**1**	**0**	**1**	**208**	**LSD024**	**?**	**j**		**47**	**LSD024.A**	**pos**	**pos**
	F 206-215		1	1	0	215	LSD025	f	m–s		47	LSD025.A	neg	
	F 206-216		**1**	**0**	**1**	**216**	**LSD026**	**?**	**ii**		**36**	**LSD026.A**	**pos**	**pos**

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Lunel-Viel (Les Horts)	38(A–C)	475–550	3	3	0	38C	LVH001	?	(a)	Double burial with remains of a third individual	M	LVH001.A	neg	
											M	LVH001.B	neg	
											M	LVH001.C	neg	
											I	LVH001.D	neg	
											I	LVH001.E	neg	
Lunel-Viel (Quartier centrale)	**1–5**	**400–600**	**6**	**6**	**0**	**1**	**LVC001**	**f**	**ea**		M	LVC001.A	neg	
											M	**LVC001.B**	**pos**	**pos**
											26	**LVC001.C**	**pos**	**pos**
						2	LVC002	?	la–em		M	LVC002.A	neg	
											M	LVC002.B	neg	
						3A	LVC003	f	a		M	LVC003.A	neg	
											M	LVC003.B	neg	
						3B	LVC004	f	a		M	LVC004.A	neg	
											M	LVC004.B	neg	
						4	**LVC005**	**m**	**m**		**M**	**LVC005.A**	**pos**	**pos**
											M	**LVC005.B**	**pos**	**pos**
											43	**LVC005.C**	**pos**	**pos**
						5	**LVC006**	**m**	**ea**		**M**	**LVC006.A**	**pos**	**pos**
											M	LVC006.B	neg	
	6/7	400–600	2	0	2	6	LVC007	f	j–ea		M	LVC007.A	neg	
											M	LVC007.B	neg	
München-Aubing	675/676/677	650–700	3	2	1	676	AUB002	f	m		13	AUB002.A	neg	
						677	AUB003	m	?		16	AUB003.A	neg	
	724/725	400–700	2	2	0	724	AUB004	m	a		46	AUB004.A	neg	
						725	AUB005	m	m		31	AUB005.A	neg	
	809/810	400–700	2	2	0	809	AUB006	m	a		16	AUB006.A	neg	
						810	AUB007	m	a		36	AUB007.A	neg	
	854/855	650–700	2	2	0	854	AUB008	m	a		47	AUB008.A	neg	
	675/676/677	650–700	3	2	1	675	AUB009	?	i		85	AUB009.A	neg	
Neuburg an der Donau	34(N/M/S)	330–360	3	3	0	34M	NEU001	m	m		13	NEU001.A	neg	
						34S	NEU002	m	a		37	NEU002.A	neg	
Peigen	18(1–2)	600–700	2	2	0	18-1	PEI001	f	a	Opposite orientation	37	PEI001.A	pos	neg
						18-2	PEI002	f	ea	Opposite orientation	47	PEI002.A	neg	
	62(1–2)	600–650	2	1	1	62-2	PEI003	f	lm		38	PEI003.A	neg	
	63(1–2)	580–620	2	2	0	63-1	PEI004	f	a		47	PEI004.A	neg	
						63-2	PEI005	m	s		45	PEI005.A	neg	
Petting	342/343/344	530–730	3	2	1	342	PET001	m	ea		37	PET001.A	neg	
						343	PET002	?	ii		74	PET002.A	neg	
	377/378	**570–600**	**2**	**2**	**0**	377	PET003	m	la–m		26	PET003.A	neg	
						378	**PET004**	**m**	**la–em**		**36**	**PET004.A**	**pos**	**pos**
	628	**530–730**	**1**	**0**	**1**	**628**	**PET005**	**?**	**ii**	**Only a few remains, directly above 630/631**	**37**	**PET005.A**	**pos**	**pos**
	630/631	**570–600**	**2**	**2**	**0**	**630**	**PET006**	**f**	**ea**		**46**	**PET006.A**	**pos**	**pos**
						631	PET007	f	em		46	PET007.A	neg	

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Regensburg	30(1–2)	350–450	2	2	0	30-1	RFF001	m	la–em	Prone position	36	RFF001.A	neg	
	53(2/3)	350–450	2	1	1	53-2	RFF002	f	j	Crouched inhumation	26	RFF002.A	neg	
						53-3	RFF003	m	em	Crouched inhumation	46	RFF003.A	neg	
Sindelsdorf	162	630–720	1	1	0	162	SIN001	m	s	162 directly above 163–165	48	SIN001.A	neg	
	163–165	630–720	3	2	1	164	SIN002	m	la–em		17	SIN002.A	neg	
						165	SIN003	f?	m		47	SIN003.A	neg	
	51/52	630–670	2	2	0	52-1	SIN006	m	s		37	SIN006.A	neg	
						52-2	SIN007	m	m		33	SIN007.A	neg	
Straubing	16/17	300–400	2	0	2	16	SAZ001	?	i		85	SAZ001.A	neg	
	54a–c	300–400	3	1	2	54a	SAZ002	m?	j		26	SAZ002.A	neg	
						54b	SAZ003	m	j		27	SAZ003.A	neg	
Unterthürheim	116(1–2)	525–680	2	1	1	116(-2)	UNT001	?	ii		16	UNT001.A	neg	
	129	**550–600**	**1**	**0**	**1**	**129**	**UNT002**	**?**	**i**	**Close to 131–134**	**54**	**UNT002.A**	**pos**	**pos**
	130	**550–600**	**1**	**1**	**0**	**130**	**UNT003**	**m**	**m**	**Close to 131–134**	**37**	**UNT003.A**	**pos**	**pos**
	131/132/133/134	**550–600**	**4**	**2**	**2**	**131**	**UNT004**	**?**	**i**		**55**	**UNT004.A**	**pos**	**pos**
						132	**UNT005**	**m**	**a**		**48**	**UNT005.A**	**pos**	**pos**
	63/64	**550–600**	**2**	**1**	**1**	**63**	**UNT006**	**f**	**a**		**36**	**UNT006.A**	**pos**	**pos**
	63/64	550–600	2	1	1	64	UNT007	?	i		84	UNT007.A	neg	

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Valencia	**41**	**500–700**	**15**	**?**	**min 4**	**1657**	**VAL001**	**?**	**?**	**Collective burial**	M	VAL001.A	neg	
											M	**VAL001.B**	**pos**	**pos**
											P	VAL001.C	neg	
											m	VAL001.D	neg	
											P	VAL001.E	neg	
											m	VAL001.F	neg	
	40	500–700	4	?	?	1728	VAL002	?	?	Burial on apse entrance	I	VAL002.A	neg	
											M	VAL002.B	neg	
											C	VAL002.C	neg	
											indet.	VAL002.D	neg	
											indet.	VAL002.E	neg	
	50	500–700	7	min 2	min 1	10039	VAL003	?	?	Collective burial next to apse	P	VAL003.A	neg	
											M	VAL003.B	neg	
											M	VAL003.C	neg	
											I	VAL003.D	neg	
											I	VAL003.E	neg	
											I	VAL003.F	neg	
						10346	VAL004	?	?	Collective burial next to apse	C	VAL004.A	neg	
	28	500–700	min 21	min 13	min 3	10307	VAL005	?	?	Collective burial next to apse	M	VAL005.A	neg	
											M	VAL005.B	neg	
											M	VAL005.C	neg	
											M	VAL005.D	neg	
											M	VAL005.E	neg	
											M	VAL005.F	neg	
											M	VAL005.G	neg	
											M	VAL005.H	neg	
											Indet.	VAL005.I	neg	
											indet.	VAL005.J	neg	
											M	VAL005.K	neg	
											indet.	VAL005.L	neg	
						10313	VAL006	?	?	Collective burial next to apse	M	VAL006.A	neg	
											m	VAL006.B	neg	
						10322	VAL007	?	?	Collective burial next to apse	M	VAL007.A	neg	
	4	500–700	4	3	1	10066	VAL008	?	?	*Tegula* burial	M	VAL008.A	neg	
											27	VAL008.B	neg	
	28	500–700	min 21	min 13	min 3	10339	VAL009	?	?	Collective burial next to apse	27	VAL009.A	neg	

Site	Group	Dating	Total	Adult	Sub-adult	Arch. ID	Lab ID ind.	Sex	Age	Features	Teeth	Lab ID sample	Scr	Seq
Waging	**199**	**500–550**	**1**	**1**	**0**	**199**	**WAG001**	**m**	**ii**	**Superposition (under 200/201)**	**55**	**WAG001.A**	**pos**	**pos**
											M	WAG001.B	neg	
	200/201	600–700	2	1	1	200	WAG002	m	em		36	WAG002.A	neg	
											M	WAG002.B	neg	
						201	WAG003	?	ii		85	WAG003.A	neg	
											M	WAG003.B	neg	
	37(1–2)	600–700	2	1	1	37-1	WAG004	f?	ii	Superposition over 39, disturbed	46	WAG004.A	neg	
						37-2	WAG005	?	j–ea		62	WAG005.A	neg	
	39(N/S)	600–700	2	1	1	39N	WAG006	f	ii		75	WAG006.A	neg	
											m	WAG006.B	neg	
						39S	WAG007	f	lm		43	WAG007.A	neg	
											P	WAG007.B	neg	
Westheim	1	500–650	3	2	1	1(-1?)	WES001	m	la–em		27	WES001.A	neg	
						1(-2?)	WES002	f?	(a)		18	WES002.A	neg	
						1(-3?)	WES003	f	lm		26	WES003.A	neg	

Table S2: Mapping results of all UDG treated and captured samples for the chromosome and all three plasmids, giving the number of unique reads after clipping, merging, quality filtering and duplicate removal, and the mean coverage. In the three columns left of the last (showing the Mahalanobis distance), the ratio of mapping reads to the respective plasmid/chromosome is calculated for each sample.

Sample	Chrom. reads	Chrom. coverage	pCD1 reads	pCD1 coverage	pMT1 reads	pMT1 coverage	pPCP1 reads	pPCP1 coverage	Ratio reads pCD1/chrom.	Ratio reads pMT1/chrom.	Ratio reads pPCP1/chrom.	Mahalanobis distance
DIR002.A	8470	0.08	10	0.01	9	0.01	3	0.01	0.00118	0.00106	0.00035	8.80
DIT003.B	877150	9.41	31613	23.44	26800	14.12	7777	49.80	0.03604	0.03055	0.00887	0.61
DIT004.A	192557	1.65	7638	4.40	6383	2.74	2043	9.06	0.03967	0.03315	0.01061	0.77
DIT004.B	21824	0.22	849	0.58	935	0.46	220	1.12	0.03890	0.04284	0.01008	1.21
EDI001.A	2720235	38.05	66158	69.86	60478	43.68	10589	95.51	0.02432	0.02223	0.00389	1.36
EDI002.A	406	0.00	0	0	0	0	1	0.00	0.00000	0.00000	0.00246	9.36
EDI003.A	482945	5.17	21554	15.80	14738	7.80	6270	37.61	0.04463	0.03052	0.01298	2.95
EDI004.A	656840	7.50	21818	17.42	18463	10.55	5507	36.90	0.03322	0.02811	0.00838	0.57
EDI005.A	1706	0.02	19	0.02	24	0.01	14	0.07	0.01114	0.01407	0.00821	3.80
LSD001.A	426234	4.77	23345	17.55	22886	12.41	4539	25.77	0.05477	0.05369	0.01065	5.28
LSD002.A	3736	0.04	99	0.07	146	0.07	23	0.11	0.02650	0.03908	0.00616	2.63
LSD007.A	107157	1.05	4145	2.70	3774	1.80	519	2.65	0.03868	0.03522	0.00484	1.83
LSD013.A	2190	0.02	57	0.04	71	0.03	22	0.11	0.02603	0.03242	0.01005	0.85
LSD019.A	380773	3.93	15515	10.87	14445	7.44	1511	8.29	0.04075	0.03794	0.00397	2.65
LSD020.A	172504	1.76	6450	4.36	7697	3.79	1229	6.13	0.03739	0.04462	0.00712	2.25
LSD021.A	209598	2.08	6261	4.17	7690	3.70	1235	6.29	0.02987	0.03669	0.00589	1.05
LSD022.A	8536	0.10	303	0.24	352	0.20	42	0.25	0.03550	0.04124	0.00492	1.92
LSD023.A	591040	7.15	19573	16.25	17618	10.69	2320	15.28	0.03312	0.02981	0.00393	1.45
LSD024.A	207496	3.52	7371	8.22	6124	5.07	1771	15.05	0.03552	0.02951	0.00854	0.74
LSD026.A	1896	0.02	46	0.03	61	0.03	13	0.06	0.02426	0.03217	0.00686	1.03
LVC001.B	40317	0.41	1731	1.19	1731	0.86	1176	6.33	0.04293	0.04293	0.02917	9.14
LVC001.C	182392	1.78	8385	5.54	6156	2.96	5694	28.43	0.04597	0.03375	0.03122	10.26

Sample	Chrom. reads	Chrom. coverage	pCD1 reads	pCD1 coverage	pMT1 reads	pMT1 coverage	pPCP1 reads	pPCP1 coverage	Ratio reads pCD1/chrom.	Ratio reads pMT1/chrom.	Ratio reads pPCP1/chrom.	Mahalanobis distance
LVC005.A	162315	1.81	7300	5.56	5327	2.94	2022	12.14	0.04497	0.03282	0.01246	2.36
LVC005.B	3628	0.04	114	0.08	132	0.07	63	0.30	0.03142	0.03638	0.01736	2.48
LVC005.C	276945	2.68	8392	5.43	9594	4.55	6033	31.11	0.03030	0.03464	0.02178	5.19
LVC006.A	41524	0.43	1178	0.81	1673	0.85	277	1.44	0.02837	0.04029	0.00667	2.50
PEI001.A	1133	0.01	6	0.00	36	0.03	0	0.00	0.00530	0.03177	0.00000	11.07
PET004.A	515024	5.62	19971	14.86	17721	9.54	4442	26.78	0.03878	0.03441	0.00862	0.73
PET005.A	55916	0.49	2213	1.31	2395	1.04	589	2.62	0.03958	0.04283	0.01053	1.15
PET006.A	19144	0.20	645	0.46	746	0.38	209	1.09	0.03369	0.03897	0.01092	0.80
UNT002.A	14842	0.14	648	0.43	546	0.28	311	1.52	0.04366	0.03679	0.02095	2.62
UNT003.A	753340	7.57	21453	14.59	18224	9.13	5836	32.65	0.02848	0.02419	0.00775	0.72
UNT004.A	533323	5.19	20168	13.53	13211	6.45	4671	24.91	0.03782	0.02477	0.00876	3.02
UNT005.A	13669	0.15	578	0.40	448	0.24	171	0.90	0.04229	0.03277	0.01251	1.39
UNT006.A	7322	0.07	177	0.12	159	0.08	80	0.42	0.02417	0.02172	0.01093	1.13
VAL001.B	996679	9.62	27992	18.30	30071	14.29	5520	29.05	0.02809	0.03017	0.00554	0.42
WAG001.A	22904	0.23	733	0.50	539	0.26	351	1.93	0.03200	0.02353	0.01532	1.93

Table S3: Sequencing results for all verified *Y. pestis*-positive samples. The data of the Altenerding sample AE1175 was reprocessed without additional capture or sequencing. Data was merged in case multiple samples of the same individual were positive (DIT004, LVC001, LVC005) or when samples of multiple individuals of the same burial context were merged (LVC).

Sample	Reads prior mapping after adapter clipping and merging	Unique reads mapping to *Y. pestis* reference	Endogenous DNA (%)	Cluster Factor	Mean Coverage	Coverage 1X (%)	Median frag. Length
AE1175	67364973	1164051	3.56	2.06	17.17	93.62	64
DIT003.B	12699975	877150	22.88	3.31	9.41	91.41	48
DIT004.A	65225088	192557	1.02	3.44	1.65	69.02	38
DIT004.B	4361688	21824	1.80	3.59	0.22	17.41	45
DIT004 merged	70153762	215192	1.06	3.46	1.88	72.67	39
EDI001.A	13026793	2720235	25.15	1.20	38.05	95.24	60
EDI003.A	72165434	482945	1.00	1.49	5.17	91.62	47
EDI004.A	61443239	656840	1.47	1.38	7.50	93.92	47
EDI005.A	13587246	1706	0.02	1.77	0.02	0.52	38
LSD001.A	12697045	426234	3.80	1.13	4.77	88.41	49
LSD002.A	8163219	3736	0.05	1.12	0.04	3.49	44
LSD007.A	20960696	107157	1.79	3.50	1.05	49.74	43
LSD013.A	8836976	2190	0.03	1.21	0.02	1.8	42
LSD019.A	19755616	380773	5.15	2.67	3.93	85.09	45
LSD020.A	10566480	172504	1.95	1.20	1.76	70.87	45
LSD021.A	9622435	209598	2.82	1.29	2.08	72.64	43
LSD022.A	9748447	8536	0.10	1.15	0.10	9.07	50
LSD023.A	20406195	591040	6.78	2.34	7.15	90.37	51
LSD024.A	22016115	207496	5.47	5.80	3.52	85.63	76
LSD026.A	8778616	1896	0.03	1.18	0.02	1.52	43
LVC001.B	3593715	40317	3.65	3.25	0.41	28.23	45
LVC001.C	16970946	182392	5.76	3.29	1.78	72.48	43
LVC001 merged	20564661	220514	3.56	3.32	2.17	76.42	43
LVC005.A	3299048	162315	14.12	2.87	1.81	70.13	50
LVC005.B	3634255	3628	0.27	2.75	0.04	3.09	44
LVC005.C	20373965	276945	7.44	3.61	2.68	80.19	43
LVC005 merged	27307268	430854	5.40	3.43	4.42	86.67	45
LVC006.A	4480909	41524	3.29	3.55	0.43	29.66	46
LVC merged	52352838	658940	4.50	3.57	6.71	89.89	45
VAL001.B	65109736	996679	4.28	2.80	9.62	91.35	43

Sample	Reads prior mapping after adapter clipping and merging	Unique reads mapping to *Y. pestis* reference	Endogenous DNA (%)	Cluster Factor	Mean Coverage	Coverage 1X (%)	Median frag. Length
PET004.A	52013500	515024	2.69	2.72	5.62	92.30	48
PET005.A	89496419	55916	0.34	5.51	0.49	34.62	39
PET006.A	48802311	19144	0.12	3.02	0.20	16.10	44
UNT002.A	88974245	14842	0.20	12.15	0.14	10.97	41
UNT003.A	37303508	753340	6.11	3.03	7.57	92.65	44
UNT004.A	26118776	533323	7.07	3.46	5.19	91.22	42
UNT005.A	3613234	13669	0.49	1.29	0.15	13.24	45
UNT006.A	91966889	7322	0.08	9.64	0.07	4.98	40
WAG001.A	74283932	22904	0.13	4.31	0.23	13.89	43

Table S4: Number of reads of EDI samples assigned to the *Y. pseudotuberculosis* complex/*Y. pestis* node in MALT with 85 % identity.

Samples	Total # of reads	*Y. pseudotuberculosis* complex node		*Y. pestis* node summed
		assigned	summed	
EDI001.A	14183537	26478	36623	9886
EDI002.A	10714470	44	58	14
EDI003.A	41364951	3485	4738	1204
EDI004.A	5792548	743	991	230
EDI005.A	9029341	21	33	4
EDI006.A	7838268	6	10	1
EDI007.A	44427907	21	27	1
EDI008.A	4544207	79	110	29
EDI009.A	9918621	12	25	2
EDI010.A	4767075	0	0	0
EDI011.A	8639291	4	7	0
EDI012.A	38492923	145	259	29
EDI013.A	17140889	7	8	1
EDI014.A	8189593	0	2	0
EDI015.A	10412796	5	10	1
EDI016.A	27869518	44	69	6
EDI017.A	9565231	10	16	0
EDI018.A	2040312	3	4	1
EDI019.A	10904548	2	6	2
EDI020.A	29162638	29	45	2
EDI021.A	10637817	4	9	1
EDI022.A	9585693	23	31	4

Table S5: Basic statistics for the background (DIR002.A) and the simulated datasets (CO92) with targeted mean coverages of 3-fold, 5-fold, 10-fold and 30-fold, as well as for the datasets after merging with the background (cont.=contaminated).

Sample Name	Reads prior mapping after adapter clipping and merging	Unique reads mapping to Y. pestis reference	Endogenous DNA (%)	Cluster Factor	Mean Coverage	Coverage 1X (%)	Median frag. Length
DIR002.A	3728460	8470	0.44	1.94	0.08	1.07	41
CO92 3x A	279387	263585	94.38	1.00	2.83	89.94	48
CO92 3x B	279376	263413	94.32	1.00	2.83	90.05	48
CO92 3x C	279392	263532	94.36	1.00	2.83	89.97	48
Average	279385.00	263510.00	94.35	1.00	2.83	89.99	48
CO92 5x A	465651	438732	94.28	1.00	4.71	94.24	48
CO92 5x B	465655	439019	94.35	1.00	4.72	94.23	48
CO92 5x C	465642	438804	94.30	1.00	4.72	94.26	48
Average	465649.33	438851.67	94.31	1.00	4.72	94.24	48
CO92 10x A	931299	877513	94.35	1.00	9.43	95.06	48
CO92 10x B	931295	877631	94.36	1.00	9.43	95.07	48
CO92 10x C	931290	877187	94.31	1.00	9.43	95.07	48
Average	931294.67	877443.67	94.34	1.00	9.43	95.07	48
CO92 30x A	2793853	2625000	94.32	1.00	28.20	95.24	48
CO92 30x B	2793896	2625351	94.33	1.00	28.21	95.26	48
CO92 30x C	2793887	2625415	94.33	1.00	28.21	95.25	48
Average	2793878.67	2625255.33	94.33	1.00	28.20	95.25	48
CO92 3x A cont.	4007847	267836	6.99	1.05	2.87	90.00	48
CO92 3x B cont.	4007836	267830	6.99	1.05	2.87	90.12	48
CO92 3x C cont.	4007852	267963	6.99	1.05	2.88	90.05	48
Average	4007845.00	267876.33	6.99	1.05	2.87	90.06	48
CO92 5x A cont.	4194111	436364	10.86	1.04	4.69	94.26	48
CO92 5x B cont.	4194115	436628	10.87	1.04	4.70	94.25	48
CO92 5x C cont.	4194102	436485	10.86	1.04	4.70	94.28	48
Average	4194109.33	436492.33	10.86	1.04	4.70	94.26	48
CO92 10x A cont.	4659759	843943	19.21	1.06	9.12	95.07	48
CO92 10x B cont.	4659755	844112	19.21	1.06	9.12	95.08	48
CO92 10x C cont.	4659750	843602	19.20	1.06	9.12	95.07	48
Average	4659754.67	843885.67	19.21	1.06	9.12	95.07	48
CO92 30x A cont.	6522313	2283270	40.66	1.16	25.01	95.25	49
CO92 30x B cont.	6522356	2283502	40.66	1.16	25.02	95.27	49
CO92 30x C cont.	6522347	2283869	40.66	1.16	25.02	95.25	49
Average	6522338.67	2283547.00	40.66	1.16	25.02	95.26	49

Table S6: Numbers of false positive SNPs introduced by the background (DIR002.A) in the simulated datasets (see also SI Appendix, Fig. S2). "Filtered out" gives the number of false positive SNPs that do not pass our SNP evaluation criteria. In this table, "Ns" are equivalent to heterozygous positions.

	Background	CO92 3x A cont.	CO92 3x B cont.	CO92 3x C cont.	Average	CO92 5x A cont.	CO92 5x B cont.	CO92 5x C cont.	Average	CO92 10x A cont.	CO92 10x B cont.	CO92 10x C cont.	Average	CO92 30x A cont.	CO92 30x B cont.	CO92 30x C cont.	Average
N	1	319	313	305	312.33	413	410	412	411.67	444	448	447	446.33	395	384	386	388.33
Reference call	0	0	0	0	0.00	0	0	0	0.00	2	0	0	0.67	57	68	66	63.67
filtered out reference call	0	0	0	0	0.00	0	0	0	0.00	2	0	0	0.67	56	66	64	62.00
SNP call	451	133	139	147	139.67	39	42	40	40.33	6	4	5	5.00	0	0	0	0.00
filtered out SNP call	451	133	139	147	139.67	39	42	40	40.33	6	4	5	5.00	0	0	0	0.00

Table S7: Numbers of 418 core positions for the backbone phylogeny of modern *Y. pestis* in the simulated datasets with ("Contaminated") and without ("Clean") background (DIR002.A); see also SI Appendix, Fig. S3. Positions were either called as "reference call" or "Ns" representing positions not called due to low coverage or heterozygosity. "Positions filtered out" give the number of positions that did not pass our SNP evaluation. The positions "lost by heterozygosity" give the difference of reference calls (contaminated) to reference calls (clean). The positions "lost by contamination" give the difference of remaining calls after filtering (contaminated) to remaining calls after filtering (clean).

		DIR002.A	CO92 3x A	CO92 3x B	CO92 3x C	Average	CO92 5x A	CO92 5x B	CO92 5x C	Average	CO92 10x A	CO92 10x B	CO92 10x C	Average	CO92 30x A	CO92 30x B	CO92 30x C	Average
Clean																		
N		-	334	340	342	338.67	194	182	187	187.67	15	19	10	14.67	0	0	0	0.00
Reference call		-	84	78	76	79.33	224	236	231	230.33	403	399	408	403.33	418	418	418	418.00
filtered out		-	0	1	2	1.00	1	1	4	2.00	0	0	0	0.00	0	0	0	0.00
remaining		-	84	77	74	78.33	223	235	227	228.33	403	399	408	403.33	418	418	418	418.00
lost by filtering		-	0.00 %	1.28 %	2.63 %	1.26 %	0.45 %	0.42 %	1.73 %	0.87 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %
Contaminated																		
N		416	338	345	348	343.67	195	187	189	190.33	17	20	13	16.67	0	0	0	0.00
Reference call		2	80	73	70	74.33	223	231	229	227.67	401	398	405	401.33	418	418	418	418.00
filtered out		2	5	3	6	4.67	11	9	13	11.00	14	17	18	16.33	15	17	17	16.33
remaining		0	75	70	64	69.67	212	222	216	216.67	387	381	387	385.00	403	401	401	401.67
lost by filtering		100.00 %	6.25 %	4.11 %	8.57 %	6.28 %	4.93 %	3.90 %	5.68 %	4.83 %	3.49 %	4.27 %	4.44 %	4.07 %	3.59 %	4.07 %	4.07 %	3.91 %
lost by heterozygosity	#	-	4	5	6	5.00	1	5	2	2.67	2	1	3	2.00	0	0	0	0.00
	%	-	4.76 %	6.41 %	7.89 %	6.30 %	0.45 %	2.12 %	0.87 %	1.16 %	0.50 %	0.25 %	0.74 %	0.50 %	0.00 %	0.00 %	0.00 %	0.00 %
lost by contamination	#	-	9	7	10	8.67	11	13	11	11.67	16	18	21	18.33	15	17	17	16.33
	%	-	10.71 %	9.09 %	13.51 %	11.06 %	4.93 %	5.53 %	4.85 %	5.11 %	3.97 %	4.51 %	5.15 %	4.55 %	3.59 %	4.07 %	4.07 %	3.91 %

Table S8: Table with all non-shared chromosomal SNPs that were called in the investigated First Pandemic genomes, sorted by genome (second cell in header). 'Position' refers to the position in the reference genome CO92, followed by the respective SNP call and the reference call. The classification as potential false or true positive is based on the preceding parameters regarding a 50 bp window surrounding the SNP: Only SNPs with no uncovered position, no heterozygous SNP and an equal mean coverage with high and low stringent mapping (ratio of 1.00) are classified as true positive SNPs (highlighted in green). If a SNP is ambiguous due to conflicting classification in different samples, this is indicated with a question mark highlighted in orange. The 'comment' column with additional classification is followed by the respective calls in the other First Pandemic genomes (0=not covered, N=heterozygous, lower case letters=less than 3 reads coverage). The last column is giving the corresponding genomes, if the SNP appears as homoplastic in the SNP table.

Position	**AE1175 call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
20769	G	A	0	1	1.00	false positive		0	N	a	0	0	0	0	A	0	a	0	0	2.MED2e_7338
420208	T	G	20	0	1.00	false positive	Feldman et al. 2016: Called as SNP but not called as unique due to homoplasy	0	t	t	t	0	0	0	0	0	0	0	0	0.PE2_6904
549767	C	T	0	0	1.45	false positive	Feldman et al. 2016	C	T	T	T	c	c	c	c	C	C	c	0	
2348782	T	C	5	0	1.00	false positive		0	t	0	t	t	0	0	0	t	0	t	0	
2828136	G	A	0	3	1.39	false positive		N	N	a	0	a	0	0	0	a	a	g	0	
2980096	T	G	0	1	1.38	false positive		t	N	g	g	t	N	g	0	0	0	g	0	2.ANT2b_351001
2980097	G	A	0	1	1.37	false positive		g	N	a	a	g	N	a	0	0	0	a	0	2.ANT2b_351001
3179828	A	C	0	0	1.00	true positive	shared AE, DIT, UNT	A	C	C	C	c	C	C	C	A	A	C	0	
3223354	T	G	0	1	1.13	false positive		0	N	0	N	0	0	0	0	0	g	0	0	1.IN2i_CMCC10012
3223359	A	C	0	1	1.13	false positive		0	N	0	N	0	0	0	a	0	c	0	0	1.IN2i_CMCC10012, 1.IN1a_CMCC11001
3225856	A	C	23	0	1.24	false positive		a	N	a	0	0	0	0	0	c	a	c	0	
3588894	C	A	0	1	1.17	false positive		a	A	0	0	0	0	a	0	0	0	a	0	
3750736	A	G	0	0	0.99	true positive?	false positive in VAL001.B, LSD001.A, shared AE, DIT, LSD, LVC, PET, UNT, VAL	a	G	G	G	A	A	a	A	A	A	A	a	

Position	AE1175 call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
3755861	T	C	0	0	1.00	true positive	Feldman et al. 2016, potentially shared ancestral	0	T	t	0	0	t	t	t	t	t	0	0	
4412624	G	A	0	0	1.00	true positive?	Feldman et al. 2016, false positive in UNT003.A, true positive in PET, potentially shared DIT, LSD, PET, UNT, VAL	0	A	0	a	g	g	0	G	G	g	g	0	
Position	**DIT003.B call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
335336	C	T	0	1	1.00	false positive		N	N	N	N	N	N	0	0	N	N	c	0	0.PE4Cd_CMCC93014 2.ANT1_Nepal516, 0.PE4_Microtus91001, 0.PE5_6213, 0.PE4_6216, *Y. pseudotuberculosis*
549767	C	T	0	0	1.17	false positive		C	T	T	T	N	c	c	c	C	C	c	0	
944177	G	C	6	0	1.00	false positive		N	N	g	N	c	c	c	c	N	N	N	0	
944178	A	G	6	0	1.00	false positive		N	N	a	g	g	g	g	g	N	g	N	0	
3179828	A	C	0	0	1.00	true positive	shared AE, DIT, UNT	A	C	C	C	c	C	C	C	A	A	C	0	
3225949	G	A	21	0	1.12	false positive		g	G	g	g	0	g	a	0	0	0	g	0	
3890928	G	C	4	0	1.03	false positive		N	N	c	c	g	G	N	N	g	N	G	0	
4232217	T	C	4	0	1.00	false positive		0	T	t	t	t	T	c	0	t	0	0	0	0.PE2_1522

Position	EDI001.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
698935	G	T	0	2	1.09	false positive		t	t	0	0	0	0	t	0	t	0	t	0	
718827	C	T	14	0	1.29	false positive		c	c	0	c	0	c	c	c	c	c	C	0	Branch 2
1105569	T	A	0	2	1.23	false positive		N	a	a	0	0	0	a	0	0	a	a	0	
1105570	C	T	0	2	1.24	false positive		N	g	0	0	0	0	0	0	0	0	0	0	
1444672	G	A	0	1	1.13	false positive?	true positive in EDI004.A, potentially shared ancestral	g	0	G	G	g	G	g	N	g	g	0	0	
1599102	A	T	2	0	1.14	false positive		a	0	t	0	0	0	0	t	0	0	a	0	
1722572	A	G	1	2	1.37	false positive		a	0	N	a	a	0	0	0	a	a	0	0	
2576207	G	A	0	1	1.29	false positive		a	N	0	g	0	N	g	g	g	g	g	0	
2801707	A	G	0	0	1.05	false positive?	true positive in EDI003.A	G	G	A	a	G	G	G	G	G	g	G	0	
3005046	T	A	22	0	1.00	false positive		t	0	0	0	0	t	0	0	t	t	t	0	
3225949	G	A	21	0	1.10	false positive		g	G	g	g	0	g	a	0	0	0	g	0	
3744976	T	C	18	0	1.28	false positive		t	0	0	0	0	0	0	0	0	0	0	0	
3745057	G	A	23	0	1.09	false positive		0	0	0	0	0	0	0	0	0	0	0	0	
4232217	T	C	0	2	1.08	false positive		0	T	t	t	t	T	c	0	t	0	0	0	
4296702	T	G	0	1	1.02	false positive		G	0	0	0	0	0	0	0	0	0	0	0	

Position	EDI003.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
114434	G	A	0	1	2.17	false positive		A	A	A	A	A	a	A	a	A	A	a	A	0	
224431	T	C	0	1	1.37	false positive		C	C	N	C	C	C	T	N	C	C	C	C	T	
233897	C	T	0	1	1.34	false positive		T	T	N	T	T	t	C	T	T	T	T	t	C	
1444672	G	A	0	0	1.20	false positive?	true positive in EDI004.A, potentially shared ancestral	g	0	G	G	G	g	G	g	N	g	g	0	0	
1820043	C	T	0	0	1.25	false positive		T	T	T	T	T	T	T	t	T	T	T	T	N	
2801707	A	G	0	0	1.00	true positive?	false positive in EDI001.A	G	G	A	N	G	G	G	G	G	G	g	G	0	
3155163	C	A	0	1	4.05	false positive		A	N	A	N	0	A	N	N	N	N	A	N	C	
3849574	A	C	0	1	1.77	false positive		C	C	C	C	c	C	C	C	C	C	C	C	C	
3896289	C	A	0	1	1.90	false positive		A	A	A	A	a	N	a	N	A	A	a	a	A	
4639469	G	A	0	1	2.62	false positive		A	A	A	A	N	A	N	A	A	A	A	A	0	

Position	EDI004.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
567757	A	C	0	0	1.00	true positive	potentially shared ancestral	a	0	a	0	a	0	0	a	0	0	N	0	
1444672	G	A	0	0	1.00	true positive?	false positive in EDI003.A and EDI004.A, potentially shared ancestral	g	0	G	G	g	G	g	N	g	g	0	0	
3262779	A	C	25	0	1.01	false positive		C	0	c	c	0	a	0	0	a	0	a	0	
3480906	G	A	0	1	1.68	false positive		A	A	A	A	N	A	A	A	A	A	A	A	0	

Position	**LSD001.A call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
187146	A	G	0	1	1.00	false positive		G	G	G	G	G	G	G	G	G	g	G	0	
463390	T	C	0	0	1.00	true positive	unique LSD001.A	C	N	C	c	C	C	C	C	C	C	C	0	
481144	T	C	0	0	1.00	true positive	shared LSD023.A, LVC	C	C	C	c	C	T	T	C	C	C	C	0	
867563	A	G	0	0	1.00	true positive	unique LSD001.A	G	G	G	g	G	G	G	G	G	G	G	0	
1490914	A	C	0	0	1.00	true positive	shared LSD023.A, LVC	C	C	C	c	N	A	A	C	C	C	C	0	
1806544	C	A	0	0	1.00	true positive	shared LSD023.A	A	A	A	A	A	c	A	A	A	A	A	0	
1957012	T	G	0	0	1.00	true positive	shared LSD023.A, LVC	G	G	G	G	G	T	t	G	G	G	G	0	
2103712	G	T	0	0	1.00	true positive	unique LSD001.A	T	T	T	T	t	T	T	T	T	T	T	0	
2272078	T	G	0	0	1.00	true positive	unique LSD001.A	G	g	G	G	G	G	G	G	G	G	G	0	
2715042	T	C	0	0	1.00	true positive	shared LSD023.A	C	C	C	C	C	T	C	C	C	C	C	0	
2737457	T	C	0	0	1.00	true positive	shared LSD023.A, LVC	C	C	C	C	C	T	T	C	C	C	C	0	
2924618	A	G	0	0	1.00	true positive	shared LSD023.A, LVC	G	G	G	G	G	A	A	G	G	G	G	0	
3017615	G	A	0	0	2.43	false positive		A	N	A	N	N	A	A	N	N	N	N	N	
3111587	T	C	0	0	1.00	true positive	unique LSD001.A	C	C	C	C	C	C	c	C	C	C	C	0	
3228793	T	A	0	0	1.00	true positive	unique LSD001.A	A	A	A	A	A	A	A	A	A	A	A	a	
3295569	T	G	0	0	1.00	true positive	unique LSD001.A	G	G	G	G	G	G	G	G	G	g	G	0	
3665838	C	T	0	0	1.00	true positive	shared LSD023.A, LVC	T	T	T	T	T	C	C	T	T	T	T	0	
3750736	A	G	2	0	1.09	true positive?	false positive in VAL, shared AE, DIT, LSD, LVC, PET, UNT, VAL	A	a	G	G	G	A	a	A	A	A	A	a	
3996401	A	C	0	0	1.00	true positive	unique LSD001.A	C	C	C	C	C	C	C	C	C	C	C	0	
4066202	A	G	0	0	1.00	true positive	unique LSD001.A	G	G	G	G	G	G	G	G	G	G	G	0	
4416487	T	G	0	0	1.00	true positive	shared LSD023.A, LVC	G	G	G	G	G	T	T	G	G	G	G	0	
4563794	A	G	0	0	1.00	true positive?	false positive in LVC, shared LSD; LVC	G	G	G	G	G	A	A	G	G	G	G	0	
4633424	A	G	0	0	1.00	true positive	shared LSD023.A, LVC	G	G	G	G	G	A	A	G	G	G	G	0	

Position	**LSD023.A call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
82545	G	T	0	0	1.00	true positive	unique LSD023.A	T	t	T	t	t	0	t	T	t	t	N	t	
216655	T	C	0	0	2.38	false positive		C	c	C	C	C	C	C	C	C	C	C	T	
216673	C	T	0	0	1.73	false positive		T	T	T	T	T	T	T	T	T	T	T	C	
218890	C	T	0	1	2.93	false positive		T	T	T	T	T	T	T	T	T	T	T	0	
221608	A	G	0	2	1.28	false positive		G	G	G	G	G	G	G	G	G	G	G	A	
221638	G	A	0	0	1.21	false positive		A	A	A	N	A	A	A	A	A	A	A	G	
221690	T	G	0	0	1.44	false positive		G	G	N	N	G	G	G	N	G	G	G	T	
221705	A	G	0	0	1.66	false positive		G	G	G	G	G	G	G	G	G	G	G	N	
222442	G	A	0	1	1.54	false positive		A	A	A	N	A	A	A	A	A	A	A	G	
222625	T	G	0	0	1.51	false positive		G	G	G	N	G	G	G	N	G	G	G	N	
222670	A	C	0	0	1.41	false positive		C	C	N	N	C	N	N	N	C	C	C	A	
222672	A	G	0	0	1.40	false positive		G	G	G	N	G	G	G	N	G	G	G	N	
223252	C	T	0	1	1.47	false positive		T	T	N	T	T	T	N	N	T	T	T	C	
224431	T	C	0	0	1.36	false positive		C	C	N	T	C	C	N	C	C	C	C	T	
224434	T	C	0	0	1.34	false positive		C	C	C	N	C	C	N	C	C	C	C	N	
225009	A	T	0	1	1.15	false positive		T	T	T	T	T	T	T	t	N	t	T	N	
225080	C	A	0	2	1.29	false positive		A	A	A	N	A	A	N	A	A	A	A	C	
225083	G	A	0	2	1.33	false positive		A	N	A	N	A	A	A	A	A	A	A	N	
225458	A	T	0	0	1.95	false positive		T	T	T	N	T	T	T	t	T	N	T	N	
225479	G	T	0	0	2.17	false positive		T	T	T	T	T	T	T	t	N	t	T	G	
225752	G	A	0	0	1.25	false positive		A	A	A	N	A	A	N	A	A	A	A	G	
226189	C	T	0	0	1.37	false positive		T	T	T	T	T	T	T	T	T	T	T	C	
226794	A	C	0	0	1.05	false positive		C	C	N	N	C	C	C	N	C	C	C	A	
227010	T	G	0	1	1.74	false positive		G	G	G	G	G	G	G	N	G	G	G	T	

Position	LSD023.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
227176	T	C	0	1	1.20	false positive		C	C	C	C	C	C	N	C	C	C	C	T	
227221	T	C	0	0	1.10	false positive		C	C	C	C	C	C	C	C	C	C	C	T	
227437	C	T	0	1	2.17	false positive		T	T	T	T	T	T	T	T	T	T	T	C	
227757	A	G	0	0	1.28	false positive		G	G	G	N	G	G	G	G	G	G	G	N	
228727	C	T	0	1	2.03	false positive		T	T	T	T	N	T	T	T	T	T	T	N	
228733	T	C	0	1	1.86	false positive		C	C	C	N	N	C	C	C	C	C	C	T	
228773	C	A	0	0	1.14	false positive		A	A	A	N	A	A	A	A	A	A	A	N	
228816	G	T	0	0	1.01	false positive		T	T	N	N	T	T	T	N	T	T	T	G	
229302	T	C	0	0	1.65	false positive		C	C	C	C	C	c	C	C	C	c	c	N	
229306	A	G	0	0	1.59	false positive		G	G	G	G	G	G	G	G	G	G	G	N	
229632	T	C	0	2	1.70	false positive		C	C	C	C	C	C	C	C	C	C	C	C	
229719	G	T	0	0	1.27	false positive		T	T	N	N	T	T	T	T	T	T	T	G	
229730	G	A	0	0	1.25	false positive		A	A	N	N	A	A	A	N	A	A	A	G	
229874	G	A	0	2	2.33	false positive		A	A	A	N	A	A	A	A	A	A	A	G	
231527	A	G	0	1	1.29	false positive		G	G	G	g	G	G	G	G	G	G	G	a	
232124	A	T	0	0	1.13	false positive		T	T	T	N	T	T	T	T	T	T	T	A	
232802	C	T	0	1	3.45	false positive		T	T	T	T	T	t	T	t	T	0	T	T	
233897	C	T	0	0	1.18	false positive		T	T	N	C	T	t	T	T	T	T	t	C	
233971	G	C	0	1	1.45	false positive		C	C	N	N	C	C	C	C	C	C	C	N	
234724	C	T	0	0	2.91	false positive		T	T	T	t	T	t	T	T	T	T	T	C	
234727	A	G	0	0	2.96	false positive		G	G	G	G	G	g	G	G	G	G	G	A	
481144	T	C	0	0	1.00	true positive	shared LSD001.A, LVC	C	C	C	c	C	T	T	C	C	C	C	0	
503367	G	A	0	0	1.39	false positive		A	A	A	A	A	A	A	A	A	A	A	N	
712800	A	C	0	0	1.00	true positive	shared LSD001.A, LVC	C	C	C	C	C	a	A	C	C	C	C	0	

Position	**LSD023.A call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
881946	A	G	0	0	1.00	true positive	unique LSD023.A	G	G	G	G	G	G	G	G	G	G	G	g	
996250	A	G	0	0	1.00	true positive	unique LSD023.A	G	0	G	G	G	G	0	G	G	G	G	0	
1444672	G	A	1	0	1.00	false positive		g	0	G	G	G	g	g	N	g	g	0	0	
1490914	A	C	0	0	1.00	true positive	shared LSD001.A, LVC	C	C	C	c	N	A	A	C	C	C	C	0	
1769827	G	T	0	0	1.00	true positive	shared LSD001.A, LVC	T	T	T	T	T	g	g	T	T	t	T	0	
1806674	A	G	0	0	1.00	true positive	unique LSD023.A	G	G	G	0	G	g	g	G	G	g	G	0	
1824110	T	A	0	0	1.24	false positive		A	A	A	N	A	a	a	A	A	A	A	T	
1824116	T	G	0	0	1.21	false positive		G	G	G	t	G	g	g	g	G	G	G	T	
1957012	T	G	0	0	1.00	true positive	shared LSD001.A, LVC	G	G	G	G	G	T	t	G	G	G	G	0	
2283822	T	C	0	0	1.00	true positive	unique LSD023.A	C	C	C	C	C	C	C	C	C	C	C	0	
2338880	A	T	0	0	1.00	true positive	unique LSD023.A	T	T	T	T	T	t	T	T	T	T	T	0	
2715042	T	C	0	0	1.00	true positive	shared LSD001.A	C	C	C	C	C	T	C	C	C	C	C	0	
2737457	T	C	0	0	1.00	true positive	shared LSD001.A, LVC	C	C	C	C	C	T	T	C	C	C	C	0	
2924618	A	G	0	0	1.00	true positive	shared LSD001.A, LVC	G	G	G	G	G	A	A	G	G	G	G	0	
3024601	T	G	0	0	1.00	true positive	shared LSD001.A, LVC	G	G	G	G	G	t	T	G	G	G	G	0	
3183745	C	T	0	0	1.09	false positive		T	T	T	T	T	T	T	T	T	T	T	0	
3299755	T	C	0	0	1.66	false positive		N	N	0	N	N	0	N	c	N	t	N	0	*Y. pseudotuberculosis*
3299756	C	T	0	0	1.67	false positive		N	N	0	N	N	0	N	t	N	c	N	0	0.PE2 strains, 3.ANT1a_7b, *Y. pseudotuberculosis*
3518210	A	G	0	2	5.80	false positive		G	G	G	g	G	g	g	G	G	N	G	N	*Y. pseudotuberculosis*
3657089	C	A	0	0	1.00	true positive	unique LSD023.A	A	A	A	A	A	A	A	A	A	A	A	0	
3665838	C	T	0	0	1.00	true positive	shared LSD001.A, LVC	T	T	T	T	T	C	C	T	T	T	T	0	
3683423	T	A	0	1	1.28	false positive		A	A	A	A	A	A	A	A	A	A	A	T	

Position	LSD023.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
3750736	A	G	0	0	1.00	true positive?	false positive in VAL, LSD001.A, shared AE, DIT, LSD, LVC, PET, UNT, VAL	A	a	G	G	G	A	a	A	A	A	A	a	
3764538	G	A	0	1	1.80	false positive		A	A	A	A	A	A	N	A	A	A	A	N	
3890928	G	C	0	2	1.24	false positive		N	G	N	c	c	g	N	N	g	N	G	0	
3956046	G	A	0	0	1.55	false positive		N	A	N	N	N	N	N	A	A	N	N	N	
4098067	G	A	0	0	3.19	false positive		A	A	A	N	N	A	N	N	N	N	N	G	
4200456	A	C	0	1	3.70	false positive		C	C	C	N	C	C	N	C	N	N	C	N	
4201407	T	C	0	1	2.66	false positive		C	c	C	C	C	c	c	C	C	c	C	T	
4201449	G	A	0	1	1.63	false positive		A	A	N	N	A	A	A	N	A	A	A	G	
4205567	T	A	0	0	1.24	false positive		A	A	A	A	N	A	N	A	A	A	a	T	
4205600	G	A	0	0	1.18	false positive		A	A	A	A	A	A	A	A	A	A	A	N	
4232217	T	C	17	0	1.00	false positive		0	T	T	t	t	t	c	0	t	0	0	0	2.MED1_1045, 2.MED1_173
4416487	T	G	0	0	1.00	true positive	shared LSD001.A, LVC	G	G	G	G	G	T	T	G	G	G	G	0	
4417092	T	G	0	1	1.00	false positive		N	N	N	N	N	N	T	N	N	N	N	0	
4563794	A	G	0	0	1.00	true positive?	false positive in LVC, shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0	
4578141	A	G	0	2	1.48	false positive		G	G	G	G	G	g	G	G	G	G	G	A	
4633424	A	G	0	0	1.00	true positive	shared LSD001.A, LVC	G	G	G	G	G	A	A	G	G	G	G	0	
4647626	T	C	0	2	1.39	false positive		C	N	C	N	C	C	C	C	C	C	C	N	

Position	LVC_merged call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
218506	A	C	0	1	3.32	false positive		C	C	C	C	C	C	C	C	C	C	C	0	
481144	T	C	0	0	1.00	true positive	shared LSD, LVC	C	C	C	N	C	T	T	C	C	C	C	0	
712800	A	C	0	0	1.00	true positive	shared LSD, LVC	C	C	C	C	C	a	A	C	C	C	C	0	
1490914	A	C	0	0	1.00	true positive	shared LSD, LVC	C	C	C	N	N	A	A	C	C	C	C	0	
2515458	T	C	0	0	1.00	true positive	unique LVC	C	C	C	C	C	C	C	C	C	C	C	0	
2546208	G	A	10	0	1.27	false positive		A	A	A	A	A	0	N	A	A	A	A	0	
2671168	T	C	0	0	1.21	false positive		C	C	C	C	C	t	t	C	0	C	C	0	
2737457	T	C	0	0	1.00	true positive	shared LSD, LVC	C	C	C	C	C	T	T	C	C	C	C	0	
Position	**LVC_merged call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
2924618	A	G	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0	
3024601	T	G	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	t	T	G	G	G	G	0	
3393110	G	A	0	0	3.73	false positive		A	A	A	N	N	A	A	a	N	0	A	g	
3665838	C	T	0	0	1.00	true positive	shared LSD, LVC	T	T	T	T	T	C	C	T	T	T	T	0	
4416487	T	G	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	T	T	G	G	G	G	0	
4417092	T	G	0	1	1.00	false positive		N	N	N	N	N	N	T	N	N	N	N	0	
4563794	A	G	0	0	1.02	false positive?	true positive in LSD001.A, LSD023.A, shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0	
4633424	A	G	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0	

Position	PET004.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE11175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
514923	T	C	0	1	1.66	false positive		C	C	C	C	C	C	C	C	C	C	C	c	
1573356	C	T	0	1	2.90	false positive		T	T	T	T	T	N	N	N	T	T	N	N	
1791420	G	A	0	0	1.00	true positive	unique PET	A	a	A	a	A	a	0	0	A	A	A	0	
3750736	A	G	0	0	1.00	true positive?	false positive in VAL, LSD001.A, shared AE, DIT, LSD, LVC, PET, UNT, VAL	A	a	G	G	G	A	A	a	A	A	A	A	
3848782	A	T	0	0	3.27	false positive		N	T	T	T	T	T	T	N	T	N	N	T	
3921330	T	C	0	0	1.00	true positive	unique PET	C	C	C	C	C	C	C	C	C	C	C	0	*Y. pseudotuberculosis*
4198065	G	A	0	0	1.53	false positive		A	A	A	A	A	A	A	A	A	A	A	0	
4198659	C	A	0	2	2.60	false positive		A	N	A	A	N	N	A	A	A	A	A	N	
4210722	T	A	0	1	2.95	false positive		A	A	A	A	A	A	N	N	A	A	A	0	

Position	PET004.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE11175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	UNT003.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
4412624	G	A	0	0	1.00	true positive?	false positive in UNT003.A, true positive in AE, potentially shared AE, DIT, LSD, LVC, PET, UNT, VAL	G	0	A	0	a	g	g	0	G	g	g	0	
4644248	T	C	0	0	1.45	false positive		C	C	C	N	C	c	C	C	C	C	C	c	

Position	UNT003.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT004.A	VAL001.B	WAG001.A	Homoplasy
549767	C	T	0	0	1.11	false positive		C	C	T	T	T	c	c	c	c	C	c	0	
1480098	T	G	0	0	1.00	true positive	potentially shared AE, DIT, homoplastic in LSD?	0	t	G	G	G	t	t	g	g	0	0	0	
3179828	A	C	0	0	1.00	true positive	shared AE, DIT, UNT	A	A	C	C	C	c	C	C	C	A	C	0	
3750736	A	G	0	0	1.00	true positive?	false positive in VAL, LSD001.A, shared AE, DIT, LSD, LVC, PET, UNT, VAL	A	a	G	G	G	A	A	a	A	A	A	a	
4412624	G	A	0	0	1.27	true positive?	true positive in AE, PET, potentially shared AE, DIT, LSD, LVC, PET, UNT, VAL	G	0	A	0	a	g	g	0	G	g	g	0	
Position	**UNT004.A call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	VAL001.B	WAG001.A	Homoplasy
225458	G	T	0	1	1.76	false positive		T	T	T	N	T	T	A	T	t	T	t	N	
231350	C	T	24	1	1.58	false positive		T	T	T	T	T	t	T	N	N	T	T	g	
3155190	C	T	0	1	1.91	false positive		T	T	T	T	T	t	T	T	T	T	N	0	
3750736	A	G	0	0	1.00	true positive?	false positive in VAL, LSD001.A, shared AE, DIT, LSD, LVC, PET, UNT, VAL	A	a	G	G	G	A	A	a	A	A	A	a	
3685814	A	G	0	0	1.59	false positive		G	g	G	0	G	0	N	G	g	g	G	G	
3685844	A	G	5	0	1.75	false positive		G	g	G	g	g	g	g	G	G	g	g	G	
4199133	A	C	0	2	2.43	false positive		C	C	C	C	C	C	C	C	C	C	C	N	
4201886	G	A	0	1	1.75	false positive		G	G	G	G	G	G	G	G	G	G	G	G	

Position	VAL001.B call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	WAG001.A	Homoplasy
718827	C	T	21	0	1.32	false positive		c	c	C	0	c	0	c	c	c	c	c	0	Branch 2
1361842	T	C	0	0	1.00	true positive	unique VAL	C	C	C	C	C	C	C	C	C	C	C	0	
2041294	G	T	22	0	1.00	false positive		T	T	T	t	T	t	t	0	T	t	t	0	
2253178	A	C	0	0	1.00	true positive	unique VAL	C	C	C	C	C	C	C	C	C	C	C	0	
2303798	C	T	16	0	1.92	false positive		T	N	T	N	N	N	N	N	N	N	N	N	
2732804	G	C	0	0	1.12	false positive		C	C	C	C	C	C	C	C	C	c	c	0	
2911604	G	A	0	0	1.00	true positive	unique VAL	A	A	A	A	A	A	A	A	A	A	A	0	
3623563	A	G	0	1	1.07	false positive		N	N	N	N	N	N	N	N	0	N	a	0	
3750736	A	G	5	0	1.00	true positive?	false positive in VAL, LSD001.A, shared AE, DIT, LSD, LVC, PET, UNT, VAL	A	a	G	G	G	A	A	a	A	A	A	a	
3890928	G	C	3	0	1.17	false positive		N	G	N	c	c	g	G	N	N	g	A	0	
4256445	G	A	0	0	1.10	false positive		0	A	A	A	A	A	A	0	A	A	N	0	
4635136	A	C	0	1	1.17	false positive		N	C	N	N	N	c	N	N	c	a	G	0	
4648185	A	G	0	0	1.03	false positive		G	G	G	G	G	G	G	G	G	G	c	G	

Table S9: Table with all non-shared plasmidal SNPs that were called in the investigated First Pandemic genomes, sorted by genome (second cell in header). 'Position' refers to the position on the plasmid of the reference genome CO92, followed by the respective SNP call and the reference call. The classification as potential false or true positive is based on the preceding parameters regarding a 50 bp window surrounding the SNP: only SNPs with no uncovered position, no heterozygous SNP and an equal mean coverage with high and low stringent mapping (ratio of 1.00) are classified as true positive SNPs (highlighted in green). If a SNP is ambiguous due to conflicting classification in different samples, this is indicated with a question mark highlighted in orange. The 'comment' column with additional classification is followed by the respective calls in the other First Pandemic genomes (0=not covered, N=heterozygous, lower case letters=less than 3 reads coverage).

Plasmid	**Position**	**AE1175 call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A
pCD1	18753	C	G	0	2	1.12	false positive		c	N	0	0	N	0	0	c	0	0	0	0
pCD1	29959	G	A	0	0	1.02	false positive?	Feldman et al. 2016; false positive in DIT, true positive in LSD, LVC, PET, UNT, VAL	G	A	A	A	G	G	G	G	G	G	G	0
Plasmid	**Position**	**DIT003.B call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A
pCD1	29959	G	A	0	0	1.05	false positive?	false positive in AE, true positive in LSD, LVC, PET, UNT, VAL	G	A	A	A	G	G	G	G	G	G	G	0

Plasmid	Position	LSD001.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A
pCD1	21278	A	G	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	N	a	N	0	g	N	0	a	0	0
pCD1	46912	G	A	6	0	1.43	false positive		N	g	N	a	N	0	g	N	0	a	0	0
pCD1	51106	A	G	0	0	1.00	true positive	unique LSD001.A	G	G	G	G	G	G	G	G	G	G	G	0

Plasmid	Position	LSD023.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A
pCD1	21278	G	A	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	A	A	A	G	G	G	G	G	G	0
pMT1	81057	T	A	0	2	1.00	false positive		N	N	N	N	N	N	N	N	N	N	N	t
pPCP1	8528	C	G	0	0	1.00	false positive*	*see screenshot Fig. S10	0	T	N	t	t	0	N	T	T	t	T	0
pPCP1	8529	G	T	0	0	1.00	false positive*	*see screenshot Fig. S10, false positive in LVC	0	T	N	t	t	g	G	T	T	t	T	0

Plasmid	Position	LVC_merged call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	PET004.A	UNT003.A	UNT004.A	VAL001.B	WAG001.A
pCD1	21278	G	A	0	0	1.00	true positive	shared LSD, LVC	G	G	G	G	G	A	A	G	G	G	G	0
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	A	A	A	G	G	G	G	G	G	0
pPCP1	8529	G	T	0	1	1.00	false positive		0	T	N	t	t	g	G	T	T	t	T	0
Plasmid	**Position**	**PET004.A call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	UNT003.A	UNT004.A	VAL001.B	WAG001.A
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	A	A	G	G	G	G	G	G	G	0
Plasmid	**Position**	**UNT003.A call**	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE1175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT004.A	VAL001.B	WAG001.A
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	A	A	A	G	G	G	G	G	G	0
pMT1	59223	A	T	0	1	1.39	false positive		N	N	T	T	N	N	T	N	N	a	T	0

Plasmid	Position	UNT004.A call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE11175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	VAL001.B	WAG001.A
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	A	N	A	G	G	G	G	G	G	0
pMT1	81058	G	A	0	3	1.14	false positive		N	G	N	N	a	N	N	g	N	N	G	0

Plasmid	Position	VAL001.B call	Reference call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	Classification	Comment	AE11175	DIT003.B	EDI001.A	EDI003.A	EDI004.A	LSD001.A	LSD023.A	LVC_merged	PET004.A	UNT003.A	UNT004.A	WAG001.A
pCD1	29959	G	A	0	0	1.00	true positive	false positive in AE, DIT, true positive in LSD, LVC, PET, UNT, VAL	G	G	A	T	A	T	T	G	T	G	G	0
pMT1	59982	C	A	0	0	1.00	true positive	unique VAL	T	T	T	T	a	T	T	T	T	T	0	0

Table S10: True SNPs identified in the genomes of Edix Hill (EDI), Dittenheim (DIT), Saint-Doulchard (LSD), Lunel-Viel (LVC), Petting (PET), Unterthürheim (UNT) and Valencia (VAL). For all identified unique SNPs including potential false positives, see SI Appendix, Table S5 and S6. *This SNP appears in only one or two reads (below our SNP calling threshold of min. 3 reads).

Chromosome										
Position	Genome	Ref	SNP	Tree	SNP type	Codon change	AA change	Gene ID	Gene name	Function
3750736	AE, DIT, LSD001.A, LSD023.A, LVC*, PET, UNT, VAL, WAG*	G	A	yes	Intergenic	-	-	-	-	-
3179828	AE, DIT, UNT	C	A	yes	Synonymous	tcC/tcA	S/S	YPO2847	*yegM*	Multidrug efflux system subunit MdtA
1444672	AE, EDI, LSD001.A*, LSD023.A, LVC*, UNT*	A	G	no	Intergenic	-	-	-	-	-
3755861	AE, EDI, LSD023.A*, LVC*, PET*, UNT*	C	T	no	Intergenic	-	-	-	-	-
4412624	AE, LSD001.A*, LSD023.A*, PET, UNT, VAL*	A	G	no	Intergenic	-	-	-	-	-
567757	AE*, EDI, LSD001.A*, PET*	C	A	no	Non-synonymous	Ctt/Att	L/I	YPO0523	-	Hypothetical protein
1480098	DIT*, LSD001.A*, LSD023.A*, UNT	G	T	no	Intergenic	-	-	-	-	-
481144	LSD001.A, LSD023.A, LVC	C	T	yes	Synonymous	Ctg/Ttg	L/L	YPO0457	creA	Hypothetical protein
712800	LSD001.A*, LSD023.A, LVC	C	A	yes	Synonymous	cgC/cgA	R/R	YPO0653	glnE	Bifunctional glutamine-synthetase adenylyltransferase/deadenyltransferase
1490914	LSD001.A, LSD023.A, LVC	C	A	yes	Upstream: 63 bases	-	-	YPO1324	-	Undecaprenyl pyrophosphate phosphatase

Chromosome										
Position	Genome	Ref	SNP	Tree	SNP type	Codon change	AA change	Gene ID	Gene name	Function
1769827	LSD001.A, LSD023.A*, LVC*	T	G	yes	Non-synonymous	Acc/Ccc	T/P	YPO1554	-	Pseudogene (sugar transport ATP-binding protein)
1957012	LSD001.A, LSD023.A, LVC*	G	T	yes	Non-synonymous	ttG/ttT	L/F	YPO1713	ogl	Oligogalacturonate lyase
2737457	LSD001.A, LSD023.A, LVC	C	T	yes	Non-synonymous	gCg/gTg	A/V	YPO2437	marC	multiple drug resistance protein MarC
2924618	LSD001.A, LSD023.A, LVC	G	A	yes	Synonymous	gcC/gcT	A/A	YPO2603	rodA	Cell wall shape-determining protein
3024601	LSD001.A*, LSD023.A, LVC	G	T	yes	Non-synonymous	atG/atT	M/I	YPO2696	phrB	3',5'-Cyclic-nucleotide phosphodiesterase
3665838	LSD001.A, LSD023.A, LVC	T	C	yes	Non-synonymous	aAa/aGa	K/R	YPO3285	tyrA	Bifunctional chorismate mutase/prephen-ate dehydrogenase
4416487	LSD001.A, LSD023.A, LVC	G	T	yes	Non-synonymous	atG/atT	M/I	YPO3929	ppc	Phosphoenolpyruvate carboxylase
4563794	LSD001.A, LSD023.A, LVC	G	A	yes	Intergenic	-	-	-	-	-
4633424	LSD001.A, LSD023.A, LVC	G	A	yes	Non-synonymous	aCg/aTg	T/M	YPO4112	-	Hypothetical protein
1806544	LSD001.A, LSD023.A*	A	C	yes	Non-synonymous	ttT/ttG	F/L	YPO1582	lacY	Galactoside permease
2715042	LSD001.A, LSD023.A	C	T	yes	Synonymous	caG/caA	Q/Q	YPO2414	lplA	Lipoate-protein ligase A
2801707	EDI	G	A	yes	Synonymous	acC/acT	T/T	YPO2493	-	Dioxygenase subunit alpha
463390	LSD001.A	C	T	yes	Synonymous	gcC/gcT	A/A	YPO0442	serB	Phosphoserine phosphatase
867563	LSD001.A	G	A	yes	Intergenic	-	-	-	-	-
2103712	LSD001.A	T	G	yes	Non-synonymous	Ttt/Gtt	F/V	YPO1856	-	Hypothetical protein
2272078	LSD001.A	G	T	yes	Non-synonymous	Caa/Aaa	Q/K	YPO1999	-	Decarboxylase
3111587	LSD001.A	C	T	yes	Non-synonymous	Ccc/Tcc	P/S	YPO2777	hisP	histidine/lysine/arginine/ornithine trans-porter subunit
3228793	LSD001.A	A	T	yes	Non-synonymous	ttA/ttT	L/F	YPO2887	yapB	Pseudogene

Chromosome										
Position	Genome	Ref	SNP	Tree	SNP type	Codon change	AA change	Gene ID	Gene name	Function
3295569	LSD001.A	G	T	yes	Upstream: 38 bases	-	-	YPO2949	-	Hypothetical protein
3996401	LSD001.A	C	A	yes	Intergenic	-	-	-	-	-
4066202	LSD001.A	G	A	yes	Downstream: 98 bases	-	-	-	-	-
82545	LSD023.A	T	G	yes	Non-synonymous	Atc/Ct	I/L	YPO0073	cpxA	Two-component sensor kinase
881946	LSD023.A	G	A	yes	Synonymous	agG/agA	R/R	YPO0801		Hypothetical protein
996250	LSD023.A	G	A	yes	Intergenic	-	-	-	-	-
1806674	LSD023.A	G	A	yes	Non-synonymous	gCc/gTc	A/V	YPO1582	lacY	Galactoside permease
2283822	LSD023.A	C	T	yes	Synonymous	ctG/ctA	L/L	YPO2011	pth	Peptidyl-tRNA hydrolase
2338880	LSD023.A	T	A	yes	Non-synonymous	Agt/Tgt	S/C	YPO2060	znuC	High-affinity zinc transporter ATPase
3657089	LSD023.A	A	C	yes	Synonymous	cgT/cgG	R/R	YPO3275	clpB	Protein disaggregation chaperone
2515458	LVC	C	T	yes	Non-synonymous	cCg/cTg	P/L	YPO2238	-	Hypothetical protein
1791420	PET	A	G	yes	Intergenic	-	-	-	-	-
3921330	PET	C	T	yes	Non-synonymous	cGc/cAc	R/H	YPO3510	-	Hypothetical protein
1361842	VAL	C	T	yes	Non-synonymous	gGt/gAt	G/D	YPO1209	tyrP	Tyrosine specific transport protein
2253178	VAL	C	A	yes	Non-synonymous	Cgc/Agc	R/S	YPO1984	-	Hypothetical protein
2911604	VAL	A	G	yes	Non-synonymous	Agt/Ggt	S/G	YPO2588	-	ABC transport protein
Plasmid pCD1										
Position	Genome	Ref	SNP	SNP type		Codon change	AA change	Gene ID	Gene name	Function
29959	AE, DIT, LSD, LVC, PET, UNT, VAL	A	G	Non-synonymous		aAc/aGc	N/S	YPCD1.41	yscO	Putative type III secretion protein
21278	LSD, LVC	G	A	Non-synonymous		gCa/gTa	A/V	YPCD1.29c	yopB	Yersinia outer membrane protein
51106	LSD001.A	G	A	Intergenic		-	-	-	-	-
Plasmid pMT1										
Position	Genome	Ref	SNP	SNP type		Codon change	AA change	Gene ID	Gene name	Function
59982	VAL	A	C	Non-synonymous		aTc/aGc	I/C	YPMT1.59C	-	Putative DNA-binding protein

Table S11: Table with all shared SNPs that were called in the First Pandemic genomes, giving the results of the SNP evaluation for each genome separately. Position is referring to the position in the reference genome CO92, followed by the reference call. For each genome, the call at the SNP position is given (0=not covered, N=heterozygous, lower case letters=less than 3 reads coverage). The classification as potential false or true positive is based on the preceding parameters regarding a 50 bp window surrounding the SNP: Only SNPs with no uncovered position, no heterozygous SNP and an equal mean coverage with high and low stringent mapping (ratio of 1.00) are classified as a candidates for true positive SNPs. SNPs that are supported by less than half of the genomes (<6, last column) are excluded as inconclusive (highlighted in orange). Although the WAG001.A sample was not included in this analysis due to low coverage, all positions were checked

		AE1175				DIT003.B				EDI001.A				EDI003.A			
Position	Reference call	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS
86824	A	G	0	0	1.00	0	43	0	1.00	G	0	0	1.00	g	12	0	1.00
189912	A	G	0	0	1.03	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00
260148	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.02	T	0	0	1.00
271114	C	A	0	0	1.00	A	0	1	1.26	A	0	0	1.01	A	0	0	1.00
485976	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.02	T	9	0	1.00
557841	C	T	0	0	1.00	T	0	0	1.03	T	0	0	1.02	0	11	0	1.00
727741	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	0	0	1.00
779365	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
898980	A	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00	t	0	0	1.00
1067966	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.01	A	0	0	1.00
1211729	A	C	0	0	1.00	C	0	0	1.00	C	0	0	1.01	C	0	0	1.00
1296743	C	T	0	0	1.00	T	0	0	1.08	T	0	0	1.00	T	0	0	1.00
1387701	C	T	0	0	1.00	t	9	0	1.00	T	0	0	1.00	t	0	0	1.00
1387756	A	G	0	0	1.00	0	42	0	1.00	G	0	0	1.00	G	0	0	1.00
1413031	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.01	A	0	1	1.00
1434752	C	A	0	0	1.03	a	0	0	1.00	A	0	0	1.00	A	0	0	1.00
1489055	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
1530658	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	0	0	1.00
1609461	T	C	0	0	1.00	C	0	0	1.10	C	0	0	1.01	0	27	0	1.00
1754708	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.04	t	0	0	1.33
1868678	G	T	0	0	1.00	T	0	0	1.15	T	0	0	1.04	0	7	0	1.00
1956162	T	C	0	0	1.00	0	41	0	1.00	C	0	0	1.00	c	0	0	1.00
2092152	C	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00	t	9	0	1.00
2097520	G	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
2352174	T	G	0	0	1.00	G	0	0	1.00	G	0	0	1.04	G	0	0	1.00

	EDI004.A				LSD001.A				LSD023.A				LVC_merged			
Position	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS
86824	G	0	0	1.00	g	8	0	1.00	g	11	0	1.00	0	48	0	1.00
189912	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00
260148	T	0	0	1.00	0	20	0	1.00	T	0	0	1.00	t	6	0	1.29
271114	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	0	0	1.69
485976	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
557841	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00	T	0	0	1.06
727741	A	0	0	1.00	0	25	0	1.15	a	0	0	1.00	A	0	0	1.00
779365	0	16	0	1.00	t	0	0	1.00	T	0	0	1.00	T	0	0	1.00
898980	t	0	0	1.00	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00
1067966	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00
1211729	C	0	0	1.05	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00
1296743	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.04
1387701	T	0	0	1.00	T	7	0	1.00	T	0	0	1.00	0	31	0	1.00
1387756	0	4	0	1.00	0	50	0	NA	0	42	0	1.00	g	20	0	1.00
1413031	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	2	0	1.00
1434752	A	0	0	1.00	A	0	0	1.19	A	0	0	1.00	A	0	0	1.00
1489055	T	0	0	1.00	T	0	0	1.05	T	0	0	1.00	N	0	0	1.00
1530658	A	0	0	1.00	a	0	0	1.00	a	0	0	1.00	a	11	0	1.00
1609461	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00
1754708	T	0	0	1.00	t	0	0	1.00	t	0	0	1.00	T	0	0	1.00
1868678	T	0	0	1.00	t	1	0	1.00	T	0	0	1.00	T	0	0	1.00
1956162	C	0	0	1.00	c	12	0	1.00	c	23	0	1.00	0	30	0	1.00
2092152	T	0	0	1.00	0	50	0	NA	T	0	0	1.00	0	12	0	1.00
2097520	T	0	0	1.00	T	7	0	1.00	T	0	0	1.00	N	7	1	1.00
2352174	g	0	0	1.00	G	0	0	1.07	g	0	0	1.00	G	0	0	1.00

	PET004.A				UNT003.A				UNT004.A				VAL001.B						
Position	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	SNP Call	Uncovered positions in 50 bp	Heterozygous SNPs in 50 bp	Mean coverage ratio LS/HS	DA101 call	WAG001.A call	No. of genomes passing criteria
86824	G	0	0	1.00	g	0	0	1.00	g	0	0	1.00	0	48	0	1.00	A	0	6
189912	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00	A	0	11
260148	t	0	0	1.00	T	0	0	1.00	T	0	0	1.00	t	0	0	1.00	T	0	9
271114	A	0	0	1.00	A	0	0	1.00	A	0	0	1.23	A	0	0	1.00	C	a	8
485976	t	1	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	C	0	9
557841	t	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	C	0	8
727741	a	0	0	1.00	A	0	0	1.00	a	0	0	1.28	A	0	0	1.00	G	a	10
779365	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	C	0	11
898980	t	0	0	1.00	T	0	0	1.12	t	0	0	1.00	t	0	0	1.00	A	0	11
1067966	A	0	0	1.00	A	0	0	1.00	A	0	0	1.09	A	0	0	1.00	C	0	10
1211729	C	0	0	1.08	C	0	0	1.00	C	0	0	1.00	C	0	0	1.04	A	0	8
1296743	T	0	0	1.00	T	0	0	1.00	T	0	0	1.06	T	0	0	1.09	C	t	8
1387701	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00	T	0	0	1.00	c	0	10
1387756	0	27	0	1.00	g	11	0	1.00	G	0	0	1.00	0	35	0	1.00	A	0	4
1413031	a	5	0	1.00	a	0	0	1.00	A	0	0	1.00	a	0	0	1.00	C	0	8
1434752	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	C	0	10
1489055	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	C	t	11
1530658	a	0	0	1.00	A	0	0	1.00	a	0	0	1.00	a	0	0	1.00	C	0	11
1609461	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00	T	0	9
1754708	T	0	0	1.00	T	0	0	1.00	0	25	0	1.00	t	0	0	1.00	C	0	9
1868678	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	G	0	9
1956162	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00	C	0	0	1.00	T	c	9
2092152	t	15	0	1.00	T	0	0	1.00	0	9	0	1.00	t	0	0	1.00	C	0	7
2097520	T	0	0	1.00	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00	G	t	10
2352174	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00	T	0	10

Position	Reference call	AE1175 SNP Call	AE1175 Uncovered positions in 50 bp	AE1175 Heterozygous SNPs in 50 bp	AE1175 Mean coverage ratio LS/HS	DIT003.B SNP Call	DIT003.B Uncovered positions in 50 bp	DIT003.B Heterozygous SNPs in 50 bp	DIT003.B Mean coverage ratio LS/HS	EDI001.A SNP Call	EDI001.A Uncovered positions in 50 bp	EDI001.A Heterozygous SNPs in 50 bp	EDI001.A Mean coverage ratio LS/HS	EDI003.A SNP Call	EDI003.A Uncovered positions in 50 bp	EDI003.A Heterozygous SNPs in 50 bp	EDI003.A Mean coverage ratio LS/HS
2419529	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	0	0	1.00
2495165	C	A	0	0	1.00	0	50	0	1.00	A	0	0	1.00	0	4	0	1.00
2725715	C	T	0	0	1.00	t	14	0	1.00	T	0	0	1.03	t	0	0	1.00
2753572	A	T	0	0	1.00	T	0	0	1.00	T	0	0	1.02	T	0	0	1.00
2977542	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00
3078807	C	A	0	0	1.05	A	0	0	1.03	A	0	0	1.00	A	0	0	1.00
3274298	A	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	t	25	0	1.00
3360963	A	C	0	0	1.08	C	0	0	1.27	C	0	0	1.09	0	9	0	1.00
3360984	C	T	0	0	1.06	T	0	0	1.28	T	0	0	1.11	T	6	0	1.00
3398153	G	A	0	0	1.00	A	0	0	1.05	A	0	0	1.02	a	0	0	1.00
3409414	T	C	0	0	1.00	c	6	0	1.00	C	0	0	1.00	c	0	0	1.00
3500922	T	G	0	0	1.00	G	0	0	1.00	G	0	0	1.00	G	0	0	1.05
3535148	G	T	0	0	1.00	T	0	0	1.00	T	0	0	1.01	T	0	0	1.00
3560088	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.01	A	0	0	1.00
3568597	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.03	T	0	0	1.08
3755861	C	T	0	0	1.00	0	29	0	1.00	T	0	0	1.01	t	23	0	1.00
3843195	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.02	a	24	0	1.00
3892488	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.02	T	0	0	1.00
4066494	C	T	0	0	1.00	t	17	0	1.37	T	0	0	1.00	T	0	0	1.30
4307755	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	0	0	1.19
4423366	G	A	0	0	1.00	A	0	0	1.04	A	0	0	1.00	A	0	0	1.00
4460688	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
4465967	C	A	0	0	1.04	A	0	0	1.00	A	0	0	1.06	A	0	0	1.00
4628496	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.02	A	0	0	1.00
4629169	G	A	0	0	1.03	A	0	0	1.00	A	0	0	1.02	A	0	0	1.00

Position	Reference call	EDI004.A SNP Call	EDI004.A Uncovered positions in 50 bp	EDI004.A Heterozygous SNPs in 50 bp	EDI004.A Mean coverage ratio LS/HS	LSD001.A SNP Call	LSD001.A Uncovered positions in 50 bp	LSD001.A Heterozygous SNPs in 50 bp	LSD001.A Mean coverage ratio LS/HS	LSD023.A SNP Call	LSD023.A Uncovered positions in 50 bp	LSD023.A Heterozygous SNPs in 50 bp	LSD023.A Mean coverage ratio LS/HS	LVC_merged SNP Call	LVC_merged Uncovered positions in 50 bp	LVC_merged Heterozygous SNPs in 50 bp	LVC_merged Mean coverage ratio LS/HS
2419529	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00
2495165	C	A	0	0	1.00	0	38	0	1.00	a	0	0	1.00	a	12	0	1.00
2725715	C	T	0	0	1.00	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00
2753572	A	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
2977542	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.10	A	0	0	1.07
3078807	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00
3274298	A	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
3360963	A	C	0	0	1.00	C	0	0	1.16	C	0	0	1.08	C	0	0	1.12
3360984	C	T	0	0	1.00	T	0	0	1.17	T	0	0	1.18	T	0	0	1.12
3398153	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	a	0	0	1.00
3409414	T	C	0	0	1.00	0	37	0	1.00	0	31	0	1.00	0	12	0	1.05
3500922	T	G	0	0	1.00	N	9	0	1.00	G	0	0	1.00	G	0	0	1.00
3535148	G	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
3560088	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.16	A	0	0	1.00
3568597	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
3755861	C	0	23	0	1.00	0	49	0	1.00	t	12	0	1.00	t	22	0	1.00
3843195	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00
3892488	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00
4066494	C	T	0	0	1.00	0	38	0	1.00	t	0	0	1.00	T	2	0	1.24
4307755	G	A	0	0	1.00	a	0	0	1.00	A	0	0	1.00	A	0	0	1.00
4423366	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.11
4460688	C	t	0	0	1.00	T	0	0	1.19	T	0	0	1.00	T	0	0	1.00
4465967	C	A	0	0	1.04	A	0	0	1.00	A	0	0	1.03	A	0	0	1.00
4628496	C	A	0	0	1.00	a	0	0	1.00	A	0	0	1.00	A	0	0	1.00
4629169	G	A	0	0	1.00	a	0	0	1.22	A	0	0	1.00	A	0	0	1.16

Position	Reference call	PET004.A SNP Call	PET004.A Uncovered positions in 50 bp	PET004.A Heterozygous SNPs in 50 bp	PET004.A Mean coverage ratio LS/HS	UNT003.A SNP Call	UNT003.A Uncovered positions in 50 bp	UNT003.A Heterozygous SNPs in 50 bp	UNT003.A Mean coverage ratio LS/HS	UNT004.A SNP Call	UNT004.A Uncovered positions in 50 bp	UNT004.A Heterozygous SNPs in 50 bp	UNT004.A Mean coverage ratio LS/HS	VAL001.B SNP Call	VAL001.B Uncovered positions in 50 bp	VAL001.B Heterozygous SNPs in 50 bp	VAL001.B Mean coverage ratio LS/HS	DA101 call	WAG001.A call	No. of genomes passing criteria
2419529	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.10	G	a	11
2495165	C	a	0	0	1.00	a	14	0	1.00	A	0	0	1.00	0	50	0	NA	C	0	6
2725715	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	10
2753572	A	T	0	0	1.07	t	0	0	1.00	T	0	0	1.00	T	0	0	1.00	A	0	10
2977542	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.08	A	0	9
3078807	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	C	0	10
3274298	A	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	t	0	11
3360963	A	C	0	0	1.00	0	16	0	1.56	0	43	0	1.00	C	0	0	1.07	A	0	2
3360984	C	T	0	0	1.00	t	16	0	1.63	t	22	0	1.00	T	0	0	1.17	C	0	2
3398153	G	A	0	0	1.00	A	0	0	1.00	a	1	0	1.00	A	0	0	1.00	G	0	9
3409414	T	0	26	0	1.00	C	0	0	1.00	c	7	0	1.00	0	26	0	1.00	T	0	5
3500922	T	G	0	0	1.00	G	0	0	1.00	g	0	0	1.00	G	0	0	1.00	T	0	10
3535148	G	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	G	0	11
3560088	G	A	0	0	1.00	A	0	0	1.00	A	0	1	1.00	A	0	0	1.00	G	0	9
3568597	C	T	0	0	1.00	T	0	0	1.05	T	0	0	1.00	T	0	0	1.00	C	0	9
3755861	C	t	1	0	1.00	t	15	0	1.00	t	7	1	1.00	0	37	0	1.00	C	0	1
3843195	C	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	C	0	10
3892488	C	T	0	0	1.19	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	C	0	10
4066494	C	0	16	0	1.00	T	0	0	1.00	t	0	0	1.00	T	0	0	1.00	C	0	7
4307755	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	G	0	11
4423366	G	A	0	0	1.07	A	0	0	1.00	A	0	0	1.00	A	0	0	1.00	G	0	9
4460688	C	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	0	1.00	T	0	11
4465967	C	A	0	0	1.00	A	0	0	1.04	A	0	0	1.00	A	0	0	1.00	c	0	7
4628496	C	A	0	0	1.00	A	0	0	1.00	a	0	0	1.00	A	0	0	1.00	C	0	11
4629169	G	A	0	0	1.00	A	0	0	1.00	A	0	0	1.12	A	0	0	1.00	G	0	7

Table S12: List of all modern and ancient *Y. pestis* genomes with accession number, sample origin and the corresponding publication.

Modern strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
0.ANT1a_42013	ADPG00000000	China	Cui et al. 2013
0.ANT1b_CMCC49003	ADQX00000000	China	Cui et al. 2013
0.ANT1c_945	ADPV00000000	China	Cui et al. 2013
0.ANT1d_164	ADOW00000000	China	Cui et al. 2013
0.ANT1e_CMCC8211	ADRD00000000	China	Cui et al. 2013
0.ANT1f_42095	ADPJ00000000	China	Cui et al. 2013
0.ANT1g_CMCC42007	ADQV00000000	China	Cui et al. 2013
0.ANT1h_CMCC43032	ADQW00000000	China	Cui et al. 2013
0.ANT2_B42003004	AAYU00000000	China	Cui et al. 2013
0.ANT2a_2330	ADOY00000000	China	Cui et al. 2013
0.ANT3_231	JMUF00000000	FSU	Eroshenko et al. 2017
0.ANT3_790	CP006806	FSU	Zhgenti et al. 2015
0.ANT3_A-1486	LYMP00000000	FSU	Eroshenko et al. 2017
0.ANT3a_CMCC38001	ADQU00000000	China	Cui et al. 2013
0.ANT3b_A1956001	ADPX00000000	China	Cui et al. 2013
0.ANT3c_42082	ADPH00000000	China	Cui et al. 2013
0.ANT3d_CMCC21106	ADQP00000000	China	Cui et al. 2013
0.ANT3e_42091b	ADPI00000000	China	Cui et al. 2013
0.ANT5_262	QAGF00000000	FSU	Kutyrev et al. 2018
0.ANT5_5M	LYMQ00000000	FSU	Kutyrev et al. 2018
0.ANT5_A-1691	LYMQ00000000	FSU	Eroshenko et al. 2017
0.ANT5_A-1836	LYOL00000000	FSU	Eroshenko et al. 2017
0.PE2_1412	CP006783	FSU	Zhgenti et al. 2015
0.PE2_1413	CP006762	FSU	Zhgenti et al. 2015
0.PE2_14735	AYLS00000000	FSU	Zhgenti et al. 2015
0.PE2_1522	CP006758	FSU	Zhgenti et al. 2015
0.PE2_1670	CP006806	FSU	Zhgenti et al. 2015
0.PE2_3067	CP006754	FSU	Zhgenti et al. 2015
0.PE2_3544	LZNH00000000	FSU	Kutyrev et al. 2018
0.PE2_3551	MBSJ00000000	FSU	Kutyrev et al. 2018
0.PE2_3770	CP006751	FSU	Zhgenti et al. 2015
0.PE2_835_BPC	LYOJ00000000	FSU	Kutyrev et al. 2018
0.PE2_8787	CP006748	FSU	Zhgenti et al. 2015
0.PE2_C-197	LIYX00000000	FSU	Kislichkina et al. 2015
0.PE2_C-235	LIYY00000000	FSU	Kislichkina et al. 2015
0.PE2_C-267	LIYZ00000000	FSU	Kislichkina et al. 2015
0.PE2_C-290	LIYU00000000	FSU	Kislichkina et al. 2015
0.PE2_C-291	LIZC00000000	FSU	Kislichkina et al. 2015
0.PE2_C-346	LIZE00000000	FSU	Kislichkina et al. 2015
0.PE2_C-359	LIZB00000000	FSU	Kislichkina et al. 2015
0.PE2_C-370	MIDX00000000	FSU	Kislichkina et al. 2017
0.PE2_C-535	MIDY00000000	FSU	Kislichkina et al. 2017
0.PE2_C-537	LIYP00000000	FSU	Kislichkina et al. 2015
0.PE2_C-590	LIYQ00000000	FSU	Kislichkina et al. 2015
0.PE2_C-666	LIZF00000000	FSU	Kislichkina et al. 2015
0.PE2_C-678	MIDZ00000000	FSU	Kislichkina et al. 2017
0.PE2_C-700	MIEA00000000	FSU	Kislichkina et al. 2017
0.PE2_C-712	MTZW00000000	FSU	Kislichkina et al. 2017
0.PE2_C-739	MTZX00000000	FSU	Kislichkina et al. 2017
0.PE2_C-741	LPTX00000000	FSU	Kutyrev et al. 2018
0.PE2_C-746	MTZY00000000	FSU	Kislichkina et al. 2017
0.PE2_C-824	MTZZ00000000	FSU	Kislichkina et al. 2017

Modern strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
0.PE2_KM874	LZTG00000000	FSU	Kutyrev et al. 2018
0.PE2_M-986	LYMO00000000	FSU	Kutyrev et al. 2018
0.PE2_PEST-F	NC_009381	FSU	Cui et al. 2013
0.PE2b_G8786	ADSG00000000	FSU	Cui et al. 2013
0.PE4_5307-Gis	LIYS00000000	FSU	Kislichkina et al. 2015
0.PE4_A-1804	LIYW00000000	FSU	Kislichkina et al. 2015
0.PE4_A-1807	LIYT00000000	FSU	Kislichkina et al. 2015
0.PE4_A-513	LIZA00000000	FSU	Kislichkina et al. 2015
0.PE4_I-3134	LIYR00000000	FSU	Kislichkina et al. 2015
0.PE4_I-3442	NHYH00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3443	MIED00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3446	NHYI00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3447	MIEE00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3455	LIYV00000000	FSU	Kislichkina et al. 2015
0.PE4_I-3515	NHYJ00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3516	NHMW00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3517	NHMX00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3518	NHMY00000000	FSU	Kislichkina et al. 2018a
0.PE4_I-3519	NHMZ00000000	FSU	Kislichkina et al. 2018a
0.PE4_M0000002	ADST00000000	China	Cui et al. 2013
0.PE4_Microtus91001	NC_005810	China	Cui et al. 2013
0.PE4a_B1313	LYMS00000000	FSU	Kutyrev et al. 2018
0.PE4a_I-2751-55	LYCL00000000	FSU	Kutyrev et al. 2018
0.PE4a_I-2998	LYMR00000000	FSU	Kutyrev et al. 2018
0.PE4Aa_12	ADOV00000000	China	Cui et al. 2013
0.PE4Ab_9	ADPT00000000	China	Cui et al. 2013
0.PE4Ba_PestoidesA	ACNT00000000	FSU	Cui et al. 2013
0.PE4Ca_CMCCN010025	ADRT00000000	China	Cui et al. 2013
0.PE4Cc_CMCC18019	ADQO00000000	China	Cui et al. 2013
0.PE4Cd_CMCC93014	ADRM00000000	China	Cui et al. 2013
0.PE4Ce_CMCC91090	ADRJ00000000	China	Cui et al. 2013
0.PE4h_A-1249	LYMN00000000	FSU	Eroshenko et al. 2017
0.PE4m_I-3086	LZNY00000000	Mongolia	Kutyrev et al. 2018
0.PE4t_A-1815	LPTY00000000	FSU	Eroshenko et al. 2017
0.PE5_I-2231	PVLX00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-2236	PVLZ00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-2238	PVLX00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-2239	LIZD00000000	Mongolia	Kislichkina et al. 2015
0.PE5_I-2422a	LIZG00000000	Mongolia	Kislichkina et al. 2015
0.PE5_I-2422b	QANK00000000	Mongolia	Kutyrev et al. 2018
0.PE5_I-2457	PVMB00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-3189	LIYO00000000	Mongolia	Kislichkina et al. 2015
0.PE5_I-3190	PVLY00000000	Mongolia	Kislichkina et al. 2018b
0.PE7b_620024	ADPM00000000	China	Cui et al. 2013
1.ANT1_Antiqua	NC_008150	Congo	Cui et al. 2013
1.ANT1_UG05-0454	AAYR00000000	Uganda	Cui et al. 2013
1.IN1a_CMCC11001	ADQK00000000	China	Cui et al. 2013
1.IN1b_780441	ADPS00000000	China	Cui et al. 2013
1.IN1c_K21985002	ADSS00000000	China	Cui et al. 2013
1.IN2a_CMCC640047	ADRA00000000	China	Cui et al. 2013
1.IN2b_30017	ADPC00000000	China	Cui et al. 2013
1.IN2c_CMCC31004	ADQR00000000	China	Cui et al. 2013
1.IN2d_C1975003	ADPZ00000000	China	Cui et al. 2013
1.IN2e_C1989001	ADQB00000000	China	Cui et al. 2013
1.IN2f_710317	ADPP00000000	China	Cui et al. 2013
1.IN2g_CMCC05013	ADQF00000000	China	Cui et al. 2013

Modern strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
1.IN2h_5	ADPK00000000	China	Cui et al. 2013
1.IN2i_CMCC10012	ADQG00000000	China	Cui et al. 2013
1.IN2j_CMCC27002	ADQQ00000000	China	Cui et al. 2013
1.IN2k_970754	ADPW00000000	China	Cui et al. 2013
1.IN2l_D1991004	ADRX00000000	China	Cui et al. 2013
1.IN2m_D1964002b	ADRV00000000	China	Cui et al. 2013
1.IN2n_CMCC02041	ADQC00000000	China	Cui et al. 2013
1.IN2o_CMCC03001	ADQD00000000	China	Cui et al. 2013
1.IN2p_D1982001	ADRW00000000	China	Cui et al. 2013
1.IN2q_D1964001	ADRU00000000	China	Cui et al. 2013
1.IN3a_F1954001	ADSC00000000	China	Cui et al. 2013
1.IN3b_E1979001	AAYV00000000	China	Cui et al. 2013
1.IN3c_CMCC84038b	ADRF00000000	China	Cui et al. 2013
1.IN3d_YN1683	ADTD00000000	China	Cui et al. 2013
1.IN3e_YN472	ADTH00000000	China	Cui et al. 2013
1.IN3f_YN1065	ADTC00000000	China	Cui et al. 2013
1.IN3g_E1977001	ADRY00000000	China	Cui et al. 2013
1.IN3h_CMCC84033	ADRE00000000	China	Cui et al. 2013
1.IN3i_CMCC84046	ADRG00000000	China	Cui et al. 2013
1.ORI1_CA88	ABCD00000000	USA	Cui et al. 2013
1.ORI1_CO92	NC_003143	USA	Cui et al. 2013
1.ORI1a_CMCC114001	ADQL00000000	China	Cui et al. 2013
1.ORI1b_India195	ACNR00000000	India	Cui et al. 2013
1.ORI1c_F1946001	ADSB00000000	China	Cui et al. 2013
1.ORI2_F1991016	ABAT00000000	China	Cui et al. 2013
1.ORI2a_YN2179	ADTE00000000	Myanmar	Cui et al. 2013
1.ORI2c_YN2551b	ADTF00000000	China	Cui et al. 2013
1.ORI2d_YN2588	ADTG00000000	China	Cui et al. 2013
1.ORI2f_CMCC87001	ADRH00000000	China	Cui et al. 2013
1.ORI2g_F1984001	ADSD00000000	China	Cui et al. 2013
1.ORI2h_YN663	ADTI00000000	China	Cui et al. 2013
1.ORI2i_CMCCK100001a	ADRR00000000	China	Cui et al. 2013
1.ORI2i_CMCCK110001b	ADRS00000000	China	Cui et al. 2013
1.ORI3_IP275	AAOS00000000	Madagascar	Cui et al. 2013
1.ORI3_MG05-1020	AAYS00000000	Madagascar	Cui et al. 2013
1.ORI3a_EV76	ADSA00000000	Madagascar	Cui et al. 2013
2.ANT1_Nepal516	ACNQ00000000	Nepal	Cui et al. 2013
2.ANT1a_34008	ADPD00000000	China	Cui et al. 2013
2.ANT1b_34202	ADPE00000000	China	Cui et al. 2013
2.ANT2a_2	ADOX00000000	China	Cui et al. 2013
2.ANT2b_351001	ADPF00000000	China	Cui et al. 2013
2.ANT2c_CMCC347001	ADQS00000000	China	Cui et al. 2013
2.ANT2d_G1996006	ADSE00000000	China	Cui et al. 2013
2.ANT2e_G1996010	ADSF00000000	China	Cui et al. 2013
2.ANT2f_CMCC348002	ADQT00000000	China	Cui et al. 2013
2.ANT3_KM682	LPVG00000000	FSU	Kutyrev et al. 2018
2.ANT3a_CMCC92010	ADRL00000000	China	Cui et al. 2013
2.ANT3b_CMCC95001	ADRN00000000	China	Cui et al. 2013
2.ANT3c_CMCC96001	ADRO00000000	China	Cui et al. 2013
2.ANT3d_CMCC96007	ADRP00000000	China	Cui et al. 2013
2.ANT3e_CMCC67001	ADRB00000000	China	Cui et al. 2013
2.ANT3f_CMCC104003	ADQH00000000	China	Cui et al. 2013
2.ANT3g_CMCC51020	ADQY00000000	China	Cui et al. 2013
2.ANT3h_CMCC106002	ADQI00000000	China	Cui et al. 2013
2.ANT3i_CMCC64001	ADQZ00000000	China	Cui et al. 2013
2.ANT3j_H1959004	ADSI00000000	China	Cui et al. 2013

Modern strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
2.ANT3k_5761	ADPL00000000	Russia	Cui et al. 2013
2.ANT3l_735	ADPR00000000	Russia	Cui et al. 2013
2.MED0_C-627	MBSI00000000	FSU	Kutyrev et al. 2018
2.MED1_1045	CP006794	FSU	Zhgenti et al. 2015
2.MED1_1116-D	LPXS00000000	FSU	Kutyrev et al. 2018
2.MED1_1240	LZNI00000000	FSU	Kutyrev et al. 2018
2.MED1_139	QAPA00000000	FSU	Kutyrev et al. 2018
2.MED1_173	LQAZ00000000	FSU	Kutyrev et al. 2018
2.MED1_1906	LYOM00000000	FSU	Kutyrev et al. 2018
2.MED1_244	LZND00000000	FSU	Kutyrev et al. 2018
2.MED1_261	LZNG00000000	FSU	Kutyrev et al. 2018
2.MED1_2944	CP006792	FSU	Zhgenti et al. 2015
2.MED1_44	LZNF00000000	FSU	Kutyrev et al. 2018
2.MED1_A-1763	LQAW00000000	FSU	Kutyrev et al. 2018
2.MED1_A-1809	LYMF00000000	FSU	Eroshenko et al. 2017
2.MED1_A-1825	LYCM00000000	FSU	Kutyrev et al. 2018
2.MED1_A-1920	LYCO00000000	FSU	Kutyrev et al. 2018
2.MED1_C-791	LQAU00000000	FSU	Kutyrev et al. 2018
2.MED1_K1973002	AAYT00000000	China	Cui et al. 2013
2.MED1_KIM10	NC_004088	Iran	Cui et al. 2013
2.MED1_KM816	LPXU00000000	FSU	Kutyrev et al. 2018
2.MED1_KM918	LPQY00000000	FSU	Kutyrev et al. 2018
2.MED1_M-1448	LYCN00000000	FSU	Kutyrev et al. 2018
2.MED1_M-1453	LQAY00000000	FSU	Kutyrev et al. 2018
2.MED1_M-1484	LQAV00000000	FSU	Kutyrev et al. 2018
2.MED1_M-1524	LYCP00000000	FSU	Kutyrev et al. 2018
2.MED1_M-1773	LYMG00000000	FSU	Kutyrev et al. 2018
2.MED1_M-1864	LOHR00000000	FSU	Kutyrev et al. 2018
2.MED1_M-519	LQAX00000000	FSU	Kutyrev et al. 2018
2.MED1_M-549	LQBA00000000	FSU	Kutyrev et al. 2018
2.MED1_M-595	LYOH00000000	FSU	Kutyrev et al. 2018
2.MED1_M-978	LPXT00000000	FSU	Kutyrev et al. 2018
2.MED1b_2506	ADPA00000000	China	Cui et al. 2013
2.MED1c_2654	ADPB00000000	China	Cui et al. 2013
2.MED1d_2504	ADOZ00000000	China	Cui et al. 2013
2.MED2b_91	ADPU00000000	China	Cui et al. 2013
2.MED2c_K11973002	AAYT00000000	China	Cui et al. 2013
2.MED2d_A1973001	ADPY00000000	China	Cui et al. 2013
2.MED2e_7338	ADPQ00000000	China	Cui et al. 2013
2.MED3a_J1963002	ADSP00000000	China	Cui et al. 2013
2.MED3b_CMCC125002b	ADQN00000000	China	Cui et al. 2013
2.MED3c_I1969003	ADSK00000000	China	Cui et al. 2013
2.MED3d_J1978002	ADSQ00000000	China	Cui et al. 2013
2.MED3f_I1970005	ADSL00000000	China	Cui et al. 2013
2.MED3g_CMCC99103	ADRQ00000000	China	Cui et al. 2013
2.MED3h_CMCC90027	ADRI00000000	China	Cui et al. 2013
2.MED3i_CMCC92004	ADRK00000000	China	Cui et al. 2013
2.MED3j_I2001001	ADSO00000000	China	Cui et al. 2013
2.MED3k_CMCC12003	ADQM00000000	China	Cui et al. 2013
2.MED3l_I1994006	ADSN00000000	China	Cui et al. 2013
2.MED3m_SHAN11	ADTA00000000	China	Cui et al. 2013
2.MED3n_SHAN12	ADTB00000000	China	Cui et al. 2013
2.MED3o_I1991001	ADSM00000000	China	Cui et al. 2013
2.MED3p_CMCC107004	ADQJ00000000	China	Cui et al. 2013
3.ANT1a_7b	ADPN00000000	China	Cui et al. 2013
3.ANT1b_CMCC71001	ADRC00000000	China	Cui et al. 2013

Modern strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
3.ANT1c_C1976001	ADQA00000000	China	Cui et al. 2013
3.ANT1d_71021	ADPO00000000	China	Cui et al. 2013
3.ANT2a_MGJZ6	ADSX00000000	Mongolia	Cui et al. 2013
3.ANT2b_MGJZ7	ADSY00000000	Mongolia	Cui et al. 2013
3.ANT2c_MGJZ9	ADSZ00000000	Mongolia	Cui et al. 2013
3.ANT2d_MGJZ11	ADSU00000000	Mongolia	Cui et al. 2013
3.ANT2e_MGJZ3	ADSW00000000	Mongolia	Cui et al. 2013
4.ANT_1454	LZNC00000000	FSU	Kutyrev et al. 2018
4.ANT_338	LZNX00000000	FSU	Kutyrev et al. 2018
4.ANT_517	LYMH00000000	FSU	Kutyrev et al. 2018
4.ANT_KM932	LZNE00000000	FSU	Kutyrev et al. 2018
4.ANT_M-1944	LYOK00000000	FSU	Kutyrev et al. 2018
4.ANT1a_MGJZ12	ADSV00000000	Mongolia	Cui et al. 2013
Ancient strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
DA101	PRJEB25891	Kyrgyzstan	Damgaard et al. 2018
Altenerding	PRJEB14851	Germany	Feldman et al. 2016
8124/8291/11972	SRR341961, SRR341962, SRR341963	United Kingdom	Bos et al. 2011
Ellwangen	PRJEB13664	Germany	Spyrou et al. 2016
Bolgar	PRJEB13664	Russia	Spyrou et al. 2016
OBS107	PRJEB12163	France	Bos et al. 2016
OBS110	PRJEB12163	France	Bos et al. 2016
OBS116	PRJEB12163	France	Bos et al. 2016
OBS124	PRJEB12163	France	Bos et al. 2016
OBS137	PRJEB12163	France	Bos et al. 2016

Table S13: Radiocarbon and archaeological dates of all individuals from plague-positive contexts in this study and of two published sites. Raw dates were calibrated with IntCal13 in OxCal v4.3.2. Archaeological dates are based on associated grave goods and cemetery stratigraphy. The probability plots for all dates are shown in SI Appendix, Fig. S8.

Site	Arch. ID	Lab ID	Radiocarbon dating lab ID	Skeletal element	Conventional Radiocarbon Age BP	$\delta^{13}C$ AMS [‰]	cal AD 2σ	Arch. Dating	Source
	Previously published individuals								
Altenerding	1175	AE1175	MAMS21886	Metacarpus	1541±19	-21.1	427–573	530–570	(3)
Altenerding	1176	AE1176	MAMS21887	Metacarpus	1563±20	-22.9	425–546	530–570	(3)
Aschheim	58	A58	MAMS15897	Rib	1563±18	-18.4	426–544	530–570	(19, 56)
Aschheim	76	A76	MAMS15898	Rib	1536±18	-21.6	428–580	530–570	(19)
Aschheim	120	A120	MAMS16593	Rib	1525±21	-17.2	430–599	525–550	(2)
	Individuals identified in this study								
Dittenheim	18A	DIT003	MAMS34551	Rib	1533±25	-31.2	428–595	500–600	This study
Edix Hill	Sk405	EDI001	MAMS38612	Rib	1497±24	-28.7	474–637	500–550	This study
Lunel-Viel	1	LVC001	MAMS34561	Femur	1492±23	-13.1	539–632	400–600	This study
Lunel-Viel	2	LVC002	MAMS34562	Femur	1474±24	-21.7	550–640	400–600	This study
Lunel-Viel	3A	LVC003	MAMS34563	Femur	1471±23	-19.6	555–641	400–600	This study
Lunel-Viel	4	LVC005	MAMS34554	Femur	1505+24	-24.9	434–618	400–600	This study
Lunel-Viel	5	LVC006	MAMS34565	Femur	1485 ±24	-17.6	543–635	400–600	This study
Lunel-Viel	6	LVC007	MAMS34566	Femur	1453±24	-25.1	567–648	400–600	This study
Petting	378	PET004	MAMS31028	Rib	1517±17	-22.3	434–600	570–600	This study
Petting	628	PET005	MAMS31033	Rib	1487±21	-23.8	542–630	530–730	This study
Petting	630	PET006	MAMS31034	Rib	1551±20	-19.6	427–558	570–600	This study
Saint-Doulchard	F206-40	LSD003	Ly-15568	Bone	1307±31	-20.6	656–770	650–800	(20)
Saint-Doulchard	F206-138	LSD027	Ly-15581	Bone	1235±30	-20.8	686–880	650–800	(20)
Saint-Doulchard	F206-145	LSD028	Ly-15625	Bone	1269±28	-20.2	665–853	650–800	(20)
Unterthürheim	129	UNT002	MAMS31035	Rib	1555±20	-19.0	427–553	550–600	This study
Unterthürheim	130	UNT003	MAMS31029	Rib	1520±17	-21.9	433–599	550–600	This study
Unterthürheim	131	UNT004	MAMS31036	Rib	1587±20	-20.9	419–537	550–600	This study
Unterthürheim	132	UNT005	MAMS34558	Rib	1505±24	-24.6	434–618	550–600	This study
Unterthürheim	134	UNT008	MAMS34568	Rib	1576±23	-23.2	421–542	550–600	This study
Unterthürheim	63	UNT006	MAMS31037	Rib	1507±20	-22.7	437–611	550–600	This study
Valencia	41/1657	VAL001	MAMS29429	Tooth	1512±22	-18.2	432–610	500–700	This study
Waging	199	WAG001	MAMS31030	Rib	1547±16	-15.7	427–563	500–550	This study

Table S14: Table giving all unique and shared SNPs that were called in the re-analyzed Aschheim genome (A120), showing the classification as assessed to the other genomes. The classification is based on the preceding parameters regarding a 50 bp window surrounding the SNP: Only SNPs with no uncovered position, no heterozygous SNP and an equal mean coverage with high and low stringent mapping (ratio of 1.00) would be classified as true positive SNPs (highlighted in green, and blue if the coverage is below 5).

Unique SNPs				
Position	SNP coverage	Uncovered pos.	Heterozygous SNPs	Mean coverage ratio LS/HS
22040	4	0	0	1.11
68401	4	6	1	1.11
90292	4	0	3	1.00
111916	3	24	0	1.27
118450	3	0	1	1.00
118467	5	0	1	1.06
119550	4	0	1	1.16
164957	4	5	1	1.24
206282	5	0	1	1.00
208068	4	0	2	1.00
208083	4	0	2	1.00
216157	8	0	1	3.93
221811	10	20	0	2.88
221849	11	0	2	2.76
222670	4	0	2	1.17
227200	4	0	2	1.09
228754	4	4	1	1.00
229308	4	8	0	2.08
229316	4	6	0	2.09
290016	4	0	2	1.00
318963	3	13	0	1.00
344159	4	0	0	1.00
352307	4	0	2	1.19
362357	4	17	0	5.11
363792	5	0	1	1.00
395774	37	0	0	4.65
439022	3	0	0	1.00
439036	3	0	0	1.00
473451	4	20	1	2.76
490767	4	7	1	1.24
500745	4	0	1	1.00
546706	3	0	0	1.03
567757	3	0	0	1.00
573139	4	0	1	0.99
573636	3	0	0	1.00
609900	5	13	1	2.54
610872	3	23	0	2.11
610908	6	19	1	2.13
705739	3	0	0	1.00
717675	4	24	2	1.44
723724	4	0	2	1.00
740842	3	6	0	1.15
788396	4	0	1	1.00
869369	4	0	2	1.26
912841	4	0	2	1.00
951452	4	0	1	1.00
1000454	4	2	1	1.00

Unique SNPs				
Position	SNP coverage	Uncovered pos.	Heterozygous SNPs	Mean coverage ratio LS/HS
1017104	4	0	0	1.00
1029706	5	1	1	1.00
1179456	4	0	1	1.00
1278057	3	22	1	1.30
1287721	5	0	1	1.00
1308293	37	0	1	1.75
1349463	4	0	2	1.00
1349472	4	0	2	1.00
1365939	4	0	1	1.00
1371025	10	2	0	3.22
1389874	3	0	1	1.09
1421032	27	0	2	1.75
1444672	4	0	1	1.15
1451459	4	0	2	1.27
1487354	4	0	1	1.30
1498866	3	0	0	1.00
1534269	4	0	1	1.00
1543236	4	10	2	1.00
1555873	4	0	2	1.00
1572689	3	12	0	12.78
1610886	3	9	0	1.00
1711880	3	0	0	1.27
1715867	5	0	0	1.20
1722937	6	0	1	1.21
1745586	4	0	1	1.13
1761696	4	0	1	1.00
1806990	4	7	2	1.06
1854844	4	14	1	1.28
1864422	4	0	2	1.00
1927370	3	8	0	1.00
1943523	6	7	1	1.57
1959365	4	0	1	1.00
2009537	4	1	1	1.00
2072340	4	0	1	1.33
2072914	5	0	1	2.38
2084670	4	0	1	1.10
2336322	3	12	0	1.00
2345720	5	0	1	1.00
2350460	4	0	1	1.26
2361259	4	5	2	1.46
2415931	5	0	1	1.00
2417705	4	0	1	1.13
2459674	4	0	1	1.00
2513785	4	0	2	1.00
2586708	4	0	1	1.00
2591088	4	0	3	2.86
2664231	4	0	2	1.00
2714188	5	0	1	1.37
2836744	4	6	1	1.00
2847001	4	0	1	1.36
2861901	3	21	0	1.00
2861902	3	20	0	1.00
2863957	4	0	2	1.00
2865494	3	3	0	3.19
2972201	4	0	1	1.00
3070014	44	0	3	1.68

Unique SNPs				
Position	SNP coverage	Uncovered pos.	Heterozygous SNPs	Mean coverage ratio LS/HS
3070065	56	0	1	1.38
3074956	3	23	0	1.16
3090759	4	0	1	1.00
3131243	3	15	0	1.00
3152585	4	0	3	1.39
3205018	4	0	3	1.15
3242622	3	2	0	1.21
3243987	4	0	1	1.41
3247821	4	0	0	1.00
3249954	4	1	2	1.00
3263455	4	0	1	1.24
3266505	3	10	1	1.31
3286005	3	0	1	1.00
3286017	4	0	1	1.00
3288015	4	0	2	1.00
3342782	4	6	1	1.00
3373238	5	0	2	1.00
3392897	18	2	1	3.16
3428541	4	0	1	1.00
3429285	6	0	1	3.25
3479274	4	0	2	1.00
3498406	4	0	1	1.03
3505837	4	0	1	1.00
3531163	4	0	1	1.40
3531197	4	0	1	1.14
3540557	4	0	3	1.87
3543444	3	0	1	1.00
3569690	4	0	1	1.25
3579698	4	0	1	1.32
3588681	3	8	0	1.00
3656870	4	0	1	1.11
3672205	8	0	1	1.35
3813424	7	23	0	2.51
3860098	4	0	1	1.07
3934112	3	7	0	1.03
3956001	9	0	2	2.96
3956018	4	0	5	3.46
4170791	45	0	1	1.80
4199220	30	0	0	6.01
4203620	23	0	1	6.18
4232217	3	8	0	1.00
4342798	4	12	1	1.78
4509055	26	0	0	2.46
4550310	3	0	1	1.26
4575345	33	0	0	4.52

Shared SNPs				
Position	SNP coverage	Uncovered pos.	Heterozygous SNPs	Mean coverage ratio LS/HS
86824	2	0	0	1.00
189912	0	23	0	1.00
260148	0	34	0	2.39
271114	3	0	0	1.48
485976	3	0	0	1.00
549767	2	0	0	1.00
557841	1	0	0	1.80
727741	5	0	2	1.53
779365	2	0	0	1.00
898980	5	0	1	1.16
1067966	6	0	0	1.11
1211729	1	0	0	1.00
1296743	1	0	0	1.58
1387701	2	0	0	1.00
1413031	0	24	0	4.23
1434752	2	6	1	1.00
1489055	2	1	0	1.42
1530658	1	16	0	1.38
1609461	2	0	2	1.29
1754708	2	0	1	1.50
1868678	3	0	1	1.00
1956162	5	0	1	1.24
2092152	2	0	0	1.85
2097520	1	0	0	1.00
2352174	1	0	0	1.21
2419529	0	24	0	1.00
2725715	3	0	0	1.00
2753572	1	0	0	1.00
2977542	0	15	0	1.00
3078807	1	18	0	1.48
3274298	3	0	1	1.00
3360963	0	41	0	4.04
3360984	0	50	0	NA
3398153	1	0	0	1.00
3409414	18	0	0	1.00
3500922	3	0	0	1.17
3535148	2	0	1	1.10
3560088	3	0	0	1.00
3568597	3	0	0	1.47
3750736	3	0	1	1.58
3843195	3	0	0	1.08
3892488	1	0	2	1.02
4066494	5	0	1	1.12
4307755	3	0	0	1.00
4412624	2	0	0	1.28
4423366	4	0	0	1.37
4460688	8	0	0	1.07
4465967	8	0	0	1.02
4628496	4	0	1	1.00
4629169	10	0	0	1.18

References

1. Renaud G, Hanghøj K, Willerslev E, Orlando L (2017) Sequence analysis gargammel: a sequence simulator for ancient DNA. *Bioin* 33:577–579.
2. Wagner DM, et al. (2014) Yersinia pestis and the Plague of Justinian 541–543 AD: A genomic analysis. *Lancet Infect Dis* 14(4):319–326.
3. Feldman M, et al. (2016) A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol Biol Evol* 33(11):2911–2923.
4. Deng Z, et al. (2015) TyrR, the regulator of aromatic amino acid metabolism, is required for mice infection of Yersinia pestis. *Front Microbiol* 6(110).
5. Chain PSG, et al. (2006) Complete genome sequence of Yersinia pestis strains Antiqua and Nepal516: Evidence of gene reduction in an emerging pathogen. *J Bacteriol* 188(12):4453–4463.
6. Chain PSG, et al. (2004) Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. *Proc Natl Acad Sci* 101(38):13826–13831.
7. Schmid A, et al. (2009) Cross-talk between type three secretion system and metabolism in Yersinia. *J Biol Chem* 284(18):12165–12177.
8. Coleman MA, et al. (2016) Expression and Association of the Yersinia pestis Translocon Proteins, YopB and YopD, Are Facilitated by Nanolipoprotein Particles. *PLoS One* 11(3):e0150166.
9. Deng W, et al. (2002) Genome Sequence of Yersinia pestis KIM. *J Bacteriol* 184(16):4601–4611.
10. Zhou D, et al. (2004) Genetics of Metabolic Variations between Yersinia pestis Biovars and the Proposal of a New Biovar, microtus. *J Bacteriol* 186(15):5147–5152.
11. Harper K (2017) *The Fate of Rome: Climate, Disease, and the End of an Empire* (Princeton University Press, Princeton).
12. Bos KI, et al. (2011) A draft genome of Yersinia pestis from victims of the Black Death. *Nature* 478(7370):506–10.
13. Spyrou MA, et al. (2016) Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host Microbe* 19(6):874–881.
14. McCormick M (2002) *Origins of the European Economy* (Cambridge University Press, Cambridge).
15. McCormick M (1998) Bateaux de vie, bateaux de mort. Maladie, commerce, transports annonaires et le passage économique du Bas-Empire au Moyen Âge. *Morfologie Sociali e Culturali in Europa* (Centro Italiano di Studi sull'Alto Medioevo, Spoleto), pp 35–118.
16. Benedictow OJ (2004) *The Black Death 1346–1353: The Complete History* (Boydell & Brewer, Woodsbridge).
17. Yue RPH, Lee HF, Wu CYH (2016) Navigable rivers facilitated the spread and recurrence of plague in pre-industrial Europe. *Sci Rep* 6(1):34867.
18. Roosen J, Curtis DR (2018) Dangers of noncritical use of historical plague data. *Emerg Infect Dis* 24(1):103–110.
19. Harbeck M, et al. (2013) Yersinia pestis DNA from Skeletal Remains from the 6th Century AD Reveals Insights into Justinianic Plague. *PLoS Pathog* 9(5):e1003349.
20. Maçon P, Durand R, Salin M, Dominin F (2011) *Le Pressoir - Rapport final d'opération de fouille préventive* (Bourges).
21. Reimer PJ, et al. (2013) IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* 55(4):1869–1887.
22. Ramsey CB (2017) Methods for Summarizing Radiocarbon Datasets. *Radiocarbon* 59(6):1809–1833.
23. Ascough PL, et al. (2010) Temporal and spatial variations in freshwater 14C reservoir effects: Lake Mývatn, northern Iceland. *Radiocarbon* 52(2–3):1098–1112.

24. Ascough P, Cook G, Dugmore A (2005) Methodological approaches to determining the marine radiocarbon reservoir effect. *Prog Phys Geogr* 4:532–547.
25. Barta P, Štolc S (2007) HBCO Correction: Its Impact on Archaeological Absolute Dating. *Radiocarbon* 49(2 "Proceedings of the 19th international Radiocarbon Conference, Keble College, Oxford, England"):465–472.
26. Wamser L (2010) *Karfunkelstein und Seide. Neue Schätze aus Bayerns Frühzeit* (Pustet, Regensburg).
27. Grupe G, Harbeck M, McGlynn GC (2015) *Prähistorische Anthropologie* (Springer, Berlin Heidelberg).
28. Pöllath R (2002) *Karolingerzeitliche Gräberfelder in Nordostbayern: eine archäologisch-historische Interpretation mit der Vorlage der Ausgrabungen von K. Schwarz in Weismain und Thurnau-Alladorf* (München).
29. Leinthaler B (1988) Der Karolingisch-ottonische Ortsfriedhof Alladorf, Ldkr. Kulmbach. Die Grabungskampagne 1984. *Zeitschrift für Archäologie des Mittelalters* 16/17:7–122.
30. Christlein R (1971) *Das alamannische Gräberfeld von Dirlewang bei Mindelheim* (Kallmünz).
31. Haas-Gebhard B (1998) *Ein frühmittelalterliches Gräberfeld bei Dittenheim (D)* (Montagnac).
32. Malim T, Hines J (1998) *The Anglo-Saxon cemetery at Edix Hill (Barrington A), Cambridgeshire* (Council for British Archaeology, York).
33. Hines JA, Bayliss AL (2013) *Anglo-Saxon graves and grave goods of the Sixth and Seventh Centuries AD: A chronological framework.* (Society for Medieval Archaeology, Leeds). SMA Monogr.
34. Kropf E, Zintl S (2007) Eine frühmittelalterliche Vierfachbestattung aus Forchheim: Markt Pförring, Landkreis Eichstätt, Oberbayern. *Das archäologische Jahr Bayern* 2006:120–123.
35. von Freeden U (1983) Das frühmittelalterliche Gräberfeld von Grafendobrach in Oberfranken. *Bericht der Römisch-Germanischen Kommission* 64:417–507.
36. Pescheck C (1996) *Das fränkische Reihengräberfeld von Kleinlangheim, Lkr. Kitzingen/Nordbayern* (Mainz).
37. Daim F (1987) *Das awarische Gräberfeld von Leobersdorf, Niederösterreich* (Wien).
38. McCormick M (2016) Tracking mass death during the fall of Rome's empire (II): a first inventory of mass graves. *J Rom Archaeol* 29:1004–1046.
39. Raynaud C (2010) *Les nécropoles de Lunel-Viel (Hérault) de l'Antiquité au Moyen Âge. Revue archéologique de Narbonnaise Supplément 40* (Presses universitaires de la Méditerranée, Montpellier).
40. Barruol G, Garnotel A, Raynaud C (2017) Maguelone de l'Antiquité à nos jours: histoire, archéologie et environnement. *Palavas-Les-Flots: Les Compagnons de Maguelone*, pp 50–51.
41. Dannheimer H (1998) *Das baiuwarische Reihengräberfeld von Aubing, Stadt München* (Stuttgart).
42. Keller E (1979) *Das spätrömische Gräberfeld von Neuburg an der Donau* (Kallmünz). Materialhefte zur bayerischen Vorgeschichte
43. Bierbrauer V (2002) Neuburg. *Reallexikon Der Germanischen Altertumskunde, Vol. 21* (Berlin), pp 106–108.
44. von Freeden U, Lehmann D (2005) *Das frühmittelalterliche Gräberfeld von Peigen, Gem. Pilsting* (Landau a. d. Isar).
45. Haas-Gebhard B, Weindauer F (2013) Die Fibelgräber der frühmittelalterlichen Nekropole von Petting (Oberbayern). *Bayerische Vorgeschichtsblätter* 78:205–234.
46. Reimann D (1991) Byzantinisches aus dem Rupertiwinkel – Zum Ohrringpaar von Petting. *Das archäologische Jahr Bayern* (1991):143–145.
47. Loré F (2012) Gräber und kein Ende? Neue Grabungen im Großen Gräberfeld von Regensburg: Oberpfalz. *Das archäologische Jahr Bayern* 2011:96–98.
48. Menke H, Menke M (2013) *Das frühmittelalterliche Gräberfeld von Sindolvesdorf/Sindelsdorf, Lkr. Weilheim-Schongau* (Kallmünz).
49. Moosbauer G (2005) *Kastell und Friedhöfe der Spätantike in Straubing: Römer und Germanen auf dem Weg zu den ersten Bajuwaren* (Rahden). Passauer Universitätsschriften zur Archäologie.

50. Grünewald C (1988) *Das alamannische Gräberfeld von Unterthürheim, Bayerisch-Schwaben* (Kallmünz).
51. Alapont Martín L, Ribera i Lacomba AV (2009) Topografía y jerarquía funeraria en la Valencia tardo-antigua. *Morir en el Mediterráneo Medieval: Actas del III Congreso Internacional de Arqueología, Arte e Historia de La Antigüedad Tardía y Alta Edad Media Peninsular* (Oxford), pp 59–88.
52. Ribera i Lacomba AV, Alapont Martín L (2006) Cementerios tardoantiguos de Valencia: arqueología y antropología. *An Arqueol Cordobesa* 17(2):161–194.
53. Ribera i Lacomba AV, Soriano R (1996) *Los cementerios de época visigoda* (València).
54. Knöchlein R (1998) *Das Reihengräberfeld von Waging am See* (Liliom, Waging am See).
55. Reiß R (1994) *Der merowingerzeitliche Reihengräberfriedhof von Westheim (Kreis Weißenburg-Gunzenhausen): Forschungen zur frühmittelalterlichen Landesgeschichte im südwestlichen Mittelfranken* (Nürnberg).
56. Haas-Gebhard B (2017) Die Pest des Frühen Mittelalters in der Münchner Schotterebene. *Bayerische Vorgeschichtsblätter* 82:191–210.

Supplementary Material for Manuscript B

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

Maria A. Spyrou*, Marcel Keller*, Rezeda I. Tukhbatova, Christiana L. Scheib, Elizabeth A. Nelson, Aida Andrades Valtueña, Gunnar U. Neumann, Don Walker, Amelie Alterauge, Niamh Carty, Craig Cessford, Hermann Fetz, Michaël Gourvennec, Robert Hartle, Michael Henderson, Kristin von Heyking, Sarah A. Inskip, Sacha Kacki, Felix M. Key, Elizabeth L. Knox, Christian Later, Prishita Maheshwari-Aplin, Joris Peters, John E. Robb, Jürgen Schreiber, Toomas Kivisild, Dominique Castex, Sandra Lösch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos & Johannes Krause
*contributed equally

Published in *Nature Communications* 10, 4470 (2019) doi:10.1038/s41467-019-12154-0

Supplementary Note 1

Archaeological context information

"Laishevo III cemetery", Laishevo Republic of Tatarstan, Russia

The Laishevo III cemetery was discovered southeast of the Laishevo town (Laishevo district, Tatarstan Republic, Russia), on the confluence of the Volga and Kama rivers. The site was excavated in 1979, with a large quantity of "Bulgarian" pottery from the Golden Horde period collected from both an eroded part of the site and from burials. It is suggested that the cemetery was located on the site of the settlement, called the Laishevo III settlement. Excavations were carried out along the eroding part of the cemetery. A total of 34 Muslim-type burials were excavated. The individuals were buried in supine position; their heads were oriented to the west, with minor deviations to the south. Hands were usually positioned on the abdomen or the chest. Apart from pottery, few additional artefacts were identified within the burials, such as belt buckles and beads, as well as a single bronze earring. Based on the ceramic findings and the bronze earring, it is suggested that the cemetery most likely dates to the Golden Horde period of Volga Bulgaria (which mainly encompassed the 13th and 14th centuries AD) – though the identified assemblage was suggested to likely not date earlier than the 14th century. Anthropological examination identified skeletal remains of 40 individuals: 14 children of up to 10 years, 13 males and 13 females.

Burial no. 10, where individual LAI010 was discovered, was partially destroyed, and had a depth of 40 cm. Two skeletons were found in this burial. Individual LAI010 **(**skeleton A from burial 10) was located in the southern part of the grave pit, and only the upper part of the skeleton was preserved. Through anthropological examination, the individual was suggested to be a 35–45 year-old male.

The triple burial 27, where individual LAI009 was found, was also at a depth of 40 cm. The skeletal remains of individual LAI009 (skeleton B from burial 27) were not well preserved. The skeleton was disarticulated with the skull found on the western part of grave pit and the remaining skeletal elements found between skeletons A and C. An iron buckle was found inside the skull, and the only bronze earring discovered on this site was also found at closest proximity to this skull than any other skeleton. Although from anthropological examination the individual was suggested to be a male of 30–40 years of age, genetic sexing suggests that it belonged to a female.

"16 rue des Trente Six Ponts", Toulouse, France

The archaeological rescue excavation that took place at 16 rue des Trente Six Ponts in Toulouse has uncovered a funerary space used between the 5th century and the end of the 14th century. The funeral area of the Late Middle Ages is the most important as it consists of 109 graves that include 29 multiple graves for a total of 444 individuals of which 306 are buried in three mass graves.

The medieval funeral complex corresponds to the eastern extension of Saint-Michel graveyard discovered in 2002[29]. This cemetery is divided into two parallel rows. These rows are separated by a vacant space, without graves or archaeological structures, allowing movement within the funerary area. The three mass graves which are characteristic of so-called "mortality crisis" graves, are found at the eastern ends of these rows. The graves can be dated by several methods. First, the orientation of the burials corresponds to the one observed for the Late Middle Ages at the excavation of the Saint-Michel cemetery in 2002[29]. Second, the artefacts uncovered in the deposits within the same strata as these graves can be dated between the second half of the 13th century and the end of the 14th century.

Moreover, some dating evidence allows us to understand more precisely the evolution of this funeral complex in the Middle Ages. A first period, the end of the 13th century to the first half of the 14th century corresponds to the beginning of the use of this space as a cemetery. This first phase is dated by a few coin hoards excavated with the buried

bodies. Two graves contain coins from the end of the 13th century (1245–1270). Six burials were dated by ^{14}C, three of which belong to a period between the second half of the 13th century and the first quarter of the 14th century.

The second phase corresponds to a "crisis" cemetery, probably linked to a plague epidemic that affected Toulouse between 1347–1350. This hypothesis is supported by several coin finds. Their contemporaneity with the Black Death can be demonstrated by depositions within a double burial (SP2806) as well as in two of the mass graves. The first deposit is found within the double burial and consists of 34 "double tournois" of Philippe VI and of two "double parisis" of Philip VI that were found to be associated with one of the bodies. These coins are thought to have been issued between March 1347 and August 1348. The second deposit, related to the second burial, consists of two "double parisis" of Philip VI issued between April 1347 and August 1350. The third deposit is associated with one of the mass graves (SP1827) and contains 36 coins ("double tournois" and "double parisis" of Philippe VI) also issued between March 1347 and August 1348. Finally, the second mass grave (SP3408) contains a deposit of four "double tournois" of Philip VI issued between August and December 1348.

The grave SP1350 corresponds to a double burial dug in stratigraphic deposits post-dating the 13th century and has the same orientation as the other multiple graves dated to the same period. It is therefore contemporaneous with the plague episode that affected Toulouse during the late 1340s. The first, skeleton SQ1352 (TRP002), corresponds to an adult woman, aged 20 to 49 years, buried at the same time as the individual SQ1353, identified as an infant (0–1 years). Skeleton SQ1352 was buried in a flexible shroud type material unlike SQ1353, which does not appear to have been similarly treated. Both individuals were likely buried in a coffin.

"The New Churchyard", London, England

Population growth in 16th-century London, coupled with a severe outbreak of plague in 1563, led to the 'New Churchyard' opening in 1569, a municipal, non-parochial burial ground[30]. The chosen site was outside the north wall of the city, on land, which had previously been part of the priory of St Mary Bethlehem, later better known as 'Bethlem' or 'Bedlam' Hospital. Indeed, the ground was often referred to as the 'Bethlem' or 'Bedlam' burial ground because of its location.

When the New Churchyard closed to burials in 1739, it was densely packed with graves primarily including the remains of the city's poor and those on the fringes of society. The use of the ground covered a period in which London suffered several plague epidemics, particularly in 1603, 1625, 1636 and 1665, with parish records confirming that hundreds of victims of the disease were sent to the New Churchyard for burial.

During the Crossrail Central development at the Broadgate ticket hall worksite at Liverpool Street, London, in 2011–15, archaeological investigations by MOLA (Museum of London Archaeology)[31] resulted in the excavation of 3354 of the estimated *c.* 25,000 burials within the New Churchyard. This work provided an opportunity to explore health and disease of the city's inhabitants during a period of considerable population growth fed by migration to the expanding metropolis.

The discovery of a mass pit in the central area of the southern half of the burial ground, containing at least 42 individuals, provided an opportunity to investigate the first archaeologically excavated 17th-century plague burial in London. The pit contained stacks of coffined and un-coffined burials, up to eight individuals deep. It was filled in a single event. The head ends of the coffins were alternated to allow the maximum use of space within the pit. A single perpendicular line of burials filled a gap at one end.

The archaeological dating, although imprecise and complicated by some intrusive finds introduced by later grave cuts, is consistent with a late 16th- or early 17th-century date. The east-west alignment of the pit was typical of burials from the earlier period of the burial ground's use, 1569 – 1670. The fill contained a small pottery assemblage dated between 1550 and 1610, and coffins within the pit were of a type that appeared in the last quarter of the 16th century and was ubiquitous from 1650 onwards. While radiocarbon dating was not sufficiently precise to distinguish between specific

plague events, the dates clearly indicated that the Great Plague of 1665 was too late to have been responsible for the mass burial and the outbreaks in 1603, 1625 and 1636 were the most likely.

Augustinian Friary, Cambridge, England

Documentary evidence indicates that the Augustinian Friary in Cambridge was founded between 1277 and 1289, as a mendicant institution with a strong focus on learning and continued until the Dissolution of the Monasteries in 1538. The friary probably acquired burial rights in 1290 if not earlier and in 1302 these were extended to include individuals who were not members of the Augustinian order. Excavations in 2016–17 by the Cambridge Archaeological Unit as part of the redevelopment of the New Museums site by the University of Cambridge revealed two phases of friary related activity, both with associated burials[32].

The earlier phase mainly comprised a cemetery located south of the friary church, where 32 burials that appear to be a mixture of friars and lay individuals were excavated. General archaeological evidence, including belt buckles associated with some of the burials, suggests that the cemetery spans the late 13th to mid-14th century, with radiocarbon dating of three burials in a stratigraphic sequence indicating that burial continued until at least the 1330s (all radiocarbon dates quoted to 95% probability). Ancient DNA analysis of one of the skeletons from the cemetery F.355 (NMS003) yielded a whole *Y. pestis* genome that likely has an identical genotype to other European genomes attributed to the Black Death episode (see Supplementary data 3), indicating that this burial likely dates to 1348/49. Stratigraphically F.355 (NMS003) predates two radiocarbon-dated burials F.332 and F.328. F.332 dates to 1286–1397 cal AD, but when factors including the marine reservoir effect, bone turnover/remodeling and stratigraphic sequence are considered it is dated to 1323–1426.

The second phase related to the friary is marked by the construction of the claustral ranges. Typologically the cloister arcade is dated to between *c*. 1330 and *c*. 1390. This construction is unlikely to have predated the late 1330s, as it was not until then that the friary managed to acquire enough properties to embark on this major project. Additionally, the legs of F.355 (NMS003), which was likely buried in 1348/49, were removed when the foundation trench was dug for the wall of the cloisters. As this process did not disturb the rest of the bones it is probable that some time had passed since the interment, conservatively suggesting a date of no earlier than *c*. 1360 for the wall footing.

Six individuals were buried in the Chapter House in the eastern claustral range of the friary *c.* 1360/90–1538. As with the earlier cemetery the burials appear to be a mixture of friars and lay individuals. Three of the burials from the Chapter House tested positive for *Yersinia pestis*; F.230 (NMS002) yielded a whole-genome, whereas F.190 (NMS001) and F.310 (NMS005) remain to be confirmed. There is no stratigraphic evidence to indicate the relative sequence of these three burials and the *Y. pestis* DNA could relate to either a single plague outbreak or multiple outbreaks. F.230 (NMS002) has a copper alloy symmetrical double oval frame belt buckle, dating to between *c*. 1350 and the 16th century and the grave fill contained some 15th- to 16th-century lead window came. F.230 has been radiocarbon dated to 1437–1619 cal AD, but when various factors are allowed–including the marine reservoir effect, the bone turnover/remodeling, the stratigraphic sequence and the fact that the burials date to 1538 at the latest–the probable date range is 1475–1536. There are several documented plague outbreaks in England during the late 15th and early 16th centuries, with major outbreaks in 1471 and 1479–80 and smaller scale recurrences in 1499–1500, 1509–21, 1526–32 and 1535–36, though it is uncertain which of these outbreaks affected Cambridge. The only specific dating evidence associated with F.190 (NMS001) is a small amount of pottery that indicates a 16th century date, and the location of the grave suggests that it may be contemporary with F.230 (NMS002). F.310 is also accompanied by a copper alloy symmetrical double oval frame belt buckle, dating to between 1350 and the 16th century.

"Sankt Johans Freidhof", Nabburg, Germany

The "Sankt Johans Freidhof" churchyard (sic, see Hensch 2014)[33] in the city of Nabburg in Southern Germany was excavated between 2008 and 2012 next to the hospital church "St. Maria", consecrated in 1420. However, the earliest graves are attributed to the older parish church "St. Johannes der Täufer" from the end of the 13th century. The ~200 excavated graves were found in a maximum of eight layers and could be classified in two groups: The older graves from the late 13th to late 14th centuries are dug in a strict and regular layout, the younger graves from the 15th century to the closure in 1597 are more irregular, presumably due to the lack of space. Only in four cases, the use of a coffin could be attested by the finds of iron nails and wooden remains.

Towards the western border of the cemetery, in total nine multiple burials were found with between two and four individuals. The four individuals that tested positive for *Y. pestis* originate from three multiple burials: NAB002 (arch. ID 451, early juvenile male) and NAB003 (arch. ID 452, early adult female) were found in a triple burial with an additional young woman, all piled up in a narrow grave pit in supine position. NAB004 (arch. ID 457, presumably female of 6–12 years, supine position) was buried in a simultaneous double burial together with an adult female on top in the opposite orientation. NAB005 (arch. ID 471, early mature female) was also buried in a simultaneous double burial, here with an adult male individual beneath in prone position.

A connection to the second plague pandemic was suggested by the archaeologists given the grave characteristics, despite the fact that there are no historical records for the Black Death in Nabburg. However, epidemic records exist for the nearby towns of Amberg, Sulzbach, Burglengenfeld and Regenstauf from 1349 onwards.

St. Leonhardi, Manching-Pichl, Germany

During the renovation works of 1984/85 at St. Leonhardi, a mass grave was found under the sacristy. The mass grave revealed a minimum number of 75 individuals in four layers. The construction of the mass grave suggests that it was not dug into the ground but the individuals were placed on ground level without coffins and afterwards covered with earth[28]. However, the removal of the ground floor within the nave also revealed the remains of six additional individuals with some of them considered to be the church donors. The high amount on disarticulated skeletal elements within the mass grave hints towards a more intensive use of the site as a burial ground. The construction of the sacristy can be dated to the second half of the 15th century, which would give a *terminus ante quem*, assuming that the sacristy was built after the deposition. Radiocarbon dates of earlier studies gave two contradicting intervals, one spanning roughly the 12th century and another spanning the 14th century[25-27], which could be explained by erroneously assigned scattered remains of earlier burials in this area.

The mass grave was repeatedly subject to aDNA studies that investigated the presence of *Y. pestis* DNA with PCR and qPCR assays[25-28,34]. In this study, individuals MAN008 (arch. ID MPS03-IV, adult female) and MAN015 (arch. ID MP56-X) tested positive for *Y. pestis* through qPCR and high-throughput sequencing (Supplementary table 2). Due to the apparently difficult assignment of skeletal elements to the individuals in the mass grave, only teeth in situ were sampled, so the jawbone could be used for ^{14}C dating (Supplementary table 1). The resultant ^{14}C dates (2-sigma: 1283–1390 calAD), revealed a 14th century interval for the deposition. Plague is thought to have first entered Bavaria in 1348 during the Black Death and continued to affect this region with epidemics during the second half of the 14th century, after the Black Death. Based on our current genomic data, the MAN008 isolate from Manching is genetically divergent from all other Black Death genomes from Europe, including one from the nearby city of Nabburg (Figure 2). As such, these results are in support of MAN008 being representative of a post-Black Death outbreak that occurred in Bavaria likely during the second half of the 14th century.

Possenhofener Str. 3, Starnberg, Germany

From 2007 to 2009 the southern and southeastern parts of the churchyard of the former parish church St. Benedikt in Achheim – todays Starnberg (Upper Bavaria) – were excavated. St. Benedikt was first mentioned in 1226 and demolished between 1764 and 1816[35]. The excavation revealed 365 burials, estimated to represent only one third of the cemetery and dating mostly between the 8th/9th and the 18th centuries. However, five stone plate graves date between the middle of the 7th and the early 9th centuries. This indicates, that the cemetery was originally founded as a private burial ground by an aristocratic family, probably living nearby in a manor, in the mid-7th century. The first church can be dated to the middle of the 7th century, too, and is actually one of the oldest known churches in Bavaria.

After the church was donated to Benediktbeuren Abbey in the 8th or 9th century, it was turned into a parish church for the surrounding village of Achheim. The Merovingian church was replaced by a larger one in the late 10th or early 11th century. This medieval church was once again completely rebuilt in the first half of the 15th century.

Achheim stayed a small village at the shore of Lake Starnberg until the mid-19th century, when it became part of the growing modern town of Starnberg. Starnberg Castle (first mentioned in 1245), which surmounts the former village of Achheim, was a temporary summer residence of the dukes of Bavaria during late medieval and early modern times, especially in the 15th and 16th centuries. During this period, several members of the aristocracy, closely connected to the ducal court or the administration of the castle, were buried in a side chapel of St. Benedikt. A couple of parish priests were buried in the church, too. Nevertheless, most burials can be identified with the common population of Achheim and probably some outlying homesteads – mostly fishermen, peasants or craftsmen. Thus, St. Benedikt represents a typical parochial churchyard of a rural community in Southern Bavaria.

The medieval graves are characterized by the complete absence of grave goods and can only be dated by stratigraphy, radiocarbon analysis and their arm posture. In the years around 1300 burial rites changed and few grave goods can be found again, especially rosaries as well as occasional coins, buckles or finger rings.

One triple burial – situated just a few meters southeast of the sacristy and belonging to the late medieval period – was the only multiple burial found at this site. Due to this, it was examined here for the presence of *Y. pestis*. Individual STA001 (arch. ID 207, infans II to juvenile, aged 12–14, presumably male) tested positive for the bacterium through qPCR and high throughput sequencing (Supplementary table 2). In its hands the individual held a rosary, as shown by few remains of blue glass beads, which cannot be dated precisely. The juvenile was simultaneously buried together with two other children aged 3–5 (arch. ID 208, infans I) and 12 (arch. ID 214, infans II), in regular supine position. The rosary found with STA003 (arch. ID 214) can be dated between the late 14th and the early 16th centuries. This special long form of a rosary was composed of approximately 150 ring-shaped bone beads (15 mm in diameter), represented by 122 intact beads and 38 fragments. Taking into account the 1-sigma interval of the radiocarbon dating, the date of the triple burial can be narrowed down to 1433–1494 (59,6 %), respectively to 1420–1523 (74,5 %) when considering the 2-sigma interval. Thus, these three subadults died during a historically undocumented plague outbreak in the 15th or early 16th century.

Kirchhof St. Johannis, Landsberg am Lech, Germany

After Landsberg was granted the town privilege in 1315 the small trading post on the crossroads of the salt route soon started to prosper[36]. Despite the outbreak of the plague in 1349 the town's population grew and at the beginning of the 17th century it reached an estimated number of around 4500 inhabitants. To take care of the salvation of the growing population the parish church was extended in the mid-15th century, resulting in a shortage of available burial space. As a result, a new cemetery was established within the town walls. The usage time of the St. Johannis churchyard could be narrowed down by historical documents to between 1507 and 1806. In this timespan Landsberg witnessed several plague epidemics: 1586, 1592, 1607, 1627/28 and 1634. While there are no mortality estimates available for the 16th

century epidemics, in 1607 74 victims were counted. From September 1627 until January 1628 plague claimed 284 victims and in 1634 760 deceased (including a large number of non-locals) were registered at the town's obituary columns. It is estimates that epidemics as well as the Thirty Years' War reduced the population by almost 80%.

Partially excavated between 2015 and 2016, the churchyard of St. Johannis revealed remains of more than 900 individuals[37,38]. On the lowest of eight burial levels, six multiple burials with four to eleven individuals were found. All together the mass graves revealed 47 individuals in up to four layers. Within the multiple burials, an equal proportion in the number of males and females can be observed; specifically, six individuals could be determined as male, eight as likely male, ten as female and further seven as likely female. 16 individuals are of indeterminate morphological sex (mostly subadults). Furthermore, in most of the multiple burials and one single burial (likely female, 11 to 14 years) the individuals were laying in or were covered with quicklime. In combination with the high percentage of subadult individuals (nearly 60%) and the absence of any evidence for trauma on the skeletons, an epidemic context was suspected by the excavators.

Imprints in the quicklime show that the dead were wrapped in shrouds and were then piled up in the burial pits. In contrast to the regular burials, which contained the typical furnishing (rosaries and multiple types of devotional objects) of catholic funerals of that period, just a few non-datable finds came from these epidemic burials.

Three individuals tested qPCR positive for *Y. pestis*. These were LBG002 (arch. ID 460, presumably male juvenile) was buried in a 11-fold burial, LBG005 (arch. ID 572, presumably female 6-12 years) from a single burial with quicklime, and LBG007 (arch. ID 598, presumably a female juvenile) buried in a quintuple burial with quicklime.

Nägeligasse, Stans, Switzerland

The partial excavation of the churchyard of "St. Peter und Paul" took place between 2015 and 2016. The cemetery was in use from the 8th century until the 1850s AD. Besides 122 single burials, the excavation revealed three multiple burials with a minimum number of four (pos. 369 – multiple burial 3), 16 (pos. 230 – multiple burial 1) and 26 individuals (pos. 232 – multiple burial 2) respectively[39,40]. The multiple burial pits were located at the margin of the cemetery, with individuals being buried in supine extended position mainly in a west-east orientation. The excavators were able to reconstruct that the individuals were likely buried simultaneously within a large pit, including some burial planks. Metal hooks indicate that the individual were buried in clothing. The excavation was accompanied by physical anthropologists that sampled the individuals on site (1 tooth per individual). A publication about the multiple burials anthropological analysis, including radiocarbon data is currently in preparation (Somers et al., in prep.)

Historical sources report the presence of plague during a minimum of seven succeeding outbreaks between 1493 and 1630 in Stans[39]. However, none of the recorded outbreaks could be assigned specifically to the multiple burials. The present molecular results suggest that individuals from all three multiple burials were positive for the *Y. pestis* bacterium. These were: STN002 (arch. ID 71), STN004 (arch. ID 91), STN005 (arch. ID 92), STN007 (arch. ID 97/251), STN008 (arch. ID 98/250), STN011 (arch. ID 102) and STN012 (arch. ID 103) from multiple burial 1; STN013 (arch. ID 85), STN014 (arch. ID 104), STN015 (arch. ID 105), STN016 (arch. ID 106), STN018 (arch. ID 121), STN019 (arch. ID 123), STN020 (arch. ID 124), STN021 (arch. ID 125) and STN026 (arch. ID 134) from multiple burial 2; as well as STN031 (arch. ID 165) and STN032 (arch. ID 167) from multiple burial 3.

Domlinden 12, Brandenburg an der Havel, Germany

During a survey in 2011, the burial of three individuals was found on the "Dominsel", the centre of the old town of Brandenburg, situated between two streams of the Havel. It is noteworthy that no context of a contemporary cemetery was found[41]. Instead, it appears that the individuals were buried in the backyard of a bourgeois house. One of the individuals was buried with a clay pipe bowl with the initials of the Dutch manufacturer Samuel Collier, setting the

terminus post quem between 1630 and 1640. All three individuals were morphologically identified as adult males. Moreover, one of them showed a healed impression fracture of the skull, indicative of interpersonal violence. Due to the poor health status and the hasty burial that did not follow the Christian burial practices, the individuals have been characterized as foreigners with low social status. Indeed, the oxygen and strontium isotope data show a non-local signature consistent with Scandinavia or the Baltic region. Therefore, it was hypothesized that the individuals were foreign soldiers housed with civilians during the occupation of the city by Swedish troops between 1626 and 1631 as part of the Thirty Years' War (1618 - 1648 AD). Plague waves have been reported for Brandenburg a. d. H. in 1625-1627 and 1631, supporting the year 1631 for the burial.

Y. pestis DNA was identified before in these individuals based on PCR SNP typing[26,27]. Here, we reconstructed whole genomes from two of those individuals, namely BRA001 (arch. ID 1, late adult male) and BRA003 (arch. ID 3, 18-20 years old male).

Supplementary Figures & Tables

Supplementary figure 1 – The picture shows the triple burial no. 27, where individual LAI009 was unearthed, in the Laishevo III cemetery, Laishevo, Russia. Picture credits: Rafail' M. Fattahov

Supplementary figure 2 – Picture of double burial from the Toulouse archaeological site "16 rue des Trente Six Ponts" showing individuals SQ1352 (TRP002) and SQ1353. Picture credits: Archeodunum SAS, Gourvennec Michaël.

Supplementary figure 3 – The picture shows the mass grave identified in the New Churchyard cemetery in London, England, including all the BED individuals described in this study. Picture credits: Crossrail / courtesy of Museum of London Archaeology (MOLA).

Supplementary figure 4 – The burial of individual F.230 (NMS002) unearthed at the Augustinian Friary, as part of the New Museums site in Cambridge, England. Picture credits: Cambridge Archaeological Unit

Supplementary figure 5 – The picture indicates the double burial with individuals 451 (NAB002) and 452 (NAB003) unearthed at the Sankt Johans Freidhof in Nabburg, Germany. Picture credits: Mathias Hensch (Schauhütte archaeology)

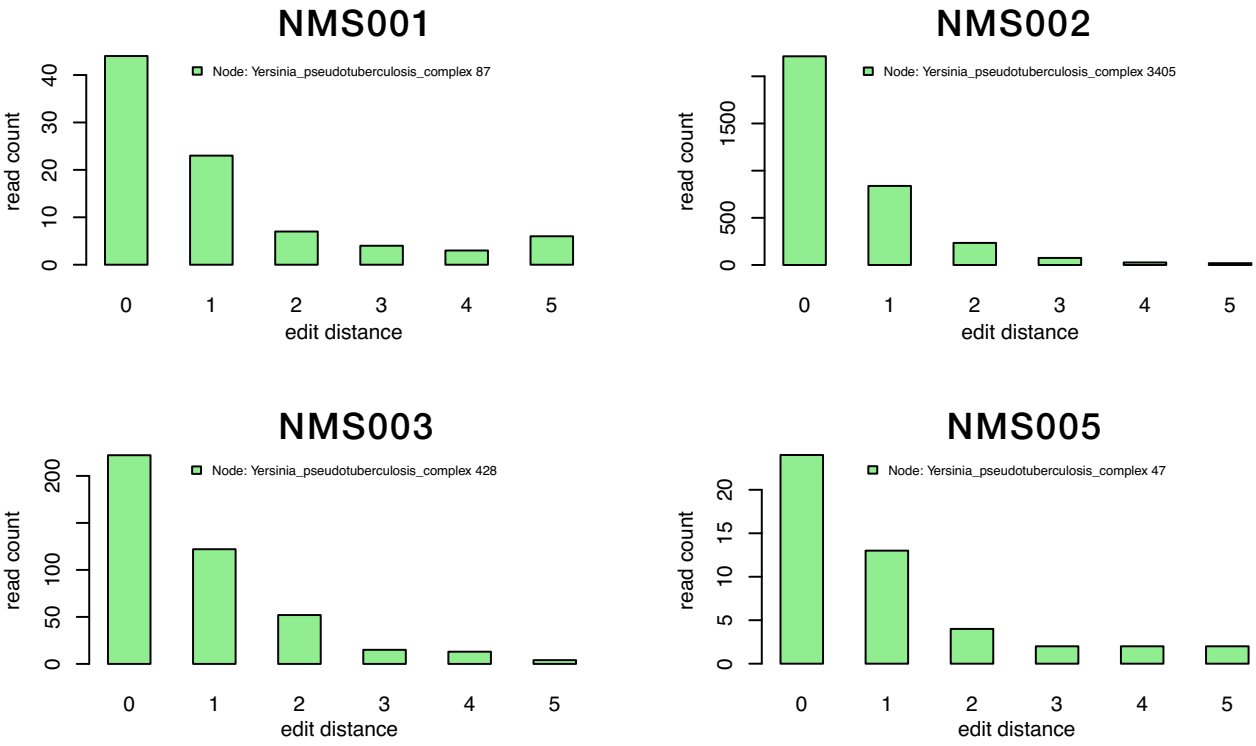
Supplementary figure 6 – The picture indicates the mass burial discovered at St. Leonhardi, Manching-Pichl, Germany, including individuals MAN008 and MAN015. Picture credits: Bavarian State Department of Monuments and Sites

Supplementary figure 7 – Picture of the triple burial (including individual STA001) discovered next to the former parish church of St. Benedikt, Starnberg, Germany. Picture credits: Bavarian State Department of Monuments and Sites

Supplementary figure 8 – Picture of a multiple burial (including individual LBG002) discovered at the St. Johannis churchyard in Landsberg am Lech, Germany. Picture credits: J. Schreiber (Dig it!)

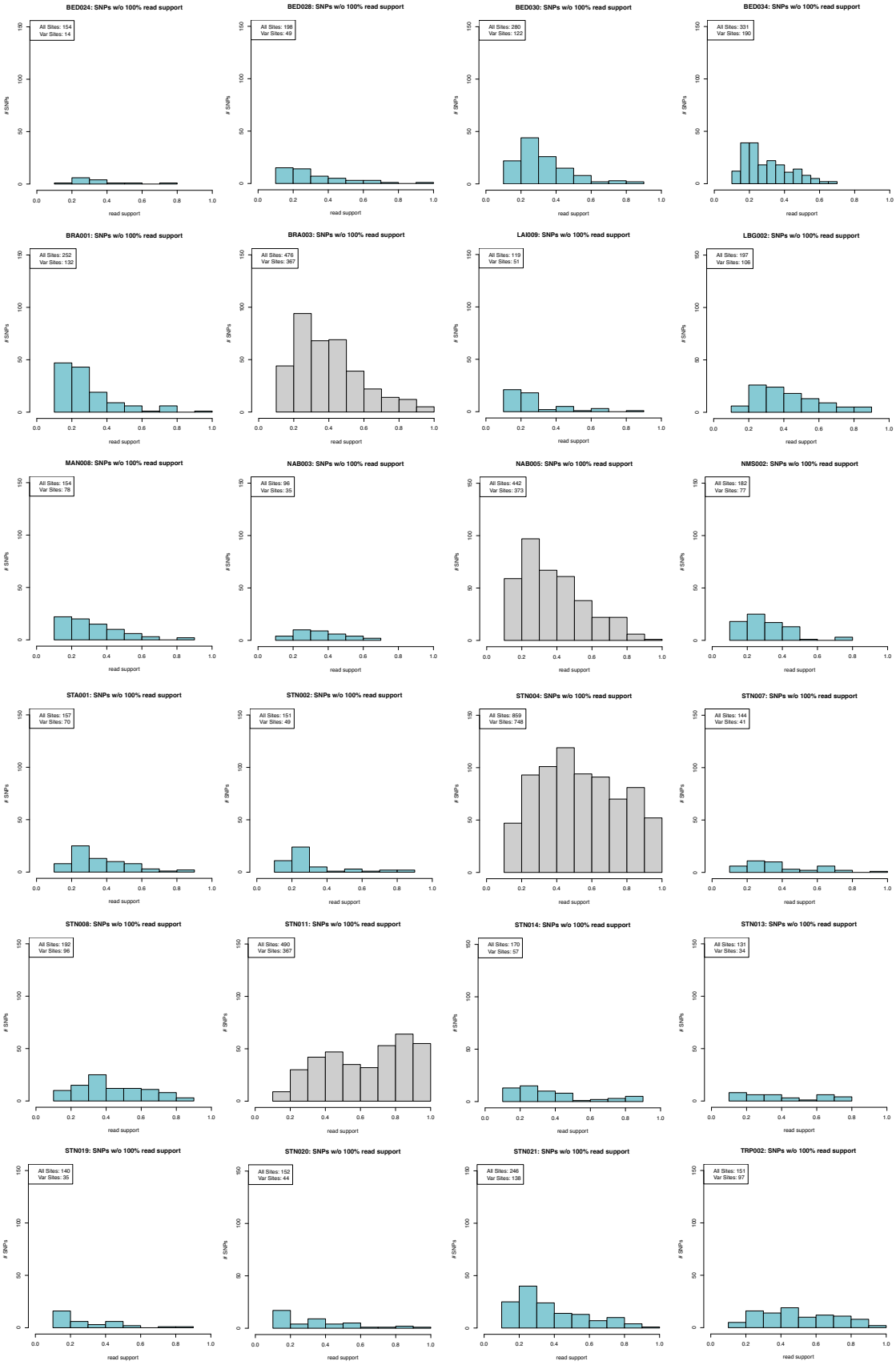
Supplementary figure 9 – Picture of multiple burial 1 (Pos. Nr. 230) unearthed in Nägeligasse, Stans NW, Switzerland, which included individuals STN002, STN004, STN005, STN007, STN008, STN011 and STN012. Picture credits: Department of Physical Anthropology, Institute for Forensic Medicine, University of Bern

Supplementary figure 10 – Picture of triple burial (including individuals BRA001 and BRA003) discovered in Brandenburg an der Havel, Germany, which has been attributed to the period of the Thirty Years War (1618–1648 AD). Picture credits: S. Dalitz (Untere Denkmalschutzbehörde der Stadt Brandenburg an der Havel).

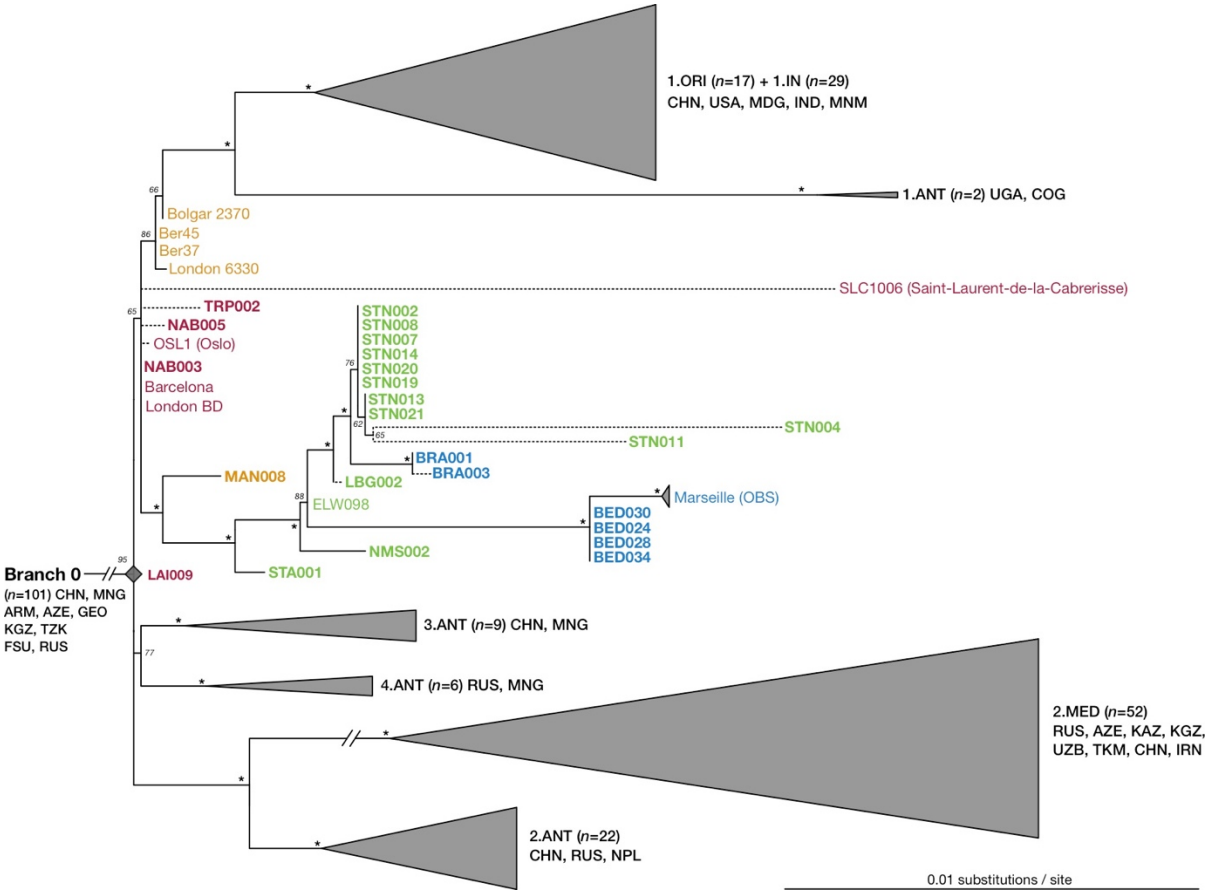


Supplementary figure 11 – Edit distance histograms of shotgun NGS reads calculated with HOPS[1] for all NMS (Augustinian friary) specimens that yielded putative *Y. pestis* matches. The edit distance is calculated for each dataset according to the closest matching reference genome within the *Y. pseudotuberculosis* complex.

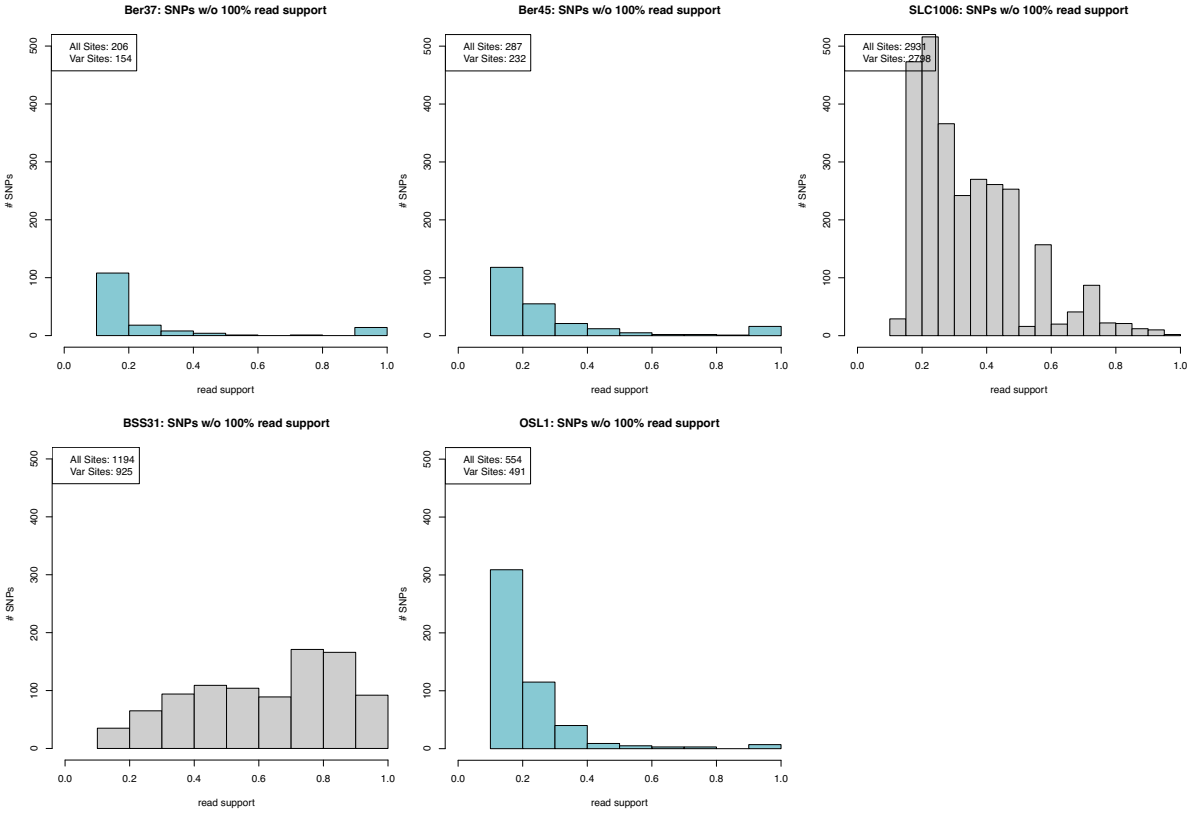
Supplementary figure 12 – Maximum likelihood phylogeny[2] (97% partial deletion) of all *Y. pestis* genomes used in this study. A total of 6,058 SNP positions were considered for the phylogeny. The tree is comprised of 233 modern isolates[3-9], 35 second pandemic isolates, one first pandemic isolate[10], one 2nd- to 3rd-century isolate[11], three Bronze Age isolates[12,13], and a *Y. pseudotuberculosis* isolate (IP32953)[14] that was used as out-group for rooting the tree. Bootstrap values of 95 or higher are shown. All early 14th-century genomes are shown in red, late 14th-century genomes are shown in orange, 15th–17th century genomes are shown in green and 17th- to18th-century genomes are shown in blue.



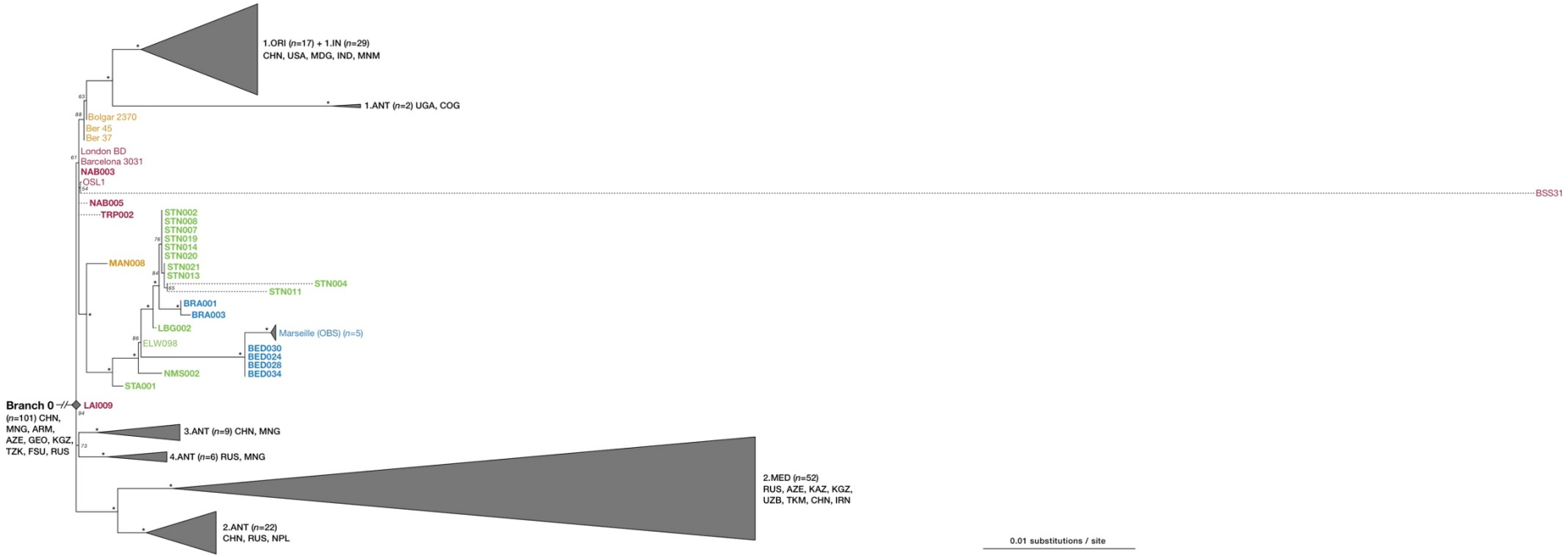
Supplementary figure 13 – Histograms showing the distribution of SNP allele frequencies in newly produced *Y. pestis* genomes with ≥5-fold coverage. The histograms were constructed using R version 3.4.1[15]. Histograms shown in grey represent genomes that were excluded from subsequent analyses.



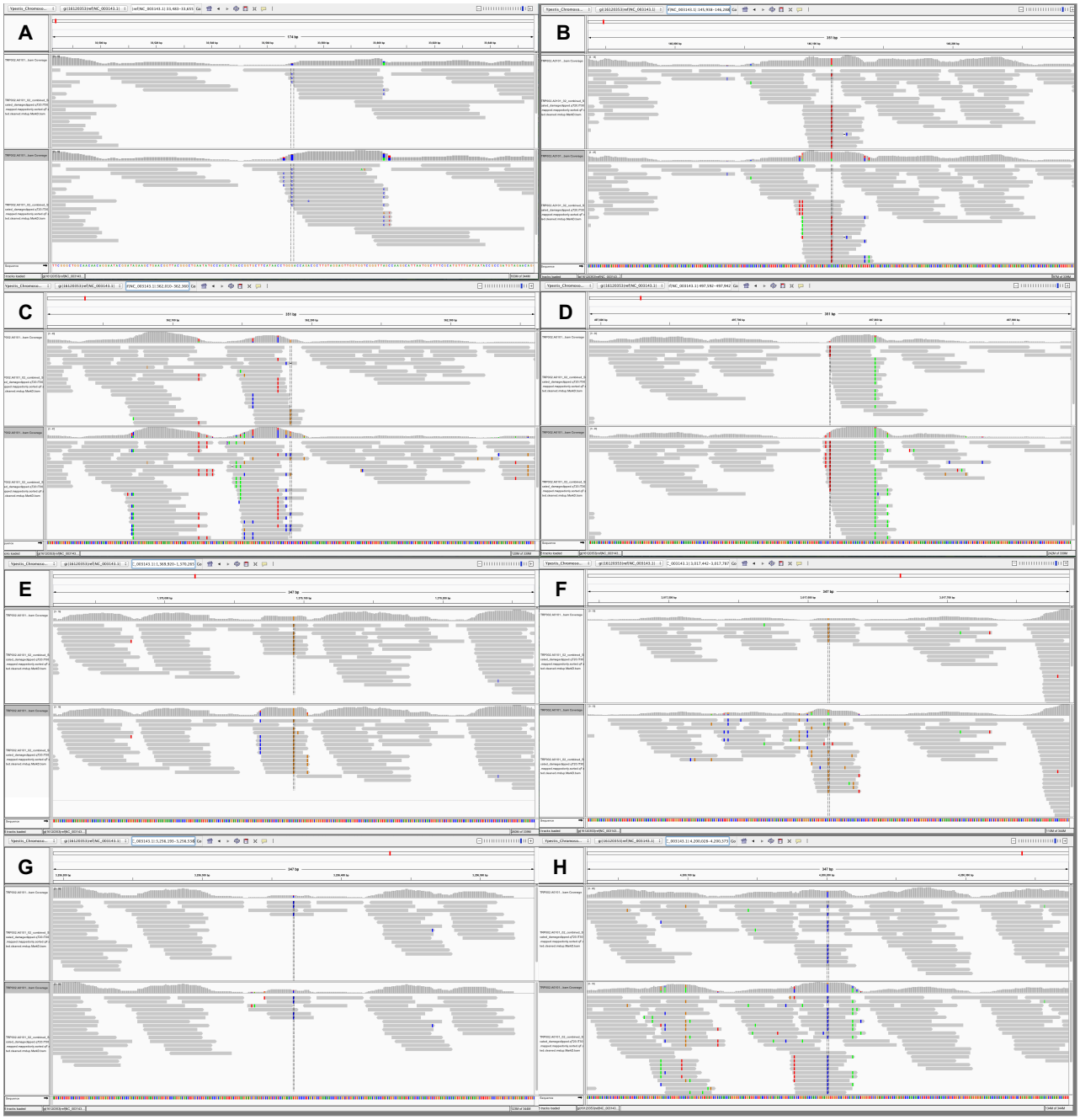
Supplementary figure 14 – A maximum likelihood phylogeny with 98% partial deletion was generated using RaxML[2]. The figure shows a graphical representation of Branches 14, and more specifically the phylogenetic positioning of all previously published and new second pandemic strains (14^{th}–18^{th} centuries). 5,736 SNP positions were used to generate the phylogeny. 1000 bootstrap iterations were carried out to determine node support. Sub-clades were collapsed to enhance tree clarity. Numbers in brackets indicate the number of strains contained in each collapsed branch. Isolates that showed evidence of false-positive SNPs to be influencing their branch lengths have their branches represented as dotted lines. 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), as well as unspecified regions of the Former Soviet Union (FSU).



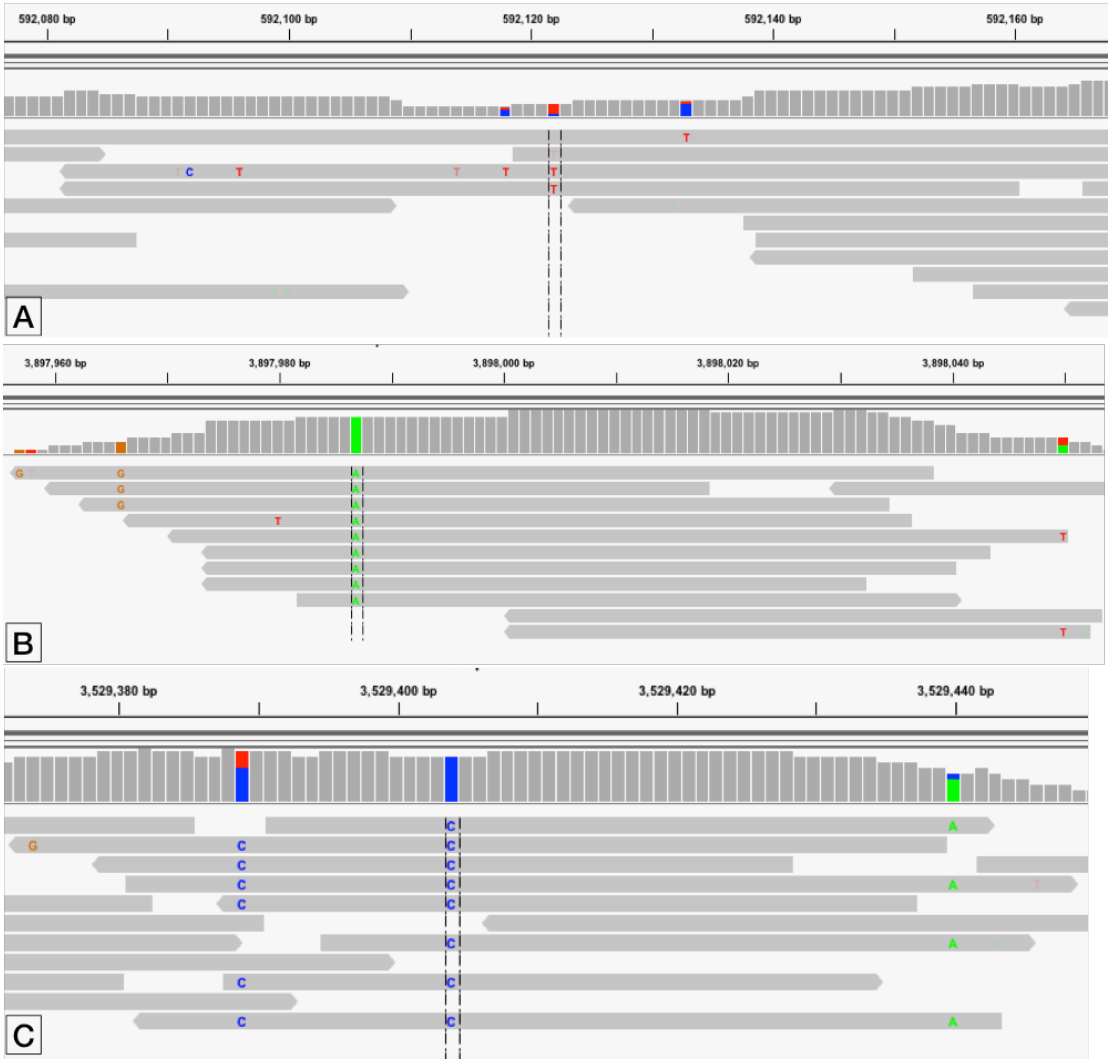
Supplementary figure 15 – Histograms showing the distribution of SNP allele frequencies in the 14th-century *Y. pestis* genomes from Namouchi et al[16]. Histograms shown in grey represent genomes that were excluded from further analysis due to high heterozygosity. The histograms were constructed using R version 3.4.1[15].



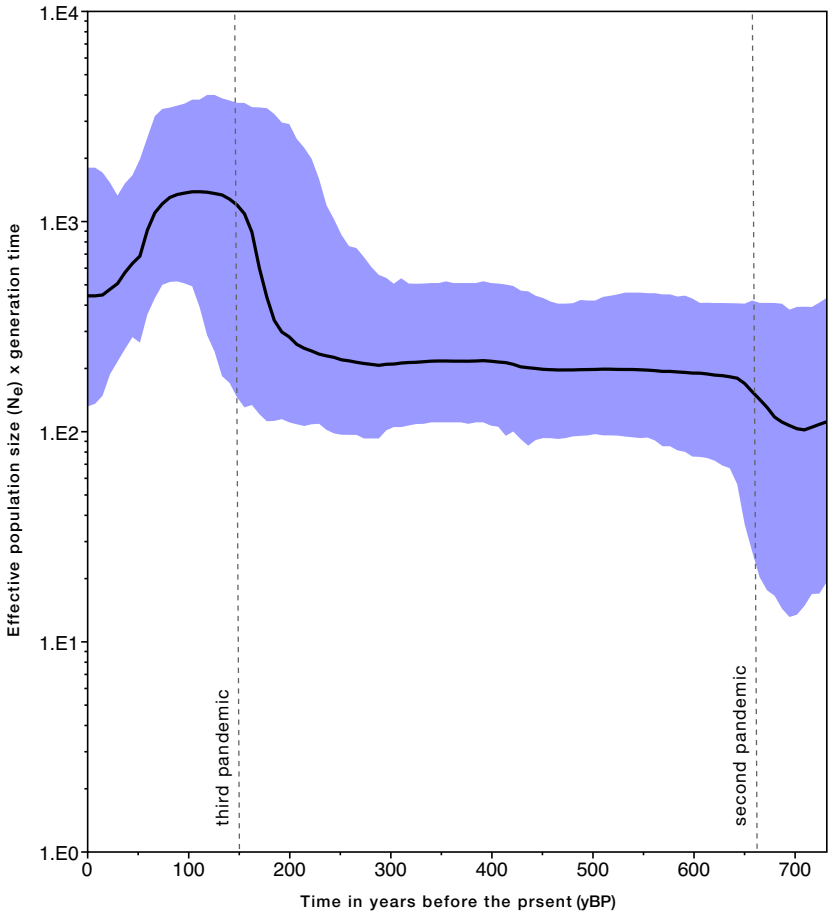
Supplementary figure 16 – A maximum likelihood phylogeny with 98% partial deletion was generated using RaxML[2] in order to determine the phylogenetic positioning of BSS31. Note that the low coverage of BSS31 (3-fold) appears to have affected the phylogenetic resolution and topologies in certain parts of the tree. For a more robust phylogeny (excluding BSS31), refer to Figure 2 in the main text. 5,867 SNP positions were used for generating the depicted tree. 1000 bootstrap iterations were carried out to determine node support. Sub-clades were collapsed to enhance tree clarity. Numbers in brackets indicate the number of strains comprising each collapsed branch. Isolates that showed evidence of environmental contamination to be influencing their SNP assignment have their branches represented by dotted lines. 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), as well as unspecified regions of the Former Soviet Union (FSU).



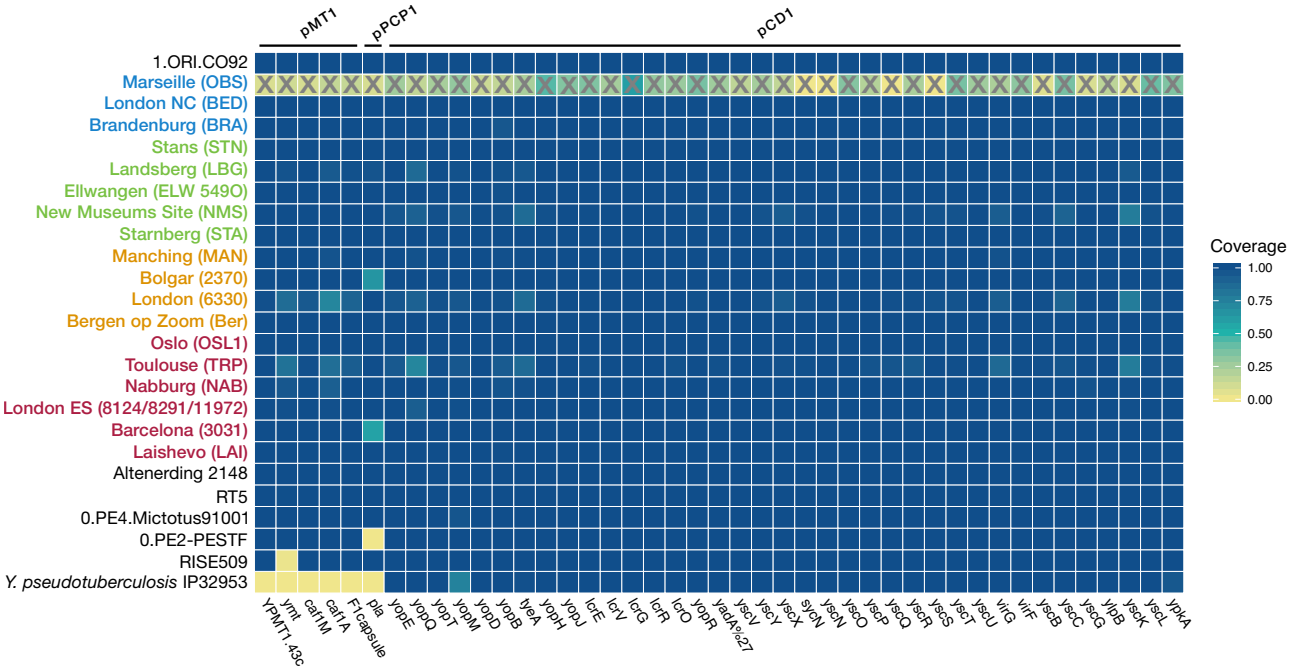
Supplementary figure 17 – Visual inspection of the eight private SNP positions identified in TRP002. The screenshots are made in IGV[17], with the cursor placed on the genomic position where each SNP was identified. BWA mapping was carried out using both stringent (A-H upper panels, -n 0.1) and lenient parameters (A-H lower panels, -n 0.01).



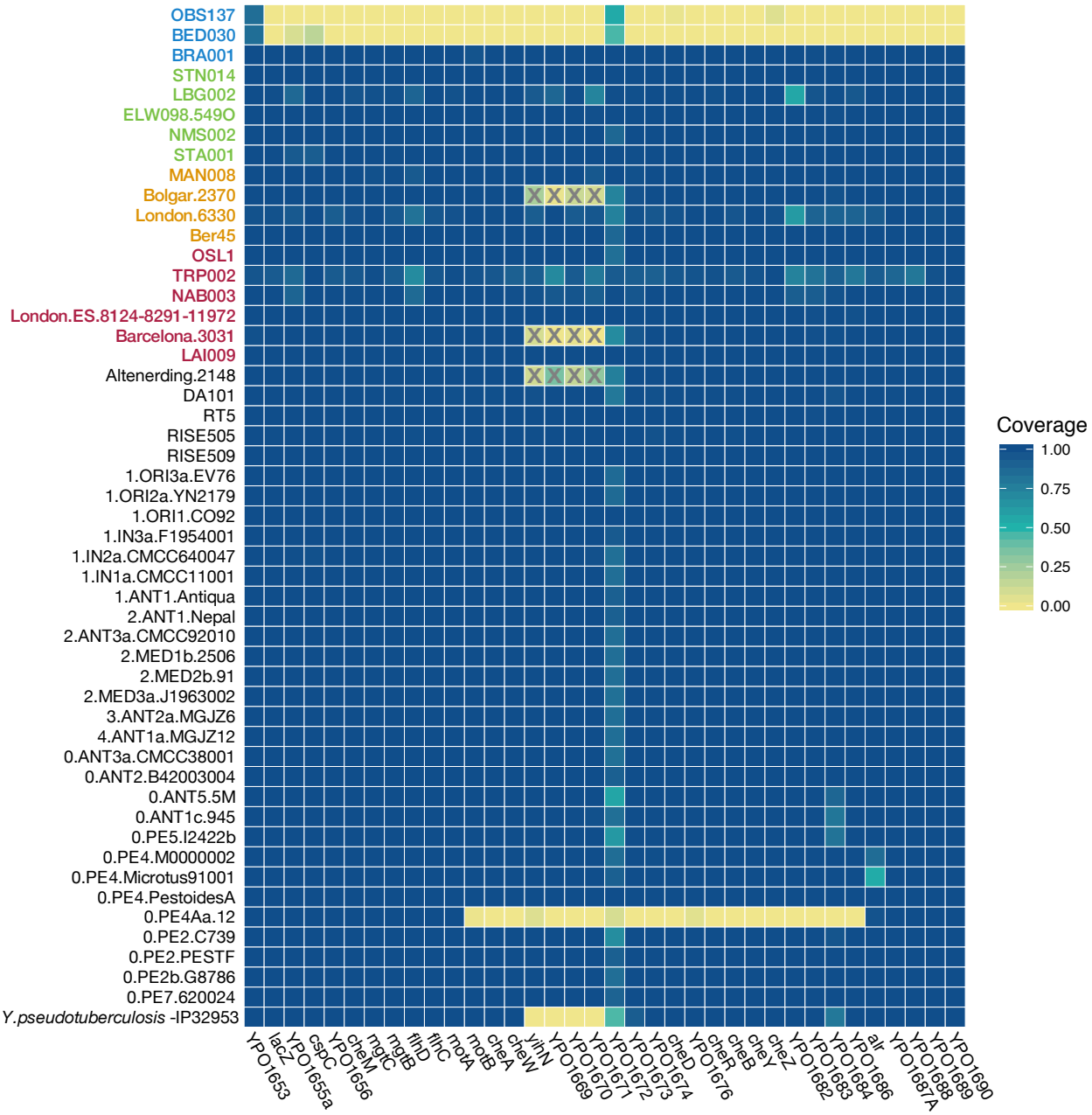
Supplementary figure 18 – Visual inspection of private SNPs in the previously published Black Death-associated *Y. pestis* genomes OSL1 and BSS31[16]. SNPs were inspected using the program IGV[17] (screenshots shown). Panel (A): C592122T indicates a private SNP identified in this study during re-analysis of OSL1, whereas (B): T3897987A and (C): T3529404C indicate previously identified private SNPs in BSS31[16].



Supplementary figure 19 – Bayesian coalescent skyline. The figure is a depiction of effective population size *(N_e)* changes across the history of *Y. pestis* Branch 1 (*n*=80), as estimated by the coalescent skyline model[18] implemented in BEAST v1.8[19]. The plot was produced in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/), after subsampling the MCMC states at lower frequency (subsampling every 100,000 states) using LogCombiner[19]. Dotted lines represent the proposed initiations of the second (1346 AD) and third (1855 AD) plague pandemics. The divergence date for Branch 1 was estimated to 731 yBP (95% HPD: 672 - 823).



Supplementary figure 20 – Analysis of coverage across virulence-associated and evolutionary determinant genes located on the pCD1, pMT1 and pPCP1 plasmids of *Y. pestis*. The presence/absence of previously identified genes[20,21] was assessed for the newly reconstructed second pandemic genomes in relation to previously published representatives of ancient and modern strains. The heatmap was constructed using the ggplot2[22] package in R version 3.4.1[15]. Genomes are arranged from top to bottom according to their inferred divergence ages (from youngest to oldest), and all second pandemic strains are shown in red, orange, green or blue based on their age (see Figure 1). A lack of coverage across the plasmids in the Marseille OBS isolates is apparent since those regions were not used as part of the probe set for enrichment of these loci and are marked here with "X"[23]. Instead, missing regions in all other isolates represent real gene absences.



Supplementary figure 21 – Analysis of coverage across genes identified as absent in 17th–18th century second pandemic genomes BED and OBS. The presence or absence of genes located on the *Y. pestis* chromosome in representative ancient and modern genomes is shown. Specifically, second pandemic isolates from all studied archaeological sites (one representative chosen when more than one identical genome was recovered from a site), one published first pandemic genome from Germany (Altenerding 2148), one 2nd- to 3rd-century genome from the Tian Shan region (DA101) and three Bronze Age genomes from Russia (RT5, RISE505, RISE509) are shown for comparison. In addition, modern representatives from all phylogenetic clades are shown. The heatmap was constructed using the ggplot2[22] package in R version 3.4.1[15]. Genomes are arranged from top to bottom according to their inferred divergence ages (from youngest to oldest), and all second pandemic strains are shown in red, orange, green or blue based on their age (see Figure 1). A lack of coverage across certain *Y. pestis*-specific regions in Altenerding 2148, Barcelona 3031, and Bolgar 2370 are marked here with "X" since they were not used as part of the probe set for their enrichment[10,24]. Instead, missing regions in all other isolates represent real gene absences.

Supplementary table 1 – Sample description and quantification of *pla* through qPCR

Sample Name	Site	Country	Archaeological IDs	Archaeological date	Cal 2-sigma (95.4%) radiocarbon date (cal AD)	*pla* qPCR quantification (copies/µl)
BED030.A0102	New Churchyard "Bedlam", London	United Kingdom	8127	1600-1700	1560–1635 (combined date)	62.60
BED028.A0102	New Churchyard "Bedlam", London	United Kingdom	8103	1600-1700	1560–1635 (combined date)	15.70
BED034.A0102	New Churchyard "Bedlam", London	United Kingdom	8198	1600-1700	1560–1635 (combined date)	5.90
BED024.A0102	New Churchyard "Bedlam", London	United Kingdom	8052	1600-1700	1560–1635 (combined date)	1.38
BED038.A0102	New Churchyard "Bedlam", London	United Kingdom	8216	1600-1700	1560–1635 (combined date)	0.65
BRA001.A0101	Domlinden 12, Brandenburg	Germany	1*	1618-1648	N/A	1.64
BRA003.A0101	Domlinden 12, Brandenburg	Germany	3*	1618-1648	N/A	1.02
LAI009.A0101	Laishevo III, Laishevo	Russia	RT88 (burial 27B)	1300-1400	N/A	140.20
LAI010.A0101	Laishevo III, Laishevo	Russia	RT89 (burial 10A)	1300-1400	N/A	1.14
LBG002.A0101	Kirchhof St. Johannis, Landsberg	Germany	Bef. 460	N/A	1455-1632	1.52
LBG005.A0101	Kirchhof St. Johannis, Landsberg	Germany	Bef. 572	N/A	N/A	0.02
LBG007.A0101	Kirchhof St. Johannis, Landsberg	Germany	Bef. 598	N/A	N/A	0.32
MAN008.B0101	St. Leonhardi, Manching-Pichl	Germany	MPS03-I*	N/A	1283-1390	171.80
MAN015.A0101	St. Leonhardi, Manching-Pichl	Germany	MP56-X*	N/A	N/A	58.58
NAB005.A/B0101	"Sankt Johans Freidhof" Nabburg	Germany	471	N/A	1298-1398	0.043/2.174
NAB003.A/B0101	"Sankt Johans Freidhof" Nabburg	Germany	452	N/A	1292-1392	19.5/17.18
NAB004.A/B0101	"Sankt Johans Freidhof" Nabburg	Germany	457	N/A	1317-1420	0.1069/0.1069
NAB002.A0101	"Sankt Johans Freidhof" Nabburg	Germany	451	N/A	N/A	0.5980
STA001.A0101	Possenhofener Str. 3, Starnberg	Germany	207	1433-1523	1420-1630	3.60
STN014.A0101	Nägeligasse, Stans	Switzerland	Grave 104 (mult. bur. 2)	N/A	1485-1635 (combined date)	129.60
STN020.A0101	Nägeligasse, Stans	Switzerland	Grave 124 (mult. bur. 2)	N/A	1485-1635 (combined date)	66.93
STN021.A0101	Nägeligasse, Stans	Switzerland	Grave 125 (mult. bur. 2)	N/A	1485-1635 (combined date)	11.71
STN019.A0101	Nägeligasse, Stans	Switzerland	Grave 123 (mult. bur. 2)	N/A	1485-1635 (combined date)	34.75
STN007.A0101	Nägeligasse, Stans	Switzerland	Grave 97/POS.251 (mult. bur. 1)	N/A	1485-1635 (combined date)	26.47
STN002.A0101	Nägeligasse, Stans	Switzerland	Grave 71 (mult. bur. 1)	N/A	1485-1635 (combined date)	10.65
STN008.A0101	Nägeligasse, Stans	Switzerland	Grave 98/POS.250 (mult. bur. 1)	N/A	1485-1635 (combined date)	15.87
STN013.A0101	Nägeligasse, Stans	Switzerland	Grave 85 (mult. bur. 1)	N/A	1485-1635 (combined date)	3.07
STN011.A0101	Nägeligasse, Stans	Switzerland	Grave 102 (mult. bur. 2)	N/A	1485-1635 (combined date)	2.07
STN004.A0101	Nägeligasse, Stans	Switzerland	Grave 91 (mult. bur. 1)	N/A	1485-1635 (combined date)	1.24
STN032.A0101	Nägeligasse, Stans	Switzerland	Grave 167 (mult. bur. 3)	N/A	1485-1635 (combined date)	1.89
STN031.A0101	Nägeligasse, Stans	Switzerland	Grave 165 (mult. bur. 3)	N/A	1485-1635 (combined date)	12.52
STN005.A0101	Nägeligasse, Stans	Switzerland	Grave 92 (mult. bur. 1)	N/A	1485-1635 (combined date)	0.89
STN018.A0101	Nägeligasse, Stans	Switzerland	Grave 121 (mult. bur. 2)	N/A	1485-1635 (combined date)	0.17
STN012.A0101	Nägeligasse, Stans	Switzerland	Grave 103 (mult. bur. 1)	N/A	1485-1635 (combined date)	0.16
STN015.A0101	Nägeligasse, Stans	Switzerland	Grave 105 (mult. bur. 2)	N/A	1485-1635 (combined date)	0.13
STN026.A0101	Nägeligasse, Stans	Switzerland	Grave 134 (mult. bur. 2)	N/A	1485-1635 (combined date)	0.28
STN016.A0101	Nägeligasse, Stans	Switzerland	Grave 106 (mult. bur. 2)	N/A	1485-1635 (combined date)	0.40
TRP002.A0101	Trente-Six Ponts str. 16, Toulouse	France	Ind. 1352	1347-1350	1288-1394	0.02

Supplementary table 2 – Summarised HOPS statistics for the analysed samples from the Augustinian Friary in Cambridge (NMS).

Sample ID	Archaeo-logical ID	Total assigned reads in MALT	Reads assigned to the *Y. pseudotuberculosis* complex	Percentage C-to-T 5' end	Percentage G-to-A 3' end	Average fragment length
NMS001	F. 190	1,574,633	110	3%	13%	49
NMS002	F. 230	2,572,436	3,795	15.1%	13.5%	49
NMS003	F. 355	589,608	434	20.3%	12.1%	51
NMS005	F. 310	860,459	50	0%	5.3%	50

Supplementary table 3 – Sequencing statistics after whole-genome *Y. pestis* capture for all samples that yielded coverage ≥ 1-fold

Sample IDs	Archaeological IDs	Pre-processed reads	Mapping reads	Unique mapping reads after quality filtering	Endogenous DNA (%)	Cluster factor	Mean coverage	Std. dev. coverage	Coverage ≥ 5X (%)	Average frag. length	Median frag. length	Percentage GC (%)
BED030.A0102	8127	19,242,747	7,335,361	3,624,482	36.2	1.9	80.1	35.1	93.6	102.9	101.0	48.5
BED028.A0102	8103	27,963,243	6,830,578	2,665,238	22.2	2.3	37.2	23.7	91.4	65.0	66.0	49.0
BED034.A0102	8198	29,358,235	3,489,557	1,371,698	10.5	2.2	18.3	10.3	89.1	62.2	62.0	49.2
BED024.A0102	8052	11,064,780	2,209,748	1,000,524	18.1	2.0	12.6	7.7	84.7	58.5	56.0	49.1
BED038.A0102	8216	31,240,177	1,974,588	458,397	5.3	3.6	4.9	4.5	44.3	50.0	49.0	47.6
BRA001.A0101	1*	20,535,183	5,297,674	2,387,557	23.2	2.0	23.8	12.7	92.0	46.4	44.0	47.5
BRA003.A0101	3*	43,129,525	2,568,078	849,170	3.4	1.7	9.1	5.5	79.2	49.8	46.0	47.4
LAI009.A0101	RT88 (burial 27B)	23,417,187	6,144,041	2,549,926	23.9	2.2	28.4	15.9	92.1	51.8	50.0	48.0
LAI010.A0101	RT89 (burial 10A)	64,934,571	1,269,665	133,256	1.0	4.8	1.3	1.4	2.8	43.9	42.0	48.4
LBG002.A0101	Bef. 460	5,696,055	1,793,631	621,713	27.9	2.6	7.2	5.3	66.4	54.2	53.0	49.9
MAN008.B0101	MPS03-I*	8,084,687	3,936,101	1,974,399	44.9	1.8	25.8	16.1	88.7	60.8	62.0	50.8
MAN015.A0101	MP56-X*	1,299,105	178,023	121,546	10.7	1.1	1.6	1.7	6.7	62.0	64.0	49.9
NAB005.A/B0101	471	66,111,306	2,757,864	786,575	2.9	2.4	8.3	5.6	73.3	49.1	47.0	47.5
NAB003.B0101	452	6,034,650	2,230,536	684,029	33.5	3.0	8.1	5.7	70.2	54.8	53.0	49.7
NAB004.A0101	457	79,055,317	1,491,516	202,056	0.7	2.8	1.9	1.8	8.2	42.8	41.0	46.6
NMS002.A0101	F. 230 / 507	68,353,310	2,022,552	855,185	2.3	1.8	12.5	5.5	91.7	67.9	72	47.3
NMS003.A0101	F. 355 / 517	22,641,646	2,550,272	289,755	9.8	7.7	4.2	2.6	42.6	67.7	66	47.1
STA001.A0101	207	28,712,935	3,283,300	1,110,049	9.9	2.6	11.7	7.2	84.3	49.0	46.0	44.7
STN014.A0101	Grave 104 (mult. bur.2)	15,736,877	8,069,865	3,822,030	48.2	2.0	55.3	26.6	93.0	67.3	74.0	48.9
STN020.A0101	Grave 124 (mult. bur.2)	7,368,937	3,502,272	2,020,769	44.3	1.6	28.2	16.9	90.3	64.8	70.0	48.5
STN021.A0101	Grave 125 (mult. bur.2)	8,852,058	3,441,583	1,588,442	35.1	2.0	21.7	13.8	88.6	63.7	66.0	48.5
STN019.A0101	Grave 123 (mult. bur.2)	5,534,501	2,114,414	1,325,076	35.3	1.5	18.7	11.8	87.1	65.8	72.0	49.1
STN007.A0101	Grave 97/POS.251 (mult. bur.1)	6,824,093	2,538,685	1,293,507	32.8	1.7	18.0	11.5	86.7	64.8	70.0	49.4
STN002.A0101	Grave 71 (mult. bur.1)	6,912,906	2,089,426	935,795	27.6	2.0	12.7	8.1	83.3	63.0	66.0	48.4
STN008.A0101	Grave 98/POS.250 (mult bur.1)	4,372,116	1,469,871	875,153	30.2	1.5	11.7	8.6	77.7	62.5	65.0	50.1
STN013.A0101	Grave 85 (mult. bur.2)	5,801,670	1,582,429	714,482	24.5	2.0	9.2	6.6	73.8	59.9	60.0	49.0
STN011.A0101	Grave 102 (mult. bur.1)	9,187,148	2,938,862	671,509	24.4	3.3	8.1	6.6	69.1	56.2	54.0	48.3
STN004.A0101	Grave 91 (mult. bur.1)	7,045,866	1,913,895	531,605	22.6	3.0	6.2	5.5	57.2	54.6	53.0	49.5
STN032.A0101	Grave 167 (mult. bur.3)	4,130,111	566,869	237,195	9.0	1.6	3.4	2.9	29.5	65.8	73.0	49.3
STN031.A0101	Grave 165 (mult. bur.3)	3,217,122	337,111	127,950	6.7	1.7	1.8	1.8	8.1	64.3	68.0	49.7
STN005.A0101	Grave 92 (mult. bur.1)	3,678,388	299,437	108,298	6.0	2.0	1.3	1.5	3.7	55.4	54.0	48.5
STN018.A0101	Grave 121 (mult. bur.2)	5,528,516	326,360	94,481	4.0	2.3	1.2	1.5	3.5	58.5	58.0	48.4
STN012.A0101	Grave 103 (mult. bur.1)	5,680,142	432,404	86,278	4.9	3.2	1.1	1.4	2.6	61.3	63.0	48.1
TRP002.A0101/02	Ind. 1352	22,619,908	5,173,193	632,303	19.8	7.1	5.9	5.4	50.9	43.2	42.0	48.7
ELW098.A0101	549_O**	48,518,869	5,078,186	1,435,225	9.0	3.0	14.2	7.8	89.1	46.0	44.0	46.6

* Samples from sites that were analysed in previous aDNA studies, where Y. pestis was detected by PCR[25-28]

Supplementary table 4 – Genes affected by the 49 kb deletion identified among BED and OBS isolates (coordinates on CO92 reference genome: 1,879,467 – 1,928,869)

Gene name	Start position (CO92)	End position (CO92)	Function
YPO1690	1928171	1928404	hypothetical protein
YPO1689	1926760	1928016	lipoprotein
YPO1688	1926119	1926367	hypothetical protein
YPO1687A	1925346	1925540	hypothetical protein
alr	1923715	1924944	alanine racemase: converts L-alanine to D-alanine which is used in cell wall bi-osynthesis
YPO1686	1922143	1923096	hypothetical protein
YPO1684	1919412	1921622	surface protein
YPO1683	1918291	1919055	N-acetylmuramoyl-L-alanine amidase
YPO1682	1918040	1918159	hypothetical protein
cheZ	1917082	1917726	chemotaxis regulator
cheY	1916683	1917072	chemotaxis regulatory protein
cheB	1915533	1916582	chemotaxis-specific methylesterase
cheR	1914661	1915533	chemotaxis methyltransferase
YPO1676	1912878	1914488	pseudogene
cheD	1911021	1912694	methyl-accepting chemotaxis protein
YPO1674	1910079	1910780	hypothetical protein
YPO1673	1907134	1910010	hypothetical protein
YPO1672	1902800	1906606	hypothetical protein
YPO1671	1901648	1902325	DNA-binding protein
YPO1670	1900484	1901470	hypothetical protein
YPO1669	1900110	1900487	hypothetical protein
yihN	1898862	1900097	hypothetical protein
cheW	1898073	1898570	purine-binding chemotaxis protein
cheA	1895739	1897910	chemotaxis protein
motB	1894445	1895728	flagellar motor protein
motA	1893561	1894448	flagellar motor protein
flhC	1892772	1893356	transcriptional activator
flhD	1892411	1892771	pseudogene
mgtB	1888533	1891232	Mg^{2+} transport ATPase protein B
mgtC	1887422	1888132	Mg^{2+} transport ATPase protein C
cheM	1884574	1886275	pseudogene
YPO1656	1884200	1884391	hypothetical protein
cspC	1883773	1883985	cold shock protein
YPO1655a	1883592	1884034	ncRNA-cspA thermoregulator
lacZ	1880050	1883232	beta-D-galactosidase
YPO1653	1878633	1879643	zinc-binding dehydrogenase

Supplementary Data 1 – Variant genomic positions identified across previously published and new second pandemic *Y. pestis* genomes sequenced to date

Position	Reference (CO92)	LAI009	Barcelona 3031	TRP002	NAB003	London ES 8124/8291/11972	OSL1	Ber37	Ber45	London 6330	Bolgar 2370	MAN008	STA001	NMS002	ELW098/549_O	LBG002	STN002	STN007	STN013	STN014	STN020	STN021	STN008	STN019	BRA001	BED024	BED028	BED030	BED034	OBS124	OBS107	OBS116	OBS137	OBS110
19828	C	.	.	N	.	.	.	.	.	.	.	.	.	A	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
20767	G	N	N	N	N	N	T	N	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
22699	C	.	.	N	N	.	.	.	.	N	.	N	.	T	.	N	N	N	N	N	N	N	N	N	N	N	.	.	N	N	.	.	.	.
29368	G	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
33569	G	.	.	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
79499	C	.	.	N	.	.	.	.	.	.	.	.	T	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
100263	C	.	N	.	N	N	N	N	A	N	N	N	N	N	N	N	N	N	N	N	N	.	N	.	N	N	N	N	.	N	.	N	N	N
100383	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
131125	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
146113	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
164513	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	T	N	N	N	T	T	T	T	T	.	.	.	.	.	.	.	.	.
169412	C	.	.	N	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
173032	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
182517	G	.	.	.	.	.	.	.	.	N	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
185051	A	.	.	N	.	.	.	.	.	.	.	.	.	.	.	T	N	N	T	N	N	N	N	N	N	.	.	.	.	.	.	.	.	.
186060	C	.	.	.	.	.	.	.	.	.	.	.	.	T	T	N	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
189227	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
189659	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	N	.	.	.
190040	C	.	N	N	N	.	.	.	N	.	.	.	N	.	N	N	N	.	N	.	.	.	N	N	.	N	.	N	.	N	N	N	N	A
190041	T	.	N	N	N	.	N	.	G	.	.	.	N	.	N	N	N	.	N	N	N	N	N	N	N	N	N	N	N	N	G	N	N	G
190049	G	N	A	N	N	N	N	.	A	N	N	N	N	.	N	N	N	.	N	.	N	N	N	N	.	N	N	.	N	N	N	N	N	N
200723	C	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
214613	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
217009	G	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
225436	T	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	A	A	A	N	N	N	N	A	A	A	N	A	A	A	A	A
226722	C	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T
228816	T	.	N	.	.	.	.	.	.	N	.	N	.	.	N	.	N	.	N	.	N	G	N	N	N	.	.	.	.	.	.	.	.	.
232790	C	.	.	N	.	.	.	.	.	N	.	.	.	.	T	N	.	.	N	.	N	N	N	.	.	.	.	.	.	.	N	.	.	.
234892	T	.	.	N	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.
274423	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
300041	C	.	.	N	.	.	.	.	.	.	.	.	.	N	.	N	.	.	.	.	.	.	.	.	.	N	N	N	N	N	N	N	T	N
300043	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	N	A	N	N	N	N	N	.	.
309535	G	.	.	.	.	.	.	.	.	N	.	.	.	.	.	N	A	A	N	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
362185	T	.	N	G	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.
400143	G	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
460999	G	.	.	N	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	N	.	T	.	.	.	.	.	.	.	.	.
477107	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
480773	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
482327	G	.	.	N	N	.	.	.	.	.	.	.	.	.	.	N	.	N	N	.	.	.	N	.	.	N	T	T	T	T	T	T	T	T
497767	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
528975	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	.	.	N	N	C	N	N
548198	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
592122	C	.	N	.	.	.	T	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
699494	A	G	G	G	G	G	G	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
700976	A	N	.	N	.	.	.	N	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.
862385	T	.	.	N	.	.	.	.	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
867712	C	.	.	.	N	.	.	.	.	N	.	.	.	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
868549	G	.	.	.	.	.	N	.	N	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C
869820	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	G	G	G	G	G
899158	C	.	N	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	N	.	.	T	T	T	T	T	T	T	T	T
944177	C	N	N	N	N	N	.	N	.	N	N	N	N	N	N	N	N	G	N	N	N	G	N	N	N	N	N	N	N	N	N	N	N	N
944178	G	N	N	N	N	N	.	.	.	N	N	N	N	N	N	N	N	N	N	N	N	A	N	N	N	N	N	N	N	N	N	N	N	.
951295	C	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
961795	C	.	.	.	.	.	.	.	.	N	.	.	.	T	T	N	T	T	T	T	T	T	N	T	T	T	T	T	T	T	T	T	T	N
965281	C	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A

Position	Reference (CO92)	LAI009	Barcelona 3031	TRP002	NAB003	London ES 8124/8291/11972	OSL1	Ber37	Ber45	London 6330	Bolgar 2370	MAN008	STA001	NMS002	ELW098/549_O	LBG002	STN002	STN007	STN013	STN014	STN020	STN021	STN008	STN019	BRA001	BED024	BED028	BED030	BED034	OBS124	OBS107	OBS116	OBS137	OBS110
983623	C	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
997286	C	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	N	.	.	.
1061182	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
1159539	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
1168951	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
1189479	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
1201122	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1230108	T	.	N	N	C	N	N	.	.	N	.	.	N	.	N	N	.	.	N	.	.	.	N	N	.	.	N	.	N	.	N	.	.	N
1232222	C	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
1254157	C	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
1304271	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1308719	G	.	.	N	.	.	.	.	.	N	.	A	A	A	A	N	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
1364191	C	.	.	.	.	.	.	.	.	.	.	.	.	A	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1370093	T	.	.	G	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1378105	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
1384546	A	.	.	N	N	.	.	.	.	.	.	.	N	N	N	N	.	N	.	.	.	N	N	.	G	N	.	.	.	.	.	.	.	.
1439084	T	.	.	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	N	.	.	.	N	N	N	A	.	N	N	N
1439085	A	.	.	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	N	.	.	.	.	N	N	C	.	N	N	N
1440851	G	.	.	N	N	.	.	.	.	.	.	.	.	.	.	N	N	N	N	.	.	.	N	N	.	N	.	.	.	T	T	T	T	T
1450916	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	N	A	A	N	.	.	.	.	.
1451124	T	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	N	G	G	G	G
1458573	T	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A
1466798	T	.	.	.	.	.	.	.	.	.	.	.	C	N	C	N	N	C	N	N	C	N	N	C	C	C	C	N	C	C	N	C	C	N
1466799	C	.	.	.	.	.	.	.	.	.	.	.	N	T	N	T	N	N	N	N	N	N	N	.	N	.	N	N	N	N	N	N	N	N
1473704	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	C	A	N	N	.	.	.	.	.
1478233	C	.	.	N	N	.	.	.	.	N	.	N	T	.	.	N	.	.	N	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.
1481292	C	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	T	T	N	N	N	T	T	T
1481297	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	A	N	N	N	N	N	N	N
1481381	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	A	A	A	N	A	N	N	N
1481393	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	N	A	N
1511518	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	G	N	N	N	N	G	N
1549630	A	.	.	.	.	.	.	.	.	N	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G
1576532	G	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	N	T	N	T	N	N	N	N	T	T	N	N	N	N	N	N
1577025	T	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	N	N	N	N	N	N	N	N	A	N	N	N	N	N	N
1586982	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
1589543	C	.	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
1614945	T	.	.	.	.	.	.	.	.	N	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
1644408	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A
1650020	G	.	.	N	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	N	.	A	.	.	.	.	.	.	.	.	.
1708192	C	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
1711356	A	.	N	N	N	N	N	N	G	N	N	N	N	N	N	N	N	N	N	N	.	N	N	N	N	.	N	N	N	N	N	N	.	N
1711357	C	.	N	N	N	N	N	N	G	N	.	N	N	N	N	N	N	N	N	N	.	N	N	N	N	.	N	N	N	N	N	N	.	N
1713927	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	N	.	.	.	.	.	.	N	A	A	A	A
1722566	T	.	.	N	N	.	.	.	.	N	.	N	.	N	.	N	N	N	N	N	G	N	N	N	.	N	.	.	.	.	.	.	.	.
1722572	G	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	N	N	N	A	N	N	N	N	A	N	.	.	.	.	.	.	.	.
1724647	C	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
1739399	G	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	N	N	N	T	N	N	N	N	T	.	.	.	.	N	.	.	.	.
1754168	T	.	.	N	N	.	.	.	.	.	.	.	.	.	.	N	G	G	N	G	N	N	N	N	G	.	.	.	.	.	.	.	.	.
1780119	C	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	T	N	N	N	N	N	N
1780122	G	.	.	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	T	N	N	N	N	N	N
1781245	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	A	N	N	N	N	N	N
1781246	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	C	N	N	N	N	N	N
1791458	A	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	N	N	N	N	N	N	N	N	G	N	.	.	N	.	.	.	.	.
1802254	G	.	.	N	.	.	.	.	.	N	.	.	.	.	.	N	N	A	N	A	A	A	N	A	A	.	.	.	.	.	.	.	.	.
1870763	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	N	A	A	N	N	N	N	N	N
1870766	G	.	.	.	.	.	.	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	N	N	T	N	N	N	N	N	N
1871129	T	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	C	N	N	C
1883743	C	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	N	N	N	.	.	.	N	.	.	N	N	N	N	N	T	N	N	N
1883750	A	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	N	.	N	.	.	.	N	.	.	N	N	N	N	N	T	N	N	N
1935112	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
1952848	G	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Position	Reference (CO92)	LA1009	Barcelona 3031	TRP002	NAB003	London ES 8124/8291/11972	OSL1	Ber37	Ber45	London 6330	Bolgar 2370	MAN008	STA001	NMS002	ELW098/549_O	LBG002	STN002	STN007	STN013	STN014	STN020	STN021	STN008	STN019	BRA001	BED024	BED028	BED030	BED034	OBS124	OBS107	OBS116	OBS137	OBS110
2010427	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
2071670	G	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
2076353	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
2077729	C	N	N	N	N	N	N	.	.	N	N	N	N	N	N	T	N	N	N	T	T	N	N	T	N	T	T	N	N	N	.	N	N	N
2077732	G	N	N	A	N	N	N	.	.	N	N	A	N	N	N	A	N	N	N	A	A	A	N	A	N	A	A	N	A	N	.	N	N	N
2105332	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	N	.	.	.	N	.	.	.	.	.	.	N	T	T	T	T
2105376	A	.	.	N	N	.	.	.	.	N	.	N	.	.	.	.	.	.	N	.	.	.	N	.	.	.	.	.	.	G	N	G	G	G
2238702	A	.	.	N	N	.	.	.	.	N	N	.	.	.	.	N	.	.	N	.	.	.	.	.	T	.	.	.	.	N	.	.	.	.
2262577	T	G	G	G	G	G	G	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
2264654	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A
2281061	C	.	N	.	.	.	.	.	.	.	N	.	A	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	N	A	A	A
2292030	C	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
2414599	T	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	N	.	N	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C
2471721	G	.	.	N	N	.	.	.	.	.	.	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	T	N	N	N	N	N	N
2471722	T	.	.	N	N	.	.	.	.	.	.	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	G	N	N	N	N	N	N
2472073	G	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.
2472383	A	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	G	G	G	G	G	G	G	G	G
2507983	T	.	.	N	N	.	.	.	.	.	.	.	.	.	.	.	N	.	N	.	.	.	N	N	.	G	G	G	G	G	G	G	G	G
2519931	C	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
2540517	A	.	.	.	.	.	.	.	.	.	.	.	N	N	N	T	N	N	N	T	N	N	N	N	T	N	N	N	N	N	N	N	N	N
2542508	A	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	C	N	N	N	N	N	N
2542509	G	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	N	N	A	N	N	N	N	N	N	N	A	N	N	N	N	N	N
2596736	A	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	G
2671194	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
2681179	G	.	.	N	.	.	.	.	.	N	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2727385	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	G
2863092	G	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
2877295	G	.	.	N	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2911615	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.
2918297	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	G	G	G	G	G	G	G	G	G
2930644	C	.	.	.	.	.	.	.	.	N	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N
2952582	A	.	.	N	.	N	.	.	.	N	.	.	.	C	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	N	.	.	.
2964936	A	.	.	N	N	N	.	.	.	N	.	N	.	.	.	N	N	N	N	.	.	.	N	N	.	G	G	G	G	G	G	G	G	G
2973013	C	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T
2992718	C	.	.	.	.	.	.	.	.	N	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3017615	A	N	.	G	.	N	N	.	.	N	.	.	.	N	N	N	.	.	.	.	.	N	.	N	.	.	.	.	.	.	.	.	.	.
3025157	G	.	A	N	N	N	.	.	N	N	N	N	N	N	N	N	N	N	.	N	N	N	N	.	N	N	N	N	N	.	N	N	N	N
3030042	G	.	.	.	.	.	.	.	.	N	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
3035749	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3070014	G	N	N	C	N	N	N	.	N	N	N	.	N	N	N	.	N	N	.	.	.	N	N	N	N	.	.	.	N	N	.	.	N	.
3098104	C	.	N	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	N	A	A	A
3105712	T	.	.	.	.	.	.	.	.	N	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3206734	T	.	.	.	.	.	.	.	.	N	.	A	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3223359	C	N	N	N	N	N	N	N	N	N	N	N	N	A	.	N	A	N	N	A	N	A	N	N	N	N	N	N	A	N	N	N	N	N
3229407	T	.	N	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C
3254908	G	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
3256366	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3269577	G	.	.	.	.	.	.	.	.	N	.	.	T	T	T	N	T	T	T	T	T	T	T	T	T	T	T	T	T	T	N	T	T	T
3269613	G	.	.	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	.	.	A	N	.	N
3269615	C	.	.	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	.	N	T	N	N	N
3299755	C	N	T	N	.	N	N	N	N	N	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
3377812	G	.	.	.	.	.	.	.	.	N	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3387542	C	N	T	N	N	N	N	.	.	N	N	N	N	N	T	N	N	T	N	N	N	N	N	T	N	N	N	N	N	N	N	N	N	N
3407572	A	.	N	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
3448508	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
3515933	A	.	.	.	N	.	.	.	.	.	.	.	.	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	.	N	N	N	N
3535033	C	.	.	.	.	.	.	.	.	N	.	.	N	N	N	N	N	N	N	N	T	N	N	N	N	N	N	N	N	N	.	N	N	N
3535034	C	.	.	.	.	.	.	.	.	N	.	.	N	N	N	N	N	N	N	N	G	N	N	N	N	N	N	N	N	N	.	.	N	N
3540139	G	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Position	Reference (CO92)	LA1009	Barcelona 3031	TRP002	NAB003	London ES 8124/8291/11972	OSL1	Ber37	Ber45	London 6330	Bolgar 2370	MAN008	STA001	NMS002	ELW098/549_O	LBG002	STN002	STN007	STN013	STN014	STN020	STN021	STN008	STN019	BRA001	BED024	BED028	BED030	BED034	OBS124	OBS107	OBS116	OBS137	OBS110
3568853	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3610371	C	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
3613964	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	N	.	A	A	A	A	A	A	A	A	A
3620114	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
3620500	G	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
3642484	C	.	.	.	.	.	.	.	.	N	.	A	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3643387	G	.	.	N	.	.	N	.	N	N	T	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3644733	G	.	.	.	.	.	N	.	.	N	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.
3698192	G	.	.	N	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3761046	G	N	.	N	.	N	N	N	N	N	N	N	.	N	N	N	N	N	N	N	.	N	N	N	N	N	N	N	N	N	A	N	N	N
3764396	C	N	N	N	N	N	.	.	.	N	.	N	A	A	A	N	N	N	N	A	A	N	N	N	N	N	A	N	N	A	A	A	A
3782640	G	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A
3806677	C	T	T	T	T	T	T	T	T	T	.	T	T	T	T	T	T	T	N	T	T	T	N	T	T	T	T	T	T	T	T	T	T
3824821	G	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A
3826553	A	N	N	N	N	N	N	N	.	N	N	N	N	N	.	N	G	N	N	N	N	N	N	N	.	N	N	N	N	.	.	N	N	N
3860572	G	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	A	N	N	A	A	A	N	N	A	.	.	.	.	.	.	.	.	.
3872698	C	.	.	N	.	.	.	.	.	N	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
3888808	C	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
3920393	G	.	.	.	.	.	.	.	.	N	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
3944305	C	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
3973901	G	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
3988141	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T
3989422	C	.	.	N	N	.	.	.	.	.	.	.	.	.	A	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4134121	A	.	.	.	N	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
4150574	C	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
4190286	C	.	N	.	.	.	.	.	.	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
4200201	G	.	.	C	.	.	N	.	.	N	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4200639	C	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
4208536	A	.	.	N	N	.	.	.	.	N	.	.	.	.	.	N	.	.	N	.	.	.	N	N	.	G	G	G	G	G	G	G	G	G
4214590	A	.	.	.	.	.	.	.	.	N	.	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	T	N	N	N	N	N	N
4232240	C	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	T	N	N	N
4236782	C	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	N	.	N	.	N	N	T	N	N
4236789	C	.	.	N	.	.	N	N	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	N	.	N	N	N	N	G	N	N
4242260	G	.	.	N	.	.	.	.	.	N	.	.	.	T	T	N	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
4271997	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.	.
4282033	G	.	.	.	.	.	.	.	.	N	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4301295	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4343973	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
4363505	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
4396236	G	.	.	.	.	.	N	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
4421278	G	.	N	N	N	N	.	.	N	N	N	.	N	N	.	N	N	N	N	N	.	N	N	N	N	N	N	N	N	A	N	N	N	N
4456212	C	.	N	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
4491395	C	N	N	N	T	N	.	.	.	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	.	N	N	N	N	N	N	N	N	N
4567317	C	.	.	.	.	.	.	.	.	N	.	.	.	N	N	N	N	N	N	N	N	N	.	N	.	.	N	.	N	N	.	A	N	N
4590845	A	.	.	N	.	N	.	.	.	N	.	.	.	.	.	.	.	.	N	.	.	.	.	.	G	N	.	.	.	.	.	.	.	.
4616904	T	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C
4631182	G	.	.	N	N	.	.	.	.	N	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.
4642828	G	.	.	N	N	N	.	.	.	N	.	N	.	.	N	N	N	N	N	N	N	N	N	N	.	N	N	A	N	A	A	A	A	A
4646438	G	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Supplementary Data 2 – Assessment of unique SNPs in all newly reconstructed genomes that exhibited a minimum of 50% genomic coverage at 5-fold. Genomes that are not listed within the table did not yield any private SNP calls (LAI009, STN014, STN021, STN019, STN007, STN002, STN008, STN013, BED034, BED024).

Sample	SNP coordinates on reference (CO92)	Ref	SNP	Occurs in # of samples	Mean coverage (strict mapping - 50bp window)	Mean coverage (lenient mapping - 50bp window)	Mapping ratio (lenient/strict - 50bp window)	Heterozygous positions (50bp window)	Positions not covered (50bp window)	Score	Classification (True / False)
BED028	300 043	G	A	1	10.02	11.32	1.13	1	0	2.13	FALSE
	1 481 297	G	A	1	8.18	8.9	1.09	0	24	25.09	FALSE
BED030	1 577 025	T	A	1	4.12	5.8	1.41	0	24	25.41	FALSE
	1 780 119	C	T	1	11.08	13.26	1.20	0	20	21.20	FALSE
	1 780 122	G	T	1	9.06	11.12	1.23	0	23	24.23	FALSE
	1 781 245	G	A	1	5.96	6.54	1.10	0	24	25.10	FALSE
	1 781 246	G	C	1	6.34	6.94	1.09	0	23	24.09	FALSE
	1 870 766	G	T	1	4.76	5.38	1.13	0	21	22.13	FALSE
	2 471 721	G	T	1	7.04	9.74	1.38	0	22	23.38	FALSE
	2 471 722	T	G	1	7.62	10.38	1.36	0	21	22.36	FALSE
	2 542 508	A	C	1	8.12	8.74	1.08	0	24	25.08	FALSE
	4 214 590	A	T	1	10.36	11.24	1.08	0	0	1.08	FALSE
BRA001	1 384 546	A	G	1	20.82	17.94	0.86	2	0	2.86	FALSE
	1 650 020	G	A	1	6.56	6.56	1.00	0	0	1.00	TRUE
	1 791 458	A	G	1	6.82	6.82	1.00	0	0	1.00	TRUE
	2 238 702	A	T	1	8.54	9.08	1.06	0	0	1.06	FALSE
BRA003	379 431	A	C	1	17.4	28.6	1.64	1	0	2.64	FALSE
	731 010	C	T	1	21.44	45.72	2.13	2	0	4.13	FALSE
	1 254 820	G	A	1	20.8	25.64	1.23	1	0	2.23	FALSE
	3 518 204	G	A	1	13.36	16.6	1.24	1	0	2.24	FALSE
	4 165 599	A	G	1	11.2	22.6	2.02	1	0	3.02	FALSE
	4 200 582	T	C	1	11.5	24.7	2.15	1	0	3.15	FALSE
	4 201 835	G	A	1	80.96	100.24	1.24	0	0	1.24	FALSE
	4 371 559	G	A	1	29.62	66.9	2.26	1	0	3.26	FALSE
	4 598 920	A	G	1	3.8	4.48	1.18	0	11	12.18	FALSE
LBG002	234 892	T	C	1	3.54	4.2	1.19	2	19	22.19	FALSE
	700 976	A	G	1	9.46	13.08	1.38	0	0	1.38	FALSE
MAN008	1 201 122	G	T	1	20.82	20.82	1.00	0	0	1.00	TRUE
	2 681 179	G	A	1	18.92	18.92	1.00	0	0	1.00	TRUE
	3 035 749	G	T	1	25.26	26.4	1.05	0	0	1.05	FALSE
	3 206 734	T	A	1	12.62	12.62	1.00	0	0	1.00	TRUE
	3 642 484	C	A	1	8.7	8.7	1.00	0	0	1.00	TRUE
	3 644 733	G	A	1	62.16	62.46	1.00	0	0	1.00	TRUE
	3 698 192	G	T	1	18.08	18.08	1.00	0	0	1.00	TRUE
	4 282 033	G	T	1	37.16	37.16	1.00	0	0	1.00	TRUE
NAB003	1 230 108	T	C	1	24.02	27.82	1.16	2	0	3.16	FALSE
NAB005	1 989 349	A	T	1	6.56	10.62	1.62	1	0	2.62	FALSE
	3 518 201	A	G	1	4.06	16.74	4.12	1	0	5.12	FALSE
	3 782 920	T	A	1	10.64	20.02	1.88	1	0	2.88	FALSE
	4 204 373	T	G	1	4.88	5.2	1.07	1	4	6.07	FALSE
NMS002	19 828	C	A	1	14.6	14.6	1.00	0	0	1.00	TRUE
	22 699	C	T	1	4.08	4.08	1.00	0	0	1.00	TRUE
	1 304 271	C	T	1	13.02	13.02	1.00	0	0	1.00	TRUE
	1 364 191	C	A	1	7.96	7.96	1.00	0	0	1.00	TRUE
	2 472 073	G	T	1	10.84	10.84	1.00	0	0	1.00	TRUE
	2 930 644	C	T	1	17.42	17.42	1.00	0	0	1.00	TRUE
	2 952 582	A	C	1	11.6	11.6	1.00	0	0	1.00	TRUE
	2 992 718	C	A	1	16.78	16.78	1.00	0	0	1.00	TRUE
	3 105 712	T	C	1	12.66	12.66	1.00	0	0	1.00	TRUE
	3 377 812	G	A	1	14.12	14.12	1.00	0	0	1.00	TRUE
	3 515 933	A	T	1	3.94	3.94	1.00	1	0	2.00	FALSE
	4 646 438	G	A	1	16.38	16.38	1.00	0	0	1.00	TRUE

Sample	SNP coordinates on reference (CO92)	Ref	SNP	Occurs in # of samples	Mean coverage (strict mapping - 50bp window)	Mean coverage (lenient mapping - 50bp window)	Mapping ratio (lenient/strict - 50bp window)	Heterozygous positions (50bp window)	Positions not covered (50bp window)	Score	Classification (True / False)
STN004	25 876	A	G	1	9.4	10.84	1.15	2	0	3.15	FALSE
	216 391	T	C	1	21.92	35.42	1.62	2	0	3.62	FALSE
	221 703	T	C	1	62.44	84.96	1.36	1	0	2.36	FALSE
	224 303	C	T	1	51.76	80.26	1.55	1	0	2.55	FALSE
	224 360	G	A	1	29.36	55.46	1.89	0	0	1.89	FALSE
	225 263	C	T	1	17.42	22.1	1.27	1	0	2.27	FALSE
	225 266	T	A	1	18.98	23.52	1.24	1	0	2.24	FALSE
	225 479	T	G	1	52.18	64.78	1.24	1	0	2.24	FALSE
	225 578	A	T	1	21.92	24.76	1.13	0	0	1.13	FALSE
	226 897	A	G	1	46.5	80.9	1.74	0	0	1.74	FALSE
	228 481	C	T	1	18.52	43.74	2.36	1	0	3.36	FALSE
	232 268	T	C	1	14.8	23.02	1.56	4	0	5.56	FALSE
	234 154	G	A	1	73.86	110.82	1.50	2	0	3.50	FALSE
	234 724	T	C	1	47.74	72.98	1.53	1	0	2.53	FALSE
	234 757	G	A	1	13.64	39.4	2.89	2	0	4.89	FALSE
	235 238	T	G	1	57.86	84.26	1.46	1	0	2.46	FALSE
	403 948	T	C	1	28.14	36.62	1.30	1	0	2.30	FALSE
	404 185	T	G	1	56.76	86.14	1.52	0	0	1.52	FALSE
	404 188	T	C	1	58.48	87.14	1.49	0	0	1.49	FALSE
	404 236	T	C	1	43.1	61.24	1.42	0	0	1.42	FALSE
	404 260	T	C	1	20.82	32.28	1.55	0	0	1.55	FALSE
	404 419	C	T	1	66.94	95	1.42	0	0	1.42	FALSE
	405 954	C	T	1	64.06	82.64	1.29	0	0	1.29	FALSE
	405 996	G	A	1	91.2	96.98	1.06	0	0	1.06	FALSE
	406 044	A	G	1	82.86	93.34	1.13	0	0	1.13	FALSE
	406 089	A	G	1	24.52	37.82	1.54	1	0	2.54	FALSE
	407 289	G	C	1	60.64	72.5	1.20	1	0	2.20	FALSE
	407 329	T	C	1	45.54	58.68	1.29	0	0	1.29	FALSE
	512 211	G	A	1	3.8	11.48	3.02	0	0	3.02	FALSE
	593 142	C	T	1	22.06	31.66	1.44	1	0	2.44	FALSE
	1 185 588	A	T	1	31	40.58	1.31	0	1	2.31	FALSE
	1 185 603	A	G	1	53.8	64.78	1.20	0	0	1.20	FALSE
	1 185 681	T	C	1	36.52	56.82	1.56	0	0	1.56	FALSE
	1 491 878	C	T	1	3.86	4.78	1.24	1	0	2.24	FALSE
	1 573 394	T	C	1	39.8	94.36	2.37	2	0	4.37	FALSE
	2 032 807	A	G	1	5.36	12.9	2.41	1	1	4.41	FALSE
	2 728 074	G	A	1	14	23.56	1.68	1	0	2.68	FALSE
	2 861 354	C	T	1	5.72	6.18	1.08	1	0	2.08	FALSE
	2 976 906	G	A	1	4.08	4.08	1.00	1	0	2.00	FALSE
	3 241 788	C	A	1	5.18	11.62	2.24	0	0	2.24	FALSE
	3 518 159	C	T	1	5.96	6.62	1.11	3	0	4.11	FALSE
	3 518 180	A	C	1	7	7.88	1.13	2	0	3.13	FALSE
	3 529 395	T	C	1	26.7	33.7	1.26	2	0	3.26	FALSE
	3 897 621	T	C	1	12.66	16.22	1.28	2	0	3.28	FALSE
	3 908 319	T	G	1	8.38	8.82	1.05	1	0	2.05	FALSE
	3 948 512	A	G	1	3.36	6.08	1.81	2	1	4.81	FALSE
	3 949 187	A	G	1	36.56	54.12	1.48	2	0	3.48	FALSE
	4 197 768	T	G	1	25.86	32.56	1.26	1	0	2.26	FALSE
	4 197 774	C	A	1	21.76	27.28	1.25	1	0	2.25	FALSE
	4 201 293	T	C	1	19.36	65.7	3.39	2	0	5.39	FALSE
	4 201 296	G	C	1	20.7	70.28	3.40	2	0	5.40	FALSE
	4 202 867	A	G	1	14.54	88.32	6.07	3	0	9.07	FALSE
	4 207 929	C	T	1	6.68	20.64	3.09	2	0	5.09	FALSE
	4 305 353	A	G	1	31.88	59.82	1.88	0	0	1.88	FALSE
	4 305 362	A	G	1	33.04	63.52	1.92	0	0	1.92	FALSE
	4 649 029	C	T	1	26.78	28.48	1.06	1	0	2.06	FALSE

Sample	SNP coordinates on reference (CO92)	Ref	SNP	Occurs in # of samples	Mean coverage (strict mapping - 50bp window)	Mean coverage (lenient mapping - 50bp window)	Mapping ratio (lenient/strict - 50bp window)	Heterozygous positions (50bp window)	Positions not covered (50bp window)	Score	Classification (True / False)
STN011	165 491	C	T	1	115.9	119.16	1.03	0	0	1.03	FALSE
	221 690	G	T	1	41.42	74.62	1.80	1	0	2.80	FALSE
	221 939	T	C	1	97.9	133.26	1.36	1	0	2.36	FALSE
	222 025	G	A	1	60.78	124.2	2.04	0	0	2.04	FALSE
	222 028	C	A	1	59.64	121.2	2.03	0	0	2.03	FALSE
	222 670	C	A	1	105.56	113.58	1.08	0	0	1.08	FALSE
	223 252	T	C	1	95.28	105.36	1.11	0	0	1.11	FALSE
	223 601	T	C	1	19.82	66.48	3.35	2	0	5.35	FALSE
	224 384	A	G	1	8.82	27.14	3.08	0	14	17.08	FALSE
	224 443	G	A	1	20.26	81.72	4.03	1	0	5.03	FALSE
	224 650	C	T	1	37.72	58.02	1.54	1	0	2.54	FALSE
	224 760	G	T	1	94.2	119.8	1.27	0	0	1.27	FALSE
	225 006	T	C	1	32.56	66.5	2.04	1	0	3.04	FALSE
	225 021	C	T	1	49.82	83.5	1.68	1	0	2.68	FALSE
	225 056	G	A	1	44.52	69.72	1.57	1	0	2.57	FALSE
	225 452	G	A	1	49.72	84.34	1.70	1	0	2.70	FALSE
	225 699	G	A	1	28.3	41.96	1.48	3	0	4.48	FALSE
	227 745	T	C	1	90.06	112.02	1.24	0	0	1.24	FALSE
	227 772	A	G	1	75.04	101.58	1.35	0	0	1.35	FALSE
	228 415	G	C	1	74.24	83.3	1.12	0	0	1.12	FALSE
	228 733	C	T	1	45.72	85.36	1.87	0	0	1.87	FALSE
	228 754	T	A	1	89.16	120	1.35	0	0	1.35	FALSE
	228 855	A	T	1	62.02	89.08	1.44	1	0	2.44	FALSE
	229 316	C	T	1	62.96	70.46	1.12	2	0	3.12	FALSE
	229 707	A	G	1	64.72	121.06	1.87	4	0	5.87	FALSE
	229 730	A	G	1	69.22	118.74	1.72	4	0	5.72	FALSE
	229 790	A	T	1	49.08	55.28	1.13	1	0	2.13	FALSE
	230 402	C	G	1	64.56	83.48	1.29	1	0	2.29	FALSE
	232 637	G	C	1	93.5	126.86	1.36	0	0	1.36	FALSE
	232 661	G	A	1	76.6	117.98	1.54	0	0	1.54	FALSE
	232 787	C	T	1	94.5	108.52	1.15	0	0	1.15	FALSE
	232 817	A	G	1	62.46	77.06	1.23	0	0	1.23	FALSE
	232 938	C	A	1	21.54	76.74	3.56	0	0	3.56	FALSE
	233 890	A	G	1	74.3	107.3	1.44	0	0	1.44	FALSE
	233 897	T	C	1	87.62	118.36	1.35	0	0	1.35	FALSE
	233 956	A	T	1	133.3	137.52	1.03	0	0	1.03	FALSE
	235 120	A	G	1	79.06	107.38	1.36	0	0	1.36	FALSE
	700 937	G	T	1	71.98	102.22	1.42	1	0	2.42	FALSE
	700 988	A	G	1	55.6	74.24	1.34	1	0	2.34	FALSE
	701 012	G	A	1	44.48	61.66	1.39	1	0	2.39	FALSE
	701 071	G	A	1	63.8	69.5	1.09	0	0	1.09	FALSE
	1 169 671	G	A	1	11.1	16.48	1.48	2	0	3.48	FALSE
	1 264 029	G	A	1	35.86	44.94	1.25	1	0	2.25	FALSE
	1 264 032	A	G	1	32.1	40.88	1.27	1	0	2.27	FALSE
	1 306 824	C	T	1	17.86	25.54	1.43	1	0	2.43	FALSE
	1 573 013	T	A	1	5.48	22.6	4.12	1	0	5.12	FALSE
	1 573 061	G	A	1	33.28	52	1.56	0	0	1.56	FALSE
	1 573 067	C	T	1	42.18	63.94	1.52	0	0	1.52	FALSE
	1 573 116	T	C	1	56.7	116.76	2.06	1	0	3.06	FALSE
	1 573 307	C	T	1	22.46	49.98	2.23	2	0	4.23	FALSE
	1 819 520	C	T	1	5.16	22.18	4.30	2	0	6.30	FALSE
	1 820 043	T	C	1	35.14	71.84	2.04	0	0	2.04	FALSE
	1 943 496	C	T	1	24.12	42.42	1.76	1	0	2.76	FALSE
	1 943 523	C	A	1	54.42	76.36	1.40	0	0	1.40	FALSE
	3 453 103	A	T	1	35.56	44.02	1.24	1	0	2.24	FALSE

Sample	SNP coordinates on reference (CO92)	Ref	SNP	Occurs in # of samples	Mean coverage (strict mapping - 50bp window)	Mean coverage (lenient mapping - 50bp window)	Mapping ratio (lenient/strict - 50bp window)	Heterozygous positions (50bp window)	Positions not covered (50bp window)	Score	Classification (True / False)
STN011	3 672 205	C	T	1	87.14	89.94	1.03	0	0	1.03	FALSE
	3 767 380	G	A	1	33.4	61.7	1.85	0	0	1.85	FALSE
	3 767 381	G	A	1	34.88	63.34	1.82	0	0	1.82	FALSE
	3 975 802	A	C	1	57.94	73.56	1.27	1	0	2.27	FALSE
	4 190 390	G	A	1	35.46	37.24	1.05	1	0	2.05	FALSE
	4 200 114	A	G	1	10.56	26.1	2.47	1	0	3.47	FALSE
	4 205 578	A	G	1	41.22	46.56	1.13	2	0	3.13	FALSE
	4 619 670	T	A	1	62.46	70.78	1.13	1	0	2.13	FALSE
	4 619 706	C	A	1	81.76	92.48	1.13	0	0	1.13	FALSE
	4 645 024	T	C	1	48.16	62.22	1.29	0	0	1.29	FALSE
	4 646 078	T	A	1	34.04	42.8	1.26	0	0	1.26	FALSE
	4 647 149	C	T	1	50.02	68.46	1.37	1	0	2.37	FALSE
	4 647 188	A	G	1	66.08	81.94	1.24	0	0	1.24	FALSE
	4 647 596	G	A	1	75.88	93.02	1.23	1	0	2.23	FALSE
	4 647 647	G	A	1	120.68	125.8	1.04	0	0	1.04	FALSE
	4 648 263	G	A	1	68.56	80.3	1.17	0	0	1.17	FALSE
	4 648 985	T	C	1	33.2	66.14	1.99	1	0	2.99	FALSE
STN020	1 722 566	T	G	1	3.6	4.6	1.28	3	1	5.28	FALSE
	3 535 034	C	G	1	6.42	6.56	1.02	0	0	1.02	FALSE
TRP002	33 569	G	C	1	6.58	10.28	1.56	1	2	4.56	FALSE
	146 113	A	T	1	15	21.26	1.42	2	0	3.42	FALSE
	362 185	T	G	1	19.22	27.16	1.41	1	0	2.41	FALSE
	497 767	G	T	1	7.58	12.5	1.65	0	5	6.65	FALSE
	1 370 093	T	G	1	7.6	11.84	1.56	0	0	1.56	FALSE
	3 017 615	A	G	1	7.26	16.84	2.32	0	0	2.32	FALSE
	3 256 366	T	C	1	3.74	5.2	1.39	1	0	2.39	FALSE
	4 200 201	G	C	1	12.24	26.76	2.19	1	0	3.19	FALSE

Supplementary Data 3 – SNP profile of the low-coverage NMS003 isolate

Description	Position in CO92	Reference call	Coverage in NMS003	NMS003	LA1009	London BD	Barcelona 3031	OSL1 (Oslo)	TRP002	NAB003	Ber37	Ber45	London 6330	Bolgar 2370	MAN008	STA001	NMS002	ELW098	LBG002	STN002	STN007	STN013	STN014	STN020	STN021	STN008	STN019	BRA001	BED024	BED028	BED030	BED034	OBS124	OBS107	OBS116	OBS137	OBS110
Black Death diagnostic positions	189 227	C	2X	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	1 871 476	G	N/A	N	.	.	.	.	N	N	.	.	.	.	.	.	.	.	N	N	.	N	.	.	.	N	N	.	.	.	.	.	.	.	.	.	.
Derived Branch 1 diagnostic positions	699 494	A	≥3X	G	G	G	G	G	G	G	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
	2 262 577	T	≥3X	G	G	G	G	G	G	G	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
	3 643 387	G	≥3X	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	3 806 677	C	≥3X	T	T	T	T	T	T	T	T	T	T	.	T	T	T	T	T	T	T	N	T	T	T	N	T	T	T	T	T	T	T	T	T	T	T
	4 301 295	G	≥3X	.	.	.	.	N	N	.	.	N	N	T	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Post-Black Death lineage diagnostic SNP positions	29 368	G	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	100 383	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	169 412	C	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	173 032	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	186 060	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	T	T	N	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	200 723	C	≥3X	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	217 009	G	≥3X	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	225 436	T	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N	A	A	A	N	N	N	N	A	A	A	A	N	A	A	A	A	A	A
	226 722	C	2X	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T
	400 143	G	≥3X	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	477 107	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	480 773	C	2X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	482 327	G	1X	.	.	.	.	.	N	N	.	.	.	.	.	.	.	.	N	.	N	N	.	.	.	N	.	.	N	T	T	T	T	T	T	T	T	T
	862 385	T	2X	.	.	.	.	.	N	.	.	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
	867 712	C	≥3X	.	.	.	.	.	.	N	.	.	N	.	.	.	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	868 549	G	≥3X	.	.	.	.	N	.	.	.	N	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C	C
	869 820	A	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	G	G	G	G	G	G
	899 158	C	1X	.	.	.	N	.	N	.	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	N	.	.	T	T	T	T	T	T	T	T	T	T
	951 295	C	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	965 281	C	2X	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	1 159 539	T	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	1 168 951	G	2X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	1 189 479	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	1 232 222	C	2X	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	1 254 157	C	2X	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	1 308 719	G	≥3X	.	.	.	.	.	N	.	.	.	N	.	A	A	A	A	N	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	1 378 105	G	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	1 440 851	G	≥3X	.	.	.	.	.	N	N	.	.	.	.	.	.	.	.	N	N	N	N	.	.	.	N	N	.	N	.	.	.	T	T	T	T	T	T
	1 458 573	T	≥3X	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A
	1 586 982	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	1 614 945	T	2X	.	.	.	.	.	.	.	.	.	N	.	.	.	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
	1 644 408	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A
	1 708 192	C	2X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	1 724 647	C	≥3X	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	1 935 112	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	1 952 848	G	2X	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	2 071 670	G	≥3X	.	.	.	.	N	.	.	.	.	N	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	2 076 353	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	2 264 654	C	2X	.	.	.	.	.	N	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A
	2 292 030	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	2 414 599	T	2X	.	.	.	.	.	N	N	.	.	N	.	.	.	.	.	N	N	.	N	.	.	.	.	.	C	C	C	C	C	C	C	C	C	C	C
	2 472 383	A	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	G	G	G	G	G	G	G	G	G	G	G
	2 507 983	T	2X	.	.	.	.	.	N	N	.	.	.	.	.	.	.	.	.	N	.	N	.	.	.	N	N	.	G	G	G	G	G	G	G	G	G	G
	2 519 931	C	1X	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	2 596 736	A	≥3X	.	.	.	.	.	N	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	G	G
	2 671 194	G	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	2 727 385	A	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	G	G	G	G
	2 877 295	G	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	2 918 297	T	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	G	G	G	G	G	G	G	G	G	G
	2 964 936	A	≥3X	.	.	N	.	.	N	N	.	.	N	.	N	.	.	.	N	N	N	N	.	.	.	N	N	.	G	G	G	G	G	G	G	G	G	G
	2 973 013	C	2X	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T
	3 030 042	G	≥3X	.	.	.	.	.	.	.	.	.	N	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T

Description	Position in CO92	Reference call	Coverage in NMS003	NMS003	LA1009	London BD	Barcelona 3031	OSL1 (Oslo)	TRP002	NAB003	Ber37	Ber45	London 6330	Bolgar 2370	MAN008	STA001	NMS002	ELW098	LBG002	STN002	STN007	STN013	STN014	STN020	STN021	STN008	STN019	BRA001	BED024	BED028	BED030	BED034	OBS124	OBS107	OBS116	OBS137	OBS110
Post-Black Death lineage diagnostic SNP positions	3 229 407	T	≥3X	.	.	.	N	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C
	3 254 908	G	2X	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	3 407 572	A	≥3X	.	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
	3 540 139	G	≥3X	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	3 610 371	C	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T
	3 613 964	C	≥3X	.	.	.	.	.	N	.	.	.	N	.	.	.	.	.	N	.	.	.	.	.	N	.	.	A	A	A	A	A	A	A	A	A	A
	3 620 114	G	2X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	3 620 500	G	≥3X	.	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	3 782 640	G	≥3X	.	.	.	.	.	N	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	3 824 821	G	2X	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A
	3 872 698	C	≥3X	.	.	.	.	.	N	.	.	.	N	.	.	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	3 944 305	C	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	3 973 901	G	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	3 988 141	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T
	4 134 121	A	≥3X	.	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	4 150 574	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
	4 190 286	C	≥3X	.	.	.	N	.	.	.	.	.	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	4 200 639	C	≥3X	.	.	.	.	.	.	.	.	.	N	.	.	N	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	4 208 536	A	2X	.	.	.	.	.	N	N	.	.	N	.	.	.	.	.	N	.	.	N	.	.	.	N	N	G	G	G	G	G	G	G	G	G
	4 242 260	G	≥3X	.	.	.	.	.	N	.	.	.	N	.	.	.	T	T	N	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	4 363 505	C	≥3X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	4 396 236	G	≥3X	.	.	.	.	N	.	.	.	.	N	.	.	.	.	.	N	.	.	.	.	.	.	.	.	T	T	T	T	T	T	T	T	T	T
	4 456 212	C	≥3X	.	.	.	N	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	A	A	A	A	A	A	A	A
	4 616 904	T	≥3X	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	C	C	C	C	C	C	C	C
	4 642 828	G	2X	.	.	N	.	.	N	N	.	.	N	.	N	.	.	N	N	N	N	N	N	N	N	N	N	.	N	N	A	N	A	A	A	A	A

Supplementary Data 4 – BLASTn analysis of reads overlapping private SNP position T3897987A, previously identified in the BSS31 genome

Hit name	Accession	Query length (bp)	Identity (%)	Query coverage (%)	Start position on reference	End position on reference	E value	Max Score
Enterobacter roggenkampii strain BP10374 chromosome, complete genome	CP038471.1	82	100	100	1252390	1252471	1.20E-32	149
Citrobacter sp. SNU WT2 chromosome, complete genome	CP038469.1	82	100	100	3967104	3967023	1.20E-32	149
Escherichia coli isolate f9610206-5e81-11e8-bf7f-3c4a9275d6c8 genome assembly, chromosome: VREC0864	LR536431.1	82	100	100	638933	638852	1.20E-32	149
Escherichia coli isolate f974b26a-5e81-11e8-bf7f-3c4a9275d6c8 genome assembly, chromosome: VREC0761	LR536430.1	82	100	100	1904244	1904325	1.20E-32	149
Citrobacter sp. LY-1 chromosome, complete genome	CP037864.1	82	100	100	4641762	4641843	1.20E-32	149
Klebsiella sp. PO552 chromosome, complete genome	CP037441.1	82	100	100	578048	577967	1.20E-32	149
Citrobacter freundii strain CAV1857 chromosome, complete genome	CP037734.1	82	100	100	575711	575630	1.20E-32	149
Klebsiella pneumoniae strain 18CPO060 chromosome, complete genome	CP034778.1	82	100	100	544280	544199	1.20E-32	149
Klebsiella pneumoniae strain BA4656 chromosome, complete genome	CP035905.1	82	100	100	1319182	1319263	1.20E-32	149
Klebsiella pneumoniae strain R46 chromosome, complete genome	CP035777.1	82	100	100	397416	397335	1.20E-32	149
Escherichia coli strain U15A chromosome, complete genome	CP035720.1	82	100	100	3539529	3539610	1.20E-32	149
Enterobacter cloacae strain CZ-1 chromosome, complete genome	CP035738.1	82	100	100	584575	584494	1.20E-32	149
Shigella sonnei strain AUSMDU00008333 genome assembly, chromosome: 1	LR213458.1	82	100	100	4349018	4349099	1.20E-32	149
Shigella flexneri strain AUSMDU00008332 genome assembly, chromosome: 1	LR213455.1	82	100	100	666303	666222	1.20E-32	149
Shigella flexneri strain AUSMDU00008355 genome assembly, chromosome: 1	LR213452.1	82	100	100	578912	578831	1.20E-32	149
Shigella sonnei strain AUSMDU00008361 genome assembly, chromosome: 1	LR213449.1	82	100	100	338848	338767	1.20E-32	149
Shigella flexneri strain SFL1520 chromosome, complete genome	CP032513.1	82	100	100	3277372	3277453	1.20E-32	149
Enterobacter cloacae strain EN3600 chromosome, complete genome	CP035633.1	82	100	100	3731832	3731913	1.20E-32	149
Citrobacter freundii strain R17 chromosome, complete genome	CP035276.1	82	100	100	577140	577059	1.20E-32	149
Klebsiella pneumoniae strain BA33875 chromosome, complete genome	CP035179.1	82	100	100	4240037	4239956	1.20E-32	149
Escherichia coli strain WCHEC020032 chromosome, complete genome	CP034966.1	82	100	100	607908	607827	1.20E-32	149
Escherichia coli strain SCEC020026 chromosome, complete genome	CP034958.1	82	100	100	623706	623625	1.20E-32	149
Klebsiella michiganensis strain M82255 chromosome, complete genome	CP035214.1	82	100	100	2272846	2272927	1.20E-32	149
Klebsiella pneumoniae strain 2N3 chromosome, complete genome	CP025541.2	82	100	100	2910854	2910773	1.20E-32	149
Kosakonia cowanii strain FBS 223 chromosome, complete genome	CP035129.1	82	100	100	532684	532603	1.20E-32	149
Escherichia coli strain EC25 chromosome, complete genome	CP035123.1	82	100	100	631008	630927	1.20E-32	149
Klebsiella variicola strain 15WZ-82 chromosome, complete genome	CP032354.1	82	100	100	1003744	1003663	1.20E-32	149
Klebsiella pneumoniae strain L5-2 chromosome, complete genome	CP025684.1	82	100	100	559842	559761	1.20E-32	149
Escherichia coli strain L103-2 chromosome, complete genome	CP034843.1	82	100	100	584027	583946	1.20E-32	149
Escherichia coli strain MS14387 genome assembly, chromosome: 1	LR130564.1	82	100	100	606497	606416	1.20E-32	149
Escherichia coli strain MS14384 genome assembly, chromosome: 1	LR130562.1	82	100	100	605935	605854	1.20E-32	149
Escherichia coli strain MS14385 genome assembly, chromosome: 1	LR130555.1	82	100	100	623502	623421	1.20E-32	149
Escherichia coli strain MS14386 genome assembly, chromosome: 1	LR130552.1	82	100	100	589787	589706	1.20E-32	149
Klebsiella pneumoniae strain KPC2 genome assembly, chromosome: 1	LR130548.1	82	100	100	571507	571426	1.20E-32	149
Escherichia coli strain B36 genome assembly, chromosome: 1	LR130545.1	82	100	100	643956	643875	1.20E-32	149
Klebsiella variicola strain 03-311-0071 genome assembly, chromosome: 1	LR130544.1	82	100	100	563394	563313	1.20E-32	149
Klebsiella variicola strain 04153260899A genome assembly, chromosome: 1	LR130543.1	82	100	100	555652	555571	1.20E-32	149
Klebsiella pneumoniae strain AJ218 genome assembly, chromosome: 1	LR130541.1	82	100	100	566759	566678	1.20E-32	149
Klebsiella variicola strain AJ055 genome assembly, chromosome: 1	LR130539.1	82	100	100	585267	585186	1.20E-32	149
Klebsiella variicola strain AJ292 genome assembly, chromosome: 1	LR130538.1	82	100	100	553463	553382	1.20E-32	149

Hit name	Accession	Query length (bp)	Identity (%)	Query coverage (%)	Start position on reference	End position on reference	E value	Max Score
Klebsiella pneumoniae strain 08EU827 chromosome, complete genome	CP025576.1	82	100	100	552279	552198	1.20E-32	149
Escherichia coli strain E-1246 chromosome, complete genome	CP025573.1	82	100	100	3959512	3959593	1.20E-32	149
Klebsiella sp. LY chromosome, complete genome	CP022444.1	82	100	100	1906634	1906715	1.20E-32	149
Escherichia coli strain 06-3462 chromosome, complete genome	CP034794.1	82	100	100	3789270	3789351	1.20E-32	149
Escherichia coli strain 2009C-4687 chromosome, complete genome	CP034799.1	82	100	100	2763347	2763266	1.20E-32	149
Escherichia coli strain 08-3918 chromosome, complete genome	CP034797.1	82	100	100	2289130	2289211	1.20E-32	149
Escherichia coli strain 08-3914 chromosome, complete genome	CP034808.1	82	100	100	2761300	2761219	1.20E-32	149
Escherichia coli strain 2009C-3554 chromosome, complete genome	CP034803.1	82	100	100	3732118	3732037	1.20E-32	149
Escherichia coli strain 2010C-3347 chromosome, complete genome	CP034806.1	82	100	100	4523304	4523385	1.20E-32	149
Escherichia coli strain 2010C-3142 chromosome, complete genome	CP034801.1	82	100	100	2756523	2756442	1.20E-32	149
Escherichia coli strain 2009C-3378 chromosome, complete genome	CP034792.1	82	100	100	5209340	5209259	1.20E-32	149
Escherichia coli strain ECCNB20-2 chromosome, complete genome	CP034787.1	82	100	100	568551	568470	1.20E-32	149
Escherichia coli strain ECZP248 chromosome, complete genome	CP034784.1	82	100	100	573681	573600	1.20E-32	149
Klebsiella pneumoniae strain NB5306 chromosome, complete genome	CP034760.1	82	100	100	570406	570325	1.20E-32	149
Enterobacter sp. N18-03635 chromosome, complete genome	CP034769.1	82	100	100	4010236	4010317	1.20E-32	149
Escherichia coli strain L65 chromosome, complete genome	CP034738.1	82	100	100	2577276	2577357	1.20E-32	149
Escherichia coli strain L41-1 chromosome, complete genome	CP034727.1	82	100	100	576717	576636	1.20E-32	149
Escherichia coli strain L100 chromosome, complete genome	CP034745.1	82	100	100	576717	576636	1.20E-32	149
Escherichia coli strain L53 chromosome, complete genome	CP034734.1	82	100	100	574974	574893	1.20E-32	149
Enterobacter hormaechei subsp. xiangfangensis strain UM_CRE-14 chromosome	CP023430.1	82	100	100	4178348	4178429	1.20E-32	149
Escherichia coli strain C4435 chromosome	CP027851.1	82	100	100	242037	242118	1.20E-32	149
Escherichia coli strain EC17GD31 chromosome, complete genome	CP031293.1	82	100	100	651010	650929	1.20E-32	149
Klebsiella quasipneumoniae strain D120-1 chromosome, complete genome	CP034678.1	82	100	100	551743	551662	1.20E-32	149
Escherichia coli strain WCHEC035053S1G0 chromosome, complete genome	CP034595.1	82	100	100	432214	432133	1.20E-32	149
Escherichia coli strain L37 chromosome, complete genome	CP034589.1	82	100	100	4092161	4092242	1.20E-32	149
Klebsiella pneumoniae strain T4 chromosome	CP034540.1	82	100	100	1056907	1056826	1.20E-32	149
Escherichia coli DSM 30083 = JCM 1649 = ATCC 11775 chromosome, complete genome	CP033092.2	82	100	100	576774	576693	1.20E-32	149
Klebsiella pneumoniae strain TOP52_1721_U1 chromosome	CP031938.1	82	100	100	953487	953568	1.20E-32	149
Citrobacter youngae strain NCTC13709 genome assembly, chromosome: 1	LR134485.1	82	100	100	539907	539826	1.20E-32	149
Klebsiella aerogenes strain NCTC9735 genome assembly, chromosome: 1	LR134475.1	82	100	100	548730	548649	1.20E-32	149
Escherichia coli strain NCTC11133 genome assembly, chromosome: 1	LR134340.1	82	100	100	3814717	3814798	1.20E-32	149
Klebsiella oxytoca strain NCTC13727 genome assembly, chromosome: 1	LR134333.1	82	100	100	913098	913017	1.20E-32	149
Escherichia coli strain NCTC9064 genome assembly, chromosome: 1	LR134315.1	82	100	100	555255	555174	1.20E-32	149
Escherichia coli strain NCTC9113 genome assembly, chromosome: 1	LR134311.1	82	100	100	558597	558516	1.20E-32	149
Escherichia coli strain NCTC9041 genome assembly, chromosome: 1	LR134296.1	82	100	100	589595	589514	1.20E-32	149
Escherichia coli strain NCTC9080 genome assembly, chromosome: 1	LR134295.1	82	100	100	573029	572948	1.20E-32	149
Klebsiella aerogenes strain NCTC9793 genome assembly, chromosome: 5	LR134280.1	82	100	100	4590261	4590180	1.20E-32	149
Escherichia coli strain NCTC8196 genome assembly, chromosome: 1	LR134270.1	82	100	100	4219393	4219474	1.20E-32	149
Klebsiella aerogenes strain NCTC9644 genome assembly, chromosome: 1	LR134254.1	82	100	100	565605	565524	1.20E-32	149
Escherichia coli strain NCTC10537 genome assembly, chromosome: 1	LR134248.1	82	100	100	600332	600251	1.20E-32	149
Escherichia coli strain NCTC9040 genome assembly, chromosome: 1	LR134247.1	82	100	100	607253	607172	1.20E-32	149
Escherichia coli strain NCTC9702 genome assembly, chromosome: 1	LR134246.1	82	100	100	621733	621652	1.20E-32	149
Escherichia coli strain NCTC9107 genome assembly, chromosome: 1	LR134240.1	82	100	100	620639	620558	1.20E-32	149
Escherichia coli strain NCTC9100 genome assembly, chromosome: 1	LR134239.1	82	100	100	568486	568405	1.20E-32	149
Escherichia coli strain NCTC9044 genome assembly, chromosome: 1	LR134238.1	82	100	100	1063950	1064031	1.20E-32	149

Hit name	Accession	Query length (bp)	Identity (%)	Query coverage (%)	Start position on reference	End position on reference	E value	Max Score
Escherichia coli strain NCTC9022 genome assembly, chromosome: 1	LR134237.1	82	100	100	594426	594345	1.20E-32	149
Escherichia coli strain NCTC9008 genome assembly, chromosome: 1	LR134236.1	82	100	100	571277	571196	1.20E-32	149
Klebsiella aerogenes strain NCTC9668 genome assembly, chromosome: 1	LR134235.1	82	100	100	3207114	3207033	1.20E-32	149
Escherichia coli strain NCTC8623 genome assembly, chromosome: 1	LR134234.1	82	100	100	2614038	2614119	1.20E-32	149
Escherichia coli strain NCTC9087 genome assembly, chromosome: 1	LR134231.1	82	100	100	569030	568949	1.20E-32	149
Klebsiella aerogenes strain NCTC8846 genome assembly, chromosome: 2	LR134230.1	82	100	100	3630739	3630820	1.20E-32	149
Escherichia coli strain NCTC9699 genome assembly, chromosome: 1	LR134228.1	82	100	100	596083	596002	1.20E-32	149
Escherichia coli strain NCTC9102 genome assembly, chromosome: 1	LR134227.1	82	100	100	566649	566568	1.20E-32	149
Escherichia coli strain NCTC9088 genome assembly, chromosome: 1	LR134226.1	82	100	100	560523	560442	1.20E-32	149
Escherichia coli strain NCTC9054 genome assembly, chromosome: 1	LR134225.1	82	100	100	571774	571693	1.20E-32	149
Klebsiella aerogenes strain NCTC9652 genome assembly, chromosome: 2	LR134224.1	82	100	100	552538	552457	1.20E-32	149
Escherichia coli strain NCTC11129 genome assembly, chromosome: 1	LR134222.1	82	100	100	558796	558715	1.20E-32	149
Escherichia coli strain NCTC11113 genome assembly, chromosome: 1	LR134221.1	82	100	100	669713	669632	1.20E-32	149
Escherichia coli strain NCTC11121 genome assembly, chromosome: 1	LR134220.1	82	100	100	625662	625581	1.20E-32	149
Klebsiella aerogenes strain NCTC10317 genome assembly, chromosome: 1	LR134217.1	82	100	100	559316	559235	1.20E-32	149

Supplementary Data 5 – BLAST analysis of reads overlapping private SNP position T3529404C, previously identified in the BSS31 genome

Hit name	Accession	Query length (bp)	Identity (%)	Query coverage (%)	Start position on reference	End position on reference	E value	Max Score
Lelliottia amnigena strain NCTC12124 genome assembly, chromosome: 1	LR134135.1	68	100.0	100	937392	937325	3.45E-25	123
Lelliottia amnigena strain FDAARGOS_395 chromosome, complete genome	CP023529.1	68	100.0	100	4425252	4425185	3.45E-25	123
Lelliottia jeotgali strain PFL01 chromosome, complete genome	CP018628.1	68	100.0	100	1057995	1057928	3.45E-25	123
Enterobacter sp. 638, complete genome	CP000653.1	68	100.0	100	998161	998094	3.45E-25	123
Serratia marcescens strain BWH-35 chromosome, complete genome	CP020507.1	68	98.5	100	1035039	1034972	1.47E-23	119
Serratia marcescens strain 95 chromosome, complete genome	CP020503.1	68	98.5	100	1062626	1062559	1.47E-23	119
Serratia marcescens strain BWH-23 chromosome, complete genome	CP020501.1	68	98.5	100	994208	994141	1.47E-23	119
Leclercia adecarboxylata strain 16008513 chromosome, complete genome	CP036199.1	68	98.5	100	4424379	4424446	1.47E-23	119
Serratia rubidaea strain NCTC9419 genome assembly, chromosome: 1	LR134155.1	68	98.5	100	3315345	3315278	1.47E-23	119
Serratia sp. FDAARGOS_506 chromosome, complete genome	CP033831.1	68	98.5	100	2493245	2493178	1.47E-23	119
Serratia sp. LS-1 chromosome, complete genome	CP033504.1	68	98.5	100	396218	396151	1.47E-23	119
Serratia marcescens AS-1 DNA, complete genome	AP019009.1	68	98.5	100	1061873	1061806	1.47E-23	119
Serratia marcescens strain KS10 chromosome	CP027798.1	68	98.5	100	3701490	3701557	1.47E-23	119
Serratia marcescens strain EL1 chromosome	CP027796.1	68	98.5	100	4240385	4240318	1.47E-23	119
Serratia rubidaea strain NCTC10036 genome assembly, chromosome: 1	LR134493.1	68	98.5	100	458815	458882	1.47E-23	119
Serratia marcescens strain N4-5 chromosome, complete genome	CP031316.1	68	98.5	100	1034687	1034620	1.47E-23	119
Serratia rubidaea strain NCTC10848 genome assembly, chromosome: 1	LS483492.1	68	98.5	100	3841750	3841817	1.47E-23	119
Serratia marcescens strain AR_0131 chromosome, complete genome	CP029715.1	68	98.5	100	5127851	5127784	1.47E-23	119
Serratia marcescens strain AR_0122, complete genome	CP029746.1	68	98.5	100	3705185	3705118	1.47E-23	119
Providencia rettgeri strain AR_0082 chromosome, complete genome	CP029736.1	68	98.5	100	20426	20493	1.47E-23	119
Serratia marcescens strain 332 chromosome, complete genome	CP021164.1	68	98.5	100	1010218	1010151	1.47E-23	119
Serratia marcescens strain SGAir0764 chromosome, complete genome	CP027300.1	68	98.5	100	1098591	1098524	1.47E-23	119
Serratia marcescens strain CAV1761 chromosome, complete genome	CP029449.1	68	98.5	100	1610427	1610494	1.47E-23	119
Serratia marcescens strain AR_0121 chromosome, complete genome	CP028949.1	68	98.5	100	1332415	1332482	1.47E-23	119
Serratia marcescens strain AR_0124 chromosome, complete genome	CP028946.1	68	98.5	100	2626543	2626476	1.47E-23	119
Serratia marcescens strain AR_0130 chromosome, complete genome	CP028947.1	68	98.5	100	1076213	1076146	1.47E-23	119
Serratia marcescens strain AR_0123 chromosome, complete genome	CP028948.1	68	98.5	100	432286	432353	1.47E-23	119
Lelliottia sp. WB101 chromosome, complete genome	CP028520.1	68	98.5	100	4474947	4475014	1.47E-23	119
Serratia marcescens strain AR_0099 chromosome, complete genome	CP027539.1	68	98.5	100	5144111	5144178	1.47E-23	119
Serratia marcescens strain AR_0091 chromosome, complete genome	CP027533.1	68	98.5	100	654594	654527	1.47E-23	119
Providencia rettgeri strain FDAARGOS_330 chromosome, complete genome	CP027418.1	68	98.5	100	2281325	2281258	1.47E-23	119
Serratia sp. MYb239 chromosome, complete genome	CP023268.1	68	98.5	100	1041842	1041775	1.47E-23	119
Serratia marcescens strain AR_0027 chromosome, complete genome	CP026702.1	68	98.5	100	1906657	1906724	1.47E-23	119
Serratia marcescens isolate GN26 chromosome	CP026650.1	68	98.5	100	154670	154603	1.47E-23	119
Serratia sp. SSNIH1 chromosome, complete genome	CP026383.1	68	98.5	100	1535943	1535876	1.47E-23	119
Leclercia sp. LSNIH3 chromosome, complete genome	CP026387.1	68	98.5	100	1357297	1357364	1.47E-23	119
Serratia marcescens strain FDAARGOS_65 chromosome, complete genome	CP026050.1	68	98.5	100	1916645	1916712	1.47E-23	119
Serratia marcescens strain SOLR4 chromosome	CP025698.1	68	98.5	100	969565	969632	1.47E-23	119
Brenneria goodwinii strain FRB141, complete genome	CP014137.1	68	98.5	100	2943995	2944062	1.47E-23	119
Serratia sp. JKS000199 genome assembly, chromosome: I	LT907843.1	68	98.5	100	345645	345578	1.47E-23	119
Serratia marcescens strain UMH12, complete genome	CP018930.1	68	98.5	100	364792	364725	1.47E-23	119
Serratia marcescens strain UMH11, complete genome	CP018929.1	68	98.5	100	360747	360680	1.47E-23	119
Serratia marcescens strain UMH10, complete genome	CP018928.1	68	98.5	100	360746	360679	1.47E-23	119

Hit name	Accession	Query length (bp)	Identity (%)	Query coverage (%)	Start position on reference	End position on reference	E value	Max Score
Serratia marcescens strain UMH8, complete genome	CP018927.1	68	98.5	100	382828	382761	1.47E-23	119
Serratia marcescens strain UMH6, complete genome	CP018926.1	68	98.5	100	386125	386058	1.47E-23	119
Serratia marcescens strain UMH3, complete genome	CP018925.1	68	98.5	100	369693	369626	1.47E-23	119
Serratia marcescens strain UMH2, complete genome	CP018924.1	68	98.5	100	417352	417285	1.47E-23	119
Serratia marcescens strain UMH9, complete genome	CP018923.1	68	98.5	100	380715	380648	1.47E-23	119
Serratia marcescens strain UMH7, complete genome	CP018919.1	68	98.5	100	395699	395632	1.47E-23	119
Serratia marcescens strain UMH5, complete genome	CP018917.1	68	98.5	100	408695	408628	1.47E-23	119
Serratia marcescens strain UMH1, complete genome	CP018915.1	68	98.5	100	363421	363354	1.47E-23	119
Serratia marcescens strain S2I7 genome	CP021984.1	68	98.5	100	1144090	1144023	1.47E-23	119
Serratia marcescens strain 1274 genome	CP019927.2	68	98.5	100	1515599	1515666	1.47E-23	119
Serratia marcescens strain B3R3, complete genome	CP013046.2	68	98.5	100	1149111	1149044	1.47E-23	119
Serratia marcescens SMB2099 complete genome	HG738868.1	68	98.5	100	1095265	1095198	1.47E-23	119
Serratia marcescens strain AS1 genome	CP010584.1	68	98.5	100	4850999	4851066	1.47E-23	119
Providencia rettgeri strain RB151, complete genome	CP017671.1	68	98.5	100	868049	867982	1.47E-23	119
Serratia sp. YD25, complete genome	CP016948.1	68	98.5	100	394963	394896	1.47E-23	119
Serratia marcescens isolate PWN146_assembly genome assembly, chromosome: Chromosome	LT575490.1	68	98.5	100	452636	452569	1.47E-23	119
Serratia marcescens strain U36365, complete genome	CP016032.1	68	98.5	100	1047746	1047679	1.47E-23	119
Serratia rubidaea strain 1122, complete genome	CP014474.1	68	98.5	100	4401346	4401413	1.47E-23	119
Leclercia adecarboxylata strain USDA-ARS-USMARC-60222, complete genome	CP013990.1	68	98.5	100	3730319	3730386	1.47E-23	119
Serratia marcescens strain SmUNAM836, complete genome	CP012685.1	68	98.5	100	370223	370156	1.47E-23	119
Serratia marcescens strain RSC-14, complete genome	CP012639.1	68	98.5	100	1044690	1044757	1.47E-23	119
Serratia marcescens strain CAV1492, complete genome	CP011642.1	68	98.5	100	1498860	1498927	1.47E-23	119
Serratia sp. SCBI, complete genome	CP003424.1	68	98.5	100	1052285	1052218	1.47E-23	119
Serratia sp. FS14, complete genome	CP005927.1	68	98.5	100	5080499	5080566	1.47E-23	119
Serratia marcescens SM39 DNA, complete genome	AP013063.1	68	98.5	100	417692	417625	1.47E-23	119
Serratia marcescens subsp. marcescens Db11, complete genome	HG326223.1	68	98.5	100	360344	360277	1.47E-23	119
Serratia marcescens WW4, complete genome	CP003959.1	68	98.5	100	1144090	1144023	1.47E-23	119
Serratia sp. FGI94 chromosome, complete genome	CP003942.1	68	98.5	100	1064623	1064556	1.47E-23	119
Serratia quinivorans strain PKL:12 chromosome, complete genome	CP038467.1	68	97.1	100	2356526	2356593	1.79E-22	114
Serratia plymuthica strain NCTC8900 genome assembly, chromosome: 1	LR134151.1	68	97.1	100	1063709	1063642	1.79E-22	114
Kluyvera intermedia strain NCTC12125 genome assembly, chromosome: 1	LR134138.1	68	97.1	100	2356039	2356106	1.79E-22	114
Pectobacterium carotovorum strain 14A chromosome, complete genome	CP034276.1	68	97.1	100	1344224	1344157	1.79E-22	114
Kosakonia sp. CCTCC M2018092 chromosome, complete genome	CP034225.1	68	97.1	100	3572234	3572301	1.79E-22	114
Serratia sp. P2ACOL2 chromosome, complete genome	CP033162.1	68	97.1	100	883309	883242	1.79E-22	114
Serratia sp. 3ACOL1 chromosome, complete genome	CP033055.1	68	97.1	100	34153	34086	1.79E-22	114
Serratia sp. 1D1416 chromosome, complete genome	CP032738.1	68	97.1	100	2602510	2602443	1.79E-22	114
Pectobacterium parmentieri strain IFB5432 chromosome, complete genome	CP026979.1	68	97.1	100	3441174	3441241	1.79E-22	114
Pectobacterium parmentieri strain IFB5441 chromosome, complete genome	CP026980.1	68	97.1	100	3553937	3554004	1.79E-22	114
Pectobacterium parmentieri strain IFB5485 chromosome, complete genome	CP026981.1	68	97.1	100	3345697	3345764	1.79E-22	114
Pectobacterium parmentieri strain IFB5604 chromosome, complete genome	CP026983.1	68	97.1	100	3327959	3328026	1.79E-22	114
Pectobacterium parmentieri strain IFB5619 chromosome, complete genome	CP026985.1	68	97.1	100	3334636	3334703	1.79E-22	114
Pectobacterium parmentieri strain IFB5623 chromosome	CP026986.1	68	97.1	100	3473452	3473519	1.79E-22	114
Serratia quinivorans strain NCTC13188 genome assembly, chromosome: 1	LR134494.1	68	97.1	100	1123495	1123428	1.79E-22	114
Serratia fonticola strain NCTC13193 genome assembly, chromosome: 1	LR134492.1	68	97.1	100	1363587	1363520	1.79E-22	114
Serratia plymuthica strain NCTC8015 genome assembly, chromosome: 1	LR134478.1	68	97.1	100	1074986	1074919	1.79E-22	114
Dickeya dianthicola strain ME23 chromosome, complete genome	CP031560.1	68	97.1	100	1233169	1233102	1.79E-22	114
Serratia plymuthica strain NCTC12961 genome assembly, chromosome: 1	LS483469.1	68	97.1	100	1103102	1103035	1.79E-22	114

Hit name	Accession	Query length (bp)	Identity (%)	Query coverage (%)	Start position on reference	End position on reference	E value	Max Score
Providencia alcalifaciens strain NCTC10286 genome assembly, chromosome: 1	LS483467.1	68	97.1	100	784759	784692	1.79E-22	114
Dickeya sp. Secpp 1600 chromosome, complete genome	CP023484.1	68	97.1	100	1203841	1203774	1.79E-22	114
Dickeya dadantii strain DSM 18020 chromosome, complete genome	CP023467.1	68	97.1	100	1216020	1215953	1.79E-22	114
Pectobacterium carotovorum strain 3-2 chromosome, complete genome	CP024842.1	68	97.1	100	1296507	1296440	1.79E-22	114
Serratia liquefaciens strain FDAARGOS_125 chromosome, complete genome	CP014017.2	68	97.1	100	1531155	1531222	1.79E-22	114
Dickeya fangzhongdai strain DSM 101947 chromosome, complete genome	CP025003.1	68	97.1	100	1290272	1290205	1.79E-22	114
Pectobacterium atrosepticum strain 36A chromosome, complete genome	CP024956.1	68	97.1	100	1308052	1307985	1.79E-22	114
Serratia grimesii isolate BXF1 genome assembly, chromosome	LT883155.1	68	97.1	100	1091788	1091721	1.79E-22	114
Pectobacterium carotovorum subsp. brasiliense strain BZA12 chromosome, complete genome	CP024780.1	68	97.1	100	2024292	2024359	1.79E-22	114
Dickeya dianthicola RNS04.9, complete genome	CP017638.1	68	97.1	100	868234	868167	1.79E-22	114

Supplementary Data 6 – Isolation, radiocarbon and archaeological dates used for BEAST v1.8 substitution rate variation analysis. *Years BP (YBP) refers to the time until year 2005, which is the most recently isolated modern strain

Strain Name	Strain ID	Pubication / NCBI Accession	Year of isolation (modern strains)	Years BP*	Lower bound (YBP*)	Upper bound (YBP*)
1.ANTa	Antiqua	Chain et al., 2006	1965	40	N/A	N/A
1.ANTb	UG05	Morelli et al., 2010	2004	1	N/A	N/A
1.IN1a	CMCC11001	Cui et al., 2013	1954	51	N/A	N/A
1.IN1b	780441	Cui et al., 2013	1978	27	N/A	N/A
1.IN1c	K21985002	Cui et al., 2013	1985	20	N/A	N/A
1.IN2a	CMCC640047	Cui et al., 2013	1964	41	N/A	N/A
1.IN2b	30017	Cui et al., 2013	1976	29	N/A	N/A
1.IN2c	CMCC31004	Cui et al., 2013	1990	15	N/A	N/A
1.IN2d	C1975003	Cui et al., 2013	1975	30	N/A	N/A
1.IN2e	C1989001	Cui et al., 2013	1989	16	N/A	N/A
1.IN2f	710317	Cui et al., 2013	1971	34	N/A	N/A
1.IN2g	CMCC05013	Cui et al., 2013	1988	17	N/A	N/A
1.IN2h	5	Cui et al., 2013	2004	1	N/A	N/A
1.IN2i	CMCC10012	Cui et al., 2013	1964	41	N/A	N/A
1.IN2j	CMCC27002	Cui et al., 2013	1991	14	N/A	N/A
1.IN2k	970754	Cui et al., 2013	1997	8	N/A	N/A
1.IN2l	D1991004	Cui et al., 2013	1991	14	N/A	N/A
1.IN2m	D1964002	Cui et al., 2013	1964	41	N/A	N/A
1.IN2n	CMCC02041	Cui et al., 2013	1965	40	N/A	N/A
1.IN2o	CMCC03001	Cui et al., 2013	1954	51	N/A	N/A
1.IN2p	D1982001	Cui et al., 2013	1982	23	N/A	N/A
1.IN2q	D1964001	Cui et al., 2013	1964	41	N/A	N/A
1.IN3a	F1954001	Cui et al., 2013	1954	51	N/A	N/A
1.IN3b	E1979001	Cui et al., 2013	1979	26	N/A	N/A
1.IN3c	CMCC84038	Cui et al., 2013	1982	23	N/A	N/A
1.IN3d	YN1683	Cui et al., 2013	1977	28	N/A	N/A
1.IN3e	YN472	Cui et al., 2013	1957	48	N/A	N/A
1.IN3f	YN1065	Cui et al., 2013	1954	51	N/A	N/A
1.IN3g	E1977001	Cui et al., 2013	1977	28	N/A	N/A
1.IN3h	CMCC84033	Cui et al., 2013	1979	26	N/A	N/A
1.IN3i	CMCC84046	Cui et al., 2013	1984	21	N/A	N/A
1.ORI1a	CMCC114001	Cui et al., 2013	1952	53	N/A	N/A
1.ORI1b	India195	Morelli et al., 2010	1898	107	N/A	N/A
1.ORI1c	F1946001	Cui et al., 2013	1946	59	N/A	N/A
1.ORI1d	CA88	Auerbach et al., 2007	1988	17	N/A	N/A
1.ORI1e	CO92	Parkhill et al., 2001	1992	13	N/A	N/A
1.ORI2a	YN2179	Cui et al., 2013	1995	10	N/A	N/A
1.ORI2b	CMCCK110001	Cui et al., 2013	1991	14	N/A	N/A
1.ORI2c	YN2551	Cui et al., 2013	2002	3	N/A	N/A
1.ORI2d	YN2588	Cui et al., 2013	2000	5	N/A	N/A
1.ORI2e	F1991016	Eppinger et al., 2009	1991	14	N/A	N/A
1.ORI2f	CMCC87001	Cui et al., 2013	1982	23	N/A	N/A
1.ORI2g	F1984001	Cui et al., 2013	1984	21	N/A	N/A
1.ORI2h	YN663	Cui et al., 2013	1982	23	N/A	N/A
1.ORI2i	CMCCK100001	Cui et al., 2013	1991	14	N/A	N/A
1.ORI3a	EV76	Cui et al., 2013	1922	83	N/A	N/A
1.ORI3b	MG05	Morelli et al., 2010	2005	0	N/A	N/A
1.ORI3c	IP275	Morelli et al., 2010	1995	10	N/A	N/A
2.MED1a	KIM	Cui et al., 2013	1968	37	N/A	N/A
ancient Branch 1	Laishevo LAI009	this study	N/A	655	605	705
ancient Branch 1	London BD	Bos et al., 2011	N/A	656	655	657
ancient Branch 1	Barcelona 3031	Spyrou et al., 2016	N/A	645	585	705
ancient Branch 1	Nabburg NAB003	this study	N/A	663	613	713
ancient Branch 1	Bolgar 2370	Spyrou et al., 2016	N/A	624	605	643
ancient Branch 1	Bergen 37/45	Namouchi et al., 2018	N/A	645	642	647
ancient Branch 1	London 6330	Bos et al., 2011	N/A	630	605	655
ancient Branch 1	Manching MAN008	this study	N/A	669	615	722
ancient Branch 1	Starnberg STA001	this study	N/A	527	482	572
ancient Branch 1	Cambridge NMS002	this study	N/A	500	469	530
ancient Branch 1	Ellwangen 549 O	Spyrou et al., 2016	N/A	449	378	520
ancient Branch 1	Landsberg LBG002	this study	N/A	462	373	550
ancient Branch 1	Stans STN014, STN020, STN021, STN019, STN007, STN002, STN008, STN013	this study	N/A	445	370	520
ancient Branch 1	Brandenburg BRA001	this study	N/A	372	357	387
ancient Branch 1	New Churchyard BED030, BED028, BED034, BED024	this study	N/A	408	370	445
ancient Branch 1	Marseille OBS137, OBS116, OBS107, OBS110, OBS124	Bos et al., 2016	N/A	284	283	285

Supplementary Data 7 – Gene annotation and effects of SNPs located on the branch between the Ellwangen (ELW098) and New Churchyard (BED) strains

Position	Reference	SNP	Effect	Gene ID	Gene Name	Gene function	old AA/ new AA	Old codon/ New codon	Codon Num (CDS)
1481292	C	T	DOWNSTREAM: 44 bases	YPO1316	YPO1316				
480773	C	T	INTERGENIC						
482327	G	T	INTERGENIC						
1481381	G	A	INTERGENIC						
1481393	G	A	INTERGENIC						
2671194	G	A	INTERGENIC						
2918297	T	G	INTERGENIC						
2964936	A	G	INTERGENIC						
4134121	A	T	INTERGENIC						
4190286	C	A	INTERGENIC						
4208536	A	G	INTERGENIC						
4456212	C	A	INTERGENIC						
4642828	G	A	INTERGENIC						
100383	C	T	NON_SYNONYMOUS_ CODING	YPO0090	glpK	pseudogene	E/K	Gaa/Aaa	325
200723	C	T	NON_SYNONYMOUS_ CODING	YPO0182	tauA	taurine transporter substrate binding subunit	T/M	aCg/aTg	135
868549	G	C	NON_SYNONYMOUS_ CODING	YPO0792	ygeD	lysophospholipid transporter LplT	C/W	tgC/tgG	209
965281	C	A	NON_SYNONYMOUS_ CODING	YPO0880	YPO0880	primase	S/Y	tCt/tAt	95
1168951	G	T	NON_SYNONYMOUS_ CODING	YPO1028	YPO1028	cysteine sulfinate desulfinase	R/S	Cgt/Agt	143
1189479	C	T	NON_SYNONYMOUS_ CODING	YPO1048	dxr	1-deoxy-D-xylulose 5-phosphate reductoisomerase	P/S	Cca/Tca	189
1378105	G	T	NON_SYNONYMOUS_ CODING	YPO1219	YPO1219	hypothetical protein	H/N	Cat/Aat	593
1451124	T	G	NON_SYNONYMOUS_ CODING	YPO1291	YPO1291	carbohydrate kinase	V/G	gTg/gGg	394
1511518	A	G	NON_SYNONYMOUS_ CODING	YPO1347	YPO1347	hypothetical protein	N/D	Aat/Gat	167
1586982	C	A	NON_SYNONYMOUS_ CODING	YPO1403	mukF	condesin subunit F	S/Y	tCc/tAc	163
1708192	C	A	NON_SYNONYMOUS_ CODING	YPO1504	YPO1504	hypothetical protein	P/Q	cCa/cAa	285
1935112	C	A	NON_SYNONYMOUS_ CODING	YPO1696	YPO1696	outer membrane usher protein	Q/H	caG/caT	208
2414599	T	C	NON_SYNONYMOUS_ CODING	YPO2145	YPO2145	SpoVR family protein	K/E	Aaa/Gaa	85
2472383	A	G	NON_SYNONYMOUS_ CODING	YPO2196	ispZ	Involved in cell division	F/S	tTc/tCc	85
2507983	T	G	NON_SYNONYMOUS_ CODING	YPO2233	YPO2233	hypothetical protein	T/P	Acc/Ccc	72
3229407	T	C	NON_SYNONYMOUS_ CODING	YPO2887	yapB	pseudogene	V/A	gTt/gCt	374
3407572	A	T	NON_SYNONYMOUS_ CODING	YPO3049	YPO3049	binding protein-dependent transporter inner membrane protein	L/Q	cTg/cAg	216
3610371	C	T	NON_SYNONYMOUS_ CODING	YPO3244	fadE	acyl-CoA dehydrogenase	A/V	gCc/gTc	292
3613964	C	A	NON_SYNONYMOUS_ CODING	YPO3246	hmwC	accessory processing protein	D/Y	Gat/Tat	252
3620114	G	A	NON_SYNONYMOUS_ CODING	YPO3247	hmwA	adhesin	A/V	gCg/gTg	164
3782640	G	A	NON_SYNONYMOUS_ CODING	YPO3389	hemL	glutamate-1-semialdehyde aminotransferase	G/D	gGc/gAc	31
3973901	G	A	NON_SYNONYMOUS_ CODING	YPO3559	YPO3559	hypothetical protein	D/N	Gac/Aac	145
4363505	C	T	NON_SYNONYMOUS_ CODING	YPO3888	ilvC	ketol-acid reductoisomerase	V/I	Gtt/Att	118
4396236	G	T	NON_SYNONYMOUS_ CODING	YPO3914	sthA	catalyzes the conversion of NADPH to NADH	D/Y	Gat/Tat	116
4616904	T	C	NON_SYNONYMOUS_ CODING	YPO4095	recF	recombination protein F: required for DNA replication	S/G	Agt/Ggt	54
2071670	G	T	STOP_GAINED	YPO1826	fliJ	flagellar biosynthesis chaperone	S/*	tCa/tAa	142

Position	Reference	SNP	Effect	Gene ID	Gene Name	Gene function	old AA/ new AA	Old codon/ New codon	Codon Num (CDS)
173032	C	T	SYNONYMOUS_ CODING	YPO0158	cysG	siroheme synthase	K/K	aaG/aaA	178
477107	C	T	SYNONYMOUS_ CODING	YPO0452	slt	lytic murein transglycosylase	L/L	Ctg/Ttg	364
869820	A	G	SYNONYMOUS_ CODING	YPO0793	aas	acyl-ACP synthetase	F/F	ttT/ttC	503
951295	C	T	SYNONYMOUS_ CODING	YPO0863	YPO0863	hypothetical protein	L/L	Ctg/Ttg	182
1159539	T	A	SYNONYMOUS_ CODING	YPO1020	recB	helicase/nuclease	P/P	ccT/ccA	1181
1724647	C	T	SYNONYMOUS_ CODING	YPO1517	YPO1517	sugar ABC transporter	S/S	tcG/tcA	130
2076353	C	T	SYNONYMOUS_ CODING	YPO1830	fliF	flagellar MS-ring protein	T/T	acG/acA	223
3620500	G	A	SYNONYMOUS_ CODING	YPO3247	hmwA	adhesin	G/G	ggC/ggT	35
3944305	C	A	SYNONYMOUS_ CODING	YPO3531	YPO3531	iron-sulfur cluster repair di-iron protein	I/I	atC/atA	81
4200639	C	A	SYNONYMOUS_ CODING	YPO3746	rpoC	DNA-directed RNA polymerase subunit beta	L/L	ctG/ctT	282

Supplementary Data 8 – Gene annotations and effects of variants unique to 1.ANT strains

Position	Reference	SNP	SNP Effect	Gene ID	Gene name	Gene function	old AA/ new AA	Old codon/ New codon	Codon # (CDS)
1166	C	T	NON_SYNONYMOUS_CODING	YPO0002	asnC	DNA-binding transcriptional regulator AsnC	A/T	Gct/Act	34
54102	A	T	NON_SYNONYMOUS_CODING	YPO0041	ligB	NAD-dependent DNA ligase LigB	T/S	Acc/Tcc	106
178028	C	T	NON_SYNONYMOUS_CODING	YPO0162	codA	cytosine deaminase	A/V	gCg/gTg	63
225409	C	T	NON_SYNONYMOUS_CODING	YPO0216	rpsC	30S ribosomal protein S3	T/I	aCt/aTt	121
274360	C	T	NON_SYNONYMOUS_CODING	YPO0274	YPO0274	hypothetical protein	P/L	cCg/cTg	174
419209	G	A	NON_SYNONYMOUS_CODING	YPO0401	YPO0401	transcriptional regulator	D/N	Gat/Aat	289
525887	T	A	NON_SYNONYMOUS_CODING	YPO0494	surA	peptidyl-prolyl cis-trans isomerase SurA	Q/L	cAg/cTg	58
535221	T	G	NON_SYNONYMOUS_CODING	YPO0501	YPO0501	hypothetical protein	S/A	Tcc/Gcc	313
574078	C	T	NON_SYNONYMOUS_CODING	YPO0530	leuD	3-isopropylmalate dehydratase small subunit	G/S	Ggc/Agc	53
727192	G	A	NON_SYNONYMOUS_CODING	YPO0668	parE	DNA topoisomerase IV subunit B	A/T	Gcc/Acc	53
759843	C	T	NON_SYNONYMOUS_CODING	YPO0698	YPO0698	outer membrane usher protein	L/F	Ctc/Ttc	472
935991	G	A	NON_SYNONYMOUS_CODING	YPO0850a	YPO0850a	PTS system glucose/sucrose specific transporter subunit IIB	T/I	aCt/aTt	21
1151889	C	T	NON_SYNONYMOUS_CODING	YPO1018	recC	catalyzes ATP-dependent exonucleolytic cleavage	P/L	cCt/cTt	744
1212998	C	T	NON_SYNONYMOUS_CODING	YPO1069	YPO1069	hypothetical protein	G/S	Ggc/Agc	70
1265901	A	G	NON_SYNONYMOUS_CODING	YPO1118	cydB	cytochrome D ubiquinol oxidase subunit II	N/D	Aat/Gat	112
1303163	A	T	NON_SYNONYMOUS_CODING	YPO1156	uvrB	excinuclease ABC subunit B	D/V	gAt/gTt	150
1333620	C	T	NON_SYNONYMOUS_CODING	YPO1185	YPO1185	ABC transport membrane permease	T/I	aCt/aTt	5
1441794	C	T	NON_SYNONYMOUS_CODING	YPO1283	uxuA	mannonate dehydratase	A/V	gCt/gTt	195
1555356	G	A	NON_SYNONYMOUS_CODING	YPO1380	YPO1380	MFS family transporter protein	D/N	Gat/Aat	373
1705844	C	T	NON_SYNONYMOUS_CODING	YPO1502	YPO1502	alcohol dehydrogenase	A/T	Gcc/Acc	71
1707708	G	A	NON_SYNONYMOUS_CODING	YPO1504	YPO1504	hypothetical protein	G/S	Ggc/Agc	124
1756795	G	A	NON_SYNONYMOUS_CODING	YPO1541	gnd	6-phosphogluconate dehydrogenase	G/S	Ggc/Agc	174
1801784	G	A	NON_SYNONYMOUS_CODING	YPO1579	YPO1579	C4-dicarboxylate transporter substrate-binding protein	P/S	Ccc/Tcc	44
1850729	C	T	NON_SYNONYMOUS_CODING	YPO1626	lolC	outer membrane-specific lipoprotein transporter subunit LolC	P/S	Ccg/Tcg	81
2016117	T	A	NON_SYNONYMOUS_CODING	YPO1770	hpaC	4-hydroxyphenylacetate 3-monooxygenase coupling protein	L/Q	cTg/cAg	135
2115661	C	T	NON_SYNONYMOUS_CODING	YPO1868	YPO1868	hypothetical protein	E/K	Gaa/Aaa	161
2484440	A	G	NON_SYNONYMOUS_CODING	YPO2210	YPO2210	hypothetical protein	V/A	gTg/gCg	18
2681696	T	C	NON_SYNONYMOUS_CODING	YPO2387	purR	DNA-binding transcriptional repressor PurR	F/S	tTt/tCt	221
2731263	C	T	NON_SYNONYMOUS_CODING	YPO2429	pheS	phenylalanyl-tRNA synthetase subunit alpha	V/I	Gta/Ata	67
2907502	G	T	NON_SYNONYMOUS_CODING	YPO2585	YPO2585	carbohydrate kinase	D/Y	Gat/Tat	361
2924187	G	A	NON_SYNONYMOUS_CODING	YPO2602	rlpA	rare lipoprotein A	P/L	cCa/cTa	36

Position	Reference	SNP	SNP Effect	Gene ID	Gene name	Gene function	old AA/ new AA	Old codon/ New codon	Codon # (CDS)
2950840	C	T	NON_SYNONYMOUS_CODING	YPO2625	nagC	N-acetylglucosamine regulatory protein	G/E	gGg/gAg	61
3099378	G	A	NON_SYNONYMOUS_CODING	YPO2765	asd	semialdehyde dehydrogenase	R/H	cGt/cAt	76
3501556	A	C	NON_SYNONYMOUS_CODING	YPO3142	amtB	ammonium transporter	V/G	gTc/gGc	324
3586778	G	A	NON_SYNONYMOUS_CODING	YPO3222	proB	gamma-glutamyl kinase	S/F	tCt/tTt	50
3698745	G	A	NON_SYNONYMOUS_CODING	YPO3316	rbsC	sugar transport system permease	L/F	Ctc/Ttc	269
3763970	T	C	NON_SYNONYMOUS_CODING	YPO3375	sodC	superoxide dismutase	D/G	gAt/gGt	33
3771852	T	A	NON_SYNONYMOUS_CODING	YPO3381	barA	hybrid sensory histidine kinase BarA	L/Q	cTg/cAg	156
3904960	G	A	NON_SYNONYMOUS_CODING	YPO3496	infB	translation initiation factor IF-2	A/V	gCg/gTg	614
3977650	G	C	NON_SYNONYMOUS_CODING	YPO3564	YPO3564	hypothetical protein	D/E	gaC/gaG	90
4061493	C	T	NON_SYNONYMOUS_CODING	YPO3640	YPO3640	hypothetical protein	P/S	Ccc/Tcc	29
4065904	C	T	NON_SYNONYMOUS_CODING	YPO3642b	YPO3642b	ncRNA-cspA thermoregulator	A/V	gCt/gTt	134
4076324	C	T	NON_SYNONYMOUS_CODING	YPO3657	panF	sodium/panthothenate symporter	S/N	aGc/aAc	371
4318701	T	C	NON_SYNONYMOUS_CODING	YPO3848	cyaA	adenylate cyclase	D/G	gAt/gGt	157
4360567	C	T	NON_SYNONYMOUS_CODING	YPO3887	YPO3887	pseudogene	R/H	cGc/cAc	544
4427178	C	T	NON_SYNONYMOUS_CODING	YPO3940	glgC	glucose-1-phosphate adenylyltransferase	S/N	aGc/aAc	404
4477698	G	A	NON_SYNONYMOUS_CODING	YPO3976	YPO3976	hypothetical protein	A/T	Gca/Aca	237
4525085	A	G	NON_SYNONYMOUS_CODING	YPO4013	yhjW	phosphoethanolamine transferase	S/P	Tca/Cca	35
4558980	G	A	NON_SYNONYMOUS_CODING	YPO4042	YPO4042	fimbrial protein	A/V	gCa/gTa	752
588988	C	A	START_LOST	YPO0544	YPO0544	hypothetical protein	M/I	atG/atT	1
2016126	T	A	STOP_GAINED	YPO1770	hpaC	4-hydroxyphenylacetate 3-monooxygenase coupling protein	L/*	tTg/tAg	138
2751771	G	A	STOP_GAINED	YPO2451	YPO2451	hypothetical protein	Q/*	Cag/Tag	68
168535	A	G	SYNONYMOUS_CODING	YPO0154	dam	DNA adenine methylase	R/R	cgA/cgG	22
367638	C	T	SYNONYMOUS_CODING	YPO0357	frdD	fumarate reductase subunit D	L/L	ttG/ttA	107
537476	C	T	SYNONYMOUS_CODING	YPO0504	YPO0504	hypothetical protein	A/A	gcC/gcT	150
586040	C	T	SYNONYMOUS_CODING	YPO0540	ilvH	acetolactate synthase small subunit	G/G	ggC/ggT	98
904076	C	T	SYNONYMOUS_CODING	YPO0824	YPO0824	hypothetical protein	L/L	Ctg/Ttg	433
1028008	G	A	SYNONYMOUS_CODING	YPO0935	gshB	glutathione synthetase	G/G	ggG/ggA	271
1200873	C	T	SYNONYMOUS_CODING	YPO1059	dnaE	DNA polymerase III subunit alpha	L/L	Ctg/Ttg	10
1578826	C	T	SYNONYMOUS_CODING	YPO1395	msbA	lipid transporter ATP-binding protein/permease	L/L	ctC/ctT	575
1748342	G	A	SYNONYMOUS_CODING	YPO1536	YPO1536	iron-siderophore transporter substrate-binding protein	L/L	Ctg/Ttg	120
2112974	G	A	SYNONYMOUS_CODING	YPO1866	uvrC	excinuclease ABC subunit C	G/G	ggG/ggA	183
2542828	A	G	SYNONYMOUS_CODING	YPO2262	YPO2262	hypothetical protein	I/I	atT/atC	233
2596092	C	T	SYNONYMOUS_CODING	YPO2308	rstA	DNA-binding transcriptional regulator RstA	R/R	cgC/cgT	214
3154240	G	A	SYNONYMOUS_CODING	YPO2827	upp	uracil phosphoribosyltransferase	I/I	atC/atT	76
3348338	C	T	SYNONYMOUS_CODING	YPO2998	YPO2998	two-component system response regulator	G/G	ggG/ggA	185
3371630	T	C	SYNONYMOUS_CODING	YPO3016	nanT	sialic acid transporter	K/K	aaA/aaG	25
3607609	A	G	SYNONYMOUS_CODING	YPO3241	yafK	hypothetical protein	K/K	aaA/aaG	205
3624499	A	T	SYNONYMOUS_CODING	YPO3249	YPO3249	allantoate amidohydrolase	T/T	acT/acA	106

Position	Reference	SNP	SNP Effect	Gene ID	Gene name	Gene function	old AA/ new AA	Old codon/ New codon	Codon # (CDS)
4201619	T	C	SYNONYMOUS_ CODING	YPO3747	rpoB	DNA-directed RNA polymerase subunit beta	E/E	gaA/gaG	1341
1578826	C	T	UPSTREAM: 21 bases	YPO1396	lpxK	tetraacyldisaccharide 4'-kinase			
3634070	A	T	UPSTREAM: 24 bases	YPO3262	YPO3262	hypothetical protein			
1306347	C	T	UPSTREAM: 26 bases	YPO1158	YPO1158	hypothetical protein			
1607551	C	T	UPSTREAM: 27 bases	YPO1415	pyrD	dihydroorotate dehydrogenase 2			
990954	G	A	UPSTREAM: 44 bases	YPO0904	YPO0904	hypothetical protein			
4256238	A	G	UPSTREAM: 47 bases	YPO3790	yigM	hypothetical protein			
1216366	C	T	UPSTREAM: 83 bases	YPO1073	metN	DL-methionine transporter ATP-binding protein			
4343118	G	A	UPSTREAM: 96 bases	YPO3867	rho	transcription termination factor Rho			
867596	C	T	UPSTREAM: 97 bases	YPO0790a	YPO0790a	hypothetical protein			
1333620	C	T	DOWNSTREAM: 10 bases	YPO1184	YPO1184	ABC transport membrane permease			
511653	G	T	DOWNSTREAM: 19 bases	YPO0480	dapB	4-hydroxy-tetrahydrodipicolinate reductase			
1037476	C	T	DOWNSTREAM: 3 bases	YPO0947	YPO0947	virulence determinant			
419209	G	A	DOWNSTREAM: 4 bases	YPO0402	YPO0402	PTS system fructose-family transporter subunit IIB			
3634070	A	T	DOWNSTREAM: 5 bases	YPO3261	YPO3261	amidase			
2413154	C	T	DOWNSTREAM: 56 bases	YPO2144	fadR	fatty acid metabolism regulator			
4123916	G	T	DOWNSTREAM: 66 bases	YPO3687	YPO3687	ribonuclease			
2828838	C	T	DOWNSTREAM: 90 bases	YPO2519	YPO2519	SAM-dependent methyltransferase			
3607609	A	G	DOWNSTREAM: 91 bases	YPO3242	yafK	hypothetical protein			
309324	C	T	INTERGENIC						
511653	G	T	INTERGENIC						
867596	C	T	INTERGENIC						
990954	G	A	INTERGENIC						
1037476	C	T	INTERGENIC						
1216366	C	T	INTERGENIC						
1306347	C	T	INTERGENIC						
1344950	A	G	INTERGENIC						
1607551	C	T	INTERGENIC						
2018367	G	A	INTERGENIC						
2413154	C	T	INTERGENIC						
2828838	C	T	INTERGENIC						
3634070	A	T	INTERGENIC						
3784042	T	A	INTERGENIC						
4123916	G	T	INTERGENIC						
4147421	G	T	INTERGENIC						
4256238	A	G	INTERGENIC						
4343118	G	A	INTERGENIC						
4393272	C	T	INTERGENIC						

Supplementary Data 9 – Gene annotations and effects of SNPs located on the branch between the 1.IN and 1.ORI strains

Position	Reference	SNP	Effect	Gene_ID	Gene Name	Gene function	old AA/ new AA	Old codon/ New codon	Codon # (CDS)
286528	T	A	SYNONYMOUS_ CODING	YPO0285	YPO0285	hypothetical protein	R/R	cgA/cgT	46
699647	T	C	SYNONYMOUS_ CODING	YPO0643	rpoD	RNA polymerase sigma factor RpoD	Q/Q	caA/caG	468
1735263	A	C	SYNONYMOUS_ CODING	YPO1526	YPO1526	assembly protein (YegA)	A/A	gcA/gcC	307
1749443	T	C	SYNONYMOUS_ CODING	YPO1537	YPO1537	iron-siderophore receptor	P/P	ccA/ccG	536
2575152	G	A	SYNONYMOUS_ CODING	YPO2291	YPO2291	virulence factor	L/L	ttG/ttA	282
2739149	C	A	SYNONYMOUS_ CODING	YPO2439	yfeA	substrate-binding protein	T/T	acC/acA	137
4082562	T	C	SYNONYMOUS_ CODING	YPO3662	YPO3662	sulfite oxidase subunit YedY	Q/Q	caA/caG	294
4518401	G	A	SYNONYMOUS_ CODING	YPO4007	uhpC	membrane protein regulates uhpT expression	N/N	aaC/aaT	270
1025278	T	G	NON_SYNONYMOUS_ CODING	YPO0932	YPO0932	hypothetical protein	S/A	Tct/Gct	137
1098675	A	C	NON_SYNONYMOUS_ CODING	YPO0989	iucA	pseudogene	S/R	Agc/Cgc	380
2508389	T	C	NON_SYNONYMOUS_ CODING	YPO2234	cstA	carbon starvation protein A	T/A	Aca/Gca	623
2744933	A	G	NON_SYNONYMOUS_ CODING	YPO2446	YPO2446	YniC; 2-deoxyglucose-6-phosphatase	I/V	Att/Gtt	97
2903882	T	G	NON_SYNONYMOUS_ CODING	YPO2582	YPO2582	sugar transport ATP-binding protein	C/G	Tgt/Ggt	413
2936268	G	A	NON_SYNONYMOUS_ CODING	YPO2614	gltJ	glutamate/aspartate transport system permease	L/F	Ctt/Ttt	82
3085079	A	G	NON_SYNONYMOUS_ CODING	YPO2752	mepA	D-alanyl-D-alanine endopeptidase	T/A	Act/Gct	85
3362591	A	G	NON_SYNONYMOUS_ CODING	YPO3009	YPO3009	two-component response regulator	S/G	Agc/Ggc	59
3421335	A	G	NON_SYNONYMOUS_ CODING	YPO3064	bcp	bacterioferritin comigratory protein	M/V	Atg/Gtg	136
3564026	C	T	NON_SYNONYMOUS_ CODING	YPO3201	proY	permease	C/Y	tGt/tAt	84
3616733	A	G	NON_SYNONYMOUS_ CODING	YPO3247	hmwA	adhesin	L/P	cTa/cCa	1291
4194600	G	A	NON_SYNONYMOUS_ CODING	YPO3742	thiG	thiazole synthase	V/I	Gtt/Att	98
4421633	T	C	NON_SYNONYMOUS_ CODING	YPO3937	glpD	glycerol-3-phosphate dehydrogenase	V/A	gTg/gCg	125
2277583	G	A	INTERGENIC						
2684793	A	G	INTERGENIC						
3324959	A	G	DOWNSTREAM: 24 bases	YPO2977	glk	glucokinase			

Supplementary References

1 Huebler, R. *et al.* HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *bioRxiv*, 534198 (2019).

2 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).

3 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).

4 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).

5 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).

6 Kislichkina, A. A. *et al.* Six Whole-Genome Assemblies of Yersinia pestis subsp. microtus bv. ulegeica (Phylogroup 0. PE5) Strains Isolated from Mongolian Natural Plague Foci. *Genome announcements* **6**, e00536-00518 (2018).

7 Kislichkina, A. A. *et al.* Nine Whole-Genome Assemblies of Yersinia pestis subsp. microtus bv. Altaica Strains Isolated from the Altai Mountain Natural Plague Focus (No. 36) in Russia. *Genome announcements* **6**, e01440-01417 (2018).

8 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).

9 Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proceedings of the National Academy of Sciences* **110**, 577-582 (2013).

10 Feldman, M. *et al.* A high-coverage Yersinia pestis genome from a sixth-century Justinianic plague victim. *Mol. Biol. Evol.* **33**, 2911-2923 (2016).

11 de Barros Damgaard, P. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369 (2018).

12 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).

13 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).

14 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, doi:10.1073/pnas.0404012101 (2004).

15 R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2015).

16 Namouchi, A. *et al.* Integrative approach using Yersinia pestis genomes to revisit the historical landscape of plague during the Medieval Period. *Proceedings of the National Academy of Sciences*, 201812865 (2018).

17 Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14**, 178-192 (2013).

18 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).

19 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).

20 Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in Yersinia pestis. *Infect. Immun.* **77**, 2242-2250, doi:10.1128/IAI.01477-08 (2009).
21 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).
22 Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).
23 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).
24 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).
25 Wiechmann, I., Harbeck, M. & Grupe, G. Yersinia pestis DNA sequences in late medieval skeletal finds, Bavaria. *Emerging Infect. Dis.* **16**, 1806 (2010).
26 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).
27 Seifert, L. *et al.* Strategy for sensitive and specific detection of Yersinia pestis in skeletons of the Black Death pandemic. *Plos one* **8**, e75742 (2013).
28 Garrelt, C. & Wiechmann, I. Detection of Yersinia pestis DNA in early and late medieval Bavarian burials. *Decyphering ancient bones; the research potential of bioarchaeological collections. Documenta Archaeobiologiae*, 247-254 (2003).
29 Paya, D. & Catalo, J. *Le cimetière Saint-Michel de Toulouse*. (CNRS, 2011).
30 Hartle, R., Carty, N., Henderson, M., Knox E. L., Walker, D. The New Churchyard: from Moorfields marsh to Bethlem burial ground, Brokers Row and Liverpool Street. *Crossrail Archaeol Ser* **10** (2017).
31 Keily, J. *Tunnel: the archaeology of Crossrail*. (2017).
32 Cressford, C. Former Old Examination Hall, North Range Buildings, New Museums Site, Cambridge: an archaeological excavation *Cambridge Archaeological Unit report no. 1377* (2017).
33 Hensch, M. *Sankt Johans Freidhof in Nabburg - Gewöhnliche und ungewöhnliche Einblicke in die spätmittelalterliche Begräbniskultur Ostbayerns.*, 423–440 (2014).
34 Garrelt, C. *Molekulargenetische Untersuchung der Bestatteten eines vermuteten Pestfriedhofs des 14. Jahrhunderts (Manching-Pichl)*, Diplomarbeit LMU München, (2002).
35 Later, C. Merowingerzeitliche Tuffplattengräber und frühmittelalterliche Kirchenbauten – Zu den Anfängen der ehemaligen Pfarrkirche St. Benedikt in Starnberg. *Bericht Der Bayerischen Bodendenkmalpflege* **52**, 373–402 (2010).
36 Lichtenstern, A. Landberg am Lech. *Geschichte und Kultur, Mering* (2012).
37 Schreiber, J., Carlichi-Witjes, N., von Heyking, K. & Immler, F. 1507-1806 : Hunderte Gräber vom Kirchhof St. Johannis in Landsberg am Lech. *Das Archäologische Jahr in Bayern*, 181–182 (2016).
38 Schreiber, J. Grabungen im Neuzeitlichen Kirchhof St. Johannis an der Brudergasse. *Dokumenta historiae* **18**, 105-145 (2017).
39 Krämer, D. Bevölkerung und Wegenetz: Leben in Abgeschiedenheit. *Geschichte des Kantons Nidwalden. Von der Urzeit bis 1850. Band 1.*, 160-172 (2014).
40 Homberger, V. R., K. Stans NW, Kreuzung Nägeligasse /Knirigasse. *Jahrbuch Archäologie Schweiz* **100**, 276-277 (2017).
41 Dalitz, S., Grupe, G. & Jungklaus, B. Das kleinste Massengrab Brandenburgs. Drei Tote aus dem Dreißigjährigen Krieg auf der Dominsel der Stadt Brandenburg an der Havel. *Historischer Verein Brandenburg (Havel) eV* **21**, 2011-2012 (2012).

Supplementary Material for Manuscript C

Manuscript C: Ancient *Yersinia pestis* genomes provide no evidence for the origins or spread of the Justinianic Plague

Marcel Keller, Maria A. Spyrou, Michael McCormick, Kirsten I. Bos, Alexander Herbig, Johannes Krause

Submitted to *bioRxiv* on 31st of October 2019, revised on 12th of November 2019, doi:10.1101/819698

Supplementary Information

Fig. S1: Maximum likelihood tree based on 3885 SNPs of 167 modern and 19 ancient genomes. Main branches are collapsed for clarity, numbers on nodes indicate bootstrap support. Highlighted are the Justinianic genome from Altenerding (green), the investigated Tian Shan genome DA101 (blue). DA147 and its artificial branch are shown in red, all branches relevant for the positioning of DA147 are shown in light blue (see Fig. 1A and C).

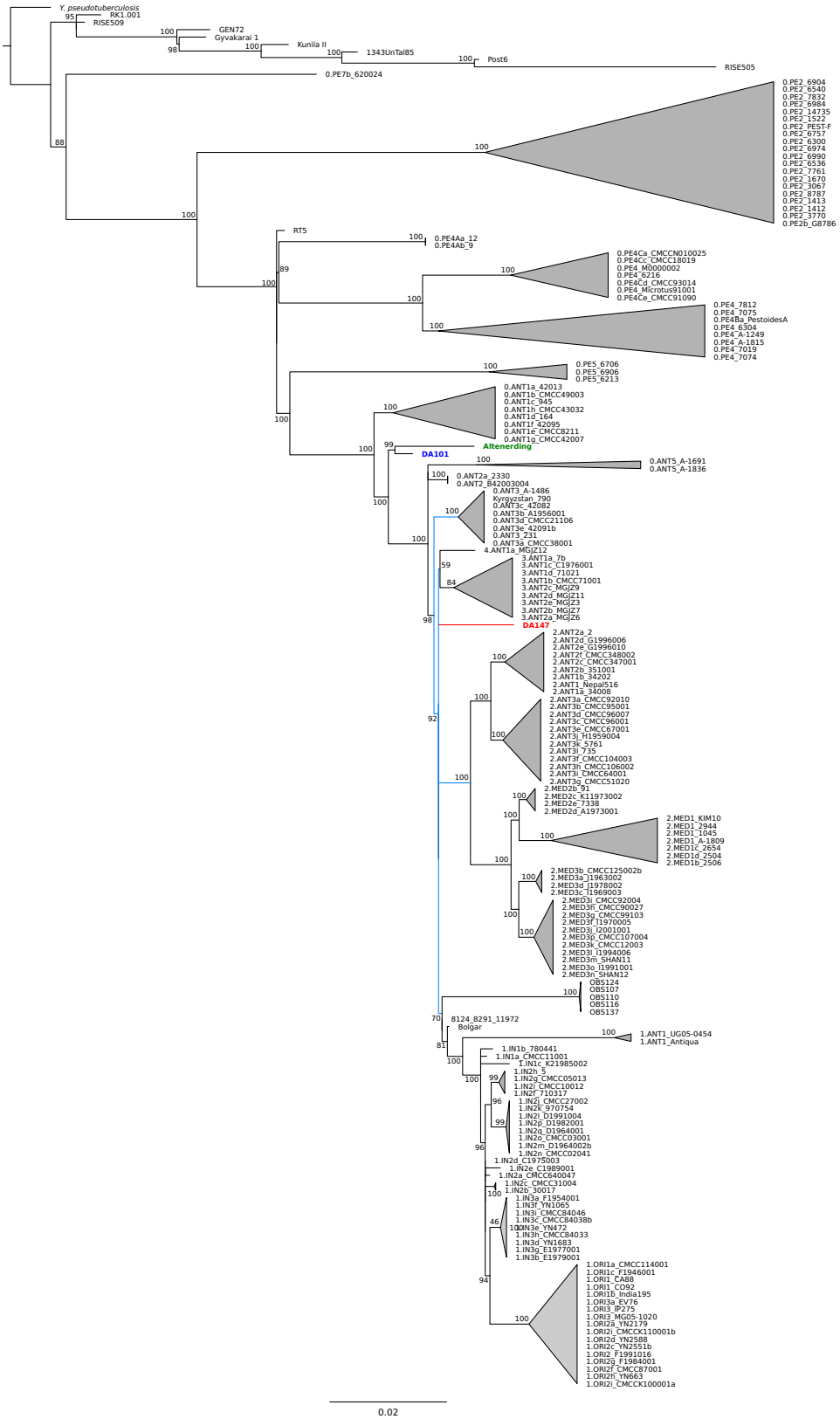


Figure S2: Maximum clade credibility tree showing substitution rate across all branches of the *Y. pestis* phylogeny, years BP are indicated on the bottom scale. Since the analysis is based on a SNP alignment with partial deletion (95%).

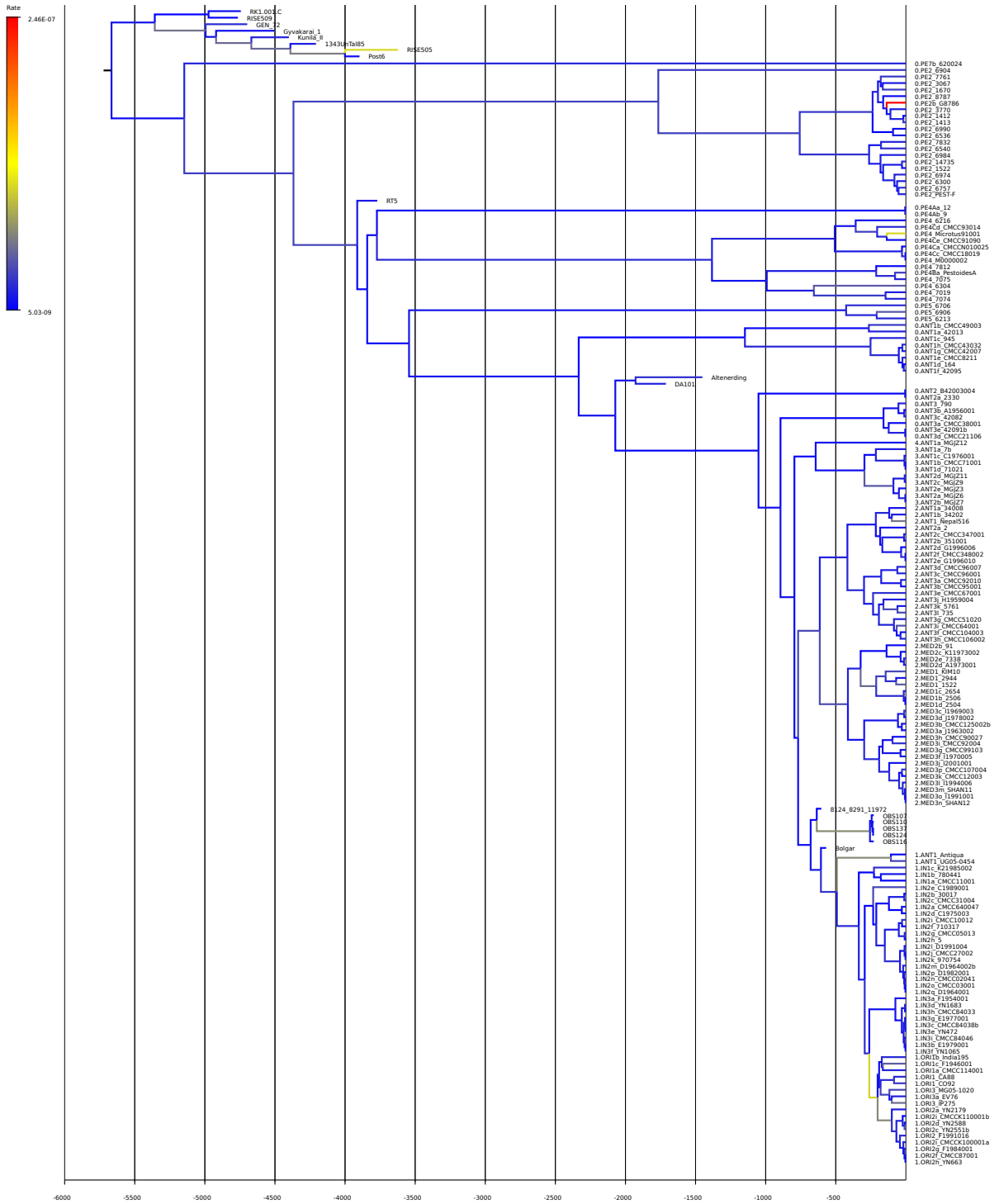


Table S1: List of all *Y. pestis* genomes with accession number, sample origin and the corresponding publication.

Strain ID	Accession No.	Origin	Publication
Modern			
0.ANT1a_42013	ADPG00000000	Xinjiang, China	Cui et al. 2013
0.ANT1b_CMCC49003	ADQX00000000	Xinjiang, China	Cui et al. 2013
0.ANT1c_945	ADPV00000000	Xinjiang, China	Cui et al. 2013
0.ANT1d_164	ADOW00000000	Xinjiang, China	Cui et al. 2013
0.ANT1e_CMCC8211	ADRD00000000	Xinjiang, China	Cui et al. 2013
0.ANT1f_42095	ADPJ00000000	Xinjiang, China	Cui et al. 2013
0.ANT1g_CMCC42007	ADQV00000000	Xinjiang, China	Cui et al. 2013
0.ANT1h_CMCC43032	ADQW00000000	Xinjiang, China	Cui et al. 2013
0.ANT2_B42003004	AAYU00000000	Xinjiang, China	Cui et al. 2013
0.ANT2a_2330	ADOY00000000	Xinjiang, China	Cui et al. 2013
0.ANT3_231	JMUF00000000	Kyrgyzstan	Eroshenko et al. 2017
0.ANT3_790	CP006806	Kyrgyzstan	Zhgenti et al. 2015
0.ANT3_A_1496	LYMP00000000	Kyrgyzstan	Eroshenko et al. 2017
0.ANT3a_CMCC38001	ADQU00000000	Xinjiang, China	Cui et al. 2013
0.ANT3b_A1956001	ADPX00000000	Xinjiang, China	Cui et al. 2013
0.ANT3c_42082	ADPH00000000	Xinjiang, China	Cui et al. 2013
0.ANT3d_CMCC21106	ADQP00000000	Xinjiang, China	Cui et al. 2013
0.ANT3e_42091b	ADPI00000000	Xinjiang, China	Cui et al. 2013
0.ANT5_A_1691	LYMQ00000000	Kyrgyzstan	Eroshenko et al. 2017
0.ANT5_A_1836	LYOL00000000	Kyrgyzstan	Eroshenko et al. 2017
0.PE2_14735	AYLS00000000	Armenia	Zhgenti et al. 2015
0.PE2_1522	CP006758	Armenia	Zhgenti et al. 2015
0.PE2_1412	CP006783	Dalidag, Georgia	Zhgenti et al. 2015
0.PE2_1413	CP006762	Ninotsminda, Georgia	Zhgenti et al. 2015
0.PE2_1670	AYLR00000000	Ninotsminda, Georgia	Zhgenti et al. 2015
0.PE2_3067	CP006754	Akhalkalaki, Georgia	Zhgenti et al. 2015
0.PE2_3770	CP006751	Ninotsminda, Georgia	Zhgenti et al. 2015
0.PE2_8787	CP006748	Ninotsminda, Georgia	Zhgenti et al. 2015
0.PE2_6300	LIZC00000000	Pre-Araks, Armenia, Azerbaijan	Kislichkina et al. 2015
0.PE2_6536	LIZE00000000	Transcaucasian Highland, Armenia, Georgia	Kislichkina et al. 2015
0.PE2_6540	LIZF00000000	Transcaucasian Highland, Armenia, Azerbaijan	Kislichkina et al. 2015
0.PE2_6757	LIYY00000000	Pre-Araks, Armenia, Azerbaijan	Kislichkina et al. 2015
0.PE2_6904	LIYP00000000	Dagestan-highland, Russia	Kislichkina et al. 2015
0.PE2_6974	LIYX00000000	Transcaucasian Highland, Armenia, Azerbaijan	Kislichkina et al. 2015
0.PE2_6984	LIYZ00000000	Transcaucasian Highland, Armenia, Azerbaijan	Kislichkina et al. 2015
0.PE2_6990	LIYU00000000	Transcaucasian Highland, Armenia, Georgia	Kislichkina et al. 2015
0.PE2_7761	LIYQ00000000	Transcaucasian Highland, Armenia, Georgia	Kislichkina et al. 2015
0.PE2_7832	LIZB00000000	Transcaucasian Highland, Armenia, Azerbaijan	Kislichkina et al. 2015
0.PE2_PEST-F	NC_009381	Former Soviet Union	Cui et al. 2013
0.PE2b_G8786	ADSG00000000	Georgia	Cui et al. 2013
0.PE4_6216	LIYR00000000	Bayan-Khongor, Mongolia	Kislichkina et al. 2015
0.PE4_6304	LIYS00000000	Gissar, Tadjikistan, Uzbekistan	Kislichkina et al. 2015
0.PE4_7019	LIYW00000000	Talas, Kyrgyzstan	Kislichkina et al. 2015
0.PE4_7074	LIYT00000000	Talas, Kyrgyzstan	Kislichkina et al. 2015
0.PE4_7075	LIZA00000000	Mountain-Altai, Russia	Kislichkina et al. 2015
0.PE4_7812	LIYV00000000	Mountain-Altai, Russia	Kislichkina et al. 2015
0.PE4_Microtus91001	NC_005810	Inner Mongolia, China	Cui et al. 2013
0.PE4Aa_12	ADOV00000000	Qinghai, China	Cui et al. 2013
0.PE4Ab_9	ADPT00000000	Qinghai, China	Cui et al. 2013
0.PE4b_M0000002	ADST00000000	Qinghai, China	Cui et al. 2013
0.PE4Ba_PestoidesA	ACNT00000000	Former Soviet Union	Cui et al. 2013
0.PE4Ca_CMCCN010025	ADRT00000000	Sichuan, China	Cui et al. 2013
0.PE4Cc_CMCC18019	ADQO00000000	Qinghai, China	Cui et al. 2013
0.PE4Cd_CMCC93014	ADRM00000000	Inner Mongolia, China	Cui et al. 2013
0.PE4Ce_CMCC91090	ADRJ00000000	Inner Mongolia, China	Cui et al. 2013
0.PE4h_A_1249	LYMN00000000	Tadzhikistan	Eroshenko et al. 2017
0.PE4t_A_1815	LPTY00000000	Kyrgyzstan	Eroshenko et al. 2017
0.PE5_6213	LIZD00000000	Northeast Mongolia, Gobi Desert	Kislichkina et al. 2015
0.PE5_6706	LIYO00000000	Northeast Mongolia, Gobi Desert	Kislichkina et al. 2015
0.PE5_6906	LIZG00000000	Northeast Mongolia, Gobi Desert	Kislichkina et al. 2015
0.PE7b_620024	ADPM00000000	Qinghai, China	Cui et al. 2013
1.ANT1_Antiqua	NC_008150	Congo	Cui et al. 2013
1.ANT1_UG05-0454	AAYR00000000	Uganda	Cui et al. 2013
1.IN1a_CMCC11001	ADQK00000000	Qinghai, China	Cui et al. 2013
1.IN1b_780441	ADPS00000000	Qinghai, China	Cui et al. 2013

1.IN1c_K21985002	ADSS00000000	Xinjiang, China	Cui et al. 2013
1.IN2a_CMCC640047	ADRA00000000	Qinghai, China	Cui et al. 2013
1.IN2b_30017	ADPC00000000	Tibet, China	Cui et al. 2013
1.IN2c_CMCC31004	ADQR00000000	Tibet, China	Cui et al. 2013
1.IN2d_C1975003	ADPZ00000000	Qinghai, China	Cui et al. 2013
1.IN2e_C1989001	ADQB00000000	Qinghai, China	Cui et al. 2013
1.IN2f_710317	ADPP00000000	Qinghai, China	Cui et al. 2013
1.IN2g_CMCC05013	ADQF00000000	Qinghai, China	Cui et al. 2013
1.IN2h_5	ADPK00000000	Qinghai, China	Cui et al. 2013
1.IN2i_CMCC10012	ADQG00000000	Qinghai, China	Cui et al. 2013
1.IN2j_CMCC27002	ADQQ00000000	Qinghai, China	Cui et al. 2013
1.IN2k_970754	ADPW00000000	Qinghai, China	Cui et al. 2013
1.IN2l_D1991004	ADRX00000000	Qinghai, China	Cui et al. 2013
1.IN2m_D1964002b	ADRV00000000	Qinghai, China	Cui et al. 2013
1.IN2n_CMCC02041	ADQC00000000	Qinghai, China	Cui et al. 2013
1.IN2o_CMCC03001	ADQD00000000	Qinghai, China	Cui et al. 2013
1.IN2p_D1982001	ADRW00000000	Gansu, China	Cui et al. 2013
1.IN2q_D1964001	ADRU00000000	Qinghai, China	Cui et al. 2013
1.IN3a_F1954001	ADSC00000000	Yunnan, China	Cui et al. 2013
1.IN3b_E1979001	AAYV00000000	Yunnan, China	Cui et al. 2013
1.IN3c_CMCC84038b	ADRF00000000	Yunnan, China	Cui et al. 2013
1.IN3d_YN1683	ADTD00000000	Yunnan, China	Cui et al. 2013
1.IN3e_YN472	ADTH00000000	Yunnan, China	Cui et al. 2013
1.IN3f_YN1065	ADTC00000000	Yunnan, China	Cui et al. 2013
1.IN3g_E1977001	ADRY00000000	Yunnan, China	Cui et al. 2013
1.IN3h_CMCC84033	ADRE00000000	Yunnan, China	Cui et al. 2013
1.IN3i_CMCC84046	ADRG00000000	Yunnan, China	Cui et al. 2013
1.ORI1_CA88	ABCD00000000	California, USA	Cui et al. 2013
1.ORI1_CO92	NC_003143	Colorado, USA	Cui et al. 2013
1.ORI1a_CMCC114001	ADQL00000000	Fujian, China	Cui et al. 2013
1.ORI1b_India195	ACNR00000000	India	Cui et al. 2013
1.ORI1c_F1946001	ADSB00000000	Fujian, China	Cui et al. 2013
1.ORI2_F1991016	ABAT00000000	Yunnan, China	Cui et al. 2013
1.ORI2a_YN2179	ADTE00000000	Myanmar	Cui et al. 2013
1.ORI2c_YN2551b	ADTF00000000	Yunnan, China	Cui et al. 2013
1.ORI2d_YN2588	ADTG00000000	Guangxi, China	Cui et al. 2013
1.ORI2f_CMCC87001	ADRH00000000	Yunnan, China	Cui et al. 2013
1.ORI2g_F1984001	ADSD00000000	Yunnan, China	Cui et al. 2013
1.ORI2h_YN663	ADTI00000000	Yunnan, China	Cui et al. 2013
1.ORI2i_CMCCK100001a	ADRR00000000	Yunnan, China	Cui et al. 2013
1.ORI2i_CMCCK110001b	ADRS00000000	Yunnan, China	Cui et al. 2013
1.ORI3_IP275	AAOS00000000	Madagascar	Cui et al. 2013
1.ORI3_MG05-1020	AAYS00000000	Madagascar	Cui et al. 2013
1.ORI3a_EV76	ADSA00000000	Madagascar	Cui et al. 2013
2.ANT1_Nepal516	ACNQ00000000	Nepal	Cui et al. 2013
2.ANT1a_34008	ADPD00000000	Tibet, China	Cui et al. 2013
2.ANT1b_34202	ADPE00000000	Tibet, China	Cui et al. 2013
2.ANT2a_2	ADOX00000000	Qinghai, China	Cui et al. 2013
2.ANT2b_351001	ADPF00000000	Tibet, China	Cui et al. 2013
2.ANT2c_CMCC347001	ADQS00000000	Tibet, China	Cui et al. 2013
2.ANT2d_G1996006	ADSE00000000	Tibet, China	Cui et al. 2013
2.ANT2e_G1996010	ADSF00000000	Tibet, China	Cui et al. 2013
2.ANT2f_CMCC348002	ADQT00000000	Tibet, China	Cui et al. 2013
2.ANT3a_CMCC92010	ADRL00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3b_CMCC95001	ADRN00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3c_CMCC96001	ADRO00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3d_CMCC96007	ADRP00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3e_CMCC67001	ADRB00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3f_CMCC104003	ADQH00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3g_CMCC51020	ADQY00000000	Jilin, China	Cui et al. 2013
2.ANT3h_CMCC106002	ADQI00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3i_CMCC64001	ADQZ00000000	Inner Mongolia, China	Cui et al. 2013
2.ANT3j_H1959004	ADSI00000000	Jilin, China	Cui et al. 2013
2.ANT3k_5761	ADPL00000000	St.Petersburg, Russia	Cui et al. 2013
2.ANT3l_735	ADPR00000000	St.Petersburg, Russia	Cui et al. 2013
2.MED1_A_1809	LYMF00000000	Kyrgyzstan	Eroshenko et al. 2017
2.MED1_1045	CP006794	Azerbaijan	Zhgenti et al. 2015
2.MED1_2944	CP006792	Kabardino-Balkaria, Russia	Zhgenti et al. 2015
2.MED1b_2506	ADPA00000000	Xinjiang, China	Cui et al. 2013
2.MED1c_2654	ADPB00000000	Xinjiang, China	Cui et al. 2013
2.MED1d_2504	ADOZ00000000	Xinjiang, China	Cui et al. 2013
2.MED2_KIM10	NC_004088	Iran/Kurdistan	Cui et al. 2013

2.MED2b_91	ADPU00000000	Xinjiang, China	Cui et al. 2013
2.MED2c_K11973002	AAYT00000000	Xinjiang, China	Cui et al. 2013
2.MED2d_A1973001	ADPY00000000	Xinjiang, China	Cui et al. 2013
2.MED2e_7338	ADPQ00000000	Xinjiang, China	Cui et al. 2013
2.MED3a_J1963002	ADSP00000000	Gansu, China	Cui et al. 2013
2.MED3b_CMCC125002b	ADQN00000000	Ningxia, China	Cui et al. 2013
2.MED3c_I1969003	ADSK00000000	Ningxia, China	Cui et al. 2013
2.MED3d_J1978002	ADSQ00000000	Ningxia, China	Cui et al. 2013
2.MED3f_I1970005	ADSL00000000	Inner Mongolia, China	Cui et al. 2013
2.MED3g_CMCC99103	ADRQ00000000	Inner Mongolia, China	Cui et al. 2013
2.MED3h_CMCC90027	ADRI00000000	Inner Mongolia, China	Cui et al. 2013
2.MED3i_CMCC92004	ADRK00000000	Inner Mongolia, China	Cui et al. 2013
2.MED3j_I2001001	ADSO00000000	Shaanxi, China	Cui et al. 2013
2.MED3k_CMCC12003	ADQM00000000	Qinghai, China	Cui et al. 2013
2.MED3l_I1994006	ADSN00000000	Hebei, China	Cui et al. 2013
2.MED3m_SHAN11	ADTA00000000	Shaanxi, China	Cui et al. 2013
2.MED3n_SHAN12	ADTB00000000	Shaanxi, China	Cui et al. 2013
2.MED3o_I1991001	ADSM00000000	Inner Mongolia, China	Cui et al. 2013
2.MED3p_CMCC107004	ADQJ00000000	Inner Mongolia, China	Cui et al. 2013
3.ANT1a_7b	ADPN00000000	Qinghai, China	Cui et al. 2013
3.ANT1b_CMCC71001	ADRC00000000	Gansu, China	Cui et al. 2013
3.ANT1c_C1976001	ADQA00000000	Gansu, China	Cui et al. 2013
3.ANT1d_71021	ADPO00000000	Gansu, China	Cui et al. 2013
3.ANT2a_MGJZ6	ADSX00000000	Dornogovi, Mongolia	Cui et al. 2013
3.ANT2b_MGJZ7	ADSY00000000	Dornogovi, Mongolia	Cui et al. 2013
3.ANT2c_MGJZ9	ADSZ00000000	Govi-Altai, Mongolia	Cui et al. 2013
3.ANT2d_MGJZ11	ADSU00000000	Bayan-Ölgii, Mongolia	Cui et al. 2013
3.ANT2e_MGJZ3	ADSW00000000	Govi-Altai, Mongolia	Cui et al. 2013
4.ANT1a_MGJZ12	ADSV00000000	Bayan-Ölgii, Mongolia	Cui et al. 2013
Ancient			
8124_8291_11972	SRR341961, SRR341962, SRR341963	United Kingdom	Bos et al. 2011
RISE505	PRJEB10885	Russia	Rasmussen et al. 2015
RISE509	PRJEB10885	Russia	Rasmussen et al. 2015
OBS107	PRJEB12163	France	Bos et al. 2016
OBS110	PRJEB12163	France	Bos et al. 2016
OBS116	PRJEB12163	France	Bos et al. 2016
OBS124	PRJEB12163	France	Bos et al. 2016
Altenerding	PRJEB14851	Germany	Feldman et al. 2016
Bolgar	PRJEB13664	Russia	Spyrou et al. 2016
GEN72	PRJEB19335	Croatia	Andrades Valtueña et al. 2017
RK1.001.C	PRJEB19335	Russia	Andrades Valtueña et al. 2017
Kunila II	PRJEB19335	Estonia	Andrades Valtueña et al. 2017
Gyvakarai 1	PRJEB19335	Lithuania	Andrades Valtueña et al. 2017
6Post	PRJEB19335	Germany	Andrades Valtueña et al. 2017
1343UnTal85	PRJEB19335	Germany	Andrades Valtueña et al. 2017
DA101	PRJEB25891	Kyrgyzstan	Daamgard et al. 2018
DA147	PRJEB25891	Russia	Daamgard et al. 2018
RT5	PRJEB24296	Russia	Spyrou et al. 2018

Table S2: Unique and shared SNPs of DA101 and Altenerding. *Erroneously classified as ancestral in Feldman et al. 2016. **Potential damage site (C>T).

Shared SNPs				
Position	**Reference (CO92)**	**Altenerding**	**DA101**	**DA147**
260148	C	T	T	N
2725715	C	T	T	N
2977542	C	A	A	N
Unique SNPs Altenerding				
Position	**Reference (CO92)**	**Altenerding**	**DA101**	**DA147**
86824	A	G	N	N
189912	A	G	.	N
420208*	G	T	.	.
271114	C	A	.	N
485976	C	T	.	.
557841	C	T	N	N
727741	G	A	.	N
779365	C	T	.	N
898980	A	T	N	.
1067966	C	A	.	N
1211729	A	C	.	N
1296743	C	T	.	N
1387701	C	T	N	N
1387756	A	G	.	.
1413031	C	A	N	.
1434752	C	A	.	N
1489055	C	T	.	N
1530658	C	A	.	N
1609461	T	C	N	N
1754708	C	T	.	T*
1868678	G	T	N	N
1956162	T	C	.	N
2092152	C	T	.	N
2097520	G	T	.	N
2352174	T	G	.	.
2419529	G	A	.	N
2495165	C	A	.	N
2753572	A	T	.	N
3078807	C	A	.	N
3179828	C	A	.	N
3274298	A	T	N	N
3360963	A	C	.	N
3360984	C	T	.	N
3398153	G	A	N	N
3409414	T	C	N	N
3500922	T	G	.	N
3535148	G	T	.	N
3560088	G	A	.	.
3568597	C	T	.	N
3750736	G	A	N	N
3755861	C	T	N	.
3843195	C	A	.	N
3892488	C	T	.	N
4066494	C	T	N	N
4307755	G	A	.	N
4412624	A	G	.	.
4423366	G	A	N	N

4460688	C	T	N	N
4465967	C	A	N	N
4628496	C	A	.	N
4629169	G	A	N	N
Unique SNPs DA101				
Position	**Reference (CO92)**	**Altenerding**	**DA101**	**DA147**
100945	C	.	T	N
178109	C	.	T	N
3066176	C	.	T	N
3075453	G	.	T	.
3550137	C	.	T	N
3592088	C	.	A	N
3641873	G	.	A	N
3686087	C	.	T	N
3729628	G	.	A	.

Table S3: SNPs defining possible positions of DA147 in the phylogenetic tree.

Position	**Reference (CO92)**	**0.ANT3**	**DA147**	**State**
561567	G	A	N	not covered
1440879	A	G	N	not covered
1698435	A	G	N	not covered
1719187	C	T	N	not covered
2003542	C	T	N	not covered
2082310	C	T	N	not covered
2130133	G	T	N	not covered
2425991	G	T	.	ancestral
2656734	C	T	.	ancestral
3512754	T	C	.	ancestral
3727189	G	T	N	not covered
4166664	G	A	.	ancestral
4245783	C	T	N	not covered
4281601	G	T	N	not covered
4427796	T	G	N	not covered
Position	**Reference (CO92)**	**Branch 0**	**DA147**	**State**
1102174	A	G	.	derived
1251046	T	C	N	not covered
2812384	G	T	.	derived
Position	**Reference (CO92)**	**Branch 1**	**DA147**	**State**
189227	C	.	T	ancestral
1871476	G	.	N	not covered
Position	**Reference (CO92)**	**Branch 2**	**DA147**	**State**
97226	G	A	.	ancestral
282762	G	A	N	not covered
335833	C	T	N	not covered
443673	G	T	N	not covered
710909	C	T	N	not covered
718827	T	C	N	not covered
759797	G	T	.	ancestral
1971665	G	A	N	not covered
2082100	C	T	N	not covered
2493895	C	T	N	not covered
2799652	T	C	N	not covered
2847692	C	T	N	not covered
2934864	C	T	N	not covered
3314013	G	T	N	not covered
3376387	A	G	.	ancestral
3539709	C	T	N	not covered
3551089	A	C	N	not covered
4311918	C	T	N	not covered
4595001	C	T	.	ancestral
4598609	G	A	N	not covered
Position	**Reference (CO92)**	**Branch 3+4**	**DA147**	**State**
3021936	C	A	.	ancestral

Table S4: Mean divergence dates and 95 % HPD intervals for important nodes in the phylogenetic tree.

Lineage divergence	Mean (95 % HPD) in years BP including RT5	Mean (95 % HPD) in years BP excluding RT5
Tree root	5782 (4941–6863)	5449 (4869–6200)
0.PE7	5232 (4229–6672)	4490 (3235–5912)
0.PE2	4428 (3904–5073)	3483 (1642–4453)
0.PE4 (incl. RT5)	3932 (3742–4153)	2803 (2175–3492)
0.PE5	3478 (2763–4004)	2605 (2058–3235)
0.ANT1	2372 (1924–2891)	2139 (1846–2476)
Altenerding+DA101	2104 (1797–2477)	1977 (1770–2244)
DA101	1959 (1729–2268)	1881 (1722–2091)
0.ANT2	1085 (741–1503)	986 (703–1348)
0.ANT3	920 (676–1218)	844 (652–1101)
Black Death	654 (602–763)	637 (602–712)

Table S5: Mean substitution rates and their respective 95% HPD for the whole Y. pestis phylogeny and the Altenerding branch.

	Mean (95 % HPD) including RT5	Mean (95 % HPD) excluding RT5
Full tree	1.48E-08 (1.23E-08–1.73E-08)	1.75E-08 (1.42E-08–2.08E-08)
Altenerding	2.67E-08 (1.08E-08–4.36E-08)	3.08E-08 (1.08E-08–4.82E-08)