The evolutionary history of plague as revealed through the analysis of ancient *Yersinia pestis* genomes

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Maria Alexandra Spyrou

aus Glasgow/Großbritannien

Tübingen

2018

Acknowledgements

I would like to express my gratitude to several people that have supported me throughout this intense journey.

My foremost appreciation goes to my primary supervisor, Johannes Krause, for his support throughout this entire PhD, for always being available to share his outstanding scientific knowledge and for giving me the freedom to discover my own assets.

To Nick Conard, for kindly agreeing to supervise my thesis and supported me during the last stages of my PhD.

To Kirsten Bos, for training me in the lab, for her supervision in several projects as well as for her mentorship throughout my PhD, and to Alexander Herbig for supervising the computational analysis portion of my projects as well as for his continuous availability.

To Wolfgang Haak, for his help with mitochondrial DNA analysis and for always bringing a great atmosphere to the work environment.

To my office-mates: Marcel Keller, who agreed to translate the summary of this thesis and has been a reliable collaborator, and to Michal Feldman who has been a true friend and always open to scientific and personal exchange.

To colleagues and friends: Rezeda Tukhbatova, for our teamwork in several fascinating projects, Kathrin Nägele, for being my lab assistant during some of the most intense working periods and Betsy Nelson for her lab support and for sharing her palaeopathological expertise.

To my Jena friends and colleagues, especially Rodrigo, for all the help and fun moments in the office and to my Tübingen friends Solmaz, Peter, Alvise and Saad who I deeply missed while in Jena and I am glad to be spending more time with.

To Argyris, Athena and Kaiti, for being my most reliable companions during the last 12 years and the best friends one could have ever wished for.

To Cosimo, for his constant encouragement, active support and patience! You believed in me more than I did in myself. We did this together!

To my aunt Vasia for her sincere love and continuous presence.

Last but not least, I would like to thank my family; my brother Giannis and my parents Eleni and Kostas, for their unconditional love as well as their unfailing care and support without which I would have never made it this far.

Table of Contents

Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
AD	Anno Domini
aDNA	Ancient Deoxyribonucleic Acid
bp	base pair(s)
C	Cytosine
CCR5	C-C motif chemokine receptor 5
DNA	Deoxyribonucleic acid
EBA	Early Bronze Age
EPT	Early phase transmission
HIV	Human Immunodeficiency Virus
LNBA	Late Neolithic and Early Bronze Age
mgtB	Magnesium transporter B
mgtC	Magnesium transporter C
MLBA	Middle/Late Bronze Age
mtDNA	Mitochondrial Deoxyribonucleic Acid
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
PDE	Phosphodiesterase
PhoP/Q	Non-specific acid phosphatase
pla	Plasminogen activator
qPCR	Quantitative Polymerase Chain Reaction
Rcs	Regulator of capsule synthesis
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
T	Thymine
TLR1-6-10	Toll-like receptor 1-6-10
ureD	Urease D
yBP	Years before the present
ymt	Yersinia murine toxin
U	Uracil
UDG	Uracil-DNA-glycosylase

Summary

The study of ancient pathogen genomics has the potential to reveal key insights into the history of infectious disease as well as into the processes that have triggered pathogen emergence leading to epidemics/pandemics. A characteristic historical example of abrupt infectious disease emergence is that of plague during the Black Death (1346-1353 AD), which claimed the lives of up to 60% of the European population in only five years. In the present thesis, I have investigated the genetic history of its causative agent, *Yersinia pestis*, by tracing back some of the earliest phases of its evolution, but also by studying its infamous, and historically documented, medieval epidemics. This work has been possible through the adoption of both *in vitro* and *in silico* methodologies for the detection of *Y. pestis* DNA in ancient human remains as well as for its enrichment, NGS sequencing and genome reconstruction. One of the sections of this thesis reveals key insights into the origin of flea-mediated transmission in *Y. pestis*, which is known to lead to the typical disease presentation referred to as bubonic plague. Contrary to previous research suggesting that during its early evolution *Y. pestis* was unable to transmit via the flea vector, I describe strains from ~4,000 years ago that show genetic compatibility with this mode of transmission. This traces the origins of bubonic plague back to the Bronze Age and shows the parallel existence of several bacterial forms that had different transmission and disease potentials. After the Bronze Age, bubonic plague was the main culprit of two historical pandemics (6th – 8th and 14th – 18th centuries AD). In this thesis, I have performed genetic screening of human remains from both of these periods in order to assess the presence of *Y. pestis*. As a result, I am able to present a systematic study of the Black Death and its ensuing medieval plague outbreaks that occurred in Europe until the 18th-century by achieving a five-fold increase in the available genomic data that exist from that time. These data reveal a low genetic diversity in the bacterium during the Black Death, and support a link between this first wave and modern-day plague occurrences around the world. Moreover, the genomes reconstructed from later European epidemics suggest that *Y. pestis* persisted locally for 400-years after the Black Death, where it further diversified into multiple genetically distinct clades. These clades differ in their evolutionary rates and virulence profiles and seem to have no identified modern descendants. Finally, apart from *Y. pestis* genomes, this thesis also describes the analysis of human DNA from part of the

studied individuals, most of which derive from epidemic contexts. Analysis of mitochondrial and nuclear DNA from Bronze Age and early medieval individuals suggests that human mobility during historic and prehistoric times may have facilitated the spread of pathogens, such as *Y. pestis*, across Eurasia.

Zusammenfassung

Die Erforschung alter Pathogen-Genome hat das Potenzial, nicht nur fundamentale Einblicke in die Geschichte von Infektionskrankheiten zu gewähren, sondern auch die Entwicklungsprozesse der Erreger aufzudecken, die schließlich zu Epidemien und Pandemien führten. Ein charakteristisches historisches Beispiel für das plötzliche Auftreten einer Infektionskrankheit ist das der Pest während des Schwarzen Todes (1346-1353 n. Chr.), welcher für den Tod von bis zu 60% der europäischen Bevölkerung innerhalb von nur fünf Jahren verantwortlich gemacht wird. In dieser Arbeit habe ich die genetische Geschichte ihres Erregers *Yersinia pestis* untersucht, einerseits durch Zurückverfolgung der frühesten Phasen seiner Evolution, andererseits aber auch durch Studien zu den berüchtigten, historisch dokumentierten mittelalterlichen Epidemien. Dies war möglich durch die Anpassung von *in-vitro*- und *in-silico*-Methoden zum Nachweis von *Y.-pestis*-DNA in menschlichen Überresten sowie deren Anreicherung, NGS-Sequenzierung und Genom-Rekonstruktion. Ein Teil dieser Arbeit widmet sich dem Ursprung der Übertragbarkeit von *Y. pestis* mittels Flöhen, was die charakteristische Manifestation der Pest verursacht, die als Beulenpest bezeichnet wird. Im Gegensatz zu früheren Studien, welche nahegelegt haben, dass *Y. pestis* in frühen Stadien seiner Evolution nicht über Flöhe übertragbar war, beschreibe ich ca. 4000 Jahre alte Stämme, welche genetisch zu diesem Übertragungsweg in der Lage waren. Dies datiert die Entstehung der Beulenpest bis in die Bronzezeit zurück und bezeugt die gleichzeitige Anwesenheit mehrerer Erregertypen, die unterschiedliche Übertragungs- und Krankheitspotenziale hatten. Nach der Bronzezeit war die Beulenpest für zwei historische Pandemien verantwortlich (6. – 8. Jh. und 14. – 18. Jh. n. Chr.). Für diese Arbeit habe ich menschliche Überreste beider Pandemien genetisch auf *Y. pestis* untersucht. Als Ergebnis kann ich eine systematische Studie zum Schwarzen Tod und den darauf folgenden mittelalterlichen Pestausbrüchen in Europa, die bis in das 18. Jh. andauerten, präsentieren, welche durch die Verfünffachung der

verfügbaren genomischen Daten aus dieser Zeit möglich wurde. Diese Daten zeigen eine geringe genetische Diversität des Bakteriums während des Schwarzen Todes und unterstützen eine Verbindung dieser ersten Welle mit weltweiten Pestfällen der heutigen Zeit. Darüber hinaus stützen die rekonstruierten Genome späterer europäischer Epidemien die These, dass sich *Y. pestis* für 400 Jahre nach dem Schwarzen Tod in Europa halten konnte und sich in mehrere, genetisch verschiedene Kladen diversifiziert hat. Diese Kladen unterscheiden sich in ihren evolutiven Raten und Virulenzprofilen und haben keine bekannten modernen Nachfahren. Schließlich beinhaltet diese Arbeit neben den Analysen zu *Y.-pestis*-Genomen auch Untersuchungen der menschlichen DNA eines Teils der untersuchten Individuen, von denen die meisten aus einem epidemischen Kontext stammen. Die Untersuchung der mitochondrialen und Kern-DNA von bronzezeitlichen und frühmittelalterlichen Individuen deutet darauf hin, dass menschliche Mobilität in historischer und vorgeschichtlicher Zeit die Ausbreitung von Erregern wie *Y. pestis* durch Eurasien begünstigt hat.

List of publications

The following four manuscripts are presented and discussed in this thesis and will be referred to as papers A-D throughout the text.

A.   **Maria A. Spyrou**, Rezeda I. Tukhbatova, Chuan-Chao Wang, Aida Andrades Valtueña, Aditya K. Lankapalli, Vitaly V. Kondrashin, Victor A. Tsibin, Aleksandr Khokhlov, Denise Kühnert, Alexander Herbig, Kirsten I. Bos and Johannes Krause. "Analysis of 3,800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague" Published in *Nature Communications*, 2018, 9(1): 2234

B.   **Maria A. Spyrou**, Alessandra Sperduti, Åshild J. Vågene, Lorenzo M. Bondioli, Henrike Heyne, Eva Fernández-Domínguez, Luca Bondioli, Wolfgang Haak, Kirsten I. Bos and Johannes Krause "Ancient DNA recovery and maternal lineage diversity of early medieval Venosa in southern Italy" Manuscript ready for submission

C.   **Maria A. Spyrou**, Rezeda I. Tukhbatova, Michal Feldman, Joanna Drath, Sacha Kacki, Julia Beltrán de Heredia, Susanne Arnold, Airat G. Sitdikov, Dominique Castex, Joachim Wahl, Ilgizar R. Gazimzyanov, Danis K. Nurgaliev, Alexander Herbig, Kirsten I. Bos and Johannes Krause "Historical *Y. pestis* genomes reveal the European Black Death as the source of ancient and modern plague pandemics" Published in *Cell Host & Microbe*, 2016, 19(6), pp. 874-81

D.   **Maria A. Spyrou***, Marcel Keller*, Rezeda Tukhbatova, Elisabeth Nelson, Don Walker, Amelie Alterauge, Hermann Fetz, Joris Peters, Niamh Carty, Robert Hartle, Michael Henderson, Elizabeth L. Knox, Sacha Kacki, Michaël Gourvennec, Dominique Castex, Sandra Lösch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos and Johannes Krause "A phylogeography of the second plague pandemic revealed through historical *Y. pestis* genomes" Manuscript ready for submission
*denotes equal contribution

Reprints were made available with permission of the respective publishers.

Own contributions

A.   I performed the screening portion of the laboratory work and the preparation for targeted enrichment together with Rezeda I. Tukhbatova. In addition, I performed all *Y. pestis* data analysis with assistance from Aida Andrades Valtueña and Aditya K. Lankapalli under the supervision of Alexander Herbig. I wrote all sections of the paper together with Kirsten I. Bos under the supervision of Johannes Krause.

B.   I performed all laboratory work under the supervision of Kirsten I. Bos and with contributions from Åshild Vågene throughout the entire workflow. I analysed all mitochondrial DNA under the supervision of Wolfgang Haak and with advise from Eva Fernández-Domínguez. I wrote all sections of the paper with contributions from Lorenzo M. Bondioli, Luca Bondioli and Wolfgang Haak under the supervision of Kirsten I. Bos and Johannes Krause.

C.   I performed all laboratory work for this study with assistance from Rezeda I. Tukhbatova for the material from Bolgar City and with the contribution of Michal Feldman on part of qPCR and targeted DNA enrichments assays. I performed all data analysis under the supervision of Alexander Herbig. I wrote the paper together with Kirsten I. Bos under the supervision of Johannes Krause.

D.   I performed the screening portion of the laboratory work and the preparation for targeted enrichment together with Marcel Keller and Rezeda Tukhbatova as well as with support from Elisabeth Nelson whom I supervised through this process. I performed the data analysis with support from Marcel Keller and Alexander Herbig. I wrote all sections of the paper with contributions by Marcel Keller for the main text and by Don Walker and Michaël Gourvennec for the Supplementary Archaeological Information under the supervision of Kirsten I. Bos and Johannes Krause.

1 Introduction

1.1 Past and present infectious disease

Across their evolution, humans have continuously faced infectious microorganisms and their interactions have been maintained in a dynamic equilibrium (Dawkins and Krebs, 1979). Pathogenic microbes are known to have influenced all regions and populations of the world and, through major epidemics and pandemics, have altered the course of human history on numerous occasions (Karlsson et al., 2014). As a result, newly emerging and re-emerging infections remain a challenge for public health and their prevention has always been at the forefront of combating priorities (Bloom et al., 2017). During the last 100 years alone, pandemics attributed to infectious diseases have contributed to substantial population declines. Notable examples include the Spanish flu (1918-1920), which claimed the lives of 50-100 million people during the First World War (Johnson and Mueller, 2002), and the on-going tuberculosis and HIV pandemics that are currently among the leading causes of death worldwide (WHO, 2016b, WHO, 2016a).

The interaction between humans and their microbes is not considered a recent phenomenon and, according to the epidemiological transitions theory, our changing environments and subsistence strategies seem to have determined the varieties of those encountered (Barrett et al., 1998). For example, during the Neolithic revolution, the transition into a sedentary lifestyle and the closer interaction with domesticated animals and plants may have facilitated exposure to a novel range of zoonotic pathogens, compared to those encountered by relatively small and mobile hunter-gatherer groups (Armelagos et al., 1996). Understanding, therefore, the natural history of pathogens and their first encounters with humans is not entirely feasible when solely accounting for their modern prevalence and geographic distribution. Traditionally, the study of past disease has focused on the morphological assessment of skeletal remains (Ortner, 2003). However, only a small number of infectious diseases are known to produce skeletal pathognomonic changes, and of those traces can only be found on a subset of the individuals that suffered chronically (Wood et al., 1992, Roberts and Buikstra, 2003, Ortner, 2008). In addition, given that based on morphology alone, the inference of evolutionary relationships becomes challenging, the recent synergy that began between

the fields of paleopathology and ancient DNA has benefited from the adoption of molecular methods for determining past infectious disease diagnosis and evolution. In this thesis, the last decade's genomic innovations were utilised to gain important insights into infectious disease history and, more specifically, into that of the notorious plague pathogen, *Yersinia pestis*.

1.2 Ancient DNA (aDNA)

1.2.1 The past and present of aDNA

Since its initiation more than three decades ago (Higuchi et al., 1984, Pääbo, 1985), the field of aDNA has greatly expanded the amount of information that can be retrieved from archaeological and/or paleontological remains. Some of the pioneering research on aDNA characterisation established that, unlike modern DNA, this type of genetic material has a low molecular weight (Pääbo, 1989) due to post-mortem enzymatic cleavage, microbial attack and hydrolytic de-purination (cleavage of *N*-glycosyl bond between the sugar and the adenine or guanine bases of DNA), all leading to macromolecular degradation (Eglinton and Logan, 1991, Lindahl, 1993). Over the years, systematic observation has revealed that the majority of aDNA molecules within a specimen are typically shorter than 100 base pairs (bp) (Sawyer et al., 2012). In addition, hydrolytic deamination of cytosine (C) residues into uracils (U) is another form of post-mortem damage characteristic of aDNA, which causes nucleotide changes, or miscoding lesions, across the DNA sequence and is often used for its authentication (Hofreiter et al., 2001, Briggs et al., 2007, Sawyer et al., 2012, Dabney et al., 2013b). Although hydrolytic deamination is shown to be positively correlated with a specimen's age (Sawyer et al., 2012), molecular preservation in ancient remains is also significantly influenced by factors such as the depositional context (i.e. soil acidity), temperature and humidity (Lindahl, 1993).

Such DNA damage characteristics posed limitations to the most prominent laboratory technique used in this field, the Polymerase Chain Reaction (PCR) (Saiki et al., 1985, Saiki et al., 1988). This method allows for targeted amplification of DNA fragments, using at least two primer sequences of typically 18-22 bp that flank the region of interest. Although high sensitivity is one of its advantages, its specificity is highly dependable on the target region as well as on the content and length of the primers used.

Hence, the amplification of targets shorter than ~50 bp was not feasible via PCR, which was often a limiting factor when analysing poorly preserved aDNA specimens, leading to the amplification of better-preserved and longer contaminant molecules (Pääbo and Wilson, 1988). As a result, the field's early years were challenged by discoveries that lacked suitable aDNA authentication criteria (Cooper and Poinar, 2000), which often lead to controversy regarding the validity of their results (Cano et al., 1993, Poinar et al., 1993, Weyand and Bunnell, 1994, Cano and Borucki, 1995).

The technological innovations of the last decade, the so-called Next Generation Sequencing (NGS), enabled the massively parallel sequencing of millions of untargeted DNA molecules (Margulies et al., 2005, Bentley et al., 2008, Mardis, 2008), which transformed the field of aDNA. This method allowed, for the first time, the development of quantitative estimates of modern-DNA contamination (Krause et al., 2010a, Fu et al., 2013, Korneliussen et al., 2014, Renaud et al., 2015), which, coupled with new methods for the retrieval of highly degraded DNA and whole genome targeting (Hodges et al., 2009, Meyer and Kircher, 2010, Dabney et al., 2013a, Fu et al., 2013, Gansauge and Meyer, 2013), has greatly expanded the amount of data that can be retrieved from ancient specimens and has drastically increased our knowledge of the human past. As a result, genome reconstruction from a range of ancient specimens, such as bones (Green et al., 2010, Krause et al., 2010b), teeth (Reich et al., 2010, Bos et al., 2011), dental calculus (Adler et al., 2013, Warinner et al., 2014), hair (Rasmussen et al., 2010), plant remains (Mascher et al., 2016, Swarts et al., 2017), soft tissue (Kay et al., 2015, Schuenemann et al., 2017), as well as soil specimens from archaeological contexts (Slon et al., 2017), has aided the understanding of evolutionary processes in real time with particular focus on the influence of ancient (or extinct) species in the genomic composition of their extant relatives. In only the last 10 years, the field has provided key insights into the past, which would have been impossible to gain from short PCR amplicons, or from modern DNA alone. Some of its most important discoveries on early human evolution include the genome sequencing of Neanderthals, which revealed evidence of their hybridization with modern humans (Green et al., 2010), the characterization of a previously unknown hominin group unearthed in Denisova cave, Siberia (Krause et al., 2010b, Reich et al., 2010), the discovery of today extinct modern human lineages that existed in Europe 42,000 years ago (Fu et al., 2015)

as well as the Eurasian genetic components that contributed to the first human expansions into the Americas (Raghavan et al., 2014, Rasmussen et al., 2014, Raghavan et al., 2015, Skoglund et al., 2015). In addition, the major cultural transitions observed in the last 10,000 years have also been delineated by ancient DNA discoveries. These include the establishment of the role of migrations in promoting the Neolithic transition in Europe (Haak et al., 2005, Skoglund et al., 2012, Lazaridis et al., 2014), the timing and dynamics of this transition in different parts of Eurasia (Lazaridis et al., 2016, Mittnik et al., 2018), the genetic processes involved in animal (Frantz et al., 2015) and crop (Mascher et al., 2016) domestication, and the discovery of substantial Bronze Age migrations into Europe, ~5,000 - 3,500 years ago, which may have been responsible for the spread of Indo-European languages (Allentoft et al., 2015, Haak et al., 2015).

1.2.2 Ancient pathogen genetics and genomics

The prospect of studying long-term pathogen evolution through aDNA became attractive during the field's early days, as it could potentially elucidate the history of infectious disease, as well as offer new lines of evidence for palaeopathological diagnoses of skeletal assemblages. In addition, it could shed light on historical epidemics and pandemics, particularly since in most acute mortality events the victims showed no indications of diagnostic skeletal lesions (Ortner, 2008). Like with the human studies, initial research on ancient pathogens, such as *Yersinia pestis* (Drancourt et al., 2004, Drancourt et al., 2007), *Mycobacterium tuberculosis* (Hershkovitz et al., 2008) and *Salmonella enterica* (Papagrigorakis et al., 2006), struggled with the limitations of PCR detection. Most research did not follow published guidelines on appropriate PCR typing compatible with aDNA (Cooper and Poinar, 2000), and their results could not be validated by independent analyses (Gilbert et al., 2004, Barnes and Thomas, 2006), which in certain cases led to extensive criticism and disbelief (Shapiro et al., 2006). In these experiments the potential contaminant amplicons likely resulted from microbial taxa associated with the researcher's own microbiome or from environmental microbes associated with the depositional context that bear high similarity to their targeted pathogenic relatives (Wilbur et al., 2009, Müller et al., 2016).

The high-throughput sequencing revolution (Koboldt et al., 2013) had a tremendous effect on the study of ancient infectious disease, since it enabled the reconstruction and

authentication of entire ancient pathogen genomes. Owing to this technology, the first ancient pathogen genome, released in 2011 (Bos et al., 2011), was that of *Y. pestis*, the causative agent of the infamous Black Death (1346-1353 AD). This study did not only put to rest alternative theories for the Black Death's causative agent (Duncan and Scott, 2005, Cohn, 2008), it also provided a comprehensive whole genome analysis of the ancient bacterium and its relationship to modern *Y. pestis* diversity (Bos et al., 2011, Cui et al., 2013). Similarly, *Y. pestis* was later also confirmed to be the causative agent of the Plague of Justinian (6th – 8th centuries AD) (Wagner et al., 2014, Feldman et al., 2016). In addition, numerous studies have characterised other ancient bacterial pathogens at the genomic level. The study of medieval *Mycobacterium leprae* (Schuenemann et al., 2013, Mendum et al., 2014) suggested a ~3,000 yBP origin of the bacterium and showed a European origin for leprosy in the Americas as well as a link between strains from medieval Europe and the present-day Middle East. The sequencing of *Mycobacterium tuberculosis* from the pre-Columbian Americas (Bos et al., 2014) revealed seals as a possible zoonotic transmission host and estimated the *M. tuberculosis* complex divergence time to only ~6,000 yBP, which contradicted studies on present-day strains suggesting that the bacterium accompanied human dispersals out-of-Africa (Comas et al., 2013). The study of ancient *Salmonella enterica* subsp. *enterica* serovar Paratyphi C (Vågene et al., 2018) suggested enteric fever as an epidemic candidate for the indigenous population declines recorded in the Americas during the early European contact period. Moreover, *Helicobacter pylori,* isolated from the stomach contents of a ~5,300-year-old ice-preserved mummy (Ötzi) (Maixner et al., 2016), showed that the bacterial population colonising this European Neolithic famer was solely of Asian origin, contrary to modern-day *H. pylori* populations in Europe that are hybrids between Asian and African strains. Regarding the same time period, the detection of *Y. pestis* 5,000 years ago showed the existence of the bacterium across Eurasia during the Late Neolithic and Bronze Age (LNBA) period and revealed important functional differences to its historical and modern relatives (Rasmussen et al., 2015, Andrades Valtueña et al., 2017). Other studies include a recent analysis of *Vibrio cholera,* which was retrieved from a 19th-century preserved intestinal specimen and demonstrated the potential of investigating historical medical collections (Devault et al., 2014a); the identification of medieval *Brucella melitensis*, which emphasized the rich

pathogen DNA content present in calcified nodules (Kay et al., 2014); and *M. tuberculosis* from an 18th-century Hungarian mummy (Kay et al., 2015), showing the preservation potential of ancient pathogen DNA in mummified tissue.

Regarding eukaryotic pathogens, recent studies have analysed DNA from the oomycete *Phitophthora infestans* from historical plant remains to reveal important features regarding the virulence and demography of the Irish potato famine (Martin et al., 2013, Yoshida et al., 2013). In addition, *Plasmodium falciparum* and *P. vivax* mitochondrial sequences have been reconstructed from 20th-century blood stains (Gelabert et al., 2016), potentiating novel DNA sources for pathogen retrieval.

Moreover, the study of virus evolution using aDNA has started to gain traction in recent years. Examples include the detection and genome reconstruction of Hepatitis B (Baele and Barbier, 1967, Kahila Bar‑Gal et al., 2012, Krause-Kyora et al., 2018, Mühlemann et al., 2018, Ross et al., 2018) and Variola (Biagini et al., 2012, Duggan et al., 2016) viruses retrieved both from bone and mummified tissue. Finally, although the survival of RNA in ancient specimens is still largely unexplored (Fordyce et al., 2013), analysis of decades-old formalin-fixed and frozen samples have provided insights into the geographic expansions of important RNA viruses. These include the initial introduction of HIV into the Americas (Worobey et al., 2016) and the analysis of influenza virus genomes that suggests an avian source for the Spanish flu pandemic (Taubenberger et al., 2005, Tumpey et al., 2005).

1.2.3 Methods for ancient pathogen detection and authentication
In the high-throughput era, almost all ancient specimens constitute complex metagenomic datasets (Warinner et al., 2017). Given that ancient endogenous DNA is usually present in low abundance, to date, the detection and authentication of ancient pathogens follow two main trajectories, namely (1) *in-vitro* pathogen detection using DNA enrichment techniques and (2) *in-silico* screening of shotgun datasets using metagenomic platforms. Both options require common laboratory processing methodologies, which are widely used in the ancient DNA field. These include the extraction (Rohland and Hofreiter, 2007, Dabney et al., 2013a) and NGS library preparation of DNA from an ancient specimen (Meyer and Kircher, 2010). In the latter step, libraries may be chosen either to retain the post-mortem deamination

characteristics of aDNA (Briggs et al., 2007) or to be treated with the enzyme uracil-DNA-glycosylase (UDG) for the complete (Briggs et al., 2010) or partial (Rohland et al., 2015) cleavage of uracil residues. Finally, an index combination is usually attached to DNA libraries through a PCR reaction for multiplex sequencing (Kircher et al., 2012).

In-vitro pathogen screening has followed both a specific and a general approach. In the specific approach, genes or plasmids are targeted in order to evaluate the presence of a single microorganism either via qPCR detection (Schuenemann et al., 2011, Harkins et al., 2015) or in-solution based DNA capture (Maricic et al., 2010, Schuenemann et al., 2011, Bouwman et al., 2012, Schuenemann et al., 2013, Bos et al., 2014). This method is relevant when the target organism is known (Schuenemann et al., 2013, Bos et al., 2014). Instead, studies have also relied on a general approach involving the simultaneous detection of several pathogens via array-based capture (Devault et al., 2014b, Bos et al., 2015). Such a method is particularly useful in cases where the target organism is not known, i.e when the pathological features identified on an ancient skeleton are unspecific or when mortality events have not left any diagnostic skeletal changes (Vågene et al., 2018).

In-silico pathogen detection has recently gained popularity within aDNA, offering a more general approach towards disease diagnosis and metagenomic profiling of ancient datasets. In addition to the traditionally used tools, such as the Basil Local Alignment tool (BLAST) (Altschul et al., 1990), some of the recently developed and widely applied metagenomic frameworks in aDNA also rely on competitive read alignment approaches. Examples of these are the MEGAN ALignment Tool (MALT) (Vågene et al., 2018) and MetaPhlAn (Segata et al., 2012) that is also integrated into the metagenomic pipeline metaBIT (Louvel et al., 2016, Librado et al., 2017). MALT is a taxonomic binning algorithm able to utilize custom whole-genome databases, such as the National Center for Biotechnology Information's RefSeq database (O'Leary et al., 2015), achieving specificity in read assignment to different taxonomic levels. Instead, MetaPhlan relies on a database of thousands of marker genes, which are selected to distinguish specific microbial clades.

It is important to note that detection of microbial taxa of interest by metagenomic tools alone cannot be considered particularly informative or diagnostic of authentic ancient pathogen DNA. These analyses should always be coupled with post-processing validation of reads, i.e through the evaluation of their genetic distance to the matching taxa (or else called edit distance) as well as through assessment of post-mortem damage signatures consistent with the presence aDNA (Key et al., 2017). In this thesis, I have utilised the aforementioned methodologies for the detection, authentication and, ultimately, the genome reconstruction of the highly pathogenic bacterium *Yersinia pestis* from ancient remains.

1.3 The plague bacterium, *Yersinia pestis*

1.3.1 Modern ecology and disease

Although plague is often used as a generic term for describing a pestilence, it also represents a well-defined infectious disease caused by the Gram-negative bacterium *Yersinia pestis* (Perry and Fetherston, 1997). The bacterium's first description is commonly attributed to Alexandre Emile Jean Yersin, who was able to culture it under unfavorable circumstances and demonstrate its common presence in humans and black rats in Hong Kong (1894) during the third plague pandemic (Treille and Yersin, 1894, Pollitzer, 1954). This, most recent of all plague pandemics began in 1855 in the Yunnan province of China, reached Hong Kong and Canton in 1894 and from there spread worldwide via steamships (Pollitzer, 1954). By the year 1903 it had claimed the lives of at least 10 million people in India alone. Although in the 1950s the pandemic was officially declared to be over, its legacy remains in certain parts of the world and is occasionally conveyed through small-scale epidemics (Bertherat, 2016). Currently, the most severely affected region is Madagascar, where the bacterium has persisted ever since its first introduction (1898) and is responsible for ~500 cases annually (Brygoo, 1966, Vogler et al., 2011, Vogler et al., 2013, Bertherat, 2016, Vogler et al., 2017).

Despite its ability to cause severe disease in humans, *Y. pestis* is not human-adapted. Instead, it is sustained among sylvatic rodents such as great gerbils, marmots, prairie dogs and mice (Pollitzer, 1960, Gage and Kosoy, 2005). These carrier populations are termed rodent reservoirs, or foci, and are found nearly worldwide. While most foci were established during the third pandemic (i.e. the ones in the U.S.A and in Madagascar)

(Morelli et al., 2010), others (such as the ones found south of the Ural Mountains as well as in Central and East Asia) are thought to have persisted for millennia (Tikhomirov, 1999, Anisimov et al., 2004, Cui et al., 2013).

The transmission of *Y. pestis* is facilitated by an arthropod vector. Although the oriental rat flea, *Xenopsylla cheopis,* is the best yet described, others such as *Oropsylla montana*, which is the North American ground squirrel flea, are likely to also play an important role in the bacterium's maintenance and dissemination (Eisen et al., 2009, Hinnebusch et al., 2017a). This continuous cycle of the *Y. pestis* transmission between wild rodents and their fleas is mainly driven by ecological and climatic factors as well as by population density (Keeling and Gilligan, 2000, Davis et al., 2004, Gage and Kosoy, 2005, Stenseth et al., 2006). A disruption of this equilibrium is what causes the so-called "epizootic" phases, which are best described as rodent epidemics. During this time rodent populations decrease, forcing fleas to seek alternative hosts, and it is the time when secondary transmission to human, domestic rodents (rats) and other mammals is most likely to occur (Gage and Kosoy, 2005).

Human disease can manifest in three clinical forms, namely the bubonic, pneumonic and septicaemic forms, with bubonic plague being the most common (Perry and Fetherston, 1997). Subsequent to the bite of an infected flea, bacteria travel via the lymphatic system to the nearest lymph node, where they replicate excessively giving rise to the so-called "buboes", the most typical disease symptoms (Wren, 2003). From there, bacteria can then penetrate other tissues. Invasion of the lung tissue manifests as secondary pneumonic plague, which causes up to 100% mortality when left untreated (Dennis et al., 1999). In addition, it is the only form of the disease that is capable of direct human-to-human transmission via aerosols. Finally, septicaemic plague can result both from (1) bacterial invasion of the blood directly from the site of the fleabite (primary septicaemic form, 10-30% of all plague cases), and (2) as a secondary manifestation after primary bubonic plague (Dennis et al., 1999, Stenseth et al., 2008).

1.3.2 *Y. pestis* evolution and earliest molecular signatures
The closest known relative of *Y. pestis* is the soil dwelling and mild enteric disease-causing bacterium *Y. pseudotuberculosis* (Achtman et al., 1999). Their divergence time is not know, as the currently estimated date interval is very broad (13,000 - 79,000 yBP)

(Rasmussen et al., 2015, Andrades Valtueña et al., 2017), and thus difficult to elucidate with the available data. The earliest *Y. pestis* genomic signatures have been identified in humans during the Late Neolithic/Bronze Age (LNBA) period, ~5,000 years ago (Rasmussen et al., 2015), which suggests that *Y. pestis* emerged prior to that time, although its precise geographic source is also unknown. Despite their genomic similarity reaching 97% among protein coding genes (Chain et al., 2004), *Y. pestis* and *Y. pseudotuberculosis* are clearly distinct in terms of their pathogenicity and transmission mechanisms. *Y. pestis* is known to have acquired its high virulence and complex niche mainly by gene loss and the acquisition of two plasmids, pPCP1 and pMT1 (Parkhill et al., 2001). Plasmid pPCP1 functions to enhance bacterial dissemination across mammalian tissues (Zimbler et al., 2015), whereas pMT1 is one of the components that confer the bacterium's adaptation to the flea vector (Hinnebusch et al., 2000, Du et al., 2002, Hinnebusch et al., 2002).

The bacterium's ability to colonize the digestive tract of fleas is central to its most commonly causing bubonic disease form. After a flea receives a blood meal from an infected individual, bacteria colonize a portion of its foregut, the proventriculus, where they form a biofilm (Sun et al., 2014, Hinnebusch et al., 2016). As bacteria replicate within the biofilm, it becomes enlarged, blocking the proventriculus and preventing blood meals from passing into the flea's midgut (Hinnebusch et al., 2016). Consequently, blocked fleas that attempt to feed end up regurgitating bacteria into the bite site, hence facilitating transmission to new hosts.

At present, the genes identified as enabling flea colonisation and biofilm-dependent blockage in *Y. pestis* are *ymt* (on the pMT1 plasmid) as well as PDE-2, PDE-3, *rcsA* and *ureD* (on the chromosome) (Sun et al., 2014). Interestingly the oldest *Y. pestis* strains identified so far, dating to 5,000 - 3,500 years ago, do not possess any of the genetic components consistent with flea-borne transmission (Rasmussen et al., 2015, Andrades Valtueña et al., 2017). These strains share a common phylogenetic branch and are ancestral to all historical and extant *Y. pestis* lineages sequenced to date. In addition, although they seem to have been widespread across Eurasia during that time (Rasmussen et al., 2015, Andrades Valtueña et al., 2017), their virulence and transmission mechanisms to human populations remain unclear. The earliest *Y. pestis*

signatures compatible with flea adaptation were identified in an Iron Age individual (~3,000 yBP) from the Caucasus region. This isolate, however, has limited phylogenetic resolution due to its low genomic coverage (Rasmussen et al., 2015), and therefore, the evidence regarding the phylogeography and timing of flea-adaptation in *Y. pestis* is still incomplete.

1.3.3 The first historically recorded pandemic

The "Plague of Justinian" (541 - 750 AD) is the first historically recorded pandemic attributed to *Y. pestis* (Little, 2007). While its geographic origin is not known, some of the first reported outbreaks occurred in the Eastern Mediterranean region, specifically in the Egyptian port city of Pelusium, in 541 AD. From there, it moved eastward and soon after reached the capital of the Byzantine (or Eastern Roman) Empire, Constantinople, in 542 AD (Little, 2007, Stoclet, 2007). Within the next two years it spread through southern Europe, reaching as far west as Spain and Southern France (Kulikowski, 2007). After the pandemic's first wave, which ceased at 545 AD, subsequent outbreaks continued to affect Europe until the mid 8th century (~750 AD) (Morony and Little, 2007). The disease is thought to have claimed a significant portion of the Byzantine population at the time. Still, the overall mortality estimates are highly debated mainly due to the absence of archival data and the lack of mass burials in Europe dating to the 5th - 8th centuries AD, though the latter has recently been revisited (McCormick, 2015).

Molecular signatures of *Y. pestis* in Justinianic plague victims were identified through a series of PCR studies (Wiechmann and Grupe, 2005, Harbeck et al., 2013). More recently, whole genomes were published (Wagner et al., 2014, Feldman et al., 2016) from victims buried in two early medieval cemeteries in Bavaria, southern Germany. Interestingly, this region lacked historical accounts of the pandemic, hence, demonstrating the significance of aDNA analysis and its potential to provide information beyond the historical record. In addition, phylogenetic analysis showed that the plague of Justinian was associated with a previously unidentified *Y. pestis* lineage that today is seemingly extinct (Wagner et al., 2014). This lineage appears most closely related to contemporary isolates from China, hinting towards an East Asian origin, rather than Africa, its historically proposed source (Wagner et al., 2014). However, given the profound East Asian bias in the available modern data (Cui et al., 2013)

(Figure 1), as well as the limited genomic information that exists for the first pandemic (Wagner et al., 2014, Feldman et al., 2016), the question of its origin remains open.

1.3.4 The second historically recorded pandemic

The mass burials frequently discovered throughout Europe and attributed to the repeated medieval plague epidemics of the 14th – 18th centuries emphasize the demographic impact of the disease at the time. This period, referred to as the second plague pandemic, began with the infamous "Black Death" of Europe (1346-1353) (Gottfried, 1983, Benedictow, 2004). Although its birthplace has not yet been identified, the first historically recorded outbreaks occurred in 1346 in the Lower Volga region, particularly the cities of Sarai and Astakhan (Gottfried, 1983, Benedictow, 2004). The initiation of the pandemic is popularly marked by the famous "Siege of Kaffa" in the Crimean peninsula. The disease is thought to have erupted among the Mongol-Tatar army of the Golden Horde, which attempted to displace Genoese merchants from their trade centre by hurling plague-infected corpses into the besieged city (Wheelis, 2002). Consequently, the surviving Genoese who managed to flee the city are believed to have introduced the disease into Southern Europe. As a result, over the next five years the bacterium spread across the entire continent and claimed the lives of an estimated 60% of the European population (Benedictow, 2004).

The first *Y. pestis* genome from this time was released in 2011 from Black Death victims buried in the East Smithfield cemetery in London (Bos et al., 2011). This, and a follow-up study focusing on modern isolates from East Asia, revealed that the Black Death was temporally associated with a sudden expansion of *Y. pestis* diversity (Cui et al., 2013). Nevertheless, in comparison to its modern relatives, this strain bears no differences in genetic signatures associated with an increased virulence, making the reasons for its high mortality and fast spread unclear (Bos et al., 2011). In addition, phylogenetic analysis revealed that modern strains that are immediately ancestral to the London Black Death strain have been isolated from China near historical trade routes, suggesting an East Asian origin for the pandemic (Cui et al., 2013). However, this argument is still contentious due to the scarcity of historical and archaeological evidence from China dating to this time period (Sussman, 2011).

After 1353 *Y. pestis* caused thousands of successive outbreaks throughout Europe until the late 18th century AD (Biraben, 1975, Buntgen et al., 2012), when the disease essentially disappeared for yet unknown reasons (Appleby, 1980). Notable examples of some of the last epidemics recorded in the continent include the Great Plague of London (1665-1666 AD) and the Great Plague of Marseille (1720-22 AD). The question of whether such recurrent outbreaks were a result of the bacterium's long-term local persistence or instead were caused by its multiple introductions from distant plague foci has been a topic of active debate and is one of the main research focuses of this thesis. Published PCR-based SNP typing has proposed the presence of two distinct genotypes in Europe during the 14th century AD, suggesting that they entered through different pulses (Haensch et al., 2010). In addition, a study based on climatic proxies has recently suggested that post-Black Death outbreaks recorded in some of the main Mediterranean ports overlap temporally with extreme climatic fluctuations in Central Asia (Schmid et al., 2015). As a result, it was suggested that climatic fluctuations in that region may have disrupted the existing rodent foci and driven recurrent introductions of the disease into Europe (Schmid et al., 2015). By contrast, a PCR-based study has proposed the long-term persistence of plague in Europe between the 14th-18th centuries AD (Seifert et al., 2016), which is in line with some historical sources (Carmicheal, 2014), and a recent whole genome analysis of isolates from the Great Plague of Marseille (1720-1722 AD) revealed the existence of a previously unidentified *Y. pestis* lineage that is today extinct. Phylogenetic analysis showed that this lineage originated from strains associated with the European Black Death, providing evidence for a shared genetic history between *Y. pestis* isolates in Europe between the 14th - 18th centuries. However, since Marseille was a highly active Mediterranean port at the time (Signoli et al., 1998), distinguishing between a European and an external source for this outbreak has been challenging.

2 Thesis objectives

In this thesis, the primary investigational goal was to gain a better understanding of the genetic history of *Yersinia pestis*, through a systematic screening of mass and single burials from potential plague contexts, spanning the last 5,000 years of human history. The oldest time period investigated here was the Bronze Age. Phylogenetic analysis of published *Y. pestis* genomes shows a tremendous parallelism to the intensified human migrations of that time but suggests that the bacterium was unable to efficiently transmit via the flea vector between 5,000 - 3,000 yBP. As part of paper A, I have analysed human teeth from Samara in Russia (~4,000 yBP) to address the emergent question concerning the timing of flea adaptation in *Y. pestis*, its association with human populations from that time as well as its relationship to extant *Y. pestis* lineages from the same geographic region.

The first historically recorded plague pandemic, the "Plague of Justinian" (6th – 8th centuries AD), is considered to have had a tremendous demographic impact on the Byzantine Empire. Nevertheless, the only *Y. pestis* genomic data that exists from this time period derives from single burials in Southern Germany where no historical record of the pandemic survives. One of the aims of paper B was to better understand the diversity of the bacterium in the Mediterranean as well as to characterise the maternal lineage diversity of its human population during the early medieval period.

The most broadly studied period in this thesis is the second plague pandemic (14th – 18th centuries AD). Some of the main topics investigated in papers C and D concern its initial wave, the Black Death (1346-1353 AD), and are: What was the level of *Y. pestis* genetic diversity present in Europe during the Black Death and did this event contribute to extant diversity found around the world today? In addition, given that plague outbreaks continued to occur in Europe for almost three centuries after 1353, were they a result of recurrent introductions of *Y. pestis* into the continent or did the bacterium persist locally during that time? Finally, is there any genetic basis to explain plague's disappearance from Europe in the 18th century? To address these questions I present a comprehensive analysis of whole-genome *Y. pestis* data from single and mass graves in modern-day Spain, France, Germany, Switzerland, England and Russia dating from the 14th – 17th centuries AD.

3 Results

3.1 Earliest genomic evidence of bubonic plague during the Bronze Age (paper A)

Title: "Analysis of 3,800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague"

Maria A. Spyrou, Rezeda I. Tukhbatova, Chuan-Chao Wang, Aida Andrades Valtueña, Aditya K. Lankapalli, Vitaly V. Kondrashin, Victor A. Tsibin, Aleksandr Khokhlov, Denise Kühnert, Alexander Herbig, Kirsten I. Bos and Johannes Krause

Published in *Nature Communications*, 2018, 9(1): 2234

Study synopsis

The earliest identified molecular evidence of *Y. pestis* dates back to the Bronze Age period (~5,000-3,500 yBP). These strains are associated with human infections; yet, their mode of transmission to human hosts remains unclear, since they lack all necessary genomic signatures that confer adaptation to the flea vector. Using limited genomic evidence, previous studies have suggested that flea-adaptation in *Y. pestis* was not acquired before ~3,000 years ago. The present study offers solid evidence regarding the timing of flea-adaptation in *Y. pestis* and contrasts previous assumptions through three key findings:

1.   I have reconstructed two ~3,800-year-old genomes from Samara, in the present-day Russian steppe region, and identified an additional *Y. pestis* lineage present during the Bronze Age that is distinct from the one previously reported (LNBA). Virulence factor analysis showed that this lineage encompasses all genetic signatures of arthropod-adaption, resembling extant and historical bubonic-disease-causing strains. This result suggests the existence of two parallel lineages in Eurasia during the Bronze Age, which may have differed in their transmission and disease potentials.

2.   A Bayesian statistical framework was used to date the divergence time of this lineage from the rest of *Y. pestis* to ~4,000 yBP. Demographic analysis examining population size changes over time revealed that this time coincides with a sudden expansion of *Y. pestis* diversity, which may signify a prehistoric epidemic event. In addition, this analysis revealed that several lineages surviving until the present day in Asia, which are linked to bubonic plague cases, also diverged during the Bronze Age (>5,000 yBP). Taken together, the current results suggest that the ability of *Y. pestis* to

cause bubonic plague was present at least 1,000 years earlier than previously suggested.

3.   Analysis of the human genome from one of the *Y. pestis*-positive individuals (with 4.2-fold average coverage) showed a close genetic similarity to Late Bronze Age populations from the same region but also to contemporaneous groups from central Europe and East Asia. This result supports the idea of intense human mobility in Eurasia during the Bronze Age, which may have facilitated the spread of infectious diseases such as plague.

3.2 Maternal lineage diversity in early medieval Southern Italy (paper B)

Title: "Ancient DNA recovery and maternal lineage diversity of early medieval Venosa in southern Italy"

Maria A. Spyrou, Alessandra Sperduti, Åshild J. Vågene, Lorenzo M. Bondioli, Henrike Heyne, Eva Fernández-Domínguez, Luca Bondioli, Wolfgang Haak, Kirsten I. Bos and Johannes Krause

Manuscript ready for submission, (2018)

Study synopsis

The Italian peninsula has played a leading role in the history of the Mediterranean region, especially during antiquity and the medieval period. As such, its ports are considered main entry points of infectious disease into Europe. For example, it is considered one of the primarily affected regions during the first (6th – 8th centuries AD) and second (14th – 18th centuries AD) plague pandemics. In addition, although a great amount of human connectivity between southern Europe, Northern Africa and the Near East are known to have existed within the Mediterranean, their influence on the genetic make-up of the Southern Italian populations from these periods is not known and has been mainly inferred from modern-day genetic data. Today, the Italian population contains a geographically linked genetic structure, with its south showing a greater affinity to populations from the geographically neighbouring Near and Middle East as well as Northern Africa.

Here, I analysed material unearthed from multiple graves at the "Thermae di Venosa" in the Basilicata region of Southern Italy. The material was radiocarbon dated to the 7th – 8th centuries AD, thus suggesting that the multiple graves might have been the result of

a Justinianic plague epidemic in that region. The main results of this paper are the following:

1.   Screening of all individuals via a *Y. pestis*-specific DNA enrichment of the pPCP1 plasmid did not yield any evidence on that account, hence the diagnosis possibilities for this epidemic remain unclear.

2.   In addition, I performed human DNA analysis on the remains of 22 individuals, using whole mitochondrial genomes, to assess the overall DNA preservation at the site and examine the mtDNA diversity of this Southern Italian population during the early medieval period. Most of the detected mitochondrial haplogroups are in accordance with reported European haplogroup frequencies. However, the additional detection of some types that are rare in Europe today, namely L2b1a, R0a1a and N3a, suggests contacts of this region with Northern Africa and the Near East. Analysis of pairwise genetic distances of early medieval Venosa to both contemporary and historical populations from Europe, the Near and Middle East and Northern Africa show non-significant differences to the majority of the comparative groups.

Although the links detected here between Southern Italy, the Near East and Northern Africa, ought to be confirmed using nuclear DNA analysis in the future, the present manuscript suggests that the maternal lineage diversity of the Basilicata region could be traced as far back as the 7th century AD.

3.3 A genomic link between medieval and modern plagues (paper C)

Title: "Historical *Y. pestis* genomes reveal the European Black Death as the source of ancient and modern plague pandemics"

Maria A. Spyrou, Rezeda I. Tukhbatova, Michal Feldman, Joanna Drath, Sacha Kacki, Julia Beltrán de Heredia, Susanne Arnold, Airat G. Sitdikov, Dominique Castex, Joachim Wahl, Ilgizar R. Gazimzyanov, Danis K. Nurgaliev, Alexander Herbig, Kirsten I. Bos and Johannes Krause

Published in *Cell Host & Microbe*, 2016, 19(6), pp. 874-81

Study synopsis

The second plague pandemic has arguably been the most destructive of all plague pandemics, causing a population decline of up to 60% in Europe during the Black Death (1346-1353 AD) alone. In addition, the source of the successive epidemics recorded in

Europe after this initial wave and until the late 18th century is unknown. Currently climatic modelling approaches and genetic data offer contrasting evidence on this topic. In this study, I present genomic evidence from epidemic burials around Europe to better understand the relationship between the Black Death and post-Black Death epidemics/outbreaks. Using three newly reconstructed ancient *Y. pestis* genomes from Spain, Germany and Russia, the major findings of this paper include:

1.   A *Y. pestis* genome obtained from mid-14th-century individuals from Barcelona, Spain, shows an identical genotype to London Black Death genomes published previously, thus, suggesting a single entry and low genetic diversity of the bacterium in Europe during the initial pandemic wave.

2.   Whole genome sequencing of a post-Black Death strain from Ellwangen, Germany (1485-1627 AD) shows its close genetic similarity to previously published strains from the Great Plague of Marseille (1720-1722 AD). These isolates group on the same phylogenetic branch, which derives from the aforementioned Black Death isolates from Barcelona and London. This result supports the idea of a shared genetic history between European *Y. pestis* strains during this 400-year-long pandemic and, therefore, adds legitimacy to the notion of the bacterium's local persistence during that time.

3.   The key finding of this study stems from the reconstruction of a late 14th-century genome from Bolgar City (1363-1400 AD), in the Volga region of Russia. This genome is distinct from the aforementioned Black Death genomes from Barcelona and London, showing shared derived variants with 19th-century strains from China. These strains are known to have caused the most recent plague pandemic that began in the Yunnan province in 1855 and spread globally in the 19th and 20th centuries and which continues to cause epidemics in certain parts of the world, such as in Madagascar, until today. Overall, this finding provides evidence for a genetic association between the 14th-century Black Death and modern plague.

In summary, the results of this study give insights regarding the repeated epidemics recorded in Europe between the 14th – 18th centuries and suggest its local persistence during that time. Importantly, they also reveal the European Black Death as a source of contemporary outbreaks around the world.

3.4 A detailed genomic analysis of the second plague pandemic (paper D)

Title: "A phylogeography of the second plague pandemic revealed through historical *Y. pestis* genomes"

Maria A. Spyrou, Marcel Keller, Rezeda Tukhbatova, Elisabeth Nelson, Don Walker, Amelie Alterauge, Hermann Fetz, Joris Peters, Niamh Carty, Robert Hartle, Michael Henderson, Elizabeth L. Knox, Sacha Kacki, Michaël Gourvennec, Dominique Castex, Sandra Lösch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos and Johannes Krause

Manuscript ready for submission, 2018

Study synopsis

Although the previously described study revealed important phylogeographic features of the second plague pandemic (Spyrou et al., 2016), certain key questions still remain unaddressed. These include plague's route of entry into Europe during the early phases of the Black Death, the level of genetic diversity in the bacterium thereafter and the possible formation of a single or multiple reservoirs in Europe between the 14th and 18th centuries. Some of the main challenges in addressing these questions have been the sparse geographical sampling and the low sample size of *Y. pestis* genomes obtained from this time period. In this manuscript I present a substantial sampling effort, which together with a recently established *in-vitro* DNA enrichment technique allowed for the quadrupling of the genomic data that exists from the second pandemic. Specifically, I present 32 new whole *Y. pestis* genome sequences (>1-fold coverage) from nine epidemic burial sites in Germany, Switzerland, France, England and Russia. Analysis of this data has led to the following key findings:

1.   Sampling of additional Black Death isolates from Germany and France shows that they are identical to previously reported genomes from London and Barcelona, therefore confirming the idea of a single entry and a near absence of genetic diversity in the bacterium during the Black Death.

2.   The sequencing of two isolates from the 14th-century Laishevo, in the Volga region of Russia, revealed its ancestral genotype compared to the aforementioned Black Death isolates from Southern, Central and Western Europe. This result provides, for the

first time, genetic evidence for the bacterium's entry into the continent through Eastern Europe, an observation that is also supported by historical records.

3.   Moreover, additional genomes were reconstructed from later stages of the pandemic, namely from post-Black Death epidemics in Germany, Switzerland and England. Phylogenetic comparison of these new genomes with published extant and historical *Y. pestis* diversity remains in support of plague's persistence in Europe through local reservoir(s) between the 14th – 18th centuries and reveals an increased genetic diversity within this post-Black Death lineage.

4.   Analysis of the newly reported post-Black Death genomic diversity suggests the existence of two distinct phylogenetic clades within Europe. Bayesian reconstruction revealed that the distinction of these clades is not only corroborated by their geographic separation but also by their differences in substitution rates. The first clade, which seems to be evolving at a faster mutation rate, also appears more widespread within Europe. The second, instead, seems by the current data to be confined to local epidemics in Germany and Switzerland.

5.   The fastest evolving phylogenetic clade contains strains from some of the last epidemics recorded in Europe, namely from 17th century England and 18th century France. Intriguingly, through virulence factor analysis, I have identified the loss of virulence-associated genes, specifically the magnesium transporter genes *mgtB* and *mgtC*, within the genomes grouping on this clade. A full functional characterisation of these factors would be required in order to explain the implications of their loss in bacterial fitness.

In summary, this study demonstrates that increased sampling efforts were able to elucidate key aspects of *Y. pestis* evolution and provide novel insights into the initiation and progression of one of the most destructive historical pandemics.
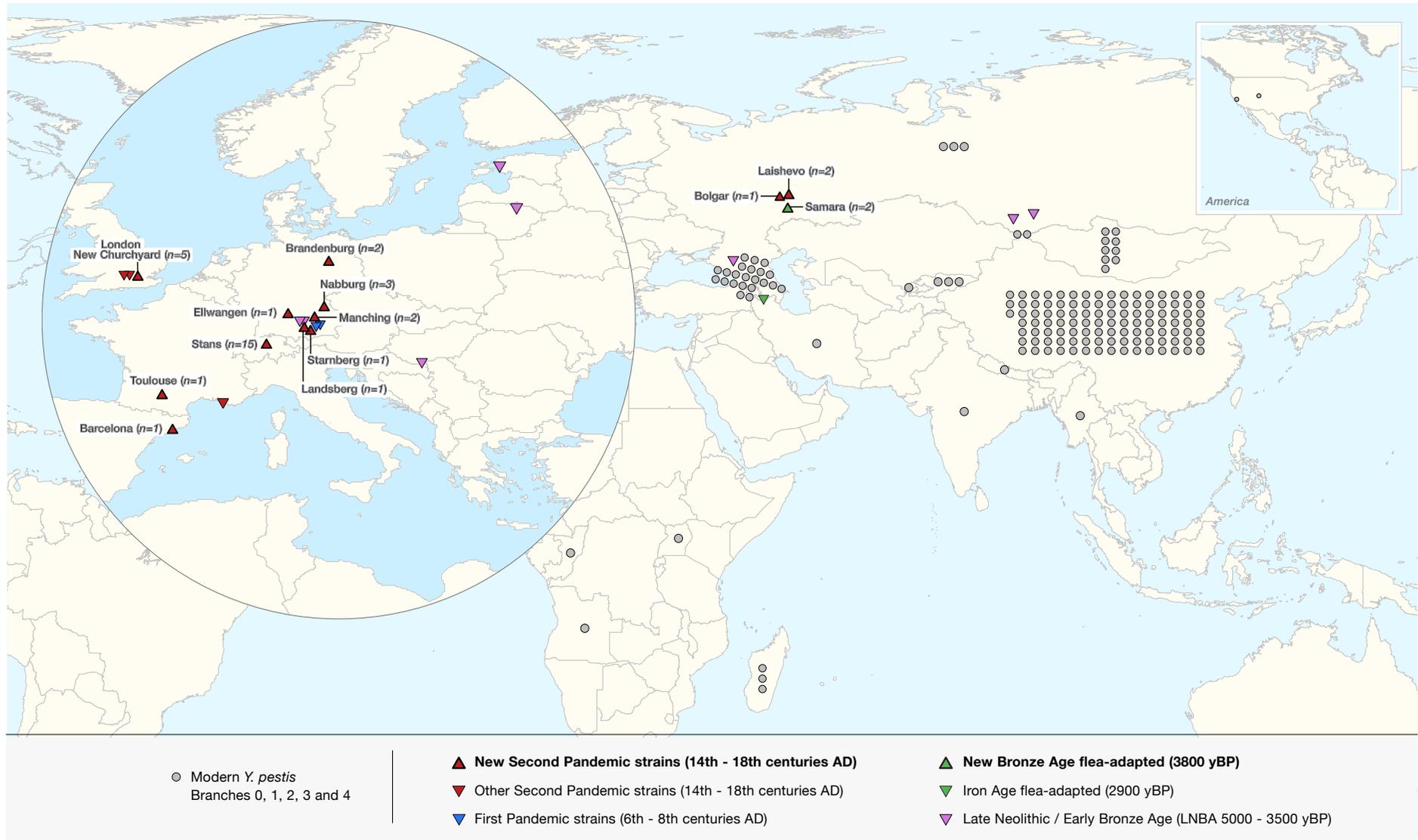
Figure 1 – Map showing all ancient and modern *Y. pestis* strains used in this thesis. The geographic locations of ancient *Y. pestis* strains reconstructed for this thesis are shown in triangles (dark black outlines) with the names of their locations and number of retrieved genomes (*n*) indicated. All other ancient strains are depicted using inverted triangles. Modern *Y. pestis* strains are shown in circles within the geographic region of their isolation.

This figure was created with graphical assistance from Michelle O'Reilly and Hans Sell.

4 Discussion

4.1 *Y. pestis* detection and genome retrieval (papers A, B, C & D)

Y. pestis was the first bacterium whose evolution was investigated through ancient genomics (Bos et al., 2011) and since then the last seven years of research have arguably taught us more about its history than for any other ancient pathogen (Wagner et al., 2014, Rasmussen et al., 2015, Bos et al., 2016, Feldman et al., 2016, Spyrou et al., 2016, Andrades Valtueña et al., 2017, Spyrou et al., 2018). Nevertheless, only 30 ancient genomes have been published to date, as their retrieval is still challenged by the low aDNA abundance and the interference of environmental contamination. Therefore, one of the main drivers of the currently presented research has been the adoption of sensitive and cost-efficient techniques for *Y. pestis* detection and genome reconstruction.

The human skeletal element that has been preferentially used for detection of *Y. pestis* in this thesis is teeth. This is because, as previously suggested (Drancourt et al., 1998, Schuenemann et al., 2011), blood-borne pathogens are most likely to reside in the desiccated blood vessels of the pulp chamber. Consistent with this hypothesis, the highest yields of *Y. pestis* DNA have to date been shown in DNA extracts from the pulp-chambers of tooth specimens (Margaryan et al., 2018).

As a first line of screening, a qPCR assay developed for the detection of 52 bp of the *Y. pestis*-specific *pla* gene present on the pPCP1 plasmid (Schuenemann et al., 2011), was often used in this thesis to screen DNA extracts and has proven useful for reducing large sample collections down to the few interesting specimens that harbour molecular signatures of the bacterium (papers C & D) (Bos et al., 2011, Schuenemann et al., 2011, Spyrou et al., 2016). In addition, a capture approach targeting the entire pPCP1 plasmid (Schuenemann et al., 2011) has also been used in this thesis for the same screening purpose (paper B). Contrary to the qPCR approach, the pPCP1 capture method gives the opportunity for NGS data generation that can be used for aDNA authentication and for the reconstruction of the entire *Y. pestis*-specific plasmid sequence (~9,000 base pairs). However, it should also be noted that this technique is more laborious and costly compared to the qPCR approach. Instead, in cases where shotgun-sequencing data were available, *in-silico* approaches have also been used (paper A). These can be

species/genus specific, i.e the competitive mapping approach used to distinguish *Y. pestis*-specific reads from those matching other *Yersinia* species (Andrades Valtueña et al., 2017), or following a more general approach through the use of metagenomic tools such as MALT (Vågene et al., 2018).

Genome reconstruction is the most important step for evolutionary inference and has been achieved here by both deep shotgun sequencing (paper A) (Rasmussen et al., 2015, Andrades Valtueña et al., 2017) and DNA enrichment methods (papers C & D) (Bos et al., 2011, Wagner et al., 2014, Bos et al., 2016, Feldman et al., 2016). While shotgun sequencing can reveal potential biases that may be introduced by targeted approaches (Cruz-Dávalos et al., 2017) (paper A) and could give insight into the acquisition of unknown genetic elements, it has to date not proven to be data and cost-effective. For example, in a recent study the generation of more than nine billion shotgun-sequenced reads distributed over seven specimens were necessary for the reconstruction of only two whole genomes with >1-fold coverage (Rasmussen et al., 2015). In comparison, the efficiency of the DNA enrichment method was most apparent in the last described study of this thesis (paper D) where, after capture, ~640 million sequenced reads yielded 32 new *Y. pestis* genomes with higher than 1-fold coverage (Figure 1). This is especially relevant for ancient DNA studies in general, where the genomic portion of interest usually makes up less than 1% of total DNA in a specimen. A direct comparison of yield achieved from shotgun-sequenced and captured data has been published for human DNA data, where the capture approach seems to have generated on average a four-fold higher genomic coverage using ~36-fold lower sequencing depth (Mathieson et al., 2015).

Two hybridization capture approaches were employed in this thesis, namely microarray-based capture and in-solution capture. The principal behind both techniques involves the *in-vitro* selection/enrichment of genomic loci of interest by exposure to a set of probes that are complimentary to the desired region (Hodges et al., 2009, Burbano et al., 2010, Fu et al., 2013). While microarray-based capture relies on the hybridization reaction happening on a glass slide where probes are immobilised (Hodges et al., 2009), the in-solution-based approach allows for the unrestricted movement of DNA probes within a reaction (Gasc et al., 2016), which may permit a greater hybridization capacity. Their comparative efficacies have not been published to date, though the in-solution

hybridization enrichment method is starting to gain popularity as it allows for the parallelisation of sample processing in 96-well plates where each sample is assigned an equal set of probes (papers A & D). Instead with microarray-based techniques several samples may be loaded on a single array and, therefore, competition over the same set of probes can influence their capture efficiencies. Such an effect was apparent in this thesis, where a single isolate from 16th-century Ellwangen (549_O) was first processed with others via microarray-based capture (Spyrou et al., 2016), yielding an 4.9-fold coverage, and later individually re-processed via in-solution capture (Paper D), which allowed for a three-fold increase in its genomic depth (14.1-fold coverage). In the future, a systematic comparison of enrichment techniques in order to explore their relative efficiencies will potentially allow the accessing of data from specimens with limited DNA contents and aid the study of *Y. pestis* evolution in the highest possible detail.

4.2 *Y. pestis* early evolution and links to human dispersals (paper A)

The Bronze Age (~5,000-3,200 years ago) was a period of major transformations in Eurasia, both from a cultural (Cunliffe, 2008) and a genetic perspective (Allentoft et al., 2015, Haak et al., 2015). Recent aDNA studies of human remains have demonstrated a large-scale migration of Yamnaya-related pastoralist groups from the Pontic-Caspian steppe both into Europe and into Central Asia (Allentoft et al., 2015, Haak et al., 2015). Their eastward migration is reflected in Early Bronze Age (EBA) populations from Central Asia and the Altai region, such as those associated with the "Afanasievo culture", which appear genetically indistinguishable from "Yamnaya" groups (Allentoft et al., 2015). In addition, their westward migration into Europe is considered to have had a genetic effect similar to, or in some parts even higher than, the preceding Neolithic transition (Haak et al., 2015). This component contributed to the genetic make-up of the Early Bronze Age (EBA) European populations such as the "Corded-ware-complex" and is present in different proportions across the entire continent until today (Allentoft et al., 2015, Haak et al., 2015). Finally, subsequent migrations from Europe back into the Eurasian steppe during the Middle and Late Bronze Age period (MLBA) are reflected through the appearance of European farmer-related ancestry in the genetic composition of Western and Central Asian populations such as 'Srubnaya', 'Sintashta' and 'Andronovo' (Allentoft et al., 2015).

This thesis contributes genomic evidence to show that the Bronze Age was not only a period of extensive human movement in Eurasia but also one of intense communicable disease transmission. Specifically, it deals with the genetic characterisation of *Y. pestis* during that time, revealing the parallel presence of two distinct lineages (Figure 2) that may have had different virulence and transmission characteristics. The first, also termed the 'LNBA lineage' has been previously analysed and seems to have been widespread across Eurasia between 5,000 and 3,500 years ago (Rasmussen et al., 2015, Andrades Valtueña et al., 2017). It consists of strains that occupy the most basal branch of the *Y. pestis* phylogeny and lack genetic signatures that cause blockage of the flea vector leading to biofilm-dependent transmission (Rasmussen et al., 2015, Andrades Valtueña et al., 2017). While their transmission mechanism is, therefore, unclear, intriguingly their phylogeographic patterns accurately parallel the above-mentioned human migrations happening during the same period (Andrades Valtueña et al., 2017). In addition, although the disease outcome of such strains has not been functionally assessed, their identification in remnants of human dental pulp suggests that they were able to invade human blood and were likely causing fatal illness. The second lineage presented here (Spyrou et al., 2018) was identified in Samara, Russia, and is distinct from the 'LNBA lineage' (Figures 1 and 2). It is phylogenetically more derived and falls close to extant strains identified today in Asia (i.e. lineages 0.PE2, 0.PE4 and 0.PE7). When using the specimens ^{14}C age as a tip calibration point for Bayesian divergence dating, its origin was estimated to ~4,000 years ago and that of it's extant relatives was estimated to ~5,200 - 4,400 years ago, suggesting that they also arose during the Bronze Age and were potentially circulating contemporaneously.

The lineage isolated from Samara seems to have all the necessary prerequisites for flea-borne transmission leading to a bubonic plague phenotype. These are the *ymt* gene present on the virulence plasmid pMT1 that is essential for colonisation of the flea's midgut (Hinnebusch et al., 2002, Sun et al., 2014), the pseudogenes PDE-2 and PDE-3 whose active forms are involved in biofilm degradation (Sun et al., 2011), the pseudogene *rscA* whose active form is a component of the Rcs pathway that inhibits biofilm formation (Sun et al., 2008) and the urease gene *ureD* that causes toxicity in fleas by releasing ammonia and is also a pseudogene in *Y. pestis* (Erickson et al., 2007, Sun et al., 2014). By contrast the 'LNBA lineage' seems to lack the *ymt* gene and

possesses the active forms of all the above-mentioned chromosomal genes, suggesting that it was unable to efficiently transmit via the flea vector.

Figure 2 – Maximum likelihood phylogenetic tree of ancient and modern *Y. pestis*. A Maximum Likelihood phylogenetic tree (95% partial deletion) was constructed with the program RaxML (Stamatakis, 2014), using (1) a worldwide dataset of modern *Y. pestis* strains, (2) previously reconstructed ancient genomes and (3) ancient *Y. pestis* genomes reconstructed for this thesis. The tree was run using 1000 bootstrap replicates and is based on 3,446 SNP positions. The number of strains within each branch is shown in brackets (*n*). The temporal transect covered by the ancient isolates ranges from ~4,800 – 230 years ago. Ancient *Y. pestis* strains reconstructed for this thesis are shown in bold print. Geographic region abbreviations are as follows: CHN (China), USA (United States of America), MDG (Madagascar), IND (India), IRN (Iran), MNM (Myanmar), RUS (Russia), GB (Great Britain), DE (Germany), SZ (Switzerland), FRA (France), MNG (Mongolia), NPL (Nepal), FSU (Former Soviet Union), CGO (Congo), and UGA (Uganda), LTU (Lithuania), EST (Estonia) and CRO (Croatia).

Although the evolutionary path that led to the biofilm-dependent transmission pathway in *Y. pestis* has been previously characterised and seems to be key for the bacterium's virulence (Sun et al., 2014), the alternative early-phase transmission (EPT) has recently been suggested as a potential propagation mechanism for the earlier 'LNBA lineage' (Andrades Valtueña et al., 2017). This form of transmission seems to occur in the initial stages of flea-colonisation by *Y. pestis* and is considered to be biofilm-independent (Vetter et al., 2010). While research has suggested EPT as a potential mechanism of *Y. pestis* transmission during plague epizootics among dense rodent populations (Eisen et al., 2015), its comparative efficiency in different flea and rodent species has only recently started to be systematically studied (Hinnebusch et al., 2017a, Hinnebusch et al., 2017b, Bland et al., 2018). In addition, although the silenced variant of *ureD* is also an essential prerequisite for flea-colonisation, its functionality has instead only been assessed independently of the genes involved in biofilm formation. Previous studies have shown that while an active *ureD* gene causes toxicity-induced mortality in fleas, ~20-40% of infected fleas can still survive (Erickson et al., 2007, Chouikha and Hinnebusch, 2014). It is, therefore, becoming increasingly important that the collective effect of those genes be studied in order to better understand the precise transitions that occurred in *Y. pestis* virulence and transmission during prehistory.

In recent years, it has become apparent that the demographic history of *Y. pestis* during the Bronze Age is a complex one. The frequency in which the 'LNBA lineage' has been identified suggests that it predominated in Eurasia at least between 5,000 and 3,500 years ago. Nevertheless, given that its genomic characteristics have not been identified in any *Y. pestis* isolate across the world today, this implies that it most likely became extinct sometime after 3,500 years ago. Although the exact reasons for that are not known, there is a clearly arising hypothesis that involves its inefficient transmission. In addition, the present thesis provides evidence that additional types similar to modern bubonic-disease-causing strains began to appear in Eurasia > 4,000 years ago, however their frequency of occurrence is still uncertain. It is, therefore, possible that the ancestral LNBA lineage was readily replaced by others that were better adapted to the arthropod-vector (Spyrou et al., 2018) and, hence, had the opportunity for a more efficient propagation to mammalian hosts/reservoirs.

From a wider perspective, the Bronze Age was a period of intense population and disease movement across Eurasia. *Y. pestis* may have been only one of several prevalent infectious pathogens, and their identification will help build a more complete picture of disease transmission during that time. In addition, infectious diseases are known to pose strong selective pressures on human populations (Karlsson et al., 2014). A recent study has given a first perspective on the potential infectious-disease-associated immune variants that may have risen in frequency during the Neolithic and Bronze Age periods, such as some on the MHC-complex and on pathogen recognition receptors (i.e. *TLR1-6-10*) (Mathieson et al., 2015). Other variants like the famous CCR5-Δ32 deletion that is at high frequency in Northern European populations today and is known to give complete or partial resistance to HIV infections, were also found to have risen in frequency >5,000 years ago although possibly through neutral processes (Sabeti et al., 2005). The re-evaluation of previously described as well as newly identified human genetic variants using larger aDNA data cohorts, and in association with known disease occurrence during that time, will help pinpoint the key events that led to modern pathogen susceptibility and prevalence across Eurasia.

4.3 *Y. pestis* and the historical plagues (papers B, C & D)

4.3.1 The period of the first pandemic (paper B)

After the Bronze Age, two historical plague pandemics are known to have devastated Europe between Late Antiquity and the Early Modern Era. The first pandemic, or Plague of Justinian, occurred between 541-750 AD and is thought to have mainly spread across the Mediterranean. Yet, the only genomic signatures of *Y. pestis* that exist to date are from two early medieval cemeteries in southern Germany. In this thesis (paper B), I screened human teeth from a 7th - 8th-century epidemic cemetery in Venosa, Southern Italy, for *Y. pestis* molecular signatures but was unable to detect the presence of the bacterium's DNA in the 22 studied remains. Yet, analysis of the human mitochondrial DNA from these individuals suggested that the Southern Italian population from that time was diverse and revealed potential links with nearby Northern African and the Near Eastern populations. Although these links may have been preserved in the population from earlier time periods, such as the Roman times or during the first phase of the Migration Period, their sources could potentially be

specified with more precision using nuclear DNA analysis. Nevertheless, this study gives a view of the early medieval matrilineal diversity in Southern Italy and the interconnectivity that may have existed within different regions of the Mediterranean during historical times.

Such contacts may have also facilitated the spread of pathogens other than *Y. pestis*. For example, a disease that is considered to have been prevalent during that time in the Mediterranean is malaria (Sallares et al., 2004). DNA from the malaria parasite has been previously detected in human remains from Southern Italy dating to the Roman period (Marciniak et al., 2016) as well as from more recent times (Gelabert et al., 2016), which shows the feasibility of further detection in ancient human remains. In addition, other epidemic agents that may have been present and more difficult to identify in ancient material are the RNA viruses, some of which have epidemic potential as they can cause haemorrhagic fevers. RNA viruses may be more challenging to recover in aDNA datasets given the unstable structure of their genetic material (Fordyce et al., 2013). The future development of approaches that aid the recovery of shorter and single stranded molecules (Gansauge and Meyer, 2013, Gansauge et al., 2017), or the employment of proteomics to detect specific RNA virus peptides/proteins (Hendy et al., 2018), may help identify the causative agent responsible for the possible Venosa epidemic.

4.3.2 Insights into the second plague pandemic (papers C & D)

The second plague pandemic, which lasted between the 14th and 18th centuries AD, has been a major focus of this thesis (papers C & D). The geographic origin of the pandemic remains hypothetical, as the last 50 years of research have revealed a substantial scholarly disagreement over this topic. One of the famous historical accounts by William H. McNeill (McNeill, 1998) suggested the Yunnan-Burma plague focus as the source for the second pandemic. According to that account, the Mongol army invaded this region in 1252 and then brought infected fleas back into their homeland, the eastern Eurasian steppe. It was either from there or from the original Yunnan-Burma focus that the pandemic likely erupted in 1331. In addition, analysis of modern *Y. pestis* genomic sequences is supportive of this view and has suggested the Qinghai-Tibet Plateau as a potential origin for the pandemic (Cui et al., 2013). Sylvatic rodent reservoirs within

this region are located near historical trade routes and seem to harbour strains that are directly ancestral to the ones isolated from the Black Death epidemic cemeteries in Europe (Cui et al., 2013). By contrast, others have come to challenge these ideas by suggesting that there is no clear evidence of epidemics in East Asia that predate the Black Death in Europe (Alexander, 1980, Benedictow, 2004, Sussman, 2011, Benedictow, 2016). In fact, the only known archaeological evidence of a pre-Black Death epidemic in that region is from a Christian cemetery excavated in the 19th century near lake Issyk-Kul, in east Kyrgyzstan, although its association with plague seems to be arbitrary (Norris, 1976, Benedictow, 2004). In addition, it is of note that most of the modern *Y. pestis* diversity that has been sequenced to date was isolated from China (Figure 1), which creates a profound bias in the available comparative data and ignores the genetic diversity harboured within the highly dense plague foci of Western and Central Asia (i.e those in modern-day Kazakhstan) (Anisimov et al., 2004). In addition, contrary to McNeill's proposals, some of the Central and West Asian foci are known to have existed for millennia and were likely established well before the 14th century (Kislichkina et al., 2015, Zhgenti et al., 2015, Eroshenko et al., 2017).

The first confidently recorded plague epidemics of the second pandemic occurred in Eastern Europe, specifically in 1346 around the Lower Volga and Black Sea regions of modern-day Russia (Benedictow, 2004). Consistent with this scenario, this thesis provides direct molecular evidence of the bacterium entering Europe through western Russia through the reconstruction of a 14th-century strain from the city of Laishevo in the Middle-Volga region (Figure 1). Intriguingly, this strain's phylogenetic placement is only one point mutation (SNP) away from a rapid *Y. pestis* lineage expansion that accounts for ~70% of the strain's diversity sampled around the world today (Figure 2). This radiation event is often attributed to the initiation of the Black Death (Cui et al., 2013). In addition, the strain from Laishevo appears to be a direct ancestor of Black Death isolates from Spain, England, France and Germany (papers C & D) that are genetically identical (Figure 2). Collectively, the evidence presented here from the Black Death period (1346-1353) suggests that the disease entered Europe through the east and did not accumulate genetic diversity after reaching the Mediterranean (Spyrou et al., 2016). Although a more exhaustive sampling would be necessary to verify this claim, the current data suggests a single wave entry, in contrast to previous assumptions

of multiple strain introductions during this first pandemic wave (Haensch et al., 2010).

The Black Death's aftermath continued in Europe for ~400 years through thousands of regional epidemics that lasted until the Early Modern Era (Biraben, 1975, Buntgen et al., 2012). Their source has long been contentious, especially since today no plague foci exist west of the Black Sea. Previous research has suggested that these plague epidemics may have resulted from independent introductions of the bacterium into Europe, potentially through climatic fluctuation that affected plague foci in Central Asia (Schmid et al., 2015). In papers C and D of this thesis, I present the analysis of genomic data from seven European sites associated with post-Black Death epidemics. Analyses of 27 ancient *Y. pestis* strains from those sites showed that they group into two genetically distinct lineages. Both lineages derive from Black Death strains isolated in Europe, thus suggesting that they arose within the continent and were likely not the result of multiple disease introductions.

The first of the two lineages has to date been isolated from late 14th-century isolates in London (St. Mary Graces, 1350-1400 AD) (Bos et al., 2011, Bos et al., 2016) and the city of Bolgar in Russia (1362-1400 AD) (Spyrou et al., 2016) as well as from modern day Sub-Saharan Africa (1.ANT) and from strains associated with the third plague pandemic (Hong Kong plague) that began in China during the late 19th century and disseminated globally within the next 50 years (Cui et al., 2013) (Figure 2). Although the bacterium's arrival route to Sub-Saharan Africa is intriguing, it is to date unresolved. Nevertheless, the fact that medieval strains from Europe are on the same evolutionary trajectory as those that caused the third plague pandemic in East Asia suggests that, subsequent to the Black Death, the bacterium spread eastward and later became the source of contemporary plague epidemics. Such a claim is consistent with historical records, suggesting that plague only reached the cities of Pskov and Novgorod in Eastern Europe between 1352-1353 after cycling through Southern, Central and Northern Europe between 1347 and 1351 (Alexander, 1980, Benedictow, 2016). At present, the data relevant to plague's eastward expansion derive mainly from medieval Europe and modern day China (Cui et al., 2013, Bos et al., 2016, Spyrou et al., 2016). Therefore, bridging the temporal and geographic gaps between Europe and East Asia will help delineate plague's medieval transmission routes and perhaps uncover the foci it established during this expansion.

Moreover, the second lineage that derives from Black Death strains seems confined to isolates from 14th - 18th-century Europe and has no modern relatives, suggesting that it is today likely extinct (Figure 2). Specifically, it includes strains from 14th - 17th-century Germany and Switzerland, as well as 17th and 18th-century England and France (Figures 1 and 2) that phylogenetically show a precise temporal structure up until a split point that coincides with an isolate from Ellwangen (1486-1627 calAD) (Spyrou et al., 2016). Based on the current data, this split seems to have given rise to two genetically and geographically distinct clades (paper D) (Figure 2). The first clade is comprised of closely related strains from Germany and Switzerland. Research on historical records has proposed that an Alpine plague focus may have facilitated the spread of plague in that region and has suggested the endemic marmot as a possible host reservoir species (Carmicheal, 2014). While the data presented in this thesis do not have the resolution to resolve the employed scenario, the genetic similarity of isolates within this clade lends some support to the existence of localized transmission routes in Central Europe. Instead, the second clade is seemingly more widespread and gave rise to strains isolated from England and the Mediterranean coast of France, both of which stem from some of the last documented European epidemics. The strains from England are from an epidemic in London and are dated to either 1603 or 1625 AD, which are among the last outbreaks recorded in that region. The strains from France derive from the well-described Great Plague of Marseille (1720-1722 AD), which marks one of the last major continental outbreaks (Bos et al., 2016).

Paper D shows that the clade encompassing strains from London and Marseille is not only geographically distinct from the one confined to Germany and Switzerland, but that it also evolved with a higher mutation rate. An overdispersion of evolutionary rates across the *Y. pestis* genealogy has been described in previous research, where higher substitution rates were thought to coincide with epidemic events (Cui et al., 2013). Such substitution increases are not considered to be associated with natural selection, but rather with the higher replication rates per unit time that occur during epidemics in comparison to enzootic bacterial maintenance (Cui et al., 2013). In addition, such accelerations have also been suggested to be more apparent during long-distance bacterial transmission (Morelli et al., 2010). Therefore, our results support previous suggestions, though the means of transmission for these strains is unclear. Although

their source was likely within Europe or its vicinity, the fact that both outbreaks (London and Marseille) occurred in central port cities may suggest a within-Europe marine transmission route.

A third point of distinction between the two clades concerns the genetic make-up of isolates that encompass them. The genomes from 17th-century London and 18th-century Marseille, which also show a higher mutation rate, seem to lack two important virulence factors, namely the genes *mgtB* and *mgtC* (paper D). Both genes are important for the early phase of infection since, as part of the PhoP/PhoQ regulatory system, they function to sequester Mg^{2+} ions in low magnesium environments such as the ones encountered within macrophages (Grabenstein et al., 2004). A recent study has shown that depletion of *mgtB* is associated with a high attenuation of *Y. pestis* in mice, likely by compromising their macrophage uptake (Ford et al., 2014). Additional functional assessment of both factors in various rodent species may reveal key aspects of their virulence significance, which is especially relevant since they are associated with some of the last recorded outbreaks of the second pandemic.

The second plague pandemic ended with plague's disappearance from Europe during the Early Modern Era (18th century) for reasons that are currently not understood. One of the most popular theories concerns the change of domestic rodent populations during the 18th century in Europe, specifically the replacement of *Rattus rattus* (black rat) by *Rattus norvegicus* (brown rat). Although both species are highly susceptible to plague (Anderson et al., 2009), the black rat is known to reside in closer proximity to humans and therefore may have been a better intermediate host for disease transmission. Currently the extent to which the black rat had a main role in facilitating its spread within Europe is unclear. Instead, the phylogeographic patterns of *Y. pestis* described in this thesis suggest that, after the Black Death, plague persisted in Europe likely through more than one reservoir and therefore, it is possible that more than one rodent species was involved in its maintenance and transmission. Today, the black rat still plays a significant transmission role in regions where plague outbreaks occur, such as in Madagascar, although sylvatic rodent species are also considered as contributing factors to its persistence (Duplantier et al., 2005). Moreover, other plausible scenarios for plague's disappearance include the improvement of sanitary conditions during the Early Modern Era, the increased quarantine measures during the 18th century, the improved

human nutrition that may have decreased susceptibility to the disease, and the gradual acquisition of immunity by both surviving rats and humans (Appleby, 1980).

It is certain that *Y. pestis* genomics research contributes one important piece of information to a subject that requires a combined approach in order to be comprehended. The integration of additional lines of evidence, such as the revision of historical epidemic records (Buntgen et al., 2012) which may contribute important epidemiological data, the archaeological identification of medieval rodent populations (Antoine, 2008) and their genomic analysis as well as disease modelling efforts that encompass vector dynamics (Keeling and Gilligan, 2000), climatic (Schmid et al., 2015, Dean et al., 2018) and genomic (Bos et al., 2011, Bos et al., 2016, Spyrou et al., 2016) data, will certainly provide a more complete picture and help disentangle the processes behind the initiation, progression and abrupt end of the second plague pandemic.

4.4 Additional notes and open questions about *Y. pestis* research

In this section, I outline a few additional points of discussion that may potentially trigger future research directions.

The bacterium *Y. pestis* diverged from its closest ancestor, *Y. pseudotuberculosis*, sometime earlier than 5,000 years ago, and went on to cause three catastrophic pandemics thereafter. Genomic data from all three pandemics as well as from enzootic foci have helped to provide key insight into the evolutionary history of this bacterium (Figures 1 and 2). However, still little is known about the time and place of its emergence. Collectively, the confidence interval produced for the divergence time of *Y. pestis* and *Y. pseudotuberculosis* by the most recent publications is very wide and is estimated between 13,000 and 79,000 yBP (Rasmussen et al., 2015, Andrades Valtueña et al., 2017). Such analysis is likely challenged by the different evolutionary properties and histories of the two bacteria. While the two species are very closely related, with an identity of up to 97% within chromosomal protein coding regions (Achtman et al., 1999, Chain et al., 2004), they are vastly distinct in terms of their diversification patterns. *Y. pestis* is a highly virulent, clonal and monomorphic pathogen (Achtman, 2012) that displays little genetic diversity. Therefore, its evolution can be unambiguously inferred through whole-genome phylogenetic analysis, although, as previously mentioned, its substitution rate can vary between phases of epidemic

eruption and enzootic persistence (Cui et al., 2013) (paper D). By contrast, *Y. pseudotuberculosis* is a soil dwelling organism that can occasionally cause small scale food-borne outbreaks (Williamson et al., 2017) and whose evolutionary history has been vastly dependent on both recombination and mutation (Laukkanen-Ninios et al., 2011). Therefore, the different genomic properties of these species should be accounted for in future combined data analysis. In addition, no ancient calibration points exist for *Y. pseudotuberculosis*, as no study has to date reported ancient DNA reads for this bacterium. Additional genome sequencing of ancient and modern *Y. pseudotuberculosis* may help such inferences by increasing the available calibration points that can be incorporated into dating analyses. These efforts may aid a more accurate inference of their joint evolutionary history and divergence time.

The second point of discussion concerns plague's recurrent disappearance from Europe. The bacterium is known to have been intermittently present in the continent since the Bronze Age, when it likely endured for 1,500 years, and all the way until the 18th century, which was preceded by 400 years of epidemic persistence. Yet today no plague foci are found in Europe west of the Black Sea. Until recently, plague's disappearance from Europe at the end of the second pandemic (18th century) was considered the only such occasion, and the wish for its explanation has triggered vast amounts of speculation (Appleby, 1980). Instead, it has now become increasingly evident that, over its history, plague lineages became extinct in Europe more than once, in fact at least three times. The first was likely towards the end of the Bronze Age, the second was at the end of the first pandemic and the third was at the end of the second pandemic (Wagner et al., 2014, Rasmussen et al., 2015, Bos et al., 2016). While today there is a considerable diversity of *Y. pestis* hosts and vectors that make up reservoirs around the world (Anisimov et al., 2004), little is known about plague susceptibility in sylvatic rodent populations that exist west of the Black Sea. In addition, previous research has suspected a restriction of arthropod vectors in the region (Cohn, 2008, Hufthammer and Walløe, 2013), although the extent to which *X. cheopis* was able to continuously survive in temperate Europe may stand as one aspect of its seemingly problematic maintenance. Future research encompassing ecological and climatic data may help to better understand the capabilities of vector and host dynamics through time in this region.

My final point of discussion involves the role of syndemics during pandemic events. Syndemics refer to the synergistic contribution of more than one disease factor (i.e additional disease agents, ecological factors or social aspects) to an epidemic/pandemic, which lead to an exacerbated disease prognosis and outcome (Singer et al., 2017). A typical modern-day example of such an interaction is the HIV and *M. tuberculosis* co-infection, which is expected to accelerate AIDS progression and severity, but also increase a patient's risk for developing active tuberculosis (Singer et al., 2017, Bell and Noursadeghi, 2018). Regarding the second plague pandemic, there is increasing evidence from osteological research that indicates a poor underlying health in the medieval population of Europe as well as some degree of selective mortality during the second pandemic (DeWitte, 2016). Particularly, it has been shown that individuals buried in the East Smithfield cemetery in London had an overrepresentation of physiological stress markers on their skeletons, suggesting that those with stress indicators were more likely to die during the Black Death compared with ones without any such depictions (DeWitte and Wood, 2008, DeWitte and Hughes-Morey, 2012). Molecular research such as shotgun genomics can contribute to these investigations. Metagenomic pipelines have shown their potential as untargeted screening approaches for detecting pathogenic organisms in human remains (Vågene et al., 2018) (paper A). As such, the hundreds of already produced datasets may be used in the future to identify patterns of syndemic infections that may have contributed to the increased disease susceptibility and mortality recorded during the plague pandemics.

5 Outlook

In this thesis, I have used both targeted and broad methodological approaches combined with NGS to conduct a detailed investigation of ancient *Y. pestis* genomes that has revealed key aspects of its last 5,000 years of evolution. In addition, the analysis of human mitochondrial and nuclear genomes from epidemic contexts demonstrates patterns of human diversity and migration that may have facilitated the spread of pathogens, such as *Y. pestis*, in the past.

The ability of aDNA to contribute insights in cases where historical or archaeological data are absent or inconclusive has become evident through this thesis. The investigated historic and prehistoric plague epidemics/pandemics are characteristic examples of rapid disease emergence, and the presented genomic results reveal key details about their history and propagation. Some of the identified lineages, such as the one present during the Bronze Age or the one that endured in Europe after the Black Death, are today extinct. Others, such as the one that spread eastward after the Black Death, was clearly more successful as it seems to have been a sources for worldwide plague epidemics during the 19th century and thereafter. As such, future sampling efforts may reveal additional links between the time periods in which *Y. pestis* has been identified, such as those between the Bronze Age and the first plague pandemic as well as between the first and second pandemics.

Overall, the sequencing of bacterial genomes from ancient contexts has just begun and the case of *Y. pestis* is only one example of the plethora of insights that can be retrieved from this type of genomic analysis. Apart from ancient diversity, future research will have to also invest in the comprehensive sampling of modern genomic data. Efforts for centralised repository systems of bacterial genomes have already been initiated (i.e. EnteroBase) (Alikhan et al., 2018), and the available sequencing power has enabled the production of thousands of whole-genome data points for phylogeographic inferences (Domman et al., 2017, Weill et al., 2017). Their incorporation into ancient frameworks as well as into different demographic models will be valuable for retracing past infectious disease phylogeography.

6 References

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A. and Carniel, E. (1999) 'Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis', *Proc Natl Acad Sci U S A,* 96(24), pp. 14043-8.

Achtman, M. (2012) 'Insights from genomic comparisons of genetically monomorphic bacterial pathogens', *Philosophical Transactions of the Royal Society B: Biological Sciences,* 367(1590), pp. 860-867.

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J., Townsend, G., Sołtysiak, A. and Alt, K. W. (2013) 'Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions', *Nature genetics,* 45(4), pp. 450.

Alexander, J. T. (1980) *Bubonic plague in early modern Russia: public health and urban disaster.* Johns Hopkins University Press, Baltimore, Maryland 21218, USA.

Alikhan, N.-F., Zhou, Z., Sergeant, M. J. and Achtman, M. (2018) 'A genomic overview of the population structure of Salmonella', *PLoS genetics,* 14(4), pp. e1007261.

Allentoft, M. E., Sikora, M., Sjogren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlstrom, T., Vinner, L., Malaspinas, A. S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Della Casa, P., Dabrowski, P., Duffy, P. R., Ebel, A. V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolar, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Palfi, G., Pokutta, D., Pospieszny, L., Price, T. D., Saag, L., Sablin, M., Shishlina, N., Smrcka, V., Soenov, V. I., Szeverenyi, V., Toth, G., Trifanova, S. V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Ponten, T., Brunak, S., Nielsen, R., Kristiansen, K. and Willerslev, E. (2015) 'Population genomics of Bronze Age Eurasia', *Nature,* 522(7555), pp. 167-72.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *Journal of molecular biology,* 215(3), pp. 403-410.

Anderson, D. M., Ciletti, N. A., Lee-Lewis, H., Elli, D., Segal, J., DeBord, K. L., Overheim, K. A., Tretiakova, M., Brubaker, R. R. and Schneewind, O. (2009) 'Pneumonic plague pathogenesis and immunity in Brown Norway rats', *The American journal of pathology,* 174(3), pp. 910-921.

Andrades Valtueña, A. A., Mittnik, A., Key, F. M., Haak, W., Allmńe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., Massy, K., Novak, M., Pfrengle, S., Reinhold, S., Šlaus, M., Spyrou, M. A., Szècsènyi-Nagy, A., Tõrv,

M., Hansen, S., Bos, K. I., Stockhammer, P. W., Herbig, A. and Krause, J. (2017) 'The Stone Age Plague and Its Persistence in Eurasia', *Current biology,* 27(23), pp. 3683-3691. e8.

Anisimov, A. P., Lindler, L. E. and Pier, G. B. (2004) 'Intraspecific diversity of Yersinia pestis', *Clin Microbiol Rev,* 17(2), pp. 434-64.

Antoine, D. (2008) '5 The Archaeology of "Plague"', *Medical History,* 52(S27), pp. 101-114.

Appleby, A. B. (1980) 'The disappearance of plague: a continuing puzzle', *Econ Hist Rev,* 33(2), pp. 161-73.

Armelagos, G. J., Barnes, K. C. and Lin, J. (1996) 'Disease in human evolution: the reemergence of infectious disease in the third epidemiological transition'.

Baele, G. and Barbier, F. (1967) '[Fibrinopenia in a case of subacute viral hepatitis]', *Tijdschr Gastroenterol,* 10(2), pp. 175-82.

Barnes, I. and Thomas, M. G. (2006) 'Evaluating bacterial pathogen DNA preservation in museum osteological collections', *Proceedings of the Royal Society of London B: Biological Sciences,* 273(1587), pp. 645-653.

Barrett, R., Kuzawa, C. W., McDade, T. and Armelagos, G. J. (1998) 'Emerging and re-emerging infectious diseases: the third epidemiologic transition', *Annual review of anthropology,* 27(1), pp. 247-271.

Bell, L. C. and Noursadeghi, M. (2018) 'Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection', *Nature Reviews Microbiology,* 16(2), pp. 80.

Benedictow, O. J. (2004) *The Black Death, 1346-1353: The complete history.* Boydell & Brewer.

Benedictow, O. J. (2016) *The Black Death and Later Plague Epidemics in the Scandinavian Countries: Perspectives and Controversies.* Walter de Gruyter GmbH & Co KG.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L. and Bignell, H. R. (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry', *nature,* 456(7218), pp. 53.

Bertherat, E. (2016) 'Plague around the world, 2010–2015', *Wkly Epidemiol Rec,* 91, pp. 89-93.

Biagini, P., Thèves, C., Balaresque, P., Geraut, A., Cannet, C., Keyser, C., Nikolaeva, D., Gerard, P., Duchesne, S. and Orlando, L. (2012) 'Variola virus in a 300-year-old Siberian mummy', *New England Journal of Medicine,* 367(21), pp. 2057-2059.

Biraben, J.-N. (1975) 'Les hommes et la peste en France et dans les pays européens et méditerranéens des origines à 1850', *Population,* 30(30), pp. 1143-1144.

Bland, D. M., Jarrett, C. O., Bosio, C. F. and Hinnebusch, B. J. (2018) 'Infectious blood source alters early foregut infection and regurgitative transmission of Yersinia pestis by rodent fleas', *PLoS pathogens,* 14(1), pp. e1006859.

Bloom, D. E., Black, S. and Rappuoli, R. (2017) 'Emerging infectious diseases: a proactive approach', *Proceedings of the National Academy of Sciences,* 114(16), pp. 4055-4059.

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J., Herring, D. A., Bauer, P., Poinar, H. N. and Krause, J. (2011) 'A draft genome of Yersinia pestis from victims of the Black Death', *Nature,* 478(7370), pp. 506-10.

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Wolfe Steadman, D. L., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J. E., Gagneux, S., Stone, A. C. and Krause, J. (2014) 'Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis', *Nature,* 514(7523), pp. 494-7.

Bos, K. I., Jäger, G., Schuenemann, V. J., Vågene, Å. J., Spyrou, M. A., Herbig, A., Nieselt, K. and Krause, J. (2015) 'Parallel detection of ancient pathogens via array-based DNA capture', *Phil. Trans. R. Soc. B,* 370(1660), pp. 20130375.

Bos, K. I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S. A., Klunk, J., Schuenemann, V. J., Poinar, D., Kuch, M., Golding, G. B., Dutour, O., Keim, P., Wagner, D. M., Holmes, E. C., Krause, J. and Poinar, H. N. (2016) 'Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus', *Elife,* 5, pp. e12994.

Bouwman, A. S., Kennedy, S. L., Müller, R., Stephens, R. H., Holst, M., Caffell, A. C., Roberts, C. A. and Brown, T. A. (2012) 'Genotype of a historic strain of Mycobacterium tuberculosis', *Proceedings of the National Academy of Sciences,* 109(45), pp. 18511-18516.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. and Paabo, S. (2007) 'Patterns of damage in genomic DNA sequences from a Neandertal', *Proc Natl Acad Sci U S A,* 104(37), pp. 14616-21.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M. and Paabo, S. (2010) 'Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA', *Nucleic Acids Res,* 38(6), pp. e87.

Brygoo, E.-R. (1966) 'Epidémiologie de la peste à Madagascar', *Les Archives de l'Institut Pasteur de Madagascar*.

Buntgen, U., Ginzler, C., Esper, J., Tegel, W. and McMichael, A. J. (2012) 'Digitizing historical plague', *Clin Infect Dis,* 55(11), pp. 1586-8.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J. and Paabo, S. (2010) 'Targeted investigation of the Neandertal genome by array-based sequence capture', *Science,* 328(5979), pp. 723-5.

Cano, R. and Borucki, M. (1995) 'Revival and identification of bacterial-spores in 25-million-year-old to 40-million-year-old dominican amber (VOL 268, PG 1060, 1995)', *Science,* 268(5215), pp. 1265-1265.

Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A. and Poinar Jr, G. O. (1993) 'Amplification and sequencing of DNA from a 120–135-million-year-old weevil', *Nature,* 363(6429), pp. 536.

Carmicheal, A. G. (2014) *Plague persistence in western Europe: A hypothesis. Pandemic diesease in the medieval world: rethinking the Black Death*: ARC Medieval Press, p. 157-192.

Chain, P. S., Carniel, E., Larimer, F. W., Lamerdin, J., Stoutland, P. O., Regala, W. M., Georgescu, A. M., Vergez, L. M., Land, M. L., Motin, V. L., Brubaker, R. R., Fowler, J., Hinnebusch, J., Marceau, M., Medigue, C., Simonet, M., Chenal-Francisque, V., Souza, B., Dacheux, D., Elliott, J. M., Derbise, A., Hauser, L. J. and Garcia, E. (2004) 'Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis', *Proc Natl Acad Sci U S A,* 101(38), pp. 13826-31.

Chouikha, I. and Hinnebusch, B. J. (2014) 'Silencing urease: a key evolutionary step that facilitated the adaptation of Yersinia pestis to the flea-borne transmission route', *Proc Natl Acad Sci U S A* , 111(52), pp. 18709-18714.

Cohn, S. K. (2008) *4 Epidemiology of the Black Death and Successive Waves of Plague. Medical History*: Cambridge University Press, p. 74-100.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S. and Thwaites, G. (2013) 'Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans', *Nature genetics,* 45(10), pp. 1176.

Cooper, A. and Poinar, H. N. (2000) 'Ancient DNA: do it right or not at all', *Science,* 289(5482), pp. 1139-1139.

Cruz‑Dávalos, D. I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin‑Orlando, A., Pruvost, M. and Alfarhan, A. H. (2017)

'Experimental conditions improving in‑solution target enrichment for ancient DNA', *Molecular ecology resources,* 17(3), pp. 508-522.

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. and Yang, R. (2013) 'Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis', *Proc Natl Acad Sci U S A,* 110(2), pp. 577-82.

Cunliffe, B. (2008) *Europe between the Oceans 9000 BC–AD 1000. New Haven-London*: Yale University Press.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J. L. and Meyer, M. (2013a) 'Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments', *Proc Natl Acad Sci U S A,* 110(39), pp. 15758-63.

Dabney, J., Meyer, M. and Pääbo, S. (2013b) 'Ancient DNA damage', *Cold Spring Harbor perspectives in biology,* 5(7), pp. a012567.

Davis, S., Begon, M., De Bruyn, L., Ageyev, V. S., Klassovskiy, N. L., Pole, S. B., Viljugrein, H., Stenseth, N. C. and Leirs, H. (2004) 'Predictive thresholds for plague in Kazakhstan', *Science,* 304(5671), pp. 736-738.

Dawkins, R. and Krebs, J. R. (1979) 'Arms races between and within species', *Proc. R. Soc. Lond. B,* 205(1161), pp. 489-511.

Dean, K. R., Krauer, F., Walløe, L., Lingjærde, O. C., Bramanti, B., Stenseth, N. C. and Schmid, B. V. (2018) 'Human ectoparasites and the spread of plague in Europe during the Second Pandemic', *Proceedings of the National Academy of Sciences*, pp. 201715640.

Dennis, D. T., Gage, K. L., Gratz, N. G., Poland, J. D., Tikhomirov, E. and Organization, W. H. (1999) 'Plague manual: epidemiology, distribution, surveillance and control'.

Devault, A. M., Golding, G. B., Waglechner, N., Enk, J. M., Kuch, M., Tien, J. H., Shi, M., Fisman, D. N., Dhody, A. N. and Forrest, S. (2014a) 'Second-pandemic strain of Vibrio cholerae from the Philadelphia cholera outbreak of 1849', *New England Journal of Medicine,* 370(4), pp. 334-340.

Devault, A. M., McLoughlin, K., Jaing, C., Gardner, S., Porter, T. M., Enk, J. M., Thissen, J., Allen, J., Borucki, M. and DeWitte, S. N. (2014b) 'Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array', *Scientific reports,* 4, pp. 4245.

DeWitte, S. N. and Wood, J. W. (2008) 'Selectivity of Black Death mortality with respect to preexisting health', *Proceedings of the National Academy of Sciences,* 105(5), pp. 1436-1441.

DeWitte, S. N. and Hughes-Morey, G. (2012) 'Stature and frailty during the Black Death: the effect of stature on risks of epidemic mortality in London, AD 1348–1350', *Journal of archaeological science,* 39(5), pp. 1412-1419.

DeWitte, S. N. (2016) 'The anthropology of plague: insights from bioarcheological analyses of epidemic cemeteries', *The Medieval Globe,* 1(1), pp. 6.

Domman, D., Quilici, M.-L., Dorman, M. J., Njamkepo, E., Mutreja, A., Mather, A. E., Delgado, G., Morales-Espinosa, R., Grimont, P. A. and Lizárraga-Partida, M. L. (2017) 'Integrated view of Vibrio cholerae in the Americas', *Science,* 358(6364), pp. 789-793.

Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O. and Raoult, D. (1998) 'Detection of 400-year-old Yersinia pestis DNA in human dental pulp: an approach to the diagnosis of ancient septicemia', *Proceedings of the National Academy of Sciences,* 95(21), pp. 12637-12640.

Drancourt, M., Roux, V., La Vu Dang, L. T.-H., Castex, D., Chenal-Francisque, V., Ogata, H., Fournier, P.-E., Crubézy, E. and Raoult, D. (2004) 'Genotyping, Orientalis-like Yersinia pestis, and plague pandemics', *Emerging infectious diseases,* 10(9), pp. 1585.

Drancourt, M., Signoli, M., La Vu Dang, B. B., Roux, V., Tzortzis, S. and Raoult, D. (2007) 'Yersinia pestis Orientalis in remains of ancient plague patients', *Emerging infectious diseases,* 13(2), pp. 332.

Du, Y., Rosqvist, R. and Forsberg, Å. (2002) 'Role of fraction 1 antigen of Yersinia pestis in inhibition of phagocytosis', *Infection and immunity,* 70(3), pp. 1453-1460.

Duggan, A. T., Perdomo, M. F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Emery, M. V., Buchmann, J. P., Duchêne, S., Jankauskas, R. and Humphreys, M. (2016) '17th century variola virus reveals the recent history of smallpox', *Current Biology,* 26(24), pp. 3407-3412.

Duncan, C. J. and Scott, S. (2005) 'What caused the black death?', *Postgraduate medical journal,* 81(955), pp. 315-320.

Duplantier, J. M., Duchemin, J. B., Chanteau, S. and Carniel, E. (2005) 'From the recent lessons of the Malagasy foci towards a global understanding of the factors involved in plague reemergence', *Vet Res,* 36(3), pp. 437-53.

Eglinton, G. and Logan, G. A. (1991) 'Molecular preservation', *Phil. Trans. R. Soc. Lond. B,* 333(1268), pp. 315-328.

Eisen, R. J., Eisen, L. and Gage, K. L. (2009) 'Studies of vector competency and efficiency of North American fleas for Yersinia pestis: state of the field and future research needs', *J Med Entomol,* 46(4), pp. 737-44.

Eisen, R. J., Dennis, D. T. and Gage, K. L. (2015) 'The role of early-phase transmission in the spread of Yersinia pestis', *Journal of medical entomology,* 52(6), pp. 1183-1192.

Erickson, D. L., Waterfield, N. R., Vadyvaloo, V., Long, D., Fischer, E. R., Ffrench‑Constant, R. and Hinnebusch, B. J. (2007) 'Acute oral toxicity of Yersinia pseudotuberculosis to fleas: implications for the evolution of vector‑borne transmission of plague', *Cellular microbiology,* 9(11), pp. 2658-2666.

Eroshenko, G. A., Nosov, N. Y., Krasnov, Y. M., Oglodin, Y. G., Kukleva, L. M., Guseva, N. P., Kuznetsov, A. A., Abdikarimov, S. T., Dzhaparova, A. K. and Kutyrev, V. V. (2017) 'Yersinia pestis strains of ancient phylogenetic branch 0. ANT are widely spread in the high-mountain plague foci of Kyrgyzstan', *PloS one,* 12(10), pp. e0187230.

Feldman, M., Harbeck, M., Keller, M., Spyrou, M. A., Rott, A., Trautmann, B., Scholz, H. C., Paffgen, B., Peters, J., McCormick, M., Bos, K., Herbig, A. and Krause, J. (2016) 'A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim', *Mol Biol Evol,* 33(11), pp. 2911-2923.

Ford, D. C., Joshua, G. W., Wren, B. W. and Oyston, P. C. (2014) 'The importance of the magnesium transporter MgtB for virulence of Yersinia pseudotuberculosis and Yersinia pestis', *Microbiology,* 160(12), pp. 2710-2717.

Fordyce, S. L., Kampmann, M.-L., Van Doorn, N. L. and Gilbert, M. T. P. (2013) 'Long-term RNA persistence in postmortem contexts', *Investigative genetics,* 4(1), pp. 7.

Frantz, L. A., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R. P., Larson, G. and Groenen, M. A. (2015) 'Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes', *Nature genetics,* 47(10), pp. 1141.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J. and Pääbo, S. (2013) 'DNA analysis of an early modern human from Tianyuan Cave, China', *Proceedings of the National Academy of Sciences,* 110(6), pp. 2223-2227.

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prufer, K., Meyer, M., Kelso, J., Reich, D. and Paabo, S. (2015) 'An early modern human from Romania with a recent Neanderthal ancestor', *Nature,* 524(7564), pp. 216-9.

Gage, K. L. and Kosoy, M. Y. (2005) 'Natural history of plague: perspectives from more than a century of research', *Annu Rev Entomol,* 50, pp. 505-28.

Gansauge, M.-T. and Meyer, M. (2013) 'Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA', *Nature protocols,* 8(4), pp. 737.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A. and Meyer, M. (2017) 'Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase', *Nucleic acids research,* 45(10), pp. e79-e79.

Gasc, C., Peyretaillade, E. and Peyret, P. (2016) 'Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms', *Nucleic acids research,* 44(10), pp. 4504-4518.

Gelabert, P., Sandoval-Velasco, M., Olalde, I., Fregel, R., Rieux, A., Escosa, R., Aranda, C., Paaijmans, K., Mueller, I. and Gilbert, M. T. P. (2016) 'Mitochondrial DNA from the eradicated European Plasmodium vivax and P. falciparum from 70-year-old slides from the Ebro Delta in Spain', *Proceedings of the National Academy of Sciences,* 113(41), pp. 11495-11500.

Gilbert, M. T. P., Cuccui, J., White, W., Lynnerup, N., Titball, R. W., Cooper, A. and Prentice, M. B. (2004) 'Absence of Yersinia pestis-specific DNA in human teeth from five European excavations of putative plague victims', *Microbiology,* 150(2), pp. 341-354.

Gottfried, R. S. (1983) *The Black Death; natural and human disaster in Medieval Europe.* New York: Simon & Schuster.

Grabenstein, J. P., Marceau, M., Pujol, C., Simonet, M. and Bliska, J. B. (2004) 'The response regulator PhoP of Yersinia pseudotuberculosis is important for replication in macrophages and for virulence', *Infection and immunity,* 72(9), pp. 4973-4984.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D. and Paabo, S. (2010) 'A draft sequence of the Neandertal genome', *Science,* 328(5979), pp. 710-22.

Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tänzer, M., Villems, R., Renfrew, C., Gronenborn, D. and Alt, K. W. (2005) 'Ancient DNA from the first European farmers in 7500-year-old Neolithic sites', *Science,* 310(5750), pp. 1016-1018.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Banffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Rojo Guerra, M. A., Roth, C., Szecsenyi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W. and Reich, D. (2015) 'Massive migration from the steppe was a source for Indo-European languages in Europe', *Nature,* 522(7555), pp. 207-11.

Haensch, S., Bianucci, R., Signoli, M., Rajerison, M., Schultz, M., Kacki, S., Vermunt, M., Weston, D. A., Hurst, D., Achtman, M., Carniel, E. and Bramanti, B. (2010) 'Distinct clones of Yersinia pestis caused the black death', *PLoS Pathog,* 6(10), pp. e1001134.

Harbeck, M., Seifert, L., Hänsch, S., Wagner, D. M., Birdsell, D., Parise, K. L., Wiechmann, I., Grupe, G., Thomas, A. and Keim, P. (2013) 'Yersinia pestis DNA from skeletal remains from the 6th century AD reveals insights into Justinianic Plague', *PLoS Pathogens,* 9(5), pp. e1003349.

Harkins, K. M., Buikstra, J. E., Campbell, T., Bos, K. I., Johnson, E. D., Krause, J. and Stone, A. C. (2015) 'Screening ancient tuberculosis with qPCR: challenges and opportunities', *Phil. Trans. R. Soc. B,* 370(1660), pp. 20130622.

Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C. and Collins, M. J. (2018) 'A guide to ancient protein studies', *Nature ecology & evolution*, pp. 1.

Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., Galili, E., Eshed, V., Greenblatt, C. L. and Lemma, E. (2008) 'Detection and molecular characterization of 9000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean', *PloS one,* 3(10), pp. e3426.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. and Wilson, A. C. (1984) 'DNA sequences from the quagga, an extinct member of the horse family', *Nature,* 312(5991), pp. 282.

Hinnebusch, B. J., Rudolph, A. E., Cherepanov, P., Dixon, J. E., Schwan, T. G. and Forsberg, Å. (2002) 'Role of Yersinia murine toxin in survival of Yersinia pestis in the midgut of the flea vector', *Science,* 296(5568), pp. 733-735.

Hinnebusch, B. J., Chouikha, I. and Sun, Y. C. (2016) 'Ecological Opportunity, Evolution, and the Emergence of Flea-Borne Plague', *Infect Immun,* 84(7), pp. 1932-40.

Hinnebusch, B. J., Bland, D. M., Bosio, C. F. and Jarrett, C. O. (2017a) 'Comparative Ability of Oropsylla montana and Xenopsylla cheopis Fleas to Transmit Yersinia pestis by Two Different Mechanisms', *PLoS Negl Trop Dis,* 11(1), pp. e0005276.

Hinnebusch, B. J., Jarrett, C. O. and Bland, D. M. (2017b) '"Fleaing" the Plague: Adaptations of Yersinia pestis to Its Insect Vector That Lead to Transmission', *Annual review of microbiology,* 71, pp. 215-232.

Hinnebusch, J., Cherepanov, P., Du, Y., Rudolph, A., Dixon, J., Schwan, T. and Forsberg, Å. (2000) 'Murine toxin of Yersinia pestis shows phospholipase D activity but is not required for virulence in mice', *International journal of medical microbiology,* 290(4-5), pp. 483-487.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W. and Hannon, G. J. (2009) 'Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing', *Nat Protoc,* 4(6), pp. 960-74.

Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. v. and Pääbo, S. (2001) 'DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA', *Nucleic acids research,* 29(23), pp. 4793-4799.

Hufthammer, A. K. and Walløe, L. (2013) 'Rats cannot have been intermediate hosts for Yersinia pestis during medieval plague epidemics in Northern Europe', *Journal of Archaeological Science,* 40(4), pp. 1752-1759.

Johnson, N. P. and Mueller, J. (2002) 'Updating the accounts: global mortality of the 1918-1920" Spanish" influenza pandemic', *Bulletin of the History of Medicine,* 76(1), pp. 105-115.

Kahila Bar‐Gal, G., Kim, M. J., Klein, A., Shin, D. H., Oh, C. S., Kim, J. W., Kim, T. H., Kim, S. B., Grant, P. R. and Pappo, O. (2012) 'Tracing hepatitis B virus to the 16th century in a Korean mummy', *Hepatology,* 56(5), pp. 1671-1680.

Karlsson, E. K., Kwiatkowski, D. P. and Sabeti, P. C. (2014) 'Natural selection and infectious disease in human populations', *Nat Rev Genet,* 15(6), pp. 379-93.

Kay, G. L., Sergeant, M. J., Giuffra, V., Bandiera, P., Milanese, M., Bramanti, B., Bianucci, R. and Pallen, M. J. (2014) 'Recovery of a medieval Brucella melitensis genome using shotgun metagenomics', *MBio,* 5(4), pp. e01337-14.

Kay, G. L., Sergeant, M. J., Zhou, Z., Chan, J. Z.-M., Millard, A., Quick, J., Szikossy, I., Pap, I., Spigelman, M. and Loman, N. J. (2015) 'Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe', *Nature communications,* 6, pp. 6717.

Keeling, M. J. and Gilligan, C. A. (2000) 'Bubonic plague: a metapopulation model of a zoonosis', *Proceedings of the Royal Society B: Biological Sciences,* 267(1458), pp. 2219-30.

Key, F. M., Posth, C., Krause, J., Herbig, A. and Bos, K. I. (2017) 'Mining metagenomic data sets for ancient DNA: recommended protocols for authentication', *Trends in Genetics,* 33(8), pp. 508-520.

Kircher, M., Sawyer, S. and Meyer, M. (2012) 'Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform', *Nucleic Acids Res,* 40(1), pp. e3.

Kislichkina, A. A., Bogun, A. G., Kadnikova, L. A., Maiskaya, N. V., Platonov, M. E., Anisimov, N. V., Galkina, E. V., Dentovskaya, S. V. and Anisimov, A. P. (2015) 'Nineteen Whole-Genome Assemblies of Yersinia pestis subsp. microtus, Including Representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica', *Genome Announc,* 3(6), pp. e01342-15.

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. and Mardis, E. R. (2013) 'The next-generation sequencing revolution and its impact on genomics', *Cell,* 155(1), pp. 27-38.

Korneliussen, T. S., Albrechtsen, A. and Nielsen, R. (2014) 'ANGSD: analysis of next generation sequencing data', *BMC bioinformatics,* 15(1), pp. 356.

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A. and Pääbo, S. (2010a) 'A complete mtDNA genome of an early modern human from Kostenki, Russia', *Current Biology,* 20(3), pp. 231-236.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P. and Pääbo, S. (2010b) 'The complete mitochondrial DNA genome of an unknown hominin from southern Siberia', *Nature,* 464(7290), pp. 894.

Krause-Kyora, B., Susat, J., Key, F. M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D. and Franzenburg, S. (2018) 'Neolithic and Medieval virus genomes reveal complex evolution of Hepatitis B', *eLife,* 7, pp. e36666.

Kulikowski, M. (2007) 'Plague in Spanish late antiquity', *Plague and the End of Antiquity,* pp. 150.

Laukkanen‐Ninios, R., Didelot, X., Jolley, K. A., Morelli, G., Sangal, V., Kristo, P., Brehony, C., Imori, P. F., Fukushima, H. and Siitonen, A. (2011) 'Population structure of the Yersinia pseudotuberculosis complex according to multilocus sequence typing', *Environmental microbiology,* 13(12), pp. 3114-3127.

Lazaridis, I. and Patterson, N. and Mittnik, A. and Renaud, G. and Mallick, S. and Kirsanow, K. and Sudmant, P. H. and Schraiber, J. G. and Castellano, S. and Lipson, M. and Berger, B. and Economou, C. and Bollongino, R. and Fu, Q. and Bos, K. I. and Nordenfelt, S. and Li, H. and de Filippo, C. and Prufer, K. and Sawyer, S. and Posth, C. and Haak, W. and Hallgren, F. and Fornander, E. and Rohland, N. and Delsate, D. and Francken, M. and Guinet, J. M. and Wahl, J. and Ayodo, G. and Babiker, H. A. and Bailliet, G. and Balanovska, E. and Balanovsky, O. and Barrantes, R. and Bedoya, G. and Ben-Ami, H. and Bene, J. and Berrada, F. and Bravi, C. M. and Brisighelli, F. and Busby, G. B. and Cali, F. and Churnosov, M. and Cole, D. E. and Corach, D. and Damba, L. and van Driem, G. and Dryomov, S. and Dugoujon, J. M. and Fedorova, S. A. and Gallego Romero, I. and Gubina, M. and Hammer, M. and Henn, B. M. and

Hervig, T. and Hodoglugil, U. and Jha, A. R. and Karachanak-Yankova, S. and Khusainova, R. and Khusnutdinova, E. and Kittles, R. and Kivisild, T. and Klitz, W. and Kucinskas, V. and Kushniarevich, A. and Laredj, L. and Litvinov, S. and Loukidis, T. and Mahley, R. W. and Melegh, B. and Metspalu, E. and Molina, J. and Mountain, J. and Nakkalajarvi, K. and Nesheva, D. and Nyambo, T. and Osipova, L. and Parik, J. and Platonov, F. and Posukh, O. and Romano, V. and Rothhammer, F. and Rudan, I. and Ruizbakiev, R. and Sahakyan, H. and Sajantila, A. and Salas, A. and Starikovskaya, E. B. and Tarekegn, A. and Toncheva, D. and Turdikulova, S. and Uktveryte, I. and Utevska, O. and Vasquez, R. and Villena, M. and Voevoda, M. and Winkler, C. A. and Yepiskoposyan, L. and Zalloua, P. and Zemunik, T. and Cooper, A. and Capelli, C. and Thomas, M. G. and Ruiz-Linares, A. and Tishkoff, S. A. and Singh, L. and Thangaraj, K. and Villems, R. and Comas, D. and Sukernik, R. and Metspalu, M. and Meyer, M. and Eichler, E. E. and Burger, J. and Slatkin, M. and Paabo, S. and Kelso, J. and Reich, D. and Krause, J. (2014) 'Ancient human genomes suggest three ancestral populations for present-day Europeans', *Nature,* 513(7518), pp. 409-13.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J. M., Gregg, M., Eshed, V., Mizrahi, A.-S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Blüher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, M., Gilbert, E., Kerr, S. M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D. A., O'Reilly, S., Richard, M. B., Semino, O., Shamoon-Pour, M., Stefanescu, G., Stumvoll, M., Tönges, A., Torroni, A., Wilson, J. F., Yengo, L., Hovhannisyan, N. A., Patterson, N., Pinhasi, R. and Reich, D. (2016) 'Genomic insights into the origin of farming in the ancient Near East', *Nature,* 536(7617), pp. 419-424.

Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., Fages, A., Khan, N., Schubert, M. and Jagannathan, V. (2017) 'Ancient genomic changes associated with domestication of the horse', *Science,* 356(6336), pp. 442-445.

Lindahl, T. (1993) 'Instability and decay of the primary structure of DNA', *nature,* 362(6422), pp. 709.

Little, L. K. (2007) *Plague and the end of antiquity: the pandemic of 541-750.* Cambridge University Press.

Louvel, G., Der Sarkissian, C., Hanghøj, K. and Orlando, L. (2016) 'metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high‐throughput sequencing shotgun data', *Molecular ecology resources,* 16(6), pp. 1415-1427.

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M. R., Hallows, J. L., Kusebauch, U., Vigl, E. E., Malfertheiner, P. and Megraud, F. (2016) 'The

5300-year-old Helicobacter pylori genome of the Iceman', *Science,* 351(6269), pp. 162-165.

Marciniak, S., Prowse, T. L., Herring, D. A., Klunk, J., Kuch, M., Duggan, A. T., Bondioli, L., Holmes, E. C. and Poinar, H. N. (2016) 'Plasmodium falciparum malaria in 1st–2nd century CE southern Italy', *Current Biology,* 26(23), pp. R1220-R1222.

Mardis, E. R. (2008) 'Next-generation DNA sequencing methods', *Annu. Rev. Genomics Hum. Genet.,* 9, pp. 387-402.

Margaryan, A., Hansen, H. B., Rasmussen, S., Sikora, M., Moiseyev, V., Khoklov, A., Epimakhov, A., Yepiskoposyan, L., Kriiska, A. and Varul, L. (2018) 'Ancient pathogen DNA in human teeth and petrous bones', *Ecology and Evolution*.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J. and Chen, Z. (2005) 'Genome sequencing in microfabricated high-density picolitre reactors', *Nature,* 437(7057), pp. 376.

Maricic, T., Whitten, M. and Paabo, S. (2010) 'Multiplexed DNA sequence capture of mitochondrial genomes using PCR products', *PLoS One,* 5(11), pp. e14004.

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S. Y. and Dietrich, F. S. (2013) 'Reconstructing genome evolution in historic samples of the Irish potato famine pathogen', *Nature communications,* 4, pp. 2172.

Mascher, M., Schuenemann, V. J., Davidovich, U., Marom, N., Himmelbach, A., Hübner, S., Korol, A., David, M., Reiter, E. and Riehl, S. (2016) 'Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley', *Nature genetics,* 48(9), pp. 1089.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A., Roodenberg, J., Verges, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R. and Reich, D. (2015) 'Genome-wide patterns of selection in 230 ancient Eurasians', *Nature,* 528(7583), pp. 499-503.

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

McNeill, W. H. (1998) 'Plagues and peoples. 1976', *Garden City, NY: Anchor P*.

Mendum, T. A., Schuenemann, V. J., Roffey, S., Taylor, G. M., Wu, H., Singh, P., Tucker, K., Hinds, J., Cole, S. T. and Kierzek, A. M. (2014) 'Mycobacterium

leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic', *BMC genomics,* 15(1), pp. 270.

Meyer, M. and Kircher, M. (2010) 'Illumina sequencing library preparation for highly multiplexed target capture and sequencing', *Cold Spring Harb Protoc,* 2010(6), pp. pdb prot5448.

Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V. and Tõrv, M. (2018) 'The genetic prehistory of the Baltic Sea region', *Nature communications,* 9(1), pp. 442.

Morelli, G., Song, Y., Mazzoni, C. J., Eppinger, M., Roumagnac, P., Wagner, D. M., Feldkamp, M., Kusecek, B., Vogler, A. J., Li, Y., Cui, Y., Thomson, N. R., Jombart, T., Leblois, R., Lichtner, P., Rahalison, L., Petersen, J. M., Balloux, F., Keim, P., Wirth, T., Ravel, J., Yang, R., Carniel, E. and Achtman, M. (2010) 'Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity', *Nat Genet,* 42(12), pp. 1140-3.

Morony, M. G. and Little, L. K. (2007) ''For Whom Does the Writer Write?'The First Bubonic Plague Pandemic According to Syriac Sources', *Plague and the End of Antiquity*, pp. 59.

Mühlemann, B., Jones, T. C., de Barros Damgaard, P., Allentoft, M. E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I. P., Boldgiv, B. and Bazartseren, T. (2018) 'Ancient hepatitis B viruses from the Bronze Age to the Medieval period', *Nature*, pp. 1.

Müller, R., Roberts, C. A. and Brown, T. A. (2016) 'Complications in the study of ancient tuberculosis: Presence of environmental bacteria in human archaeological remains', *Journal of Archaeological Science,* 68, pp. 5-11.

Norris, J. (1976) 'East or west? The geographic origin of the Black Death', *Bulletin of the History of Medicine,* 51(1), pp. 1-24.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B. and Ako-Adjei, D. (2015) 'Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation', *Nucleic acids research,* 44(D1), pp. D733-D745.

Ortner, D. J. (2003) *Identification of pathological conditions in human skeletal remains.* Academic Press.

Ortner, D. J. (2008) 'Differential diagnosis of skeletal lesions in infectious disease', *Advances in human palaeopathology*, pp. 189-214.

Pääbo, S. (1985) 'Molecular cloning of ancient Egyptian mummy DNA', *nature,* 314(6012), pp. 644-645.

Pääbo, S. and Wilson, A. C. (1988) 'Polymerase chain reaction reveals cloning artefacts', *Nature,* 334(6181), pp. 387.

Pääbo, S. (1989) 'Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification', *Proceedings of the National Academy of Sciences,* 86(6), pp. 1939-1943.

Papagrigorakis, M. J., Yapijakis, C., Synodinos, P. N. and Baziotopoulou-Valavani, E. (2006) 'DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens', *International Journal of Infectious Diseases,* 10(3), pp. 206-214.

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., Baker, S., Basham, D., Bentley, S. D., Brooks, K., Cerdeno-Tarraga, A. M., Chillingworth, T., Cronin, A., Davies, R. M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A. V., Leather, S., Moule, S., Oyston, P. C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B. G. (2001) 'Genome sequence of Yersinia pestis, the causative agent of plague', *Nature,* 413(6855), pp. 523-7.

Perry, R. D. and Fetherston, J. D. (1997) 'Yersinia pestis--etiologic agent of plague', *Clin Microbiol Rev,* 10(1), pp. 35-66.

Poinar, H. N., Cano, R. J. and Poinar Jr, G. O. (1993) 'DNA from an extinct plant', *Nature,* 363(6431), pp. 677.

Pollitzer, R. (1954) 'The Plague. Geneva', *World Health Organization*, pp. 26.

Pollitzer, R. (1960) 'A review of recent literature on plague', *Bull World Health Organ,* 23, pp. 313-400.

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford Jr, T. W., Orlando, L. and Metspalu, E. (2014) 'Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans', *Nature,* 505(7481), pp. 87-91.

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C. and Malaspinas, A.-S. (2015) 'Genomic evidence for the Pleistocene and recent population history of Native Americans', *Science,* 349(6250), pp. aab3884.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T. and Gupta, R. (2010) 'Ancient human genome sequence of an extinct Palaeo-Eskimo', *Nature,* 463(7282), pp. 757.

Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford Jr, T. W., Rasmussen, S., Moltke, I., Albrechtsen, A. and Doyle, S. M. (2014) 'The genome of a Late Pleistocene human from a Clovis burial site in western Montana', *Nature,* 506(7487), pp. 225.

Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjogren, K. G., Pedersen, A. G., Schubert, M., Van Dam, A., Kapel, C. M., Nielsen, H. B.,

Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M. V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Ponten, T., Foley, R. A., Lahr, M. M., Nielsen, R., Kristiansen, K. and Willerslev, E. (2015) 'Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago', *Cell,* 163(3), pp. 571-82.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M. and Paabo, S. (2010) 'Genetic history of an archaic hominin group from Denisova Cave in Siberia', *Nature,* 468(7327), pp. 1053-60.

Renaud, G., Slon, V., Duggan, A. T. and Kelso, J. (2015) 'Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA', *Genome Biol,* 16, pp. 224.

Roberts, C. A. and Buikstra, J. E. (2003) *The bioarchaeology of tuberculosis: a global perspective on a re-emerging disease.* University Press of Florida.

Rohland, N. and Hofreiter, M. (2007) 'Ancient DNA extraction from bones and teeth', *Nature protocols,* 2(7), pp. 1756.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. and Reich, D. (2015) 'Partial uracil-DNA-glycosylase treatment for screening of ancient DNA', *Philos Trans R Soc Lond B Biol Sci,* 370(1660), pp. 20130624.

Ross, Z. P., Klunk, J., Fornaciari, G., Giuffra, V., Duchêne, S., Duggan, A. T., Poinar, D., Douglas, M. W., Eden, J.-S. and Holmes, E. C. (2018) 'The paradox of HBV evolution as revealed from a 16th century mummy', *PLoS pathogens,* 14(1), pp. e1006750.

Sabeti, P. C., Walsh, E., Schaffner, S. F., Varilly, P., Fry, B., Hutcheson, H. B., Cullen, M., Mikkelsen, T. S., Roy, J. and Patterson, N. (2005) 'The case for selection at CCR5-Δ32', *PLoS biology,* 3(11), pp. e378.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. (1985) 'Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia', *Science,* 230(4732), pp. 1350-1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) 'Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase', *Science,* 239(4839), pp. 487-491.

Sallares, R., Bouwman, A. and Anderung, C. (2004) 'The spread of malaria to Southern Europe in antiquity: new approaches to old problems', *Medical history,* 48(3), pp. 311-328.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. and Paabo, S. (2012) 'Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA', *PLoS One,* 7(3), pp. e34131.

Schmid, B. V., Buntgen, U., Easterday, W. R., Ginzler, C., Walloe, L., Bramanti, B. and Stenseth, N. C. (2015) 'Climate-driven introduction of the Black Death and successive plague reintroductions into Europe', *Proc Natl Acad Sci U S A,* 112(10), pp. 3020-5.

Schuenemann, V. J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B. K., Wood, J. W., Earn, D. J., White, W., Krause, J. and Poinar, H. N. (2011) '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(38), pp. E746-52.

Schuenemann, V. J., Singh, P., Mendum, T. A., Krause-Kyora, B., Jager, G., Bos, K. I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J. L., Kjellstrom, A., Wu, H., Stewart, G. R., Taylor, G. M., Bauer, P., Lee, O. Y., Wu, H. H., Minnikin, D. E., Besra, G. S., Tucker, K., Roffey, S., Sow, S. O., Cole, S. T., Nieselt, K. and Krause, J. (2013) 'Genome-wide comparison of medieval and modern Mycobacterium leprae', *Science,* 341(6142), pp. 179-83.

Schuenemann, V. J., Peltzer, A., Welte, B., van Pelt, W. P., Molak, M., Wang, C.-C., Furtwängler, A., Urban, C., Reiter, E. and Nieselt, K. (2017) 'Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods', *Nature communications,* 8, pp. 15694.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O. and Huttenhower, C. (2012) 'Metagenomic microbial community profiling using unique clade-specific marker genes', *Nature methods,* 9(8), pp. 811.

Seifert, L., Wiechmann, I., Harbeck, M., Thomas, A., Grupe, G., Projahn, M., Scholz, H. C. and Riehm, J. M. (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(1), pp. e0145194.

Shapiro, B., Rambaut, A. and Gilbert, M. T. P. (2006) 'No proof that typhoid caused the Plague of Athens (a reply to Papagrigorakis et al.)', *International journal of infectious diseases,* 10(4), pp. 334-335.

Signoli, M., Bello, S. and Dutour, O. (1998) '[Epidemic recrudescence of the Great Plague in Marseille (May-July 1722): excavation of a mass grave]', *Med Trop (Mars),* 58(2 Suppl), pp. 7-13.

Singer, M., Bulled, N., Ostrach, B. and Mendenhall, E. (2017) 'Syndemics and the biosocial conception of health', *The Lancet,* 389(10072), pp. 941-950.

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T. P., Götherström, A. and Jakobsson, M. (2012) 'Origins and genetic legacy

of Neolithic farmers and hunter-gatherers in Europe', *Science,* 336(6080), pp. 466-469.

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N. and Reich, D. (2015) 'Genetic evidence for two founding populations of the Americas', *Nature,* 525(7567), pp. 104.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V. and Miller, R. (2017) 'Neandertal and Denisovan DNA from Pleistocene sediments', *Science,* 356(6338), pp. 605-608.

Spyrou, M. A., Tukhbatova, R. I., Feldman, M., Drath, J., Kacki, S., Beltran de Heredia, J., Arnold, S., Sitdikov, A. G., Castex, D., Wahl, J., Gazimzyanov, I. R., Nurgaliev, D. K., Herbig, A., Bos, K. I. and 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(6), pp. 874-81.

Spyrou, M. A., Tukhbatova, R. I., Wang, C.-C., Valtueña, A. A., Lankapalli, A. K., Kondrashin, V. V., Tsybin, V. A., Khokhlov, A., Kühnert, D. and Herbig, A. (2018) 'Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague', *Nature Communications,* 9(1), pp. 2234.

Stamatakis, A. (2014) 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics,* 30(9), pp. 1312-3.

Stenseth, N. C., Samia, N. I., Viljugrein, H., Kausrud, K. L., Begon, M., Davis, S., Leirs, H., Dubyanskiy, V., Esper, J. and Ageyev, V. S. (2006) 'Plague dynamics are driven by climate variation', *Proceedings of the National Academy of Sciences,* 103(35), pp. 13110-13115.

Stenseth, N. C., Atshabar, B. B., Begon, M., Belmain, S. R., Bertherat, E., Carniel, E., Gage, K. L., Leirs, H. and Rahalison, L. (2008) 'Plague: past, present, and future', *PLoS medicine,* 5(1), pp. e3.

Stoclet, A. J. (2007) 'Consilia humana, ops divina, superstitio Seeking Succor and Solace in Times of Plague, with Particular Reference to Gaul in the Early Middle Ages', *Plague and the End of Antiquity: The Pandemic of 541,* 750, pp. 135-49.

Sun, Y.-C., Hinnebusch, B. J. and Darby, C. (2008) 'Experimental evidence for negative selection in the evolution of a Yersinia pestis pseudogene', *Proceedings of the National Academy of Sciences,* 105(23), pp. 8097-8101.

Sun, Y.-C., Koumoutsi, A., Jarrett, C., Lawrence, K., Gherardini, F. C., Darby, C. and Hinnebusch, B. J. (2011) 'Differential control of Yersinia pestis biofilm formation in vitro and in the flea vector by two c-di-GMP diguanylate cyclases', *PLoS One,* 6(4), pp. e19267.

Sun, Y. C., Jarrett, C. O., Bosio, C. F. and Hinnebusch, B. J. (2014) 'Retracing the evolutionary path that led to flea-borne transmission of Yersinia pestis', *Cell Host Microbe,* 15(5), pp. 578-86.

Sussman, G. D. (2011) 'Was the black death in India and China?', *Bull Hist Med,* 85(3), pp. 319-55.

Swarts, K., Gutaker, R. M., Benz, B., Blake, M., Bukowski, R., Holland, J., Kruse-Peeples, M., Lepak, N., Prim, L. and Romay, M. C. (2017) 'Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America', *Science,* 357(6350), pp. 512-515.

Taubenberger, J. K., Reid, A. H., Lourens, R. M., Wang, R., Jin, G. and Fanning, T. G. (2005) 'Characterization of the 1918 influenza virus polymerase genes', *Nature,* 437(7060), pp. 889.

Tikhomirov, E. (1999) 'Epidemiology and distribution of plague', *World Health Organisation*.

Treille, G.-F. and Yersin, A. 'La peste bubonique à Hong Kong'. *VIIIe Congrès international d'hygiène et de démographie*: Pesti konyvnyomda-részvénytársaság, 310-311.

Tumpey, T. M., Basler, C. F., Aguilar, P. V., Zeng, H., Solórzano, A., Swayne, D. E., Cox, N. J., Katz, J. M., Taubenberger, J. K. and Palese, P. (2005) 'Characterization of the reconstructed 1918 Spanish influenza pandemic virus', *science,* 310(5745), pp. 77-80.

Vågene, A. J., Herbig, A., Campana, M. G., Robles Garcia, N. M., Warinner, C., Sabin, S., Spyrou, M. A., Andrades Valtuena, A., Huson, D., Tuross, N., Bos, K. I. and Krause, J. (2018) 'Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico', *Nat Ecol Evol,* 2(3), pp. 520-528.

Vetter, S. M., Eisen, R. J., Schotthoefer, A. M., Montenieri, J. A., Holmes, J. L., Bobrov, A. G., Bearden, S. W., Perry, R. D. and Gage, K. L. (2010) 'Biofilm formation is not required for early-phase transmission of Yersinia pestis', *Microbiology,* 156(7), pp. 2216-2225.

Vogler, A. J., Chan, F., Wagner, D. M., Roumagnac, P., Lee, J., Nera, R., Eppinger, M., Ravel, J., Rahalison, L., Rasoamanana, B. W., Beckstrom-Sternberg, S. M., Achtman, M., Chanteau, S. and Keim, P. (2011) 'Phylogeography and molecular epidemiology of Yersinia pestis in Madagascar', *PLoS Negl Trop Dis,* 5(9), pp. e1319.

Vogler, A. J., Chan, F., Nottingham, R., Andersen, G., Drees, K., Beckstrom-Sternberg, S. M., Wagner, D. M., Chanteau, S. and Keim, P. (2013) 'A decade of plague in Mahajanga, Madagascar: insights into the global maritime spread of pandemic plague', *MBio,* 4(1), pp. e00623-12.

Vogler, A. J., Andrianaivoarimanana, V., Telfer, S., Hall, C. M., Sahl, J. W., Hepp, C. M., Centner, H., Andersen, G., Birdsell, D. N. and Rahalison, L. (2017) 'Temporal phylogeography of Yersinia pestis in Madagascar: Insights into the long-term maintenance of plague', *PLoS neglected tropical diseases,* 11(9), pp. e0005887.

Wagner, D. M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J. W., Enk, J., Birdsell, D. N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M., Golding, B., Riehm, J. M., Earn, D. J. D., DeWitte, S., Rouillard, J.-M., Grupe, G., Wiechmann, I., Bliska, J. B., Keim, P. S., Scholz, H. C., Holmes, E. C. and Poinar, H. (2014) 'Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis', *Lancet Infectious Diseases,* 14(4), pp. 319-326.

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y. and Fiddyment, S. (2014) 'Pathogens and host immunity in the ancient human oral cavity', *Nature genetics,* 46(4), pp. 336.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., Orlando, L. and Krause, J. (2017) 'A robust framework for microbial archaeology', *Annual review of genomics and human genetics,* 18, pp. 321-356.

Weill, F.-X., Domman, D., Njamkepo, E., Tarr, C., Rauzier, J., Fawal, N., Keddy, K. H., Salje, H., Moore, S. and Mukhopadhyay, A. K. (2017) 'Genomic history of the seventh pandemic of cholera in Africa', *Science,* 358(6364), pp. 785-789.

Weyand, N. and Bunnell, M. (1994) 'DNA sequence from Cretaceous period bone fragments', *Science,* 266(5188), pp. 1229-1232.

Wheelis, M. (2002) 'Biological warfare at the 1346 siege of Caffa', *Emerging infectious diseases,* 8(9), pp. 971.

WHO (2016a) 'Global A.I.D.S. Update 2016', *Geneva: UNAIDS.*

WHO (2016b) 'Global tuberculosis report 2016'.

Wiechmann, I. and Grupe, G. (2005) 'Detection of Yersinia pestis DNA in two early medieval skeletal finds from Aschheim (Upper Bavaria, 6th century AD)', *American journal of physical anthropology,* 126(1), pp. 48-55.

Wilbur, A. K., Bouwman, A. S., Stone, A. C., Roberts, C. A., Pfister, L.-A., Buikstra, J. E. and Brown, T. A. (2009) 'Deficiencies and challenges in the study of ancient tuberculosis DNA', *Journal of Archaeological Science,* 36(9), pp. 1990-1997.

Williamson, D. A., Baines, S. L., Carter, G. P., da Silva, A. G., Ren, X., Sherwood, J., Dufour, M., Schultz, M. B., French, N. P. and Seemann, T. (2017) 'Genomic insights into a sustained national outbreak of Yersinia pseudotuberculosis', *Genome biology and evolution,* 8(12), pp. 3806-3814.

Wood, J. W., Milner, G. R., Harpending, H. C., Weiss, K. M., Cohen, M. N., Eisenberg, L. E., Hutchinson, D. L., Jankauskas, R., Cesnys, G. and Česnys, G. (1992) 'The

osteological paradox: problems of inferring prehistoric health from skeletal samples [and comments and reply]', *Current anthropology,* 33(4), pp. 343-370.

Worobey, M., Watts, T. D., McKay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., Koblin, B. A., Heneine, W., Lemey, P. and Jaffe, H. W. (2016) '1970s and 'Patient 0'HIV-1 genomes illuminate early HIV/AIDS history in North America', *Nature,* 539(7627), pp. 98.

Wren, B. W. (2003) 'The yersiniae: a model genus to study the rapid evolution of bacterial pathogens', *Nat Rev Microbiol,* 1(1), pp. 55-64.

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S. and Krause, J. (2013) 'The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine', *Elife,* 2, pp. e00731.

Zhgenti, E., Johnson, S. L., Davenport, K. W., Chanturia, G., Daligault, H. E., Chain, P. S. and Nikolich, M. P. (2015) 'Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus Region', *Genome Announc,* 3(5), pp. e01030-15.

Zimbler, D. L., Schroeder, J. A., Eddy, J. L. and Lathem, W. W. (2015) 'Early emergence of Yersinia pestis as a severe respiratory pathogen', *Nature Communications,* 6, pp. 7487.

7 Figures

Figure 1- Map of ancient and modern *Y. pestis* strains used in this thesis (Appears on page 27)

Figure 2- Maximum likelihood phylogenetic tree of ancient and modern *Y. pestis* (Appears on page 32)

8 Appendix

A. **Maria A. Spyrou**, Rezeda I. Tukhbatova, Chuan-Chao Wang, Aida Andrades Valtueña, Aditya K. Lankapalli, Vitaly V. Kondrashin, Victor A. Tsibin, Aleksandr Khokhlov, Denise Kühnert, Alexander Herbig, Kirsten I. Bos and Johannes Krause. "Analysis of 3,800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague" Published in *Nature Communications*, 2018, 9(1): 2234

B. **Maria A. Spyrou**, Alessandra Sperduti, Åshild J. Vågene, Lorenzo M. Bondioli, Henrike Heyne, Eva Fernández-Domínguez, Luca Bondioli, Wolfgang Haak, Kirsten I. Bos and Johannes Krause "Ancient DNA recovery and maternal lineage diversity of early medieval Venosa in southern Italy" Manuscript ready for submission

C. **Maria A. Spyrou**, Rezeda I. Tukhbatova, Michal Feldman, Joanna Drath, Sacha Kacki, Julia Beltrán de Heredia, Susanne Arnold, Airat G. Sitdikov, Dominique Castex, Joachim Wahl, Ilgizar R. Gazimzyanov, Danis K. Nurgaliev, Alexander Herbig, Kirsten I. Bos and Johannes Krause "Historical *Y. pestis* genomes reveal the European Black Death as the source of ancient and modern plague pandemics" Published in *Cell Host & Microbe*, 2016, 19(6), pp. 874-81

D. **Maria A. Spyrou***, Marcel Keller*, Rezeda Tukhbatova, Elisabeth Nelson, Don Walker, Amelie Alterauge, Hermann Fetz, Joris Peters, Niamh Carty, Robert Hartle, Michael Henderson, Elizabeth L. Knox, Sacha Kacki, Michaël Gourvennec, Dominique Castex, Sandra Lösch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos and Johannes Krause "A phylogeography of the second plague pandemic revealed through historical *Y. pestis* genomes" Manuscript ready for submission

*denotes equal contribution

Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague

Maria A. Spyrou [1,2], Rezeda I. Tukhbatova[1,3], Chuan-Chao Wang [1,4], Aida Andrades Valtueña[1], Aditya K. Lankapalli[1], Vitaly V. Kondrashin[5], Victor A. Tsybin[6], Aleksandr Khokhlov[7], Denise Kühnert[1,8], Alexander Herbig [1], Kirsten I. Bos[1] & Johannes Krause [1,2]

The origin of *Yersinia pestis* and the early stages of its evolution are fundamental subjects of investigation given its high virulence and mortality that resulted from past pandemics. Although the earliest evidence of *Y. pestis* infections in humans has been identified in Late Neolithic/Bronze Age Eurasia (LNBA 5000–3500y BP), these strains lack key genetic components required for flea adaptation, thus making their mode of transmission and disease presentation in humans unclear. Here, we reconstruct ancient *Y. pestis* genomes from individuals associated with the Late Bronze Age period (~3800 BP) in the Samara region of modern-day Russia. We show clear distinctions between our new strains and the LNBA lineage, and suggest that the full ability for flea-mediated transmission causing bubonic plague evolved more than 1000 years earlier than previously suggested. Finally, we propose that several *Y. pestis* lineages were established during the Bronze Age, some of which persist to the present day.

[1] Max Planck Institute for the Science of Human History, Kahlaische Str. 10, 07745 Jena, Germany. [2] Institute for Archaeological Sciences, University of Tübingen, Rümelinstrasse 23, 72070 Tübingen, Germany. [3] Center of Excellence "Archaeometry", Kazan Federal University, Kazan 420008, Russian Federation. [4] Department of Anthropology and Ethnology, Xiamen University, 361005 Xiamen, China. [5] LLC "Gefest", Michurina Str. 4, Samara 443030, Russia. [6] State Institute of Culture, Agency for Preservation of the Historical and Cultural Heritage of the Samara Region, Samara 443010, Russia. [7] Samara State University of Social Sciences and Education, Maxim Gorky Str., Samara 443090, Russia. [8] Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, 8091 Zurich, Switzerland. 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)

Yersinia pestis, the causative agent of bubonic, pneumonic and septicaemic plague, evolved from the closely related environmental progenitor Y. pseudotuberculosis[1]. Although primarily a coloniser of sylvatic rodents via flea-dependent transmission, ancient DNA studies have demonstrated its status as an infectious disease in humans for the last 5000 years[2,3], and have confirmed its involvement in some of the most devastating historical pandemics[4,5]. The first historically recorded plague pandemic began with the Plague of Justinian (AD 541–543), and persisted until the eighth century AD[5,6]. The second pandemic occurred between the 14th and 18th centuries AD, began with the infamous Black Death of Europe in 1347[4,7] and was a precursor of modern-day plague epidemics over a wide geographic range[8].

Today, plague has a near-worldwide distribution and is maintained within sylvatic rodent populations[9]. Although several of these rodent reservoirs were established during the third plague pandemic that began in 19th century China[10,11], many of those identified in Central and East Asia and most notably those of the Caspian Sea region harbour Y. pestis strains that occupy basal positions in the global phylogeny (i.e., 0.PE2)[12]. This supports the idea of these foci having persisted for millennia[12–15]. What remains unknown, however, is the time period and processes involved in their establishment, and the level of Y. pestis genetic diversity harboured within them during the early phases of its evolution.

After its divergence from Y. pseudotuberculosis, Y. pestis acquired its high pathogenicity and distinct niche mainly by chromosomal gene loss[16] as well as the acquisition of two virulence-associated plasmids, pMT1 and pPCP1[1,17,18]. Throughout this process, one of the most crucial evolutionary adaptations related to its pathogenicity was its ability to colonise arthropods, a phenotypic/functional gain mediated by a combination of chromosomal and plasmid loci[19,20]. These genetic changes are central to the most common "bubonic" form of the disease, where bacteria enter the body via the bite of an infected flea, travel via the lymph to the closest lymph node and replicate while evading host defences. Recent ancient genomic investigations of Y. pestis have identified its earliest known variants in Eurasia during the Late Neolithic/Bronze Age period (LNBA) that show genetic characteristics incompatible with arthropod adaptation. These strains, therefore, have been considered incapable of an efficient flea-based transmission[2]; however, the alternative early-phase transmission could have provided an independent means of arthropod dissemination[2,3,21]. To date, the earliest evidence of a Y. pestis strain with signatures associated with flea adaptation has been reported during the Iron Age through shotgun sequencing of an ~2900-year-old genome from Armenia (strain RISE397), though at a coverage too low (0.25-fold) to permit confident phylogenetic positioning[2]. Although the mechanism by which the LNBA lineage caused human disease is unclear, its frequency in Eurasia during the Bronze Age[2,3] and its phylogeographic pattern that mimics contemporaneous human migrations are noteworthy[3].

The Bronze Age in Eurasia was a period of technological transition among human populations, often associated with the initiation of cultural and societal complexity[22]. Recent aDNA analysis of human remains from the time period between 5500 and 3200 BP has linked such transitions to a large-scale expansion of "Yamnaya" pastoralists from the Pontic–Caspian steppe both westwards into Europe, giving rise to the so-called "Corded–ware complex", and eastwards into Central Asia and the Altai region, represented by Early Bronze Age (EBA) cultures such as the "Afanasievo"[23,24]. Specifically in Europe, the "Yamnaya" migrations resulted in admixture with the local Neolithic farmer populations, forming the gene pool that appears to constitute European populations to the present day[23–25]. In addition, recent studies propose subsequent population expansions from Europe back into Central Asia during the Middle and Late Bronze Age (MLBA), which is genetically reflected by the appearance of European farmer-related ancestry among Late Bronze Age (LBA) steppe populations such as "Sintashta", "Srubnaya", "Potapovka" and "Andronovo"[24,26].

The central steppe region seems to have played a significant role as a migration corridor during the entire Bronze Age, and as such, it likely facilitated the spread of human-associated pathogens, such as Y. pestis, across Eurasia. Here, we explore additional Y. pestis diversity in that region by isolating strains from LBA Samara, in Russia. We identify a Y. pestis lineage contemporaneous to the LNBA strains with genomic variants consistent with flea adaptation. This reveals the co-circulation of two Y. pestis lineages during the Bronze Age with different properties in terms of their transmission and disease potentials.

Results

Y. pestis and human endogenous DNA screening. We screened a total of 64 million shotgun next-generation sequencing (NGS) reads (Supplementary Table 1) from nine teeth of nine individuals recovered from kurgan burials in the Samara region (see Supplementary Methods) to assess the endogenous human DNA content and the possible presence of Y. pestis. Our Y. pestis screening procedure involved (1) mapping of all reads against the CO92 reference genome (NC_003143.1), and (2) taxonomically classifying the reads to bacterial species using the metagenomic tool MALT[27] with a special focus on those assigned to both the entire Y. pseudotuberculosis complex, as well as Y. pestis specifically (Supplementary Table 1). As MALT has previously proven to be an efficient tool in binning reads from complex metagenomic datasets into their respective bacterial taxa and has been successfully used for identifying pathogen DNA in archaeological material[27], we considered individuals as putatively positive only when reads were assigned to Y. pestis by both conventional read-mapping and MALT. Our screening revealed four potentially positive individuals (Supplementary Table 1), one of which, individual RT5 (Fig. 1a), exhibited the highest amounts of endogenous Y. pestis DNA (0.11%, Supplementary Table 1). Notably, individual RT5 also exhibited exceptional human DNA preservation (31.3% of endogenous DNA, Supplementary Table 2). A shotgun-sequencing approach was, therefore, used for retrieving the entire Y. pestis and human genomes from this specimen. In addition, an in-solution Y. pestis enrichment approach was employed for putatively positive individuals.

Human uniparental and genomic analyses. Shotgun sequencing of RT5 resulted in 1.14 billion raw reads and a 4.2-fold average human genomic coverage (Supplementary Table 3). Genetic sex identification assigned RT5 to a male, which is in line with the anthropological assignment (Fig. 1a, Supplementary Methods). Nuclear contamination estimates based on X-chromosomal heterozygosity were low in RT5, estimated at an average of 0.5% (Supplementary Table 4), which permitted the usage of all generated human data. Y-chromosomal and mitochondrial assignment revealed the individual carrying R1a1a1b and U2e2a haplogroups, respectively (Supplementary Methods, and Supplementary Table 5). To gain insight into the ancestry of RT5, we performed principal component analysis (PCA)[28–30] and ADMIXTURE[31] analysis, where previously published ancient Eurasian populations[32–35] were used as comparative datasets (Fig. 1 b, c). Our PCA (Fig. 1b) and ADMIXTURE (Fig. 1b, c) plots show RT5 to have close genetic affinity to ancient populations from EBA Europe and the MLBA steppe, which are
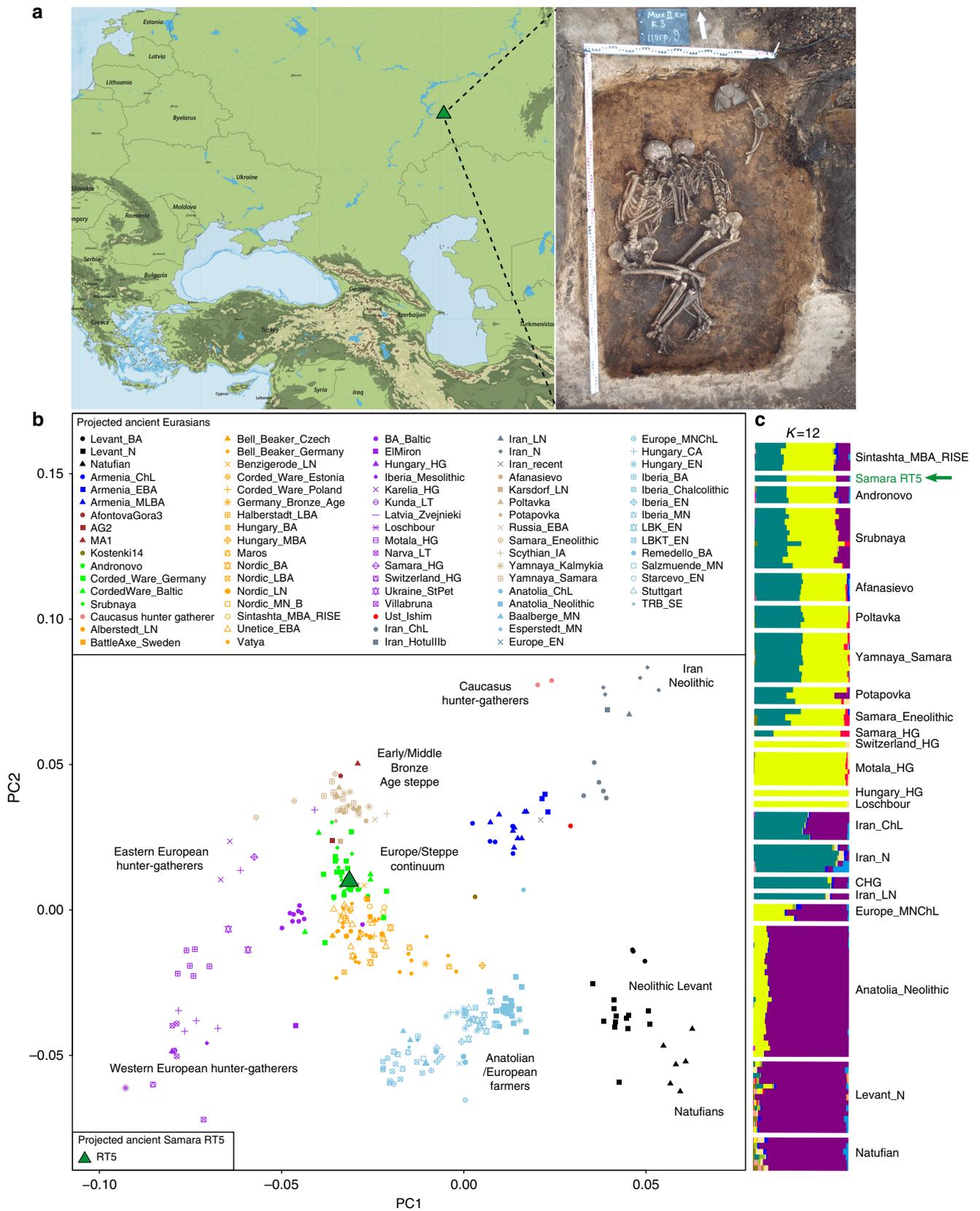
Fig. 1 Population genetic analysis to infer the ancestry of RT5. **a** Geographic location (map purchased from vectormaps.de) and picture of RT5 burial in the Mikhailovsky II site (picture credits to V.V. Kondrashin and V.A. Tsybin). **b** Principal component analysis (PCA) of modern-day western Eurasian populations (not shown) and projected ancient populations ($n = 82$, see population labels), including the newly sequenced RT5 individual from Samara and **c** estimation of ancestral admixture components using ADMIXTURE analysis ($K = 12$) (see Supplementary Methods)

genetically distinct from EBA populations from the Central Asian steppe, as they encompass early European farmer-related ancestry as part of their genetic composition (Fig. 1c). Examples of such groups include the European "Corded Ware"-associated populations, the "Andronovo" from the Altai region and the Samara-region "Srubnaya" culture with which RT5 has been archaeologically associated.

Y. pestis quality assessment and genome reconstruction. Although _Y. pestis_ has been previously identified in Bronze Age individuals[2,3], its presence in the Volga region "Srubnaya"-associated populations has not been characterised to date. After _Y. pestis_ capture, samples RT5 and RT6 yielded an average genomic coverage greater than 1-fold, with RT6 reaching 1.9-fold and RT5 reaching 32.3-fold (Supplementary Table 3). In addition, we retrieved a 9.2-fold _Y. pestis_ genome from RT5 shotgun sequencing alone (Supplementary Table 3). Comparison between the RT5-captured and deep-shotgun-sequenced _Y. pestis_ reads revealed nearly identical GC contents and deamination profiles (Supplementary Table 3 and Supplementary Fig. 1), but significantly different read-length distributions (Supplementary Fig. 2), with an average fragment length increase of 3.6 bp after capture (_t_-test, _P_-value < 2.2e-16). Assessment of SNP profiles inferred from captured and shotgun-sequenced RT5 reads shows no SNP-calling inconsistencies between the two datasets (Supplementary Data 1).

Moreover, during a first assessment of coverage across the plasmids (Supplementary Table 3), it became apparent that RT5 and RT6 harbour the 1.8 kb Yersinia murine toxin (_ymt_) gene locus on the pMT1 plasmid (Fig. 2a), which encodes for a virulence factor essential for the colonisation of the flea's midgut. This gene is absent in all previously sequenced LNBA strains[2,3], though it has been identified in a later Iron Age individual (~2900 BP) from modern-day Armenia (RISE397) (Fig. 2a)[2].

Y. pestis phylogenetic analysis. In order to reconstruct a phylogeny using a maximum likelihood (ML) approach, we analysed our RT5 _Y. pestis_ isolate against a total of 177 genomes, including previously published ancient strains, as well as a worldwide modern _Y. pestis_ dataset (see Methods and Supplementary Data 2)[2–4,6,8,12,14,15,36]. _Y. pseudotuberculosis_ (strain IP32953)[18] was used as an outgroup for rooting the tree. Our ML tree (Fig. 2b) is consistent with previously published phylogenies, where the LNBA isolates occupy the most basal _Y. pestis_ branch[2,3]. Although ^{14}C dates place RT5 within the LNBA time range (Fig. 2c, Supplementary Table 6), it occupies an unexpected position in the phylogeny, appearing further derived along branch 0 and being part of a polytomy that gave rise to three independent lineages (Fig. 2b). We explored the possibility of this polytomy reflecting a limited phylogenetic resolution at that node, given the exclusion of missing data (complete deletion) in our analysis. By contrast, we were able to replicate this topology after inclusion of all data in our phylogenetic reconstruction (Supplementary Fig. 3). In addition, since the coverage of RT6 was too low for confident SNP-calling, its genotype was manually explored after clipping 2 bp from the 3′ and 5′ ends of all reads to avoid the interference of post-mortem damage with our SNP assignments (see Methods). As RT5 possesses five unique, non-homoplastic, SNPs, we assessed their similarity with RT6. For all such positions covered in RT6, it possesses the identical SNP genotype as RT5 (Supplementary Table 7), suggesting that the two strains are likely identical, or are at minimum closely related. Such a result is expected since the two individuals derive from a double burial (Fig. 1a).

We further assessed the relatedness of RT5 with the previously published Iron Age isolate (RISE397)[2]. The coverage of RISE397 was too low (0.25-fold) to permit robust phylogenetic analysis. Therefore, manual genotyping of phylogenetically "diagnostic" SNPs was performed to infer its possible positioning (see Methods). As 22.5% of the RISE397 genome was covered ≥1-fold, and only 3.7% of the genome was covered ≥2-fold, we considered all mapping reads for this analysis. Reads covering informative positions were authenticated based on whether they carried diagnostic SNPs or/and carried additional substitutions that were consistent with terminal deamination, which is characteristic of aDNA (Supplementary Data 3)[37]. RISE397 DNA reads cover 26% (12/46) of the diagnostic positions leading to the RT5 node, making it clearly distinct from the ancestral 0. PE2 and 0.PE7 genomes since it possesses the derived-state (CO92 reference) allele in positions where the 0.PE2 and 0.PE7 genomes have ancestral variants (Supplementary Data 3, Supplementary Fig. 4). In addition, 30% (7/23) of SNPs extending from RT5 to the Justinian 2148 branch were covered in RISE397 (Supplementary Data 4). All such positions were identical to RT5, exhibiting the ancestral alleles, and hence supporting that RISE397 is basal to lineage 0.ANT1 (Supplementary Data 3, Supplementary Fig. 4). Finally, only one of five private-derived RT5 SNPs was covered in RISE397, where it instead matched the reference sequence (Supplementary Table 7). Although achieving a higher coverage would be necessary to verify its precise positioning, our analysis suggests that RISE397 and RT5 are closely related strains and potentially originated from the same progenitor (Supplementary Fig. 4). The node which gave rise to RT5 and perhaps also RISE397 seems to have initiated a radiation event that gave rise to all historical and extant _Y. pestis_ lineages that have been isolated to date, with the exception of the more basal 0.PE2, 0.PE7 and LNBA (Fig. 2b, Supplementary Fig. 4).

Y. pestis divergence time estimates and demographic analysis. To estimate the time to the most recent common ancestor (tMRCA) of all _Y. pestis_ strains, we employed the coalescent constant size, and the coalescent Bayesian skyline models implemented in BEASTv1.8[38-40]. According to our marginal likelihood (MLE) estimates computed via path sampling (PS) and stepping stone sampling (SS)[41], the Bayesian skyline model is the more suitable demographic model for the current dataset (Supplementary Table 8 and Methods). Nevertheless, both analyses produced overlapping divergence date intervals (Supplementary Fig. 5). While the constant size method is unlikely to represent a realistic demographic history model for epidemic pathogens, it has often been a preferred dating method[2,3,12]. Here, it produced a coalescent date estimate of 6797y BP (HPD 95%: 5299–8743) for _Y. pestis_, which is about 2000 years older than the oldest strains thus far identified[2,3] (Supplementary Table 9). In contrast, the coalescent skyline method, which allows for population size changes through time, produced a narrower interval and a younger tMRCA estimated at 5727y BP (HPD 95%: 4909–6842) (Supplementary Fig. 6, Supplementary Table 9).

In addition, using the Bayesian skyline model, we estimated effective population size (N_e) changes across the evolutionary history of _Y. pestis_. Our skyline plot (Supplementary Fig. 7) reveals an initial population expansion at ~4000y BP. Such an increase corresponds temporally with the RT5 polytomy described here (Fig. 2b) that we date to 4011y BP (HPD 95%: 3760–4325) (Supplementary Fig. 6). Although such effect could arise as a result of sampling bias in the data, it is of note that from 159 modern _Y. pestis_ strains considered for the present analysis, only 13.2% are phylogenetically ancestral to the described
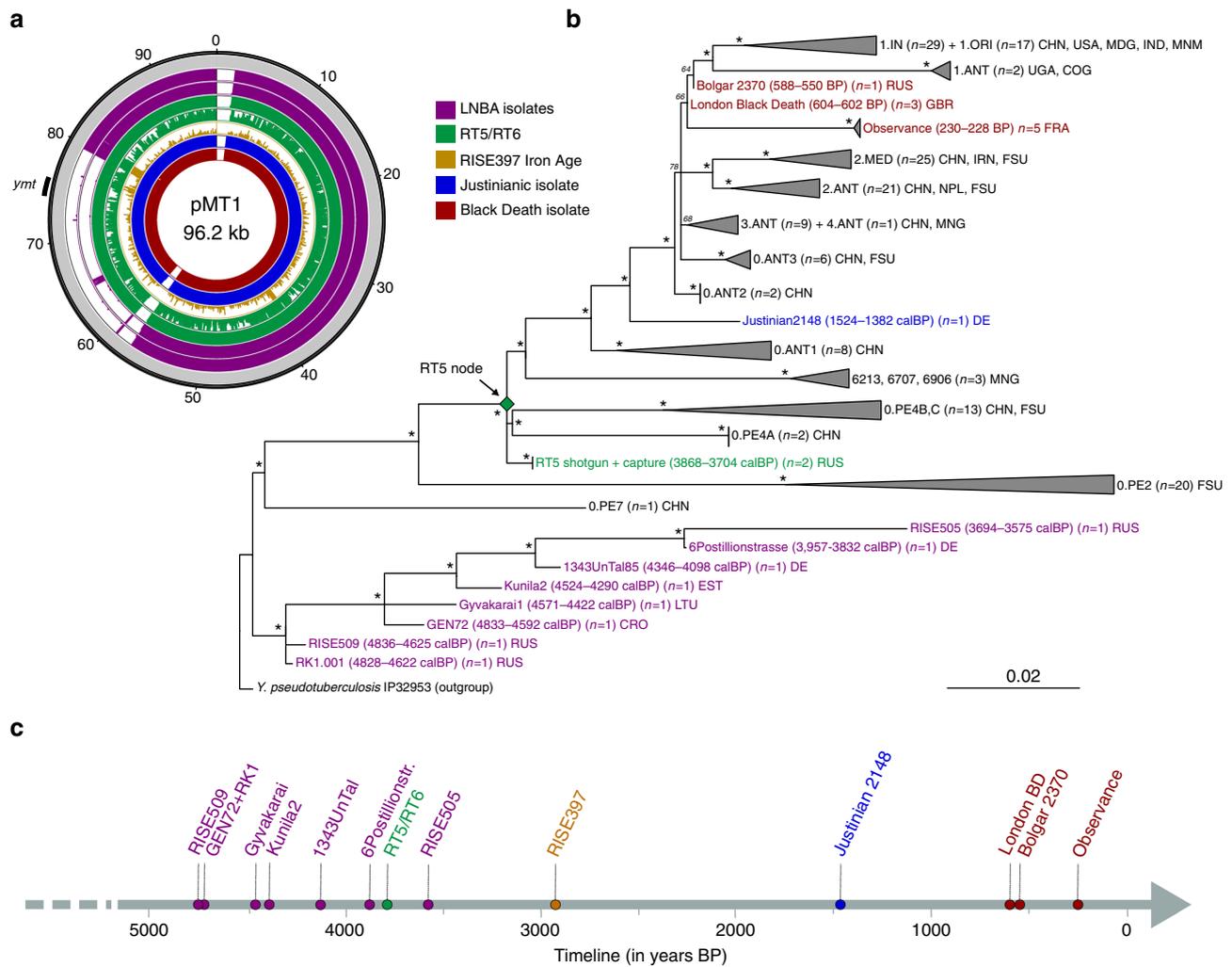
Fig. 2 *Y. pestis* genomic characterisation and maximum likelihood phylogeny. **a** pMT1 coverage plots made with the Circos[80] software. The plots were constructed to a maximum coverage of three-fold, and average coverage was calculated over 100-bp windows. The presented strains are in the following order starting from the outmost: CO92 pMT1 (reference) in grey, the oldest (RISE509) and youngest (RISE505) isolates from the LNBA lineage are shown in purple, RT5 and RT6 are shown in green, the Iron Age RISE397 isolate is shown in brown, a Justinianic isolate from Altenerding (Germany) is shown in blue and a London Black Death isolate is shown in red. The position of the *ymt* gene within the pMT1 plasmid is indicated on the plot. **b** A worldwide dataset of *Y. pestis* ancient and present-day chromosomal genomes ($n = 179$) was used to reconstruct the phylogenetic tree, considering 1054 SNP positions (see Supplementary Fig. 3 for a phylogeny using all sites). The main branches were collapsed to enhance the clarity of the phylogeny, and branch lengths are shown as number of substitutions per site. The newly sequenced RT5 strain (green) was included in the phylogeny alongside eight Bronze Age strains belonging to the LNBA lineage (purple), a single Justinianic strain (blue), and nine second pandemic strains (red). Asterisks denote bootstrap values >95 (1000 bootstrap iterations carried out). The two-sigma (95.4%) radiocarbon or archaeological dates of Bronze Age and historical strains are shown. Country or geographical region abbreviations are as follows: CHN (China), USA (United States of America), MDG (Madagascar), IND (India), IRN (Iran), MNM (Myanmar), RUS (Russia), GB (Great Britain), DE (Germany), FRA (France), MNG (Mongolia), NPL (Nepal), FSU (Former Soviet Union), CGO (Congo), UGA (Uganda), LTU (Lithuania), EST (Estonia) and CRO (Croatia). See also Supplementary Fig. 4 for the inferred phylogenetic positioning of RISE397. **c** Timeline spanning radiocarbon and archaeological dates, from which *Y. pestis* genomic data have been included in this study. Points on the timeline indicate median dates

polytomy, while the majority derive from it (Fig. 2b). Subsequently, our skyline plot reveals a population decline starting at ~300y BP that is immediately followed by an increase (Supplementary Fig. 7). This result is likely associated with the coalescence times of the most extensively sampled modern isolates, particularly those related to the third plague pandemic (Fig. 2b, Supplementary Fig. 6).

***Y. pestis* virulence factor analysis**. The presence of several *Y. pestis* virulence-associated genes was evaluated in RT5 (Fig. 3). While the LNBA, 0.PE2 and 0.PE4 strains seem to lack certain virulence determinants, RT5 harbours all known virulence factors

with the exception of the filamentous prophage (YpfΦ), which is, however, most consistently identified among 1.ORI strains (Fig. 3, Fig. 2b)[42,43].

Another important gene for *Y. pestis* virulence is *pla* located on the species-specific pPCP1 plasmid[44]. Although the gene is largely conserved among *Y. pestis* strains, an isoleucine (ancestral) to threonine (derived) alteration at amino acid position 259 has been used to differentiate the most basal isolates (LNBA, 0.PE7, 0.PE2 and 0.PE4) from the rest of *Y. pestis*[44]. In RT5, the *pla* genotype was manually explored, and was found to exhibit the ancestral allele (Supplementary Fig. 8). Although the ancestral *pla* allele has been associated with a less-efficient bacterial dissemination in mammals[44], modern strains from lineages 0.PE4 and 0.
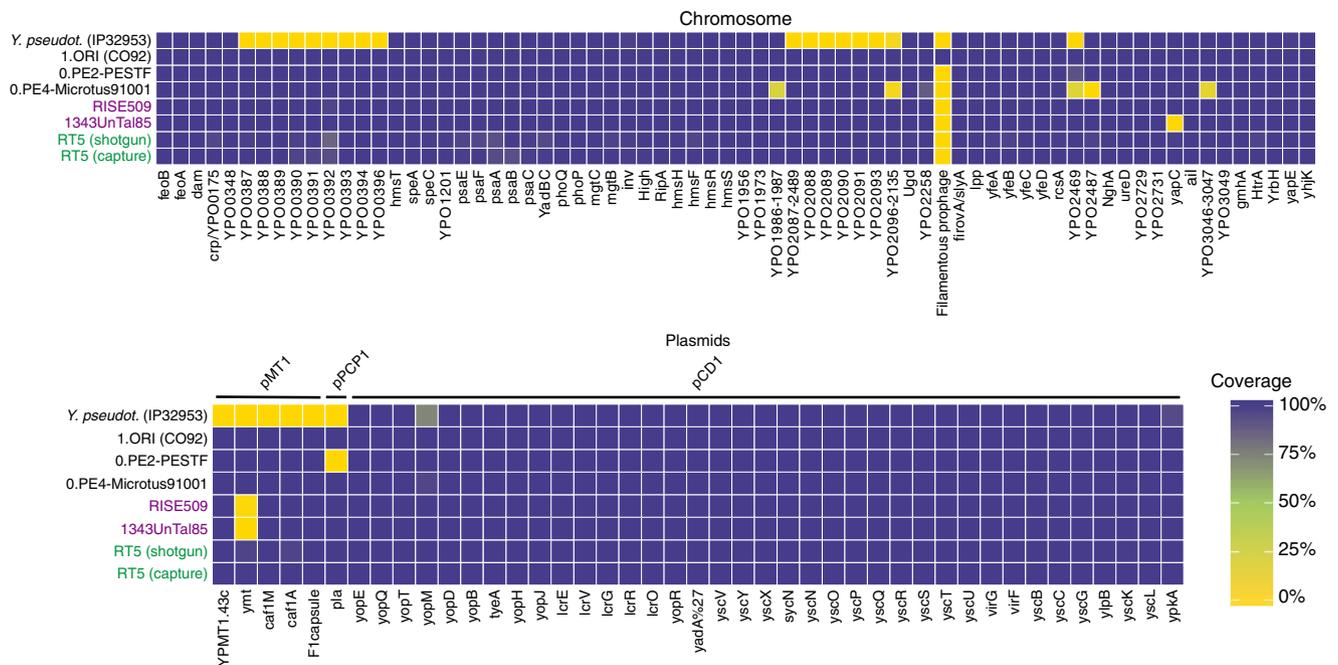
Fig. 3 Heat map of coverage across virulence-associated genes. The virulence potential of RT5 shotgun-sequenced and captured genomes is compared to representative strains from the LNBA lineage, namely, RISE509 (whose virulence profile is identical to that of isolates RK1.001, GEN72, Gyvakarai1 and Kunila2[3]) and 1343UnTal85 (whose virulence profile is identical to 6Postillionstrasse and RISE505[3]). In addition it is compared to modern-isolate representatives 0.PE4-Microtus91001, 0.PE2-PestoidesF and 1.OR1-CO92, and to *Y. pseudotuberculosis* (strain IP32953). The virulence factors inspected were located on the *Y. pestis* chromosome, as well as on the pMT1, pPCP1 and pCD1 plasmids. The percentage of each gene covered (scale bar) was computed and plotted in the form of a heatmap using the ggplot2[79] package in R

PE7 have proven to be potent inducers of bubonic plague in humans[12].

In addition, we manually explored the status of *ureD*, PDE-2, PDE-3 and *rcsA*, all of which have been either lost or inactivated in *Y. pestis* by substitution or single-nucleotide InDels. The inactivation of these genes contributes to *Y. pestis*' ability to colonise, block and be transmitted via fleas (for more details, see Methods)[19,45]. Their active variants have been identified in previously published LNBA strains[2,3], thus suggesting either an inability or a lower efficiency in arthropod-based transmission. By contrast, we find that RT5 possessed the inactive form of all those genes, with the exception of a nonsense mutation in PDE-3 where it shows the ancestral allele (Supplementary Fig. 9). Together with the active *ymt* locus on the pMT1 plasmid (Fig. 2a), this suggests that RT5 was already adapted to the flea vector during the Bronze Age. Moreover, immune evasion by suppression of flagellar genes in *Y. pestis* is considered an important evolutionary advantage associated to a more complex niche adaptation that is absent in its closest ancestor *Y. pseudotuberculosis*[46]. The *flhD* regulatory gene is part of the *flhDC* operon and is expressed in a temperature-dependent manner in *Y. pseudotuberculosis*[46], but is inactive in all extant and historical *Y. pestis* strains sequenced to date. Although the strains belonging to the LNBA lineage encompass the active variant of *flhD*, RT5 contains the derived, inactive form (Supplementary Fig. 9).

As presented in the phylogeny (Fig. 2b), RT5 appears closely related to lineage 0.PE4 also referred to as "microtus". Since certain "microtus" strains have been associated with a decreased pathogenicity in humans, we inspected genes previously identified as responsible for this phenotype and verified their status in RT5 (Supplementary Table 10)[47,48]. Some of these loci seem to have been lost in 0.PE4 (Fig. 3 and Supplementary Table 10), and others have been disrupted by insertion sequences (IS) (*IS100* and *IS285*) or inactivated by substitutions/InDels (Supplementary

Table 10). In RT5, all such genes appear in their active form (Fig. 3, Supplementary Fig. 10), and therefore we have no evidence that this strain had a decreased virulence in humans.

Discussion

Our results contribute to investigations regarding the evolution of *Y. pestis* and its disease potential in past human populations. We used shotgun sequencing and in-solution capture to reconstruct *Y. pestis* genomes from Bronze Age individuals (RT5 and RT6) in the Samara region. In addition, we retrieved a 4.2-fold human genome from individual RT5 through shotgun sequencing. Population genetic analysis identified individual RT5 as having close genetic affinity to EBA European populations and MLBA populations from the Eurasian steppe region (Fig. 1b). In particular, we show the presence of Yamnaya-related as well as farmer-related ancestry in RT5 (Fig. 1c). Our genomic characterisation is in line with previously described BA Eurasian populations, including "Srubnaya" individuals from Samara[26], where European farmer-related ancestry becomes present in Central Asia during the MBA as a result of population movements from Europe back into the steppe region[24,26].

We identify two ~3800-year-old individuals (Fig. 1a, Supplementary Table 3) out of nine analysed that were infected with *Y. pestis* at the time of their deaths. The Samara *Y. pestis* genomes presented here reveal greater lineage diversity during the Bronze Age than was previously described. Compared to the recently published LNBA isolates[2,3], RT5 and RT6 form a distinct branch in the *Y. pestis* phylogeny (Fig. 2b), deriving from a polytomy that gave rise to at least three separate lineages, two of which have persisted to the present day. RT5 falls only five derived SNPs away from the described polytomy (Supplementary Table 7). Our dating analyses consistently suggest the presence of this putative ancestor at ~4000y BP (Supplementary Fig. 6)

followed by a population expansion shortly after that time (Supplementary Fig. 7). Though its place of origin is not yet empirically identified, given the close genetic and temporal affinity to RT5 (Fig. 2b), a steppe source is plausible. Given that previous research has proposed a relationship between rapid *Y. pestis* expansions and historical plague epidemics in humans[12], future investigations of lineage diversity from modern and ancient sources may reveal additional details on this ancient radiation event.

Apart from RT5, a second known lineage that emerged as part of the described polytomy is "microtus" (0.PE4)[12,49]. 0.PE4 is found today in Central and East Asia, with some isolates being associated with bubonic plague infections in humans[12], and others being characterised as avirulent to humans[47,48,50,51]. Despite the variation in pathogenicity of 0.PE4 isolates, the phylogenetically related RT5 strain seems unaffected by genetic alterations/disruptions associated with reduced virulence (Fig. 3, Supplementary Fig. 10 and Supplementary Table 10). In addition, the third and most diverse branch established during this radiation event is one that gave rise to a multitude of *Y. pestis* lineages, which survive to the present day (Fig. 2b). These include 0.ANT along with the strain that caused the first historically documented plague pandemic (Plague of Justinian––sixth century)[5,6]; the entire branch 1 that includes strains responsible for the second (Black Death—fourth century)[4,8,36] and third (China—19th century) plague pandemics[11] and branches 2, 3 and 4 that are typically isolated from modern sylvatic rodents in Central and East Asia (Fig. 2b)[11,12].

A recent study has suggested that flea-adapted *Y. pestis*, along with its potential to cause bubonic plague in humans, likely originated around 3000y BP[2]. Contrary to such conclusions, the lineage giving rise to our *Y. pestis* isolates (RT5 and RT6) likely arose ~4000 years ago (Supplementary Tables 6 and 9), and possessed all vital genetic characteristics required for flea-borne transmission of plague in rodents, humans and other mammals. These include a fully incorporated *ymt* locus (Fig. 2a), and the inactive forms of *ureD*[52], PDE-2, PDE-3 and *rcsA* genes[19], as well as an inactive *flhD* flagellin gene (Supplementary Fig. 9). Moreover, our analysis of the previously published Iron Age RISE397 strain from modern-day Armenia[2] revealed its close relationship to RT5 and RT6 (Supplementary Fig. 4). Note that the modern 0.PE2 and 0.PE7 lineages, which are known to possess all genomic characteristics that confer adaptation to fleas[19], fall ancestral to RT5 (Fig. 2b) and RISE397 (Supplementary Fig. 4), but are more derived than the LNBA lineage. Our phylogenetic and dating results thus suggest that 0.PE2 and 0.PE7 also originated during the Bronze Age, with their mean divergence here estimated to 4474 (HPD 95%: 3936–5158) and 5237 (HPD 95%: 4248–6346) years BP, respectively, based on the Bayesian skyline model (Supplementary Table 9). While these lineages may have been confined to sylvatic rodent reservoirs during the EBA, the possibility that they co-circulated among human populations contemporaneously with the LNBA lineage should be considered. Although the places of origin of 0.PE2 and 0.PE7 are not known, today, their strains are isolated from modern-day China and the Caucasus region. In terms of their disease potential, both 0.PE2 and 0.PE7 possess pMT1 plasmids with fully functional *ymt* genes, but 0.PE2 strains lack pPCP1[44], and though frequently recovered from sylvatic rodent reservoirs, their virulence in humans is not known. On the other hand, the more basal 0.PE7 contains pPCP1[2] and has previously been associated with human bubonic plague[12]. It is, therefore, tempting to hypothesise that efficient flea adaptation in *Y. pestis*, as well as the potential for bubonic disease, might have evolved earlier than 5000 years ago.

Overall, the detection of *Y. pestis* in Bronze Age human remains from Eurasia has suggested the presence of the pathogen in this vast geographic area along with its ability to cause bubonic plague millennia before the first historically documented plague pandemic[2,3]. It seems possible that already in the Bronze Age, with the establishment of transport and trade networks, the interconnectivity between Europe and Asia that is also reflected in the ancient human genomes, likely contributed to the spread of infectious disease. Similarly, the abundant trade routes of the medieval period are considered the main conduit for plague's movement between Asia and Europe[8,12]. Our current data suggest a more complex model, where at least two human-associated lineages (LNBA and RT5) with different transmission potentials were established in Eurasia during the Bronze Age (Fig. 2b, c). Whether these lineages had equal prevalence among human populations, and the extent to which human practices contributed to their dissemination, are concepts requiring further investigation. Additional Bronze Age/Iron Age genomes could provide further insights into the early stages of *Y. pestis* evolution, and help pinpoint key events that contributed to the high virulence and spread of one of humankind's most notorious pathogens.

Methods

Sampling and extraction. All laboratory procedures were performed in the dedicated ancient DNA facilities of the Max Planck Institute for the Science of Human History in Jena, Germany.

Teeth from nine individuals (one tooth from each), buried in the Mikhaylovka II tombs of the Samara region in Russia, were sectioned in the cementoenamel junction using a coping saw and 50–100 mg of dental pulp was removed from each tooth using a dental drill.

Extraction of 50–60 mg of dental pulp from each tooth sample was performed using a previously described protocol; optimised for the recovery of short DNA fragments, most typical of ancient DNA[53]. An initial lysis step was performed over a 12–16 h incubation of the dental pulp powder in 1 ml of extraction buffer (0.45 M EDTA, pH 8.0, and 0.25 mg ml^{-1} proteinase K) at 37 °C. Following extraction, DNA was bound to a silica membrane using a binding buffer containing guanidine hydrochloride (protocol previously described in ref. [53]) and purified in combination with 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 and 0.05% Tween20). One extraction blank and one positive extraction control (previously assessed cave-bearing specimen) were taken along for the extraction slot.

Illumina library preparation and sequencing. To screen all samples for the presence of *Y. pestis* and human endogenous DNA, 10 µl of each extract was converted into double-stranded Illumina NGS libraries, using a previously described protocol[54], without initial uracil-DNA-glycosylase (UDG) treatment[55]. A positive control (cave-bearing specimen) and a negative library control (H$_2$O) were taken along for the experiment. A total of 1 µl from each library was subsequently quantified using IS7/IS8 primers. A combination of two unique indexes (8 bp length of each index sequence) also containing the universal IS5/IS6 priming sites was assigned to each sample for subsequent multiplex sequencing[56]. The libraries were then indexed through a ten-cycle amplification reaction using the *Pfu Turbo Cx Hotstart DNA Polymerase* (Agilent). Indexed PCR products were purified using a Qiagen MinElute kit (Qiagen), eluted in TET (10 mM Tris-HCl, 1 mM EDTA, pH 8.0 and 0.05% Tween20) and then qPCR quantified using IS5/IS6 primers, to assess the efficiency of the indexing reaction. After this, indexed libraries were amplified for different amounts of cycles, to achieve a total of 10^{13} DNA copies per reaction in order to avoid polymerase saturation and heteroduplex formation. PCR products were again purified using a Qiagen MinElute kit (Qiagen) and eluted in TET (10 mM Tris-HCl, 1 mM EDTA, pH 8.0 and 0.05% Tween20). The concentration (ng µl^{-1}) of the indexed and amplified libraries was then measured using a 4200 Agilent Tape Station Instrument (Agilent). Finally, all samples were diluted and pooled at equimolar ratios to achieve a final 10 nM pool that would serve as template for sequencing.

In silico screening for *Y. pestis* reads. The sample pool was single-read sequenced on a HiSeq 4000 platform using a 1 × 76+8+8 cycles chemistry kit according to the manufacturer's protocol, to produce between 5,969,436 and 8,215,620 raw demultiplexed reads per sample. Pre-processing of reads was performed using the automated pipeline EAGER v1.92[57] to clip adaptors (using ClipAndMerge) and to filter reads for sequencing quality (minimum base quality 20) and length (keeping all reads ≥30 bp). Mapping was performed using BWA[58] implemented in EAGER to *Y. pestis* CO92 (NC_003143.1), using a −n parameter of 0.01, a −l seedlength of 16 and subsequently using SAMtools to filter for reads with

a mapping quality (−q) of 37. The MarkDuplicates tool in Picard (1.140, http://broadinstitute.github.io/picard/) was used to remove duplicates.

In addition, the Megan ALignment Tool (MALT)[27] was used to assess the metagenomic composition of the samples, as well as a screening tool for the identification of *Y. pestis*. All bacterial genomes available at GenBank were used as a reference database for the programme (NCBI RefSeq, December 2015). Pre-processed reads were used as input for MALT (version 0.3.6), and the parameters were set to 85 for the minimum percent identity (--minPercentIdentity), 0.01 for the minimum support parameter (--minSupport), using a top percent value of 1 (--topPercent) and the semi-global alignment mode. All the remaining parameters were set to default. The results were viewed in MEGAN6[59]. Putatively positive *Y. pestis* samples were evaluated by comparing the number of reads mapping to *Y. pestis* CO92 (NC_003143.1) to the reads assigned by MALT on the *Y. pestis* and *Y. pseudotuberculosis complex* nodes (Supplementary Table 1).

In-solution *Y. pestis* capture and deep-shotgun sequencing. Rich double-stranded DNA libraries were prepared for in-solution capture and deep-shotgun sequencing of putatively positive *Y. pestis* samples, using 50 µl of extract (or 2 × 25 µl of extract), according to a previously described protocol[54], with an initial partial-UDG treatment step[60], where UDG in combination with endonuclease VIII (USER enzyme, New England Biolabs) were used to remove all deaminated cytosines (uracils) with the exception of terminal uracil nucleotides that lack 5′ phosphate. Double-indexing and subsequent library amplification steps were carried out as mentioned in the previous section "Illumina library preparation and sequencing". At this stage, the sample RT5 was diluted to 10 nM for deep-shotgun sequencing on a HiSeq 4000 platform using a 1 × 76+8+8 cycles chemistry kit. In addition, 1–2 µg of samples RT5 and RT6 were in-solution captured as described previously[3], where a combination of the following *Y. pestis* and *Y. pseudotuberculosis* genomes were used as templates for probe design: CO92 chromosome (NC_003143.1), CO92 plasmid pMT1 (NC_003134.1), CO92 plasmid pCD1 (NC_003131.1), KIM 10 chromosome (NC_004088.1), Pestoides F chromosome (NC_009381.1) and *Y. pseudotuberculosis* IP32953 chromosome (NC_006155.1). Samples were captured in separate wells of a 96-well plate, whereas extraction and library blanks (data not shown) with non-overlapping index combination were pooled and captured in a single well. Sequencing was performed on a HiSeq 4000 platform using both single-end (1 × 76+8+8 cycles) as well as paired-end (2 × 76+8+8 cycles) chemistry kits.

***Y. pestis* read authentication and genome reconstruction.** Sequencing resulted in up to 1,140,960,213 raw reads per sample. Adaptor trimming of raw, demulti-plexed, reads was performed using leeHom[61]. Subsequently, length and quality-filtering steps were performed in EAGER, as mentioned in the previous section "In silico screening for *Y. pestis* reads". After pre-processing, captured paired-end and single-end reads from the same individuals were merged into a single file for mapping. BWA[58] integrated in EAGER was used for mapping against the *Y. pestis* CO92 reference (NC_003143.1)[62] using the following parameters: −n 0.1, −l 32, and subsequently SAMtools was used to filter for reads with mapping quality of 37 (−q 37). Mean coverages were estimated using QualiMap v.2.2.1[63] and DNA deamination profiles typical of aDNA were calculated using MapDamage 2.0[64]. For genome reconstruction, and for downstream SNP calling, the same pipeline was followed with a single alteration: after adaptor trimming, reads were inputted into EAGER and 2 bp were trimmed from each end using ClipAndMerge prior to length filtering and mapping to eliminate post-mortem damage that might affect downstream SNP calling.

Read-length comparison of capture and shotgun *Y. pestis* reads. Two datasets derived from the same individual (RT5), sequenced using the 1 × 76+8+8 cycles kit parameters, were used to compare the read-length distributions of shotgun-sequenced reads and captured reads. For this analysis, datasets were limited to the same genomic coverage (~9-fold), to ensure uniform comparison and avoid any biases that might arise from unequal coverage. Reads shorter or equal to 74 bp were considered for the analysis, to avoid the incorporation of reads that still contain traces of adaptor, or are longer than 76 bp. Box-plot comparisons and Student's *t*-test were calculated using R version 3.4.1[65].

***Y. pestis* SNP calling and phylogenetic analysis.** For SNP calling, we used the UnifiedGenotyper of the Genome Analysis Toolkit (GATK)[66]. The newly produced RT5 shotgun-sequenced and captured genomes were analysed alongside 177 previously published *Y. pestis* genomes (179 in total), including one previously published historical strain from the Plague of Justinian[6], nine genomes from the second plague pandemic[4,8,36], eight previously published LNBA genomes[2,3] and a global dataset of 159 modern *Y. pestis* genomes (Supplementary Data 2)[11,12,14,15,47,62,67–70]. A *Y. pseudotuberculosis* strain (IP32953)[18] was used as an outgroup. A vcf file was produced for every sample using the "EMIT_ALL_SITES" in GATK[66], which generated a call for all positions in the reference genome. In addition, the custom Java programme MultiVCFAnalyzer v0.85[71] (https://github.com/alexherbig/MultiVCFAnalyzer) was used to produce a combined SNP table, including all variable positions across our dataset, with the exclusion of previously defined noncore regions and homoplasies, as well as repeat regions, tRNAs, rRNAs and tmRNAs[11,12]. In addition, we used ClonalFrameML[72] to identify additional

homoplasies or recombinant regions in our dataset by using a full-genome alignment and a RAxML[73] ML SNP tree as input, as well as the -em and the -ignore_incomplete_sites options for running the programme. Through this analysis, we identified seven additional regions (resulting in 26 SNPs) and two homoplastic SNPs (28 SNPs in total), which were also excluded from the comparative SNP analysis (Supplementary Table 11). For the remaining data, SNPs were filtered according to the following criteria: (1) homozygous SNPs and reference alleles were called when covered at least three-fold with a minimum genotyping quality of 30, (2) in cases of heterozygous positions, a SNP or reference base was called when supported by at least 90% of the reads covering the respective position and (3) if none of the criteria could be fulfilled, a "N" was inserted in the respective position. A total of 3821 SNP positions were called in the current dataset.

From the resulting SNP alignment, two ML phylogenetic trees were inferred with RAxML (version 8.2.9)[73], using the generalised time-reversible (GTR) substitution model with gamma-distributed rates (six rate categories). The first included all data. For the second tree reconstruction, all columns with missing data were excluded (complete deletion), which resulted in a total of 1054 SNP positions to be considered for the phylogeny. A total of 1000 bootstrap replicates were carried out to estimate the topology support of each tree.

In addition, the phylogenetic positioning of RT6 (present study) and RISE397[2] was manually explored using the following methods:

- For RT6: To check whether RT5 and RT6 form the same phylogenetic branch, SNPs specific to the RT5 genome ($n = 5$, Supplementary Table 7) were assessed for their presence in RT6. RT6 possesses four out of five SNP positions covered at least one-fold. All positions covered encompass identical alleles to RT5.
- For RISE397: The SNP table produced by MultiVCFAnalyzer was filtered for all diagnostic positions leading from the root of the tree towards the RT5 node ($n = 46$ SNP positions). In addition, the SNP table was independently filtered for all positions leading from the RT5 node to the Justinianic node (branch represented by Justinian 2148 strain) ($n = 23$ SNP positions), for which RT5 appears to have ancestral alleles. Missing data (N's) were excluded from this SNP analysis. The state of all alleles in RISE397 was then manually inspected using the Integrative Genomics Viewer[74]. The reads covering all respective positions were visually authenticated by assessing whether they include terminal substitutions that could be explained by aDNA damage (Supplementary Data 3, 4).

Divergence date and demographic analyses. TempEst v1.5[75] was used to assess for the presence of temporal signal in the dataset. Inclusion of the ^{14}C or archaeological dates for all ancient isolates resulted in a 0.6 correlation coefficient, which permitted the proceeding with molecular dating analysis. The software package BEAST v1.8[39] was used to estimate the divergence time of *Y. pestis* lineages, using the coalescent constant size[40] and the coalescent Bayesian skyline[38] models. The SNP alignment including only *Y. pestis* strains was used as input, after removal of all missing data (complete deletion). The tip dates of all modern *Y. pestis* strains were set to 0 years before present (BP). The ages of all ancient and historical *Y. pestis* genomes were estimated with their prior uniform distributions based either on the two-sigma (95.4%) ^{14}C date interval[2,3,6] or the archaeological dates[4,8,36] in years BP, as follows: RISE509 (4836–4625, median: 4729), RK1.001 (4828–4622, median: 4720), GEN72 (4833–4592, median: 4721), Gyvakarai1 (4571–4422, median: 4485), Kunila2 (4524–4290, median: 4427), 1343UnTal (4346–4098, median: 4203), 6Post (3957–3832, median: 3873), RT5 (3868–3704, median: 3789), RISE505 (3694–3575, median: 3635), London Black Death (602–604, median: 603), observance (228–230, median: 229), Bolgar 2370 (550–588, median: 569) and Justinian 2148 (1382–1524, median: 1453). In addition, we used MEGA7[76] to test whether there is an equal evolutionary rate across our phylogeny. The strict clock rate was rejected, and therefore we applied a lognormal relaxed clock[77] for all dating analyses, along with the GTR model of nucleotide substitution (six gamma categories), as previously performed[2,3]. An ML phylogeny was reconstructed using RaxML[73] and was used as a starting tree. The LNBA lineage and the rest of *Y. pestis* strains were constrained to be two separate monophyletic groups in BEAUti v1.8[39]. A single chain of 300,000,000 states was run for each model setup, sampling every 10,000 states. In addition, we estimated MLE using PS/SS sampling[41] as part of the same set-up in BEAUti to assess which of the two demographic models is best fit for our data. We run this analysis for an extra 300,000,000 states divided between 100 steps (3,000,000 states each), using an alpha parameter of 0.3. After run completion, the molecular dating results were viewed in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure that all expected sample sizes were above 200. TreeAnnotator was used to produce a maximum clade credibility tree with a 10% burn-in (excluding the first 3000 trees), which resulted in processing of 27,001 trees for each analysis with a Jeffreys prior distribution (1.0) for the population sizes. In addition, for the coalescent skyline analysis, we used 20 as the dimension for the population and group sizes. Once the chain was complete, we used LogCombiner[39] to resample MCMC states at lower frequency (every 300,000) with a 10% burn-in, and the resultant .log and .tree files were used as input for the skyline plot construction in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

Analysis of virulence factors. As map-quality filtering may influence read mappability in certain chromosomal and plasmid regions, we used a mapping quality filter of 0 (–q parameter) to evaluate the presence or absence of chromosomal and plasmid virulence-associated genes in RT5 compared to previously published modern and LNBA *Y. pestis* genomes[48]. Bedtools[78] were used to calculate the percentage of gene covered across each region, and a heatmap was plotted using the ggplot2[79] package of R version 3.4.1[65].

In addition, the virulence-associated genes *flhD*, PDE-2, PDE-3, *ureD* and *rcsA*, which are known to have become inactivated in *Y. pestis* by either mutation or single-nucleotide insertions/deletions[19], were instead manually explored using IGV[74]. Gene *flhD*, associated with flagellar biosynthesis and whose silencing contributes to immune evasion, is inactivated by a frameshift caused by a T insertion, present at position 1,892,659 in CO92[46]. PDE-2, a phosphodiesterase gene contributor in biofilm degradation is inactivated by a T insertion in a six-T stretch at position 1,434,044 in CO92. In addition, PDE-3, also part of the same biofilm-degradation mechanism, is affected by two mutations, a C > T change (also called the PDE-*pe*' allele[19]), and a nonsense G > A substitution, which are, respectively, shown at positions 3,944,166 and 3,944,534 in *Y. pseudotuberculosis* IP32953[18]. The urease enzyme, *ureD*, that causes toxicity in fleas, is inactivated in *Y. pestis* by a G insertion in a six-G stretch, shown at position 2,997,296 in CO92. Finally, the *rcsA* gene, a component of the Rcs system that functions as an inhibitor to biofilm formation, is known to have become inactivated in *Y. pestis* by a 30 bp internal duplication, previously described in the strain KIM (NC_004088.1)[19].

In addition, as certain *Y. pestis* strains present in the closely related 0.PE4 "microtus" lineage have been previously associated to a reduced pathogenicity, genes associated to this attenuated phenotype[48] were explored in RT5, in relation to 0.PE4. The following regions were integrated into the presence/absence heatmap analysis described previously, as they are absent in microtus: YPO1986 to YPO1987, YPO2096 to YPO2135, YPO2469, YPO2487 to YPO2489 and YPO3046 to YPO3047 (region annotations are given as they appear in CO92) (Fig. 3). Additional genes, which appear to have been disrupted by IS elements, or by mutations/InDels (Supplementary Table 9), were visually inspected in IGV (Supplementary Fig. 10)[74].

References

1. 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).
2. Rasmussen, S. et al. Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
3. Andrades Valtueña, A. A. et al. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691 e3688 (2017).
4. Bos, K. I. et al. A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506–510 (2011).
5. 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).
6. Feldman, M. et al. A high-coverage yersinia pestis genome from a sixth-century Justinianic plague victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).
7. Benedictow, O. J. *The Black Death, 1346-1353: The Complete History* (Boydell & Brewer, Woodbridge, UK, 2004).
8. Spyrou, M. A. et al. Historical Y. pestis genomes reveal the European Black Death as the source of ancient and modern plague pandemics. *Cell Host Microbe* **19**, 874–881 (2016).
9. 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).
10. Pollitzer, R. *The Plague* (World Health Organization, Geneva, 1954).
11. Morelli, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).
12. Cui, Y. et al. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc. Natl Acad. Sci. USA* **110**, 577–582 (2013).
13. Anisimov, A. P., Lindler, L. E. & Pier, G. B. Intraspecific diversity of Yersinia pestis. *Clin. Microbiol. Rev.* **17**, 434–464 (2004).
14. Kislichkina, A. A. et al. Nineteen whole-genome assemblies of Yersinia pestis subsp. microtus, including representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and uleiqeica. *Genome Announc.* **3**, e01342-15 (2015).
15. Zhgenti, E. et al. Genome assemblies for 11 Yersinia pestis strains isolated in the Caucasus region. *Genome Announc.* **3**, e01030-15 (2015).
16. McNally, A., Thomson, N. R., Reuter, S. & Wren, B. W. 'Add, stir and reduce': Yersinia spp. as model bacteria for pathogen evolution. *Nat. Rev. Microbiol.* **14**, 177–190 (2016).
17. Perry, R. D. & Fetherston, J. D. Yersinia pestis—etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).
18. 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).
19. Sun, Y. C., Jarrett, C. O., Bosio, C. F. & Hinnebusch, B. J. Retracing the evolutionary path that led to flea-borne transmission of Yersinia pestis. *Cell Host Microbe* **15**, 578–586 (2014).
20. Hinnebusch, B. J. et al. Role of Yersinia murine toxin in survival of Yersinia pestis in the midgut of the flea vector. *Science* **296**, 733–735 (2002).
21. Hinnebusch, B. J., Chouikha, I. & Sun, Y. C. Ecological opportunity, evolution, and the emergence of flea-borne plague. *Infect. Immun.* **84**, 1932–1940 (2016).
22. Cunliffe, B. *Europe between the Oceans 9000 BC–AD 1000* (Yale University Press, New Haven, CT, 2008).
23. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
24. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
25. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
26. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
27. 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).
28. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
29. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
30. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
31. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
32. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
33. Mittnik, A. et al. The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9**, 442 (2018).
34. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
35. Jones, E. R. et al. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576–582 (2017).
36. Bos, K. I. et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).
37. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
38. 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).
39. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
40. Kingman, J. F. C. The coalescent. *Stoch. Process Their Appl.* **13**, 235–248 (1982).
41. Baele, G., Lemey, P. & Vansteelandt, S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinforma.* **14**, 85 (2013).
42. Gonzalez, M. D., Lichtensteiger, C. A., Caughlan, R. & Vimr, E. R. Conserved filamentous prophage in Escherichia coli O18: K1: H7 and Yersinia pestis biovar orientalis. *J. Bacteriol.* **184**, 6050–6055 (2002).
43. Derbise, A. & Carniel, E. YpfΦ: a filamentous phage acquired by Yersinia pestis. *Front. Microbiol.* **5**, 701 (2014).
44. Zimbler, D. L., Schroeder, J. A., Eddy, J. L. & Lathem, W. W. Early emergence of Yersinia pestis as a severe respiratory pathogen. *Nat. Commun.* **6**, 7487 (2015).
45. Chouikha, I. & Hinnebusch, B. J. 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 (2014).
46. Minnich, S. A. & Rohde, H. N. A rationale for repression and/or loss of motility by pathogenic Yersinia in the mammalian host. *Genus Yersinia* **603**, 298–311 (2007).
47. 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).

48. Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in Yersinia pestis. *Infect. Immun.* **77**, 2242–2250 (2009).

49. Song, Y. et al. Complete genome sequence of Yersinia pestis strain 91001, an isolate avirulent to humans. *DNA Res.* **11**, 179–197 (2004).

50. Viola, R. E., Yerman, L., Fowler, J. M., Arvidson, C. G. & Brubaker, R. R. A missense mutation causes aspartase deficiency in Yersinia pestis. *Microbiology* **154**, 1271–1280 (2008).

51. Bearden, S. W. et al. Attenuated enzootic (pestoides) isolates of Yersinia pestis express active aspartase. *Microbiology* **155**, 198–209 (2009).

52. Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E. & Simonet, M. Silencing and reactivation of urease in Yersinia pestis is determined by one G residue at a specific position in the ureD gene. *Infect. Immun.* **69**, 170–176 (2001).

53. 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).

54. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb prot5448 (2010).

55. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).

56. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).

57. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).

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

59. Huson, D. H. et al. MEGAN community edition—interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **12**, e1004957 (2016).

60. 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).

61. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).

62. Parkhill, J. et al. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413**, 523–527 (2001).

63. Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).

64. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

65. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2015).

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

67. Garcia, E. et al. Pestoides F, an atypical Yersinia pestis strain from the former Soviet Union. *Genus Yersinia* **603**, 17–22 (2007).

68. Deng, W. et al. Genome sequence of Yersinia pestis KIM. *J. Bacteriol.* **184**, 4601–4611 (2002).

69. 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).

70. 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).

71. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).

72. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comp. Biol.* **11**, e1004041 (2015).

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

74. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinforma.* **14**, 178–192 (2013).

75. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

76. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

77. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).

78. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

79. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York, 2009).

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

Acknowledgements

Author contributions

M.A.S., R.I.T., K.I.B. and J.K. designed the study; R.I.T., V.V.K., V.A.T. and A.K. provided access to human archaeological material; M.A.S. and R.I.T. performed laboratory work; M.A.S., C.C.W., A.A.V., A.K.L., D.K. and A.H. performed data analyses; and M.A.S., K.I.B. and J.K. wrote the manuscript with input from all co-authors.

Additional information

Supplementary Information

Title: Analysis of 3,800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague

Spyrou et al.

Supplementary Figures: 1-10
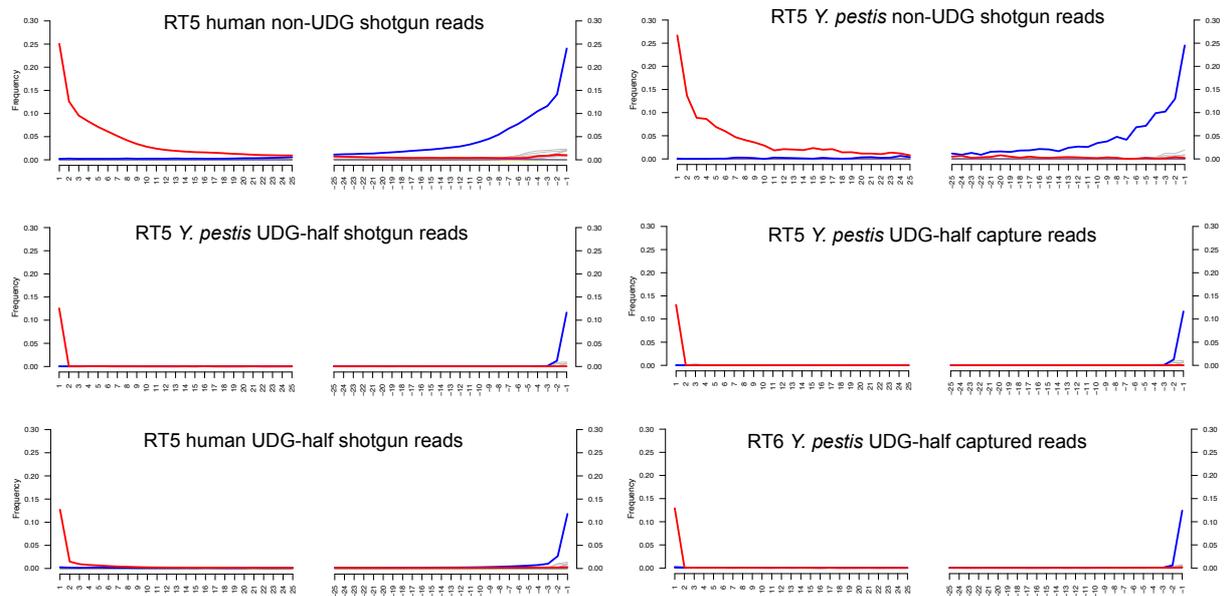
Supplementary Figure 1 – Ancient DNA damage profiles
Deamination profiles obtained from mapping of whole-genome shotgun sequenced and captured reads against the *Y. pestis* (CO92, chromosomal) and human (*hg19*, nuclear) reference genomes. Illumina double-stranded libraries were constructed using non-UDG[1] and partial UDG[2] protocols. Deamination profiles were produced using mapDamage2.0[3]. C > T changes at 5' end of fragments are depicted in red, whereas G > A substitutions at 3' end of fragments are depicted in blue.
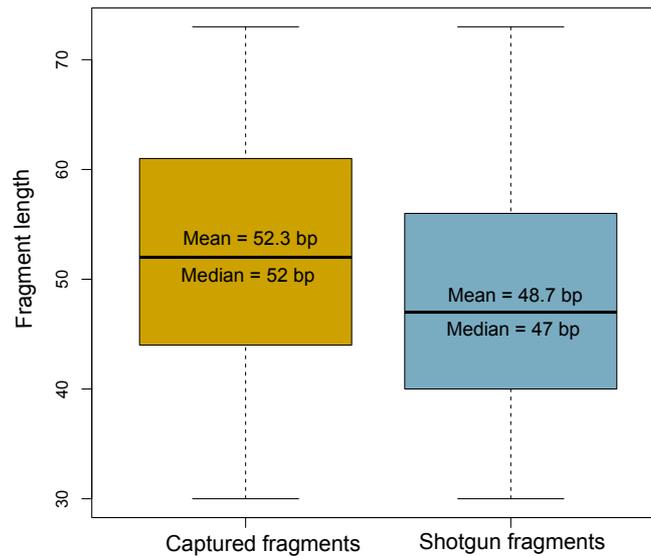
Supplementary Figure 2 – Shotgun sequenced vs captured fragment length distribution
Comparisons between RT5 *Y. pestis* fragment length distributions after deep shotgun
sequencing, and after capture. Fragments used for this analysis were sequenced as single-
ended 75 bp reads. A read length range of 30 - 74 bp was used for comparing the two
datasets, where 77.7% and 89.4% of total mapped reads were retained from the capture and
shotgun dataset, respectively. Comparison of these fractions revealed a 3.6 bp read length
increase after capture (t-test, *P*-value < 2.2e-16). The plot was produced in R version 3.4.1[4].
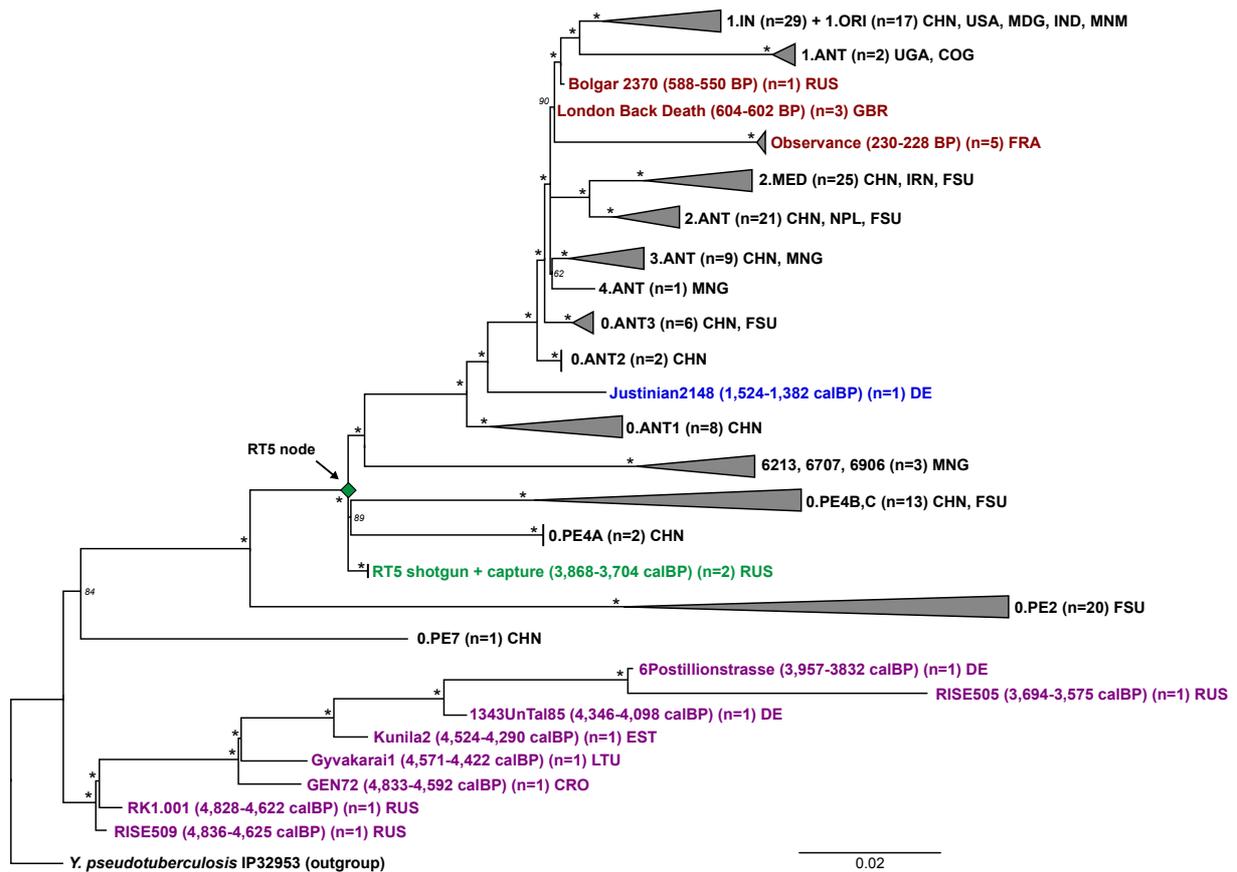
Supplementary Figure 3 – Maximum Likelihood phylogeny using all sites

The phylogenetic positioning of RT5 was assessed using a Maximum Likelihood approach and 1000 bootstrap iterations. The phylogenetic tree was produced using RaxML[5]. A worldwide dataset of *Y. pestis* chromosomal genomes (n=179) was used to reconstruct the phylogenetic tree, considering all variable positions (3,821 SNP positions). Main branches were collapsed to enhance the clarity of the phylogeny. The newly sequenced RT5 strain (green) was included in the phylogeny alongside eight Bronze Age strains belonging to the LNBA lineage (purple), a single Justinianic strain (blue), and nine second pandemic strains (red). Asterisks (*) denote bootstrap values > 95. The 2-sigma (95.4%) radiocarbon or archaeological dates of Bronze Age and historical strains are shown. Country or geographical region abbreviations are as follows: CHN (China), USA (United States of America), MDG (Madagascar), IND (India), IRN (Iran), MNM (Myanmar), RUS (Russia), GB (Great Britain), DE (Germany), FRA (France), MNG (Mongolia), NPL (Nepal), FSU (Former Soviet Union), CGO (Congo), and UGA (Uganda), LTU (Lithuania), EST (Estonia) and CRO (Croatia). See also Supplementary fig. 4 for the inferred phylogenetic positioning of RISE397.
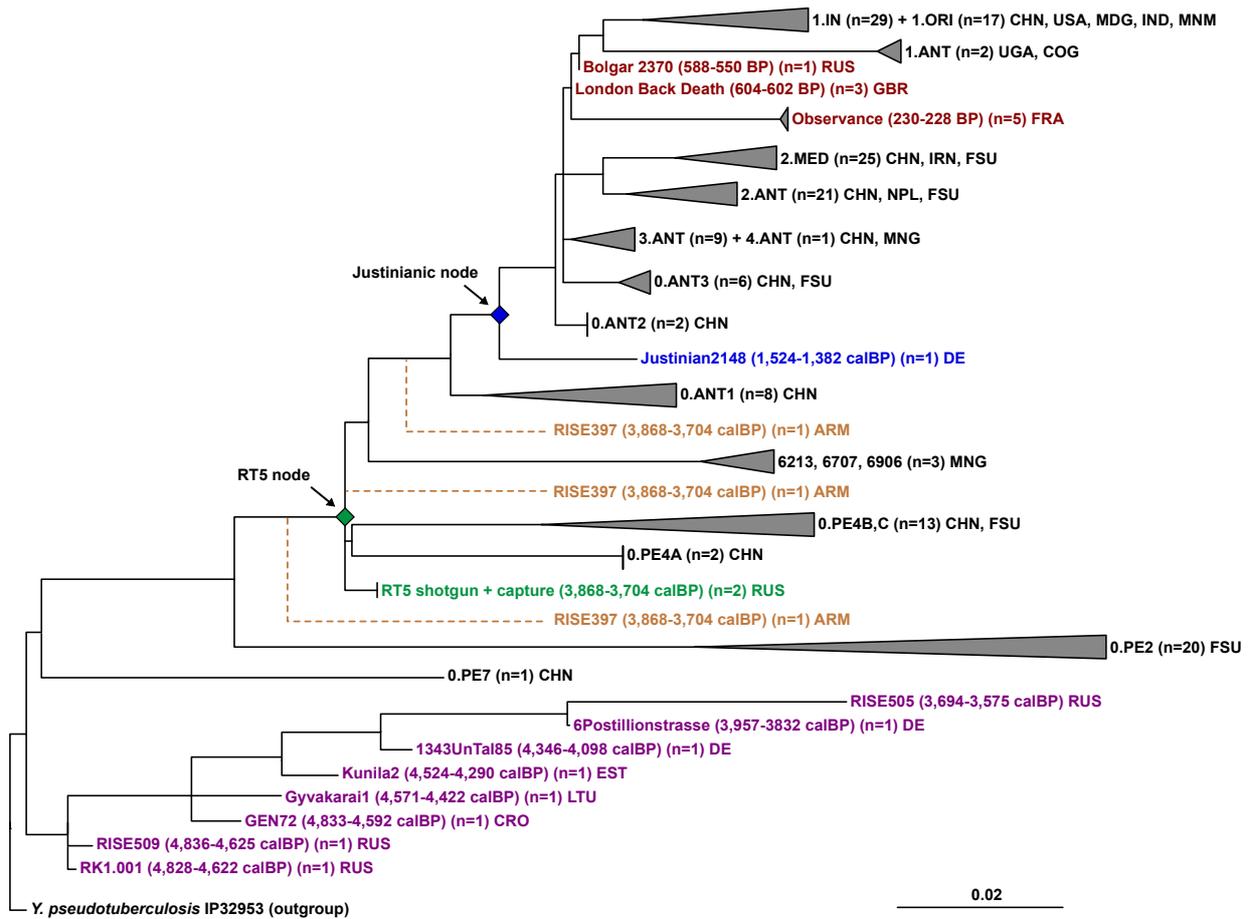
Supplementary Figure 4 – Inferred phylogenetic positioning of RISE397
The potential phylogenetic positioning of RISE397 was estimated by visual inspection of diagnostic SNPs and is illustrated here by dashed lines (see also Supplementary Data 3, 4) on the inferred Maximum Likelihood phylogeny (see Figure 2).
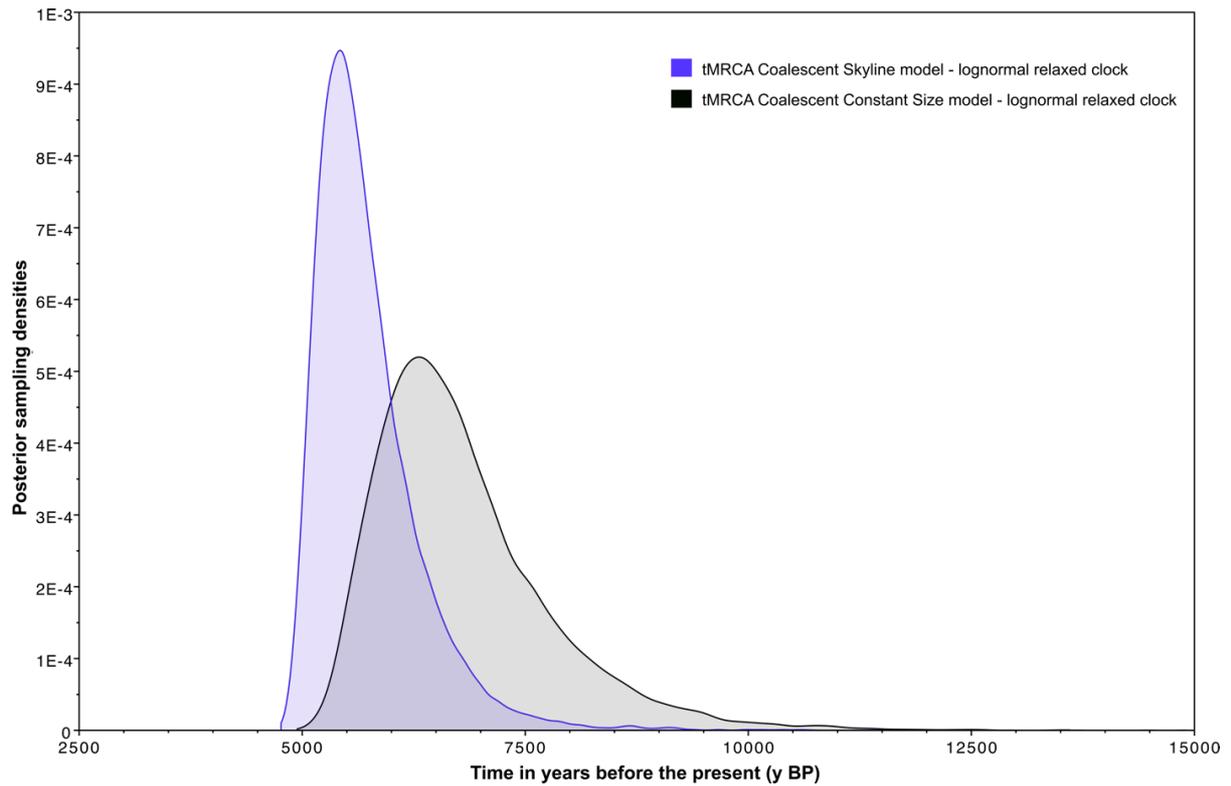
Supplementary Figure 5 – Posterior distributions of tMRCAs using Tracer v1.6
Overlapping posterior distributions of the time to the most recent common ancestor (tMRCA)
for all *Y. pestis* using the Coalescent Constant Size[6] and Coalescent Bayesian Skyline[7] models
in BEAST v1.8[8].
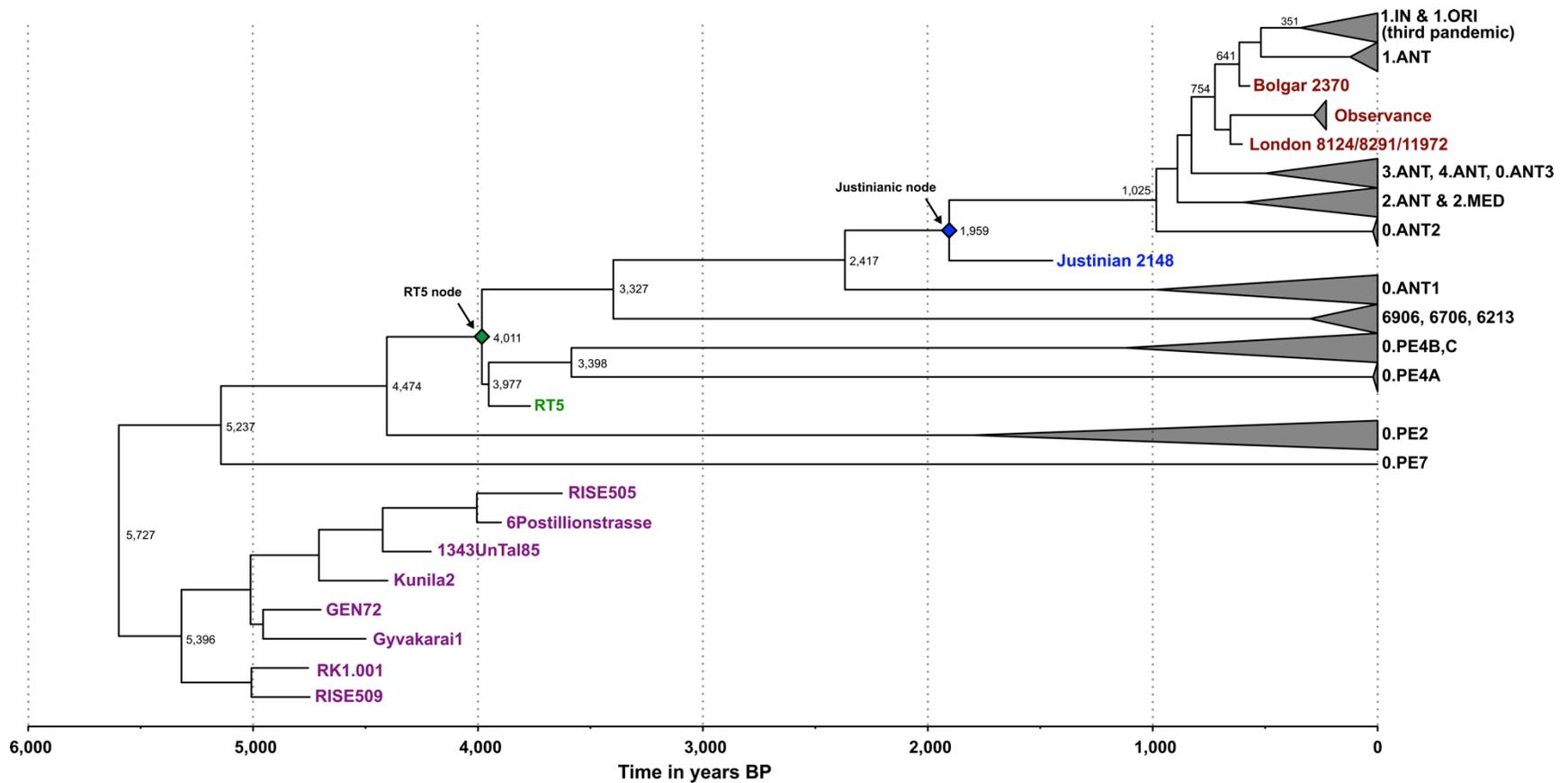
Supplementary Figure 6 - Maximum Clade Credibility tree
The MCC tree was produced using TreeAnnotator of BEAST v1.8[8] and is a product of demographic analysis based on the Coalescent Skyline model, summarizing 27,001 trees. The tree was visualized in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). It is presented in a temporal scale between 6,000 and 0 yBP, and the mean divergence dates of major *Y. pestis* lineages are indicated on each corresponding node.
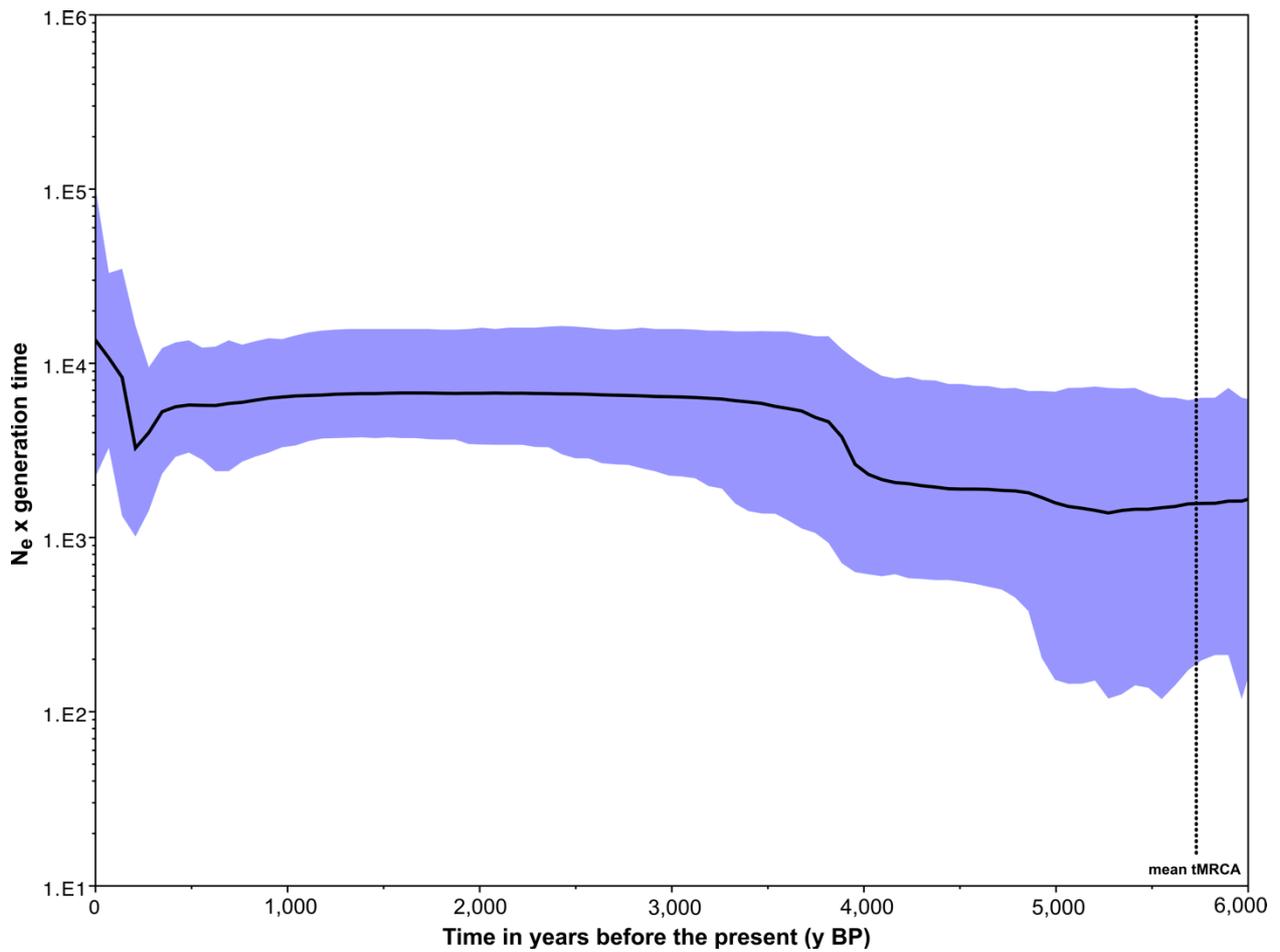
Supplementary Figure 7 – Coalescent Skyline plot
Depiction of effective population size (N_e) changes over time, as estimated by the Bayesian Coalescent Skyline model, implemented in BEAST v1.8[8]. The plot was produced in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). *Y. pestis* strains used for this analysis, overlap with ones used for the phylogenetic analysis (see Figure 2), excluding the outgroup (see Methods). The mean divergence date of all *Y. pestis* strains (5,727y BP) is shown with a dotted line.
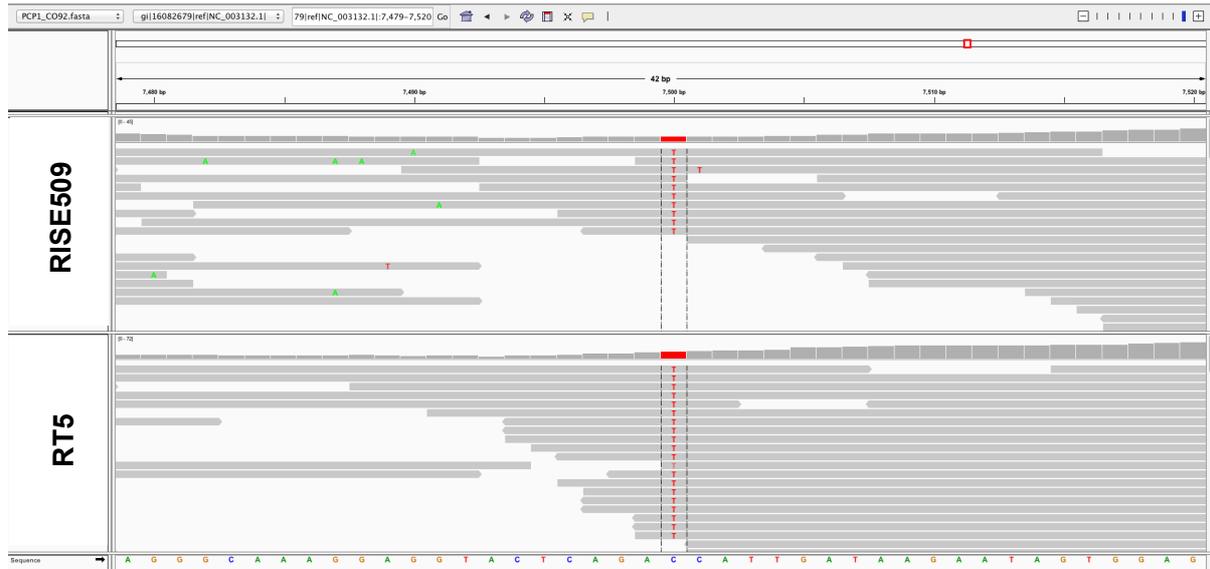
Supplementary Figure 8 – Pla genotype

RT5 and RISE509 *pla* genotype (pPCP1 plasmid), shown using an IGV[9] screenshot. The ancestral 'T' variant at position 7,500 on pPCP1 (CO92) confers an isoleucine at amino acid position 259 in *pla*[10].
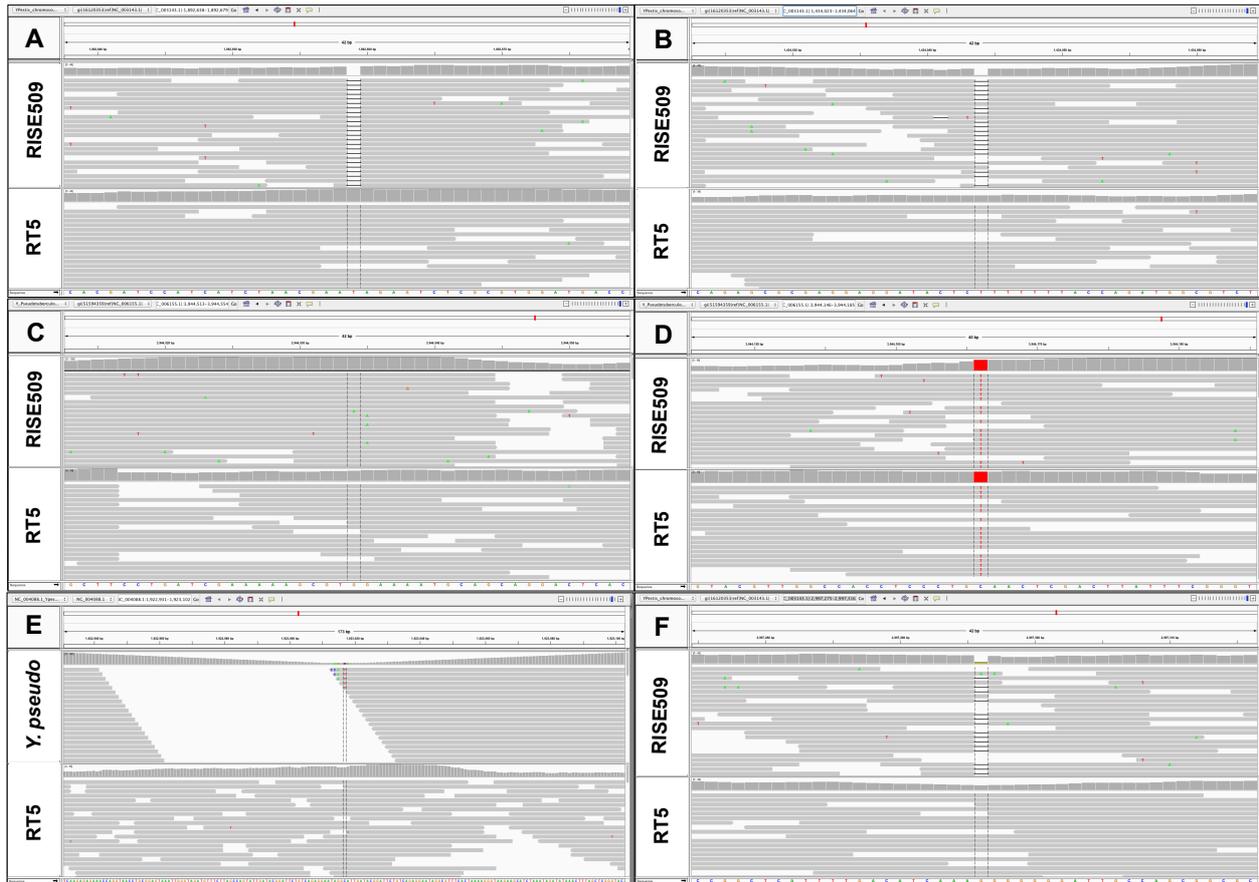
Supplementary Figure 9 – Visual inspection of virulence factors associated with *Y.pestis* immune evasion and flea adaptation. IGV[9] screenshots of virulence genes affected by substitutions, single nucleotide InDels or within gene duplications. RT5 was compared to either RISE509 or *Y. pseudotuberculosis* IP32953 for the following genes: (A) *flhD*, (B) PDE-2, (C) PDE-3, (D) PDE-3-*pe'*, (E) *rcsA*, (F) *ureD*. See main Methods section for additional information on the affected regions.
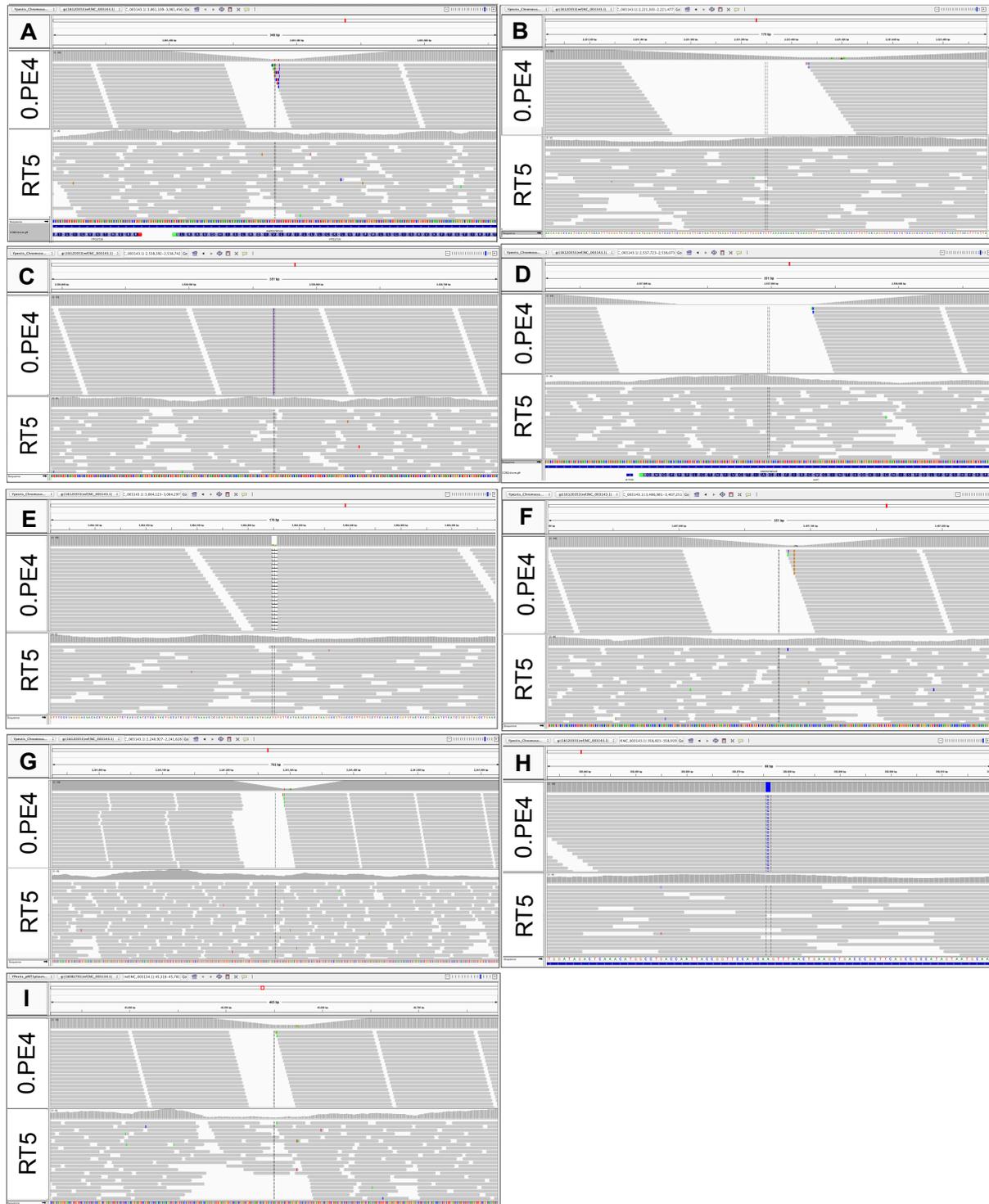
Supplementary Figure 10 – Visual inspection of 0.PE4-affected genes in RT5
The depicted genetic loci have been suggested to contribute to a decreased virulence in 0.PE4 Microtus 91001. Here we show IGV[9] screenshots of the RT5 genotype compared to that of 0.PE4 Microtus 91001 strain. The genes shown here are the ones either disrupted by IS elements, affected by small deletions, or by substitutions, in 0.PE4 Microtus 91001: (A) YPO2729 (B) YPO1956, (C) YPO2258 (G insertion) (D) YPO2258 (or *araC*) (E) YPO2731 (F) YPO3049 (or *hutC*), (G) YPO1973 (H) YPO0348 (I) YPMT1.43c. Refer to Supplementary Table 10 and Figure 3 for a description of these and other affected regions.

Supplementary Tables: 1-11

Supplementary Table 1 - *Yersinia pestis* screening statistics for all non-UDG Samara libraries. Specimens considered as putatively positive are highlighted.

Sample	Raw Reads	Pre-processed Reads	Unique Mapping Reads (CO92)	Endogenous DNA (%)	Reads assigned by MALT to *Y. pestis/Y. pseudotuberculosis* complex
RT1	7,070,200	6,862,786	10	0.000	0/0
RT2	7,760,547	7,477,994	42	0.001	0/0
RT3	6,541,880	6,229,799	38	0.001	0/37
RT4	6,255,907	6,043,051	10	0.000	0/0
RT5	7,258,101	7,003,670	7,186	0.117	1,423/7,627
RT6	7,017,656	6,706,316	82	0.001	0/71
RT7	7,928,892	7,638,467	26	0.000	0/0
RT8	5,969,436	5,766,028	34	0.001	0/35
RT9	8,215,620	7,957,649	15	0.000	0/0

Supplementary Table 2 - Human DNA screening of all non-UDG Samara libraries after mapping against *hg19*. The specimen that was pursued for further analysis is highlighted.

Sample	Raw Reads	Pre-Processed Reads	Unique Mapped Reads	Endogenous DNA (%)	Duplication Factor	Mean Fragment Length	GC content (%)
RT1	7,070,200	6,862,786	2,128,479	33.5	1.08	62.1	45.8%
RT2	7,760,547	7,477,994	17,842	0.3	1.07	60.1	43.6%
RT3	6,541,880	6,229,799	10,065	0.2	1.08	52.7	42.7%
RT4	6,255,907	6,043,051	6,687	0.1	1.07	61.7	44.9%
RT5	7,258,101	7,003,670	2,008,157	31.3	1.09	53.5	46.2%
RT6	7,017,656	6,706,316	77,968	1.2	1.07	60.5	43.0%
RT7	7,928,892	7,638,467	13,526	0.2	1.10	61.8	43.5%
RT8	5,969,436	5,766,028	17,314	0.3	1.08	52.3	46.4%
RT9	8,215,620	7,957,649	5,432	0.1	1.08	64.0	44.2%

Supplementary Table 3 - Genome reconstruction statistics after capture and deep shotgun sequencing of partially UDG-treated libraries.

Data type	Sample	Reference used for mapping	Unique Mapping Reads	Endogenous DNA (%)	Duplication Factor	Mean Coverage	Coverage across reference >= 3X	GC content (%)
Shotgun Reads	RT5	CO92 Chrom.	826,585	0.094	1.21	9.2	89.02%	48.61
		CO92 pCD1	27,688	0.005	1.79	21.4	93.45%	46.89
		CO92 pMT1	21,991	0.003	1.56	11.8	92.34%	50.85
		CO92 pPCP1	9,228	0.003	3.17	57.0	87.41%	45.85
		Human HG19	251,410,787	27.40	1.22	4.2	54.93%	47.62
Captured Reads	RT5	CO92 Chrom.	2,566,811	33.85	2.07	32.3	93.52%	48.38
		CO92 pCD1	64,303	1.738	4.23	57.6	95.03%	46.50
		CO92 pMT1	61,412	1.671	2.69	37.4	95.70%	50.50
		CO92 pPCP1	11,000	0.784	11.15	80.3	87.25%	46.21
	RT6	CO92 Chrom.	191,079	7.161	7.48	1.9	28.58%	48.82
		CO92 pCD1	7,467	0.380	10.15	5.0	55.40%	48.12
		CO92 pMT1	6,554	0.264	8.04	3.2	46.00%	50.76
		CO92 pPCP1	2,938	0.175	13.41	14.9	67.98%	49.00

Supplementary Table 4 – Results of X-chromosomal contamination estimates for RT5.

Methods	Version	MoM	SE(MoM)	ML	SE(ML)
Method1	old_llh	0.005003	6.786513e-04	0.005958	3.950846e-13
Method1	new_llh	0.004986	6.808932e-04	0.005957	2.329628e-13
Method2	old_llh	0.004085	1.209904e-03	0.004209	4.900471e-14
Method2	new_llh	0.004071	1.213368e-03	0.004206	2.082066e-13

Supplementary Table 5 - Y-chromosomal SNP positions present in RT5.

Haplogroup	SNP name	Other names	rsID	GRCh37	Ancestral	Derived	RT5	Coverage
R1a	L62	M513; PF6200	rs17222573	17,891,241	A	G	G	2
R1a	L63	M511; PF6203	rs17307677	18,162,834	T	C	C	1
R1a	L146	M420; PF6229	rs17250535	23,473,201	T	A	A	4
R1a1	Page65.2	PF6234; SRY1532.2; SRY10831.2	rs2534636	2,657,176	C	T	T	2
R1a1	M459	PF6235	rs17316227	6,906,074	A	G	G	1
R1a1	L120	M516; PF6236	rs17307105	15,879,017	A	G	G	1
R1a1a	M515	NA	rs17221601	14,054,623	T	A	A	2
R1a1a	L168	NA		16,202,177	A	G	G	3
R1a1a1	M417	NA	rs17316771	8,533,735	G	A	A	3
R1a1a1	Page7	NA	rs34297606	14,498,990	C	T	T	4
R1a1a1b	S441	Z647	rs112284571	7,683,058	G	A	A	2
R1a1a1b	S224	Z645	rs111731595	8,245,045	C	T	T	3
R1a1a1b2	F992	S202; Z93		7,552,356	G	A	A	1

Supplementary Table 6 - Radiocarbon dating results of two individuals from the Mikhailovsky II burial site.

Individual	Material	Labno. MAMS	^{14}C age [yrs BP]	Cal 2-sigma (95.4%)
RT5	Tooth	29430	3,517 ± 27	3,868-3,704 calBP
RT6	Tooth	29431	3,499 ± 25	3,842-3,696 calBP

Supplementary Table 7 - Unique *Y. pestis* SNP positions identified in individual RT5, and their respective genotype in ancient genomes included in this study.

Position	Gene /Locus	aa* change	Reference (CO92)	RT5	RT6	GEN72	RK1.001	Kunila2	Gyvakarai1	RISE505	RISE509	6Post	1343UnTal85	Justinian 2148	Black Death 8124/8291/11972
1,339,231	YPO1189	N/A	C	A	A	.	.	.	.	.	.	.	.	.	.
2,961,749	*chb*	H > T	C	T	N	.	.	.	.	.	.	.	.	.	.
3,369,465	*cysP*	F > L	G	A	A	.	.	.	.	.	.	.	.	.	.
3,620,361	*hmwA*	G > R	C	A	A	.	.	N	.	.	.	N	.	.	.
4,182,150	*purH*	A > S	G	T	T	.	.	.	.	.	.	.	.	.	.

*Amino acid

Supplementary Table 8 – Demographic model comparisons using PS/SS sampling[11] in BEASTv1.8[8]. The table shows the marginal likelihood estimates (MLE) produced for the two models by both analyses.

	Coalescent Constant Size	Coalescent Bayesian Skyline
Path sampling MLE	-9535.76	-9,526.68
Stepping stone sampling MLE	-9535.46	-9,528.37

Supplementary Table 9 - *Y. pestis* dating result comparisons between Coalescent Constant Size and Coalescent Skyline demographic models using BEASTv1.8[8]

Lineage divergence	Coalescent Constant Size[6] 95% HPD (mean y BP)	Coalescent Bayesian Skyline[7] 95% HPD (mean y BP)
MRCA[*]	6,797 (5,299-8,743)	5,727 (4,909-6,842)
0.PE7	6,225 (4,569-8,076)	5,237 (4,248-6,346)
0.PE2	4,879 (4,003-6,009)	4,474 (3,936-5,158)
RT5	4,089 (3,761-4,513)	4,011 (3,760-4,325)
Justinianic	1,925 (1,501-2,440)	1,959 (1,500-2,536)
Black Death	780 (604-1,029)	754 (603-989)

[*]MRCA here refers to the most recent common ancestor of all *Y. pestis* isolates, i.e. the root of the tree.

Supplementary Table 10 - List of regions/loci affected in 0.PE4 Microtus 91001 and considered as contributors to its decreased virulence.

Locus name in CO92	Position Start	Position End	Locus name in Mictorus 91001	Description[12,13]
YPO0348	358,526	359,962	*aspA*	C to A substitution at position 358,876 (CO92)[*]
YPO1956	2,221,367	2,221,645	YP1700	IS*285* insertion (disrupted)[*]
YPO1973	2,240,953	2,241,720	YP1715	IS*100* insertion (disrupted)[*]
YPO2258 (*araC*)	2,537,797	2,538,729	YP2054	112 bp deletion 26 bp downstream of start codon + G insertion 775 bp downstream of start codon[*]
YPO2729	3,061,203	3,061,550	YP2435	IS*285* insertion (disrupted)[*]
YPO2731	3,062,324	3,064,363	YP2433	2 bp deletion 153 bp downstream of start codon[*]
YPO3049	3,406,449	3,408,218	YP2671	7 bp deletion 1,125 bp downstream of start codon[*]
YPO1986-1987	2,254,897	2,257,189	N/A	Disrupted/Absent[**]
YPO2096-2135	2,370,557	2,403,216	N/A	Disrupted/Absent[**]
YPO2469	2,769,322	2,769,654	N/A	Disrupted/Absent[**]
YPO2487-2489	2,788,665	2,790,115	N/A	Disrupted/Absent[**]
YPO3046-3047	3,402,002	3,404,908	N/A	Disrupted/Absent[**]
YPMT1.43c (pMT1)	4,5098	45,742	pMT056	20 bp deletion 69 bp downstream of start codon[*]

[*]See Supplementary Figure 10 for visual inspection of all InDels and mutations in 0.PE4
[**]See Figure 3 of main text for loci absent in 0.PE4

Supplementary Table 11 - Newly identified SNPs that were excluded from comparative variant calling analysis in addition to previously defined regions.

Position in CO92	Reference	Variant	Description	Strains identified in
1,029,500	A	G	SNP[*]	Georgia 1413 (Zhgenti *et al.*, 2015[14])
1,029,502	A	C	SNP[*]	Georgia 1413 (Zhgenti *et al.*, 2015[14])
1,029,503	A	C	SNP[*]	Georgia 1413 (Zhgenti *et al.*, 2015[14])
1,361,705	T	C	SNP[*]	6304 (Kislichkina *et al.*,2015[15])
1,361,707	A	G	SNP[*]	6304 (Kislichkina *et al.*,2015[15])
1,361,719	G	T	SNP[*]	6304 (Kislichkina *et al.*,2015[15])
1,687,299	G	T	SNP[*]	6904 (Kislichkina *et al.*,2015[15])
1,687,300	T	C	SNP[*]	6904 (Kislichkina *et al.*,2015[15])
1,687,301	T	C	SNP[*]	6904 (Kislichkina *et al.*,2015[15])
3,489,416	G	A	SNP[*]	1.ANT1_UG05-0454[**]
3,489,419	G	A	SNP[*]	1.ANT1_UG05-0454[**]
3,489,428	G	T	SNP[*]	1.ANT1_UG05-0454[**]
3,489,429	A	T	SNP[*]	1.ANT1_UG05-0454[**]
3,860,629	A	G	SNP[*]	0.PE7b_620024[16]
3,860,637	A	C	SNP[*]	0.PE7b_620024[16]
3,860,639	A	T	SNP[*]	0.PE7b_620024[16]
4,273,931	C	A	SNP[*]	0.ANT3a_CMCC38001[16]
4,273,933	C	T	SNP[*]	0.ANT3a_CMCC38001[16]
4,273,941	A	T	SNP[*]	0.ANT3a_CMCC38001[16]
4,273,942	A	T	SNP[*]	0.ANT3a_CMCC38001[16]
4,355,693	C	A	SNP[*]	6216 (Kislichkina *et al.*,2015[15])
4,355,759	T	C	SNP[*]	6216 (Kislichkina *et al.*,2015[15])
4,355,760	A	G	SNP[*]	6216 (Kislichkina *et al.*,2015[15])
3,939,869	T	A	SNP[*]	6304 (Kislichkina *et al.*,2015[15])
3,939,870	T	C	SNP[*]	6304 (Kislichkina *et al.*,2015[15])
3,939,872	T	C	SNP[*]	6304 (Kislichkina *et al.*,2015[15])
358,876	A	C	homoplastic SNP	shared between LNBA, 0.PE4 and 2.MED3i
1,805,037	C	A	homoplastic SNP	shared between 0.PE2 and 0.PE4

[*]SNPs that appear within potentially recombining regions as defined by ClonalFrameML[17]
[**]NCBI Accession NZ_AAYR00000000 (See Supplementary Data 2)

Supplementary Methods

Archaeological context and sample information

Mikhailovsky II burial site, Samara, Russia

Between the end of the Middle Bronze Age (MBA) and the beginning of the Late Bronze Age (LBA) there is a cultural expansion observed in the Eurasian steppes that results in a series of genetically and culturally related populations extending from the Altai all the way to Europe during the LBA[18-20]. The two most widespread LBA cultures are the 'Andronovo' and 'Srubnaya' and together their subsistence economy is often described as 'complex agro-pastoralism'[21]. While the 'Andronovo' is found to occupy most of the area east of the Ural Mountains extending from the southern Urals to the Altai, the 'Srubnaya' or 'Timber grave' culture was dominant in the European steppes, west of the Ural Mountains. Here, we analyse material from the Mikhailovsky II burial site, which was excavated in 2015 and is one of numerous kurgan cemeteries identified in the Samara Oblast. It consists of seven kurgan burials, and is chronologically associated to the 'Pokrovka' phase (3,900-3,750 BP) of the 'Srubnaya' culture (3,850-3,150 BP) (radiocarbon dates produced in this study provided in Supplementary Table 6), also referred to as the 'proto-Srubnaya' that is considered the earliest phase of the LBA in the Samara Oblast. All sex and age groups were represented in this cemetery. We analysed nine individuals buried in three kurgans and identified two individuals buried in the same kurgan (see Supplementary Figure 1) to be positive for *Y. pestis*. According to anthropological analysis these were a 30-40 year-old male (RT5) and 35-45 year-old female (RT6).

Read-processing, human DNA screening and genome reconstruction

The sequenced, non-UDG, double-stranded libraries, prepared for screening from 10 μl of extract, were pre-processed using EAGER v1.92[22], for removing adaptors, length filtering (kept all reads of length > 30 bp), and quality filtering (q 20). To assess the amount of endogenous DNA in each dataset, reads where then mapped with BWA[23], which is implemented in EAGER, against the human *hg19* genome, using a seedlength of 32, –n 0.01, and a minimum mapping quality of 0.

The deep shotgun-sequenced, partial-UDG treated[2], RT5 library was pre-processed for genome reconstruction in the following way: Adaptors were removed using leeHom[24], and reads were then used as input in EAGER for subsequent length and quality filtering, as well as BWA mapping as mentioned above. Finally, 2 bp were clipped from both the 3' and 5'

ends of reads after mapping, to avoid the interfering of damaged bases with subsequent SNP calling.

RT5 genetic sex determination & contamination estimation

We performed sex identification of the deep-sequenced RT5 using two well-tested methods called Rx and Ry previously described in Skoglund et al[25] and Mittnik et al[26]. We calculated the number of reads mapping against the X and Y chromosome and compared the rates to the total number of reads mapping against the autosomes. Based on this ratio, we then decided upon a threshold to assess whether our specimen could be identified as male or female. Individual RT5 showed 4,476,135 reads mapping to the X chromosome and 382,860 reads mapping onto the Y chromosome. The Rx and Ry are calculated to be 0.449 and 0.079 respectively, which are consistent RT5 being a male.

We performed a X-chromosomal contamination test following an approach introduced by Rasmussen et al[27] and implemented in the ANGSD software suite[28]. We found a total number of 21,425 SNPs on chromosome X covered at least twice and applied the "MoM" and "ML" estimate from "Method 1" and "Method 2" likelihood computation. We determined a consistent contamination estimate of about 0.5% from different methods (Supplementary table 4).

RT5 Y-chromosomal haplogroup identification

We were able to determine the Y chromosomal haplogroup by examining a set of diagnostic positions on chromosome Y using the ISOGG database (http://isogg.org/). For this, we restricted our analysis to only include reads with a mapping quality higher than 30. We determined haplogroups by identifying the most derived Y chromosomal SNP in individual RT5. The sample has a derived allele at R1a1a1b2-F992: G→A, however only with coverage of 1-fold. As the low coverage at that position might cause a misidentification of the haplogroup, our individual was not assigned to haplogroup R1a1a1b2. We were instead able to find multiple upstream mutations assigning our individual to R1a1a1b and R1a1a1, suggesting that placement of individual RT5 in Y chromosomal haplogroup R1a1a1b is ultimately correct.

RT5 mitochondrial haplogroup analysis

We used EAGER v1.92[29] to reconstruct the RT5 mitochondrial genome and created a quality filtered (q=30) consensus sequence using the tool schmutzi[30]. Using MitoTool[31], we determined the maternal haplogroup of this individual to be U2e2a with the following variants:

73, 152, 217, 263, 508, 750, 1438, 1719, 1811, 2706, 3106A, 3720, 3849, 4553, 4736A, 4746d, 4749, 4750G, 4754-4771d, 5390, 5426, 6045, 6152, 7028, 7058, 8473, 10876, 11467, 11719, 12308, 12372, 12557, 13020, 13734, 14766, 15326, 15891, 15907, 16051, 16092, 16129C, 16183C, 16189, 16362, 16519.

1240K pulldown

We used a dataset of 1240K SNPs and pulled down variants for these alleles as previously published in Lazaridis et al.[32] and obtained a mean depth of 4.2-fold and a total coverage of 1,100,070 SNPs. On this dataset, we performed several population genetic analyses to investigate the genetic ancestry of RT5 and its relationship to modern and ancient populations from West Eurasia[32-35].

Principal Component Analysis

We used smartpca (version: 13050)[36], part of the EIGENSOFT package[37,38], to carry out Principal Component Analysis (PCA). We performed PCA on present-day West Eurasian populations and then projected the ancient samples using the lsqproject: YES option, which accounts for samples with substantial missing data. We did not perform any outlier removal iterations (numoutlieriter: 0). We set all other options to default.

ADMIXTURE analysis

We performed an ADMIXTURE[39] analysis after pruning the data for linkage disequilibrium in PLINK[40] with the parameters --indep-pairwise 200 25 0.4 retaining 298,692 SNPs for the Human Origin dataset. We ran ADMIXTURE with default 5-fold cross validation, varying the number of ancestral populations between $K = 2$ and $K = 16$ in bootstraps of 100 with different random seeds. We used 331 ancient samples as well as 2583 present-day individuals from worldwide populations for our analysis. The lowest cross-validation errors could be observed with $K=12$.

Supplementary References

1   Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).

2   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci* **370**, 20130624, doi:10.1098/rstb.2013.0624 (2015).

3   Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684, doi:10.1093/bioinformatics/btt193 (2013).

4   R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2015).

5   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).

6   Kingman, J. F. C. The coalescent. *Stoch Process Their Appl* **13**, 235-248 (1982).

7   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).

8   Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).

9   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).

10   Zimbler, D. L., Schroeder, J. A., Eddy, J. L. & Lathem, W. W. Early emergence of Yersinia pestis as a severe respiratory pathogen. *Nat. Commun.* **6**, 7487 (2015).

11   Baele, G., Lemey, P. & Vansteelandt, S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics* **14**, 85, doi:10.1186/1471-2105-14-85 (2013).

12   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).

13   Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in Yersinia pestis. *Infect. Immun.* **77**, 2242-2250, doi:10.1128/IAI.01477-08 (2009).

14   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).

15   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).

16   Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc Natl Acad Sci U S A* **110**, 577-582, doi:10.1073/pnas.1205750110 (2013).

17   Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comp. Biol.* **11**, e1004041 (2015).

18   Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, doi:10.1038/nature14317 (2015).

19   Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172, doi:10.1038/nature14507 (2015).

20   Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, doi:10.1038/nature16152 (2015).

21   Anthony, D. W. *et al.* The Samara valley project. *Late Bronze Age economy and ritual in the Russian steppes. Eurasia Antiq* **11**, 395-417 (2005).

22   Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).

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

24   Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.*, e141 (2014).

25   Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci* **40**, 4477-4482 (2013).

26   Mittnik, A., Wang, C.-C., Svoboda, J. & Krause, J. A Molecular Approach to the Sexing of the Triple Burial at the Upper Paleolithic Site of Dolní Věstonice. *PloS One* **11**, e0163019 (2016).

27   Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98 (2011).

28   Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

29   Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).

30   Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* **16**, 224, doi:10.1186/s13059-015-0776-0 (2015).

31   Fan, L. & Yao, Y.-G. An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. *Mitochondrion* **13**, 360-363 (2013).

32   Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424 (2016).

33   Mittnik, A. *et al.* The genetic prehistory of the Baltic Sea region. *Nat Commun* **9**, 442, doi:10.1038/s41467-018-02825-9 (2018).

34   Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, doi:10.1038/nature17993 (2016).

35   Jones, E. R. *et al.* The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576-582 (2017).

36   Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

37   Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909 (2006).

38   Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459-463 (2010).

39   Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).

40   Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).

Supplementary Data 1: SNP table depicting all positions variant in RT5. The table includes all ancient genomes that were used for the analysis, as well as representatives of basal *Y. pestis* lineages 0.PE7, 0.PE2 and 0.PE4. SNPs unique to the RT5 isolate are highlighted in grey

Position	CO92 Chr. (Ref)	RT5 Capture	RT5 Shotgun	GEN72	RK1.001	Kunila2	Gyvakarai1	6Post	1343UnTal85	RISE505	RISE509	0.PE7b (620024)	0.PE2 (PEST-F)	0.PE4 (Microtus91001)	Justinian2148	London BD (8124/8291/11972)	Bolgar 2370	Ellwangen	OBS107	OBS110	OBS116	OBS124	OBS137
22702	C	T	N	N	N	N	T	T	T	T	N	T	T	T	.	.	.	.	N	.	.	.	N	.
22948	C	A	A	A	A	A	A	A	A	A	A	A	A	A	.	.	.	.	N	.	.	.	.	.
74539	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
119861	C	A	N	N	N	N	.	N	N	N	N	.	.	N	.	.	.	.	.	.	.	.	.	.
119863	A	T	N	N	N	N	.	.	.	N	N	.	.	N	.	.	.	.	.	.	.	.	.	.
126977	A	G	G	N	N	G	G	G	G	N	G	G	G	G	.	.	.	.	.	.	.	.	.	.
130643	G	A	A	A	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
155747	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G	G	G
189227	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.
190049	G	A	N	N	N	N	N	N	N	N	N	N	.	.	N	N	N	N	N	N	N	N	N	N
213082	T	C	C	C	N	N	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
213227	A	T	T	T	N	T	T	T	T	N	T	T	T	T	.	.	.	.	N	.	.	.	.	.
228268	T	G	G	G	G	G	G	G	G	G	G	G	G	.	G	.	.	.	.	.	.	.	.	.
285273	C	T	N	N	N	N	N	N	N	.	.	N	N	N	N	.	.	.	.	.	.	.	.	.
286528	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	N	A	A	A	A	A	A
325836	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
341720	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
373649	T	C	C	N	N	C	N	N	C	C	C	C	C	C	.	.	.	.	N	.	.	.	.	.
399533	C	A	A	N	N	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	N	.
434257	T	G	G	G	G	N	G	G	G	G	G	G	G	G	.	.	.	.	N	.	.	.	.	.
445431	T	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
445674	T	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	N	.	.	.	.	.
483293	T	C	C	C	N	C	C	C	C	C	C	C	C	C	.	.	.	.	.	N	.	.	.	.
488947	T	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
545488	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
547131	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
699494	A	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	.	G	G	G	G	G	G	G
699647	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
870232	C	T	T	T	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.	.
877258	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	N	C	C	C	C	C	C
882354	A	C	C	C	C	N	C	C	C	C	C	C	C	C	.	.	.	.	N	.	.	.	.	.
915325	T	A	A	A	A	A	N	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.	.
918790	C	T	N	T	N	T	T	T	T	N	T	T	T	T	T	.	.	.	N	.	.	.	.	.
983851	A	G	G	G	N	N	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.
1017647	T	C	N	C	C	N	N	C	C	C	C	C	C	C	C	C	N	N	C	C	C	C	C	C
1025278	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
1051913	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
1098675	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1102174	A	G	G	G	G	N	G	G	G	G	G	G	G	G	.	.	.	N	.	.	.	.	.	.
1178178	T	C	C	C	C	N	C	N	C	C	N	C	C	C	C	C	C	N	C	C	C	C	C	C
1178459	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1251046	T	C	C	C	C	N	C	N	C	C	C	C	C	C	.	.	.	N	.	.	.	.	.	.
1263337	G	A	N	A	N	A	N	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.
1272559	T	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1285651	C	T	T	T	N	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.	.
1306718	T	C	C	C	N	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1332800	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.
1339231	C	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1374526	T	C	N	C	N	C	N	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
1385780	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1388315	T	A	A	A	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.	.
1440842	T	C	C	N	N	C	N	N	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
1446217	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	N	N	.	.	N	.
1450348	T	C	C	C	N	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
1487493	T	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
1512930	A	G	G	G	G	N	N	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G	G	G
1540754	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
1546885	T	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	N	.	.	.	.	.
1547235	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	N	.

1548654	A	G	G	G	G	G	N	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.
1599786	C	T	T	T	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.	.
1705810	A	C	N	C	N	N	N	N	C	C	C	C	C	C	C	C	.	.	N	.	.	.	.	.
1735263	A	C	C	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1749443	T	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1804559	C	T	N	T	N	T	T	T	T	T	T	T	T	T	T	.	.	N	.	.	.	.	.	.
1808946	T	C	C	C	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
1831391	A	G	G	N	N	G	N	G	G	G	G	G	G	G	G	.	.	.	N	.	.	.	.	.
1859946	T	C	C	C	C	N	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.
1861680	T	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.
1871476	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	.	.	.	.	.	.	.	.
1875298	C	T	T	T	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.	.
1932514	C	G	G	G	N	G	G	N	G	G	G	G	G	G	.	.	.	N	N	.	.	.	.	.
1956130	T	C	C	C	N	C	C	C	C	C	C	C	.	C	.	.	.	.	.	.	.	.	.	.
2012524	T	G	G	G	G	G	G	N	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
2022335	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
2077732	G	A	N	A	N	N	.	N	N	.	.	N	N	N	N	N	N	N	.	N	N	N	N	N
2098628	T	C	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
2262577	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	G	G	G	G	G	G	G
2277583	G	A	A	N	N	A	N	A	A	N	A	A	A	A	A	A	N	A	A	A	A	A	A	A
2278317	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
2281856	A	C	C	C	C	C	C	C	C	C	C	C	C	C	N	.	N	N	.	.	.	.	.	.
2300659	T	G	G	G	N	G	N	G	G	G	G	G	G	G	G	G	.	N	G	G	G	G	G	G
2356003	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2431250	T	G	G	G	N	N	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.
2476573	C	A	A	A	A	A	A	A	A	N	A	A	A	A	A	.	.	.	.	.	.	.	.	.
2508389	T	C	C	C	N	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
2575152	G	A	A	A	N	A	A	N	A	A	A	A	A	A	A	A	A	N	A	A	A	A	A	A
2655012	C	T	N	T	N	T	T	T	T	T	T	T	T	T	T	.	.	N	.	.	.	.	.	.
2656129	T	G	G	G	G	G	N	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
2684793	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
2721828	C	A	A	A	A	A	A	N	A	A	A	A	A	A	.	.	.	.	.	.	.	.	.	.
2732999	T	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.	.
2739149	C	A	N	A	N	N	A	N	N	A	A	N	A	A	A	A	A	N	A	A	A	A	A	A
2744933	A	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	N	G	N	G	G	G
2773647	A	G	G	G	N	G	G	G	G	G	G	G	G	G	G	.	.	.	.	N	.	.	.	.
2797988	A	G	G	G	N	G	G	G	N	G	G	G	G	G	G	.	.	.	N	.	.	.	.	.
2812384	G	T	T	T	T	N	T	T	T	T	T	T	T	T	T	.	.	N	.	.	.	.	.	.
2829833	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
2836077	G	A	A	A	A	N	A	N	A	A	A	N	A	A	.	.	.	N	.	.	.	N	.	.
2903882	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
2934972	C	G	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G
2936268	G	A	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
2942734	C	G	G	G	N	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
2943162	T	C	C	C	N	C	C	N	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.
2950954	G	A	A	A	N	A	N	A	A	A	A	A	A	.	A	A	A	N	A	A	A	A	A	A
2958327	C	N	T	N	N	T	T	T	T	T	N	T	T	T	T	N	T	T	T	T	T	N	T	T
2959217	T	C	C	C	C	C	N	N	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.
2961749	C	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2995771	A	G	G	G	G	N	N	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3002442	G	A	N	A	N	N	A	N	A	A	N	A	A	A	.	.	N	N	.	.	.	.	.	.
3025936	T	C	C	C	C	N	N	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.	.	.
3085079	A	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3096319	G	A	A	A	A	N	N	A	A	N	A	A	A	A	A	.	.	.	.	.	.	.	.	.
3145523	A	C	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
3190399	A	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3210101	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3223359	C	A	N	N	N	.	N	N	N	N	N	N	N	N	A	N	N	.	N	N	N	N	N	N
3244204	A	G	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3253682	A	C	C	C	C	C	C	N	C	C	C	C	C	C	.	.	.	.	N	.	.	.	.	.
3267118	A	G	G	G	N	N	G	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G	G
3286596	C	G	G	G	N	N	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.	.

Pos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
3318458	C	T	T	T	N	T	N	N	T	T	T	T	T	T	.	.	N	N	.	.	.	.	.
3323161	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3324959	A	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3341360	G	T	N	T	N	N	N	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.	.
3362591	A	G	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3369465	G	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3371160	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	N	.	.	.	.	.
3392521	C	T	T	T	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.	.
3393976	T	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3397040	A	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3407398	T	C	N	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.	.
3421335	A	G	G	G	N	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	N	G
3421553	A	G	G	N	N	G	G	G	G	N	G	G	G	G	.	.	.	.	.	.	.	.	.
3426560	A	G	G	G	N	G	G	G	G	G	G	N	G	G	G	.	.	.	.	.	.	.	.
3442617	A	T	T	T	N	T	T	N	T	T	T	T	T	T	T	T	T	N	T	T	T	T	T
3478224	A	C	C	C	N	N	C	N	C	C	C	C	C	C	.	.	.	N	.	.	.	.	.
3558289	T	C	C	C	N	C	N	N	C	N	C	C	C	C	.	.	.	.	.	.	.	.	.
3558904	T	G	G	G	G	N	G	N	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3564026	C	T	T	T	T	T	N	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
3571531	A	G	G	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3616733	A	G	G	G	N	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3620361	C	A	A	.	.	N	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3631152	A	G	N	G	N	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.
3639873	T	G	G	G	G	N	G	N	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.
3645151	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3658233	T	G	G	G	N	G	N	G	G	G	G	G	G	G	G	.	N	.	.	.	.	.	.
3659175	G	T	N	T	N	N	N	T	T	T	T	T	T	T	.	.	.	N	.	.	.	.	.
3667806	A	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3712455	T	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3726726	A	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3767613	C	T	T	N	N	T	T	T	T	T	T	T	T	T	T	.	.	.	.	.	.	.	.
3796789	A	G	G	G	N	N	N	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
3806677	C	T	N	T	N	T	N	T	T	N	T	T	T	T	T	T	.	T	T	T	T	T	T
3824488	C	T	T	T	N	T	T	T	T	N	N	T	T	T	.	.	.	.	.	.	.	.	.
3973746	C	T	T	T	T	T	N	N	T	N	T	T	T	T	T	T	T	N	T	T	T	T	T
4080579	T	C	C	C	N	C	N	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.
4081612	T	C	C	N	N	C	C	C	C	C	C	C	C	C	C	.	.	N	.	.	.	.	.
4082562	T	C	C	C	C	C	C	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
4083536	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
4087224	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.
4139729	C	T	T	T	T	T	N	T	T	N	T	T	T	T	.	.	.	N	.	.	.	.	.
4182150	G	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4185893	A	C	C	C	C	C	C	C	C	C	C	N	C	C	.	.	.	.	.	.	.	.	.
4194600	G	A	A	A	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
4243823	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
4339366	T	G	G	G	N	G	N	G	G	G	G	G	G	G	G	.	.	.	.	N	.	.	.
4357515	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	N	.	.	.	.
4371886	A	G	G	N	G	N	N	N	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
4399470	A	G	G	N	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.
4421633	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
4421689	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G
4491395	C	T	N	T	N	N	N	N	N	N	.	N	N	N	N	N	N	N	N	N	N	N	N
4518401	G	A	A	A	N	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
4527483	A	G	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G	N	G	G	G	G
4548190	A	G	G	G	G	G	G	G	G	G	G	G	G	G	.	.	.	.	.	.	.	.	.
4579183	A	G	G	G	G	G	N	G	G	N	G	G	G	G	G	G	G	G	G	G	G	G	G
4585561	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.
4589706	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	.	.	.	.	.	.	.	.
4634287	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G

Supplementary Data 2: Table of genomes used in this study for phylogenetic analysis

Isolate	Lineage	Publication (or NCBI accession)
RT5 captured	RT5	this study
RT5 shotgun sequenced	RT5	this study
Gyvakarai1	LNBA	Andrades Valtueña et al., 2017
Kunila2	LNBA	Andrades Valtueña et al., 2017
6Postillionstrasse	LNBA	Andrades Valtueña et al., 2017
1343UnTal85	LNBA	Andrades Valtueña et al., 2017
GEN72	LNBA	Andrades Valtueña et al., 2017
RISE505	LNBA	Rasmussen et al., 2015
RISE509	LNBA	Rasmussen et al., 2015
RK1.001	LNBA	Andrades Valtueña et al., 2017
Bolgar 2370	Branch 1	Spyrou et al., 2016
London 8124/8291/11972 (pooled n=3)	Branch 1	Bos et al., 2011
OBS107	Branch 1	Bos et al., 2016
OBS110	Branch 1	Bos et al., 2016
OBS116	Branch 1	Bos et al., 2016
OBS124	Branch 1	Bos et al., 2016
OBS137	Branch 1	Bos et al., 2016
Justinian 2148 Altenerding	Banch 0	Feldman et al., 2016
4.ANT1a_MGJZ12	4.ANT	Cui et al., 2013
3.ANT1a_7b	3.ANT	Cui et al., 2013
3.ANT1b_CMCC71001	3.ANT	Cui et al., 2013
3.ANT1c_C1976001	3.ANT	Cui et al., 2013
3.ANT1d_71021	3.ANT	Cui et al., 2013
3.ANT2a_MGJZ6	3.ANT	Cui et al., 2013
3.ANT2b_MGJZ7	3.ANT	Cui et al., 2013
3.ANT2c_MGJZ9	3.ANT	Cui et al., 2013
3.ANT2d_MGJZ11	3.ANT	Cui et al., 2013
3.ANT2e_MGJZ3	3.ANT	Cui et al., 2013
2.MED1b_2506	2.MED	Cui et al., 2013
2.MED1c_2654	2.MED	Cui et al., 2013
2.MED1d_2504	2.MED	Cui et al., 2013
2.MED2_KIM10	2.MED	Deng et al., 2002 (NC_004088)
2.MED2b_91	2.MED	Cui et al., 2013
2.MED2c_K11973002	2.MED	Cui et al., 2013
2.MED2d_A1973001	2.MED	Cui et al., 2013
2.MED2e_7338	2.MED	Cui et al., 2013
2.MED3a_J1963002	2.MED	Cui et al., 2013
2.MED3b_CMCC125002b	2.MED	Cui et al., 2013
2.MED3c_I1969003	2.MED	Cui et al., 2013
2.MED3d_J1978002	2.MED	Cui et al., 2013
2.MED3f_I1970005	2.MED	Cui et al., 2013
2.MED3g_CMCC99103	2.MED	Cui et al., 2013
2.MED3h_CMCC90027	2.MED	Cui et al., 2013
2.MED3i_CMCC92004	2.MED	Cui et al., 2013
2.MED3j_I2001001	2.MED	Cui et al., 2013
2.MED3k_CMCC12003	2.MED	Cui et al., 2013
2.MED3l_I1994006	2.MED	Cui et al., 2013
2.MED3m_SHAN11	2.MED	Cui et al., 2013
2.MED3n_SHAN12	2.MED	Cui et al., 2013
2.MED3o_I1991001	2.MED	Cui et al., 2013
2.MED3p_CMCC107004	2.MED	Cui et al., 2013
Azerbaijan_1045	2.MED	Zhgenti et al., 2015
RussianFederation_2944	2.MED	Zhgenti et al., 2015
2.ANT1_Nepal516	2.ANT	N/A (NZ_ACNQ00000000)
2.ANT1a_34008	2.ANT	Cui et al., 2013
2.ANT1b_34202	2.ANT	Cui et al., 2013
2.ANT2a_2	2.ANT	Cui et al., 2013
2.ANT2b_351001	2.ANT	Cui et al., 2013
2.ANT2c_CMCC347001	2.ANT	Cui et al., 2013
2.ANT2d_G1996006	2.ANT	Cui et al., 2013
2.ANT2e_G1996010	2.ANT	Cui et al., 2013
2.ANT2f_CMCC348002	2.ANT	Cui et al., 2013
2.ANT3a_CMCC92010	2.ANT	Cui et al., 2013
2.ANT3b_CMCC95001	2.ANT	Cui et al., 2013
2.ANT3c_CMCC96001	2.ANT	Cui et al., 2013
2.ANT3d_CMCC96007	2.ANT	Cui et al., 2013
2.ANT3e_CMCC67001	2.ANT	Cui et al., 2013
2.ANT3f_CMCC104003	2.ANT	Cui et al., 2013
2.ANT3g_CMCC51020	2.ANT	Cui et al., 2013
2.ANT3h_CMCC106002	2.ANT	Cui et al., 2013
2.ANT3i_CMCC64001	2.ANT	Cui et al., 2013
2.ANT3j_H1959004	2.ANT	Cui et al., 2013
2.ANT3k_5761	2.ANT	Cui et al., 2013
2.ANT3l_735	2.ANT	Cui et al., 2013
1.ORI1_CA88	1.ORI	N/A (NZ_ABCD00000000)
1.ORI1_CO92	1.ORI	Parkhill et al., 2001
1.ORI1a_CMCC114001	1.ORI	Cui et al., 2013
1.ORI1b_India195	1.ORI	N/A (NZ_ACNR00000000)
1.ORI1c_F1946001	1.ORI	Cui et al., 2013
1.ORI2_F1991016	1.ORI	N/A (NZ_ABAT00000000)
1.ORI2a_YN2179	1.ORI	Cui et al., 2013
1.ORI2c_YN2551b	1.ORI	Cui et al., 2013
1.ORI2d_YN2588	1.ORI	Cui et al., 2013
1.ORI2f_CMCC87001	1.ORI	Cui et al., 2013
1.ORI2g_F1984001	1.ORI	Cui et al., 2013
1.ORI2h_YN663	1.ORI	Cui et al., 2013
1.ORI2i_CMCCK100001a	1.ORI	Cui et al., 2013
1.ORI2i_CMCCK110001b	1.ORI	Cui et al., 2013
1.ORI3_IP275	1.ORI	N/A (NZ_AAOS00000000)
1.ORI3_MG05-1020	1.ORI	N/A (NZ_AAYS00000000)
1.ORI3a_EV76	1.ORI	Cui et al., 2013

1.IN1a_CMCC11001	1.IN	Cui et al., 2013
1.IN1b_780441	1.IN	Cui et al., 2013
1.IN1c_K21985002	1.IN	Cui et al., 2013
1.IN2a_CMCC640047	1.IN	Cui et al., 2013
1.IN2b_30017	1.IN	Cui et al., 2013
1.IN2c_CMCC31004	1.IN	Cui et al., 2013
1.IN2d_C1975003	1.IN	Cui et al., 2013
1.IN2e_C1989001	1.IN	Cui et al., 2013
1.IN2f_710317	1.IN	Cui et al., 2013
1.IN2g_CMCC05013	1.IN	Cui et al., 2013
1.IN2h_5	1.IN	Cui et al., 2013
1.IN2i_CMCC10012	1.IN	Cui et al., 2013
1.IN2j_CMCC27002	1.IN	Cui et al., 2013
1.IN2k_970754	1.IN	Cui et al., 2013
1.IN2l_D1991004	1.IN	Cui et al., 2013
1.IN2m_D1964002b	1.IN	Cui et al., 2013
1.IN2n_CMCC02041	1.IN	Cui et al., 2013
1.IN2o_CMCC03001	1.IN	Cui et al., 2013
1.IN2p_D1982001	1.IN	Cui et al., 2013
1.IN2q_D1964001	1.IN	Cui et al., 2013
1.IN3a_F1954001	1.IN	Cui et al., 2013
1.IN3b_E1979001	1.IN	Cui et al., 2013
1.IN3c_CMCC84038b	1.IN	Cui et al., 2013
1.IN3d_YN1683	1.IN	Cui et al., 2013
1.IN3e_YN472	1.IN	Cui et al., 2013
1.IN3f_YN1065	1.IN	Cui et al., 2013
1.IN3g_E1977001	1.IN	Cui et al., 2013
1.IN3h_CMCC84033	1.IN	Cui et al., 2013
1.IN3i_CMCC84046	1.IN	Cui et al., 2013
1.ANT1_Antiqua	1.ANT	Chain et al., 2006
1.ANT1_UG05-0454	1.ANT	N/A (NZ_AAYR00000000)
0.PE7b_620024	0.PE7	Cui et al., 2013
0.PE4_Microtus91001	0.PE4	Zhou et al., 2004 (NC_005810)
0.PE4Aa_12	0.PE4	Cui et al., 2013
0.PE4Ab_9	0.PE4	Cui et al., 2013
0.PE4Ba_PestoidesA	0.PE4	N/A (NZ_ACNT00000000)
0.PE4Ca_CMCCN010025	0.PE4	Cui et al., 2013
0.PE4Cc_CMCC18019	0.PE4	Cui et al., 2013
0.PE4Cd_CMCC93014	0.PE4	Cui et al., 2013
0.PE4Ce_CMCC91090	0.PE4	Cui et al., 2013
Kislichkina2015_6213	0.PE4	Kislichkina et al., 2015
Kislichkina2015_6216	0.PE4	Kislichkina et al., 2015
Kislichkina2015_6304	0.PE4	Kislichkina et al., 2015
Kislichkina2015_6706	0.PE4	Kislichkina et al., 2015
Kislichkina2015_6906	0.PE4	Kislichkina et al., 2015
Kislichkina2015_7019	0.PE4	Kislichkina et al., 2015
Kislichkina2015_7074	0.PE4	Kislichkina et al., 2015
Kislichkina2015_7075	0.PE4	Kislichkina et al., 2015
Kislichkina2015_7812	0.PE4	Kislichkina et al., 2015
M0000002	0.PE4	Cui et al., 2013
0.PE2_PEST-F	0.PE2	Garcia et al., 2007 (NC_009381)
0.PE2b_G8786	0.PE2	Cui et al., 2013
Armenia_14735	0.PE2	Zhgenti et al., 2015
Armenia_1522	0.PE2	Zhgenti et al., 2015
Georgia_1412	0.PE2	Zhgenti et al., 2015
Georgia_1413	0.PE2	Zhgenti et al., 2015
Georgia_1670	0.PE2	Zhgenti et al., 2015
Georgia_3067	0.PE2	Zhgenti et al., 2015
Georgia_3770	0.PE2	Zhgenti et al., 2015
Georgia_8787	0.PE2	Zhgenti et al., 2015
Kislichkina2015_6300	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6536	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6540	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6757	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6904	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6974	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6984	0.PE2	Kislichkina et al., 2015
Kislichkina2015_6990	0.PE2	Kislichkina et al., 2015
Kislichkina2015_7761	0.PE2	Kislichkina et al., 2015
Kislichkina2015_7832	0.PE2	Kislichkina et al., 2015
0.ANT3a_CMCC38001	0.ANT3	Cui et al., 2013
0.ANT3b_A1956001	0.ANT3	Cui et al., 2013
0.ANT3c_42082	0.ANT3	Cui et al., 2013
0.ANT3d_CMCC21106	0.ANT3	Cui et al., 2013
0.ANT3e_42091b	0.ANT3	Cui et al., 2013
Kyrgyzstan_790	0.ANT3	Zhgenti et al., 2015
0.ANT2_B42003004	0.ANT2	Eppinger et al., 2009
0.ANT2a_2330	0.ANT2	Cui et al., 2013
0.ANT1a_42013	0.ANT1	Cui et al., 2013
0.ANT1b_CMCC49003	0.ANT1	Cui et al., 2013
0.ANT1c_945	0.ANT1	Cui et al., 2013
0.ANT1d_164	0.ANT1	Cui et al., 2013
0.ANT1e_CMCC8211	0.ANT1	Cui et al., 2013
0.ANT1f_42095	0.ANT1	Cui et al., 2013
0.ANT1g_CMCC42007	0.ANT1	Cui et al., 2013
0.ANT1h_CMCC43032	0.ANT1	Cui et al., 2013
Y. pseudotuberculosis IP32953	outgroup	Chain et al., 2004

Supplementary Data 3: SNP positions leading to RT5, and their respective alleles in RISE397.Coloured rows indicate positions covered in RISE397. Positions leading to 0.PE2 are coloured in green, whereas positions falling between 0.PE2 and RT5 are coloured in blue.

Position	CO92 Chr. (Ref)	RISE397	Coverage	C > T substitutions	G > A substitutions	RT5 capture	RT5 shotgun	GEN72	RK1.001	Kunila2	Gyvakarai1	6Post	1343UnTal85	RISE505	RISE509	0.PE4 Microtus91001	0.PE2 PEST-F	0.PE7b 620024
68330	T	N				.	.	C	C	C	C	C	C	C	C	.	.	C
143776	T	N				.	.	C	C	C	C	C	C	C	C	.	C	C
350726	A	.	2	1	2	.	.	C	C	C	C	C	C	C	C	.	.	C
356921	A	N				.	.	C	C	C	C	C	C	C	C	.	.	.
472352	A	N				.	.	G	G	G	G	G	G	G	G	.	.	G
998669	T	N				.	.	C	C	C	C	C	C	C	C	.	.	C
1050790	T	N				.	.	A	A	A	A	A	A	A	A	.	.	A
1204121	A	N				.	.	C	C	C	C	C	C	C	C	.	C	C
1345963	T	N				.	.	A	A	A	A	A	A	A	A	.	A	A
1573651	T	N				.	.	C	C	C	C	C	C	C	C	.	C	C
1748616	T	N				.	.	C	C	C	C	C	C	C	C	.	C	C
1779139	T	N				.	.	G	G	G	G	G	G	G	G	.	.	G
1828930	A	.	1	0	1	.	.	G	G	G	G	G	G	G	G	.	.	G
1946279	A	N				.	.	C	C	C	C	C	C	C	C	.	.	N
1991759	A	.	1	0	0	.	.	G	G	G	G	G	G	G	G	.	G	G
1992413	G	N				.	.	T	T	T	T	T	T	T	T	.	.	T
2040939	G	N				.	.	A	A	A	A	A	A	A	A	.	.	A
2084726	A	N				.	.	C	C	C	C	C	C	C	C	.	.	C
2355850	T	N				.	.	C	C	C	C	C	C	C	C	.	C	C
2491300	T	N				.	.	C	C	C	C	C	C	C	C	.	.	C
2498949	A	.	1	2	0	.	.	G	G	G	G	G	G	G	G	.	G	G
2502743	A	N				.	.	C	C	C	C	C	C	C	C	.	.	C
2686272	A	N				.	.	C	C	C	C	C	C	C	C	.	C	C
2713925	A	.	1	0	0	.	.	G	G	G	G	G	G	G	G	.	G	G
2722524	G	N				.	.	A	A	A	A	A	A	A	A	.	A	A
2758907	A	N				.	.	G	G	G	G	G	G	G	G	.	.	G
2946695	A	N				.	.	G	G	G	G	G	G	G	G	.	.	G
2966528	T	.	1	0	1	.	.	C	C	C	C	C	C	C	C	.	.	C
3004177	T	.	1	0	0	.	.	A	A	A	A	A	A	A	A	.	.	A
3216099	A	N				.	.	C	C	C	C	C	C	C	C	.	C	C
3236038	T	.				.	.	A	A	A	A	A	A	A	A	.	.	A
3371854	C	N				.	.	T	T	T	T	T	T	T	T	.	.	T
3481364	T	N				.	.	A	A	A	A	A	A	A	A	.	.	.
3539317	A	.	2	2	0	.	.	T	T	T	T	T	T	T	T	.	.	T
3539827	A	N				.	.	T	T	T	T	T	T	T	T	.	.	T
3558196	T	.	2	2	1	.	.	C	C	C	C	C	C	C	C	.	.	N
3607537	A	N				.	.	G	G	G	G	G	G	G	G	.	G	G
3929620	A	N				.	.	G	G	G	G	G	G	G	G	.	G	G
4210272	A	N				.	.	G	G	G	G	G	G	G	G	.	G	G
4239796	G	.	1	0	0	.	.	A	A	A	A	A	A	A	A	.	.	A
4255094	C	N				.	.	T	T	T	T	T	T	T	T	.	.	T
4260577	A	.	1	0	1	.	.	G	G	G	G	G	G	G	G	.	.	G
4330728	A	N				.	.	G	G	G	G	G	G	G	G	.	G	G
4363234	A	.	1	0	0	.	.	G	G	G	G	G	G	G	G	.	.	G
4590648	A	N				.	.	G	G	G	G	G	G	G	G	.	.	G
4630269	A	N				.	.	C	C	C	C	C	C	C	C	.	.	C

Supplementary Data 4: SNP positions between RT5 and Justinian2148, and their respective alleles in RISE397. Coloured rows indicate positions covered in RISE397. Positions leading from RT5 to 0.ANT1 are marked as orange, whereas positions falling between 0.ANT1 and Justinian 2148 are marked in purple

CO92 Chr. Position	CO92 Chr. (Ref)	RISE397	Coverage	C > T substitutions	G > A substitutions	RT5 capture	RT5 shotgun	GEN72	RK1.001	Kunila2	Gyvakarai1
22948	C	N				A	A	A	A	A	A
445431	T	N				C	C	C	C	C	C
445674	T	N				C	C	C	C	C	C
488947	T	C	2	3	0	C	C	C	C	C	C
870232	C	N				T	T	T	T	T	T
1332800	A	N				G	G	G	G	G	G
1388315	T	N				A	A	A	A	A	A
1446217	A	G	3	0	3	G	G	G	G	G	G
1487493	T	C	1	0	0	C	C	C	C	C	C
1546885	T	N				C	C	C	C	C	C
1599786	C	N				T	T	T	T	T	T
1861680	T	N				G	G	G	G	G	G
1875298	C	N				T	T	T	T	T	T
2732999	T	G	1	0	0	G	G	G	G	G	G
3323161	A	G	1	1	1	G	G	G	G	G	G
3371160	A	N				G	G	G	G	G	G
3392521	C	N				T	T	T	T	T	T
3393976	T	N				G	G	G	G	G	G
3712455	T	N				G	G	G	G	G	G
4357515	A	G	1	0	0	G	G	G	G	G	G
4548190	A	N				G	G	G	G	G	G
4585561	T	C	1	0	3	C	C	C	C	C	C
4589706	T	N				C	C	C	C	C	C

TABLE CONTINUED

CO92 Chr. Position	RISE505	RISE509	6Post	1343UnTal85	Justinian 2148	0.ANT1c_945	0.ANT1f_42095	0.PE4 Microtus91001	0.PE2 PEST-F	0.PE7b 620024
22948	A	A	A	A	.	A	A	A	A	A
445431	C	C	C	C	.	.	.	C	C	C
445674	C	C	C	C	.	.	.	C	C	C
488947	C	C	C	C	.	C	C	C	C	C
870232	T	T	T	T	.	T	T	T	T	T
1332800	G	G	G	G	.	.	.	G	G	G
1388315	A	A	A	A	.	.	.	A	A	A
1446217	G	G	G	G	.	.	.	G	G	G
1487493	C	C	C	C	.	.	.	C	C	C
1546885	C	C	C	C	.	.	.	C	C	C
1599786	T	T	T	T	.	.	.	T	T	T
1861680	G	G	G	G	.	.	.	G	G	G
1875298	T	T	T	T	.	.	.	T	T	T
2732999	G	G	G	G	.	G	G	G	G	G
3323161	G	G	G	G	.	.	.	G	G	G
3371160	G	G	G	G	.	G	G	G	G	G
3392521	T	T	T	T	.	.	.	T	T	T
3393976	G	G	G	G	.	.	.	G	G	G
3712455	G	G	G	G	.	.	.	G	G	G
4357515	G	G	G	G	.	.	.	G	G	G
4548190	G	G	G	G	.	.	.	G	G	G
4585561	C	C	C	C	.	.	.	C	C	C
4589706	C	C	C	C	.	C	C	C	C	C

Ancient DNA recovery and maternal lineage diversity of early medieval Venosa in southern Italy

Maria A. Spyrou[1,2], Alessandra Sperduti[3], Åshild J. Vågene[1,2], Lorenzo M. Bondioli[4], Henrike Heyne[5], Eva Fernández-Domínguez[6], Luca Bondioli[3], Wolfgang Haak[1], Kirsten I. Bos[1], Johannes Krause[1,2]

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

[2]Institute for Archeological Sciences, University of Tuebingen, 72020, Tuebingen, Germany

[3]Servizio di Bioarcheologia, Museo delle Civiltà, 00144, Rome, Italy

[4]Department of History, Princeton University, 129 Dickinson Hall, New Jersey, USA

[5]Institute of Human Genetics, University Hospital of Leipzig, Leipzig, Germany

[6]Department of Archaeology, Durham University, South Road, Durham, UK

Abstract

The Italian peninsula has hosted a multitude of cultures and has been an important center of European history during the Roman and early medieval periods due to its accessible location in the Mediterranean region. Little is known, however, about the genetic make up of Italy during that time. Here, we perform ancient mtDNA analysis, and *Y. pestis*-specific screening, on human remains from Venosa, in southern Italy, recovered from multiple graves and radiocarbon dated to the 7th – 8th centuries CE. Our pathogen screening did not reveal a plague epidemic as the likely cause of the multiple inhumations. In addition, we retrieve complete mitochondrial genomes from all 22 individuals, which revealed a diverse population on the maternal line. We detect haplogroups that are considered rare or of low frequency in Europe today, namely L2b1a, R0a1a and N3a, but are more common in Africa and the Near East. Using pairwise distance (Fst) analysis, we were unable to detect significant differences between 7th century individuals from Venosa and any comparative modern-day European population. Our analyses suggest that the maternal lineage diversity of the 7th century CE Basilicata region may have been influenced by pre-occurring or contemporaneous Mediterranean exchange.

Introduction

The Italian peninsula has experienced continuous occupation by humans for the last 45,000 years. The earliest morphological and molecular evidence of modern humans in the region begins in the early Upper Paleolithic period[1-3], and includes an uninterrupted interval of human residency during the Last Glacial Maximum (25 kya – 19 kya), where it likely served as one of the few southern European refugia due to its temperate climate[4]. The influence of pre-historic European population turnovers on the transformations of human genetic diversity over time has been a central focus of recent ancient DNA (aDNA) studies[3,5-8]. Although it has been demonstrated that modern Europeans do, to a great extent, trace their genetic ancestry to events preceding the Late Bronze Age[9], there is increasing evidence that later historical events have also contributed to the genetic make-up of certain regions of Europe[10,11]. Scholarly attention has focused in particular on the Migrations period, which followed the fall of the Roman Empire and marked the transition from Roman to medieval times. This time is characterized by a complex history of resettlements occurring over a period of at least four centuries (5th to 9th centuries CE)[12] that is consistently portrayed in historiography as being of substantial socio-political transformation despite the absence of scholarly consensus on its demographic, ethnic, or cultural impact[13]. Since written sources are limited in providing resolution regarding the influence of historical population movements on the local genetic structure, their impact on the people of the modern-day Italian region has still to be assessed.

The population of modern-day Italy is among the best genetically characterized populations in Europe, and has been shown to harbour a geographically-linked signal[6,14,15]. However, at present little is known about the genetic composition of its population during the late antiquity and early medieval times. The periods of the Roman and Byzantine imperial rule in Italy (27 BCE –1071 CE) are of particular interest given both empires' Mediterranean dimension. In particular, given southern Italy's strategic positioning in the Mediterranean it is possible that human movements between its shores might have played a role in shaping the contemporary demographics and genetic structure of its people. Few studies have attempted to associate the modern genetic

diversity of Italy to historical demographic events[16,17], though the lack of aDNA data from relevant time periods reduces the power of their inferences.

The skeletal collection from the 'Terme di Venosa' (Area archeologica di Venosa, Potenza, Italy) offers the opportunity for an evaluation of population-level changes in the past due to its inclusion of both single and collective burials that likely represent a transect of the population, and its wide temporal range. Based on the archaeological context, the site's features encompass the 3rd century BCE and possibly extend to the 10th century CE[18]. During excavations in the late 1980's, five adjoining multiple graves were discovered, with 7-12 individuals in each. Grave artifacts were absent from the burials though a small number of coins found in the corresponding archaeological levels allowed for a rough 8th -10th century CE date to be suggested for the deposition[18]. Anthropological examination found no evidence for skeletal trauma, which excluded interpersonal violence as cause of death of the individuals. In addition, the absence of pathological changes diagnostic for infectious diseases along with the demographic profiles and depositional context led to the hypothesis that the mass graves might be associated with an acute epidemic that affected the local population of Venosa[18]. Although this region is known to have suffered several epidemic waves and famines during that time[18,19], its early medieval population has to-date only been studied osteologically.

Here, we attempt a genetic characterisation of the population of Venosa, in Basilicata, during the early medieval period. We use complete mitochondrial genomes coupled with direct radiocarbon dating to study uniparental changes in the diversity of this early medieval population of southern Italy compared to that of historical and modern-day populations from West Eurasia and Africa. In addition, we use a *Yersinia pestis*-specific plasmid capture approach[20] to assess whether the putative epidemic could be associated with plague. Our results currently do not support a plague outbreak in the population; however, our analyses reveal a diverse mtDNA pool in the Venosa population that suggests possible genetic exchange within the Mediterranean.

Results

Tooth samples and long bone fragments were collected from multiple graves unearthed in the 'Terme di Venosa' area. Four individuals were radiocarbon dated and yielded overlapping dates ranging between 650-800 cal CE (Table 1), suggesting a contemporaneous deposition of the remains. DNA extracts were generated from the pulp chambers of twenty-two teeth, each from a different individual. Extracts were converted into Illumina double-stranded libraries (no UDG treatment) and shotgun sequenced to a minimum depth of 395,809 reads. The resulting number of reads allowed us to determine the genetic sex of those individuals with a confident call for 14/22, among which eight (57%) could be identified as male and six (43%) as female (Supplementary table 1). The remaining eight individuals gave inconclusive results due to their low endogenous nuclear DNA contents (Supplementary table 1). To assess the possibility of a plague outbreak between the 7th and 8th c. AD in Venosa, we performed a capture-based enrichment for the *Y. pestis*-specific pPCP1 plasmid in all samples as previously described[20]. After NGS sequencing of the captured products to a depth of 480,979 merged paired-end reads we did not identify any DNA reads mapping to pPCP1, suggesting that our cohort was negative for the plague bacterium. In addition, we captured the mitochondrial DNA from these libraries, also using an in-solution hybridization approach[21]. After capture and NGS sequencing to a depth of 1,692,241 reads, we were able to reconstruct complete mitochondrial genomes for all 22 individuals, yielding average coverages between 11.7 and 1339.6-fold, with 21/22 genomes being covered more than 99.5% at 5-fold (Table 2). Authenticity of the ancient DNA was assessed by estimating the percentage of cytosine deamination on the 5' and 3' ends of molecules[22] and by evaluating the relative amount of present day human contamination[23]. All samples were found to harbour 25% to 35% deamination on either terminal end, thus confirming them as ancient (Supplementary table 2). Contamination analysis yielded an estimate range between 0% and 5% of modern day human contamination across all samples, a value, considered negligible given the high-coverage of our ancient mitochondrial genomes (Table 2).

Haplogroups were determined for all 22 mitochondrial genomes (Table 2), and identified 21 distinct haplotypes (Supplementary table 3), which suggests the retrieval

of a rather diverse sample of the population. The distributions of haplogroup frequencies are shown in Figure 1a. Haplogroup H displays the highest frequency within the Venosa population (50%), which matches the expected mitochondrial DNA pattern for Western Eurasia (Table 2 and Figure 1). It includes sub-haplogroups identified frequently in Europe today, such as H1, H3, H5 and H7[24-26], but also others that seem to be rare or absent (i.e. H14 and H35). In addition, other haplogroups identified here that are today rare among European populations are L2b, which is typical for African populations[27,28], and N3a, which is more common in present day Iran (1.14%), its suggested place of origin (Table 2)[29]. While L haplogroups are present at low frequencies in southern Italy today[30], haplogroup N3a is potentially quite rare and, to our knowledge, has not been described in Italian populations to date.

We further performed a Principal Component Analysis (PCA) to assess the genetic composition of the Venosa population in relation to the mitochondrial haplogroup diversity of 39 modern-day European populations, three from the Caucasus, 14 from the Near Eastern and 24 from Africa, using publicly available frequency data compiled from previous studies (Figure 1b)[5,31] and accommodating published historic datasets from a Roman (1st-6th century CE, Sagalassos) population from Turkey[32], an early medieval population from Northern Italy (6th-8th century CE, archaeologically defined as "Lombards")[33], and a Byzantine population from Southwest Turkey (11th–13th century, Sagalassos)[34]. Based on 18 common haplogroup frequencies (Figure 1) PC1 (28.8% of variance) shows a clear separation between African and non-African populations, whereas PC2 (13.3% of variance) seems to separate populations from Europe and the Near East. Early medieval Venosa appears unrelated to African populations, and falls within the Western Eurasian cline, between eastern and southern European and Near Eastern populations. Note that populations from modern-day southern Italy, here represented by mitochondrial haplogroup frequencies from Basilicata (BAS), Calabria (CAL) and Sicily (SIC), also fall within the Western Eurasian cline, near southern and eastern European populations such as those from Cyprus (CYP), Romania (ROU) and Albania (ALB), those from the Caucasus (Azerbaijan, AZE, and Georgia, GEO), and also those from regions in the Near East such as Syria (SYR) and Turkey (TUR) (Supplementary table 4). Therefore, our PCA

suggests a mitochondrial affinity of early medieval Venosa to today's Near Eastern populations from the south and eastern Mediterranean region and, to a lesser extent, to present day European groups, including an early medieval population from Northern Italy.

We performed population comparisons via computation of pairwise genetic distances (Fst values) to compare our Venosa cohort to 53 modern-day populations from Western Eurasia and Northern Africa, the two historical groups from Turkey (Roman and Byzantine periods)[32,34] and the early medieval group from Northern Italy[33] (Supplementary table 5). For maximum compatibility we restricted our analysis to 342 bp of the hypervariable sequence I (HVS I) (see Methods) for which comparative data exists and have been previously compiled from modern-day and ancient populations[5]. Our results suggest non-significant ($P > 0.05$) differences between our early medieval population and all historical and extant comparative populations from Europe (n=32) and the Caucasus (n=3) (Figure 2, Supplementary table 5). Importantly, we also find no significant Fst values when comparing Venosa to most North African and Near Eastern populations (n=21), however with the exception of modern-day Algeria ($P=0.03613$ ±0.0055), Saudi Arabia ($P=0.04785$ +/- 0.0067), and Druze ($P=0.03906$ +/- 0.0055) (Figure 2 and Supplementary table 5). Though limited only to the HVS I region of the mitochondrial genome, as well as to a sample set of 22 individuals, we detected no significant differentiation between the mitochondrial make-up of early medieval southern Italy and modern-day Europe. This analysis would either suggest the absence of a substantial population change during this 1,300-year time period, or the lack of resolution in our analysis due to data restriction.

Discussion

Despite the genomics revolution in aDNA research, mitochondrial DNA remains one of the best-described genetic markers and is still widely used in population genetics. It is generally accepted that nuclear DNA provides a more detailed resolution of fine scale population differentiation and therefore has a greater potential for uncovering human population history[10,11]. However, when it comes to substantial population turnovers, mitochondrial DNA has also proven a useful marker for detecting pronounced

population-level changes across time, and has often been the first line of genetic evidence in uncovering large-scale genetic transformations in Europe[3,5,35,36].

Here we have isolated and sequenced complete mitochondrial genomes from 22 individuals from Venosa, an early medieval site in southern Italy. This demonstrates the successful retrieval of DNA from archaeological samples that stem from the temperate Mediterranean region (Table 2 and Supplementary table 1). Such sites are often more challenging for aDNA recovery, as the rate of its decay has been previously shown to correlate inversely with increases in average temperatures[37]. The successful reconstruction of 22 mitochondrial genomes indicates that DNA preservation did not pose a major challenge in this study (Supplementary table 1). Although our *Y. pestis* screening could not support a link to a plague outbreak, future pathogen screening for a wider range of microorganisms, including bacteria, viruses and eukaryotic parasites, using metagenomic tools[38] could potentially give an improved resolution to the epidemic burial hypothesis.

In addition, we have used autosomal DNA recovered from shallow shotgun sequencing (between 395,809 and 1,400,922 reads per sample), to determine the genetic sex of these individuals, according to a previously published method[39]. We find a nearly equal distribution of males and females (Supplementary table 1), supporting the hypothesis of a non-selective mortality event, which may reflect a cross-section of the local population of Venosa between the 7th and 8th centuries CE (Table 1).

The analysis of mitochondrial genomes provides useful pilot data on the genetic diversity of the early medieval Venosa in the Basilicata region, and is potentially representative of southern Italy in general terms. We find haplogroup H (50%) to be most abundant, as is the case in modern-day Europe (Supplementary table 3). According to our PCA analysis, Venosa (VEN) falls on the Western Eurasian cline outside of the European diversity, clearly separated from a previously published early medieval population from Piedmont, Northern Italy (PDTM), and more closely related to Near Eastern groups, including Byzantine individuals from Sagalassos (Turkey) dating to 11th-13th century CE (Figure 1)[32,34]. This result likely stems from the presence of certain haplogroups within this population that are otherwise rare or are at low frequency in Europe today, but are more frequent in the Near East, namely haplogroups L2b1a (4.5%), R0a1a (4.5%) and N3a (4.5%). Macro-haplogroup L, as well as its sub-

lineage L2*, appear with highest frequency across the African continent[40], and although L lineages have been previously identified in modern-day southern Italy, their frequencies usually do not exceed 2%[28,30,41]. In addition, haplogroup R0a reaches its highest frequency in the Near East, specifically in the Arabian Peninsula (up to 18%), but has a much lower frequency (< 2%) in modern-day European populations (Supplementary table 4)[42]. Finally, we also observed haplogroup N3a, a haplogroup virtually absent in European populations and seen at low frequencies mostly in present-day Iran[43]. Although other sub-lineages of macro-haplogroup N* are relatively widespread in the Near East, this does not seem to be the case for N3 which has been infrequently described elsewhere[29]. The fact that an early medieval population from southern Italy seems to contain haplogroups identified rarely or at low frequencies in Europe today might reflect the influx of people at a contemporaneous, or at an earlier, time. As premised, population influxes to the region both pre-dated Roman times and post-dated the waning of the empire between the 8th century BCE and the 7th-8th centuries CE[44,45]. Nonetheless, it is likely that the magnitude of such movements dramatically increased in the period of Roman military expansion and imperial consolidation, with the forced relocations of vast numbers of enslaved individuals from around the Mediterranean basin and beyond into southern Italy[46-49]. In addition, this does not exclude the influence of other population movements in contributing to the regional genetic structure, such as ancient Greek colonization or population transfers in the Byzantine period[50-52]. In general, we caution that any argument would need further support from a larger scale study, with a larger sample size as well as Y chromosomal and nuclear DNA data. At our current resolution, we cannot exclude a bias caused by excess sampling of rare haplogroups on account of our small sample size.

We investigated this effect further by computing pairwise differences between Venosa and comparative modern and ancient populations from Europe, West Asia and North Africa. Acknowledging the possibility of limited resolution in this analysis, since only a fraction of the mitochondrial genome (HVS I region - 342 bp) was used here for population comparisons, we find no significant subdivision between Venosa and most comparative populations (Figure 2). In addition, we find significant Fst values upon comparison of Venosa with modern-day Algeria, Saudi Arabia, as well as the Levantine Druze (Figure 2, Supplementary table 5). The affinity of early medieval Venosa to most

populations from the Near East and North Africa might, as described earlier, be interpreted as a consequence of the close contacts that existed in the Mediterranean regions of the pre-Roman and the Byzantine periods. Instead, the higher genetic distance of Venosa to few of these regions might reflect limited genetic contact, population isolation or additional population divergence since that time. In addition, the observed close affinity of the Venosa population to all comparative Europeans may suggest the absence of a substantial addition to the mtDNA pool between the early medieval and modern times (Fst value between Venosa and modern-day Italy being 0.0020 with a *P*-value of 0.33203 ±0.0146).

The extent to which historical events contributed to the genetic structure of today's Europe is a topic best explored using several lines of evidence, including both uniparental and autosomal DNA markers, and the accommodation of contextual evidence. Although our findings are limited by the fact that mitochondrial DNA is considered to be relatively homogeneous throughout the European continent today, we provide preliminary data for investigating possible exchange between southern Italy, North Africa, and the Near East during the early medieval or preceding times. For this, our study shows that sufficient DNA can be recovered from challenging Mediterranean climates, which opens a foreground of possibilities towards analysing material from earlier time periods to determine the period(s) at which the observed genetic diversity was formed, as well as extending this analysis to the nuclear level.

Methods

Tooth sampling and extraction

Teeth from twenty-two individuals buried in the archaeological site of the "Thermae of Venosa" (Lucany, southern Italy) were sampled for our analysis[18]. Each tooth was cut open across the cementoenamel junction using a coping saw, and powdered dentine was removed from the pulp chamber using a dremel tool (Nakanishi EMAX Evolution). Approximately 50mg of the powder was used for subsequent DNA extraction of each sample, using a protocol optimised for ancient DNA[53]. Bone powder lysis was performed for 12-16h at 37°C with the addition of proteinase K (10mg/ml). Post-lysis, DNA was purified and eluted in 100μL TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20). One negative control was included for every ten samples and one

positive extraction control (cave bear) of known expected DNA yield was taken along for every extraction set.

Library preparation, pPCP1 and mitochondrial genome enrichment

We built next generation sequencing (NGS) libraries from 10 μL of extract following an established Illumina library preparation protocol optimized for double-stranded ancient DNA[54], without UDG treatment to retain the characteristic ancient DNA deamination pattern observed at the end of molecules. One negative H_2O control (water blank) was included for every ten libraries and one positive library control of known expected DNA yield was taken along for every library preparation set. In addition, a unique double-index combination was assigned to each library[55]. Indexing was performed in a 10-cycle PCR reaction using the Pfu Turbo Cx Hotstart DNA Polymerase, and indexed products were quantified by qPCR using the IS5/IS6 primer set[56]. Subsequently, indexed libraries were further amplified using the AccuPrime Pfx DNA polymerase, for different amounts of cycles according to their copy numbers up to 10^13 copies, to avoid DNA polymerase saturation and heteroduplex formation. The concentration of each amplified library was measured using an Agilent 2100 Bioanalyzer DNA 1000 chip. At this stage, an aliquot of each library was diluted to 10nM and pooled with the rest for shotgun sequencing. Subsequently, up to six libraries were pooled to equimolar ratios to achieve a total of 2ug in each pool. Pools were then used as templates for pPCP1 and mitochondrial genome enrichment, following previously described protocols[20,21]. Overnight hybridization was performed at 65°C, and the hybridised DNA pools were then eluted from the streptavidin-coated beads using 125mM NaOH, purified using MinElute spin columns (Qiagen), and quantified via qPCR. Subsequently, enriched pools were amplified using AccuPrime Pfx DNA polymerase as described above, quantified with an Agilent 2100 Bioanalyzer DNA 1000 chip, diluted, and pooled in equimolar amounts for sequencing.

Mitochondrial genome reconstruction and pPCP1 screening

Post-enriched libraries were sequenced on an Illumina HiSeq 2500 (2x100+8+8 cycles). Pre-processing and mapping of raw reads was performed using the EAGER pipeline[57] according to the following criteria: adaptors were clipped from all raw demultiplexed reads, and reads were merged when overlapping by at least 10bp using Clip&Merge. A

subsequent base quality filter (minimum score of 20) and length filter was used to remove all reads shorter than 30bp. Mapping against the pPCP1 *Y. pestis* CO92 (NC_003132.1) was performed by masking the reference positions 3,200 - 4,000 as this region was previously found to be problematic due to it's similarity to expression vectors[20]. The following parameters were used for BWA [58] mapping against pPCP1: –n 0.01, a mapping quality of 37 and a seed length of 16. In addition, initial BWA mapping against the revised Cambridge Reference Sequence (rCRS) mitochondrial genome[59] was performed using -n 0.01, mapping quality of 30 and seeding off, followed by a realignment of mapped reads using CircularMapper, a tool implemented in EAGER and optimized for circular genomes. Average coverage of the mitochondrial genome assemblies ranged between 11.7 and 1,339.6-fold.

Sex determination

Pre-enriched libraries were shotgun sequenced on an Illumina HiSeq 2500 (2x100+8+8 cycles). Pre-processing and mapping of raw reads was performed as part of the EAGER pipeline[57] as mentioned earlier. Mapping was performed using BWA[58] against the *hg19* reference genome (-n 0.01 and mapping quality of 0). Mapped reads were then used to determine sex for all individuals, using a published approach, where the X chromosome ratio (Rx) is calculated by dividing the coverage across the X chromosome(s) by the normalised coverage across the autosomes. Assuming that shotgun sequencing is random, for females the expected Rx should be around 1.0, and for males around 0.5[39].

Ancient DNA authentication, mitochondrial genome consensus generation, and haplogroup assignment

Ancient DNA authenticity was evaluated by calculating the average fragment length of unique mapped reads, and the level of deamination typical for ancient DNA using MapDamage 2.0 tool[22]. The average fragment length of mapped reads across all samples ranged between 55 and 83 bp, indeed typical for ancient DNA that is expected to be shorter than 100bp[60], and the percentage of damage observed on the 1st position of the 5' (C>T) and 3' (G>A) end of molecules ranged between 25% and 35% (Table S, Figure S). In addition, a probabilistic approach, *schmutzi*, was used to jointly estimate present day contamination in all samples, as well as to generate mitochondrial

consensus genomes stemming from authentic ancient DNA fragments[23]. Initially, *contDeam* was used to generate a contamination prior based on the ancient DNA damage, followed by an iterative approach that uses 256 non-redundant Eurasian mtDNAs to produce a final contamination estimate (*mtCont*, applied parameters – notusepredC and –uselength). In addition, *schmutzi* was used to reconstruct a consensus sequence where a quality score is assigned to each base call, accounting for the following criteria: fragment mapping quality, fragment length, coverage at every position, DNA damage, previously estimated contamination and the positioning of the base along each DNA fragment. Finally, a quality filter was applied, where bases retained in the consensus had a quality > 20. Bases with quality < 20, were excluded and an "N" was included in their respective positions. After consensus generation, HaploGrep 2.0[61], an on-line tool that combines phylogenetic information from Phylotree (here used build 17)[62,63], was used to assign mitochondrial haplogroups to each individual.

Haplogroup frequency-based analysis: PCA

A PCA plot was created on R version 3.2.1 using the prcomp function[64], to visualise and examine the relationships of haplogroup frequencies between the studied early medieval Venosa, 81 previously published modern-day populations from Western Eurasia and Africa compiled by[31] and[5], a Roman population from Turkey[32], an early medieval population from Piedmont in Italy[33], and a Byzantine population from Turkey[34]. To allow for data compatibility between all populations, we accounted for the 18 most common haplogroups within West Asia and Africa: H, HV, I, J, K, L, M1, N, N1a, N1b, R, R0, T, T1, T2, U, W and X.

Sequence based analysis: Population pairwise F_{st}

For sequence-based analysis we trimmed our Venosa dataset to 342 bp of the HVS I (nucleotide positions: 16,064-16,405) using the program Geneious[65]. We used jModeltest 2.1[66] to calculate the best suited evolutionary model for our extracted 22 HVS I regions. Based on the outcome, a Tamura & Nei model with gamma value of 0.186 was used for subsequent model based analysis. Sequences were then combined with the respective 342 bp HVS I regions of 55 modern-day West Eurasian and African

populations[5], a Roman population from Turkey[32], an early medieval population from Piedmont in Italy[33], and a Byzantine population from Turkey[34]. A pairwise distance matrix was calculated using the software Arlequin (version 3.5.1.2)[67]. In addition, the pairwise distance values between Venosa and comparative modern and historical populations were plotted onto a map using the ggplot, ggmap, maps and mapdata packages integrated in R version 3.2.1.

Acknowledgements
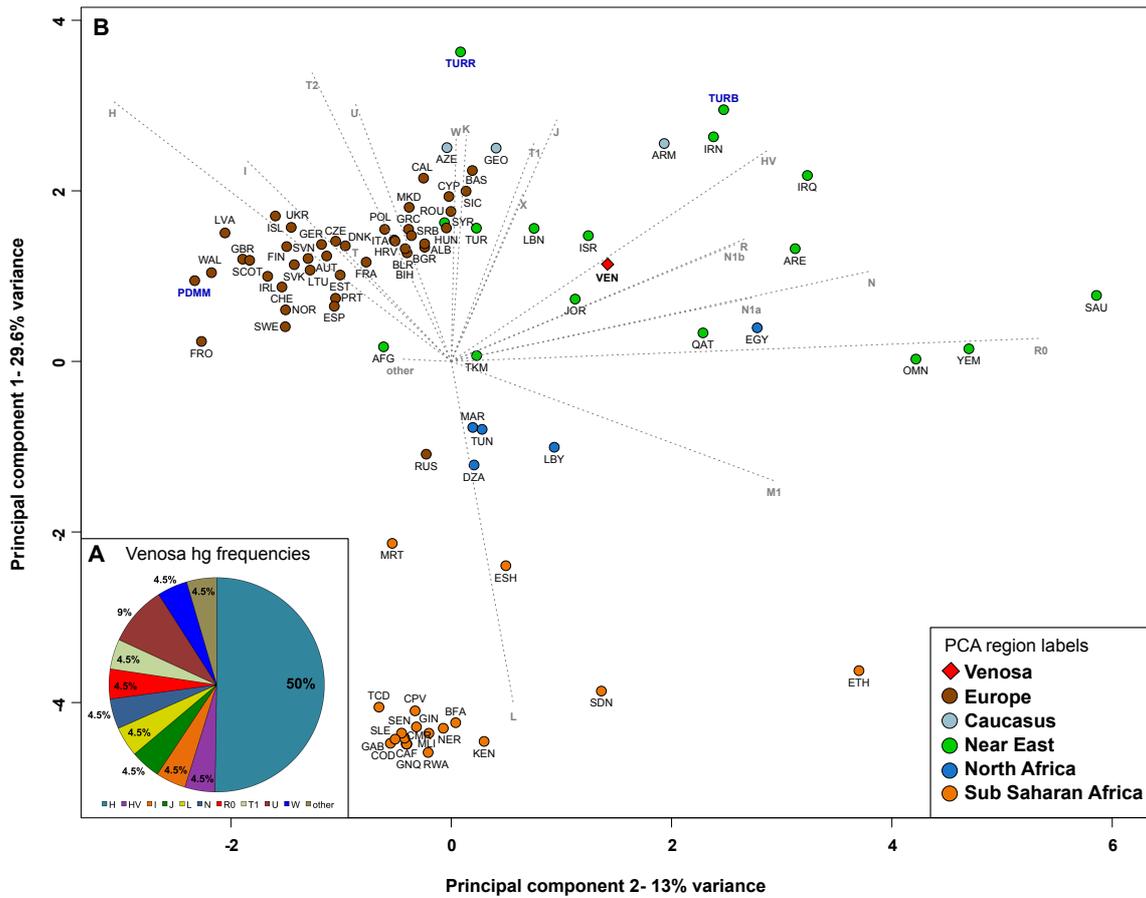
Figure 1 - (**A**) Mitochondrial haplogroup frequencies (%) of early medieval Venosa, and (**B**) PCA plot showing Venosa (VEN) relative to 81 modern-day populations from West Eurasia and Africa and three historical population from Italy and Turkey (name labels in blue) based on 18 common haplogroup frequencies (H, HV, I, J, K, L, M1, N, N1a, N1b, R, R0, T, T1, T2, U, W, X and other). For country abbreviations, refer to Supplementary table 4.
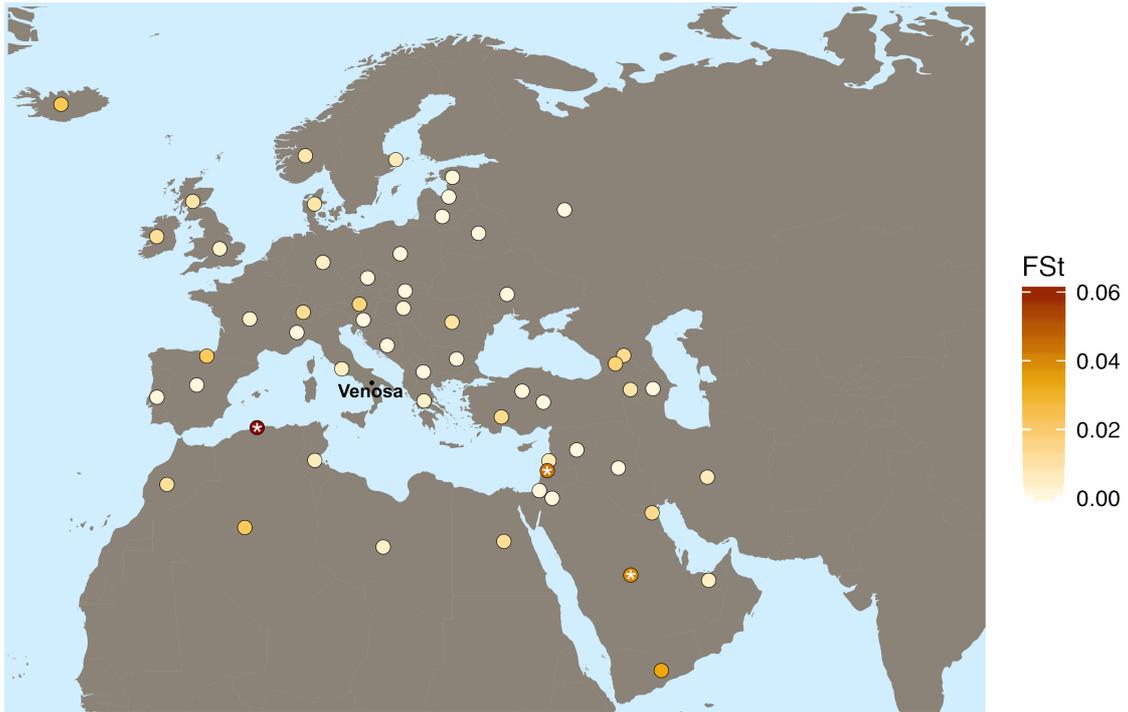
Figure 2 - Map depicting pairwise distances between Venosa and comparative historical and present-day populations from Europe, the Caucasus, the Near East and Northern Africa. Circles represent geographic location coordinates of every comparative population, and the colour coding ranges between the highest (red) and lowest (white) computed Fst distance values. A "*" is indicated in each of the circles where a significant pairwise distance was calculated ($P < 0.05$).

Table 1 - Radiocarbon dating of Venosa multiple graves. The ^{14}C ages are given in years before present (BP) and correspond to years before 1950. Calibration of dates was performed using the dataset INTCAL13[68] and the software SwissCal 1.0 (L. Wacker, ETH-Zürich).

Sample	[14]C age [yr BP]	Cal 2-sigma (95.4%)	Labno. MAMS	Material
Venosa grave 1.8	1334±23	650-763 calCE	28366	Humerus fragment
Venosa grave 3.3	1309±22	660-766 calCE	28367	Humerus fragment
Venosa grave 6.3	1263±23	670-775 calCE	28368	Femur fragment
Venosa grave 2.6	1260±22	672-800 calCE	28369	Humerus fragment

Table 2 - Summary statistics of mitochondrial genome reconstruction, contamination estimates and haplogroup assignment of the early medieval population from Venosa.

Individual	Merged paired-end reads	Unique MT Mapping reads	Duplication factor	Mean coverage	% Covered >= 5X	Average fragment length	Contamination *mtCont* (%)	Haplogroup
V. 769	428,190	136,573	1.44	619.5	100%	75.2	1-3%	N3a
V.770	280,831	29,463	1.37	133.5	99.98%	75.1	3-5%	U5a1c
V.771	98,247	11,139	1.12	51.7	99.91%	76.9	3-5%	V2
V.772	285,864	91,445	1.44	457.4	99.99%	82.9	2-4%	H1h1
V.773	102,720	19,864	1.11	93.2	99.93%	77.8	1-3%	HV0e
V.774	886,372	73,233	1.24	285.9	99.99%	64.7	1-3%	T1a
V.776	137,592	11,539	1.22	52.0	99.91%	74.7	1-3%	H1h1
V.777	1,692,241	322,197	2.70	1339.6	100%	68.9	1-3%	H
V.778	530,799	48,301	1.42	220.4	99.98%	75.6	1-3%	L2b1a
V.780	456,196	21,358	1.35	73.1	99.56%	56.7	1-3%	W6
V.781	115,806	11,509	1.18	41.6	99.56%	59.9	2-4%	R0a1a
V.782	96,515	29,205	1.27	106.5	99.97%	60.4	1-3%	J1c3
V.783	340,167	65,464	1.26	211.8	99.99%	53.6	1-3%	H14a
V.784	198,949	14,225	1.24	50.8	99.79%	59.1	1-3%	H
V.785	502,883	95,381	1.24	305.4	99.99%	53.0	0-2%	H1
V.787	303,514	22,101	1.17	74.4	99.93%	55.7	1-3%	I1b
V.788	511,462	164,051	1.39	646.3	100%	65.9	1-3%	H35
V.789	254,858	20,724	1.25	69.7	99.73%	55.7	0-2%	H5a1g1a
V.790	151,295	19,814	1.19	69.9	99.80%	58.5	1-3%	H3u
V.791	107,826	7,701	1.12	37.9	99.88%	81.5	1-3%	H5e1
V.793	454,962	74,763	1.30	257.0	99.99%	57.0	1-3%	U5a2d1
V.794	111,831	3,540	1.10	11.7	88.96%	54.8	1-3%	H7a1

References

1   Benazzi, S. *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479**, 525-528 (2011).

2   Benazzi, S. *et al.* Archaeology. The makers of the Protoaurignacian and implications for Neandertal extinction. *Science* **348**, 793-796, doi:10.1126/science.aaa2773 (2015).

3   Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**, 827-833, doi:10.1016/j.cub.2016.01.037 (2016).

4   Stewart, J. R. & Stringer, C. B. Human evolution out of Africa: the role of refugia and climate change. *Science* **335**, 1317-1321, doi:10.1126/science.1215627 (2012).

5   Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257-261, doi:10.1126/science.1241844 (2013).

6   Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).

7   Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, doi:10.1038/nature14317 (2015).

8   Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, doi:10.1038/nature17993 (2016).

9   Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, doi:10.1038/nature16152 (2015).

10   Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat Commun* **7**, 10408, doi:10.1038/ncomms10408 (2016).

11   Martiniano, R. *et al.* Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat Commun* **7**, 10326, doi:10.1038/ncomms10326 (2016).

12   Cunliffe, B. *Europe between the Oceans 9000 BC–AD 1000.* (Yale University Press, 2008).

13   Wickham, C. *Framing the Early Middle Ages: Europe and the Mediterranean 400-800.* (Oxford University Press, 2005).

14   Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98-101, doi:10.1038/nature07331 (2008).

15   Barbujani, G., Bertorelle, G., Capitani, G. & Scozzari, R. Geographical structuring in the mtDNA of Italians. *Proc Natl Acad Sci U S A* **92**, 9171-9175 (1995).

16   Boattini, A. *et al.* Traces of medieval migrations in a socially stratified population from Northern Italy. Evidence from uniparental markers and deep-rooted pedigrees. *Heredity* **114**, 155-162 (2015).

17   Sarno, S. *et al.* An ancient Mediterranean melting pot: investigating the uniparental genetic structure and population history of sicily and southern Italy. *PLoS One* **9**, e96074, doi:10.1371/journal.pone.0096074 (2014).

18   Macchiarelli, R. a. S., L. Early mediaval human skeletons from the thermae of Venosa,

Italy. Skeletal biology and life stresses in a group presumably inhumed following an epidemic. *Riv Antropol* **67**, 105–128 (1989).

19   Castex, D. in *Paleomicrobiology*   23-48 (Springer, 2008).

20   Schuenemann, V. J. *et al.* Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proc Natl Acad Sci U S A* **108**, E746-752, doi:10.1073/pnas.1105107108 (2011).

21   Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004, doi:10.1371/journal.pone.0014004 (2010).

22   Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684, doi:10.1093/bioinformatics/btt193 (2013).

23   Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* **16**, 224, doi:10.1186/s13059-015-0776-0 (2015).

24   Achilli, A. *et al.* The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *The American Journal of Human Genetics* **75**, 910-918 (2004).

25   Loogvali, E. L. *et al.* Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.* **21**, 2012-2021, doi:10.1093/molbev/msh209 (2004).

26   Roostalu, U. *et al.* Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol. Biol. Evol.* **24**, 436-448, doi:10.1093/molbev/msl173 (2007).

27   Cann, R. L., Stoneking, M. & Wilson, A. C. Mitochondrial DNA and human evolution. *Nature* **325**, 31-36, doi:10.1038/325031a0 (1987).

28   Torroni, A. *et al.* Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.* **69**, 1348-1356, doi:10.1086/324511 (2001).

29   Derenko, M. *et al.* Complete mitochondrial DNA diversity in Iranians. *PLoS One* **8**, e80673, doi:10.1371/journal.pone.0080673 (2013).

30   Ottoni, C. *et al.* Human mitochondrial DNA variation in Southern Italy. *Ann Hum Biol* **36**, 785-811, doi:10.3109/03014460903198509 (2009).

31   Brotherton, P. *et al.* Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat Commun* **4**, 1764, doi:10.1038/ncomms2656 (2013).

32   Ottoni, C. *et al.* Comparing maternal genetic variation across two millennia reveals the demographic history of an ancient human population in southwest Turkey. *R Soc Open Sci* **3**, 150250, doi:10.1098/rsos.150250 (2016).

33   Vai, S. *et al.* Genealogical relationships between early medieval and modern inhabitants of Piedmont. *PLoS One* **10**, e0116801, doi:10.1371/journal.pone.0116801 (2015).

34   Ottoni, C. *et al.* Mitochondrial analysis of a Byzantine population reveals the differential impact of multiple historical events in South Anatolia. *Eur J Hum Genet* **19**, 571-576, doi:10.1038/ejhg.2010.230 (2011).

35	Haak, W. *et al.* Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.* **8**, e1000536 (2010).

36	Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *science* **326**, 137-140 (2009).

37	Allentoft, M. E. *et al.* in *Proc. R. Soc. B.* 4724-4733 (The Royal Society).

38	Vågene, A. J. *et al.* Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol* **2**, 520-528, doi:10.1038/s41559-017-0446-6 (2018).

39	Mittnik, A., Wang, C. C., Svoboda, J. & Krause, J. A Molecular Approach to the Sexing of the Triple Burial at the Upper Paleolithic Site of Dolni Vestonice. *PLoS One* **11**, e0163019, doi:10.1371/journal.pone.0163019 (2016).

40	Silva, M. *et al.* 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Scientific reports* **5** (2015).

41	Brisighelli, F. *et al.* Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. *PLoS One* **7**, e50794 (2012).

42	Gandini, F. *et al.* Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Scientific reports* **6** (2016).

43	Kushniarevich, A. *et al.* Uniparental genetic heritage of Belarusians: encounter of rare Middle Eastern matrilineages with a Central European mitochondrial DNA pool. *PloS one* **8**, e66499 (2013).

44	Purcell, H. *The Corrupting Sea: a Study of Mediterranean History.*, (Wiley-Blackwell, 2000).

45	McCormick. *Origins of the European Economy: Communications and Commerce, A.D. 300-900.* (Cambridge University Press, 2001).

46	Laurano, A. *Peasants and Slaves: The Rural Population of Roman Italy (200 BC to AD 100).* (Cambridge University Press, 2011).

47	Scheidel, W. Human Mobility in Roman Italy, II: The Slave Population*. *The Journal of Roman Studies* **95**, 64-79 (2005).

48	Harris, W. V. Demography, Geography and the Sources of Roman Slaves. *The Journal of Roman Studies* **89**, 62-75 (1999).

49	Brunt, P. A. *Italian Manpower 225 BC-AD 14.* (Oxford University Press, 1971).

50	Burgers, G. L. M., Donnellan, L., Nizzo, V. *Contexts of Early Colonization. Acts of the conference Contextualizing Early Colonization. Archaeology, Sources, Chronology and Interpretative Models between Italy and the Mediterranean.* Vol. Vol. I. (Palombi Editori, 2016).

51	Tsetskhladze, G. R. *Greek Colonisation: An Account Of Greek Colonies and Other Settlements Overseas.* Vol. Vol.I (BRILL, 2006).

52	McCormick, M. The Imperial Edge: Italo-byzantine Identity, Movement and Integration, A.D. 650-950. *Studies on the Internal Diaspora of the Byzantine Empire*, 17-52 (1998).

53	Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* **110**, 15758-15763, doi:10.1073/pnas.1314445110 (2013).

54	Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb.prot5448, doi:10.1101/pdb.prot5448 (2010).

55   Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3, doi:10.1093/nar/gkr771 (2012).

56   Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).

57   Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).

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

59   Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147, doi:10.1038/13779 (1999).

60   Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).

61   Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58-W63 (2016).

62   van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386-394, doi:10.1002/humu.20921 (2009).

63   van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series* **5**, e392-e394 (2015).

64   Team, R. D. C. R: A Language and Environment for Statistical Computing: Vienna, Austria. (2015).

65   Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).

66   Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772, doi:10.1038/nmeth.2109 (2012).

67   Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-567, doi:10.1111/j.1755-0998.2010.02847.x (2010).

68   Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869-1887 (2013).

Supplementary table 1 Shotgun sequencing and sex determination results of 22 individuals from Venosa

Sample name	Jena IDs	Archaeological ID	Number of raw reads	Unique mapped reads	Endogenous DNA (%)	Average length (bp)	X chromosome: autosome ratio (Rx)	Rx Confidence interval (95%)	Genetic Sex
V.769	VEN001	fossa IV/V ind A	595,973	110,430	17.1	48.04	0.420335178	0.3996-0.4411	XY
V.770	VEN002	fossa 1B	1,400,922	1,268	0.1	54.55	1.210607214	1.0955-1.3257	XX
V.771	VEN003	fossa II.A	705,417	94	0.0	42.58	N/A	N/A	N/A
V.772	VEN004	fossa II	562,550	362	0.1	59.11	0.975606546	0.7151-1.2361	consistent with XX but not XY
V.773	VEN005	1.1	633,092	2,158	0.3	61.68	0.938368434	0.8571-1.0197	XX
V.774	VEN006	1.4	827,832	174,156	20.6	53.43	0.494905345	0.4701-0.5197	XY
V.778	VEN009	2.6	938,775	1,622	0.2	52.87	1.146936305	1.0704-1.2234	XX
V.780	VEN010	3.1	888,650	3,845	0.4	43.08	0.419450647	0.3842-0.4547	XY
V.781	VEN011	3.2	814,012	517	0.1	44.03	0.611791704	0.5101-0.7135	N/A
V.782	VEN012	3.3	683,250	2,342	0.3	52.71	0.419886369	0.3957-0.444	XY
V.783	VEN013	3.5	395,809	46,052	11.6	46.21	0.501155255	0.4796-0.5227	XY
V.784	VEN014	3.6	792,697	771	0.1	48.96	0.825344748	0.7505-0.9002	consistent with XX but not XY
V.785	VEN015	3.8	877,985	612,222	69.8	50.11	0.972846852	0.9285-1.0172	XX
V.787	VEN016	fossa VI.C	881,132	2,874	0.3	52.09	1.160349336	1.0675-1.2532	XX
V.788	VEN017	6.1	1,002,845	1,363	0.1	49.52	0.792656379	0.6217-0.9637	N/A
V.789	VEN018	6.2	832,916	871	0.1	46.46	0.762770504	0.6658-0.8598	N/A
V.790	VEN019	6.4	857,314	232	0.0	44.65	N/A	N/A	N/A
V.791	VEN020	6.6	730,001	349	0.0	43.94	1.034836213	0.7828-1.2869	consistent with XX but not XY
V.793	VEN021	6.8	726,206	587	0.1	47.04	0.755964368	0.651-0.8609	N/A
V.794	VEN022	6.9	872,577	240	0.0	45.13	N/A	N/A	N/A
V776	VEN007	2.4	839,500	194	0.0	47.35	N/A	N/A	N/A
V777	VEN008	2.3	1,022,918	4,203	0.4	44.46	0.567112937	0.5276-0.6066	XY

Supplementary table 2 Mitochondrial DNA deamination frequencies in ancient individuals from Venosa

Sample	C>T deamination of 1st Base 5' (%)	C>T deamination of 2nd Base 5' (%)	G>A deamination of 1st Base 3' (%)	G>A deamination of 2nd Base 3' (%)
V.769	28.3	20.9	27.9	20.9
V.770	33.7	24.8	32.6	24.2
V.771	32.2	22.8	35.4	23.4
V.772	31.7	23.6	31.1	23.7
V.773	32.0	25.4	32.3	24.6
V.774	32.9	21.8	33.1	21.4
V.776	35.1	24.0	34.9	26.0
V.777	29.2	22.2	29.0	21.9
V.778	34.2	24.2	33.6	24.3
V.780	31.8	22.6	31.0	23.0
V.781	29.5	21.9	29.5	20.5
V.782	28.8	23.2	29.5	21.5
V.783	34.5	21.1	33.8	20.7
V.784	32.0	25.0	33.7	22.8
V.785	32.4	20.1	33.1	20.5
V.787	30.2	19.6	30.0	20.6
V.788	25.6	19.1	24.8	18.3
V.789	31.1	22.7	31.4	22.3
V.790	32.0	23.9	32.0	23.9
V.791	31.9	23.6	32.5	23.5
V.793	30.6	21.5	31.7	22.8
V.794	35.5	24.1	34.0	23.5

Supplementary table 3 Haplotypes for all mitochondrial genomes retrieved from early Medieval Venosa, as retrieved from Haplogrep.

Sample ID	Haplogroup	Quality	Polymorphisms																		
V.769	N3a	0.9317	73G	146C	210G	263G	571T	750G	1005C	1438G	2706G	3107C	3357A	4769G	5048C	6366A	6806G	7028T	7897A	8653G	8860G
V.770	U5a1c	0.8945	73G	152C	263G	310C	750G	1438G	2706G	3107C	3197C	4769G	6752G	7028T	7080C	9477A	11467G	11719A	12308G	12372A	13617C
V.771	V2	0.9678	72C	263G	750G	1438G	2706G	4580A	4769G	7028T	8860G	13105G	14587G	15326G	15904T	16298C	16311C				
V.772	H1h1	0.8749	263G	310C	750G	942G	1438G	1763G	3010A	3107C	4769G	7013A	8860G	11914A	15326G	16519C					
V.773	HV0e	0.9035	72C	195C	263G	310C	750G	1438G	2706G	4769G	7028T	7600A	7705C	8860G	10609C	15326G	15454C	16298C	16311C		
V.774	T1a	0.945	73G	263G	310C	709A	750G	1438G	1888A	2706G	3107C	4216C	4769G	4917G	7028T	8697A	8701G	8860G	10463C	11251G	11719A
V.776	H1h1	0.8749	263G	310C	750G	942G	1438G	1763G	3010A	4769G	7013A	8860G	11914A	15326G	16519C						
V.777	H	0.8225	263G	310C	750G	1438G	3107C	4769G	8860G	9804A	15326G	15670C	16278T	16293C	16519C						
V.778	L2b1a	0.9521	73G	143A	150T	152C	195C	197G	198T	204C	263G	315.1C	418T	750G	769A	1018A	1438G	1442A	1706T	2332T	2358G
V.780	W6	0.8987	73G	263G	315.1C	709A	750G	1243C	1438G	2706G	3505G	4093G	4769G	5046A	5460A	7028T	8251A	8271G	8614C	8860G	8994A
V.781	R0a1a	0.9581	58C	64T	146C	263G	310C	750G	827G	1438G	2442C	2706G	3107C	3847C	4769G	6617T	7028T	8292A	8860G	11761T	13188T
V.782	J1c3	1	73G	185A	228A	263G	295T	315.1C	462T	489C	750G	1438G	2706G	3010A	3107C	4216C	4769G	7028T	8860G	10398G	11251G
V.783	H14a	0.8253	263G	310C	750G	1438G	4769G	4973C	5495C	7645C	8860G	10217G	12067T	12192A	14971C	15326G	15380G	15427G	16256T	16352C	
V.784	H	0.8196	263G	310C	368G	750G	1007A	1438G	3107C	4769G	8860G	14281T	15326G	16519C							
V.785	H1	1	263G	315.1C	750G	1438G	3010A	3106A	3107C	4769G	8860G	10049G	15326G	16519C							
V.787	I1b	0.9476	73G	199C	204C	250C	263G	310C	455.1T	750G	1438G	1719A	2706G	3645C	4529T	4769G	6227C	6734A	7028T	7637A	8251A
V.788	H35	0.9214	263G	310C	750G	1438G	3107C	3342T	4769G	5054A	8860G	15326G	16519C								
V.789	H5a1g1a	1	263G	315.1C	444G	456T	750G	1284C	1438G	3107C	4336C	4769G	7517G	8860G	9804A	15326G	15833T	16172C	16304C	16311C	
V.790	H3u	0.9152	263G	315.1C	750G	1438G	4769G	6776G	8860G	9966A	14577C	15326G	16240G	16519C							
V.791	H5e1	0.9543	263G	309.1C	315.1C	456T	593G	750G	1438G	4769G	8343G	8610C	8860G	12771A	15326G	16294T	16304C				
V.793	U5a2d1	0.9386	73G	263G	310C	549T	750G	1438G	2706G	3107C	3197C	3750T	4769G	5964C	7028T	7843G	7978T	8104C	8860G	9477A	11107T
V.794	H7a1	0.891	263G	310C	750G	1438G	1719A	4769G	4793G	7079T	8860G	14384A	15326G	16261T	16344T	16519C					

Table continued

8937C	9815T	10586A	11128G	11719A	12705T	13692T	14634C	14766T	15326G	16086C	16172C	16174T	16187T	16189C	16217C	16223T	
14274G	14766T	14793G	15218G	15326G	16180G	16192T	16256T	16270T	16320T	16399G							
12633A	13368A	14766T	14905A	15310C	15326G	15452A	15607G	15928A	16126C	16163G	16186T	16189C	16234T	16294T	16519C		
2416C	2706G	3107C	3594T	4104G	4158G	4370C	4767C	4769G	5027T	5331A	5814C	5900.1T	6026A	6713T	7028T	7119A	7256T	7521A	7624A	8080T
11674T	11719A	11947G	12414C	12705T	14766T	15326G	15884C	16223T	16292T	16325C	16519C						
13884G	14766T	15326G	16126C	16355T	16362C												
11719A	12612G	13708A	13934T	14766T	14798C	15326G	15452A	16069T	16126C								
8860G	9966A	10034C	10238C	10398G	11719A	12501A	12705T	13780G	14766T	15043A	15326G	15924G	16129A	16178C	16223T	16311C	16391A	16519C
11467G	11719A	12308G	12372A	13617C	14766T	14793G	15326G	16256T	16270T	16278T	16304C	16526A					

Table continued

8206A	8387A	8701G	8854A	8860G	9221G	9540C	10115C	10398G	10828C	10873C	11719A	11944C	12236A	12705T	12948G	13590A	13650T	13924T	14059G	14569A

Table continued

14766T	15110A	15217A	15301A	15326G	16114A	16129A	16213A	16223T	16278T	16294T	16355T	16362C	16390A

SNPs considerd as uncertain through manual inspection

Supplementary table 4 PCA input frequency table of 18 most common haplogroups across Eurasia. Frequencies of the early medieval population from Venosa are highlighted in grey.

Code	Country	Region	H	HV	I	J	K	L	M1	N	N1a	N1b	R	R0	T	T1	T2	U	W	X	other
EGY	Egypt	North Africa	0.1480	0.0530	0.0200	0.0790	0.0500	0.2320	0.0730	0.0010	0.0110	0.0300	0.0100	0.0420	0.0060	0.0500	0.0590	0.1020	0.0070	0.0160	0.0000
MAR	Morocco	North Africa	0.2990	0.0390	0.0000	0.0540	0.0470	0.2130	0.0350	0.0010	0.0000	0.0040	0.0000	0.0180	0.0040	0.0120	0.0400	0.1460	0.0060	0.0220	0.0820
LBY	Libya	North Africa	0.3150	0.0600	0.0030	0.0650	0.0350	0.2950	0.0280	0.0030	0.0030	0.0050	0.0000	0.0400	0.0000	0.0100	0.0300	0.0810	0.0000	0.0150	0.8600
TUN	Tunisia	North Africa	0.2830	0.0330	0.0030	0.0450	0.0300	0.2870	0.0270	0.0000	0.0000	0.0120	0.0010	0.0170	0.0090	0.0510	0.0260	0.1080	0.0010	0.0120	0.0560
DZA	Ageria	North Africa	0.2360	0.0080	0.0000	0.0710	0.0160	0.1730	0.0790	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0390	0.0080	0.2990	0.0000	0.0080	0.0450
TURB	Turkey Roman	Near East	0.1890	0.0590	0.0000	0.1180	0.0980	0.0000	0.0000	0.0000	0.0000	0.1130	0.0000	0.0590	0.0000	0.0000	0.0980	0.1770	0.0784	0.0200	0.0000
TURR	Turkey Byzantine	Near East	0.1250	0.0417	0.0000	0.0000	0.2500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0833	0.0833	0.2083	0.0420	0.1670	0.0000
SYR	Syria	Near East	0.2870	0.0290	0.0220	0.0960	0.0660	0.0510	0.0000	0.0000	0.0000	0.0150	0.0070	0.0370	0.0000	0.0440	0.0810	0.1910	0.0290	0.0070	0.0280
IRQ	Iraq	Near East	0.1730	0.0950	0.0100	0.1330	0.0520	0.0760	0.0120	0.0120	0.0000	0.0290	0.0290	0.0550	0.0050	0.0550	0.0360	0.1760	0.0210	0.0120	0.0450
TUR	Turkey	Near East	0.3160	0.0530	0.0160	0.0940	0.0640	0.0160	0.0000	0.0040	0.0040	0.0120	0.0100	0.0100	0.0060	0.0310	0.0410	0.1770	0.0270	0.0310	0.9660
TKM	Turkmenistan	Near East	0.2290	0.0130	0.0110	0.0660	0.0370	0.0000	0.0000	0.0030	0.0050	0.0000	0.0260	0.0160	0.0210	0.0130	0.0450	0.0950	0.0130	0.0210	0.0440
YEM	Yemen	Near East	0.0570	0.0500	0.0060	0.1660	0.0610	0.3080	0.0150	0.0190	0.0190	0.0040	0.0250	0.1070	0.0060	0.0170	0.0210	0.0710	0.0060	0.0080	0.0000
ARE	United Arab Emirates	Near East	0.0980	0.0730	0.0280	0.1120	0.0670	0.1740	0.0060	0.0220	0.0030	0.0170	0.0280	0.0560	0.0000	0.0140	0.0200	0.1400	0.0360	0.0110	0.0070
LBN	Lebanon	Near East	0.3440	0.0400	0.0160	0.0770	0.0840	0.0200	0.0100	0.0050	0.0010	0.0290	0.0040	0.0320	0.0050	0.0600	0.0450	0.1470	0.0190	0.0220	0.0960
ISR	Israel	Near East	0.2930	0.0400	0.0150	0.0790	0.1040	0.0790	0.0120	0.0090	0.0030	0.0220	0.0300	0.0130	0.0340	0.0340	0.1010	0.0100	0.0880	0.4050
OMN	Oman	Near East	0.1580	0.0110	0.0320	0.0950	0.0840	0.1790	0.0000	0.0420	0.0000	0.0000	0.0210	0.1790	0.0000	0.0000	0.0210	0.1260	0.0000	0.0000	0.0260
AFG	Afganistan	Near East	0.2330	0.0110	0.0000	0.0110	0.2220	0.0000	0.0000	0.0110	0.0000	0.0110	0.0000	0.0000	0.0110	0.0000	0.0440	0.2330	0.0000	0.0000	0.0000
SAU	Saudi Arabia	Near East	0.0910	0.0110	0.0090	0.2120	0.0390	0.1000	0.0370	0.0260	0.0230	0.0230	0.0050	0.1770	0.0000	0.0230	0.0400	0.1120	0.0110	0.0280	0.3200
QAT	Qatar	Near East	0.0790	0.0450	0.0110	0.1800	0.0340	0.1690	0.0340	0.0000	0.0110	0.0110	0.0000	0.0790	0.0000	0.0000	0.0450	0.1800	0.0340	0.0110	0.0600
IRN	Iran	Near East	0.1820	0.0880	0.0250	0.1380	0.0590	0.0130	0.0010	0.0250	0.0030	0.0140	0.0260	0.0140	0.0100	0.0390	0.0420	0.1970	0.0230	0.0190	0.0040
JOR	Jordan	Near East	0.2530	0.0600	0.0160	0.0600	0.0440	0.1370	0.0220	0.0160	0.0000	0.0220	0.0050	0.0270	0.0050	0.0050	0.0660	0.2250	0.0110	0.0110	0.1480
PDTM	Italy Piedmont Medieval	Europe	0.5000	0.0000	0.0360	0.1790	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0360	0.0000	0.0714	0.1790	0.0000	0.0000	0.0000
VEN	Italy Venosa Medieval	Europe	0.5000	0.0450	0.0450	0.0450	0.0000	0.0450	0.0000	0.0450	0.0000	0.0000	0.0000	0.0450	0.0000	0.0450	0.0000	0.0900	0.0450	0.0000	0.0450
CAL	Italy Calabria	Europe	0.3542	0.0313	0.0000	0.1458	0.0625	0.0104	0.0104	0.0000	0.0000	0.0313	0.0000	0.0104	0.0313	0.0208	0.0938	0.1042	0.0729	0.0104	0.0104
SIC	Italy Sicily	Europe	0.3333	0.0667	0.0095	0.0857	0.0762	0.0286	0.0095	0.0095	0.0000	0.0190	0.0000	0.0286	0.0381	0.0952	0.1143	0.0190	0.0286	0.0381
BAS	Italy Basilicata	Europe	0.3684	0.0526	0.0316	0.0421	0.1579	0.0211	0.0000	0.0000	0.0000	0.0000	0.0421	0.0105	0.0000	0.0211	0.0842	0.1263	0.0211	0.0211	0.0000
WAL	Wales	Europe	0.4720	0.0000	0.0520	0.1010	0.0910	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0100	0.0230	0.0490	0.1580	0.0030	0.0080	0.0370
GBR	United Kingdom	Europe	0.4350	0.0010	0.0380	0.1150	0.0770	0.0020	0.0000	0.0000	0.0020	0.0030	0.0000	0.0000	0.0070	0.0150	0.0690	0.1670	0.0150	0.0180	0.5800
PRT	Portugal	Europe	0.4280	0.0180	0.0210	0.0690	0.0600	0.0610	0.0090	0.0010	0.0010	0.0030	0.0020	0.0040	0.0030	0.0320	0.0630	0.1410	0.0200	0.0180	0.0000
GRC	Greece	Europe	0.4140	0.0300	0.0210	0.0980	0.0520	0.0010	0.0050	0.0040	0.0030	0.0120	0.0050	0.0110	0.0040	0.0360	0.0680	0.1520	0.0130	0.0430	0.8420
LVA	Latvia	Europe	0.4210	0.0220	0.0460	0.0610	0.0240	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0120	0.0170	0.0630	0.2580	0.0410	0.0020	0.0610
BIH	Bosnia Herzegovina	Europe	0.4470	0.0530	0.0320	0.0800	0.0600	0.0050	0.0020	0.0050	0.0030	0.0020	0.0110	0.0030	0.0020	0.0120	0.0380	0.1680	0.0230	0.0140	0.0590
UKR	Ukraine	Europe	0.3860	0.0410	0.0290	0.0820	0.0500	0.0010	0.0000	0.0100	0.0000	0.0000	0.0030	0.0000	0.0930	0.0040	0.0140	0.1870	0.0270	0.0130	0.6550
FRO	Faroe Islands	Europe	0.7110	0.0000	0.0580	0.1820	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0080	0.0170	0.0170	0.0000	0.1230	
FIN	Finland	Europe	0.3430	0.0040	0.0370	0.0530	0.0400	0.0000	0.0000	0.0000	0.0030	0.0010	0.0020	0.0000	0.0000	0.0190	0.0230	0.2550	0.0840	0.0120	0.0000
GER	Germany	Europe	0.4240	0.0140	0.0240	0.0920	0.0790	0.0020	0.0000	0.0010	0.0070	0.0020	0.0000	0.0000	0.0030	0.0270	0.0860	0.1570	0.0200	0.0170	0.0340
NOR	Norway	Europe	0.4190	0.0040	0.0170	0.0940	0.0490	0.0030	0.0000	0.0010	0.0030	0.0040	0.0000	0.0000	0.0040	0.0090	0.0690	0.2170	0.0160	0.0040	0.0000
SWE	Sweden	Europe	0.3800	0.0160	0.0220	0.0620	0.0520	0.0050	0.0000	0.0640	0.0000	0.0000	0.0000	0.0000	0.0000	0.0210	0.0350	0.2640	0.0100	0.0100	0.0190
POL	Poland	Europe	0.3960	0.0370	0.0160	0.1080	0.0400	0.0070	0.0000	0.0140	0.0000	0.0000	0.0000	0.0020	0.0070	0.0230	0.0770	0.2110	0.0260	0.0160	0.0580
BGR	Bulgaria	Europe	0.4120	0.0470	0.0160	0.0790	0.0510	0.0020	0.0010	0.0020	0.0050	0.0100	0.0020	0.0060	0.0130	0.0380	0.0410	0.1480	0.0220	0.0310	0.0550
HRV	Croatia	Europe	0.4060	0.0400	0.0200	0.0830	0.0510	0.0010	0.0000	0.0040	0.0120	0.0000	0.0010	0.0000	0.0000	0.0210	0.0610	0.1950	0.0350	0.0120	1.0000
CHE	Switzerland	Europe	0.4890	0.0140	0.0090	0.1190	0.0550	0.0090	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0230	0.0960	0.1230	0.0180	0.0050	0.0590
ESP	Spain	Europe	0.4320	0.0220	0.0110	0.0710	0.0600	0.0300	0.0050	0.0010	0.0000	0.0020	0.0030	0.0060	0.0020	0.0210	0.0710	0.1730	0.0130	0.0180	0.0400
CZE	Czech Republic	Europe	0.4310	0.0230	0.0260	0.1120	0.0440	0.0020	0.0000	0.0020	0.0050	0.0050	0.0000	0.0000	0.0020	0.0370	0.0860	0.1590	0.0090	0.0160	0.0600
EST	Estonia	Europe	0.4590	0.0100	0.0100	0.1020	0.0190	0.0000	0.0000	0.0000	0.0130	0.0000	0.0000	0.0000	0.0060	0.0160	0.0700	0.2480	0.0250	0.0060	0.0010
CYP	Cyprus	Europe	0.2940	0.0000	0.0350	0.0590	0.2000	0.0350	0.0240	0.0000	0.0000	0.0120	0.0470	0.0000	0.0120	0.0000	0.0590	0.1530	0.0350	0.0350	0.8390
ISL	Iceland	Europe	0.3780	0.0350	0.0420	0.1400	0.1020	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0110	0.0050	0.1000	0.1360	0.0070	0.0150	0.0050
AUT	Austria	Europe	0.4520	0.0160	0.0110	0.0880	0.0880	0.0000	0.0000	0.0050	0.0000	0.0000	0.0000	0.0030	0.0030	0.0400	0.0800	0.1710	0.0130	0.0130	0.0260
BLR	Belarus	Europe	0.3930	0.0340	0.0200	0.0890	0.0290	0.0000	0.0000	0.0150	0.0000	0.0080	0.0030	0.0000	0.0020	0.0200	0.0550	0.2200	0.0360	0.0070	0.0630
SVN	Slovenia	Europe	0.4370	0.0200	0.0220	0.1090	0.0530	0.0000	0.0000	0.0000	0.0040	0.0000	0.0000	0.0000	0.0020	0.0290	0.0730	0.1740	0.0200	0.0110	0.1430
HUN	Jungary	Europe	0.3090	0.0130	0.0180	0.0680	0.1310	0.0050	0.0000	0.0000	0.0140	0.0250	0.0030	0.0010	0.0030	0.0200	0.0790	0.0810	0.0260	0.0610	0.9230
ALB	Albania	Europe	0.4790	0.0210	0.0210	0.0890	0.0340	0.0070	0.0000	0.0000	0.0070	0.0210	0.0070	0.0140	0.0000	0.0550	0.0550	0.1640	0.0140	0.0070	0.0090
ITA	Italy	Europe	0.4210	0.0360	0.0140	0.0850	0.0770	0.0080	0.0060	0.0020	0.0020	0.0100	0.0040	0.0080	0.0060	0.0320	0.0870	0.1330	0.0170	0.0190	0.0060
MKD	Macedonia	Europe	0.4740	0.0480	0.0130	0.0610	0.0610	0.0000	0.0000	0.0000	0.0000	0.0160	0.0100	0.0000	0.0000	0.0610	0.0610	0.1380	0.0260	0.0260	0.2110
SCOT	Scotland	Europe	0.4420	0.0020	0.0410	0.1280	0.0690	0.0000	0.0010	0.0000	0.0010	0.0010	0.0000	0.0000	0.0120	0.0220	0.0590	0.1560	0.0060	0.0250	0.0040
RUS	Russia	Europe	0.1750	0.0140	0.0090	0.0430	0.0170	0.0010	0.0010	0.0070	0.0020	0.0050	0.0050	0.0010	0.0030	0.0150	0.0250	0.1070	0.0090	0.0060	0.0350
IRL	Ireland	Europe	0.4410	0.0130	0.0300	0.1070	0.1200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0070	0.0130	0.0540	0.1240	0.0230	0.0070	0.1020
SVK	Slovakia	Europe	0.3900	0.0290	0.0410	0.0910	0.0370	0.0040	0.0010	0.0000	0.0030	0.0030	0.0000	0.0000	0.0130	0.0770	0.1780	0.0190	0.0100	0.0410
FRA	France	Europe	0.4470	0.0360	0.0200	0.0780	0.0860	0.0080	0.0000	0.0030	0.0030	0.0040	0.0060	0.0040	0.0080	0.0170	0.0650	0.1430	0.0170	0.0120	0.0450
SRB	Serbia	Europe	0.4480	0.0130	0.0190	0.1230	0.0450	0.0000	0.0000	0.0000	0.0190	0.0060	0.0000	0.0000	0.0060	0.0320	0.0390	0.1360	0.0520	0.0130	0.6710
ROU	Romania	Europe	0.3520	0.0380	0.0070	0.1020	0.1000	0.0000	0.0000	0.0010	0.0010	0.0130	0.0040	0.0200	0.0040	0.0420	0.0510	0.1600	0.0470	0.0180	0.0070
LTU	Lithuania	Europe	0.4520	0.0090	0.0310	0.0620	0.0140	0.0000	0.0000	0.0000	0.0090	0.0140	0.0000	0.0000	0.0000	0.0310	0.0680	0.2100	0.0200	0.0090	0.7830
DNK	Denmark	Europe	0.4500	0.0070	0.0430	0.1290	0.0860	0.0040	0.0000	0.0000	0.0110	0.0000	0.0140	0.0000	0.0110	0.0110	0.0570	0.1360	0.0070	0.0110	0.0110
ARM	Armenia	Caucasus	0.2940	0.0600	0.0140	0.0960	0.0730	0.0000	0.0000	0.0230	0.0090	0.0280	0.0140	0.0050	0.0090	0.0550	0.0500	0.2200	0.0090	0.0370	0.0000

Code	Country	Region																				
GEO	Georgia	Caucasus	0.2320	0.0370	0.0070	0.0300	0.1330	0.0000	0.0110	0.0110	0.0000	0.0180	0.0180	0.0070	0.0300	0.0440	0.0890	0.2360	0.0330	0.0300	0.3870
AZE	Azerbaijan	Caucasus	0.2350	0.0610	0.0260	0.0610	0.0430	0.0000	0.0090	0.0090	0.0000	0.0000	0.0170	0.0000	0.0350	0.0430	0.0960	0.2090	0.0260	0.0350	0.0400
SDN	Sudan	Sub Saharan Africa	0.0520	0.0090	0.0000	0.0170	0.0090	0.7220	0.0700	0.0000	0.0000	0.0000	0.0000	0.0430	0.0000	0.0170	0.0000	0.0260	0.0000	0.0090	0.0340
MLI	Mali	Sub Saharan Africa	0.0450	0.0040	0.0000	0.0000	0.0000	0.9190	0.0180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0090	0.0000	0.0000	0.0210
CAF	Cantral African Republic	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
COD	Democratic Republic of Congo	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0330
GIN	Guinea	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	0.9380	0.0110	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0510	0.0000	0.0000	0.0000
TCD	Chad	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	0.9180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0820	0.0000	0.0000	0.9050
GNQ	Equatorial Guinea	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0130
KEN	Kenya	Sub Saharan Africa	0.0000	0.0070	0.0000	0.0070	0.0070	0.9210	0.0360	0.0000	0.0000	0.0000	0.0000	0.0070	0.0000	0.0000	0.0000	0.0070	0.0000	0.0000	0.0410
BFA	Burkina Faso	Sub Saharan Africa	0.1270	0.0000	0.0000	0.0000	0.0000	0.7390	0.0520	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0220	0.0000	0.0000	0.0350
MRT	Mauritania	Sub Saharan Africa	0.1160	0.0000	0.0000	0.0350	0.0580	0.4770	0.0000	0.0000	0.0000	0.0000	0.0000	0.0230	0.0000	0.0000	0.0000	0.2560	0.0000	0.0000	0.0250
CMR	Cameroun	Sub Saharan Africa	0.0020	0.0000	0.0000	0.0000	0.0000	0.9850	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0120	0.0000	0.0000	0.0000
GAB	The Gambia	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9350
SEN	Senegal	Sub Saharan Africa	0.0090	0.0000	0.0000	0.0000	0.0000	0.9650	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0180	0.0000	0.0000	0.0320
SLE	Sierra Leone	Sub Saharan Africa	0.0040	0.0000	0.0000	0.0000	0.0000	0.9850	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0070	0.0000	0.0000	0.5540
NER	Niger	Sub Saharan Africa	0.0320	0.0000	0.0000	0.0000	0.0000	0.8060	0.0320	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0320	0.0000	0.0000	0.0000
RWA	Rwanda	Sub Saharan Africa	0.0000	0.0000	0.0000	0.0000	0.0000	0.9820	0.0180	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3740
ESH	Western Sahara	Sub Saharan Africa	0.1670	0.0000	0.0000	0.0830	0.0000	0.5000	0.0420	0.0000	0.0000	0.0420	0.0000	0.0000	0.0000	0.0000	0.0000	0.1670	0.0000	0.0000	0.0370
ETH	Ethiopia	Sub Saharan Africa	0.0070	0.0170	0.0070	0.0120	0.0100	0.6700	0.1150	0.0020	0.0190	0.0030	0.0000	0.0800	0.0000	0.0030	0.0120	0.0310	0.0090	0.0030	0.0000
CPV	Cape Verde	Sub Saharan Africa	0.0030	0.0070	0.0000	0.0000	0.0000	0.9350	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030	0.0000	0.0000	0.0030	0.0340	0.0000	0.0070	0.0970

Supplementary table 5: Pairwise distance (Fst) calculation between Venosa and 56 comparative populations

Population	Longitude	Latitude	FSt	*P*-value
Italy Piedmont Medieval	7.515	45.052	0.0000	0.60059+-0.0129
Turkey Byzantine	32.864	39.927	0.0000	0.42480+-0.0171
Russia	37.624	55.750	0.0000	0.65820+-0.0145
Azerbaijan	47.577	40.143	0.0000	0.46582+-0.0125
Iraq	43.679	33.223	0.0000	0.48438+-0.0155
Jordan	36.238	30.585	0.0000	0.56445+-0.0173
Palestinians	34.791	31.253	0.0000	0.72754+-0.0133
Syria	38.997	34.802	0.0000	0.48535+-0.0144
Turkey	35.243	38.964	0.0000	0.60059+-0.0132
Belorussia	27.953	53.710	0.0000	0.56055+-0.0149
Czech Republic	15.473	49.817	0.0000	0.59375+-0.0148
Poland	19.145	51.919	0.0000	0.50781+-0.0144
Slovakia	19.699	48.669	0.0000	0.58496+-0.0142
Slovenia	14.995	46.151	0.0000	0.64746+-0.0168
Ukraine	31.166	48.379	0.0000	0.45312+-0.0144
Latvia	24.603	56.880	0.0000	0.45898+-0.0136
Lithuania	23.881	55.169	0.0000	0.52148+-0.0176
Portugal	-8.224	39.400	0.0000	0.69043+-0.0123
Spain	-3.749	40.464	0.0000	0.55762+-0.0180
Albania-Macedonia	21.745	41.609	0.0000	0.44434+-0.0167
Bosnia-Herzegovina-Croatia-Serbia	17.679	43.916	0.0000	0.57129+-0.0112
Bulgaria	25.486	42.734	0.0000	0.47070+-0.0146
Estonia	25.014	58.595	0.0005	0.45215+-0.0171
Hungary	19.503	47.162	0.0008	0.40527+-0.0132
France	2.214	46.228	0.0015	0.36816+-0.0153
Greece	21.824	39.074	0.0016	0.34473+-0.0142
England-Wales	-1.174	52.356	0.0016	0.36426+-0.0137
Libya	17.228	26.335	0.0020	0.34668+-0.0147
Germany	10.452	51.166	0.0020	0.33691+-0.0124
Italy	12.567	41.872	0.0020	0.33203+-0.0146
Oman-Quatar-United Arab Emirates	53.848	23.424	0.0021	0.33496+-0.0138
Tunisia	9.537	33.887	0.0026	0.29199+-0.0115
Iran	53.688	32.428	0.0028	0.31641+-0.0122
Lebanon	35.862	33.855	0.0029	0.31445+-0.0134
Sweden	18.644	60.128	0.0031	0.30273+-0.0140
Norway	8.469	60.472	0.0038	0.26855+-0.0125
Denmark	9.502	56.264	0.0043	0.25000+-0.0129
Armenia	45.038	40.069	0.0043	0.25586+-0.0142
Scotland	-4.203	56.491	0.0044	0.23438+-0.0101
Romania	24.967	45.943	0.0047	0.27539+-0.0126
Morocco	-7.093	31.792	0.0053	0.22168+-0.0111
Ireland	-8.244	53.413	0.0055	0.22949+-0.0134
Egypt	30.802	26.821	0.0056	0.21680+-0.0170
Switzerland	8.228	46.818	0.0061	0.21484+-0.0139
Ossetians	44.287	43.045	0.0063	0.18457+-0.0113
Turkey Roman	30.519	37.678	0.0063	0.29492+-0.0123
Kuwait	47.482	29.312	0.0065	0.19434+-0.0122
Austria	14.550	47.516	0.0078	0.18066+-0.0102
Georgia	43.357	42.315	0.0084	0.17578+-0.0112
Basques	-2.619	42.990	0.0104	0.11914+-0.0088
Berber	1.660	28.034	0.0106	0.14453+-0.0109
Iceland	-19.021	64.963	0.0110	0.10156+-0.0078
Yemen	48.516	15.553	0.0162	0.06934+-0.0091
Saudi Arabia	45.079	23.886	0.0185	0.04785+-0.0067
Druze	35.690	32.995	0.0208	0.03906+-0.0055
Algerian	3.042	36.753	0.0290	0.03613+-0.0055

Cell Host & Microbe

Historical *Y. pestis* Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics

Graphical Abstract
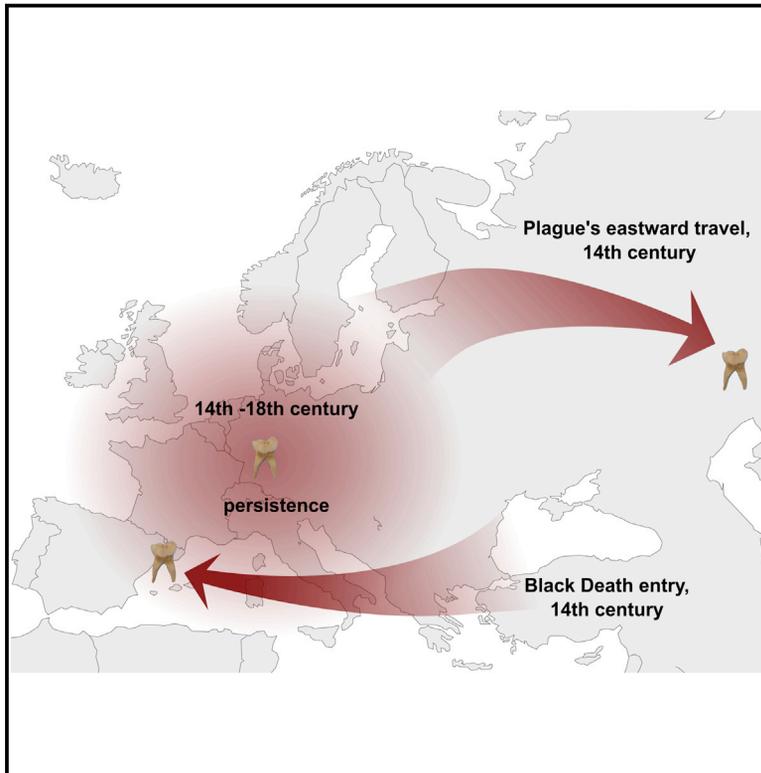
Authors

Maria A. Spyrou, Rezeda I. Tukhbatova, Michal Feldman, ..., Alexander Herbig, Kirsten I. Bos, Johannes Krause

Correspondence

herbig@shh.mpg.de (A.H.),
bos@shh.mpg.de (K.I.B.),
krause@shh.mpg.de (J.K.)

In Brief

Spyrou et al. have sequenced historical *Yersinia pestis* genomes from victims of the Black Death and subsequent outbreaks in Europe. Their data suggest a connection between the Black Death and the modern-day plague pandemic as well as the persistence of plague in Europe between the 14th and 18th centuries.

Highlights

- Three historical *Yersinia pestis* genomes from the second plague pandemic in Europe

- Low genetic diversity of the pathogen during the Black Death

- Indication for link between the Black Death and 19th century plague pandemic lineages

- Connection between post-Black Death outbreaks in Europe supports a local plague focus

CrossMark

CellPress

Historical *Y. pestis* Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics

Maria A. Spyrou,[1] Rezeda I. Tukhbatova,[2,3] Michal Feldman,[1] Joanna Drath,[4] Sacha Kacki,[5] Julia Beltrán de Heredia,[6] Susanne Arnold,[7] Airat G. Sitdikov,[2,3] Dominique Castex,[5] Joachim Wahl,[4,8] Ilgizar R. Gazimzyanov,[3] Danis K. Nurgaliev,[9] Alexander Herbig,[1,*] Kirsten I. Bos,[1,*] and Johannes Krause[1,4,*]

[1]Max Planck Institute for the Science of Human History, Jena 07743, Germany
[2]Laboratory of Paleoanthropology & Paleogenetics, Kazan Federal University, Kazan 420008, Russian Federation
[3]Institute of Archaeology named after A. Kh. Khalikov, Tatarstan Academy of Sciences, Kazan 420012, Russian Federation
[4]Department of Archeological Sciences, University of Tuebingen, Tuebingen 72070, Germany
[5]PACEA, CNRS Institute, Université de Bordeaux, Pessac 33615, France
[6]Museu de Historia de Barcelona, Barcelona 08002, Spain
[7]State Office for Cultural Heritage Management Baden-Württemberg, Esslingen 73728, Germany
[8]State Office for Cultural Heritage Management Baden-Württemberg, Osteology, Konstanz 78467, Germany
[9]Institute of Geology and Petroleum Technologies, Kazan Federal University, Kazan 420008, Russian Federation
*Correspondence: herbig@shh.mpg.de (A.H.), bos@shh.mpg.de (K.I.B.), krause@shh.mpg.de (J.K.)
http://dx.doi.org/10.1016/j.chom.2016.05.012

SUMMARY

Ancient DNA analysis has revealed an involvement of the bacterial pathogen *Yersinia pestis* in several historical pandemics, including the second plague pandemic (Europe, mid-14th century Black Death until the mid-18th century AD). Here we present reconstructed *Y. pestis* genomes from plague victims of the Black Death and two subsequent historical outbreaks spanning Europe and its vicinity, namely Barcelona, Spain (1300–1420 cal AD), Bolgar City, Russia (1362–1400 AD), and Ellwangen, Germany (1485–1627 cal AD). Our results provide support for (1) a single entry of *Y. pestis* in Europe during the Black Death, (2) a wave of plague that traveled toward Asia to later become the source population for contemporary worldwide epidemics, and (3) the presence of an historical European plague focus involved in post-Black Death outbreaks that is now likely extinct.

INTRODUCTION

Yersinia pestis evolved from the closely related zoonotic enterobacterium *Y. pseudotuberculosis* (Achtman et al., 1999) to become one of the most virulent pathogens known to humans. Its recent identification in ancient human material from Altai, Siberia suggests it caused human infections as early as 5,000 years ago, though its ability for flea-borne transmission leading to bubonic disease might have been absent in these older, divergent lineages (Rasmussen et al., 2015). To our knowledge, bubonic plague, and presumably also the pneumonic and septicemic forms, have been the likely culprit of three major pandemics, namely the Plague of Justinian (Eastern Roman Empire, 6th and 8th centuries AD), the second-wave plague pandemic (Europe,

mid-14th century Black Death until the mid-18th century AD), and the third plague pandemic that started during the late 19th century in China. Differences in mortality rate and epidemiology of the three pandemics initiated controversy over whether they shared a common etiologic agent (Cohn, 2008). In recent years, however, ancient DNA (aDNA) has confirmed a *Y. pestis* involvement in both historical pandemics (Bos et al., 2011; Haensch et al., 2010; Wagner et al., 2014).

The Black Death claimed up to 50% of the European population between 1347 and 1353 (Benedictow, 2004). The disease is thought to have arisen from plague foci in East Asia and to have spread into Europe via trade routes (Morelli et al., 2010). Its origin, however, is still contentious due to a lack of convincing archeological or documentary evidence from the early 14th century in East Asia (Sussman, 2011). Ancient *Y. pestis* genomes obtained from medieval victims have indicated the presence of a radiation event immediately preceding the Black Death that gave rise to most of the strain diversity circulating in the world today (Bos et al., 2011; Cui et al., 2013). Based on the relationship of ancient European and modern genomes, it was recently suggested that a wave of plague might have traveled from Europe toward Asia after the Black Death, eventually settling in China and later giving rise to the third pandemic (Wagner et al., 2014). Genomes from its purported route are, however, missing in the discussions, and are needed to add legitimacy to the model.

After the Black Death, plague continued to strike Europe for another four centuries through subsequent outbreaks that ceased at the end of the 18th century (Benedictow, 2004). The reasons for its sudden disappearance in Europe are unknown. Sylvatic plague foci have a nearly worldwide presence today, but are absent in Europe (Gage and Kosoy, 2005; Tikhomirov, 1999). The question of whether the recurrent European plague outbreaks of the 14th to 18th centuries were the result of multiple reintroductions of plague into Europe, or rather were attributed to now-extinct European plague foci, is still being explored. Previous studies that draw upon aDNA and climatic data favor the former hypothesis. Through a SNP-based PCR approach,

CrossMark

Figure 1. Samples and Their Respective Locations
(A) Tooth sample that was positive for *Y. pestis* (3031) and mass grave picture from the plague burial in Barcelona.
(B) *Y. pestis*-positive tooth sample and picture of infected individual (2370) from the Ust'-Ierusalimsky tomb of Bolgar City.
(C) Picture of mass grave in Ellwangen, and two tooth samples from individual 549_O, found positive for the plague bacterium.

purportedly distinct plague lineages were identified in different areas of Europe during the 14th century and were thought to have entered via different pulses (Haensch et al., 2010). In addition, plague outbreaks documented in some of the main Mediterranean ports were found to coincide with extreme climate fluctuations in Central Asia, suggesting that recurrent maritime imports of plague from Asia might have been responsible for post-Black Death plague outbreaks (Schmid et al., 2015). By contrast, others have suggested a long-term persistence of plague in Europe (Seifert et al., 2016). Using a PCR SNP-typing approach of putative plague material from Southern and Northeastern Germany, identical *Y. pestis* SNP profiles were identified in strains circulating within Europe between the Black Death and 17th century AD (Seifert et al., 2016), implying a single source population for the European plagues of that time period. A further genome-wide analysis of *Y. pestis* strains from the Great Plague of Marseille (1720–1722) has identified a previously uncharacterized lineage of *Y. pestis* that descends from a strain present during the Black Death (Bos et al., 2016). While the lineage is considered to represent an historical plague focus potentially responsible for post-Black Death European outbreaks (Bos et al., 2016), the use of material from a highly operational Mediterranean center that linked Western Europe with the East (Signoli et al., 1998) makes identification of the disease source elusive.

Here, we aim to address three outstanding questions regarding *Y. pestis* history. First, we investigate the possibility of disease entry via multiple pulses during the Black Death by comparing the genotype of a strain from the pandemic's early phase to those circulating in other areas later in the pandemic. Material from Barcelona, Spain, one of the Mediterranean cities through which plague entered southern continental Europe, is compared to Black Death genomes from London. Second, we evaluate the likelihood of the proposed eastward migration of strains from Europe to Asia after the Black Death through the analysis of human remains from a 14th century plague burial in the Volga region of Russia. Third, we take a further step toward understanding the

relationship of post-Black Death outbreaks in Europe and evaluate the likelihood of a local reservoir. For this, we investigate a 16th century plague outbreak in Southwestern Germany and compare it to both a London outbreak that occurred soon after the Black Death and to the Great Plague of Marseille, France in 1722. Following the success of previous genomic investigations of ancient bacterial disease (Bos et al., 2011, 2014, 2016; Schuenemann et al., 2011, 2013; Wagner et al., 2014), we employ similar methods of DNA capture and high-throughput sequencing to retrieve the genomes of three historical *Y. pestis* strains. Our results suggest (1) limited *Y. pestis* diversity during the early phase of the Black Death, and likely a single entry into Europe; (2) a wave of plague that traveled eastward after the Black Death and later gave rise to the 19th century pandemic; and (3) an involvement of the same plague lineage in two post-Black Death European epidemics that are 200 years apart.

RESULTS

Archaeological Sites and Dating

Samples were collected from a mass grave in Barcelona, Spain, a single grave in Bolgar City in Russia, and a mass grave in Ellwangen, Germany (Figure 1 and Supplemental Experimental Procedures). Aside from the Bolgar City site that was dated to the second half of the 14th century using coin artifacts known to have been minted after 1362 (Supplemental Experimental Procedures and Figure S1), archaeological dates were not available. To estimate or confirm the historical period during which each of the outbreaks occurred, radiocarbon dates from bone fragments and tooth roots were obtained. The dates yielded were 1300–1420 cal AD for Barcelona, 1298–1388 cal AD for Bolgar City, and 1486–1627 cal AD for Ellwangen (Figure 1 and Table S1).

Screening for *Y. pestis*

A total of 223 DNA extracts from teeth of 178 individuals were evaluated for the presence of *Y. pestis* DNA through a

A

Branch 1 (CHN, USA, MDG, IND, MNM, COG, UGA)
incl. 3rd pandemic 1.ORI
82
94
Bolgar (RUS) 1362-1400+
London (6330) (GB) 1350-1400+
Ellwangen (DE) 1485-1627+
Observance (FRA) 1720-22+
95
London (GB) 1348-50+
Barcelona (SP) 1300-1420+
Branch 4 (MNG)
98 86
Branch 3 (CHN, MNG)
99
2.ANT (CHN, NPL, FSU)
2.MED (incl 2.MED1) (CHN, FSU, IRN)
0.ANT3 (CHN, FSU)
0.ANT2 (CHN)
87
0.ANT1 (CHN)
94
0.PE4 (CHN, FSU)
0.PE3 (AGO)
0.PE2 (FSU)
0.PE7 (CHN)
90 SNPs
Y. pseudotuberculosis (outgroup)

B

1.ORI (3rd Pandemic)
(CHN, USA, MDG, IND, MNM)
1.IN3 (CHN)
1.IN2 (CHN)
1.IN2 (CHN)
1.IN2 (CHN)
1.IN2 (CHN)
1.IN1 (CHN)
1.IN1 (CHN)
1.IN1 (CHN)
p6
p3, p4
1.ANT1
(COG, UGA)
p7 Bolgar (RUS) 1362-1400+
p5 London (6330) (GB) 1350-1400+
p1, p2 Ellwangen (DE) 1485-1627+
Observance
(FRA) 1720-22+
London (GB) 1348-50+
Barcelona (SP) 1300-1420+
10 SNPs
Branches 0, 2, 3 and 4

Figure 2. *Yersinia pestis* Phylogeny
(A) Maximum Parsimony phylogenetic tree of 141 modern and 10 historical *Y. pestis* strains. 3,351 SNP positions were considered for the phylogeny. The reconstructed tree shows the topology of the new isolates from Barcelona, Bolgar City, and Ellwangen relative to previously sequenced modern and ancient *Y. pestis* strains. Asterisks (*) indicate bootstrap values of 100. Collapsed branches are represented by triangles, to enhance tree clarity. Strains belonging to Branch 1 are represented in red, Branch 2 in yellow, Branch 3 in blue, Branch 4 in orange, and Branch 0 in black. Ancient Branch 1 strains are indicated by their archaeological or radiocarbon date and by a (+). Because of the great number of derived SNP positions of the 0.PE3 lineage, its branch was reduced to adjust scaling of the tree. Geographic region abbreviations correspond to: CHN (China), USA (United States of America), MDG (Madagascar), IND (India), IRN (Iran), MNM (Myanmar), RUS (Russia), GB (Great Britain), DE (Germany), FRA (France), SP (Spain), MNG (Mongolia), NPL (Nepal), FSU (Former Soviet Union), AGO (Angola), CGO (Congo), and UGA (Uganda).
(B) A magnified version of Branch 1 is shown to enhance its resolution. The branch of lineage 1.ANT was manually reduced to adjust tree scaling. A detailed description of p1-p7 SNPs is given in Table 1 (see also Table S2, Table S3, Table S4 and Figure S2).

and the *Y. pestis* plasmids pMT1 and pCD1 as template for probe design (Supplemental Experimental Procedures). Array captures produced average genomic coverage of 10.3-fold for Barcelona 3031, 19.3-fold for Bolgar City 2370, and 4.9-fold for Ellwangen 549_O (Table S1 and Table S2). Owing to its low coverage, data presented for sample 549_O are from a pool of two independent libraries produced from two teeth of the same individual (Table S1 and Table S2).

Phylogenetic Analysis of Historical *Y. pestis* Genomes
Our ancient genomes were then added to a *Y. pestis* phylogeny constructed from previously published genomes including 130 modern genomes (Cui et al., 2013), 7 historical genomes (Bos et al., 2011, 2016), and 11 newly available modern *Y. pestis* strains from the Former Soviet Union (Zhgenti et al., 2015) (Table S3). Our maximum parsimony tree revealed that the modern Former Soviet Union genomes group with what was previously thought to be diversity restricted in China, specifically lineage 0.ANT3 (Cui et al., 2013). They also add further diversity to the 2.MED1 lineage and, importantly, to the 0.PE2 lineage, which is the second deepest branch in the *Y. pestis* phylogeny (Figure 2A, Figure S2, and Table S3). This reveals a more extensive *Y. pestis* diversity outside of China than was previously thought.

species-specific quantitative PCR (qPCR) assay targeting the plasminogen activator (*pla*) gene located on the PCP1 plasmid (Schuenemann et al., 2011) (Supplemental Experimental Procedures). Results indicated 53 potentially positive DNA extracts stemming from 32 individuals. All extraction and PCR blanks were free of amplification products. Amplification products were not sequenced, as samples from potentially positive individuals were directly turned into double-stranded next-generation sequencing libraries and were used for whole-genome array capture. After capture, three individuals had sufficient *Y. pestis* DNA for genome-level analysis. These were tooth specimens 3031 from Barcelona, 2370 from Bolgar City, and 549_O from Ellwangen (Figure 1, Table S1 and Supplemental Experimental Procedures).

Y. pestis Genome-Capture Results
Whole-genome array capture was performed using the chromosome of *Y. pseudotuberculosis* (Chain et al., 2004)

Table 1. SNP Description of Diagnostic Branch 1 Positions in the Newly Sequenced Ancient *Y. pestis* Genomes

SNP Name	Position on Chromosome CO92	CO92 (Reference)	Barcelona	Bolgar City	Ellwangen	Gene
p1	189,227	C	C	C	C	pabA
p2	1,871,476	G	G	G	G	NC[a]
p3[b]	699,494	A	G	A	G	alt (rpoD)
p4	2,262,577	T	G	T	G	YPO1990
p5	4,301,295	G	G	G	G	recQ
p6	3,806,677	C	T	C	T	b0125 (hpt)
p7	3,643,387	G	G	T	G	YPO3271

[a]Non-coding (NC).
[b]The p3 SNP corresponds to the previously described s12 position present in a Black Death plague victim from the Netherlands (Haensch et al., 2010). It is also present in a derived state isolate from London (6330), from which a complete genome is available (Bos et al., 2011).

All three reconstructed historical genomes group in Branch 1, and all possess diagnostic SNP positions here referred to as "p1" and "p2" (Table 1), which were previously identified in historical *Y. pestis* genomes from the Black Death (Bos et al., 2011) (Figure 2B, Table 1). The positioning of the strains reported here in the phylogeny confirms their authenticity as ancient. To date, all *Y. pestis* genomes isolated from the historic 2nd plague pandemic group in Branch 1.

We find no detectable differences between our Black Death strain from Barcelona and three previously genotyped strains from London 1348–1350 (Bos et al., 2011). The Bolgar City strain, however, contains additional differences in four positions compared to Black Death isolates: two of these are shared with London individual 6330 (positions p3 and p4, Figure 2B and Table 1), one is shared with all modern Branch 1 strains (p6), and one is unique to this individual (p7, Figure 2B). Additionally, the Ellwangen strain groups in a sub-branch of Branch 1, together with five strains previously typed from the Great Plague of Marseille (L'Observance), 1720–1722 (Figure 2B) (Bos et al., 2016). Our analysis reveals 20 positions shared with the strains from L'Observance and three unique SNPs (Table S4). That the Ellwangen strain is ancestral to the Observance strains comes as no surprise given the older age of the samples (Figure 2B). This "Ellwangen-Observance" lineage originates from Black Death strains currently represented by the isolates from London and Barcelona. Like the strain from Marseille, that from Ellwangen does not share additional derived positions with other ancient or modern strains (Figure 2B), as no modern descendants have as yet been identified in this sub-branch.

DISCUSSION

Our genomes from Barcelona, Bolgar City, and Ellwangen group on the same phylogenetic branch (Branch 1), adding further legitimacy to the notion that the Black Death and subsequent plague outbreaks in Europe, as well as the worldwide third pandemic, were caused by the same *Y. pestis* lineage (Figure 2, Figure S2, and Figure 3). Further analysis of ancient and modern strains of this branch could reveal important clues to explain why this particular lineage was involved in both the second and third pandemic.

Our analysis reveals that the strain from Barcelona is identical to a previously sequenced Black Death *Y. pestis* strain from London (1348–1350). Barcelona was one of the main entry points for

the Black Death into Europe, with historical reports suggesting the disease first entered there in the spring of 1348 (Gottfried, 1983). In London, the earliest reports of the illness are from autumn 1348 (Benedictow, 2004). This indicates a contemporary presence of the same strain in both southern and northern Europe, supporting the notion of a single wave entry, with low genetic diversity in the pathogen. Historical sources indicate that plague first came into view in 1347, with outbreaks in the southern islands of Crete, Sicily, and Sardinia, followed by entry into mainland Europe via the heavily trafficked ports of Genoa and Marseille. Samples from these locations and those further afield from its purported source population in East Asia may provide us with relevant details regarding the microevolution of a highly virulent pathogen at the beginning of a mass pandemic.

The key finding of our study stems from the analysis of an historical *Y. pestis* strain from the Volga region in Russia (Figure 3). This genome has added legitimacy to an important link between the second and third plague pandemics hypothesized elsewhere (Wagner et al., 2014). Under this model, *Y. pestis* spread from Europe to Asia after the Black Death and gave rise to both the 1.IN lineages of the Yunnan Province of China (1.IN3) as well as the 1.ORI strains associated with worldwide spread during the third plague pandemic (Figure 2B, Figure 3, and Table S3) (Wagner et al., 2014). That our sample from Bolgar City shares one additional Branch 1 derived position with a strain circulating in London during the second half of the 14th century provides solid evidence of plague's eastward travel subsequent to the Black Death (Table 1, Figure 2B, and Figure 3). Of note, the 1.ANT lineage today restricted to Sub-Saharan Africa possesses only an additional ten derived Branch 1 positions compared to our Bolgar lineage (Table S4). A compelling possibility is that this plague lineage was introduced via European presence in the continent: its shared ancestry with the Bolgar lineage could imply that it derives from an historical focus that existed along the eastern path that *Y. pestis* traveled after the Black Death. We therefore consider it possible that strains ancestral to these African lineages may have caused disease in Europe during the second wave and may one day be identified in ancient European skeletons.

As the geographical origins of the "p1" and "p2" SNPs are unknown (Figure 2B), the possibility of Branch 1 lineages arising from pre-existing diversity in Asia and independently dispersing into Europe must be considered (Haensch et al., 2010). This model is supported by climatic evidence, where regular
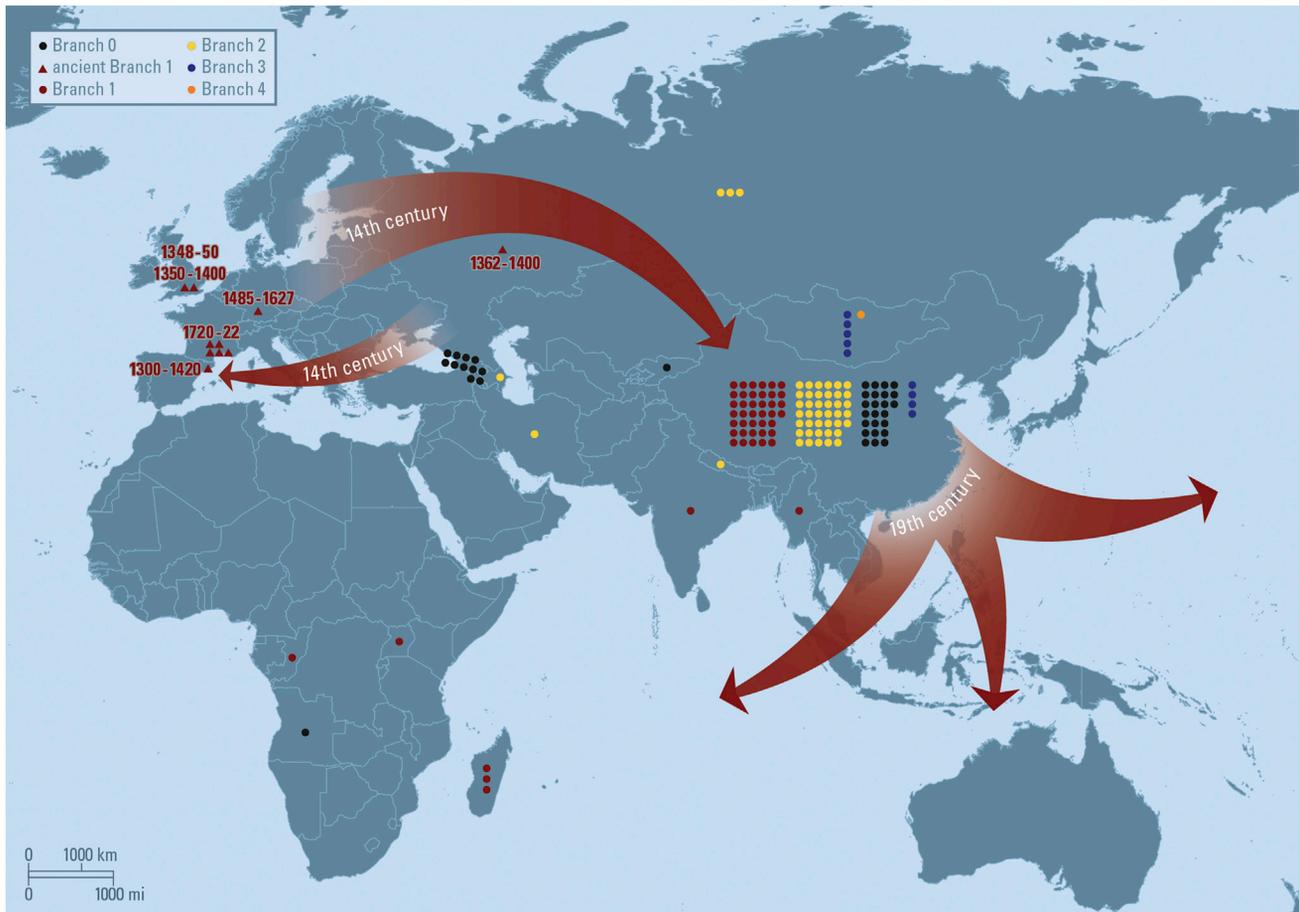
Figure 3. Plague Introduction and Dispersal

Map describing our favored dissemination pattern of *Y. pestis* during the second and third plague pandemics. All strains included in our dataset are depicted as points on the map. Branch 1 strains are in red and include both second pandemic (triangles) and modern (circles) isolates. Branch 2 strains are in yellow, Branch 3 strains are in blue, a single Branch 4 strain is in orange, and Branch 0 strains are in black. Positioning of modern strain distribution on the map corresponds to geographic location, but for the purpose of our study an accurate coordinate system was not necessary. Red arrows indicate Branch 1 cycling through Europe during the 14th century, eastward travel out of Europe after the Black Death, and global dissemination from China during the third plague pandemic (see also Table S3).

westward pulses of plague from an Asian focus throughout the second pandemic are thought possible (Schmid et al., 2015). We find this model for the second pandemic difficult to reconcile with our current data. Although it has been previously shown that *Y. pestis* has an extremely variable substitution rate (Cui et al., 2013), our Russian strain has only two additional derived substitutions (p6, p7, Figure 2) compared to London *Y. pestis* genome 6330 (Bos et al., 2011), dated to 1350–1400. This close genetic similarity suggests that our Russian strain represents a new outbreak subsequent to that which occurred in London after the Black Death. The alternative "Asian origin" model would require a minimum of four separate lineages exiting together from the same focus to account for the level of diversity observed in Europe during the Black Death and its aftermath, i.e., (1) London/Barcelona, (2) London 6330, (3) Bolgar City, and (4) Sub-Saharan Africa. We regard the likelihood of such similar strains leaving Asia in a short time frame to be low, but acknowledge it would be possible if (1) their ancestral focus was in a location particularly conducive to westward travel, or (2) there exists a biological reason for their greater ease in rapid long-distance

travel. While the above scenarios could equally explain the sole involvement of Branch 1 in contemporary plague outbreaks outside of China, we regard a single exit followed by an eastward travel as a more parsimonious explanation for the current data. Under this scenario, historical strains carrying the previously described "p3" SNP (Figure 2B) subsequently traveled east to later become established in China, whereas those giving rise to the Ellwangen-Observance lineage did not (Figure 3). Once in the Former Soviet Union, plague likely became established in rodent populations in an area accessible to western Russia and evolved locally, as evidenced by the single unique derived position in our strain from Bolgar City (Figure 2B and Figure 3). Given that all modern Branch 1 lineages descend from a close hypothetical relative of our Russian strain, these European forms may well have given rise to the third plague pandemic in China and beyond.

Consensus has not yet been reached regarding the role played by the Russian region in the introduction of plague into Europe during the Black Death (Alexander, 1980; Benedictow, 2004; McNeill, 1998; Norris, 1977; Schmid et al., 2015). Drawing

upon historical and climatic data, scholars have adopted a "proximal origin" theory, which states that the Black Death erupted from plague foci in the Caucasus and neighboring areas (Alexander, 1980; Benedictow, 2004; Norris, 1977; Sussman, 2011; Varlik, 2015). Molecular investigations of the plague bacillus, however, have pointed to China as both the birthplace of *Y. pestis* itself and the origin of the Black Death (Cui et al., 2013; Morelli et al., 2010). This is difficult to reconcile with the strong East Asian sampling bias of the available data, coupled with the fact that the second most basal *Y. pestis* lineage sampled thus far stems from a rodent focus in the Former Soviet Union (Cui et al., 2013) (Figure 3). In our current investigation, we attempted to partially overcome this limitation by integrating recently sequenced strains from the Caucasus region (Zhgenti et al., 2015) in our *Y. pestis* phylogeny. To our surprise, these strains grouped with some lineages previously thought to be mostly or entirely restricted to China (Figure 2A). We therefore highlight the need to expand the sampling region of both modern and ancient *Y. pestis* to establish a more comprehensive understanding of its evolutionary history and modern ecology.

Our plague strain from the German city of Ellwangen is ancestral to those associated with the Great Plague of Marseille (L'Observance), an epidemic that occurred in France some 200 years later (Figure 2B). This branch descends directly from the strain circulating in both London and Barcelona during the Black Death and does not possess the additional Branch 1 positions present in the London 6330 and Bolgar lineages described above. That the Ellwangen genome shares 20 positions with the Marseille strain and has three unique positions (Table S4) suggests the two share a common genetic history and diverged from the same source population in advance of the 16th century Ellwangen outbreak. A previous study has pointed to natural plague foci in Asia as likely sources of the multiple plague outbreaks in Europe following the Black Death (Schmid et al., 2015). An alternative model, however, proposes a local European source for plague, given the high number of documented sporadic epidemics in isolated rural areas throughout the second wave. Alpine rodent species are considered one possible reservoir (Carmichael, 2014). Both models are explored in recent aDNA analyses of post-Black Death European plague material (Bos et al., 2016; Seifert et al., 2016), though at a resolution too low to strongly favor one hypothesis over the other. Based on modern epidemiological data, no known plague foci exist within Europe; however, several foci are suspected to exist in areas along the former Silk Road, the most prolific of which are immediately to the east of the Caspian Sea (Gage and Kosoy, 2005). The geographical location of the city of Ellwangen, and the seemingly restricted outbreak here, however, makes the introduction of disease via trade routes outside of Europe unlikely. We rather view our data as more supportive of a European reservoir for the disease. As only a small rodent focus with limited exposure to a susceptible host species is thought to be theoretically sufficient to initiate a large-scale human plague epidemic (Keeling and Gilligan, 2000), plague's presence in this proposed European reservoir need not have been large. The Ellwangen-Observance lineage contains no known extant descendants; hence, this focus may no longer exist (Figure 2B), and its extinction may have coincided with the sudden disappearance of plague in Europe. The popular theory of an 18th century domestic rodent

replacement of *Rattus rattus* by *Rattus norvegicus* (Appleby, 1980) could still carry some traction. The black rat is a well-known harbinger of plague in several locations where *Y. pestis* infections persist today (Duplantier et al., 2005; Vogler et al., 2011), and though brown Norway rats have a similar susceptibility to plague infection (Anderson et al., 2009), their different ecological niche and comparatively reduced contact with humans in a domestic setting may have slowed the transmission of disease entering from a neighboring sylvatic population.

Our phylogeny is compatible with popular demographic scenarios wherein the Black Death cycled through the Mediterranean (Barcelona), spread to Northern Europe (London), subsequently traveled east into Russia (Bolgar), and eventually made its way into China, its presumed origin and ultimate source of the modern plague pandemic (Figure 3). The most parsimonious interpretation of our data holds that, in the course of its travels, a minimum of one plague lineage was left behind along its route that persisted long enough to later diversify and give rise to at least two subsequent epidemics—one in 16th century Germany and one in 18th century France (Bos et al., 2016). The above proposal, however, is unlikely to explain the full spectrum of *Y. pestis* diversity and plague epidemics during the notorious so-called "second wave" plague pandemic; a unidirectional dispersal of *Y. pestis* is unlikely, as multiple factors are sure to have contributed to its spread in humans and other host species. The epidemics in Germany and France, for example, stemmed from only one of possibly several historical plague foci within Europe or its vicinity. Concurrent plague foci harboring strains related to our Bolgar lineage, to the lineage identified in late 14th century London, or potentially others not yet identified may have been responsible for additional second wave plague outbreaks. Currently there is a lack of ancient *Y. pestis* data from the proposed entry and end points of the Black Death in Europe (Gottfried, 1983). Genetic analyses of putative plague material from these regions would be essential in unraveling additional key features related to the paths traveled by the Black Death and the legacy it left behind.

EXPERIMENTAL PROCEDURES

Array Design and Captures

A one-million-feature Agilent microarray was designed with an in-house probe design software using the chromosome of *Yersinia pseudotuberculosis* (NCBI: NC_006155) (Chain et al., 2004), as well as the *Y. pestis* (CO92) plasmids pMT1 (NCBI: NC_003134) and pCD1 (NCBI: NC_003131). DNA extracts from *pla*-positive samples (Supplemental Experimental Procedures) were turned into double-stranded DNA libraries as described before (Meyer and Kircher, 2010). Serial hybridization-based array capture was performed using previously established methods (Hodges et al., 2009) (Supplemental Experimental Procedures).

High-Throughput Sequencing and Read Processing

Following high-throughput sequencing on Illumina platforms, all pre-processing mapping and genotyping steps were performed using the automated pipeline EAGER (Peltzer et al., 2016). For SNP filtering, the MultiVCFAnalyzer custom java program was applied to all *vcf* files to comparatively filter all detected SNPs (Supplemental Experimental Procedures).

Phylogenetic Reconstruction

A SNP table was used as input for phylogenetic reconstruction. Phylogenetic trees were generated using the Maximum Parsimony (MP) and Maximum Likelihood (ML) methods available in MEGA6.06 (Tamura et al., 2013),

discarding alignment columns with more than 5% missing data. The three newly reconstructed *Y. pestis* strains from Barcelona, Bolgar City, and Ellwangen were analyzed alongside seven previously sequenced historical strains from the second plague pandemic (Bos et al., 2011, 2016) and 141 published modern *Y. pestis* strains (Cui et al., 2013; Zhgenti et al., 2015). A *Y. pseudotuberculosis* strain (IP32953) (Chain et al., 2004) was used as outgroup for rooting the tree, and all its derived SNPs were removed to scale branch lengths (Supplemental Experimental Procedures).

ACCESSION NUMBERS

Raw sequencing reads produced for this study have been deposited at the European Nucleotide Archive (ENA) under accession number ENA: PRJEB13664.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2016.05.012.

AUTHOR CONTRIBUTIONS

J.K., K.I.B, A.H., and M.A.S. conceived the study; M.A.S., R.I.T., M.F., and K.I.B. performed laboratory work; M.A.S., A.H., K.I.B., and J.K. analyzed data; J.B.d.H., S.A., D.C., J.W., I.R.G., A.G.S., and D.K.N. provided archaeological material and archaeological context information; J.D., S.K., D.C., J.W., and I.R.G. performed anthropological and paleopathological examination; M.A.S., K.I.B., A.H., and J.K. wrote the manuscript with contribution from all co-authors.

REFERENCES

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. (1999). Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. USA *96*, 14043–14048.

Alexander, J.T. (1980). Bubonic plague in early modern Russia: public health and urban disaster (Johns Hopkins University Press).

Anderson, D.M., Ciletti, N.A., Lee-Lewis, H., Elli, D., Segal, J., DeBord, K.L., Overheim, K.A., Tretiakova, M., Brubaker, R.R., and Schneewind, O. (2009). Pneumonic plague pathogenesis and immunity in Brown Norway rats. Am. J. Pathol. *174*, 910–921.

Appleby, A.B. (1980). The disappearance of plague: a continuing puzzle. Econ. Hist. Rev. *33*, 161–173.

Benedictow, O.J. (2004). The Black Death, 1346-1353: the complete history (Boydell & Brewer).

Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., et al. (2011). A draft genome of Yersinia pestis from victims of the Black Death. Nature *478*, 506–510.

Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., et al. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature *514*, 494–497.

Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M., et al. (2016). Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. eLife *5*, 5.

Carmichael, A.G. (2014). Plague Persistence in Western Europe: A Hypothesis. The Medieval Globe *1*, 8.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., 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.

Cohn, S.K. (2008). Epidemiology of the Black Death and Successive Waves of Plague. In Medical History, *Volume 52* (Cambridge University Press).

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu, L., et al. (2013). Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Proc. Natl. Acad. Sci. USA *110*, 577–582.

Duplantier, J.M., Duchemin, J.B., Chanteau, S., and Carniel, E. (2005). From the recent lessons of the Malagasy foci towards a global understanding of the factors involved in plague reemergence. Vet. Res. *36*, 437–453.

Gage, K.L., and Kosoy, M.Y. (2005). Natural history of plague: perspectives from more than a century of research. Annu. Rev. Entomol. *50*, 505–528.

Gottfried, R.S. (1983). The Black Death: Natural and Human Disaster in Medieval Europe (Simon & Schuster).

Haensch, S., Bianucci, R., Signoli, M., Rajerison, M., Schultz, M., Kacki, S., Vermunt, M., Weston, D.A., Hurst, D., Achtman, M., et al. (2010). Distinct clones of Yersinia pestis caused the black death. PLoS Pathog. *6*, e1001134.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., and Hannon, G.J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat. Protoc. *4*, 960–974.

Keeling, M.J., and Gilligan, C.A. (2000). Bubonic plague: a metapopulation model of a zoonosis. Proc. Biol. Sci. *267*, 2219–2230.

McNeill, W.H. (1998). Plagues and Peoples (Anchor Press).

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, t5448, http://dx.doi.org/10.1101/pdb.prot5448.

Morelli, G., Song, Y., Mazzoni, C.J., Eppinger, M., Roumagnac, P., Wagner, D.M., Feldkamp, M., Kusecek, B., Vogler, A.J., Li, Y., et al. (2010). Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. *42*, 1140–1143.

Norris, J. (1977). East or west? The geographic origin of the Black Death. Bull. Hist. Med. *51*, 1–24.

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

Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M., et al. (2015). Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. Cell *163*, 571–582.

Schmid, B.V., Büntgen, U., Easterday, W.R., Ginzler, C., Walløe, L., Bramanti, B., and Stenseth, N.C. (2015). Climate-driven introduction of the Black Death

and successive plague reintroductions into Europe. Proc. Natl. Acad. Sci. USA 112, 3020–3025.

Schuenemann, V.J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J., 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, E746–E752.

Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., et al. (2013). Genome-wide comparison of medieval and modern Mycobacterium leprae. Science 341, 179–183.

Seifert, L., Wiechmann, I., Harbeck, M., Thomas, A., Grupe, G., Projahn, M., Scholz, H.C., and Riehm, J.M. (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.

Signoli, M., Bello, S., and Dutour, O. (1998). [Epidemic recrudescence of the Great Plague in Marseille (May-July 1722): excavation of a mass grave]. Med. Trop. (Mars.) 58 (2, Suppl), 7–13.

Sussman, G.D. (2011). Was the black death in India and China? Bull. Hist. Med. 85, 319–355.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Tikhomirov, E. (1999). Epidemiology and distribution of plague. In Plague Manual: Epidemiology, Distribution, Surveillance and Control (World Health Organization).

Varlik, N. (2015). Plague and Empire in the Early Modern Mediterranean World (Cambridge University Press).

Vogler, A.J., Chan, F., Wagner, D.M., Roumagnac, P., Lee, J., Nera, R., Eppinger, M., Ravel, J., Rahalison, L., Rasoamanana, B.W., et al. (2011). Phylogeography and molecular epidemiology of Yersinia pestis in Madagascar. PLoS Negl. Trop. Dis. 5, e1319.

Wagner, D.M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J.W., Enk, J., Birdsell, D.N., Kuch, M., Lumibao, C., et al. (2014). Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. Lancet Infect. Dis. 14, 319–326.

Zhgenti, E., Johnson, S.L., Davenport, K.W., Chanturia, G., Daligault, H.E., Chain, P.S., and Nikolich, M.P. (2015). Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus Region. Genome Announc. 3, 3.

Supplemental Information

Historical *Y. pestis* Genomes Reveal

the European Black Death as the Source

of Ancient and Modern Plague Pandemics

Maria A. Spyrou, Rezeda I. Tukhbatova, Michal Feldman, Joanna Drath, Sacha Kacki, Julia Beltrán de Heredia, Susanne Arnold, Airat G. Sitdikov, Dominique Castex, Joachim Wahl, Ilgizar R. Gazimzyanov, Danis K. Nurgaliev, Alexander Herbig, Kirsten I. Bos, and Johannes Krause

Supplemental Figures and Tables

Figure S1

Supplemental Figure 1, Related to Figure 1. Archaeological artifacts associated with the Bolgar City site.
(A-L) Coin artifacts associated with the burial of individual 2370 from Bolgar City, Ust'Jerusalem tomb, were minted and released during the times of Murad Khan and Abdullah Khan ibn Uzbeg Khan and date the burial to the second half of the 14th century, in the period after 1362 AD.

Figure S2
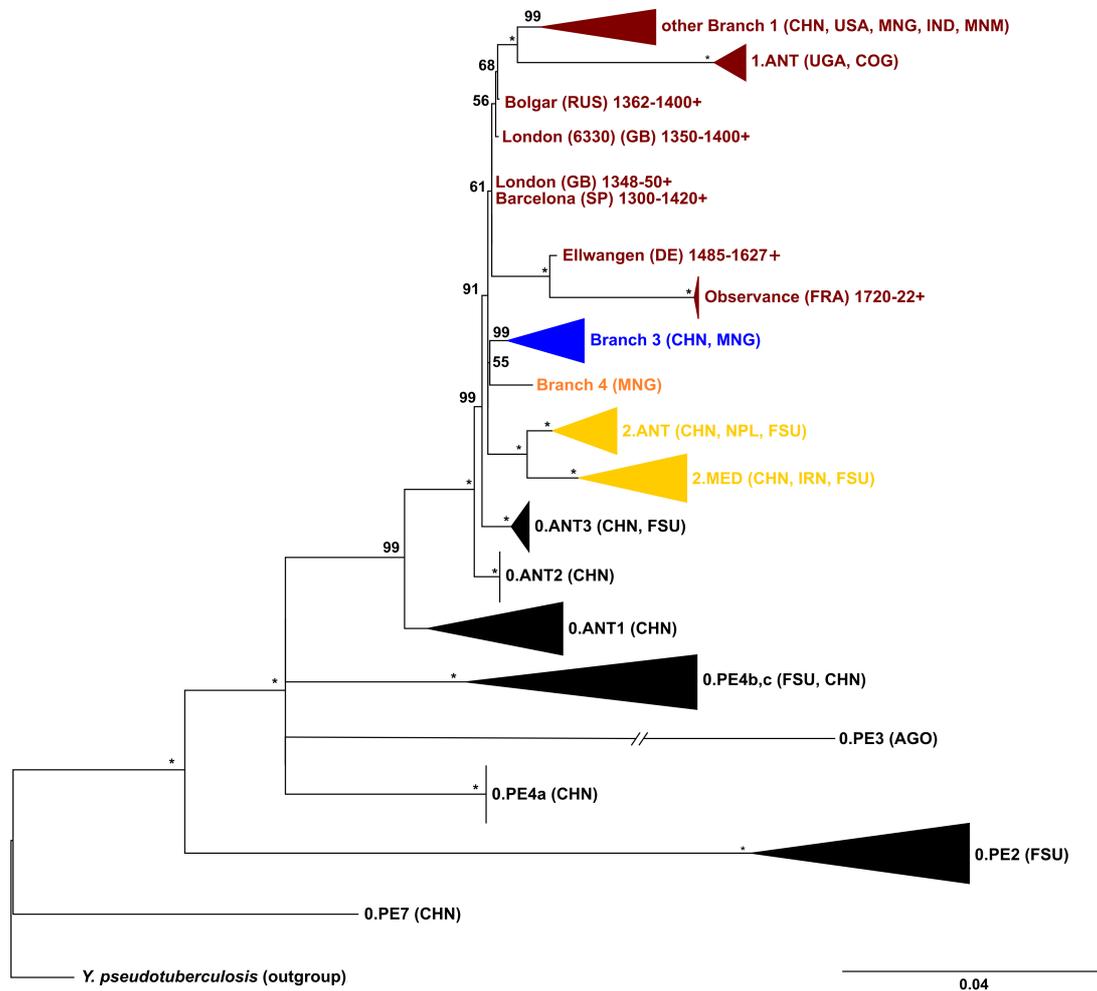
Supplemental Figure 2, Related to Figure 2 and Figure 3. Maximum Likelihood phylogeny.
Maximum Likelihood phylogenetic tree of 141 modern and 10 historical *Y. pestis* strains. A *Y. pseudotuberculosis* strain was used as outgroup (IP32953). 3351 SNP positions were considered for the phylogeny. The reconstructed tree shows the topology of the newly sequenced isolates from Barcelona, Bolgar City and Ellwangen relative to previously sequenced modern and ancient *Y. pestis* strains. Asterisks (*) indicate bootstrap values of 100. Collapsed branches are represented by triangles, to enhance tree clarity. Strains belonging to Branch 1 are represented in red, Branch 2 is represented in yellow, Branch 3 is represented in blue, Branch 4 is represented in orange and Branch 0 is represented in black. Ancient Branch 1 strains are indicated by their archaeological or radiocarbon date and by a (+). Because of the great number of derived SNP positions of the 0.PE3 lineage, its branch was reduced to adjust scaling of the tree. Geographic region abbreviations correspond to: CHN (China), USA (United States of America), MDG (Madagascar), IND (India), IRN (Iran), MNM (Myanmar), RUS (Russia), GB (Great Britain), DE (Germany), FRA (France), SP (Spain), MNG (Mongolia), NPL (Nepal), FSU (Former Soviet Union), AGO (Angola), CGO (Congo) and UGA (Uganda).

Table S1

Supplemental Table 1, Related to Figure 1 and Table S2. Sample description, screening results and sequencing statistics.

Sample ID	Site	Tooth type	Dates cal AD	Archaeological dating	Copies of pla (copies/µl)	CO92 Chromosomal fold-coverage
3031	Barcelona	Molar	1300-1420[a]	-	174	10.3x
2370	Bolgar City	Molar	1298-1388[b]	1362-1400 AD[c]	553	19.3x
549_O	Ellwangen	Molar/ Incisor	1485-1627[b]	-	3/2	4.9x[d]

[a]Published in (Beltrán de Heredia Bercero, 2014)
[b]Radiocarbon dates produced for this study, presented in calibrated year AD (1 sigma)
[c]Archaeological dates point to the second half of the 14th century, and specifically the period after 1362 AD (see also Supplemental Experimental Procedures and Figure S1)
[d]Pooled data from four independent serial capture experiments, using two libraries generated from two tooth samples belonging to the same individual

Table S2

Supplemental Table 2, Related to Figure 2. Read processing and mapping statistics

Sample ID	Site	Number of pre-processed reads before mapping	Number of mapped reads	Duplication factor	Number of mapped reads after map quality filtering	Average coverage on CO92 Chromosome	Percentage of CO92 chromosome covered 3-fold
3031	Barcelona	60,516,261	2,116,962	2.43	870,477	10.3	89.88%
2370	Bolgar	47,595,565	4,183,226	2.49	1,682,954	19.28	91.46%
549_O	Ellwangen	145,645,509	2,120,360	4.79	442,573	4.91	73.56%

Supplemental Experimental Procedures

Information on archaeological sites and aDNA specimens

Saints Màrtirs Just i Pastor, Barcelona, Spain
The Barcelona mass grave was discovered in 2012 during an excavation campaign performed in the sacristy of the Saints Màrtirs Just i Pastor church, and is likely the first Black Death burial discovered in Spain. The burial was only partially excavated due to security restrictions (i.e. it's proximity to the sacristy's walls) and its western part was destroyed by the foundations of the gothic church. The pit was 1.50 m deep, and estimated to have been 5.5 m long and 4 m wide. Individuals were deposited in 11 successive layers, each layer containing between 4 and 15 skeletons. During the excavation, the skeletons of 120 individuals were recovered, of which 50 non-adults (i.e. less than 20 year-old) and 70 adults, out of an estimated number of 350-400 individuals buried in the pit (Beltrán de Heredia Bercero, 2014). Due to partial excavation of the burial, most of the skeletons recovered were incomplete. Most of the individuals were lying on their back, with their upper limbs in a flexed position (hands on the thorax or abdomen), and their lower limbs extended. None of the individuals were in prone position, and only a few were on their side. Regarding burial methods, no coffins were used. Bodies were directly placed in the grave all together, and the pit was subsequently filled with earth. The excavation, however, revealed textile remains (linen and hemp), suggesting that bodies were wrapped in shrouds. It is noteworthy that a layer of lime was deposited on top of the accumulation of cadavers; this is the only one example of use of such prophylactic material in a medieval plague pit (Kacki, 2014; Schotsmans et al., 2015). For this study, 49 teeth were removed from 18 individuals and used for biomolecular analysis. According to osteological examination of long bone length and dental development, individual 3031 from Barcelona, from which a *Y. pestis* genome was recovered, was most likely a 6-9 year old non-adult, whose sex could not be determined. Although the skeleton was only partly preserved, there was no detectably diagnostic feature of a specific pathological condition. Stress markers such as *cribra orbitalia* and mild developmental dental defects (linear enamel hypoplasia), however, may indicate malnutrition earlier in life.

Ust'-Jerusalem necropolis and Bolgar City mausoleum, Russia
The medieval city of Bolgar was situated on the bank of the Volga River, 30 km downstream from its confluence with the Kama River and some 130 km from modern Kazan (Tatarstan, Russian Federation). Bolgar City was an early settlement of the civilization of Volga-Bolgars, which existed between the 7th and 15th centuries AD and was intermittently capital of Volga Bulgaria between the 10th and 15th centuries (Sitdikov, 2014). Bolgar City was also the first capital of the Golden Horde in the 13th century. The UNESCO World Heritage Committee declared the ancient Bolgar hill fort as a World Heritage Site in 2014. The Ust'-Jerusalem necropolis was excavated between 1996 and 2003, covers an 800 sq. m. area and includes 318 single burials (Vasiliev, 2004). Palaeodemographic analysis revealed a high infant mortality rate (over 57% of the group), which may have been attributed to unfavorable social and environmental conditions, early childbirth in women, and a significant lack of food resources (Boruckaya, 2003). For the current investigation, material was chosen from the anthropological collections of the Ust'-Ierusalimsky tombs (Figure 1b) and the Bolgar city mausoleum (Vasiliev, 2004). A total of 95 teeth were extracted from 93 individuals and used for ancient DNA analysis. A complete skeleton was recovered from the plague victim (2370) of the Ust'Jerusalem tomb in Bolgar City. Anthropological analysis revealed a 35-40 year old male, whose burial was not consistent with medieval Muslim funerary practices (Figure 1b). Coin artifacts that were discovered during excavation of the burial dated the site to the second half of the 14th century. Grave artifacts associated specifically with individual 2370 consisted of 12 silver coins (Figure S1), the earliest of which date to 1362 AD. Such coin types are considered to have been minted and released during the times of Murad Khan and Abdullah Khan ibn Uzbeg Khan of the Golden Horde (Bosworth, 1996), who ruled between 1362 and 1370 AD (Figure S1).

"Marktplatz" Ellwangen, Germany
Excavations of the Ellwangen "Marktplatz" that were initiated as a tribute to the 1250th anniversary of the city, revealed a burial ground proposed to have been used for about a millennium, with human remains unearthed spanning form the 8th to the 18th centuries AD (Arnold, 2014). The cemetery contained 3 mass graves and 14 multiple burials amongst single burials. To-date, a total of 800 individuals has been identified and unearthed. The mass graves included a total of 102 individuals, and presented a case of unstructured burial practices clearly reflecting an event of mass mortality (Wahl, 2014). It is possible that part of the multiple burials were also attributed to the same catastrophic event.

No signs of warfare were detected in the remains, and skeletal indications of infectious disease were unspecific and not uniform across the individuals. Microscopical analyses detected the presence of intestinal parasites among skeletal material, possibly indicating unhygienic living conditions of the population. Within the mass graves, 80% of individuals were determined to be non-adults (<20 years) with an average age of 9.4 years (Wahl, 2014). In addition, the multiple burials contained a total of 73, mostly incomplete, skeletons that were distributed across 14 graves with 2 to 10 individuals in each grave. In this case the average age was much higher, estimated to 17.4 years. For the present study, 79 teeth were removed from 67 individuals for ancient DNA analysis. Individual 549_O from Ellwangen, from which a *Y. pestis* genome was reconstructed, was identified as a 12-14 year old non-adult, whose sex could not be confidently determined. In this case, the skeletal material recovered was also incomplete, with non-specific bone changes and dental defects, including calculus formation (Figure 1c).

DNA Extraction from archaeological material

Extractions were performed for a total of 223 tooth samples, isolated from potential plague victims. 50 mg of pulverized dental pulp, was removed using a dental drill, as preserved pathogen DNA is more likely to reside in the dried blood vessels of the pulp chamber (Schuenemann et al., 2011). All procedures were carried out in the dedicated ancient DNA laboratory of Paleogenetics in the University of Tübingen. DNA extraction was performed according to a previously described protocol (Dabney et al., 2013), with a rotation of 12-16h at 37°C during an initial lysis step. A negative control was included for every 10 samples, and one positive extraction control for every extraction slot.

Screening for *pla*

Initial screening was performed to evaluate the presence of *Y.pestis* DNA in all samples, by using the species-specific gene plasminogen activator (*pla*). 223 DNA extracts were qPCR screened for the presence of the *pla* gene, located in the pPCP1 plasmid using a previously described approach (Schuenemann et al., 2011). 79 samples were from Ellwangen (67 individuals), 49 from Barcelona (18 individuals), 95 from Bolgar city (93 individuals). Amplification products were not sequenced. Potential *Y. pestis* positive samples were subsequently used for further screening using an established whole-genome array capture method (Hodges et al., 2009).

Array design

A one million feature Agilent microarray (Hodges et al., 2009) was designed using an in-house probe design software, both for screening and whole-genome reconstruction purposes. In order to prevent hybridization capture bias for a certain *Y. pestis* reference sequence, probes were prepared using the chromosome of the bacterium *Yersinia pseudotuberculosis* (Accession number: NC_006155), which is the closest known relative of *Y. pestis*, with up to 97% identity in chromosomal genes (Achtman et al., 1999; Chain et al., 2004). The *Y. pestis* (CO92) plasmids pMT1 (Accession number: NC_003134) and pCD1 (Accession number: NC_003131) were also included in the design. A complete set of 976,658 probes was generated for the array.

Array captures

60 µl of extract was used to produce double stranded DNA libraries, as described before (Meyer and Kircher, 2010), for all positive samples. Blank library controls and extraction blanks were also included in every library preparation slot. As the characteristic cytosine deamination accumulating within DNA molecules over time may challenge downstream analyses (Briggs et al., 2007; Sawyer et al., 2012), an initial UDG and endonuclease VIII treatment (USER enzyme) was used to remove uracil residues and subsequently repair DNA fragments (Briggs et al., 2010). Unique double index DNA barcodes (Kircher et al., 2012) were attached onto libraries through a 10-cycle amplification reaction, using universal IS5/IS6 primers. Post indexing, libraries were amplified using AccuPrime *Pfx* or Herculase II Fusion DNA polymerase to accomplish a 19 µg pool of samples. 1ug of a positive control was added to make up a final 20 µg pool that served as template for array capture. A separate pool was made for extraction and library blank controls, and was tested on a separate array to avoid cross talk between samples and blanks. Hybridization-based array capture was performed using previously established methods (Hodges et al., 2009). Hybridization of template to the probes took place over a two-night incubation step at 65 °C. Following array elution, captured template was re-amplified with universal IS5/IS6 primers using Herculase II Fusion DNA polymerase to achieve 20 µg of product that would serve as template for a subsequent capture step (serial capture). Identical array design and methodology were used for the second round of capture. Following elution, products were again

amplified as described above and subsequently diluted down to 10nM for high throughput sequencing. Sequencing was performed on HiSeq 2500 and NextSeq 500 Illumina platforms.

High throughput read pre-processing and mapping pipeline
High throughput sequencing produced up to 123,690,558 raw paired-end reads per sequencing run per library. All pre-processing, mapping and genotyping steps were performed using the automated pipeline EAGER (Peltzer et al., 2016). Adaptors were clipped from all reads produced by the Illumina platforms and overlapping reads were merged. Subsequent quality filtering and length filtering removed reads shorter that 30 bp. All reads were then mapped with BWA (Li and Durbin, 2010) using *Y. pestis* CO92 as a reference genome (Parkhill et al., 2001), the first complete *Y. pestis* genome to be sequenced (Accession number: AL590842.1). To avoid cross mapping from multiple DNA sources, mapping parameters used included a stringency of 0.1 (-n parameter) and a map quality filter of 37 (Table S1).

SNP calling
SNP calling was performed using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK) (DePristo et al., 2011) on the newly produced ancient mapped data, alongside previously published ancient and modern *Y. pestis* data. A total of 152 samples were considered for this study. A *vcf* file was produced for every sample using the "EMIT_ALL_SITES" option, which generates a call for every genomic site. The MultiVCFAnalyzer custom java program was applied to all *vcf* files to comparatively filter all the detected SNPs, and produce a multiple alignment of variable positions in which a SNP was called when present in at least one of the samples in the dataset. Homozygous SNPs were called when covered at least 3-fold with a minimum genotyping quality of 30. Respectively, a reference base was called when supported by 3 independent fragments with the same quality threshold. In case of a heterozygous position, a SNP or reference base was called when at least 90% of the reads covering the position support it. If none of the options was possible, an "N" was inserted at the corresponding positions. In the current dataset, a total of 3,444 variant positions were called.

Phylogenetic reconstruction
After SNP filtering, a SNP table was used as input for phylogenetic reconstruction encompassing a total of 3,444 SNP positons. Phylogenetic trees were generated, using the Maximum Parsimony (MP) and Maximum Likelihood (ML) methods available in MEGA6.06 (Tamura et al., 2013), discarding all alignment columns with more than 5% missing data, which caused the removal of 93 SNP positions. The Maximum Likelihood phylogeny was inferred assuming a General Time Reversible (GTR) model and a Nearest-Neighbor-Interchange (NNI) tree inference option. The total amount of remaining positions to be considered was 3,351. 1000 pseudo-replicates were carried out to assess tree robustness by the bootstrapping method for both phylogenetic methods. A total of 152 samples were used for generation of the phylogeny. The three new strains reported in this study include a Black Death strain from Barcelona, a strain from Bolgar city dating to the second half of the 14th century AD (1362-1400), and a 16th century AD strain from Ellwangen. Data from previously sequenced ancient and modern *Y. pestis* strains included five strains from the 18th century plague of Marseille (Bos et al., 2016), one sequence from Black Death victims from London 1348-1350 (8291-1197-8124), one strain from London 1350-1400 (6330) and 141 modern *Y. pestis* strains (Cui et al., 2013; Zhgenti et al., 2015). A previously published ancient strain recovered from victims of the Plague of Justinian (Wagner et al., 2014) was omitted from the analysis, as it does not contribute to the interpretation of our results, and the low coverage of this genome could negatively influence the robustness of our phylogeny. A *Y. pseudotuberculosis* strain (IP32953) (Chain et al., 2004) was used as an outgroup for rooting the tree, and all its derived SNPs were removed to scale branch lengths.

Supplemental References

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. (1999). Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proceedings of the National Academy of Sciences of the United States of America *96*, 14043-14048.

Arnold, S. (2014). Ellwangen, Ostalbkreis; Die sanierung des Marktplatzes in Ellwangen-nicht enden wollende überraschungen (Darmstadt: Theiss).

Beltrán de Heredia Bercero, J., Gibrat Pineda, I. (2014). El primer testimoni arqueològic de la pesta negra a Barcelona: la fossa comuna de la Basílica dels Sants Màrtirs Just i Pastor. Quaderns d'Arqueologia i Història de la Ciutat de Barcelona *10*, 164-179.

Boruckaya, S.B. (2003). Analysis of the physical development of the population, which left the Ust'-Jerusalem Cemetery (Bolgar city, Tatarstan) Ecology of ancient and modern communities *2*, 212-214.

Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M.*, et al.* (2016). Eighteenth century genomes reveal the long-term persistence of an historical plague focus. eLife *5*.

Bosworth, C.E. (1996). The new Islamic dynasties (Columbia University Press).

Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M.*, et al.* (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the United States of America *104*, 14616-14621.

Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Paabo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic acids research *38*, e87.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L.*, et al.* (2004). Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proceedings of the National Academy of Sciences of the United States of America *101*, 13826-13831.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu, L.*, et al.* (2013). Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Proceedings of the National Academy of Sciences of the United States of America *110*, 577-582.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L.*, et al.* (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences of the United States of America *110*, 15758-15763.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M.*, et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics *43*, 491-498.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., and Hannon, G.J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nature protocols *4*, 960-974.

Kacki, S., Castex, D. (2014). La sépulture multiple de la basilique des Saints Martyrs Just et Pastor : bio-archéologie des restes humains. Quaderns d'Arqueologia i Història de la Ciutat de Barcelona *10*, 180-199.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic acids research *40*, e3.

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

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor protocols *2010*, pdb prot5448.

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L.*, et al.* (2001). Genome sequence of Yersinia pestis, the causative agent of plague. Nature *413*, 523-527.

Peltzer, A., Jager, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome biology *17*, 60.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Paabo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PloS one *7*, e34131.

Schotsmans, E.M., Van de Vijver, K., Wilson, A.S., and Castex, D. (2015). Interpreting lime burials. A discussion in light of lime burials at St. Rombout's cemetery in Mechelen, Belgium (10th–18th centuries). Journal of Archaeological Science: Reports *3*, 464-479.

Schuenemann, V.J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J*., et al.* (2011). Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. Proceedings of the National Academy of Sciences of the United States of America *108*, E746-752.

Sitdikov, A.G., Valiev, R. R., Starkov, A. S. (2014). Archaeological investigations of Bolgar and Sviyazhsk in 2013 (in Russian) (Kazan: Institute of Archaeology).

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution *30*, 2725-2729.

Vasiliev, S., Boruckaya, S. B. (2004). Paleodemographics of the Ust'-Jerusalem tomb (Bolgar city). The Antiquity and the Middle Ages of Volga-Kama region Materials of the Third Halikov's Readings (in Russian), 38-40.

Wagner, D.M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J.W., Enk, J., Birdsell, D.N., Kuch, M., Lumibao, C*., et al.* (2014). Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis. The Lancet Infectious Diseases *14*, 319-326.

Wahl, J., Immel, A., Krause, J. (2014). Ellwangen, Ostalbkreis; Naturwissenschaftliche Untersuchungen zur Ausgrabung in Ellwangen (Darmstadt: Theiss).

Zhgenti, E., Johnson, S.L., Davenport, K.W., Chanturia, G., Daligault, H.E., Chain, P.S., and Nikolich, M.P. (2015). Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus Region. Genome announcements *3*.

Table S3, Related to Figure 2 and Figure 3. Geographical regions of *Y. pestis* genomes used for SNP calling and phylogeny

Strain identifier	Branch	Lineage	Country		
				Newly integrated Caucasus strains	
				Newly sequenced historical strains	
0.ANT1a_42013	0	0.ANT1	China		
0.ANT1b_CMCC49003	0	0.ANT1	China		
0.ANT1c_945	0	0.ANT1	China		
0.ANT1d_164	0	0.ANT1	China		
0.ANT1e_CMCC8211	0	0.ANT1	China		
0.ANT1f_42095	0	0.ANT1	China		
0.ANT1g_CMCC42007	0	0.ANT1	China		
0.ANT1h_CMCC43032	0	0.ANT1	China		
0.ANT2_B42003004	0	0.ANT2	China		
0.ANT2a_2330	0	0.ANT2	China		
0.ANT3a_CMCC38001	0	0.ANT3	China		
0.ANT3b_A1956001	0	0.ANT3	China		
0.ANT3c_42082	0	0.ANT3	China		
0.ANT3d_CMCC21106	0	0.ANT3	China		
0.ANT3e_42091b	0	0.ANT3	China		
Kyrgyzstan_790	0	0.ANT3	Kyrgyzstan		
0.PE2_PEST-F	0	0.PE2	Georgia		
0.PE2b_G8786	0	0.PE2	Georgia		
Armenia_14735	0	0.PE2	Armenia		
Armenia_1522	0	0.PE2	Armenia		
Georgia_1412	0	0.PE2	Georgia		
Georgia_1413	0	0.PE2	Georgia		
Georgia_1670	0	0.PE2	Georgia		
Georgia_3067	0	0.PE2	Georgia		
Georgia_3770	0	0.PE2	Georgia		
Georgia_8787	0	0.PE2	Georgia		
0.PE3_Angola	0	0.PE3	Angola		
0.PE4_Microtus91001	0	0.PE4	China		
0.PE4Aa_12	0	0.PE4	China		
0.PE4Ab_9	0	0.PE4	China		
0.PE4Ba_PestoidesA	0	0.PE4	Georgia		
0.PE4Ca_CMCCN010025	0	0.PE4	China		
0.PE4Cc_CMCC18019	0	0.PE4	China		
0.PE4Cd_CMCC93014	0	0.PE4	China		
0.PE4Ce_CMCC91090	0	0.PE4	China		
M0000002	0	0.PE4	China		
0.PE7b_620024	0	0.PE7	China		
1.ANT1_Antiqua	1	1.ANT1	Congo		
1.ANT1_UG05-0454	1	1.ANT1	Uganda		
1.IN1a_CMCC11001	1	1.IN1	China		
1.IN1b_780441	1	1.IN1	China		
1.IN1c_K21985002	1	1.IN1	China		
1.IN2a_CMCC640047	1	1.IN2	China		
1.IN2b_30017	1	1.IN2	China		
1.IN2c_CMCC31004	1	1.IN2	China		
1.IN2d_C1975003	1	1.IN2	China		
1.IN2e_C1989001	1	1.IN2	China		
1.IN2f_710317	1	1.IN2	China		
1.IN2g_CMCC05013	1	1.IN2	China		
1.IN2h_5	1	1.IN2	China		
1.IN2i_CMCC10012	1	1.IN2	China		
1.IN2j_CMCC27002	1	1.IN2	China		
1.IN2k_970754	1	1.IN2	China		
1.IN2l_D1991004	1	1.IN2	China		
1.IN2m_D1964002b	1	1.IN2	China		
1.IN2n_CMCC02041	1	1.IN2	China		
1.IN2o_CMCC03001	1	1.IN2	China		
1.IN2p_D1982001	1	1.IN2	China		
1.IN2q_D1964001	1	1.IN2	China		
1.IN3a_F1954001	1	1.IN3	China		
1.IN3b_E1979001	1	1.IN3	China		
1.IN3c_CMCC84038b	1	1.IN3	China		
1.IN3d_YN1683	1	1.IN3	China		
1.IN3e_YN472	1	1.IN3	China		
1.IN3f_YN1065	1	1.IN3	China		
1.IN3g_E1977001	1	1.IN3	China		
1.IN3h_CMCC84033	1	1.IN3	China		
1.IN3i_CMCC84046	1	1.IN3	China		
1.ORI1_CA88	1	1.ORI1	U.S.A		
1.ORI1_CO92	1	1.ORI1	U.S.A		
1.ORI1a_CMCC114001	1	1.ORI1	China		
1.ORI1b_India195	1	1.ORI1	India		
1.ORI1c_F1946001	1	1.ORI1	China		
1.ORI2_F1991016	1	1.ORI2	China		
1.ORI2a_YN2179	1	1.ORI2	Myanmar		
1.ORI2c_YN2551b	1	1.ORI2	China		
1.ORI2d_YN2588	1	1.ORI2	China		
1.ORI2f_CMCC87001	1	1.ORI2	China		
1.ORI2g_F1984001	1	1.ORI2	China		

1.ORI2h_YN663	1	1.ORI2	China
1.ORI2i_CMCCK100001a	1	1.ORI2	China
1.ORI2i_CMCCK110001b	1	1.ORI2	China
1.ORI3_IP275	1	1.ORI3	Madagascar
1.ORI3_MG05-1020	1	1.ORI3	Madagascar
1.ORI3a_EV76	1	1.ORI3	Madagascar
2.ANT1_Nepal516	2	2.ANT1	Nepal
2.ANT1a_34008	2	2.ANT1	China
2.ANT1b_34202	2	2.ANT1	China
2.ANT2a_2	2	2.ANT2	China
2.ANT2b_351001	2	2.ANT2	China
2.ANT2c_CMCC347001	2	2.ANT2	China
2.ANT2d_G1996006	2	2.ANT2	China
2.ANT2e_G1996010	2	2.ANT2	China
2.ANT2f_CMCC348002	2	2.ANT2	China
2.ANT3a_CMCC92010	2	2.ANT3	China
2.ANT3b_CMCC95001	2	2.ANT3	China
2.ANT3c_CMCC96001	2	2.ANT3	China
2.ANT3d_CMCC96007	2	2.ANT3	China
2.ANT3e_CMCC67001	2	2.ANT3	China
2.ANT3f_CMCC104003	2	2.ANT3	China
2.ANT3g_CMCC51020	2	2.ANT3	China
2.ANT3h_CMCC106002	2	2.ANT3	China
2.ANT3i_CMCC64001	2	2.ANT3	China
2.ANT3j_H1959004	2	2.ANT3	China
2.ANT3k_5761	2	2.ANT3	Russian Federation
2.ANT3l_735	2	2.ANT3	Russian Federation
2.MED1b_2506	2	2.MED1	China
2.MED1c_2654	2	2.MED1	China
2.MED1d_2504	2	2.MED1	China
2.MED1_KIM10	2	2.MED1	Iran/Kurdistan
RussianFederation_2944	2	2.MED1	Russian Federation
Azerbaijan_1522	2	2.MED1	Azerbaijan
2.MED2b_91	2	2.MED2	China
2.MED2c_K11973002	2	2.MED2	China
2.MED2d_A1973001	2	2.MED2	China
2.MED2e_7338	2	2.MED2	China
2.MED3a_J1963002	2	2.MED3	China
2.MED3b_CMCC125002b	2	2.MED3	China
2.MED3c_I1969003	2	2.MED3	China
2.MED3d_J1978002	2	2.MED3	China
2.MED3f_I1970005	2	2.MED3	China
2.MED3g_CMCC99103	2	2.MED3	China
2.MED3h_CMCC90027	2	2.MED3	China
2.MED3i_CMCC92004	2	2.MED3	China
2.MED3j_I2001001	2	2.MED3	China
2.MED3k_CMCC12003	2	2.MED3	China
2.MED3l_I1994006	2	2.MED3	China
2.MED3m_SHAN11	2	2.MED3	China
2.MED3n_SHAN12	2	2.MED3	China
2.MED3o_I1991001	2	2.MED3	China
2.MED3p_CMCC107004	2	2.MED3	China
3.ANT1a_7b	3	3.ANT1	China
3.ANT1b_CMCC71001	3	3.ANT1	China
3.ANT1c_C1976001	3	3.ANT1	China
3.ANT1d_71021	3	3.ANT1	China
3.ANT2a_MGJZ6	3	3.ANT2	Mongolia
3.ANT2b_MGJZ7	3	3.ANT2	Mongolia
3.ANT2c_MGJZ9	3	3.ANT2	Mongolia
3.ANT2d_MGJZ11	3	3.ANT2	Mongolia
3.ANT2e_MGJZ3	3	3.ANT2	Mongolia
4.ANT1a_MGJZ12	4	4.ANT1	Mongolia
8124_8291_11972	1	ancient Branch 1	Great Britain
6330	1	ancient Branch 1	Great Britain
OBS107	1	ancient Branch 1	France
OBS110	1	ancient Branch 1	France
OBS116	1	ancient Branch 1	France
OBS124	1	ancient Branch 1	France
OBS137	1	ancient Branch 1	France
Barcelona (3031)	1	ancient Branch 1	Spain
Bolgar (2370)	1	ancient Branch 1	Russian Federation
Ellwangen (549_O)	1	ancient Branch 1	Germany
Y.pseudotuberculosis (IP32953)	outgroup	outgroup	France

Table S4, Related to Figure 2. SNP table including non-unique and unique SNPs of all second pandemic *Y. pestis* strains sequenced to-date

Position	Reference	OBS107	OBS110	OBS116	OBS124	OBS137	Ellwangen	Bolgar	Barcelona	London (8124_8291_11972)	London (6330)
29368	G	T	T	T	T	T	T				
74539	C	T	T	T	T	T	T	T	T	T	T
100383	C	T	T	T	T	N	.	.	.	.	.
130643	G	A	A	A	A	A	A	A	A	A	A
155747	A	G	G	G	G	G	N	G	G	G	G
169412	C	T	T	T	T	T	T	.	.	.	.
173032	C	T	T	T	T	T	.	.	.	.	.
186060	C	T	T	T	T	T	T	.	.	.	.
190040	C	N	A	N	N	N	N	.	N	.	.
190041	T	G	G	N	N	N	N	.	N	.	.
190049	G	N	N	N	N	N	N	N	A	N	N
200723	C	T	T	T	T	T	.	.	.	.	N
217009	G	T	T	T	T	T	T	.	.	.	.
225436	T	A	A	A	A	A	N	.	.	.	.
226722	C	T	T	T	T	T	.	.	.	.	N
286528	T	A	A	A	A	A	N	A	A	A	A
300041	C	N	N	N	N	T	.	.	.	.	.
325836	T	C	C	C	C	C	C	C	C	C	C
400143	G	A	A	A	A	A	A	.	.	.	.
477107	C	T	T	T	T	T	.	.	.	.	.
480773	C	T	T	T	T	T	.	.	.	.	.
482327	G	T	T	T	T	T	.	.	.	.	.
528975	A	N	N	C	N	N	.	.	.	.	.
545488	T	C	C	C	C	C	C	C	C	C	C
571183	G	N	N	N	A	N	N	N	N	N	N
699494	A	G	G	G	G	G	G	.	G	G	.
699647	T	C	C	C	C	C	C	C	C	C	C
862385	T	G	G	G	G	G	N	.	.	.	.
867712	C	A	A	A	A	A	A	.	.	.	N
868549	G	C	C	C	C	C	N	.	.	.	.
869820	A	G	G	G	G	G	N	.	.	.	.
877258	T	C	C	C	C	C	N	C	C	C	C
899158	C	T	T	T	T	T	.	.	N	.	.
951295	C	T	T	T	T	T	N	.	.	.	.
961795	C	T	N	T	T	T	T	.	.	.	N
965281	C	A	A	A	A	A	.	.	.	.	.
1017647	T	C	C	C	C	C	N	N	N	C	C
1025278	T	G	G	G	G	G	G	G	G	G	G
1098675	A	C	C	C	C	C	C	C	C	C	N
1159539	T	A	A	A	A	A	.	.	.	.	.
1168951	G	T	T	T	T	T	.	.	.	.	.
1178178	T	C	C	C	C	C	N	C	C	C	C
1178459	T	C	C	C	C	C	C	C	C	C	C
1189479	C	T	T	T	T	T	.	.	.	.	.
1232222	C	T	T	T	T	T	T	.	.	.	.
1254157	C	T	T	T	T	T	T	.	.	.	.
1272559	T	C	C	C	C	C	C	C	C	C	C
1306718	T	C	C	C	C	C	C	C	C	C	N
1308719	G	A	A	A	A	A	N	.	.	.	N
1378105	G	T	T	T	T	T	.	.	.	.	.
1385780	T	C	C	C	C	C	C	C	C	C	C
1398797	C	N	N	N	N	N	A	N	N	N	N
1439084	T	.	N	N	A	N	.	.	.	.	.
1439085	A	.	N	N	C	N	.	.	.	.	.
1440851	G	T	T	T	T	T	.	.	.	.	.
1451124	T	G	G	G	N	G	N				N
1458573	T	A	A	A	A	A	.				N
1466798	T	N	N	C	C	C	N				
1481292	C	N	T	T	N	T	.	.	.	.	N
1481381	G	A	N	N	N	N	.				
1481393	G	A	N	N	A	A	.				
1511518	A	N	N	N	N	G	.	.	.	.	.
1512930	A	G	G	G	G	G	N	G	G	G	G
1549630	A	N	G	G	G	G	N	.	.	.	N
1586982	C	A	A	A	A	A	.				
1614945	T	G	G	G	G	G	N	.	.	.	N
1644408	C	A	A	A	A	A	N	.	.	.	.
1708192	C	A	A	A	A	A	.	.	.	.	.
1713927	C	A	A	A	N	A	.	.	.	.	N
1724647	C	T	T	T	T	T	.	.	.	.	.
1735263	A	C	C	C	C	C	C	C	C	C	C
1749443	T	C	C	C	C	C	C	C	C	C	C
1808946	T	C	C	C	C	C	C	C	C	C	C
1871129	T	C	C	N	N	N	.	.	.	.	.
1883743	C	T	N	N	N	N	.	.	.	.	N
1883750	A	T	N	N	N	N	.	.	.	.	N
1935112	C	A	A	A	A	A	.	.	.	.	.
1952848	G	A	A	A	A	A	A	.	.	.	.
2022335	A	C	C	C	C	C	C	C	C	C	C
2071670	G	T	T	T	T	T	N	.	.	.	N
2076353	C	T	T	T	T	T	.	.	.	.	.
2098628	T	C	C	C	C	C	C	C	C	C	C
2105332	C	T	T	T	N	T	N	.	.	.	N
2105376	A	N	G	G	G	G	.	.	.	.	N
2262577	T	G	G	G	G	G	G	.	G	G	.
2264654	C	A	A	A	A	A	.	.	.	.	N
2277583	G	A	A	A	A	A	N	A	N	A	A
2278317	A	G	G	G	G	G	G	G	G	G	G
2281061	C	N	A	A	A	A	N	N	N	.	.
2292030	C	T	T	T	T	T	T	.	.	.	.
2300659	T	G	G	G	G	G	N	G	G	G	G
2356003	T	A	A	A	A	A	A	A	A	A	A
2414599	T	C	C	C	C	C	.	.	.	.	N
2472383	A	G	G	G	G	G	.	.	.	.	.
2507983	T	G	G	G	G	G	N	.	.	.	.
2508389	T	C	C	C	C	C	C	C	C	C	C
2519931	C	T	T	T	T	T	T	.	.	.	.
2575152	G	A	A	A	A	A	N	A	A	A	A
2596736	A	G	G	G	G	G	.	.	.	.	N
2671194	G	A	A	A	A	A	N	.	.	.	.
2684793	A	G	G	G	G	G	G	G	G	G	G
2727385	A	G	G	G	G	G	.	.	.	.	.
2739149	C	A	A	A	A	A	N	A	A	A	N
2744933	A	G	N	G	G	G	N	G	G	G	G
2877295	G	A	A	A	A	A	A	.	.	.	.
2903882	T	G	G	G	G	G	G	G	G	G	G
2913027	A	.	.	.	.	.	T	.	.	.	.
2918297	T	G	G	G	G	G	.	.	.	.	.
2934972	T	G	G	G	G	G	N	G	G	G	G
2936268	G	A	A	A	A	A	A	A	A	A	A
2950954	G	A	A	A	A	A	N	A	A	A	N
2958327	C	T	T	T	N	T	T	T	T	N	N
2964936	A	G	G	G	G	G	.	.	.	N	N
2973013	C	T	T	T	T	T	N	.	N	.	.
3025157	G	N	N	N	.	N	N	N	A	N	N
3030042	G	T	T	T	T	T	T	.	.	.	N
3085079	A	G	G	G	G	G	G	G	G	G	N

Ellwangen-Observance shared SNPs N=20
Ellwangen Unique SNPs N=3
SNP positions between Bolgar and 1.ANT N=10

Position	Ref	1	2	3	4	5	6	7	8	9	10
3098104	C	N	A	A	A	A	N	.	N	.	N
3145523	A	C	C	C	C	C	C	C	C	C	C
3190399	A	G	G	G	G	G	G	G	G	G	G
3229407	T	C	C	C	C	C	N	N	N	.	.
3244204	A	G	G	G	G	G	G	G	G	G	G
3253104	G	N	N	N	N	N	T	.	.	.	.
3254908	G	A	A	A	A	A	A	.	.	.	.
3267118	A	G	G	G	G	G	N	G	G	G	G
3269577	G	N	T	T	T	T	N	.	.	.	N
3269613	G	A	N	N	.	.	.	.	.	.	.
3269615	C	T	N	N	N	N	.	.	.	.	.
3299755	C	N	N	N	N	N	N	.	T	N	N
3324959	A	G	G	G	G	G	G	G	G	G	G
3336063	C	N	.	.	.	.	T	.	.	.	N
3362591	A	G	G	G	G	G	G	G	G	G	G
3387542	C	N	N	N	N	N	N	N	T	N	N
3397040	A	G	G	G	G	G	G	G	G	G	N
3407572	A	T	T	T	T	T	.	.	N	.	N
3421335	A	G	G	G	N	G	G	G	G	G	N
3442617	A	T	T	T	T	T	N	T	T	T	N
3540139	G	A	A	A	A	A	A	.	.	.	.
3564026	C	T	T	T	T	T	T	T	T	T	T
3571531	A	G	G	G	G	G	G	G	G	G	N
3610371	C	T	T	T	T	T	.	.	.	.	.
3613964	C	A	A	A	A	A	.	.	.	.	N
3616733	A	G	G	G	G	G	G	G	G	G	N
3620114	G	A	A	A	A	A	.	.	.	.	.
3620500	G	A	A	A	A	A	N	.	.	.	N
3643387	G	.	.	.	.	.	.	T	.	.	N
3645151	C	G	G	G	G	G	G	G	G	G	G
3667806	A	G	G	G	G	G	G	G	G	G	G
3726726	A	G	G	G	G	G	G	G	G	G	N
3761046	G	A	N	N	N	N	N	N	.	N	N
3764396	C	A	A	A	N	A	A	.	N	N	N
3782640	G	A	A	A	A	A	N	.	.	.	N
3806677	C	T	T	T	T	T	T	.	T	T	T
3824821	G	A	A	A	A	A	.	.	.	.	N
3872698	C	T	T	T	T	T	T	.	.	.	N
3888808	C	.	.	T	.	.	.	.	.	.	.
3944305	C	A	A	A	A	A	.	.	.	.	.
3973746	C	T	T	T	T	T	N	T	T	T	T
3973901	G	A	A	A	A	A	.	.	.	.	.
3988141	C	T	T	T	T	T	.	.	.	.	.
3989422	C	.	.	.	.	.	A	.	.	.	.
4082562	T	C	C	C	C	C	C	C	C	C	C
4083536	A	G	G	G	G	G	G	G	G	G	G
4134121	A	T	T	T	T	T	.	.	.	.	N
4150574	C	A	A	A	A	A	A	.	.	.	.
4173149	A	C	C	C	N	C	N	C	N	N	N
4190286	C	A	A	A	A	A	N	N	N	.	N
4194600	G	A	A	A	A	A	A	A	A	A	A
4200639	C	A	A	A	A	A	N	.	.	.	N
4208536	A	G	G	G	G	G	N	.	.	.	N
4232240	C	T	N	N	N	N	.	.	.	.	.
4236782	C	N	N	T	N	N	.	.	.	.	.
4236789	C	N	N	G	N	N	N	.	.	.	.
4242260	G	T	T	T	T	T	T	.	.	.	N
4243823	A	T	T	T	T	T	T	T	T	T	N
4296702	G	N	N	N	N	N	N	T	N	N	N
4301295	G	.	.	.	.	.	.	.	.	.	T
4363505	C	T	T	T	T	T	.	.	.	.	.
4371886	A	G	G	G	G	G	G	G	G	G	G
4396236	G	T	T	T	T	T	.	.	.	.	N
4421278	G	N	N	N	A	N	N	N	N	N	N
4421633	T	C	C	C	C	C	C	C	C	C	C
4421689	A	G	G	G	G	G	N	G	G	G	G
4456212	C	A	A	A	A	A	N	N	N	.	.
4518401	G	A	A	A	A	A	A	A	A	A	A
4527483	A	G	G	G	G	G	N	G	G	G	N
4567317	C	.	N	A	N	N	N	.	.	.	N
4579183	A	G	G	G	G	G	G	G	G	G	.
4616904	T	C	C	C	C	C	.	.	.	.	.
4634287	A	G	G	G	G	G	G	G	G	G	G
4642828	G	A	A	A	A	A	N	.	.	N	N

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

Maria A. Spyrou[1,2]*, Marcel Keller[1,3]*, Rezeda Tukhbatova[1,4], Elisabeth Nelson[1], Don Walker[5], Amelie Alterauge[6], Hermann Fetz[7], Joris Peters[3], Niamh Carty[5], Robert Hartle[5], Michael Henderson[5], Elizabeth L. Knox[5], Sacha Kacki[8], Michaël Gourvennec[9], Dominique Castex[10], Sandra Lösch[6], Michaela Harbeck[3], Alexander Herbig[1], Kirsten I. Bos[1] and Johannes Krause[1,2]

Affiliations:

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

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

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

[4]Center of Excellence "Archaeometry", Kazan Federal University, Kazan 420008, Russian Federation

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

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

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

[8]Department of Archaeology, Durham University, South Rd, Durham DH1 3LE, United Kingdom

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

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

*These authors contributed equally to this work

Abstract

The second plague pandemic (14th to 18th centuries AD), caused by the bacterium *Yersinia pestis*, is infamous for its initial wave, the Black Death (1346-1353 AD), and its repeated scourges in Europe and the vicinity until the Early Modern Era. Currently, there is disagreement on the origin of the Black Death and on its relationship to the ensuing outbreaks given the contrasting evidence that exists from climatic and genomic data. Here, we nearly quadruple the number of *Y. pestis* genomes sequenced from that time period by analysing human remains from nine European epidemic cemeteries spanning the 14th to 17th centuries AD. Our data support a single wave entry of the disease from Eastern Europe and low genetic diversity in the bacterium during the pandemic's initial wave as well as during local outbreaks thereafter. In addition, our analysis of post-Black Death outbreaks reveals the local diversification of a single *Y. pestis* lineage, which may have given rise to more than one reservoir of the disease in, or close to, Europe.

Introduction

One of the most devastating pandemics of human history was the second plague pandemic, which began with the infamous Black Death (BD, 1346-1353 AD) and subsequently continued with recurrent outbreaks in Europe until the 18th century[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 sylvatic rodent populations in Asia, Africa and the Americas[4-6]. The source of the second pandemic, as well as the route that plague followed prior to its entry into Europe, remains hypothetical. Genetic characterisation of BD isolates analysed alongside modern diversity has associated its initiation with a radial diversification of *Y. pestis* lineages, and has proposed an East Asian origin as modern genomes that are directly ancestral to this event have been identified in China near historical trade routes[7,8]. However, the fact that most modern *Y. pestis* strains have been isolated in China[7] creates and evident sampling bias in the data and make the accuracy of such inferences questionable. The view of an East Asian origin is also supported by certain historical accounts, suggesting that the bacterium was transported westward either via the Mongol army or via Silk Road traders during the early 14th century[9]. Such claims, however, could not yet be

verified given the scarcity of documentary sources, as well as the lack of molecular evidence from Central and East Asia dating to the early 14th century[10].

The first clearly documented outbreaks of the second pandemic occurred in 1346 in the Lower Volga and Black Sea regions. From there, it seems to have spread through the rest of Europe within the next five years, causing reductions in the population estimated to be as high as 60%[1,11,12]. However, at present, the only ancient genetic evidence available from Eastern Europe stems from the time period shortly after the BD (Bolgar City, Russia, 1362-1400 AD)[13] during which *Y. pestis* seems to have expanded eastward to later become a precursor of the third plague pandemic that spread worldwide from China during the 19th century[13-15]. Therefore, it has been difficult to genetically assess whether this region also contributed in the entrance of the disease in Europe, which still awaits genomic evidence from the earliest phases of the second pandemic.

After the BD, plague was a common scourge in Europe as evidenced by the thousands of recorded epidemics it supposedly caused between 1353 and 1770[2,16]. Whether these were multiple introductions of the disease from an Asian source or rather its local persistence in Europe is an on-going subject of scholarly debate[13,17-19]. While data from climatic proxies are interpreted to support the former hypothesis[17], genetic evidence currently favours the latter[13,18]. Analysis of *Y. pestis* genomes from 16th century Ellwangen, Germany[13] and the Great Plague of Marseille in France (L'Observance, 1720-22)[19] revealed the existence of a previously unidentified *Y. pestis* lineage that was responsible for both of these post-Black Death (pBD) outbreaks. This lineage descends from the strain associated with the BD thus far identified in both London and Barcelona, and, therefore, likely represents plagues' legacy and persistence in or around Europe after 1353. The limited number of ancient *Y. pestis* genomes[8,19,20], however, challenges our ability to construct hypotheses of whether this lineage was solely responsible for the many pBD outbreaks in Europe and whether it derived from a single or multiple reservoirs.

Here, we take steps to overcome these limitations by greatly expanding the number of available *Y. pestis* genomes from multiple time periods and locations, in order to gain

additional knowledge on the early stages of the second pandemic, and to study the genetic diversity of the bacterium present in Europe after the BD. We analyse material recovered from nine archaeological sites located in Russia, Germany, Switzerland, England and France from which we reconstruct 32 *Y. pestis* genomes. Our data supports the entrance of the *Y. pestis* into Europe through the east during the initial wave of the pandemic and consistently demonstrates low genetic diversity of the bacterium during the BD. In addition, our genomic analysis of pBD outbreaks from Central and Western Europe suggests the local diversification of a single plague lineage between the 14th and 18th centuries that may have resided in more than one reservoir.

Results

Y. pestis qPCR screening

Human archaeological material was screened for the presence of the *Y. pestis*-specific gene, *pla*, located on the pPCP1 plasmid, using a previously described qPCR approach[21]. A total of 181 teeth were retrieved and tested from archaeological sites in the cities of Nabburg (n=12), Manching-Pichl[18] (n=28), Starnberg (n=3), Landsberg am Lech (n=10) and Brandenburg an der Havel[18] (n=3) in Germany; Stans (n=32) in Switzerland, London (n=40) in England, Toulouse (n=39) in France and the city of Laishevo (n=10) in the Volga region of Russia (Figure 1 and Supplementary Information). Extracts from 49 teeth across all sites tested positive for *pla* (Supplementary table 1). All extraction negative controls were free of amplification product. Amplification products from putatively positive individuals were not sequenced, as the presence of *Y. pestis* was subsequently assessed through whole-genome capture and high-throughput Illumina sequencing.

Y. pestis capture and whole genome reconstruction

We prepared UDG-treated libraries[22,23] from putatively positive extracts and used a *Y. pestis* in-solution capture approach[24] combined with high-throughput sequencing for the retrieval of 1,299,105 to 79,055,317 raw reads per sequenced library. All data was mapped against the *Y. pestis* CO92 reference genome (NC_003143.1)[3]. This resulted in between 86,278 to 3,822,030 unique mapping reads yielding 1.1 to 80.1-fold coverage across 32 individuals that span the time transect between the 14th and 17th centuries in

Europe (Supplementary table 2). More specifically, we could retrieve three *Y. pestis* genomes from Nabburg, two from Manching-Pichl[18], one from Starnberg, one from Landsberg am Lech, two from Brandenburg am Havel[18] (all from Germany), 15 from Stans (Switzerland), five from London (England), one from Toulouse (France) and two from Laishevo (Russia). Of those, 23 isolates showed at least 50% of the reference genome covered at 5-fold (Supplementary table 2), which allowed for their 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 using spatially immobilized array-capture[13] (Supplementary table 2).

Y. pestis phylogenetic reconstruction

To infer genetic relationships between the new and previously published *Y. pestis* isolates we constructed phylogenies using the Maximum Likelihood and Maximum Parsimony algorithms, allowing for up to 4% missing data (96% partial deletion) to accommodate lower coverage genomes. As a reference dataset we used 159 modern isolates[7,25,26], which represent most of the published *Y. pestis* genetic diversity. In addition we included all previously published second pandemic isolates (n=12)[8,19,20], a 6th century isolate from Germany[27] and three Bronze Age isolates from the Altai and Volga regions (see Methods)[28].

All newly reconstructed genomes group with Branch 1 and appear closely related to the previously published second pandemic isolates from Europe (Figure 2, Supplementary fig. 1), thus confirming their authenticity. A number of genomes were excluded from the phylogeny (NAB003, BRA003, STN011 and STN013, see Supplementary table 2) as they showed an excess amount of heterozygous alleles in comparison to other contemporaneous isolates from the same archaeological contexts (Supplementary Figure 2). Such alleles potentially arise from enrichment of non-target DNA stemming from closely related organisms, an effect frequently identified in ancient microbial datasets[27,29,30] that can result in false branch elongation (Supplementary fig. 3) and misinterpretation of phylogeographic inferences.

- *Genomes from the Black Death*

Our phylogenetic reconstruction shows that the LAI009 isolate from Laishevo (Figure 1), is ancestral to the BD isolates from Central, Western and Southern Europe, as well as to the previously published late 14th century isolates from London (6330)[8] and Bolgar City[13] (Figure 2). This genome possesses only one derived SNP distinguishing it from the radiation node (N07) that gives rise to branches 1-4[7]. Since this SNP is shared with all other second pandemic strains, this genome represents the earliest so far identified form of the strain that entered Europe during the initial wave of the second pandemic. For the Central and Western European genomes, NAB003 does not show differences compared to previously published BD genomes from London and Barcelona[8,20], whereas TRP002 from Toulouse, which dates to 1347-1350 based on archaeological evidence, is seemingly distinct from other isolates and forms its own unique branch (Figure 2). Although this isolate showed indications of high heterozygosity (Supplemental fig. 3) and suffered from comparatively lower coverage, we nevertheless considered it further in our analysis, as it is the sole genome available from this site. To investigate whether its branch arose from true unique variants or contaminant reads that resulted in erroneous SNP calling, the genome alignment was evaluated in greater detail. After visual inspection, all eight SNPs unique to the TRP002 genome appear in regions of the genome where reads from diverse sources seem to be mapping (Supplementary fig. 4). In addition, such variants exhibited all previously defined characteristics of problematic SNP assignments in low coverage genomes[27] and, therefore, were considered to be of exogenous origin (Supplementary fig. 4). We, hence, conclude that apart from LAI009, all other reconstructed genomes that date to the Black Death period have identical genotypes.

- *Post Black Death genomes*

We find a number of genomes grouping with the previously described 'pBD' lineage together with published strains from Ellwangen, Germany (1486-1630)[13], and Observance, France (1720-1722)[19], which are descended from the European BD isolates. Here, we identify the earliest evidence of this lineage in a 14th - 15th century isolate from Manching-Pichl (MAN)[18] (see Supplementary Information), which is followed by the more derived 15th - 17th century isolates from Landsberg (LBG) and Stans (STN), the 16th century Starnberg (STA), as well as the 17th century Brandenburg

an der Havel (BRA)[18] 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 (maximum one SNP difference between isolates, 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 Obervance (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 "pBD lineage" gave rise to (at least) two distinct clades within Europe, with the Ellwangen isolate being positioned closest to an apparent population split (Figure 2). From this divergence, one clade gives rise to the strains associated with outbreaks in Southern Germany and Switzerland, and the second encompasses strains from 17th century London (BED) and Observance (France). Notably, these two clades show dissimilar rates of substitution accumulation. For example, the mean SNP distance between the Ellwangen genome and the London genomes (BED, d=45) is double from that observed for Brandenburg (BRA, d=22), despite an assumption of them being contemporaneous based on archaeological data (Supplementary table 1; Figure 2).

Analysis of substitution rate variation in Y. pestis

We used the Bayesian framework BEAST v1.8 in order to make a global assessment of substitution rate variations across the entire *Y. pestis* genealogy, using all available calibration points in our worldwide dataset (Supplementary data 1). 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 lineage expansion during epidemic spread, revealed as radiation events throughout the tree[7,15]. Our analysis suggests over 30-fold variance between the fastest and slowest mutation rates observed in the tree (Figure 3b). In particular, we find the highest disparities in Branch 1 (Figure 3a), which has the widest geographical distribution of all branches and is associated with both the second and third pandemics (Figure 2). Within this clade we observe the fastest rates in three internal branches (Figure 3a). The first, which displays the fastest rate among the entire phylogeny, spans the genetic distance between the strains from Ellwangen (549_O) and London (BED), and supports the

conflicting branch lengths of BED and BRA strains described earlier (Figure 3 and Supplementary data 2). The second is the branch leading to the 1.ANT strains isolated from Africa (Congo and Uganda) (Figure 3 and Supplementary data 3). The broad history of 1.ANT and the time period associated with its establishment in Africa are unknown. The third is the branch leading to 1.ORI isolates (Figure 3 and Supplementary data 4), which is associated with the global spread of *Y. pestis* via marine routes during the third plague pandemic (1894 - 1950s)[14,15]. Our results, therefore, support the idea of faster substitution rates during epidemic spread, here particularly noticeable for lineages known to have expanded over wide geographic areas.

Virulence factor analysis

In order to investigate the virulence profiles of all newly reconstructed genomes, we analysed the presence or absence of virulence-associated genes located on the *Y. pestis* chromosome (Figure 4) and plasmids (Supplementary fig. 5)[31,32], 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'[13], and the first pandemic genome 'Altenerding 2148'[27] seem to lack coverage in certain *Y. pestis*-specific regions, since *Y. pseudotuberculosis* probes were used for enrichment of their genomes (Figure 4). Regarding the newly reconstructed strains, we find that most possess all analysed virulence determinants with the exception of the New Churchyard (BED) and Marseille (OBS) strains that seem to lack the magnesium transporter genes *mgtB* and *mgtC* (Figure 4). Since both genes are covered in all other second pandemic genomes (Figure 4), it is unlikely that they were not captured by our probe-set and, therefore, likely represent real gene losses. Magnesium transporters are considered vital for intracellular uptake and for survival of bacteria under low Mg^{2+} conditions[33], and their disruption in *Y. pestis* has been associated with a decreased virulence in mice[34]. Both *mgtB* and *mgtC* are present in all 159 modern *Y. pestis* genomes used in our comparative dataset. Therefore their loss seems to be specific to the BED and OBS strains, which group on the same phylogenetic clade and are the most derived isolates on the "pBD lineage"

(Figure 2). The former is likely associated with one of the last outbreaks that occurred in England (1560 – 1635 calAD) and the latter is representative of the Plague of Marseille (1720-1722)[19], known as one of the last major outbreaks that occurred in continental Europe[35].

Discussion

A series of studies have sufficiently demonstrated the preservation of *Y. pestis* in ancient human remains from a wide temporal transect[8,13,19,27,28,36,37]. In this study, this has given a great opportunity for the extensive sampling of multiple epidemic burials from the time period between the 14th and 17th centuries in Europe, in order to gain a more complete overview of the second plague pandemic. As a result, we nearly quadruple the amount of genomic data available from that time (Figure 1 and Supplementary tables 1, 2) and their integration with existing frameworks has revealed key insights into the initiation and progression of the second plague pandemic.

Based on historical sources alone, it has been difficult to determine the time at which *Y. pestis* first reached western Russia[12]. A commonly accepted view has its arrival in the southwest, particularly in cities of Astrakhan and Sarai, in 1346[1,11,38] and subsequent spread into southern Europe from the Crimean peninsula. On the other hand, the spread of plague into northwestern Russia (i.e. in the cities of Pskov and Novgorod[12,38]), may have followed an alternative route, occurring at the end of the BD, between 1351-1353, via the Baltic Sea[1,12,38]. Such a notion of plague's expansion from northern Europe towards the east is also supported by published ancient genomic data from Bolgar City, in the Middle Volga region of Russia[13]. Importantly, through analysis of our new strain from Laishevo (LAI009), which is phylogenetically ancestral to all second pandemic strains sequenced to date (Figure 2), we provide evidence for the bacterium's presence in the same region, ~2,000 km northeast of the Crimean peninsula, prior to reaching Southern Europe in 1347 - 1348[1,11] (here represented by strains from Barcelona and Toulouse). These results lead us to propose that the N07 derived variant previously termed "p1" (Figure 2, Supplementary fig. 1) that is common to all other second pandemic strains, was likely acquired within Europe during the BD. In addition, given the proximity of the LAI009 genome to the N07 radiation node often associated with the initiation of the BD (Figure 2, Supplementary fig. 1)[7], as well as the apparent East

Asian sampling bias of modern isolates[7], additional data will be necessary for an accurate re-evaluation of BD's geographic origin.

Such insights become valuable when attempting to reconstruct the spread of plague after 1353. Previous research based on climatic proxies[17] and PCR data[39] have proposed multiple introduction waves of the bacterium into Europe as the main source for the pBD outbreaks recorded until the 18th century. Here, using previously published[8,13,19] and new whole genome data from a total of 15 archaeological sites, we interpret our data as supporting a single entry of plague into Europe during the second pandemic. Subsequent to this entry, we observe the formation of two sister lineages (Figure 2). The first lineage is responsible for the eastward expansion of plague after the BD. It contains strains from late-14th century London (6330)[8] and Bolgar City (2370)[13], as well as extant strains from Africa (1.ANT)[40], and, most importantly, a worldwide set of isolates associated with the third pandemic (1.ORI, 19th – 20th centuries)[7,14,15] (Figure 2). The second, here termed the "pBD lineage" is characterised by a genomic diversity currently restricted to European second pandemic isolates. It is represented by historical strains isolated from Germany (MAN, STA, ELLW, LBG, and BRA), Switzerland (STN), England (BED) and France (OBS) (Figure 2), suggesting its persistence in Europe between the 14th and 18th centuries. This lineage has no identified modern descendants, which is likely related to the disappearance of plague from Europe in the 18th century, as previously suggested[13,19].

We find that the "pBD lineage" gave rise to (at least) two distinct clades within Europe that separate the strains identified in Central Europe during the 15th - 17th centuries, and those identified in 17th - 18th century England and France. Their distinction is corroborated not only by their genetic and geographic separation (Figure 2), but also by apparent differences in their virulence profiles (Figure 4) and substitution rates (Figure 3). The clade that exhibits a slower substitution rate is mainly represented by temporally and genetically closely related isolates from Southern Germany and Switzerland (Figure 2), which may indicate endemic circulation of the bacterium in that region. Such an observation is compatible with the hypothesis of an Alpine rodent reservoir facilitating the spread of plague in central Europe after the BD[41]. On the other hand, the clade that exhibits a faster substitution rate (Figure 3) seems to have had a more wide geographic distribution. Given that both Marseille and London were among the main maritime trade

centres in Europe during that time, it is likely that introduction of the disease in these areas occurred via ships. Similarly, published research has demonstrated that transmission of plague via steamships during the 19th century third pandemic played a significant role in initial introduction of the bacterium to several regions worldwide, such as in Madagascar where the bacterium persists until today[14,15,42,43]. Although marine introductions of plague into London and Marseille during the second pandemic vastly expand the location of their potential source, the phylogenetic positioning of the BED and OBS genomes within the "pBD lineage" and in relation to other second pandemic isolates suggests it was likely within Europe or the vicinity. In addition, we identify the loss of two virulence-associated genes, *mgtB* and *mgtC*, in both the BED and OBS strains but not in any other second pandemic or modern strain within our dataset (Figure 4). The inferred virulence importance of these genes is associated with the intracellular survival of *Y. pestis* within macrophages[34,44], and they have been proposed as potential drug targets for certain pathogenic bacteria[34,45]. Given that, within our dataset, the *Y. pestis* strains that lack them were responsible for some of the last recorded European plague outbreaks (see Supplementary Information), their precise functional significance will be of future importance for the characterisation of plague's virulence and maintenance within Europe towards the end of the second pandemic.

The second plague pandemic has arguably caused the highest level of mortality of the three recoded plague pandemics[46]. It serves as a classic historical example of rapid infectious disease emergence in Europe, long-term local persistence and eventual extinction for reasons that are not currently understood. We have shown that extensive sampling of ancient *Y. pestis* genomic data can provide direct molecular evidence on the genealogical relationships of strains present in Europe during that time. In addition, we provide insights into the initiation and progression of the second pandemic and suggest that a single source reservoir may be insufficient to explain the breadth of epidemics recorded during the 400-year course of the pandemic in Europe. Although certain key areas in Europe remain under-sampled for ancient DNA, namely Southern Europe and Scandinavia, vast amounts of high-quality genomic data are becoming increasingly available. Their integration into disease modelling efforts, which consider vector transmission dynamics[47,48], climatic[17,49,50] and epidemiological data[51], as well as a

critical re-evaluation of historical records[52] will become increasingly important for the better understanding of the second plague pandemic.

Methods

Laboratory work was primarily performed in the dedicated aDNA facilities of the Max Planck Institute for the Science of Human History in Jena and part of the sampling and DNA extractions were performed at aDNA facilities of the ArchaeoBioCenter of the Ludwig Maximilian University of Munich.

Tooth sampling, DNA extraction and Y. pestis qPCR screening

181 teeth from nine sites located in Germany, Switzerland, England, France and Russia spanning the 14th – 17th centuries (see Supplementary Information) were sectioned in the cementoenamel junction, and 30-70 mg of powder were removed from the surface of the pulp chamber using a dentist drill. This powder was then used for DNA extraction, using a protocol optimised for the retrieval of short fragments that are most typical for ancient DNA[53]. Tooth powder was incubated in 1 ml lysis buffer (0.45M EDTA, pH 8.0, and 0.25 mg/ml proteinase K) overnight (12-16 h) at 37°C. After, DNA was bound to the silica membrane of spin columns using 10 ml of GuHCl-based binding buffer as described before[53], and purification was performed using either the MinElute purification kit (Qiagen), or the Viral Nucleic Acid Kit (Roche). DNA was eluted in 100 μl TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20). Extraction blanks and a positive extraction control (cave bear specimen) were taken along for every extraction batch. All extracts were then evaluated for PCR inhibition, by spiking 2 μl of each extract in a qPCR reaction containing a standard of known concentration[21]. 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*) in, located on the *Y.pestis*-specific pPCP1 plasmid using a published protocol[21]. 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 PCR products.

UDG library preparation and Y. pestis whole genome capture

After *pla* screening, samples that were determined as putatively positive for *Y. pestis* were converted into Illumina double-stranded DNA libraries as described before[23], using 50-60 μl of DNA extract, with an initial USER (New England Biolabs) treatment step, where uracil-DNA-glycosylase (UDG) was used in combination with endonuclease VIII to excise uracil nucleotides that result from post-mortem DNA damage[22,54]. Subsequently full UDG-treated, or partially UDG-treated libraries were quantified on a qPCR using the IS7/IS8 primer combination. Based on the initial quantification, the libraries were then split for double indexing[55], to enhance the efficiency of the indexing reaction. 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 combination to DNA library molecules using *Pfu Turbo Cx Hotstart DNA Polymerase* (Agilent). PCR products were purified using the MinElute DNA purification kit (Qiagen), and eluted in TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.05% Tween20). After indexing all libraries were amplified using *Herculase II Fusion DNA Polymerase* (Agilent) to a concentration of 200-300 ng/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 (10mM Tris-HCl, 1 mM EDTA pH 8.0, 0.05% Tween20). In-solution *Y.pestis* capture was then performed as described previously[24], where the following genomes were used as templates for probe design: CO92 chromosome (NC_003143.1), CO92 plasmid pMT1 (NC_003134.1), CO92 plasmid pCD1 (NC_003131.1), KIM 10 chromosome (NC_004088.1), Pestoides F chromosome (NC_009381.1) and *Y. pseudotuberculosis* IP 32953 chromosome (NC_006155.1). DNA captures were carried out on 96-well plates. Each sample was either captured in its individual well, or pooled with maximum one more sample from the same site. Blanks with non-overlapping index combinations were captured together.

Sequencing and read processing

After capture all products were sequenced on the NextSeq500 platform using (1x151+8+8 cycles or 1x76+8+8 cycles) and HiSeq4000 using (1x76+8+8 cycles or 2x76+8+8 cycles). Preprocessing of de-multiplexed reads was performed on the automated pipeline EAGER[56] and involved removing Illumina adapters and red

merging performed by AdapterRemoval v2[57], as well as filtering for sequencing quality (minimum base quality of 20) and read length (to retrieve reads ≥ 30 bp). Subsequently, reads were mapped with BWA[58] implemented in EAGER against the CO92 reference genomes (NC_003143.1)[3] using stringent parameters (-n 0.1, -l 32) for genome reconstruction and more lenient parameters for (-n 0.01, -l 32) genome inspection. Reads with mapping quality lower than 37 (-q) were removed using Samtools, and duplicates were removed using the MarkDuplicates tool (http://broadinstitute.github.io/picard/).

SNP calling and phylogenetic analysis

SNP calling was performed using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK)[59]. Our newly reconstructed genomes were analysed alongside 175 previously published *Y. pestis* genomes, which included a modern-day dataset of 159 genomes[7,25,26,40], one Justinianic strain (Altenerding)[27], three Bronze Age strains[28], 12 previously published historical genomes from the second plague pandemic and a *Y. pseudotuberculosis* strain (IP32953) which was used as outgroup. A vfc file was produced for every isolate using the 'EMIT_ALL_SITES' option, which generated a call for every position present in the reference genome. Furthermore we used he custom java tool MultiVCFAnalyzer v0.85[30](https://github.com/alexherbig/MultiVCFAnalyzer) to produce a SNP table of variant positions across all genomes analysed, using the following parameters: for homozygous alleles, a SNP would be called when covered at least 3-fold with a minimum genotyping quality of 30 and for heterozygous alleles a variant would be called when 90% of reads would support it. In cases where none of the parameters would be met an "N" would be inserted in the respective genomic position. In addition, we omitted previously defined noncore regions, as well as annotated repetitive elements, homoplasies, tRNAs, rRNAs, and tmRNAs from our SNP analysis[7,15]. In the present dataset, a total of 3,993 variant positions were identified. The annotation as well as the effect of each SNP was determined through the program SnpEff[60].

For the phylogenetic analysis, we used a SNP alignment produced by MultiVCFAnalyzer v0.85 to construct phylogenetic trees using the Maximum Likelihood (ML) and Maximum Parsimony (MP) algorithms. Up to 4% missing data was included in the analysis (96% partial deletion), resulting in a total amount of 3,489

SNPs used for phylogenetic reconstruction. The MP phylogeny was produced in MEGA7[61] in order to make a first assessment of genome topologies. The ML phylogenies were reconstructed with the program RAxML (version 8.2.9)[62] using the Generalised Time Reversible (GTR)[63] model with four gamma rate categories and 1000 bootstrap replicates to assess tree topology support.

Heterozygosity estimates

Heterozygous positions in all genomes were investigated given the disparity of branch lengths observed in isolates from the same archaeological sites (see Supplementary figure 2). Our approach takes into account the "haploid" nature of prokaryotic genomes, suggesting that "heterozygous" SNPs could either arise as a result of mixed infections or from erroneous mapping of DNA reads that belong to closely related bacterial contaminants. We subsampled all newly reconstructed genomes to ~5-fold coverage, and performed SNP calling using GATK[59], as described above. We then compiled a SNP table of all variant positions across our dataset using MultiVCFAnalyzer v0.85[30] accounting for all heterozygous positions. Histograms of allele frequencies for all SNPs with <100% read support were constructed with R v3.4.1[64] using representative genomes from all sites.

Estimates of rate variation

The Bayesian framework BEASTv1.8[65] was used to assess the substitution rate variation across the *Y. pestis* phylogeny. We used all modern and ancient *Y. pestis* genomes previously included in the phylogenetic analysis (n=181), with the exception of the problematic TRP002 genome, including all data in the SNP alignment (3,993 SNPs). Our BEAUti set-up included all available archaeological, radiocarbon, and sampling dates of both ancient and modern genomes were used as calibration points for the Bayesian phylogeny (Supplementary table 1). Divergence dates for each node in the tree were estimated as years before the present, where the year 2006 was considered as the present since it represents the most recently isolated modern *Y. pestis* strains. For modern strains with unknown isolation dates, the mean isolation date from all modern isolates were used (25 years BP) with a uniform prior spanning of the entire range of modern isolation dates (0 - 108 years BP). The Bronze Age RISE509 and RISE505 genomes[28] were constrained as a monophyletic outgroup clade to the rest of *Y. pestis*.

All Branch 1 genomes were also constrained as a monophyletic clade. The Generalised Time Reversible (GTR)[63] substitution model (6 gamma rate categories), a lognormal relaxed clock (clock rate tested and strict clock rejected in MEGA7[61]) and the Coalescent Constant Size[66] & Coalescent Bayesian Skyline[67] demographic models were used for two separate analyses. For each analysis, two chains of 400,000,000 states were used to ensure run convergence. In addition, we estimated marginal likelihoods to determine the best-fitted demographic model for our dataset using path sampling and stepping stone sampling (PS/SS) implemented in BEAST v1.8[65]. Each chain was run for an additional 400,000,000 states (4,000,000 states divided into 100 steps), with an alpha parameter of 0.3 and determined the Coalescent Bayesian Skyline as better fit for the data. The results produced by this demographic model were then viewed in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) to ensure all relevant expected sample sizes (ESS) were > 150. We used TreeAnnotator[65], to produce a maximum clade credibility (MCC) phylogeny for the best-fit model using 10% burn-in, which resulted in the processing of 36,001 trees. The MCC tree was viewed and modified in FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) where branch lengths were represented as a function of age and mean rates were used to colour individual branches.

Virulence factor presence/absence analysis

In order to investigate the virulence profiles of the newly reconstructed second pandemic genomes, the highest quality (coverage) genomes from every site (LAI009, NAB003, MAN008, STA001, LBG002, STN014, BRA001, BED030) were compared to each other and to previously published representatives of ancient (London ES, Bolgar 2370, Barcelona 3031, Ellwangen549_O, OBS0137, RISE509, RT5, Altenerding 2148) and modern (1.ORI-CO92, 0.PE2-PESTF, 0.PE4-Microtus 91001) *Y. pestis* isolates. The newly reconstructed TRP002 and the published London 6330 genomes were excluded from this analysis due to their low coverage. All other listed genomes were re-mapped against the CO92 chromosomal reference genome (NC_003143.1) using a mapping quality (-q) filter of 0. The coverage across of 123 previously defined virulence-associated genes[31] was calculated using bedtools[68]. The results are plotted in the form of a heatmap of using the ggplot2[69] package in R version 3.4.1[70] and can be viewed in Figure 4.

Figures 1 - 4
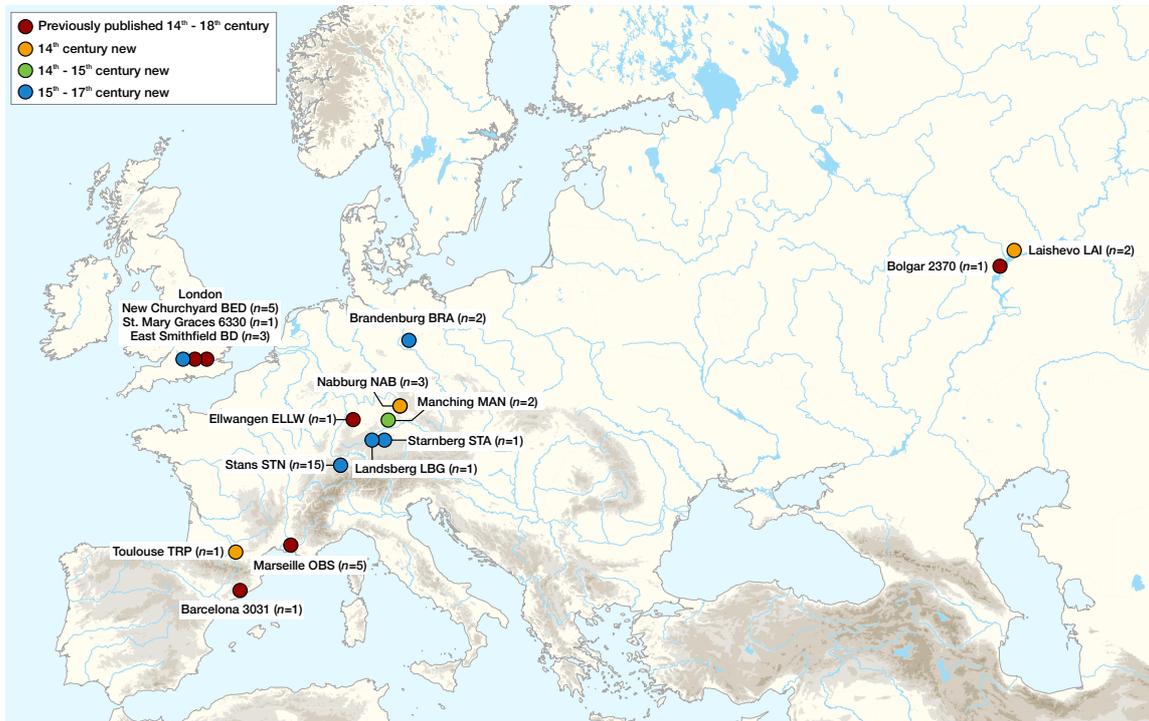
Figure 1 – Locations of previously published and new archaeological sites from where *Y. pestis* genomes have been retrieved. The geographic locations of the archaeological sites from which second pandemic (14th to 18th century) *Y. pestis* genomes have been reconstructed are shown in the above map. The number (*n*) of whole genomes obtained form each site is shown in brackets.
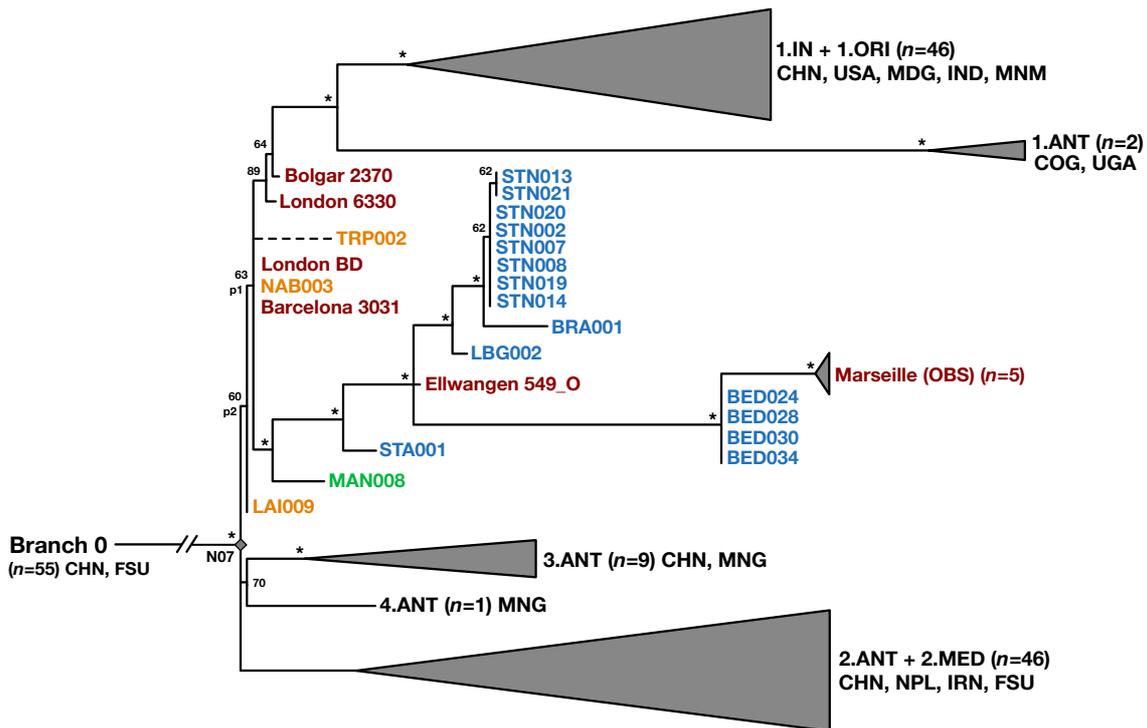
Figure 2 – Phylogenetic positioning of second pandemic strains

A maximum likelihood phylogeny was generated using 96% partial deletion, considering 3,489 SNP positions. The image shows a graphical representation of Branches 1 - 4 (see Supplementary figure 2 for a complete phylogeny), to emphasize the phylogenetic positioning of the new and previously published second pandemic strains (14th -18th centuries). Apart from new isolates, all other clades represented by five or more isolates were collapsed to enhance tree visibility. Nodes that have bootstrap values of ≥ 95 are indicated by asterisks (*). Geographic abbreviations of modern strain isolation locations are as follows: China (CHN), United States of America (USA), Madagascar (MDG), India (IND), Myanmar (MNM), Congo (COG), Uganda (UGA), Former Soviet Union (FSU), Mongolia (MNG), Nepal (NPL), Iran (IRN). Numbers in brackets indicate the number of strains represented in each collapsed branch.
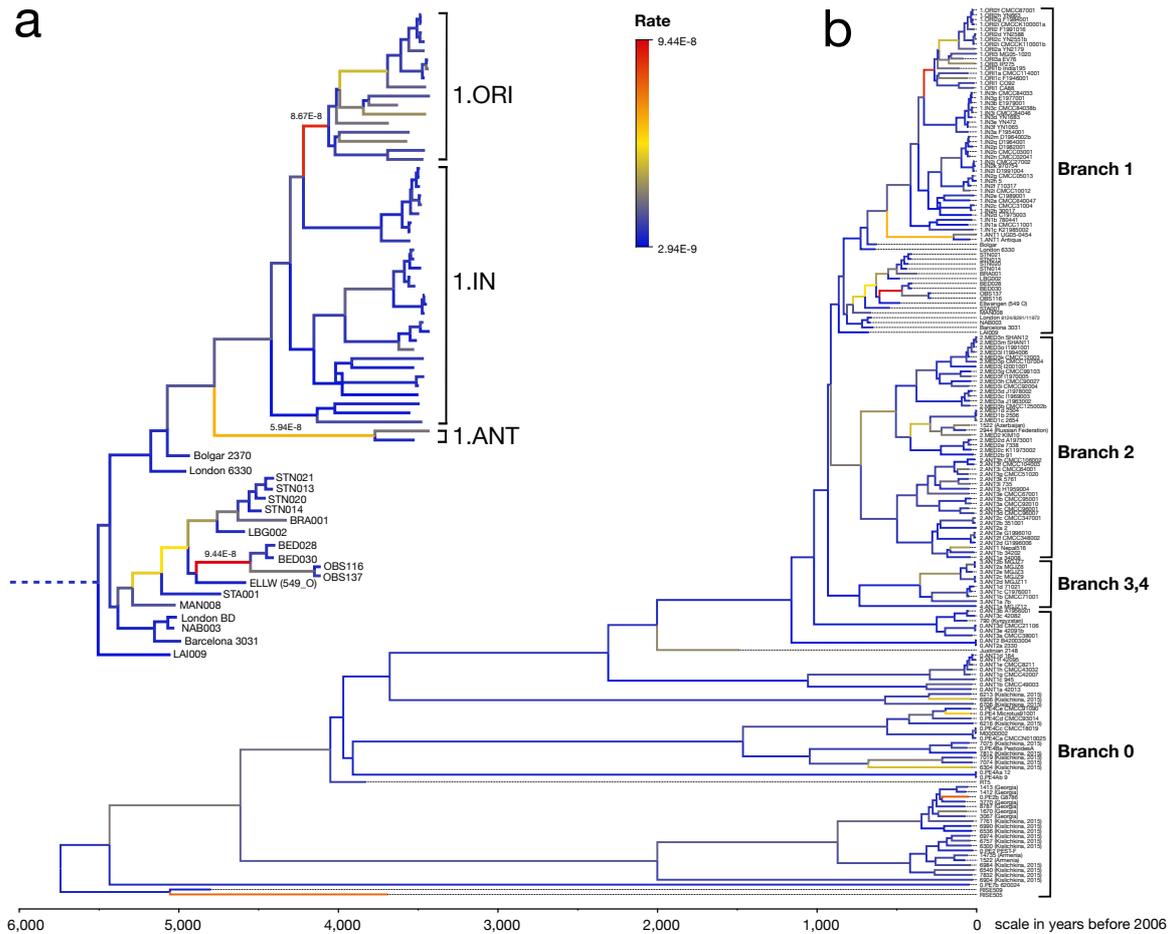
Figure 3 – Substitution rate variation across the _Y. pestis_ phylogeny

The figure presents a Maximum Clade Credibility phylogenetic tree calculated in BEASTv1.8. The tree was viewed in FigTree v1.4, and modified so that branch colours represent mean substitution rates (substitution/site/year). Panel (A) shows an expanded view of Branch 1, which is the portion of the phylogenetic tree where all second pandemic strains are grouped. Panel (B) shows the variation of substitution rates across the entire _Y. pestis_ phylogeny. The isolates used for this analysis overlap with the ones used for the SNP and Maximum Likelihood phylogenetic analysis (see Supplementary figure 1), with the exception that only two representative genomes from each archaeological site were used in cases where more identical ones were retrieved. The mean substitution rate across the entire phylogeny was calculated to 1.36E-8 substitutions/site/year. Branch lengths are scaled to represent sample ages, and the entire tree represents 5,877 years of _Y. pestis_ evolution.
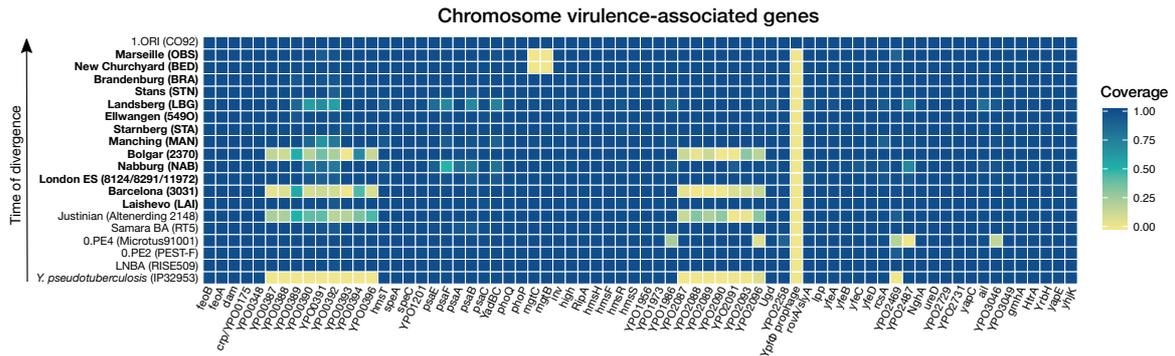
Figure 4 – Heatmap of virulence factor presence/absence analysis

A comparison of virulence genetic profiles was performed across newly reconstructed and previously published second pandemic genomes (in bold). An assessment of the presence or absence of 80 previously defined[31] chromosomal virulence-associated genes was used for this characterisation. Published *Y. pestis* representative strains from the Bronze Age period[28,71] (RISE509 and RT5), from the first pandemic[27] (6th century - Altenerding 2148), from modern-day isolates (0.PE2, 0.PE4 & 1.ORI)[7], as well as a *Y. pseudotuberculosis* (IP32953) strain[72], are also shown for comparative purposes. The colours represent a scale 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[70] using the ggplot2 package[73]. Refer to Supplementary figure 5 for presence/absence of virulence genes across the pMT1, pPCP1 and pCD1 plasmids.

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

Sample Name	Site name	Date (AD)	Uniquely mapping reads	Endogenous DNA post enrichment (%)	Mean fold coverage	Genome covered ≥ 5-fold (%)	Average fragment length (bp)	GC content (%)
BED030.A0102	New Churchyard, London	1560–1635[a]	3,624,482	36.2	80.1	93.6	102.9	48.5
BED028.A0102	New Churchyard, London	1560–1635[a]	2,665,238	22.2	37.2	91.4	65.0	49.0
BED034.A0102	New Churchyard, London	1560–1635[a]	1,371,698	10.5	18.3	89.1	62.2	49.2
BED024.A0102	New Churchyard, London	1560–1635[a]	1,000,524	18.1	12.6	84.7	58.5	49.1
BRA001.A0101	Domlinden 12, Brandenburg	1618-1648[b]	2,387,557	23.2	23.8	92.0	46.4	47.5
LAI009.A0101	Laishevo III, Laishevo	1300-1400[b]	2,549,926	23.9	28.4	92.1	51.8	48.0
LBG002.A0101	Kirchhof St. Johannis, Landsberg	1455-1632[b]	621,713	27.9	7.2	66.4	54.2	49.9
MAN008.B0101	St. Leonhardi, Manching-Pichl	1348-1450[b]	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
STA001.A0101	Possenhofener str. 3, Starnberg	1433-1500[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

References

1 Benedictow, O. J. *The Black Death, 1346-1353: The complete history*. (Boydell & Brewer, 2004).

2 Biraben, J.-N. Les hommes et la peste en France et dans les pays européens et méditerranéens des origines à 1850. *Population* **30**, 1143-1144 (1975).

3 Parkhill, J. *et al.* Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413**, 523-527, doi:10.1038/35097083 (2001).

4 Gage, K. L. & Kosoy, M. Y. Natural history of plague: perspectives from more than a century of research. *Annu. Rev. Entomol.* **50**, 505-528, doi:10.1146/annurev.ento.50.071803.130337 (2005).

5 Prentice, M. B. & Rahalison, L. Plague. *Lancet* **369**, 1196-1207, doi:10.1016/S0140-6736(07)60566-2 (2007).

6 Tikhomirov, E. Epidemiology and distribution of plague. *World Health Organisation* (1999).

7 Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc Natl Acad Sci U S A* **110**, 577-582, doi:10.1073/pnas.1205750110 (2013).

8 Bos, K. I. *et al.* A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506-510, doi:10.1038/nature10549 (2011).

9 McNeill, W. H. Plagues and peoples. 1976. *Garden City, NY: Anchor P* (1998).

10 Sussman, G. D. Was the black death in India and China? *Bull Hist Med* **85**, 319-355, doi:10.1353/bhm.2011.0054 (2011).

11 Gottfried, R. S. *The Black Death; natural and human disaster in Medieval Europe*. (Simon & Schuster, 1983).

12 Alexander, J. T. *Bubonic plague in early modern Russia: public health and urban disaster*. (Johns Hopkins University Press, Baltimore, Maryland 21218, USA, 1980).

13 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).

14 Pollitzer, R. The Plague. Geneva. *World Health Organization*, 26 (1954).

15 Morelli, G. *et al.* Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140-1143, doi:10.1038/ng.705 (2010).

16 Buntgen, U., Ginzler, C., Esper, J., Tegel, W. & McMichael, A. J. Digitizing historical plague. *Clin. Infect. Dis.* **55**, 1586-1588, doi:10.1093/cid/cis723 (2012).

17 Schmid, B. V. *et al.* Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci U S A* **112**, 3020-3025, doi:10.1073/pnas.1412887112 (2015).

18 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).

19 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).

20 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, doi:10.1016/j.chom.2016.05.012 (2016).

21 Schuenemann, V. J. *et al.* Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proc Natl Acad Sci U S A* **108**, E746-752, doi:10.1073/pnas.1105107108 (2011).

22 Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87, doi:10.1093/nar/gkp1163 (2010).

23 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).

24 Andrades Valtuena, A., *et al*. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 1-9, doi:10.1016/j.cub.2017.10.025 (2017).

25 Kislichkina, A. A. *et al*. Nineteen Whole-Genome Assemblies of Yersinia pestis subsp. microtus, Including Representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica. *Genome Announc* **3**, e01342-01315, doi:10.1128/genomeA.01342-15 (2015).

26 Zhgenti, E. *et al*. Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus Region. *Genome Announc* **3**, e01030-01015, doi:10.1128/genomeA.01030-15 (2015).

27 Feldman, M. *et al*. A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol. Biol. Evol.* **33**, 2911-2923, doi:10.1093/molbev/msw170 (2016).

28 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).

29 Vågene, A. J. *et al*. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol* **2**, 520-528, doi:10.1038/s41559-017-0446-6 (2018).

30 Bos, K. I. *et al*. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494-497, doi:10.1038/nature13591 (2014).

31 Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in Yersinia pestis. *Infect. Immun.* **77**, 2242-2250, doi:10.1128/IAI.01477-08 (2009).

32 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).

33 Groisman, E. A. *et al*. Bacterial Mg2+ homeostasis, transport, and virulence. *Annu. Rev. Genet.* **47**, 625-646 (2013).

34 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).

35 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).

36 Wagner, D. M. *et al*. Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319-326, doi:10.1016/s1473-3099(13)70323-2 (2014).

37 Andrades Valtueña, A. A. *et al*. The Stone Age Plague and Its Persistence in Eurasia. *Curr. Biol.* **27**, 3683-3691. e3688 (2017).

38 Benedictow, O. J. *The Black Death and Later Plague Epidemics in the Scandinavian Countries: Perspectives and Controversies*. (Walter de Gruyter GmbH & Co KG, 2016).

39 Haensch, S. *et al*. Distinct clones of Yersinia pestis caused the black death. *PLoS Pathog* **6**, e1001134, doi:10.1371/journal.ppat.1001134 (2010).

40 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).

41 Carmicheal, A. G. *Plague persistence in western Europe: A hypothesis*. Vol. 1 157-192 (ARC Medieval Press, 2014).

42 Brygoo, E.-R. Epidémiologie de la peste à Madagascar. *Les Archives de l'Institut Pasteur de Madagascar* (1966).

43   Vogler, A. J. *et al.* Temporal phylogeography of Yersinia pestis in Madagascar: Insights into the long-term maintenance of plague. *PLoS neglected tropical diseases* **11**, e0005887 (2017).

44   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).

45   Belon, C. *et al.* A macrophage subversion factor is shared by intracellular and extracellular pathogens. *PLoS Path.* **11**, e1004969 (2015).

46   Perry, R. D. & Fetherston, J. D. Yersinia pestis--etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35-66 (1997).

47   Dean, K. R. *et al.* Human ectoparasites and the spread of plague in Europe during the Second Pandemic. *Proceedings of the National Academy of Sciences*, 201715640 (2018).

48   Keeling, M. J. & Gilligan, C. A. Bubonic plague: a metapopulation model of a zoonosis. *Proceedings of the Royal Society B: Biological Sciences* **267**, 2219-2230, doi:10.1098/rspb.2000.1272 (2000).

49   Xu, L. *et al.* The trophic responses of two different rodent–vector–plague systems to climate change. *Proceedings of the Royal Society of London B: Biological Sciences* **282**, 20141846 (2015).

50   Xu, L. *et al.* Wet climate and transportation routes accelerate spread of human plague. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, 20133159 (2014).

51   Whittles, L. K. & Didelot, X. Epidemiological analysis of the Eyam plague outbreak of 1665–1666. *Proc. R. Soc. B* **283**, 20160618 (2016).

52   Roosen, J. & Curtis, D. R. Dangers of Noncritical Use of Historical Plague Data. *Emerging Infect. Dis.* **24**, 103 (2018).

53   Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* **110**, 15758-15763, doi:10.1073/pnas.1314445110 (2013).

54   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci* **370**, 20130624, doi:10.1098/rstb.2013.0624 (2015).

55   Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3, doi:10.1093/nar/gkr771 (2012).

56   Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).

57   Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* **5**, 337, doi:10.1186/1756-0500-5-337 (2012).

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

59   DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498, doi:10.1038/ng.806 (2011).

60   Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92 (2012).

61   Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).

62   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).

63   Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences* **17**, 57-86 (1986).

64   Team, R. D. C. R: A Language and Environment for Statistical Computing: Vienna, Austria. (2015).

65   Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).

66   Kingman, J. F. C. The coalescent. *Stoch Process Their Appl* **13**, 235-248 (1982).

67   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).

68   Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

69   Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).

70   R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2015).

71   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).

72   Chain, P. S. *et al.* Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. *Proc Natl Acad Sci U S A* **101**, 13826-13831, doi:10.1073/pnas.0404012101 (2004).

73   Wickham, H. ggplot2: Elegant Graphics for Data Analysis Springer-Verlag. *New York* (2009).

Acknowledgements

Author contributions

K.I.B, M.H., and J.K. conceived of the investigation. M.A.S., M.K., R.T. and E.N. performed laboratory work. M.S., M.K., and A.H. performed data analysis. D.W., A.A., H.F., J.P., N.C., R.H., M.H., E.L.K., S.K., M.G., D.C., S.L., and M.H. identified and provided access to archaeological material. M.S. and K.I.B wrote the paper with contribution from all co-authors.

Supplementary Information

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

Spyrou et al.

Supplementary Figures 1 - 5
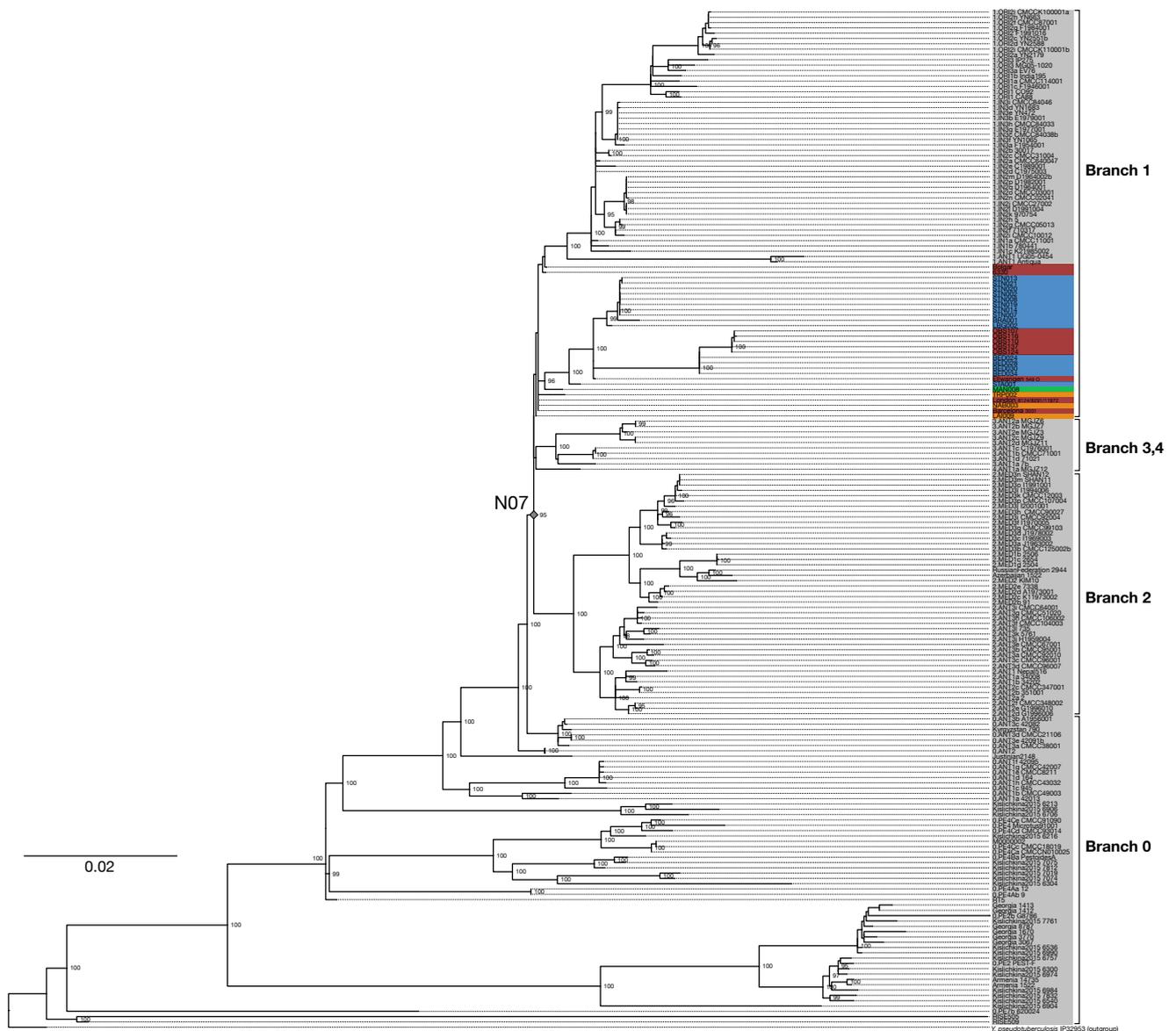
Supplementary figure 1 - Maximum likelihood phylogeny[1] (96% partial deletion) of all *Y. pestis* genomes used in this study. A total of 3,489 SNP positions were considered for the phylogeny. The tree comprises of 159 modern isolates, 31 second pandemic isolates, one first pandemic isolate, three Bronze Age isolates, and a *Y. pseudotuberculosis* isolate (IP32953)[2] that was used as outgroup for rooting the tree. Bootstrap values of 95 or higher are shown. Previously published second pandemic isolates are shown in red[3-5], newly sequenced 14th century isolates are shown in orange, the new 14-15th century MAN008 isolate is shown in green and all newly sequences 15th -17th century isolates are shown in blue.
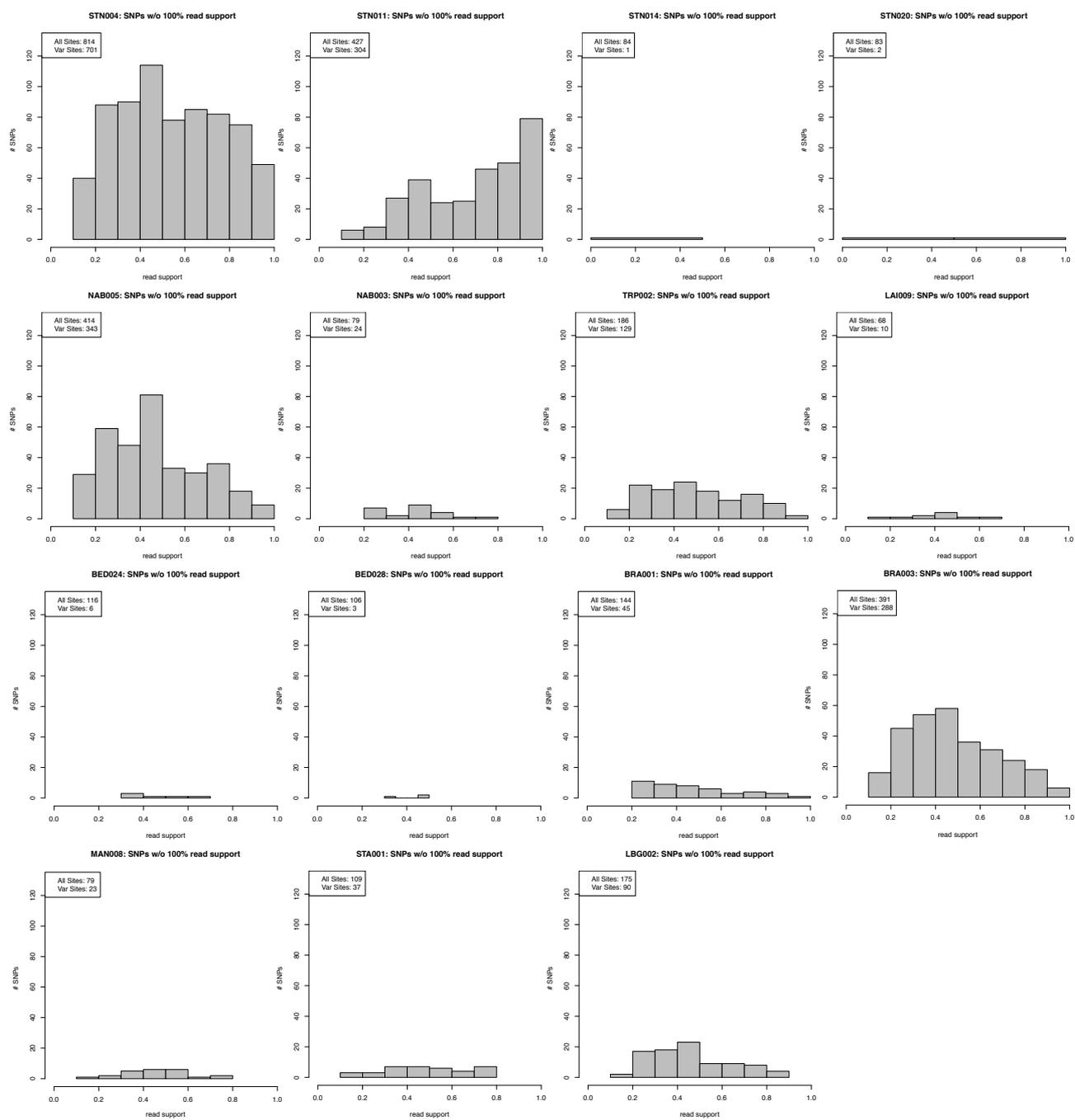
Supplementary figure 2 - Histograms showing the distribution of heterozygous positions in representative *Y. pestis* genomes from all new sites analysed in this study. All isolates were down sampled to an equal genomic coverage (5-fold) for this analysis. The histograms were constructed using R version 3.4.1[6].
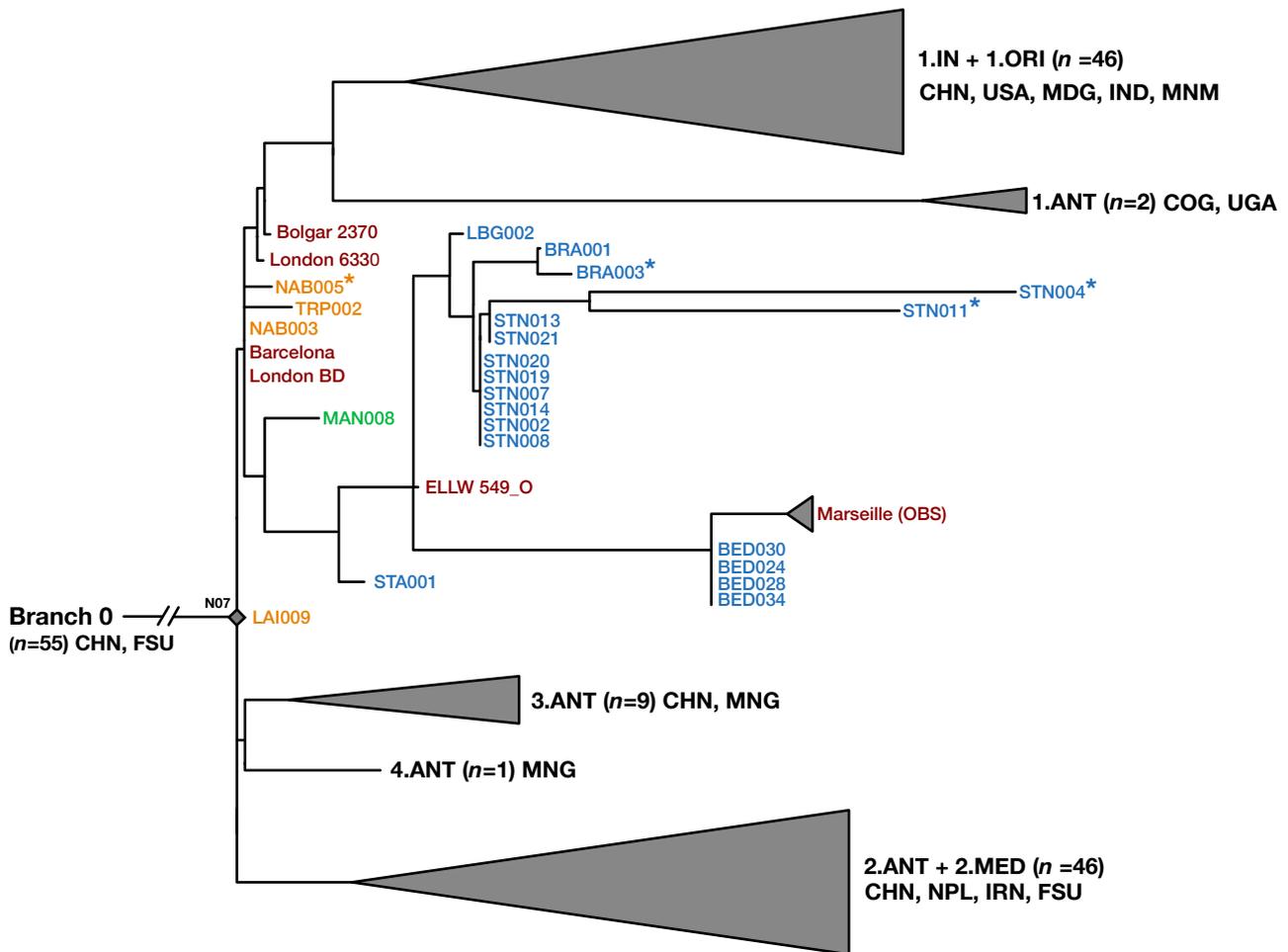
Supplementary figure 3 – A Maximum Parsimony[7] phylogeny using 96% partial deletion was generated. The figure shows a graphical representation of Branches 1-4, and more specifically the phylogenetic positioning of all previously published and new second pandemic strains (14th -18th centuries). All other clades with more than five isolates, except for the new isolates, were collapsed to enhance tree clarity. Isolates that showed evidence of environmental contamination to be influencing their SNP assignment are marked with and asterisk (*), and they were excluded from subsequent analyses in cases where more than one isolate for the same site was retrieved. 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), Former Soviet Union (FSU), Mongolia (MNG), Nepal (NPL), Iran (IRN). Numbers in brackets indicate the number of strains contained in each collapsed branch.
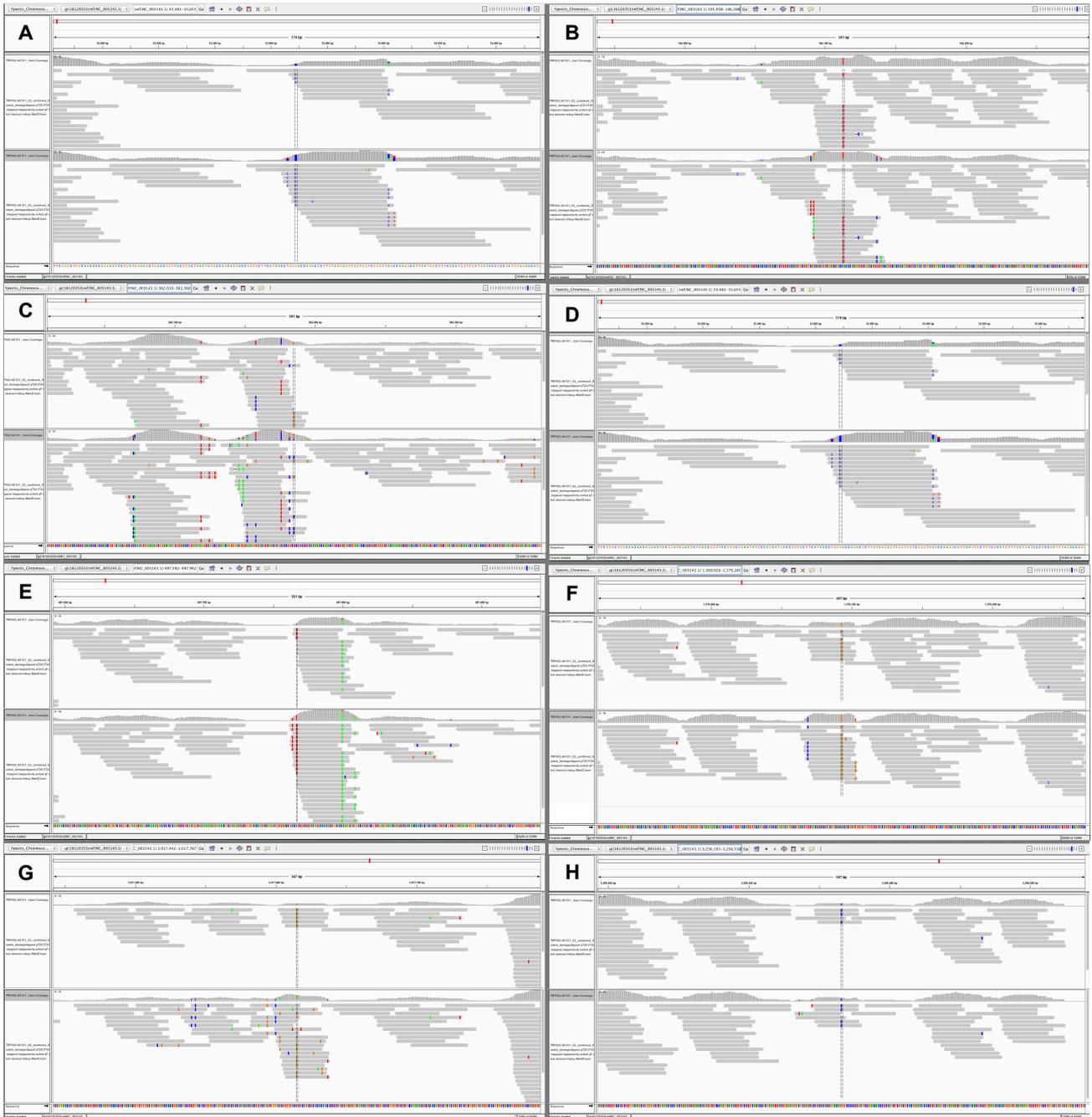
Supplementary figure 4 – Visual inspection of the eight unique SNP positions identified in TRP002. The screenshots are made in IGV[8], with the cursor placed on the genomic position where each SNP was identified. BWA mapping was carried out using both stringent (A-H upper panels, -n 0.1) and lenient parameters (A-H lower panels, -n 0.01).

Supplementary figure 5 – Analysis of coverage across plasmid virulence-associated genes. The presence/absence of previously identified virulence-associated genes[9,10] located on the *Y. pestis* plasmids 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[11] package in R version 3.4.1[12]. 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 bold. A lack of coverage across the plasmids of the Marseille OBS isolates is apparent since those regions were not used as part of the probe set for enrichment of these loci[13]. 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 radiocarbon date (95.4%)	*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	Geramany	Bef. 598	N/A	N/A	0.32
MAN008.B0101	St. Leonhardi, Manching-Pichl	Germany	MPS03-I*	1348-1450	1283-1390	171.80
MAN015.A0101	St. Leonhardi, Manching-Pichl	Germany	MP56-X*	1348-1450	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	1433-1500	1420-1630	3.60
STN014.A0101	Nägeligasse, Stans	Switzerland	Grave 104	N/A	1485-1635 (combined date)	129.60
STN020.A0101	Nägeligasse, Stans	Switzerland	Grave 124	N/A	1485-1635 (combined date)	66.93
STN021.A0101	Nägeligasse, Stans	Switzerland	Grave 125	N/A	1485-1635 (combined date)	11.71
STN019.A0101	Nägeligasse, Stans	Switzerland	Grave 123	N/A	1485-1635 (combined date)	34.75
STN007.A0101	Nägeligasse, Stans	Switzerland	Grave 97/POS.251	N/A	1485-1635 (combined date)	26.47
STN002.A0101	Nägeligasse, Stans	Switzerland	Grave 71	N/A	1485-1635 (combined date)	10.65
STN008.A0101	Nägeligasse, Stans	Switzerland	Grave 98/POS250	N/A	1485-1635 (combined date)	15.87
STN013.A0101	Nägeligasse, Stans	Switzerland	Grave 85	N/A	1485-1635 (combined date)	3.07
STN011.A0101	Nägeligasse, Stans	Switzerland	Grave 102	N/A	1485-1635 (combined date)	2.07
STN004.A0101	Nägeligasse, Stans	Switzerland	Grave 91	N/A	1485-1635 (combined date)	1.24
STN032.A0101	Nägeligasse, Stans	Switzerland	Grave 167	N/A	1485-1635 (combined date)	1.89
STN031.A0101	Nägeligasse, Stans	Switzerland	Grave 165	N/A	1485-1635 (combined date)	12.52
STN005.A0101	Nägeligasse, Stans	Switzerland	Grave 92	N/A	1485-1635 (combined date)	0.89
STN018.A0101	Nägeligasse, Stans	Switzerland	Grave 121	N/A	1485-1635 (combined date)	0.17
STN012.A0101	Nägeligasse, Stans	Switzerland	Grave 103	N/A	1485-1635 (combined date)	0.16
STN015.A0101	Nägeligasse, Stans	Switzerland	Grave 105	N/A	1485-1635 (combined date)	0.13
STN026.A0101	Nägeligasse, Stans	Switzerland	Grave 134	N/A	1485-1635 (combined date)	0.28
STN016.A0101	Nägeligasse, Stans	Switzerland	Grave 106	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

* Two individuals from Manching-Pichl and two from Brandenburg and der Havel that where investigated in previous studies, where *Y. pestis* was detected by PCR[14-17].

Supplementary table 2 – Sequencing statistics after whole genomes *Y. pestis* capture for all samples that yielded coverage ≥ 1-fold

Sample Name	Archaeological IDs	Number of Pre-processed reads	All mapping reads	Uniquely 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.A0101	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
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	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	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	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	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	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	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/POS250	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	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	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	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	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	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	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	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	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	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
JK1548	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

* Individuals that where included in previous studies, where *Y. pestis* was detected by PCR[14-17]
**Previously published Ellwangen individual[5]

Supplementary Methods

Archaeological context information

"Laishevo III cemetery", 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 the pottery, few additional artefacts were identified within the burials. These were some 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 Mongolian time (the Golden Horde period of Volga Bulgaria - not 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. It is suggested that this was a man of 35-45 years old.

Burial 27, where individual LAI009 was found, was also at a depth of 40 cm and 3 skeletons were found within. The skeletal remains of individual LAI009 (skeleton B from burial 27) were not well preserved. The skull was found disarticulated from the rest of the skeleton, laying in the western part of the grave pit, whereas the rest of the bones were placed disorderly between skeletons A and C. An iron buckle was found inside the skull, and the only bronze earring discovered in this site was also found closer to this skull. 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.

Toulouse, France, "16 rue des Trente Six Ponts"

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[18]. 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 "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[18]. Then, the artefacts uncovered in the deposits stratigraphically associated with these graves can be dated between the second half of the 13th century and the end of the 14th century.

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 monetary hoards buried with the bodies. Two graves contain coins from the end of the 13th century (1245-1270). Six burials were dated by 14C, 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 monetary finds. Their contemporaneity with the plague episode of Late Middle Ages can be demonstrated by the such depositions in a double burial as well as in two of the mass graves. Within the double burial a first deposit of 34 "double tournois" of Philippe VI issued between February 1347 and April 1350 were found with one of the bodies. The second deposit, related to the second burial, consists of two "double parisis" of Philip VI issued between April 1347 and August 1350. Then, one of the mass graves contains a deposit of 36 coins issued between March 1347 and August 1348, as well as two "double tournois" of Philip VI there issued between 1348 and 1350.

The grave SP1350 corresponds to a double burial dug in stratigraphical deposits post-dating the 13th century and has the same orientation as the other multiple graves dated to the same period. It is therefore contemporary with the plague episode that affected Toulouse during the second half of the 14th century. The first, skeleton SQ1352 (TRP002), corresponds to an adult woman, aged 20 to 49 years, buried at the same time as the individual SQ1353 that corresponds to 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.

London, England, "The New Churchyard"

Population growth in 16th-century London, coupled with a severe outbreak of plague in 1563, led to the of the 'New Churchyard' opening in 1569, a municipal, non-parochial burial ground[19]. 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 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 and 1665, with parish records confirming that many 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)[19] 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 uncoffined burials, up to eight 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 and 1625 were the most likely.

Nabburg, Germany, "Sankt Johans Freidhof"

The "Sankt Johans Freidhof" (sic, see Hensch 2014)[20] churchyard 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 century 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 9 multiple burials were found with between two and four individuals. The four individuals 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 and 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.

Manching-Pichl, Germany, St. Leonhardi

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[17]. However, the removal of the ground floor within the nave also revealed the remains of six additional individuals that were partially considered as 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 archaeological dating of the mass grave poses a challenge, since the only artefact found in association with the individuals is a fibula with two contradictive dates suggesting either the 13th or 15th century. 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[14-16], which could be explained by erroneously assigned scattered remains of earlier burials in this area.

The mass grave was repeatedly subject to aDNA studies on the presence of *Y. pestis* DNA investigated with PCR and qPCR assays[14-17,21,22]. In this study, the 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 tables 1, 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). Based on the resultant ^{14}C dates (2-sigma: 1298-1390 calAD), and the fact that plague did not enter Bavaria before 1348 during the second pandemic, we in addition consider the archaeological findings to give a plausible date interval to the deposition. The archaeological information indicates that (1) the sacristy was built during the second half of the 15th century and (2) reveals a possible 15th century artefact (fibula) found within the mass grave. We, therefore, suggest a date range of 1348 – 1450 AD for the burial. Such a date is in line with our genomic results, which revealed a more derived phylogenetic placement of the MAN008 genome compared to Black Death isolates from Europe (Figure 2).

Starnberg, Germany, Possenhofener Str. 3

The excavation Starnberg, Possenhofener Str. 3, from 2003 to 2007 revealed 365 burials of the churchyard attributed to the parish church of St. Benedikt, first mentioned in 1220 and demolished between 1764 and 1816[23]. The excavated burials are estimated to represent only one third of the cemetery and are mostly dated to the 17th to 18th century. However, five stone plate graves indicate an occupation of the site from as early as the 7th century on, also corresponding with archaeological features of earlier church buildings. One triple burial was the only multiple burial found at this site and was examined here for the presence of *Y. pestis*. Individual STA001 (arch. ID 207, infans II to juvenile, presumably male) was tested positively. It was simultaneously buried together with two children aged 0-6 and 6-12 in regular supine position. Of the two rosaries found with STA001 and STA003 (arch. ID 214), one can be dated to the 15th century. Together with the 1-sigma interval of the radiocarbon dating, the date can be narrowed down to 1433-1494.

Landsberg am Lech, Germany, Kirchhof St. Johannis

Excavated between 2015 and 2016, the churchyard of St. Johannis revealed remains of more than 900 individuals[24]. The usage time of this churchyard could be narrowed down by historical documents to between 1507 and 1806. On the lowest level, eight multiple burials with four to eleven individuals were found. Furthermore, in seven of the multiple burials and one single burial, the individuals were laid in or covered with quicklime. In combination with the high percentage of subadult individuals (nearly 60 %) and the absence of any evidence for lethal injuries, this prompted the excavators to suspect an epidemic background.

Three individuals tested qPCR positive for *Y. pestis*: 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.

Stans, Switzerland, Nägeligasse

The excavation of the churchyard of "St. Peter und Paul" took place between 2015 and 2016. Besides 122 single burials, the excavation revealed three multiple burials with a minimum number of four, 16 and 26 individuals respectively[25]. The excavators were able to reconstruct that the individuals were likely clothed and buried simultaneously.

Historical sources report the presence of plague during the Black Death as well as a a minimum of seven succeeding outbreaks between 1493 and 1630[26]. 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.

Brandenburg an der Havel, Germany, Domlinden 12

During a survey in 2011, the burial of three individuals was found on the "Dominsel", the center 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[27]. 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. Based on the context and isotope analyses on Oxygen and Strontium, hinting towards Scandinavia or the Baltic region, it was hypothesized that the individuals were foreign soldiers housed with civilians during the occupation of the city by Swedish troops in 1631 as part of the Thirty Years' War. 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[15,16]. Here, we reconstructed whole genomes from those two individuals, namely BRA001 (arch. ID 1, late adult male) and BRA003 (arch. ID 3, late juvenile male).

Supplementary References

1   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).
2   Chain, P. S. *et al.* Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. *Proc Natl Acad Sci U S A* **101**, 13826-13831, doi:10.1073/pnas.0404012101 (2004).
3   Bos, K. I. *et al.* A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506-510, doi:10.1038/nature10549 (2011).
4   Bos, K. I. *et al.* Eighteenth century genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, doi:10.7554/eLife.12994 (2016).
5   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).
6   Team, R. D. C. R: A Language and Environment for Statistical Computing: Vienna, Austria. (2015).
7   Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).
8   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).
9   Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in Yersinia pestis. *Infect. Immun.* **77**, 2242-2250, doi:10.1128/IAI.01477-08 (2009).

10   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).

11   Wickham, H. *ggplot2: elegant graphics for data analysis.* (Springer, 2016).

12   R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2015).

13   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).

14   Wiechmann, I., Harbeck, M. & Grupe, G. Yersinia pestis DNA sequences in late medieval skeletal finds, Bavaria. *Emerging Infect. Dis.* **16**, 1806 (2010).

15   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).

16   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).

17   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).

18   Paya, D. & Catalo, J. *Le cimetière Saint-Michel de Toulouse.* (CNRS, 2011).

19   Keily, J. *Tunnel: the archaeology of Crossrail.* (2017).

20   Hensch, M. *Sankt Johans Freidhof in Nabburg - Gewöhnliche und ungewöhnliche Einblicke in die spätmittelalterliche Begräbniskultur Ostbayerns.*, 423–440 (2014).

21   Popper, H. & Schaffner, F. [Eppinger and the intrahepatic cholestasis]. *Wien Klin Wochenschr* **78**, 675-680 (1966).

22   Garrelt, C. *Molekulargenetische Untersuchung der Bestatteten eines vermuteten Pestfriedhofs des 14. Jahrhunderts (Manching-Pichl)*, Diplomarbeit, (2002).

23   Later, V. C. Merowingerzeitliche Tuffplattengräber und frühmittelalterliche Kirchenbauten – Zu den Anfängen der ehemaligen Pfarrkirche St. Benedikt in Starnberg. *Bericht Der Bayerischen Bodendenkmalpflege* **52**, 373–402 (2010).

24   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).

25   Krämer, D. Geschichte des Kantons Nidwalden. Bd. 1: Von der Urzeit bis 1850. *Stans: Historischer Verein Nidwalden* (2014).

26   Krämer, D. Bevölkerung und Wegnetz: Leben in Abgeschiedenheit (1550–1850). (2014).

27   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, editor* **21**, 2011-2012 (2012).

Supplementary data 1: Isolation, radiocarbon and archaeological dates used for BEAST substitution rate variation analysis

Strain Name	Strain ID	Pubication	Year of isolation (modern strains)	Years since 2006	Lower bound (since 2006)	Upper bound (since 2006)
	Georgia_1412	Zhgenti et al., 2015	N/A*	25	0	108
	Georgia_1413	Zhgenti et al., 2015	N/A*	25	0	108
	Georgia_1670	Zhgenti et al., 2015	N/A*	25	0	108
	Georgia_3067	Zhgenti et al., 2015	N/A*	25	0	108
	Georgia_3770	Zhgenti et al., 2015	N/A*	25	0	108
	Georgia_8787	Zhgenti et al., 2015	N/A*	25	0	108
	Kyrgyzstan_790	Zhgenti et al., 2015	N/A*	25	0	108
	Armenia_14735	Zhgenti et al., 2015	N/A*	25	0	108
	Armenia_1522	Zhgenti et al., 2015	N/A*	25	0	108
	Azerbaijan_1045	Zhgenti et al., 2015	N/A*	25	0	108
	RussianFederation_2944	Zhgenti et al., 2015	N/A*	25	0	108
	6904	Kislichkina et al., 2015	1984	22	N/A	N/A
	7761	Kislichkina et al., 2015	1983	23	N/A	N/A
	6990	Kislichkina et al., 2015	1971	35	N/A	N/A
	6974	Kislichkina et al., 2015	1967	39	N/A	N/A
	6757	Kislichkina et al., 2015	1968	38	N/A	N/A
	6984	Kislichkina et al., 2015	1969	37	N/A	N/A
	7832	Kislichkina et al., 2015	1978	28	N/A	N/A
	6300	Kislichkina et al., 2015	1971	35	N/A	N/A
	6536	Kislichkina et al., 2015	1976	30	N/A	N/A
	6540	Kislichkina et al., 2015	1989	17	N/A	N/A
	7019	Kislichkina et al., 2015	1980	26	N/A	N/A
	7074	Kislichkina et al., 2015	1980	26	N/A	N/A
	7812	Kislichkina et al., 2015	2002	4	N/A	N/A
	7075	Kislichkina et al., 2015	1961	45	N/A	N/A
	6706	Kislichkina et al., 2015	1986	20	N/A	N/A
	6906	Kislichkina et al., 2015	1974	32	N/A	N/A
	6213	Kislichkina et al., 2015	1972	34	N/A	N/A
	6216	Kislichkina et al., 2015	1984	22	N/A	N/A
	6304	Kislichkina et al., 2015	1979	27	N/A	N/A
0.ANT1a	42013	Cui et al., 2013	1971	35	N/A	N/A
0.ANT1b	CMCC49003	Cui et al., 2013	1976	30	N/A	N/A
0.ANT1c	945	Cui et al., 2013	1994	12	N/A	N/A
0.ANT1d	164	Cui et al., 2013	1985	21	N/A	N/A
0.ANT1e	CMCC8211	Cui et al., 2013	1982	24	N/A	N/A
0.ANT1f	42095	Cui et al., 2013	2001	5	N/A	N/A
0.ANT1g	CMCC42007	Cui et al., 2013	1967	39	N/A	N/A
0.ANT1h	CMCC43032	Cui et al., 2013	1980	26	N/A	N/A
0.ANT2a	B42003004	Cui et al., 2013	2003	3	N/A	N/A
0.ANT2a	2330	Cui et al., 2013	2003	3	N/A	N/A
0.ANT3a	CMCC38001	Cui et al., 2013	1979	27	N/A	N/A
0.ANT3b	A1956001	Cui et al., 2013	1956	50	N/A	N/A
0.ANT3c	42082	Cui et al., 2013	1995	11	N/A	N/A
0.ANT3d	CMCC21106	Cui et al., 2013	2001	5	N/A	N/A
0.ANT3e	42091	Cui et al., 2013	1999	7	N/A	N/A
0.PE2a	PestoidesF	Cui et al., 2013	1984	22	N/A	N/A
0.PE2b	G8786	Cui et al., 2013	N/A*	25	0	108
0.PE4Aa	12	Cui et al., 2013	2004	2	N/A	N/A
0.PE4Ab	9	Cui et al., 2013	2004	2	N/A	N/A
0.PE4Ba	PestoidesA	Cui et al., 2013	N/A*	25	0	108
0.PE4Ca	CMCCN010025	Cui et al., 2013	2000	6	N/A	N/A
0.PE4Cb	M0000002	Cui et al., 2013	2001	5	N/A	N/A
0.PE4Cc	CMCC18019	Cui et al., 2013	2001	5	N/A	N/A
0.PE4Cd	CMCC93014	Cui et al., 2013	1970	36	N/A	N/A
0.PE4Ce	CMCC91090	Cui et al., 2013	1970	36	N/A	N/A
0.PE4Cf	91001	Cui et al., 2013	1970	36	N/A	N/A
0.PE7b	620024	Cui et al., 2013	1962	44	N/A	N/A
1.ANTa	Antiqua	Cui et al., 2013	1965	41	N/A	N/A
1.ANTb	UG05	Cui et al., 2013	2004	2	N/A	N/A
1.IN1a	CMCC11001	Cui et al., 2013	1954	52	N/A	N/A
1.IN1b	780441	Cui et al., 2013	1978	28	N/A	N/A
1.IN1c	K21985002	Cui et al., 2013	1985	21	N/A	N/A
1.IN2a	CMCC640047	Cui et al., 2013	1964	42	N/A	N/A
1.IN2b	30017	Cui et al., 2013	1976	30	N/A	N/A
1.IN2c	CMCC31004	Cui et al., 2013	1990	16	N/A	N/A
1.IN2d	C1975003	Cui et al., 2013	1975	31	N/A	N/A
1.IN2e	C1989001	Cui et al., 2013	1989	17	N/A	N/A
1.IN2f	710317	Cui et al., 2013	1971	35	N/A	N/A
1.IN2g	CMCC05013	Cui et al., 2013	1988	18	N/A	N/A
1.IN2h	5	Cui et al., 2013	2004	2	N/A	N/A
1.IN2i	CMCC10012	Cui et al., 2013	1964	42	N/A	N/A
1.IN2j	CMCC27002	Cui et al., 2013	1991	15	N/A	N/A
1.IN2k	970754	Cui et al., 2013	1997	9	N/A	N/A
1.IN2l	D1991004	Cui et al., 2013	1991	15	N/A	N/A
1.IN2m	D1964002	Cui et al., 2013	1964	42	N/A	N/A
1.IN2n	CMCC02041	Cui et al., 2013	1965	41	N/A	N/A
1.IN2o	CMCC03001	Cui et al., 2013	1954	52	N/A	N/A
1.IN2p	D1982001	Cui et al., 2013	1982	24	N/A	N/A
1.IN2q	D1964001	Cui et al., 2013	1964	42	N/A	N/A
1.IN3a	F1954001	Cui et al., 2013	1954	52	N/A	N/A
1.IN3b	E1979001	Cui et al., 2013	1979	27	N/A	N/A
1.IN3c	CMCC84038	Cui et al., 2013	1982	24	N/A	N/A
1.IN3d	YN1683	Cui et al., 2013	1977	29	N/A	N/A
1.IN3e	YN472	Cui et al., 2013	1957	49	N/A	N/A
1.IN3f	YN1065	Cui et al., 2013	1954	52	N/A	N/A
1.IN3g	E1977001	Cui et al., 2013	1977	29	N/A	N/A
1.IN3h	CMCC84033	Cui et al., 2013	1979	27	N/A	N/A
1.IN3i	CMCC84046	Cui et al., 2013	1984	22	N/A	N/A
1.ORI1a	CMCC114001	Cui et al., 2013	1952	54	N/A	N/A
1.ORI1b	India195	Cui et al., 2013	1898	108	N/A	N/A
1.ORI1c	F1946001	Cui et al., 2013	1946	60	N/A	N/A
1.ORI1d	CA88	Cui et al., 2013	1988	18	N/A	N/A
1.ORI1e	CO92	Cui et al., 2013	1992	14	N/A	N/A
1.ORI2a	YN2179	Cui et al., 2013	1995	11	N/A	N/A
1.ORI2b	CMCCK110001	Cui et al., 2013	1991	15	N/A	N/A
1.ORI2c	YN2551	Cui et al., 2013	2002	4	N/A	N/A
1.ORI2d	YN2588	Cui et al., 2013	2000	6	N/A	N/A
1.ORI2e	F1991016	Cui et al., 2013	1991	15	N/A	N/A
1.ORI2f	CMCC87001	Cui et al., 2013	1982	24	N/A	N/A
1.ORI2g	F1984001	Cui et al., 2013	1984	22	N/A	N/A
1.ORI2h	YN663	Cui et al., 2013	1982	24	N/A	N/A
1.ORI2i	CMCCK100001	Cui et al., 2013	1991	15	N/A	N/A
1.ORI3a	EV76	Cui et al., 2013	1922	84	N/A	N/A
1.ORI3b	MG05	Cui et al., 2013	2005	1	N/A	N/A
1.ORI3c	IP275	Cui et al., 2013	1995	11	N/A	N/A
2.ANT1a	34008	Cui et al., 2013	1968	38	N/A	N/A
2.ANT1b	34202	Cui et al., 2013	1990	16	N/A	N/A
2.ANT1c	Nepal516	Cui et al., 2013	1967	39	N/A	N/A
2.ANT2a	2	Cui et al., 2013	2004	2	N/A	N/A
2.ANT2b	351001	Cui et al., 2013	1996	10	N/A	N/A
2.ANT2c	CMCC347001	Cui et al., 2013	1996	10	N/A	N/A
2.ANT2d	G1996006	Cui et al., 2013	1996	10	N/A	N/A
2.ANT2e	G1996010	Cui et al., 2013	1996	10	N/A	N/A
2.ANT2f	CMCC348002	Cui et al., 2013	1998	8	N/A	N/A
2.ANT3a	CMCC92010	Cui et al., 2013	1971	35	N/A	N/A
2.ANT3b	CMCC95001	Cui et al., 2013	1970	36	N/A	N/A
2.ANT3c	CMCC96001	Cui et al., 2013	1957	49	N/A	N/A
2.ANT3d	CMCC96007	Cui et al., 2013	1989	17	N/A	N/A
2.ANT3e	CMCC67001	Cui et al., 2013	1959	47	N/A	N/A
2.ANT3f	CMCC104003	Cui et al., 2013	1989	17	N/A	N/A
2.ANT3g	CMCC51020	Cui et al., 2013	1954	52	N/A	N/A
2.ANT3h	CMCC106002	Cui et al., 2013	1996	10	N/A	N/A
2.ANT3i	CMCC64001	Cui et al., 2013	1957	49	N/A	N/A
2.ANT3j	H1959004	Cui et al., 2013	1959	47	N/A	N/A
2.ANT3k	5761	Cui et al., 2013	1956	50	N/A	N/A
2.ANT3l	735	Cui et al., 2013	1965	41	N/A	N/A
2.MED1a	KIM	Cui et al., 2013	1968	38	N/A	N/A
2.MED1b	2506	Cui et al., 2013	2005	1	N/A	N/A
2.MED1c	2654	Cui et al., 2013	2006	0	N/A	N/A
2.MED1d	2504	Cui et al., 2013	2005	1	N/A	N/A
2.MED2b	91	Cui et al., 2013	1987	19	N/A	N/A
2.MED2c	K11973002	Cui et al., 2013	1973	33	N/A	N/A
2.MED2d	A1973001	Cui et al., 2013	1973	33	N/A	N/A
2.MED2e	7338	Cui et al., 2013	1973	33	N/A	N/A

2.MED3a	J1963002	Cui et al., 2013	1963	43	N/A	N/A
2.MED3b	CMCC125002	Cui et al., 2013	1964	42	N/A	N/A
2.MED3c	I1969003	Cui et al., 2013	1969	37	N/A	N/A
2.MED3d	J1978002	Cui et al., 2013	1978	28	N/A	N/A
2.MED3f	I1970005	Cui et al., 2013	1970	36	N/A	N/A
2.MED3g	CMCC99103	Cui et al., 2013	1970	36	N/A	N/A
2.MED3h	CMCC90027	Cui et al., 2013	1970	36	N/A	N/A
2.MED3i	CMCC92004	Cui et al., 2013	1957	49	N/A	N/A
2.MED3j	I2001001	Cui et al., 2013	2001	5	N/A	N/A
2.MED3k	CMCC12003	Cui et al., 2013	1961	45	N/A	N/A
2.MED3l	I1994006	Cui et al., 2013	1994	12	N/A	N/A
2.MED3m	SHAN11	Cui et al., 2013	2006	0	N/A	N/A
2.MED3n	SHAN12	Cui et al., 2013	2006	0	N/A	N/A
2.MED3o	I1991001	Cui et al., 2013	1991	15	N/A	N/A
2.MED3p	CMCC107004	Cui et al., 2013	2003	3	N/A	N/A
3.ANT1a	7	Cui et al., 2013	2004	2	N/A	N/A
3.ANT1b	CMCC71001	Cui et al., 2013	1961	45	N/A	N/A
3.ANT1c	C1976001	Cui et al., 2013	1976	30	N/A	N/A
3.ANT1d	71021	Cui et al., 2013	1989	17	N/A	N/A
3.ANT2a	MGJZ6	Cui et al., 2013	1997	9	N/A	N/A
3.ANT2b	MGJZ7	Cui et al., 2013	1997	9	N/A	N/A
3.ANT2c	MGJZ9	Cui et al., 2013	1998	8	N/A	N/A
3.ANT2d	MGJZ11	Cui et al., 2013	2000	6	N/A	N/A
3.ANT2e	MGJZ3	Cui et al., 2013	1980	26	N/A	N/A
4.ANT1a	MGJZ12	Cui et al., 2013	2002	4	N/A	N/A
ancient Brach 0	RISE505	Rasmussen et al., 2015	N/A	3691	3631	3750
ancient Brach 0	RISE509	Rasmussen et al., 2015	N/A	4785	4678	4892
ancient Brach 0	RT5	Spyrou et al., 2018	N/A	3845	3760	3924
ancient Brach 0	Justinian 2148	Feldman et al., 2016	N/A	1509	1438	1544
ancient Branch 1	LAI009	this study	N/A	656	606	706
ancient Branch 1	London BD	Bos et al., 2011	N/A	659	658	660
ancient Branch 1	Barcelona 3031	Spyrou et al., 2016	N/A	646	586	706
ancient Branch 1	NAB003	this study	N/A	664	614	714
ancient Branch 1	Bolgar 2370	Spyrou et al., 2016	N/A	625	606	644
ancient Branch 1	Manching	this study	N/A	607	556	658
ancient Branch 1	STA001	this study	N/A	540	506	573
ancient Branch 1	Ellwngen ELLW549_O	Spyrou et al., 2016	N/A	450	379	521
ancient Branch 1	Landsberg LBG	this study	N/A	463	374	551
ancient Branch 1	Stans STN	this study	N/A	446	371	521
ancient Branch 1	Brandenburg BRA	this study	N/A	374	358	389
ancient Branch 1	New Churchyard BED	this study	N/A	411	376	446
ancient Branch 1	Marseille OBS	Bos et al., 2016	N/A	285	284	286
ancient Branch 1	London 6330	Bos et al., 2011	N/A	631	606	656

* For modern strains with unknown isolation dates we put 25 years of age and uniform distribution between 0 and 108

List of publications:

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).

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).

Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci U S A* **110**, 577-582, doi:10.1073/pnas.1205750110 (2013).

Bos, K. I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506-510, doi:10.1038/nature10549 (2011)

Spyrou, M. A. *et al.* Historical *Y. pestis* Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host & Microbe* **19**, 874-881 (2016)

Bos, K. I. *et al.* Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* **5**, e12994, doi:10.7554/eLife.12994 (2016)

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).

Feldman, M. *et al.* A High-Coverage *Yersinia pestis* Genome from a Sixth-Century Justinianic Plague Victim. *Mol. Biol. Evol.* **33**, 2911-2923, doi:10.1093/molbev/msw170 (2016).

Spyrou, M. A. et al. Analysis of 3800-year-old *Yersinia pestis* suggests Bronze Age origin for bubonic plague. Accepted *Nature Communications* (2018)

Supplementary data 2: Gene annotation and effects of SNPs falling between the Ellwangen (ELLW) 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	Gene_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	Gene_YPO0090	glpK%27	pseudogene	E/K	Gaa/Aaa	325
200723	C	T	NON_SYNONYMOUS_CODING	Gene_YPO0182	tauA	taurine transporter substrate binding subunit	T/M	aCg/aTg	135
868549	G	C	NON_SYNONYMOUS_CODING	Gene_YPO0792	ygeD	lysophospholipid transporter LplT	C/W	tgC/tgG	209
965281	C	A	NON_SYNONYMOUS_CODING	Gene_YPO0880	YPO0880	primase	S/Y	tCt/tAt	95
1168951	G	T	NON_SYNONYMOUS_CODING	Gene_YPO1028	YPO1028	cysteine sulfinate desulfinase	R/S	Cgt/Agt	143
1189479	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1048	dxr	1-deoxy-D-xylulose 5-phosphate reductoisomerase	P/S	Cca/Tca	189
1378105	G	T	NON_SYNONYMOUS_CODING	Gene_YPO1219	YPO1219	hypothetical protein	H/N	Cat/Aat	593
1451124	T	G	NON_SYNONYMOUS_CODING	Gene_YPO1291	YPO1291	carbohydrate kinase	V/G	gTg/gGg	394
1511518	A	G	NON_SYNONYMOUS_CODING	Gene_YPO1347	YPO1347	hypothetical protein	N/D	Aat/Gat	167
1586982	C	A	NON_SYNONYMOUS_CODING	Gene_YPO1403	mukF	condesin subunit F	S/Y	tCc/tAc	163
1708192	C	A	NON_SYNONYMOUS_CODING	Gene_YPO1504	YPO1504	hypothetical protein	P/Q	cCa/cAa	285
1935112	C	A	NON_SYNONYMOUS_CODING	Gene_YPO1696	YPO1696	outer membrane usher protein	Q/H	caG/caT	208
2414599	T	C	NON_SYNONYMOUS_CODING	Gene_YPO2145	YPO2145	SpoVR family protein	K/E	Aaa/Gaa	85
2472383	A	G	NON_SYNONYMOUS_CODING	Gene_YPO2196	ispZ	Involved in cell division	F/S	tTc/tCc	85
2507983	T	G	NON_SYNONYMOUS_CODING	Gene_YPO2233	YPO2233	hypothetical protein	T/P	Acc/Ccc	72
3229407	T	C	NON_SYNONYMOUS_CODING	Gene_YPO2887	yapB	pseudogene	V/A	gTt/gCt	374
3407572	A	T	NON_SYNONYMOUS_CODING	Gene_YPO3049	YPO3049	binding protein-dependent transporter inner membrane protein	L/Q	cTg/cAg	216
3610371	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3244	fadE	acyl-CoA dehydrogenase	A/V	gCc/gTc	292
3613964	C	A	NON_SYNONYMOUS_CODING	Gene_YPO3246	hmwC	accessory processing protein	D/Y	Gat/Tat	252
3620114	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3247	hmwA	adhesin	A/V	gCg/gTg	164
3782640	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3389	hemL	glutamate-1-semialdehyde aminotransferase	G/D	gGc/gAc	31
3973901	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3559	YPO3559	hypothetical protein	D/N	Gac/Aac	145
4363505	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3888	ilvC	ketol-acid reductoisomerase	V/I	Gtt/Att	118
4396236	G	T	NON_SYNONYMOUS_CODING	Gene_YPO3914	sthA	catalyzes the conversion of NADPH to NADH	D/Y	Gat/Tat	116
4616904	T	C	NON_SYNONYMOUS_CODING	Gene_YPO4095	recF	recombination protein F: required for DNA replication	S/G	Agt/Ggt	54
2071670	G	T	STOP_GAINED	Gene_YPO1826	fliJ	flagellar biosynthesis chaperone	S/*	tCa/tAa	142
173032	C	T	SYNONYMOUS_CODING	Gene_YPO0158	cysG	siroheme synthase	K/K	aaG/aaA	178
477107	C	T	SYNONYMOUS_CODING	Gene_YPO0452	slt	lytic murein transglycosylase	L/L	Ctg/Ttg	364
869820	A	G	SYNONYMOUS_CODING	Gene_YPO0793	aas	acyl-ACP synthetase	F/F	ttT/ttC	503
951295	C	T	SYNONYMOUS_CODING	Gene_YPO0863	YPO0863	hypothetical protein	L/L	Ctg/Ttg	182
1159539	T	A	SYNONYMOUS_CODING	Gene_YPO1020	recB	helicase/nuclease	P/P	ccT/ccA	1181
1724647	C	T	SYNONYMOUS_CODING	Gene_YPO1517	YPO1517	sugar ABC transporter	S/S	tcG/tcA	130
2076353	C	T	SYNONYMOUS_CODING	Gene_YPO1830	fliF	flagellar MS-ring protein	T/T	acG/acA	223
3620500	G	A	SYNONYMOUS_CODING	Gene_YPO3247	hmwA	adhesin	G/G	ggC/ggT	35
3944305	C	A	SYNONYMOUS_CODING	Gene_YPO3531	YPO3531	iron-sulfur cluster repair di-iron protein	I/I	atC/atA	81
4200639	C	A	SYNONYMOUS_CODING	Gene_YPO3746	rpoC	DNA-directed RNA polymerase subunit beta	L/L	ctG/ctT	282

Supplementary data 3: 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	Gene_YPO0002	asnC	DNA-binding transcriptional regulator AsnC	A/T	Gct/Act	34
54102	A	T	NON_SYNONYMOUS_CODING	Gene_YPO0041	ligB	NAD-dependent DNA ligase LigB	T/S	Acc/Tcc	106
178028	C	T	NON_SYNONYMOUS_CODING	Gene_YPO0162	codA	cytosine deaminase	A/V	gCg/gTg	63
225409	C	T	NON_SYNONYMOUS_CODING	Gene_YPO0216	rpsC	30S ribosomal protein S3	T/I	aCt/aTt	121
274360	C	T	NON_SYNONYMOUS_CODING	Gene_YPO0274	YPO0274	hypothetical protein	P/L	cCg/cTg	174
419209	G	A	NON_SYNONYMOUS_CODING	Gene_YPO0401	YPO0401	transcriptional regulator	D/N	Gat/Aat	289
525887	T	A	NON_SYNONYMOUS_CODING	Gene_YPO0494	surA	peptidyl-prolyl cis-trans isomerase SurA	Q/L	cAg/cTg	58
535221	T	G	NON_SYNONYMOUS_CODING	Gene_YPO0501	YPO0501	hypothetical protein	S/A	Tcc/Gcc	313
574078	C	T	NON_SYNONYMOUS_CODING	Gene_YPO0530	leuD	3-isopropylmalate dehydratase small subunit	G/S	Ggc/Agc	53
727192	G	A	NON_SYNONYMOUS_CODING	Gene_YPO0668	parE	DNA topoisomerase IV subunit B	A/T	Gcc/Acc	53
759843	C	T	NON_SYNONYMOUS_CODING	Gene_YPO0698	YPO0698	outer membrane usher protein	L/F	Ctc/Ttc	472
935991	G	A	NON_SYNONYMOUS_CODING	Gene_YPO0850a	YPO0850a	PTS system glucose/sucrose specific transporter subunit IIB	T/I	aCt/aTt	21
1151889	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1018	recC	catalyzes ATP-dependent exonucleolytic cleavage	P/L	cCt/cTt	744
1212998	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1069	YPO1069	hypothetical protein	G/S	Ggc/Agc	70
1265901	A	G	NON_SYNONYMOUS_CODING	Gene_YPO1118	cydB	cytochrome D ubiquinol oxidase subunit II	N/D	Aat/Gat	112
1303163	A	T	NON_SYNONYMOUS_CODING	Gene_YPO1156	uvrB	excinuclease ABC subunit B	D/V	gAt/gTt	150
1333620	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1185	YPO1185	ABC transport membrane permease	T/I	aCt/aTt	5
1441794	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1283	uxuA	mannonate dehydratase	A/V	gCt/gTt	195
1555356	G	A	NON_SYNONYMOUS_CODING	Gene_YPO1380	YPO1380	MFS family transporter protein	D/N	Gat/Aat	373
1705844	A	T	NON_SYNONYMOUS_CODING	Gene_YPO1502	YPO1502	alcohol dehydrogenase	A/T	Gcc/Acc	71
1707708	G	A	NON_SYNONYMOUS_CODING	Gene_YPO1504	YPO1504	hypothetical protein	G/S	Ggc/Agc	124
1756795	G	A	NON_SYNONYMOUS_CODING	Gene_YPO1541	gnd	6-phosphogluconate dehydrogenase	G/S	Ggc/Agc	174
1801784	G	A	NON_SYNONYMOUS_CODING	Gene_YPO1579	YPO1579	C4-dicarboxylate transporter substrate-binding protein	P/S	Ccc/Tcc	44
1850729	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1626	lolC	outer membrane-specific lipoprotein transporter subunit LolC	P/S	Ccg/Tcg	81
2016117	T	A	NON_SYNONYMOUS_CODING	Gene_YPO1770	hpaC	4-hydroxyphenylacetate 3-monooxygenase coupling protein	L/Q	cTg/cAg	135
2115661	C	T	NON_SYNONYMOUS_CODING	Gene_YPO1868	YPO1868	hypothetical protein	E/K	Gaa/Aaa	161
2484440	A	G	NON_SYNONYMOUS_CODING	Gene_YPO2210	YPO2210	hypothetical protein	V/A	gTg/gCg	18
2681696	T	C	NON_SYNONYMOUS_CODING	Gene_YPO2387	purR	DNA-binding transcriptional repressor PurR	F/S	tTt/tCt	221
2731263	C	T	NON_SYNONYMOUS_CODING	Gene_YPO2429	pheS	phenylalanyl-tRNA synthetase subunit alpha	V/I	Gta/Ata	67
2907502	G	T	NON_SYNONYMOUS_CODING	Gene_YPO2585	YPO2585	carbohydrate kinase	D/Y	Gat/Tat	361
2924187	G	A	NON_SYNONYMOUS_CODING	Gene_YPO2602	rlpA	rare lipoprotein A	P/L	cCa/cTa	36
2950840	C	T	NON_SYNONYMOUS_CODING	Gene_YPO2625	nagC	N-acetylglucosamine regulatory protein	G/E	gGg/gAg	61
3099378	G	A	NON_SYNONYMOUS_CODING	Gene_YPO2765	asd	semialdehyde dehydrogenase	R/H	cGt/cAt	76
3501556	A	C	NON_SYNONYMOUS_CODING	Gene_YPO3142	amtB	ammonium transporter	V/G	gTc/gGc	324
3586778	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3222	proB	gamma-glutamyl kinase	S/F	tCt/tTt	50
3698745	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3316	rbsC	sugar transport system permease	L/F	Ctc/Ttc	269
3763970	T	C	NON_SYNONYMOUS_CODING	Gene_YPO3375	sodC	superoxide dismutase	D/G	gAt/gGt	33
3771852	T	A	NON_SYNONYMOUS_CODING	Gene_YPO3381	barA	hybrid sensory histidine kinase BarA	L/Q	cTg/cAg	156
3904960	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3496	infB	translation initiation factor IF-2	A/V	gCg/gTg	614
3977650	G	C	NON_SYNONYMOUS_CODING	Gene_YPO3564	YPO3564	hypothetical protein	D/E	gaC/gaG	90
4061493	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3640	YPO3640	hypothetical protein	P/S	Ccc/Tcc	29
4065904	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3642b	YPO3642b	ncRNA-cspA thermoregulator	A/V	gCt/gTt	134
4076324	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3657	panF	sodium/panthothenate symporter	S/N	aGc/aAc	371
4318701	T	C	NON_SYNONYMOUS_CODING	Gene_YPO3848	cyaA	adenylate cyclase	D/G	gAt/gGt	157
4360567	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3887	YPO3887	pseudogene	R/H	cGc/cAc	544
4427178	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3940	glgC	glucose-1-phosphate adenylyltransferase	S/N	aGc/aAc	404
4477698	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3976	YPO3976	hypothetical protein	A/T	Gca/Aca	237
4525085	A	G	NON_SYNONYMOUS_CODING	Gene_YPO4013	yhjW	phosphoethanolamine transferase	S/P	Tca/Cca	35
4558980	G	A	NON_SYNONYMOUS_CODING	Gene_YPO4042	YPO4042	fimbrial protein	A/V	gCa/gTa	752
588988	C	A	START_LOST	Gene_YPO0544	YPO0544	hypothetical protein	M/I	atG/atT	1
2016126	T	A	STOP_GAINED	Gene_YPO1770	hpaC	4-hydroxyphenylacetate 3-monooxygenase coupling protein	L/*	tTg/tAg	138
2751771	G	A	STOP_GAINED	Gene_YPO2451	YPO2451	hypothetical protein	Q/*	Cag/Tag	68
168535	A	G	SYNONYMOUS_CODING	Gene_YPO0154	dam	DNA adenine methylase	R/R	cgA/cgG	22
367638	C	T	SYNONYMOUS_CODING	Gene_YPO0357	frdD	fumarate reductase subunit D	L/L	ttG/ttA	107
537476	C	T	SYNONYMOUS_CODING	Gene_YPO0504	YPO0504	hypothetical protein	A/A	gcC/gcT	150
586040	C	T	SYNONYMOUS_CODING	Gene_YPO0540	ilvH	acetolactate synthase small subunit	G/G	ggC/ggT	98
904076	C	T	SYNONYMOUS_CODING	Gene_YPO0824	YPO0824	hypothetical protein	L/L	Ctg/Ttg	433
1028008	G	A	SYNONYMOUS_CODING	Gene_YPO0935	gshB	glutathione synthetase	G/G	ggG/ggA	271
1200873	C	T	SYNONYMOUS_CODING	Gene_YPO1059	dnaE	DNA polymerase III subunit alpha	L/L	Ctg/Ttg	10
1578826	C	T	SYNONYMOUS_CODING	Gene_YPO1395	msbA	lipid transporter ATP-binding protein/permease	L/L	ctC/ctT	575
1748342	G	A	SYNONYMOUS_CODING	Gene_YPO1536	YPO1536	iron-siderophore transporter substrate-binding protein	L/L	Ctg/Ttg	120
2112974	A	G	SYNONYMOUS_CODING	Gene_YPO1866	uvrC	excinuclease ABC subunit C	G/G	ggG/ggA	183
2542828	A	G	SYNONYMOUS_CODING	Gene_YPO2262	YPO2262	hypothetical protein	I/I	atT/atC	233
2596092	C	T	SYNONYMOUS_CODING	Gene_YPO2308	rstA	DNA-binding transcriptional regulator RstA	R/R	cgC/cgT	214
3154240	G	A	SYNONYMOUS_CODING	Gene_YPO2827	upp	uracil phosphoribosyltransferase	I/I	atC/atT	76
3348338	C	T	SYNONYMOUS_CODING	Gene_YPO2998	YPO2998	two-component system response regulator	G/G	ggG/ggA	185
3376130	T	C	SYNONYMOUS_CODING	Gene_YPO3016	nanT	sialic acid transporter	K/K	aaA/aaG	25
3607609	A	G	SYNONYMOUS_CODING	Gene_YPO3241	yafK	hypothetical protein	K/K	aaA/aaG	205
3624499	A	T	SYNONYMOUS_CODING	Gene_YPO3249	YPO3249	allantoate amidohydrolase	T/T	acT/acA	106
4201619	T	C	SYNONYMOUS_CODING	Gene_YPO3747	rpoB	DNA-directed RNA polymerase subunit beta	E/E	gaA/gaG	1341
1578826	C	T	UPSTREAM: 21 bases	Gene_YPO1396	lpxK	tetraacyldisaccharide 4'-kinase			
3634070	A	T	UPSTREAM: 24 bases	Gene_YPO3262	YPO3262	hypothetical protein			
1306347	C	T	UPSTREAM: 26 bases	Gene_YPO1158	YPO1158	hypothetical protein			
1607551	C	T	UPSTREAM: 27 bases	Gene_YPO1415	pyrD	dihydroorotate dehydrogenase 2			
990954	G	A	UPSTREAM: 44 bases	Gene_YPO0904	YPO0904	hypothetical protein			
4256238	A	G	UPSTREAM: 47 bases	Gene_YPO3790	yigM	hypothetical protein			
1216366	C	T	UPSTREAM: 83 bases	Gene_YPO1073	metN	DL-methionine transporter ATP-binding protein			
4343118	G	A	UPSTREAM: 96 bases	Gene_YPO3867	rho	transcription termination factor Rho			
867596	C	T	UPSTREAM: 97 bases	Gene_YPO0790a	YPO0790a	hypothetical protein			
1333620	C	T	DOWNSTREAM: 10 bases	Gene_YPO1184	YPO1184	ABC transport membrane permease			
511653	G	T	DOWNSTREAM: 19 bases	Gene_YPO0480	dapB	4-hydroxy-tetrahydrodipicolinate reductase			
1037476	C	T	DOWNSTREAM: 3 bases	Gene_YPO0947	YPO0947	virulence determinant			
419209	G	A	DOWNSTREAM: 4 bases	Gene_YPO0402	YPO0402	PTS system fructose-family transporter subunit IIB			
3634070	A	T	DOWNSTREAM: 5 bases	Gene_YPO3261	YPO3261	amidase			
2413154	C	T	DOWNSTREAM: 56 bases	Gene_YPO2144	fadR	fatty acid metabolism regulator			
4123916	G	T	DOWNSTREAM: 66 bases	Gene_YPO3687	YPO3687	ribonuclease			
2828838	C	T	DOWNSTREAM: 90 bases	Gene_YPO2519	YPO2519	SAM-dependent methyltransferase			
3607609	A	G	DOWNSTREAM: 91 bases	Gene_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						
1344910	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 4: Gene annotations and effects of SNPs falling 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)
1025278	T	G	NON_SYNONYMOUS_CODING	Gene_YPO0932	YPO0932	hypothetical protein	S/A	Tct/Gct	137
1098675	A	C	NON_SYNONYMOUS_CODING	Gene_YPO0989	iucA	pseudogene	S/R	Agc/Cgc	380
2508389	T	C	NON_SYNONYMOUS_CODING	Gene_YPO2234	cstA	carbon starvation protein A	T/A	Aca/Gca	623
2744933	A	G	NON_SYNONYMOUS_CODING	Gene_YPO2446	YPO2446	YniC; 2-deoxyglucose-6-phosphatase	I/V	Att/Gtt	97
2903882	T	G	NON_SYNONYMOUS_CODING	Gene_YPO2582	YPO2582	sugar transport ATP-binding protein	C/G	Tgt/Ggt	413
2936268	G	A	NON_SYNONYMOUS_CODING	Gene_YPO2614	gltJ	glutamate/aspartate transport system permease	L/F	Ctt/Ttt	82
3085079	A	G	NON_SYNONYMOUS_CODING	Gene_YPO2752	mepA	D-alanyl-D-alanine endopeptidase	T/A	Act/Gct	85
3362591	A	G	NON_SYNONYMOUS_CODING	Gene_YPO3009	YPO3009	two-component response regulator	S/G	Agc/Ggc	59
3421335	A	G	NON_SYNONYMOUS_CODING	Gene_YPO3064	bcp	bacterioferritin comigratory protein	M/V	Atg/Gtg	136
3564026	C	T	NON_SYNONYMOUS_CODING	Gene_YPO3201	proY	permease	C/Y	tGt/tAt	84
3616733	A	G	NON_SYNONYMOUS_CODING	Gene_YPO3247	hmwA	adhesin	L/P	cTa/cCa	1291
4194600	G	A	NON_SYNONYMOUS_CODING	Gene_YPO3742	thiG	thiazole synthase	V/I	Gtt/Att	98
286528	T	A	SYNONYMOUS_CODING	Gene_YPO0285	YPO0285	hypothetical protein	R/R	cgA/cgT	46
699647	T	C	SYNONYMOUS_CODING	Gene_YPO0643	rpoD	RNA polymerase sigma factor RpoD	Q/Q	caA/caG	468
1735263	A	C	SYNONYMOUS_CODING	Gene_YPO1526	YPO1526	assembly protein (YegA)	A/A	gcA/gcC	307
1749443	T	C	SYNONYMOUS_CODING	Gene_YPO1537	YPO1537	iron-siderophore receptor	P/P	ccA/ccG	536
2575152	G	A	SYNONYMOUS_CODING	Gene_YPO2291	YPO2291	virulence factor	L/L	ttG/ttA	282
2739149	C	A	SYNONYMOUS_CODING	Gene_YPO2439	yfeA	substrate-binding protein	T/T	acC/acA	137
4082562	T	C	SYNONYMOUS_CODING	Gene_YPO3662	YPO3662	sulfite oxidase subunit YedY	Q/Q	caA/caG	294
3324959	A	G	DOWNSTREAM: 24 bases	Gene_YPO2977	glk	glucokinase			
2277583	G	A	INTERGENIC						
2684793	A	G	INTERGENIC						