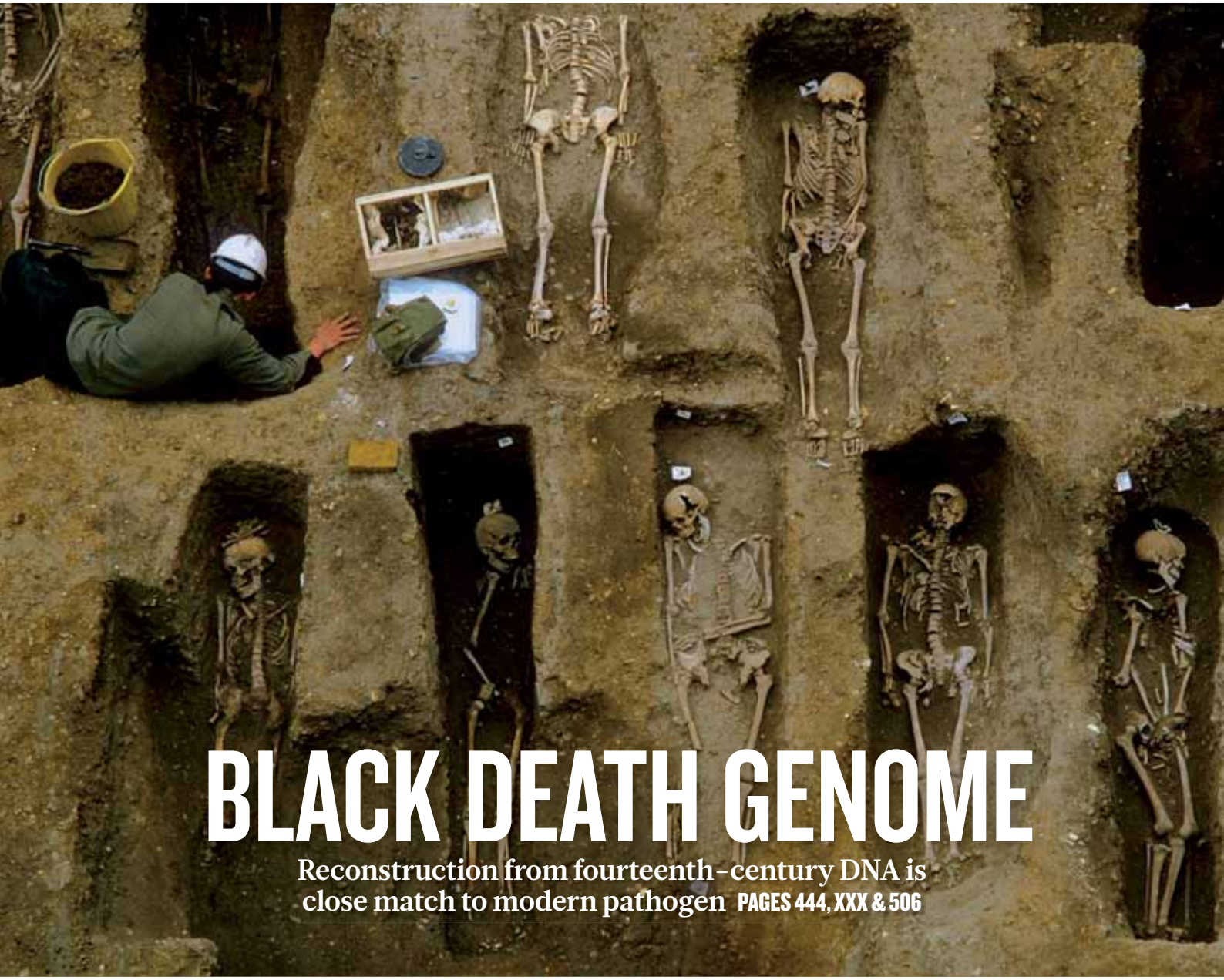
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



BLACK DEATH GENOME

Reconstruction from fourteenth-century DNA is close match to modern pathogen **PAGES 444, XXX & 506**

DWARF PLANETS

MEET PLUTO'S COOLER TWIN

Stellar occultation is chance to observe Eris

PAGES 462 & 493

CLIMATE CHANGE

FOOD SUPPLIES AT RISK

Predicted droughts threaten dustbowl conditions

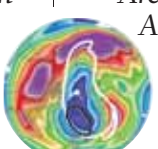
PAGE 450

ATMOSPHERIC SCIENCE

TROUBLE UP NORTH

Arctic ozone hole echoes Antarctic forerunner

PAGES XXX & 469

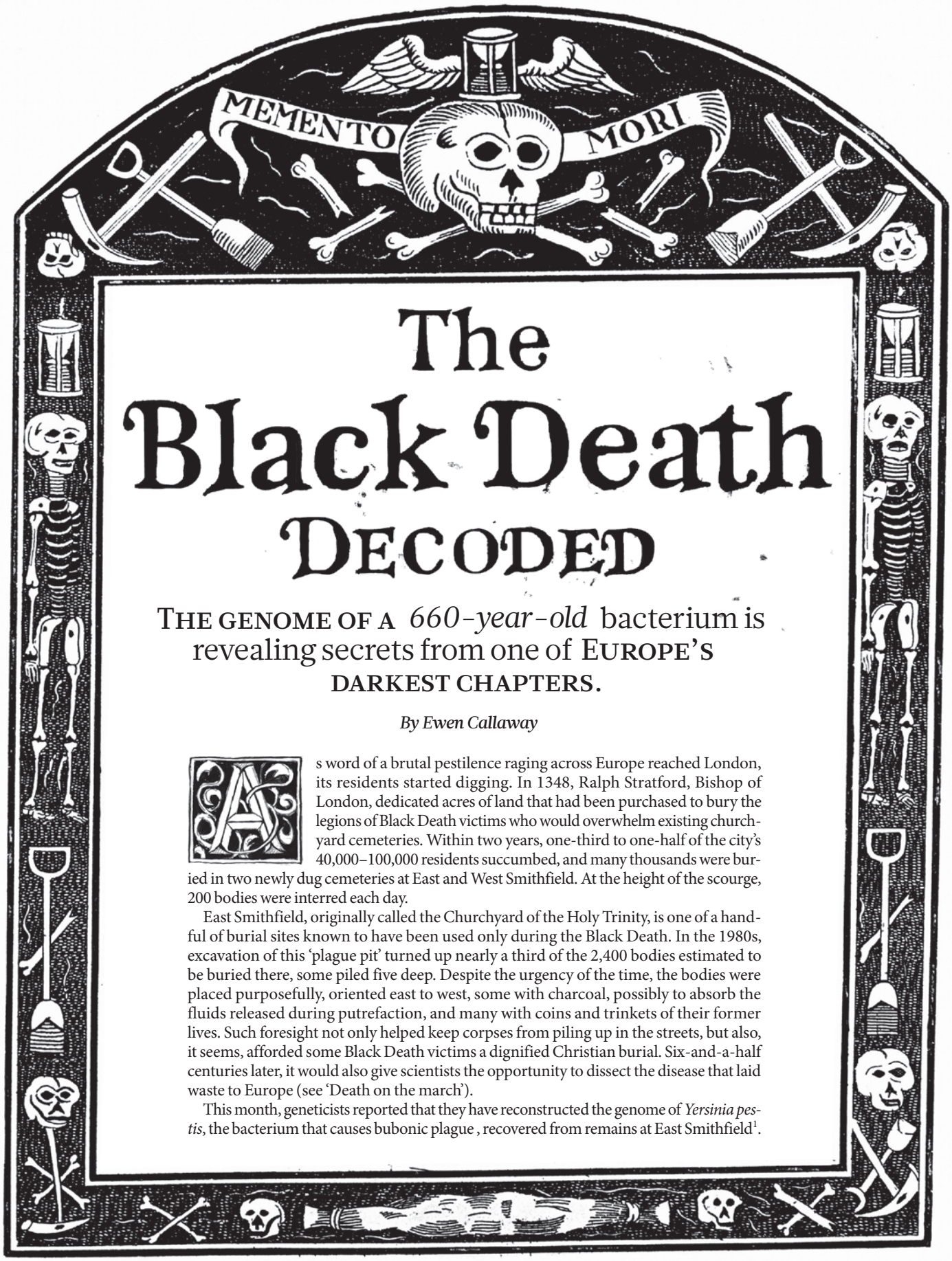


NATURE.COM/NATURE

27 October 2011 £10

Vol. 478, No. 7370

9 770028 083095 43



The Black Death Decoded

THE GENOME OF A *660-year-old* bacterium is revealing secrets from one of EUROPE'S DARKEST CHAPTERS.

By Ewen Callaway

As word of a brutal pestilence raging across Europe reached London, its residents started digging. In 1348, Ralph Stratford, Bishop of London, dedicated acres of land that had been purchased to bury the legions of Black Death victims who would overwhelm existing churchyard cemeteries. Within two years, one-third to one-half of the city's 40,000–100,000 residents succumbed, and many thousands were buried in two newly dug cemeteries at East and West Smithfield. At the height of the scourge, 200 bodies were interred each day.

East Smithfield, originally called the Churchyard of the Holy Trinity, is one of a handful of burial sites known to have been used only during the Black Death. In the 1980s, excavation of this 'plague pit' turned up nearly a third of the 2,400 bodies estimated to be buried there, some piled five deep. Despite the urgency of the time, the bodies were placed purposefully, oriented east to west, some with charcoal, possibly to absorb the fluids released during putrefaction, and many with coins and trinkets of their former lives. Such foresight not only helped keep corpses from piling up in the streets, but also, it seems, afforded some Black Death victims a dignified Christian burial. Six-and-a-half centuries later, it would also give scientists the opportunity to dissect the disease that laid waste to Europe (see 'Death on the march').

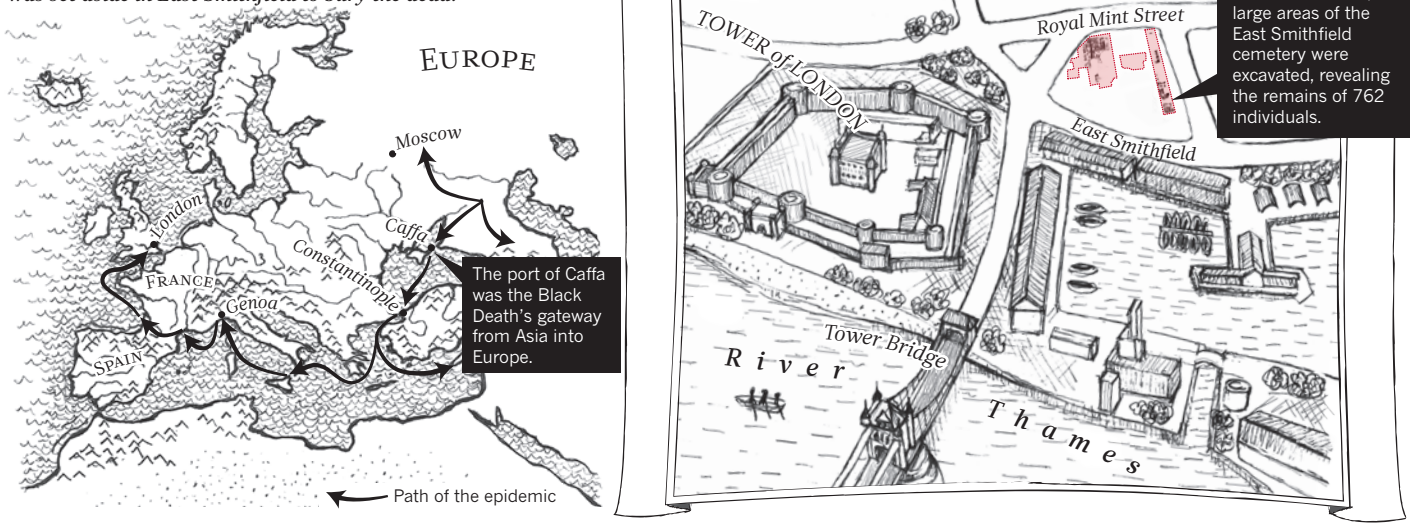
This month, geneticists reported that they have reconstructed the genome of *Yersinia pestis*, the bacterium that causes bubonic plague , recovered from remains at East Smithfield[1].

IAM/AKG-IMAGES

SOURCE: REF. 1

Death on the march

In the 1340s, a pestilence originating in Western Asia spread rapidly across Europe. Before it overtook London in 1348, land was set aside in East Smithfield to bury the dead.



The sequence — the first from an ancient bacterial pathogen — may help to explain how a disease could wreak so much havoc. It also marks a renaissance in genetic studies of ancient diseases, a field that has suffered a controversial history but that is now being revitalized. "There will be a race now for all the ancient pathogens," says Hendrik Poinar, a palaeogeneticist at McMaster University in Hamilton, Canada, who co-led the sequencing efforts.

Plagued with disbelief

When Alexandre Yersin linked *Y. pestis* to bubonic plague in 1894, many scientists surmised that the pathogen was behind not only the Black Death, but also a spate of earlier mass die-offs. The sixth-century Justinian plague devastated Constantinople and killed millions in Europe and the Near East. Plagues reared their heads periodically for the next two centuries. Black Death itself reappeared several times, even into the nineteenth century.

Clues tying *Y. pestis* to these outbreaks came largely from historical accounts of their symptoms, such as Giovanni Boccaccio's description of the Black Death in *The Decameron,* written around 1350: "It first betrayed itself by the emergence of certain tumours in the groin or the armpits, some of which grew as large as a common apple, others as an egg."

But some modern historians and scientists came to doubt that *Y. pestis* caused these ancient outbreaks. Bubonic plague epidemics known to have been caused by *Y. pestis* in the past century seemed too mild to have been caused by the same culprit as the Black Death: they killed fewer people and spread more slowly. Some 'plague revisionists' have argued that fleas, which spread *Y. pestis* to humans, would have struggled to survive the cold temperatures reported during the Black Death. And there was the speed with which it killed — Boccaccio reported that death often occurred within three days of the first symptoms appearing. Anthrax or a haemorrhagic-fever-causing virus similar to Ebola would be more likely than plague to cause such a rapid demise, say critics.

DNA evidence would seem to offer a definitive answer. In 2000, a team led by Didier Raoult, a microbiologist at the University of the Mediterranean in Marseilles, France, said it had proved the link between the bacterium and the disease. The researchers reported[2] that they had successfully recovered *Y. pestis* DNA from the teeth of a child and two adults dug up from a fourteenth-century mass burial site in Montpellier. The team identified the bacterium using a sensitive technique called the polymerase chain reaction (PCR) to amplify a portion of a gene from *Y. pestis* called *pla*. "We believe that we can end the controversy," the team wrote[2]. "Medieval Black Death was plague."

But several critics raised concerns about contamination. The PCR might instead have amplified DNA from modern *Y. pestis* used previously in the lab, or possibly the sequences from a closely related soil-dwelling bacterium. "I could never, ever replicate it," says Thomas Gilbert, an evolutionary geneticist at the University of Copenhagen in Denmark. In 2004, Gilbert and his colleagues reported no trace of *Y. pestis* DNA in 108 teeth from 61 individuals found in plague pits in France, Denmark and England (including East Smithfield)[3].

Raoult says that there was no contamination and that Gilbert's methods did not accurately replicate his[4]. Still, those who were already sceptical of the suggestion that *Y. pestis* caused the Black Death latched on to Gilbert's study.

Other studies of microbial DNA extracted from ancient human remains — including those affected by tuberculosis, syphilis and malaria — were also being scrutinized. In several cases, researchers could not replicate results, or they found methodological shortcomings. Critics said that DNA from these samples was too degraded by heat, moisture and time to detect, and the field soon divided into believers and sceptics.

"There was a complete schism," says Ian Barnes, a palaeogeneticist at Royal Holloway University of London, who says he spent two-and-a-half years trying — unsuccessfully — to find DNA evidence of syphilis or tuberculosis in bones dating from the nineteenth and early twentieth centuries[5]. "People largely ignored each other," he says.

Digging up answers

Although Poinar was dubious of claims about ancient microbial DNA, he was intrigued by the bones from East Smithfield. Nearly all of the remains are from Black Death victims, many of whom were cut down during the prime of their lives.

In a bright ground-floor laboratory of the Museum of London, a short walk from East Smithfield, osteoarchaeologist Jelena Bekvalac examines the nearly complete skeleton of one of the plague pit's former residents. Wearing a black silk scarf dotted with white skull-prints, Bekvalac handles a pelvic bone and determines that it belonged to a man who died in his late teens or early twenties. Apart from some plaque on his teeth and a gash in his skull that shows some signs of healing, the man's skeleton offers no outward evidence of Black Death.

His remains, and those from hundreds of others, represent a snapshot of life and death in London during the epidemic. Since the site's excavation, researchers have descended on the bones in search of information.

In the late 1990s, Poinar met Sharon DeWitte, then a graduate student at Pennsylvania State University in State College, who was working on a demographic analysis of the remains suggesting that Black Death preferentially killed those who were already frail. The two considered drilling into teeth

and bones to find *Y. pestis* DNA, but Poinar wasn't satisfied with the available detection tools, which were still based on PCR. "We sort of sat on the samples for a few years waiting for all the stars to align," says DeWitte, now at the State University of New York at Albany.

That alignment came from next-generation DNA sequencers, machines that read short snippets of DNA. The technology was perfect for sequencing DNA that has been damaged by spending hundreds of years underground.

The sequencers allowed Svante Pääbo, a palaeogeneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and his team to sequence a draft of the Neanderthal genome[6]. But finding and sequencing ancient pathogens in a human skeleton is much harder — like finding "needles in the football field", Poinar says — because their genomes are 1,000 times shorter than that of the Neanderthal and closely resemble those of soil microbes that have infiltrated the bones.

Another technology helped narrow the search. Pääbo and his team developed a technique, called targeted capture, in which they used lab-synthesized 'bait' DNA to snag ancient DNA strands from a bone sample[7], leaving soil-microbe and other sequences behind. "It's pretty much like fishing in a pond," says Johannes Krause, a palaeogeneticist at the University of Tübingen in Germany, who worked with Pääbo on the Neanderthal genome and co-led the Black Death project with Poinar.

In a proof-of-principle experiment published in August of this year, Krause and Poinar's team used sequences from a contemporary plague strain to fish out *Y. pestis* DNA from the teeth of victims buried at East Smithfield. From this, they sequenced a short loop of DNA, called the pPCP1 plasmid, that is partially responsible for bubonic plague's ability to infect humans.

Their results[8], along with a paper published last year[9] that found *Y. pestis* sequences in different Black Death bone samples, have convinced most scientists that bubonic plague was involved in the Black Death.

In their most recent paper[1], Poinar and Krause completed the ancient genome and showed that it sits at the root of an evolutionary tree that comprises 17 contemporary strains of *Y. pestis*. This indicates that the Black Death strain spawned many of the forms of *Y. pestis* that infect humans today.

This strain, Krause adds, probably emerged not long before the Black Death started its rampage across western Asia and Europe in the fourteenth century. "That, for me, was the biggest surprise," he says. It suggests, the authors argue, that earlier plagues were caused by either a now-extinct strain of *Y. pestis* or by an entirely different pathogen.

Mark Achtman, a plague-evolution expert at University College Cork in Ireland, calls this interpretation "absolute nonsense". Krause and Poinar's team did not consider a number of modern plague strains found in central and east Asia, which are thought to have earlier origins than the East Smithfield strain, Achtman says. Genome sequences for these strains were not available to his team, says Krause, but he is eager to see how they are related.

➲ NATURE.COM
For more on Black Death research, see *Nature*'s video:
go.nature.com/hxbtel

Mysterious scourge

Just as puzzling, however, is that *Y. pestis* seems to have changed very little over the past 660 years. The genome of the Black Death strain differs from that of the modern *Y. pestis* 'reference' strain by about 100 nucleotides, but each of these genetic differences can be found in at least one contemporary strain. "We can't find anything that makes the Black Death special," Krause says.

Historical descriptions of the Black Death have helped link *Yersinia pestis* with the disease.

The team is now looking for other genetic changes that could account for the Black Death's ferocity, such as rearrangements in the genome, which are difficult to determine from the short fragments of DNA available. To better understand how the plague worked, researchers could try to resurrect the Black Death pathogen by modifying the genomes of contemporary *Y. pestis* strains. Although this might sound alarming, research on *Y. pestis* is already carefully controlled, and even an accidental infection with such a strain could be easily treated with modern antibiotics.

Moreover, Poinar says, the Black Death was not just about the bacterium. Environmental and epidemiological factors must have aided in its vicious tear through Europe. Sick soldiers returning to Europe from Caffa, the Black Sea port that was the plague's gateway from Asia, unleashed the disease on a population that would have been weakened by malnourishment and years of cold, wet weather, he says.

Achtman says that it is possible that Black Death was not spread by rat-dwelling fleas, as *Y. pestis* is today, but by other animals, which could have enhanced transmission. Or another circulating pathogen could have contributed, as in the 'Spanish flu' pandemic that killed up to 100 million people worldwide in 1918-19, often with the help of bacterial pneumonia.

Whatever questions remain about the Black Death, scientists are now keen to apply the latest sequencing methods to other ancient epidemics. "I've completely gone from thinking, 'ancient pathogens are a load of crap,' to 'hold on, maybe some of this stuff works'," says Gilbert, whose team has started to sequence DNA from pathogens that plagued ancient crops. Researchers could identify ancient microbes and chart their spread and their evolutionary relationships with contemporary strains. For example, Europeans who travelled to the New World may have introduced new forms of tuberculosis to North America and brought syphilis back to Europe.

Ancient pathogens may help scientists understand current and future outbreaks, says Terry Brown, a biomolecular archaeologist at the University of Manchester, UK. He and Charlotte Roberts, of Durham University, UK, are charting the evolution of tuberculosis strains in Britain and Europe. "By looking over the past 1,000 years of disease in British cities, we can understand problems occurring in the Third World, where more and more people are crowding into cities," he says. Similarly, the sequencing and resurrection of the influenza strain responsible for the 1918 pandemic[10] has helped researchers to interpret the sequences of contemporary flu strains.

For all its ferocity, the Black Death left few visible marks on London. Today, the plague pit at East Smithfield is in the heart of London's financial district, buried under modern office suites and the old Royal Mint building. The only visible remnants are the crumbled ruins of St Mary Graces, a Cistercian abbey built near the site in 1350.

London may have seen its last significant bubonic plague outbreak, but catastrophic epidemics are a rule of human history, not an exception. Centuries from now, what traces will the next great scourge leave? Future archaeologists chronicling its history may find memorials, graves and probably even the bodies of victims. But another story will also lurk in its DNA, just waiting to be read. ■

Ewen Callaway *writes for* Nature *from London.*

1. Bos, K. I. *et al. Nature* **478,** 506–510 (2011).
2. Raoult, D. *et al. Proc. Natl Acad. Sci. USA* **97,** 12800–12803 (2000).
3. Gilbert, M. T. P. *et al. Microbiology* **150,** 341–354 (2004).
4. Drancourt, M. & Raoult, D. *Microbiology* **150,** 263–264 (2004).
5. Barnes, I. & Thomas, M. G. *Proc. R. Soc. B* **273,** 645–653 (2006).
6. Green, R. E. *et al. Science* **328,** 710–722 (2010).
7. Briggs, A. W. *et al. Science* **325,** 318–321 (2009).
8. Schuenemann, V. J. *et al. Proc. Natl Acad. Sci. USA* **108,** E746–E752 (2011).
9. Haensch, S. *et al. PLoS Pathog.* **6,** e1001134 (2010).
10. Taubenberger, J. K. *et al. Nature* **437,** 889-893 (2005).

has provided insight into the formation, evolution and dynamical histories of these bodies. We now expect that there are hundreds of objects that will eventually be classified as dwarf planets. Because Eris is the only body similar to Pluto for which detailed stellar-occultation data are available, Sicardy and colleagues' results[5] represent a major step forward in our knowledge about large Kuiper-belt objects.

Future attempts will certainly be made to obtain additional stellar-occultation data for Eris as well as for other Kuiper-belt objects. Whether they are called planets or not, there is clearly still much to learn about these distant, icy bodies. ■

Amanda Gulbis *is at the Southern African Large Telescope and South African Astronomical Observatory, 7935 Cape Town, South Africa.*
e-mail: amanda@salt.ac.za

1. Jewitt, D. & Luu, J. *Nature* **362,** 730–732 (1993).
2. Brown, M. E., Trujillo, C. A. & Rabinowitz, D. L. *Astrophys. J.* **635,** L97–L100 (2005).
3. Brown, M. E., Schaller, E. L., Roe, H. G., Rabinowitz, D. L. & Trujillo, C. A. *Astrophys. J.* **643,** L61–L63 (2006).
4. Bertoldi, F., Altenhoff, W., Weiss, A., Menten, K. M. & Thum, C. *Nature* **439,** 563–564 (2006).
5. Sicardy, B. *et al. Nature* **478,** 493–496 (2011).
6. Elliot, J. L. *et al. Nature* **465,** 897–900 (2010).
7. Person, M. J. *et al. Astron. J.* **136,** 1510–1518 (2008).
8. McCarthy, D. W. *et al. Astron. J.* **136,** 1519–1522 (2008).

GENOMICS

Plague's progress

The Black Death was one of the most devastating pandemics in human history. The first complete genome sequence of the causative *Yersinia pestis* bacterium provides a fresh perspective on plague evolution. SEE LETTER P.506

EDWARD C. HOLMES

The Black Death was a pandemic of almost unprecedented scale. It is estimated that 30–50% of Europe's population perished from the plague between 1347 and 1351. As might be expected from such a remarkable level of mortality, this pestilence had a profound impact on medieval society. For example, it greatly influenced the depiction of death in art (Fig. 1), and acted as the driving force for the establishment of some of the first public-health measures, although these were generally futile. It has long been supposed that the Black Death, like all plague epidemics since, was caused by the bacterium *Yersinia pestis* in a transmission cycle involving fleas, rats and humans. But the evolutionary relationships between *Y. pestis* strains from different plague epidemics have been less clear. On page 506 of this issue, Bos *et al.*[1] describe the first complete genome sequence of *Y. pestis* from Black Death victims, and show that this pandemic was a pivotal event in plague evolution.

Obtaining bacterial DNA that is almost 700 years old presents a number of challenges, prominent among them the risk of inadvertent contamination by DNA from other sources[1,2]. Bos *et al.* extracted bacterial DNA from five teeth of plague victims taken from a burial pit at East Smithfield (ES) in London, which was established at the height of the pandemic in 1348–49. The key methodological advance in their work was the use of a molecular capture assay that assisted in the detection of *Y. pestis* DNA amid a background of host and environmental DNA. The stringent laboratory procedures used by the authors, the observed patterns of mutational damage in the DNA, and the finding that the ES strain is ancestral to all contemporary *Y. pestis* strains on phylogenetic trees, strongly suggest that the obtained genome sequence is an authentic representative of the Black Death pathogen.

Figure 1 | The last rites. In this fourteenth-century picture, a victim of the Black Death is attacked by a devil, while a priest reads the last rites and God watches from above.

THE BRITISH LIBRARY/HIP/TOPFOTO

The ES strain is significant for several reasons. First, in combination with studies of partial genome sequences of *Y. pestis* from Black Death victims[3], it demonstrates beyond doubt that *Y. pestis* was the true cause of the Black Death, despite claims to the contrary[4]. Second, the fact that the ES strain falls at the base of a phylogenetic tree that links contemporary *Y. pestis* genomes suggests that the Black Death was a crucial event in plague evolution that generated all later lineages of *Y. pestis*, including those responsible for the 'Asian' (or modern) plague pandemic that has spread globally since the nineteenth century. Interestingly, the greatest genetic diversity of *Y. pestis* is observed[5] in China, suggesting that most plague epidemics originated in this region. Third, Bos *et al.*[1] suggest that the genetic similarity between the ES strain and contemporary strains of *Y. pestis* that are associated with less severe epidemics indicates that the high mortality of the Black Death was not simply a function of the bacterial strain involved. Given the small sample size of the authors' study, this assertion must be treated with caution, but the availability of a complete genome of *Y. pestis* from the Black Death should make the hypothesis experimentally testable.

Although the recovery and sequencing of the ES strain confirms the role of *Y. pestis* in the Black Death, it also raises questions about the cause of two earlier major disease pandemics previously assigned to *Y. pestis*: one that spread through parts of Africa, Asia and Europe in AD 541–542, during the reign of the Roman emperor Justinian and, more tentatively, the plague of Athens (430–426 BC), which was evocatively described in the writings of the Greek historian Thucydides. Bacterial DNA has purportedly been recovered from both of these epidemics, although in the case of the plague of Athens, the DNA was attributed[6] to the *Salmonella enterica* serovar Typhi bacterium that causes typhoid fever, rather than to *Y. pestis*. However, the Athenian *Salmonella* strain is not actually closely related to that responsible for typhoid[7], suggesting that the ancient DNA had been contaminated by DNA from a modern, soil-living *Salmonella* species. The cause of the plague of Athens therefore remains a mystery.

The DNA supposedly from the Justinian plague is certainly that of *Y. pestis*, but the close similarity between the 'Justinian' DNA and that of modern *Y. pestis* variants[8] suggests that the former strain may not be authentic. Importantly, if all contemporary strains of *Y. pestis* are derived from the Black Death, as suggested by Bos and colleagues[1], then both of the earlier epidemics were caused either by a strain of *Y. pestis* that has left no contemporary descendants, or by an entirely different organism. Ancient DNA may be central to resolving this question.

The analysis of ancient DNA is a powerful

tool, providing a unique perspective on the timescale and rates of microbial evolution, and of the host adaptation of microbial populations. Ancient DNA also provides a means to determine the cause of ancient epidemics, particularly when historical descriptions of disease symptoms yield ambiguous diagnoses. However, the spectre of contamination has loomed large, so that the impact of ancient-DNA analysis on our understanding of the evolution of infectious diseases has been relatively minor. Up until now, the most important study of historical pathogens has been the sequencing[9] of the virus responsible for the devastating influenza pandemic of 1918–19, which has provided insight into viral pathogenesis[10]. The ES strain of *Y. pestis* is one of the oldest pathogens sequenced so far, and the results are some of the most credible in the field's history. Bos and colleagues' study[1] certainly sheds light on one of the most significant events in human history, but its greatest significance may be that it heralds a new era of research into the genomics of ancient bacterial diseases. ■

Edward C. Holmes *is at the Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA, and the Fogarty International Center, Bethesda, Maryland.*
e-mail: echolmes@psu.edu

1. Bos, K. I. *et al. Nature* **478,** 506–510 (2011).
2. Gilbert, M. T. P. *et al. Microbiology* **150,** 341–354 (2004).
3. Haensch, S. *et al. PLoS Pathog.* **6,** e1001134 (2010).
4. Scott, S. & Duncan, C. J. *Biology of Plagues* (Cambridge Univ. Press, 2001).
5. Morelli, G. *et al. Nature Genet.* **42,** 1140–1143 (2010).
6. Papagrigorakis, M. J., Yapijakis, C., Synodinos, P. N. & Baziotopolou-Valavani, E. *Int. J. Infect. Dis.* **10,** 206–214 (2006).
7. Shapiro, B., Rambaut, A. & Gilbert, M. T. P. *Int. J. Infect. Dis.* **10,** 334–335 (2006).
8. Drancourt, M. *et al. Emerg. Infect. Dis.* **10,** 1585–1592 (2004).
9. Taubenberger, J. K. *et al. Nature* **437,** 889–893 (2005).
10. Tumpey, T. M. *et al. Science* **310,** 77–80 (2005).

ASTROPHYSICS

Shedding light on the fabric of space–time

The idea that space–time might be fundamentally fuzzy is much debated among theorists. A search for signatures of this effect on light from distant cosmic sources has come up empty-handed, but shows the potential of this approach.

GIOVANNI AMELINO-CAMELIA

One of the most fascinating journeys in science concerns the evolution of the description of space and time. Examples of milestones on this journey are the realization that time is relative and that space-time bends in response to the presence of heavy particles and bodies[1]. Only one feature of the initial naive conceptualization of space and time by Isaac Newton has survived: in current theories, space-time is still viewed as a smooth entity. However, even this last Newtonian pillar is now being scrutinized. The main prediction of the theory of quantum mechanics, which underlies many of the advances of fundamental physics in the twentieth century, is that the results of a large class of measurements are affected by irreducible uncertainties. And all attempts to apply this successful theory to the description of space-time suggest that, as a result of some of these uncertainties, space-time should be fundamentally fuzzy. Writing in *Astronomy & Astrophysics*, Tamburini *et al.*[2] describe how they have searched for an imprint of space-time fuzziness on light from sources located at large distances from Earth.

In a fuzzy space-time, particles would travel much like a car on a bumpy road, with the notion of a smooth trajectory emerging only if observations cannot resolve the effects of the bumps (Fig. 1). Unfortunately, the magnitude of the 'bumps' of space-time — the scale of space-time fuzziness — is horrifyingly small. This scale is expected to be of the order of the minuscule Planck length, L_p, which is about 10^{-35} metres and is defined as $L_p = (\hbar G/c^3)^{1/2}$, where G is Newton's constant of gravity, $\hbar$ is the reduced Planck constant of quantum mechanics and c is the speed of light.

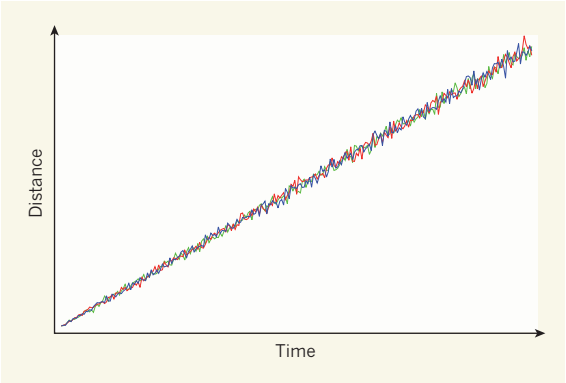


Figure 1 | Bumpy journeys. The distance travelled by free particles in space increases with time, but the exact form of the increase depends on the nature of the fabric of space-time. In a smooth space-time, the increase is exactly linear, whereas in a 'bumpy' space-time the dependence is only roughly so. Shown here, qualitatively, are three particle journeys (blue, red and green) in a bumpy space-time. The effects of space-time bumps are invisibly small for short distances, but become visible for larger distances. Tamburini *et al.*[2] looked for evidence of a bumpy space-time on light from distant cosmic sources. (Figure courtesy of N. Loret.)

Theoretical studies of space-time fuzziness have been an active area of research since at least the 1960s, when John Wheeler pointed out[3] the significance of this feature and coined the catchy term 'space-time foam' to describe it. But for a long time there was no experimental counterpart to these studies, because it seemed that 'bumps' with a magnitude of 10^{-35} metres should forever be beyond the reach of the sensitivity of any experimental apparatus. Moreover, devising detailed candidate models of space-time fuzziness is hard, because essentially it involves combining the complexity of quantum mechanics with the intricacies of general relativity[4].

As a result, phenomenological approaches to describing space-time foam started to gain momentum only quite recently. A first wave of studies[5,6] attempted to 'parameterize our ignorance' about space-time foam and to compensate for the smallness of the Planck length by studying the effects of fuzziness for particles travelling across distances of a few kilometres on Earth. But then, with a second wave of investigations[7,8], it became clear that this strategy could be generalized to the case of observations of light from certain classes of distant astrophysical sources. And it is this astrophysics version of the approach to testing space-time fuzziness that Tamburini *et al.*[2] adopt in their study.

The authors' analysis exploits the fact that, as a result of space-time fuzziness, the wavefront of a light wave from a distant source is expected to develop tiny corrugations as it travels, such that some portions of the wavefront are advanced while others are retarded. For propagation over large cosmological distances from Earth, this wavefront corrugation can add up to give a macroscopic effect. Tamburini *et al.* searched for this effect in images of around 160 distant quasars (extremely luminous galactic nuclei powered by supermassive black holes), all at redshift greater than 4. Their analysis benefits from both the large number

LETTER

doi:10.1038/nature10549

A draft genome of *Yersinia pestis* from victims of the Black Death

Kirsten I. Bos[1]*, Verena J. Schuenemann[2]*, G. Brian Golding[3], Hernán A. Burbano[4], Nicholas Waglechner[5], Brian K. Coombes[5], Joseph B. McPhee[5], Sharon N. DeWitte[6,7], Matthias Meyer[4], Sarah Schmedes[8], James Wood[9], David J. D. Earn[5,10], D. Ann Herring[11], Peter Bauer[12], Hendrik N. Poinar[1,3,5] & Johannes Krause[2,12]

Technological advances in DNA recovery and sequencing have drastically expanded the scope of genetic analyses of ancient specimens to the extent that full genomic investigations are now feasible and are quickly becoming standard[1]. This trend has important implications for infectious disease research because genomic data from ancient microbes may help to elucidate mechanisms of pathogen evolution and adaptation for emerging and re-emerging infections. Here we report a reconstructed ancient genome of *Yersinia pestis* at 30-fold average coverage from Black Death victims securely dated to episodes of pestilence-associated mortality in London, England, 1348–1350. Genetic architecture and phylogenetic analysis indicate that the ancient organism is ancestral to most extant strains and sits very close to the ancestral node of all *Y. pestis* commonly associated with human infection. Temporal estimates suggest that the Black Death of 1347–1351 was the main historical event responsible for the introduction and widespread dissemination of the ancestor to all currently circulating *Y. pestis* strains pathogenic to humans, and further indicates that contemporary *Y. pestis* epidemics have their origins in the medieval era. Comparisons against modern genomes reveal no unique derived positions in the medieval organism, indicating that the perceived increased virulence of the disease during the Black Death may not have been due to bacterial phenotype. These findings support the notion that factors other than microbial genetics, such as environment, vector dynamics and host susceptibility, should be at the forefront of epidemiological discussions regarding emerging *Y. pestis* infections.

The Black Death of 1347–1351, caused by the bacterium *Yersinia pestis*[2,3], provides one of the best historical examples of an emerging infection with rapid dissemination and high mortality, claiming an estimated 30–50% of the European population in only a five-year period[4]. Discrepancies in epidemiological trends between the medieval disease and modern *Y. pestis* infections have ignited controversy over the pandemic's aetiologic agent[5,6]. Although ancient DNA investigations have strongly implicated *Y. pestis*[2,3] in the ancient pandemic, genetic changes in the bacterium may be partially responsible for differences in disease manifestation and severity. To understand the organism's evolution it is necessary to characterize the genetic changes involved in its transformation from a sylvatic pathogen to one capable of pandemic human infection on the scale of the Black Death, and to determine its relationship with currently circulating strains. Here we begin this discussion by presenting the first draft genome sequence of the ancient pathogen.

Y. pestis is a recently evolved descendent of the soil-dwelling bacillus *Yersinia pseudotuberculosis*[7], which in the course of its evolution acquired two additional plasmids (pMT1 and pPCP1) that provide it with specialized mechanisms for infiltrating mammalian hosts. To investigate potential evolutionary changes in one of these plasmids, we reported on the screening of 46 teeth and 53 bones from the East Smithfield collection of London, England for presence of the *Y. pestis*-specific pPCP1 (ref. 3). Historical data indicate that the East Smithfield burial ground was established in late 1348 or early 1349 specifically for interment of Black Death victims[8] (Supplementary Figs 1 and 2), making the collection well-suited for genetic investigations of ancient *Y. pestis*. DNA sequence data for five teeth obtained via molecular capture of the full *Y. pestis*-specific pPCP1 revealed a C to T damage pattern characteristic of authentic endogenous ancient DNA[9], and assembly of the pooled Illumina reads permitted the reconstruction of 98.68% of the 9.6-kilobase plasmid at a minimum of twofold coverage[3].

To evaluate the suitability of capture-based methods for reconstructing the complete ancient genome, multiple DNA extracts from both roots and crowns stemming from four of the five teeth which yielded the highest pPCP1 coverage[3] were used for array-based enrichment (Agilent) and subsequent high-throughput sequencing on the Illumina GAII platform[10]. Removal of duplicate molecules and subsequent filtering produced a total of 2,366,647 high quality chromosomal reads (Supplementary Table 1a, b) with an average fragment length of 55.53 base pairs (Supplementary Fig. 4), which is typical for ancient DNA. Coverage estimates yielded an average of 28.2 reads per site for the chromosome, and 35.2 and 31.2 for the pCD1 and pMT1 plasmids, respectively (Fig. 1a, c, d and Supplementary Table 1b, c). Coverage was predictably low for pPCP1 (Fig. 1e) because probes specific to this plasmid were not included on the arrays. Coverage correlated with GC content (Supplementary Fig. 6), a trend previously observed for high-throughput sequence data[11]. The coverage on each half of the chromosome was uneven due to differences in sequencing depth between the two arrays, with 36.46 and 22.41 average reads per site for array 1 and array 2, respectively. Although greater depth contributed to more average reads per site, it did not increase overall coverage, with both arrays covering 93.48% of the targeted regions at a minimum of onefold coverage (Supplementary Table 1b). This indicates that our capture procedure successfully retrieved template molecules from all genomic regions accessible via this method, and that deeper sequencing would not result in additional data for CO92 template regions not covered in our data set.

Genome architecture is known to vary widely among extant *Y. pestis* strains[12]. To extrapolate gene order in our ancient genome, we analysed reads mapping to the CO92 reference for all extracts stemming

[1]McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada. [2]Institute for Archaeological Sciences, Rümelinstr. 23, University of Tübingen, 72070 Tübingen, Germany. [3]Biology Department, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada. [4]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [5]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada. [6]Department of Anthropology, University of South Carolina, Columbia, South Carolina 29208, USA. [7]Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA. [8]Institute of Applied Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, Texas 76107, USA. [9]Department of Anthropology and Population Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [10]Department of Mathematics and Statistics, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. [11]Department of Anthropology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada. [12]Human Genetics Department, Medical Faculty, University of Tübingen, 72070 Tübingen, Germany.
*These authors contributed equally to this work.

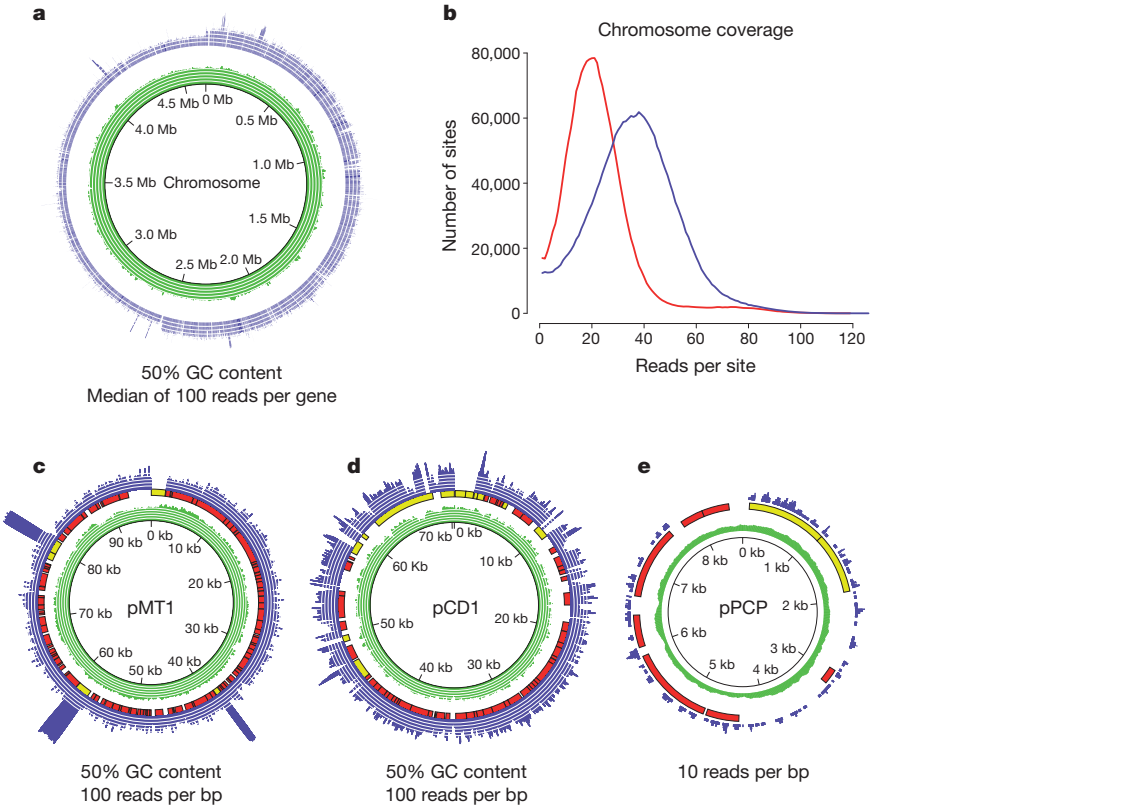


Figure 1 | Coverage plots for genomic regions sequenced. **a**, **c**–**e**, Coverage plots for the chromosome (**a**) and the plasmids pMT1 (**c**), pCD1 (**d**) and pPCP (**e**). Coverage in blue, GC content in green. Scale lines indicate 10-, 20-, 30-, 40- and 50-fold coverage and 10%, 20%, 30%, 40% and 50% GC content. For plasmids, red corresponds to coding regions, yellow to mobile elements. Chromosome shows median coverage per gene. Plasmids show each site plotted. Coverage distributions for the plasmids are shown in Supplementary Fig. 5. **b**, Distributions show chromosomal coverage of array 1 (blue) and array 2 (red), indicating that deeper sequencing increases the number of reads per site, but does not substantially influence overall coverage.

from a single individual who yielded the highest coverage (individual 8291). Despite the short read length of our ancient sequences and the highly repetitive nature of the *Y. pestis* genome, 2,221 contigs matching CO92 were extracted, comprising a total of 4,367,867 bp. To identify potential regions of the ancient genome that are architecturally distinct from CO92, all reads not mapping to the CO92 reference were in turn considered for contig construction. After filtering for a minimum length of 500 bp, 2,134 contigs remained comprising 4,013,009 bp, of which 30,959 stemmed from unmapped reads. Conventional BLAST search queried against the CO92 genome identified matches for 2,105 contigs. Evidence of altered architecture was identified in 10 contigs (Supplementary Table 2). An example of such a structural variant is shown in Fig. 2, where reference-guided assembly incorporating unmapped reads to span the breakpoint validates its reconstruction. This specific genetic orientation is found only in *Y. pseudotuberculosis* and *Y. pestis* strains Mictrotus 91001, Angola, Pestoides F and B42003004, which are ancestral to all *Y. pestis* commonly associated with human infections (branch 1 and branch 2 strains[13,14]). Furthermore, discrepancies in the arrangement of this region in branch 1 and branch 2 modern *Y. pestis* strains indicate that rearrangements occurred as separate events on different lineages.

Single-nucleotide differences between our ancient genome and the CO92 reference surprisingly consisted of only 97 chromosomal positions, and 2 and 4 positions in the pCD1 and pMT1 plasmids, respectively (Supplementary Table 3), indicating tight genetic conservation in this organism over the last 660 years. Twenty-seven of these positions were unreported in a previous analysis of extant *Y. pestis* diversity[14] (Supplementary Tables 3 and 4). Comparison of our ancient genome to its ancestor *Y. pseudotuberculosis* revealed that the medieval sequence contained the ancestral nucleotide for all 97 positions, indicating that it does not possess any derived positions absent in other *Y. pestis* strains. Two previously reported chromosomal differences[3] were not present in our genomic sequence data, suggesting that they probably

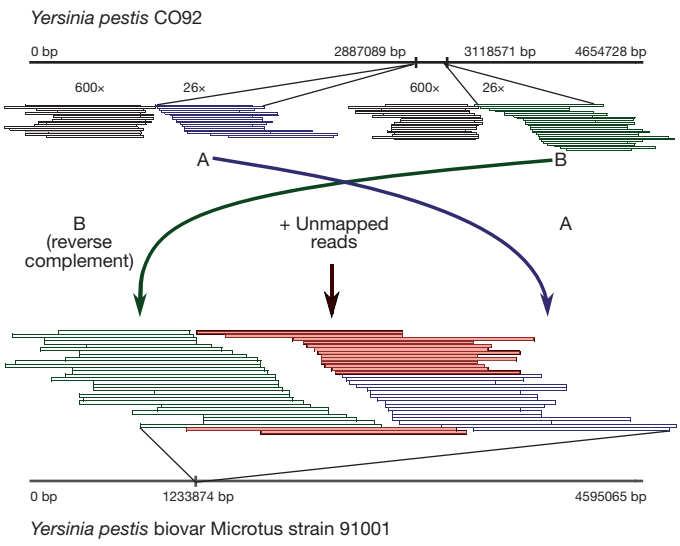


Figure 2 | Alignment of mapped reconstructed contigs against CO92 and Microtus genomes. Reads mapped at positions A (blue) and B (green) are 231 kb apart in the linearized CO92 genome. Adjacent sequence is high coverage although only 18× and 20× is shown due to space constraints (black) for A and B, respectively. The structural variant was assembled using reads that did not map to CO92 (red). Its position is shown on the linearized Microtus 91001 chromosome where the 9,096 bp contig maps with 100% identity.

derived from deaminated cytosines that would have been removed in the current investigation via uracil-DNA-glycosylase treatment before array capture.

To place our ancient genome in a phylogenetic context, we characterized all 1,694 previously identified phylogenetically informative positions[14] (Supplementary Table 4), and compared those from our ancient organism against aggregate base call data for 17 publicly available *Y. pestis* genomes and the ancestral *Y. pseudotuberculosis*. When considered separately, sequences from three of the four victims fall only two substitutions from the root of all extant human pathogenic *Y. pestis* strains (Fig. 3a), and they show a closer relationship to branch 1 *Y. pestis* than to branch 2; however, one of the four victims (individual 6330) was infected with a strain that contained three additional derived positions seen in all other branch 1 genomes[14]. This suggests either the presence of multiple strains in the London 1348–1350 pandemic or microevolutionary changes accruing in one strain, which is known to occur in disease outbreaks[15]. Additional support for *Y. pestis* microevolution is indicated by the presence of several variant positions for which sequence data from one individual shows two different nucleotides at comparable frequencies (Supplementary Table 5). Position 2896636, for example, is a known polymorphic position in extant *Y. pestis* populations[14], and this position shows the fixed derived state in one individual (6330) and the polymorphic state in another (individual 8291) at minimum fivefold coverage (Supplementary Fig. 7). This provides a remarkable example of microevolution captured during an historical pandemic. The remaining variance positions are unchanged in the 18 extant *Yersinia* genomes, thus they may be unique to the ancient organism and are, therefore, of further interest. Additional sampling

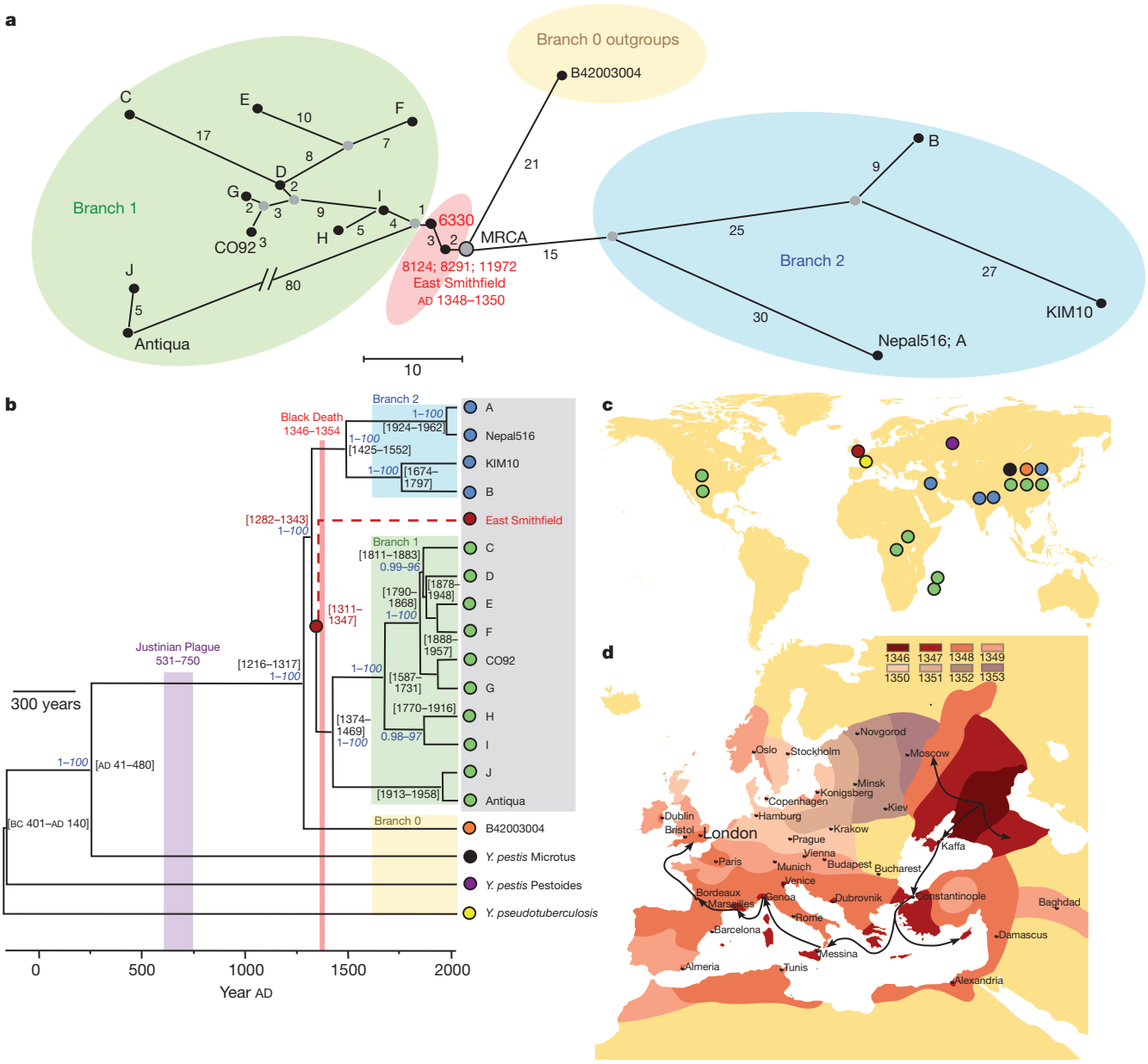


Figure 3 | Phylogenetic placement and historical context for the East Smithfield strain. a, Median network of ancient and modern *Y. pestis* based on 1,694 variant positions in modern genomes[14]. Coloured circles represent different clades as defined in ref. 13. Gray circles represent hypothetical nodes. **b**, Phylogenetic tree using 1,694 variable positions. Divergence time intervals are shown in calendar years, with neighbour-joining bootstrap support (blue italic) and Bayesian posterior probability (blue). Grey box indicates known human pathogenic strains. A, NZ ACNQ01000; Nepal516, NC 008149; KIM10, NC 004088; B, NZ AAYT01000; C, NZ ABAT01000; D, NZ ACNS01000; E, NZ AAYS01000; F, NZ AAOS02000; CO92, NC 003143; G, NZ ABCD01000; H, NZ AAYV01000; I, NC 014029; J, NZ AAYR01000; Antiqua, NC 008150. **c**, Geographical origin of genome sequences used in **a** and **b**. **d**, Geographical spread of the Black Death from infection routes reported in ref. 4.

of ancient genomes will assist in determining the frequency of these mutations in co-circulating *Y. pestis* strains, and will clarify the emergence of branch 2 strains that are as yet unreported in ancient samples.

Consistent tree topologies were produced using several construction methods and all major nodes were supported by posterior probability (pp) values of >0.96 and bootstrap values >90 (Fig. 3b and Supplementary Figs 8 and 9). The trees place the East Smithfield sequence close to the ancestral node of all extant human pathogenic *Y. pestis* strains (only two differences in 1,694 positions) and at the base of branch 1 (Fig. 3b). A secure date for the East Smithfield site of 1348–1350 allowed us to assign a tip calibration to the ancient sequence and thus date the divergence time of the modern genomes and the East Smithfield genome using a Bayesian approach. Temporal estimates indicate that all *Y. pestis* commonly associated with human infection shared a common ancestor sometime between 668 and 729 years ago (AD 1282–1343, 95% highest probability density, HPD), encompassing a much smaller time interval than recently published estimates[14] and further indicating that all currently circulating branch 1 and branch 2 isolates emerged during the thirteenth century at the earliest (Fig. 3b), potentially stemming from an Eastern Asian source as has been previously suggested[14]. This implies that the medieval plague was the main historical event that introduced human populations to the ancestor of all known pathogenic strains of *Y. pestis*. This further questions the aetiology of the sixth to eighth century Plague of Justinian, popularly assumed to have resulted from the same pathogen: our temporal estimates imply that the pandemic was either caused by a *Y. pestis* variant that is distinct from all currently circulating strains commonly associated with human infections, or it was another disease altogether.

Although our approach of using an extant *Y. pestis* reference template for bait design precluded our ability to identify genomic regions that may have been present in the ancient organism and were subsequently lost in CO92, genomic comparisons of our ancient sequence against its closest outgroups may yield valuable insights into *Y. pestis* evolution. The Microtus 91001 strain is the closest branch 1 and branch 2 relative confirmed to be non-pathogenic to humans[16], hence genetic changes may represent contributions to the pathogen's adaptation to a human host. Comparisons against this outgroup revealed 113 changes (Supplementary Table 6a, b), many of which are found in genes affecting virulence-associated functions like biofilm formation (*hmsT*), iron-acquisition (*iucD*) or adaptation to the intracellular environment (*phoP*). Similarly, although its virulence potential in humans has yet to be confirmed to our knowledge, *Y. pestis* B42003004 isolated from a Chinese marmot population[17] has been identified as the strain closest to the ancestral node of all *Y. pestis* commonly associated with human plague, and thus may provide key information regarding the organism's evolution. Full genome comparison against the East Smithfield sequence revealed only eight single-nucleotide differences (Supplementary Table 6c), six of which result in non-synonymous changes (Supplementary Table 6d). Although these differences probably do not affect virulence, the influence of gene loss, gene gain or genetic rearrangements, all of which are well documented in *Y. pestis*[12,18], is as yet undetermined. In more recent evolutionary terms, single-nucleotide differences in several known pathogenicity-associated genes were found between our ancient genome and the CO92 reference sequence (Supplementary Table 3), which may represent further adaptations to human hosts.

Through enrichment by DNA capture coupled with targeted high throughput DNA sequencing, we have reconstructed a draft genome for what is arguably the most devastating human pathogen in history, and revealed that the medieval plague of the fourteenth century was probably responsible for its introduction and widespread distribution in human populations. This indicates that the pathogen implicated in the Black Death has close relatives in the twenty-first century that are both endemic and emerging[19]. Introductions of new pathogens to populations are often associated with increased incidence and severity of disease[20] and although the mechanisms governing this phenomenon are complex[21], genetic data from ancient infectious diseases will provide invaluable contributions towards our understanding of host–pathogen coevolution. The Black Death is a seminal example of an emerging infection, travelling across Europe and claiming the lives of an estimated 30 million people in only 5 years, which is much faster than contemporary rates of bubonic or pneumonic plague infection[22] and dissemination[7,8]. Regardless, although no extant *Y. pestis* strain possesses the same genetic profile as our ancient organism, our data suggest that few changes in known virulence-associated genes have accrued in the organism's 660 years of evolution as a human pathogen, further suggesting that its perceived increased virulence in history[23] may not be due to novel fixed point mutations detectable via the analytical approach described here. At our current resolution, we posit that molecular changes in pathogens are but one component of a constellation of factors contributing to changing infectious disease prevalence and severity, where genetics of the host population[24], climate[25], vector dynamics[26], social conditions[27] and synergistic interactions with concurrent diseases[28] should be foremost in discussions of population susceptibility to infectious disease and host–pathogen relationships with reference to *Y. pestis* infections.

METHODS SUMMARY

DNA from dental pulp was extracted and converted into sequencing libraries as previously described[3]. Potential sequencing artefacts resulting from deaminated nucleotides were eliminated by treatment of the DNA extracts with uracil-DNA-glycosylase and endonuclease VIII. DNA extracts were subsequently converted into sequencing libraries and amplified to incorporate unique sequence tags on both ends of the molecule. Two Agilent DNA capture arrays were designed for capture of the full *Y. pestis* chromosome (4.6 megabases), and the pCD1 (70 kb) and pMT1 (100 kb) plasmids using the modern *Y. pestis* strain CO92 (accession numbers NC_003143, NC_003131, NC_003134) for bait design with 3 bp tiling density. Serial array capture was performed over two copies of each array using the enriched fraction from the first round of capture as a template for a second round. The resulting products were amplified and pooled in equimolar amounts. All templates were sequenced for 76 cycles from both ends on the Illumina GAII platform, and reads merged into single fragments were included in subsequent analyses only if forward and reverse sequences overlapped by a minimum of 11 bp. Reads were mapped against the CO92 genome using the software BWA, and molecules with the same start and end coordinates were removed with the *rmdup* program in the *samtools* suite. Reference-guided sequence assembly was performed using Velvet version 1.1.03, with mapped and unmapped reads supplied in separate channels. Single-nucleotide differences were determined at a minimum of fivefold coverage and base frequency of at least 95% for both a pooled data set for all individuals and one in which all individuals were treated separately. A median network was constructed on these base calls using SplitsTree4. Phylogenetic trees were constructed using parsimony, neighbour-joining (MEGA 4.1) and Bayesian methods, and coalescence dates were determined in BEAST using both a strict and a relaxed molecular clock (Supplementary Fig. 9).

Received 25 July; accepted 9 September 2011.
Published online 12 October 2011.

1. Stoneking, M. & Krause, J. Learning about human population history from ancient and modern genomes. *Nature Rev. Genet.* **12,** 603–614 (2011).
2. Haensch, S. *et al.* Distinct clones of *Yersinia pestis* caused the Black Death. *PLoS Pathog.* **6,** e1001134 (2010).
3. Schuenemann, V. J. *et al.* Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1105107108 (29 August, 2011).
4. Benedictow, O. J. *The Black Death 1346–1353: The Complete History* (Boydell Press, 2004).
5. Scott, S. & Duncan, C. J. *Biology of Plagues* (Cambridge Univ. Press, 2001).
6. Cohn, S. K. *The Black Death Transformed: Disease and Culture in Early Renaissance Europe* (Arnold Publication, 2003).
7. Achtman, M. *et al. Yersinia pestis*, the cause of the plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl Acad. Sci. USA* **96,** 14043–14048 (1999).
8. Cowal, L., Grainger, I., Hawkins, D. & Mikulski, R. 2008. The Black Death Cemetery, East Smithfield, London (Museum of London Archaeology Service, 2008).
9. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104,** 14616–14621 (2007).

10. Hodges, E. *et al.* Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols* **4,** 960–974 (2009).
11. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134,** 416–426 (2008).
12. Chain, P. S. G. *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).
13. Achtman, M. *et al.* Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc. Natl Acad. Sci. USA* **101,** 17837–17842 (2004).
14. Morelli, G. *et al. Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nature Genet.* **42,** 1140–1143 (2010).
15. Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327,** 469–474 (2010).
16. Song, Y. *et al.* Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res.* **11,** 179–197 (2004).
17. 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).
18. Pouillot, F., Fayolle, C. & Carniel, E. Characterization of chromosomal regions conserved in *Yersinia pseudotuberculosis* and lost by *Yersinia pestis*. *Infect. Immun.* **76,** 4592–4599 (2008).
19. Stenseth, N. C. *et al.* Plague: past, present, and future. *PLoS Med.* **5,** e3 (2008).
20. Baum, J. & Bar-Gal, G. K. in *Emerging Pathogens Archaeology, Ecology & Evolution of Infectious Diseases* (eds Greenblat, C. L. & Spigleman, M.) 67–78 (Oxford Univ. Press, 2003).
21. Brown, N. F. *et al.* Crossing the line: selection and evolution of virulence traits. *PLoS Pathog.* **2,** e42 (2006).
22. WHO. Interregional meeting on prevention and control of plague. ⟨http://www.who.int/csr/resources/publications/WHO_HSE_EPR_2008_3w.pdf⟩ (2008).
23. Wood, J. W., Ferrell, R. J. & DeWitte-Aviña, S. N. The temporal dynamics of the fourteenth-century Black Death: new evidence from ecclesiastical records. *Hum. Biol.* **75,** 427–448 (2003).
24. Joosten, M. H. A. J., Cosijnsen, T. J. & De Wit, P. J. G. Host resistance to a fungal tomato pathogen lost by a single base pair change in an avirulence gene. *Nature* **367,** 384–386 (1994).
25. Xu, L. *et al.* Nonlinear effect of climate on plague during the third pandemic in China. *Proc. Natl Acad. Sci. USA* (2011).
26. Keeling, M. K. & Gilligan, C. A. Metapopulation dynamics of bubonic plague. *Nature* **407,** 903–906 (2000).
27. Barrett, R., Kuzawa, C. W., McDade, T. & Armelagos, G. J. Emerging and re-emerging infectious diseases: the third epidemiologic transition. *Annu. Rev. Anthropol.* **27,** 247–271 (1998).
28. Singer, M. & Clair, S. Syndemics and public health: reconceptualizing disease in bio-social context. *Med. Anthropol. Q.* **17,** 423–441 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. White (deceased), J. Bekvalac and R. Redfern from the Museum of London Centre for Human Bioarchaeology for access to samples, M. Kircher and S. Forrest for assistance with computational analysis, G. Wright for support throughout the project, past and present members of the McMaster Ancient DNA Centre for support throughout the project, and D. Poinar for constructive comments on earlier versions of the manuscript. We also thank S. Pääbo and the Max Planck Institute of Evolutionary Anthropology for use of their clean room facilities and molecular biology lab. Funding was provided by the Carl Zeiss Foundation (J.K.), the Human Genetics department of the Medical faculty in Tübingen (J.K.), the Canada Research Chairs program (H.N.P., G.B.G.), the Canadian Institute for Health Research (H.N.P.), the Social Science and Humanities Research Council of Canada (H.N.P.), the Michael G. DeGroote Institute for Infectious Disease Research (H.N.P., B.K.C., D.J.D.E.), an Early Research award from the Ontario Ministry of Research and Education (H.N.P.), the Natural Sciences and Engineering Research Council of Canada (D.J.D.E.), the James S. McDonnell Foundation (D.J.D.E.), and the University at Albany Research Foundation and Center for Social and Demographic Analysis and the Wenner-Gren Foundation (S.N.D.).

Author Contributions K.I.B., S.N.D., D.J.D.E., J.K. and H.N.P. conceived the project. K.I.B., S.N.D., S.S. and J.W. performed skeletal sampling. K.I.B., J.K. and V.J.S. carried out laboratory work. H.A.B., K.I.B., J.K., M.M. and H.N.P. designed experiments. K.I.B., G.B.G., J.K., H.N.P., V.J.S. and N.W. analysed the data. B.K.C., D.J.D.E., D.A.H. and J.B.M. provided valuable interpretations. P.B. provided technical support. K.I.B., J.K. and H.N.P. wrote the paper.

Author Information Sequencing data have been deposited in GenBank under the accession number SRA045745.1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.K. (johannes.krause@uni-tuebingen.de) or H.N.P. (poinarh@mcmaster.ca).

doi:10.1038/nature10549

Materials and Methods

Mortality records. Figure S1 is based on data collected and published by Cohn, 2003[1] (Fig. 7.33). The City of London did not publish bills of mortality in the 14th century. Cohn counted the numbers of last wills and testaments submitted to the Court of Husting in London each month in 1348 and 1349. While only relatively prosperous individuals would have had last wills and testaments, the temporal pattern of deaths indicated by these monthly counts is likely to be strongly correlated with the full epidemic curve for the population of London as a whole.

Samples. Dental material from the four individuals was harvested from the East Smithfield collection at the Museum of London as previously described[2]. The East Smithfield cemetery of London, England (Figure S2) is one of the few excavated medieval catastrophe burials in Europe, and historical documentation clearly indicates purchase of this land in late 1348 or early 1349 specifically for interment of victims of the pestilence[3,4], who had quickly exhausted available plots in local perish cemeteries. Shortly after the Black Death, the Abbey of St. Mary Graces was established on this land in 1350. The plot later housed the nineteenth-century Royal Mint for the city of London, hence the skeletal collection has been equivocally referred to as the "Royal Mint collection" in published work.

Extraction and screening. All manipulations of samples and DNA extracts were performed in facilities specifically dedicated to the extraction of ancient DNA with no prior exposure to sources of *Y. pestis* DNA, and following stringent protocols to limit laboratory contamination[5]. Pulverisation of dentin from both roots and molars was performed as previously described[2]. DNA extraction from dental material was performed by a protein kinase mediated lysis, followed by purification via either a modified phenol chloroform method as previously described[2] (McMaster) or a guanidinium-silica method[6] (MPI). A minimum of one negative extraction blank per 8 samples was carried along throughout.

Library preparation and indexing. 75µl of each DNA extract, extraction blank control or water library blank control (a minimum of one per every 10 libraries), was used for library preparation following the protocol of Meyer and Kircher (2010)[7] with modifications for ancient DNA[8]. For extracts purified via phenol chloroform, supernatant and pellet fractions were combined, and 75µl of the resultant pool was used in the immortalisation procedure. For all libraries, the blunt-ending step was preceded by treatment with uracil-DNA-glycosylase as well as Endonuclease VIII[9] to remove deaminated cytosines that contribute to C > T damage motifs in ancient DNA[10]. The chemistry of this procedure was scaled to accommodate a larger volume of extract. Post library preparation, each library was amplified with two 5'-tailed 'indexed' PCR primers, adding sample-specific indexes to both library adapters in this process[8]. Individual PCR reactions contained a maximum of $5x10^8$ library template molecules, and were amplified in 100µl reactions with the following thermal profile: 12 minute initial denaturation at 95°C, 10 cycles consisting of a 30 second denaturation at 95°C, a 30 second annealing at 58°C, and a 45 second elongation at 72°C, ultimately followed by at 10 minute final elongation at 72°C.

Array design. We used Agilent one million (1M) custom features arrays to capture *Yersinia pestis* genomic DNA libraries. We designed two 1M arrays by tiling 60 bp probes every three bases along the complete 4.6 Mb *Y. pestis* strain CO92 genome (NC_003143), the 70Kb pCD1 (NC_003131) and 96Kb pMT1 (NC_003134) plasmids. We included additional 60 bp probes every 1 base to get extra coverage of 933 single nucleotide polymorphisms (SNPs) that have been used to performed phylogenetic analysis of different *Y. pestis* isolates[11], and eight *Y. pestis* genomic regions that contain the virulence genes *hms* (hemin storage locus, YPO1951- YPO1954), *T3SS* (Type III secretion system, YPO0255 - YPO0273), an enhancing factor (YPO0339), the putative insecticide enhancing factors (YPO3678, YPO3681), and the putative insecticide toxin (YPO2312, YPO2380, YPO3673, YPO3674). The array design pipeline for mammalian genomes we used removes probes that contain repetitive elements based on 15-mer frequency counts[12]. First a table of frequencies for every 15-mer found in a given genome (sense and antisense) is calculated. Then, for each probe, the genomic frequency values of all of the probe's 15-mers are averaged. If this average genomic frequency of the 15-mers in one probe is above a given threshold, the probe is removed. A threshold of 100 has been used to design arrays based on mammalian genomes[12, 13]. We applied here the same methodology to remove probes that contain repetitive elements. Using 100 as a threshold, no probes were discarded. The highest frequency for a 15-mer in the *Y. pestis* genome was 66. The first array we designed (array A) contained probes covering the first 2.5 Mb of the *Y. pestis* chromosome, the second array (array B) contained the last 2.1 Mb of the chromosome, the plasmid pCD1, the plasmid pMT1, several virulence genes, and all SNPs taken from Morelli et al. 2010[11].

Array capture. Capture was performed following methods previously described[17] where all amplification reactions were performed with Phusion Hi-Fidelity DNA polymerase (New England Biosystems) in 100µl reactions with the following chemistry: 50µl Phusion High Fidelity Master Mix, 400 nM each p5 – p7 bridge amplification primer, and 10µl of template. Thermal profile consisted of a 3 minute initial denaturation at 95°C, between 5 to 10 cycles of a 30 second denaturation at 95°C, 30 second annealing at 60°C, and a 45 second extension time at 72°C, and a final elongation at 72°C for 5 minutes. PCR products were spin column purified over MinElute and quantitated via an Agilent 2100 bioanalyzer following manufacturer's protocols. The products were captured on two copies of each array A and array B in parallel, four arrays in total. After the first round of capture the 490µl eluate from the two identical arrays were pooled and PCR amplified in 20 independent PCRs in 100µl reactions containing 50µl PhusionTM High-Fidelity Master Mix, 1µM of each p5 – p7 bridge amplification primer and 50 µl library template and a total of 20 cycles. PCR products were spin column purified and quantified via an Agilent 2100 bioanalyzer. A second round of array capture was performed using the same conditions as the first round, only this time each of the two library pools were put on single arrays (array A or array B). The eluted 490µl from each array was subsequently amplified in 10 reactions using the amplification conditions described above.

Sequencing. The enriched libraries were pooled in equimolar concentrations and sequenced on the *Illumina Genome Analyzer IIx* platform using 2 x 76 + 7 +7 cycles on a single sequencing lane according to

the manufacturer's instructions for multiplex sequencing (FC-104-400x v4 sequencing chemistry and PE-203-4001 cluster generation kit v4). The manufacturer's protocol was followed except that an indexed control PhiX 174 library (index 5'- TTGCCGC-3') was spiked into each lane yielding a fraction of 2-3% control reads in all lanes of the run. The sequencing data were analyzed starting from QSEQ sequence files and CIF intensity files from the Illumina Genome Analyzer RTA 1.6 software. The raw reads were aligned to the PhiX 174 reference sequence to obtain a training data set for the base caller Ibis[15]. Raw sequences called by Ibis 1.1.1 were filtered for the 'AATCTTC' index as described[15]. The paired-end reads were subjected to a fusion process (including removal of adapter sequences and adaptor dimers) by requiring at least an 11nt overlap between the two reads. In the overlapping sequence, quality scores were combined and the base with the highest base quality score was called. Only sequences merged in this way were used for further analysis. A small number of larger molecules (longer than 141nt) had lower sequence quality, and were thus discarded. The total number of merged reads for individual samples and the total pool can be found in Table S1b.

Mapping, and calculation of average fragment length and genome coverage.

Mapping. Merged reads were aligned with BWA[16] to the CO92 *Y.pestis* reference genome (NC_003143), the 70Kb pCD1 (NC_003131), the 96Kb pMT1 (NC_003134), and the 9.6Kb pPCP1 (NC_003132.1) CO92 plasmids using default parameters. Mapping alignments were converted to SAM/BAM format[17] with BWA's *samse* command. Accurate mapping into different genomic regions was complicated by the presence of a 1956-bp transposase that is present in one copy in the pCD1 and the pPCP1, and in multiple copies on the pMT1 and the chromosome. One copy of this transposase was removed from the reference plasmid sequences prior to mapping, though its multiple copies in the chromosome and pMT contributed to inflated coverage estimates for these genetic components prior to additional filtering (Table S1b). The pooled dataset of all enriched extracts generated a total of 20,512,847 merged reads, of which 13,180,582 (64.26%) mapped to the CO92 reference chromosome. Multiple amplifications of the DNA library and enriched fractions performed as part of the capture procedure made it necessary to remove all identical molecules from the merged reads such that each starting library template was considered only once in genome assembly. This poses certain limitations when working with data from a bacterial population, since accuracy in identifying a unique template sequence requires properly distinguishing true genetic variation in the population, which can accrue rapidly even within a single host during pandemic episodes, from artefacts introduced via PCR misincorporations and Illumina sequencing. Duplicate molecules were removed using *rmdup*. This method clusters all reads together that have identical start and end coordinates in the mapping alignment, and uses the read with the highest mapping quality as representative of the cluster. This process, therefore, reduced the error signal resulting from amplification and sequencing, and also reduced the number of reads mapping to the transposase of the chromosome and pMT. The total number of mapping fragments before and after duplicate removal was calculated with *flagstat*[17] (Table S1b). Duplicate removal resulted in a total of

2,812,240 (13.71%) chromosomal reads, and filtering with Q>30 reduced this number to 2,366,647 (11.53%) high quality chromosomal reads. The average fragment length and length distribution was calculated using the SAM format loaded into *Galaxy*[18, 19]. Using the *pileup* function of *samtools*[17] all BAM files were converted into *pileup* files to calculate average coverage over the genome by dividing the total number of nucleotides mapping to the CO92 genome by its length using a mapping quality of Q>30. Average coverage was estimated at varying fold coverage and with and without mapping quality filtering (Table S1a, S1b, and S1c). To determine if we have sequenced the enriched library to exhaustion we estimated how often we have sequenced each original fragment for array A and array B separately. Average coverage of the target chromosomal region for each array was calculated using a mapping quality filter of Q>30 for the total reads both before and after duplicate removal (*rmdup*). Array A gave an average coverage of 66.09 before and 36.46 after duplicate removal, and array B gave 34.87 before and 22.41 after. Thus for array A, each original fragment was seen on average approximately 1.8 times whereas for array B each fragment was seen on average 1.5 times. This indicates that array A was sequenced to a greater depth than array B. The differences in sequencing depth for the arrays could be potentially explained by differences in enrichment efficiencies or experimental variation. Regardless of different sequencing depths, each array covered a comparable number of sites for the target region at a minimum of 1-fold coverage, indicating that further sequencing would not contribute to additional information on sites not currently covered (Figure 1b, main manuscript).

Genome architecture and contig assembly. Using the read dataset for individual 8291 aligned to the CO92 chromosome with BWA[16], all regions with a minimum mapping quality of 25 were extracted. A reference-guided sequence assembly was performed in Velvet version 1.1.03[20] using the extracted reference sequences supplied to the Columbus extension for reference-aided assembly in addition to all mapped and unmapped reads for individual 8291 in separate channels. A hash length of k = 21 was identified as producing an optimal assembly after iterations using several values for k. NCBI BLAST 2.2.21 was used for local similarity searches. The reference-guided assembly produced 130,556 contigs with an N50 of 288 bp. When filtered for contigs 500bp or larger 2134 contigs remained totalling 4,013,009 bp. This consisted of 30,959 unmapped reads. To identify contigs that potentially contained novel architecture distinct from CO92 and stemming from unmapped reads, all contigs were queried against the CO92 chromosome using BLAST[21] with an E-value cutoff of 1e-50. Twenty-nine were contigs that had no hits to CO92: six had 100% identity to human mtDNA, 8 matched mobile elements that were common to several species of bacteria, and 16 had no or low significant similarity to sequences in the NCBI database. A potential variant was considered for further examination if the contig sequence produced multiple hits to the CO92 chromosome. Any such contig would have segments that occur at a distance in the CO92 chromosome that the BLAST algorithm would be unable to extend through. Ideally these segments would be non-overlapping and would occur at a distance of more than one read length (54bp). Each contig was queried against other NCBI *Yersinia* sequences to check if the sequence proposed by the contig was observed in other species in the genus. Of the 20 such contigs identified (Table S2), half were either eliminated as mapping errors or were removed due to their location within

known repetitive regions. The remaining 10 contigs provided indications of altered structure in the ancient organism as compared to the reference CO92 genome. The contig numbers, length, description of the change relative to CO92, and the number of *Yersinia* strains in which each structural variant was observed are recorded in Tables S2a and S2b. An example is provided in Figure 2 in the main manuscript using contig number 6149, which suggests a breakpoint joining sequence in the *pbpC* pseudogene, which is intact in ancestral *Yersinia* strains, but spans a gap of 231 kb in the CO92 genome. The mapping alignment shows both joined ends occurring next to regions of high coverage (insertion element IS285), though a read depth of 20 is depicted in Figure 2 owing to space restrictions and resolution. Only short overlaps of several bases are observed across the breakpoint in the mapping alignment and no read is aligned across its centre at that position. Unmapped reads have spanned this breakpoint producing a natural overlap joining these two ends in reference-guided assembly. This specific orientation is shared only with *Y. pseudotuberculosis*, Microtus, Pestoides, and B42003004, which are ancestral to all currently circulating *Y. pestis* strains commonly associated with human infection. Separate orientations exist for this region in branch 1 (*pestis* Z176003, *pestis* D182038, *pestis* D106004, CO92, Antiqua) and branch 2 (Medievalis str. Harbin 35, Nepal516, KIM) strains, indicating that rearrangements in the *pbpC* pseudogene occurred as independent events on different lineages. This ancestral character supports the authenticity of our sequence data because these contigs demonstrate marked differences from modern *Y. pestis*. Each proposed breakpoint position was examined in the 8291 mapping alignment as well as a mapping alignment for each contig to determine if 1) the reconstruction was supported by multiple reads, 2) it did not occur in a repetitive region of CO92, and 3) was not produced by mis-assembly of a collapsed repeat. Additionally, the mapping alignment needed to indicate that the breakpoints occur in relatively poorly mapped regions of CO92 such as that described for contig 6149 above. This process revealed that 10 of the 20 contigs resulted from improper assembly (Table S2b).

Genomic analysis. To identify positions that differ from the CO92 *Y.pestis* reference genome, pileup files generated from BAM files were filtered for positions with a mapping score of Q>30, a major base frequency of at least 95% and at least a 5 fold coverage using *Galaxy*[18, 19]. Using these conditions 97 nucleotide positions were found to be different between the CO92 reference genome and the East Smithfield individuals 11972, 8291, 8134, and 94 positions between 6330 and CO92 (Table S3a). All 97 (or 94) nucleotides were found to be in the ancestral state in the East Smithfield individuals when compared to *Y. pseudotuberculosis*, thus not a single unique derived position was found in the ancient bacterial genome. Furthermore, all of these positions are known variants in extant *Y. pestis* strains[11]. The same strategy was used for reference comparison against the three plasmids. The average coverage of the plasmids was found to be similar for both the pMT1 and the pCD1. The PCP1 plasmid has an estimated copy number of 186 within *Y.pestis*[22] and was not put on the designed capture arrays since it was sequenced at high coverage in a previous work[2]. The estimated low coverage of less than one fold was, therefore, expected. For the pMT1 plasmid 2 differences to CO92 were found, and 4 differences for the pCD1. As with the chromosomal positions, all differences from the reference sequences in the

plasmids were ancestral in the East Smithfield individuals. Differences from the reference sequence are recorded in Table S3a and S3b.

Variant positions were identified by comparing base call frequency data from all high-quality reads for each position with individuals treated separately. A position was considered polymorphic if at least two different nucleotides were recorded in reads with a minimum of 5-fold coverage, permitting 0.9675 confidence. All polymorphic positions are shown in Table S5.

Heat plots were generated for the CO92 genome and for the plasmid genomes to show the relationship between read depth and GC content. In each case the number of sites that have a given number of reads per site and a given GC content (averaged over the 30 neighbouring bases) is colour encoded. The heat plots (Figure S6) show that most sites are covered by a comparatively small number of reads and have intermediate GC content. There is a small skew toward less coverage when the GC content is low. This skew becomes more apparent in the plasmids.

Phylogenetic analysis. We used 1694 of the 1748 positions that were previously identified as variant in modern *Y. pestis* genomes[11], since a multiple alignment of 17 modern genomes as well as the outgroup *Y. pseudotuberculosis* using MAUVE[23] revealed only 1694 of them to be variant in extant *Y. pestis*. The genome sequence of the Angola strain (NC_010159) was excluded from phylogenetic analysis due to its high divergence compared to the *Y. pseudotuberculosis* outgroup[11]. Strain NZ_AAUB01000 was excluded from further analysis due to poor sequence quality. The 27 previously unreported variant positions discovered in our analysis that are derived in the CO92 reference sequence were not included in the 1694 variant position since they could not be confirmed via HPLC[11] and could potentially represent sequencing artefacts in the publically available reference sequence (NC_003143). All 1694 positions were called from the four East Smithfield individuals (Table S4) using *pileup* files and filtered for a minimum coverage of 5-fold and mapping quality of Q>30 using Galaxy[18, 19]. A 95% majority base filter was not used in this step. Using the software SplitsTree4[24] median networks were calculated for all 1694 positions from the 18 modern genomes as well as a subset excluding the Branch 0 genomes[11], and compared to all four East Smithfield individuals. We found that three of the four East Smithfield individuals clustered together (8124, 8291, 11972) whereas one individual (6330) fell closer to other Branch 1 strains (Figure 3, main manuscript). Phylogenetic trees using the Maximum Parsimony (MP) and the Neighbour Joining (NJ) algorithm implemented in MEGA 4.1[25] confirmed a distinct position of individual 6330 compared to the others (Figure S8a and S8b). Thus individual 6330 was excluded for further analysis and individuals 8124, 8291, and 11972 were pooled to get a higher number of informative sites (Table S4, pool 3). The resulting MP and NJ trees are well resolved based on 1000 bootstrap replicates with bootstrap values >90 (Figure S8c and S8d).

Divergence times. To estimate the divergence times of all *Y.pestis* strains the Bayesian algorithm implemented in BEAST v1.5.3 was used. The pool (8124, 8291, 11972) consensus was analyzed using a relaxed uncorrelated log-normal molecular clock as well as the strict clock option[26]. As a tip calibration point we set the prior distribution age of the East Smithfield sample to 1348 - 1350 AD as well as the

isolation years for the modern strains as described elsewhere[11]. For each analysis, we used a model that assumes a constant population size across the phylogeny and ran 50,000,000 generations of the Markov Chain Monte Carlo with the first 5,000,000 generations discarded as burn-in. We chose the HKY sequence evolution model. We performed two BEAST runs and merged both using Logcombiner[26]. The estimated mutation rate was 1.96E-08 (HPD 95% lower 1.7E-08, HPD 95% higher 2.2E-08) mutations per site and per year over the whole genome, using a strict molecular clock[26] . This estimate is similar to previously determined mutation rates for *Y.pestis*[11]. A consensus tree of all 90,000 trees was inferred using TreeAnnotater V.1.4.8. As observed for the NJ and MP tree all nodes were found to be well resolved in the Bayesian tree with posterior probabilities of >0.96 (Figure S9). The estimated coalescence times for the strict and relaxed molecular clock are shown in Figure S9.

Identifying synonymous vs non-synonymous changes.

Comparisons with related genomes. Descriptions of genomic regions for the nucleotide differences were determined using the current annotation of C092 and its plasmids. The location of the difference is given in the *mpileup* file and this was determined to be either synonymous or non-synonymous given the estimated mapping position and the location within the annotation.

A number of differences in coding regions for proteins involved in various aspects of pathogen virulence were observed between the ES strain and both *Y. pestis* Microtus and *Y. pestis* CO92 (Tables S3 and S6d). Although the specific role of these differences in virulence has not been assessed, the genes in which they appear are well known virulence factors.

Supplementary figures:

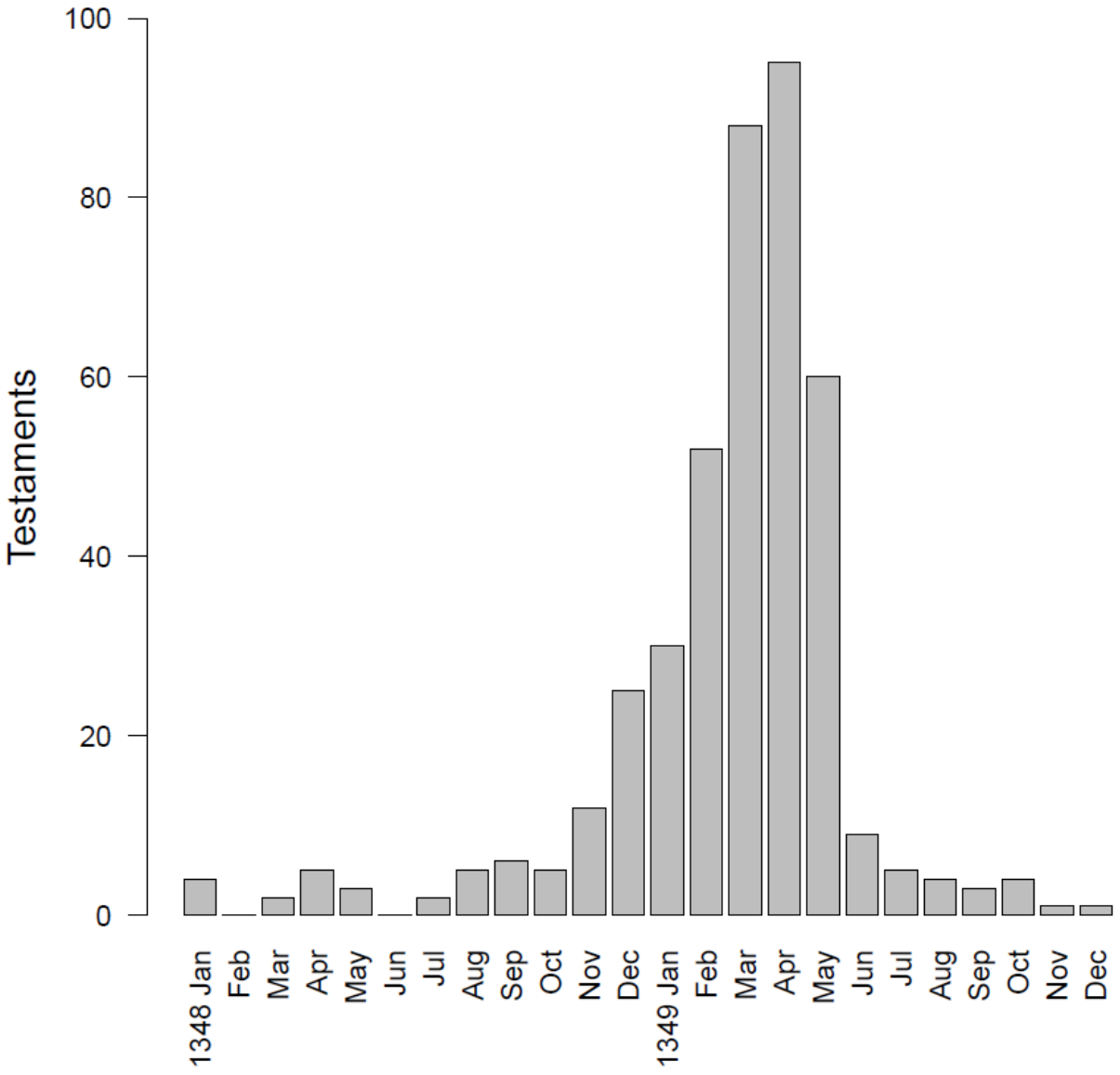


Figure S1 – Monthly numbers of Last Wills and Testaments submitted to the Court of Husting in London in 1348 and 1349, as counted by Cohn, 2003[1]. Deaths reported in this way would have represented a small sample of all deaths in London, but the structure of the epidemic curve is clear nevertheless.

LONDON, 1593. By JOHN NORDEN.

Figure S2 – John Norden's map of London in 1593. East Smithfield is indicated on the far right of the map in red, just north of the river.

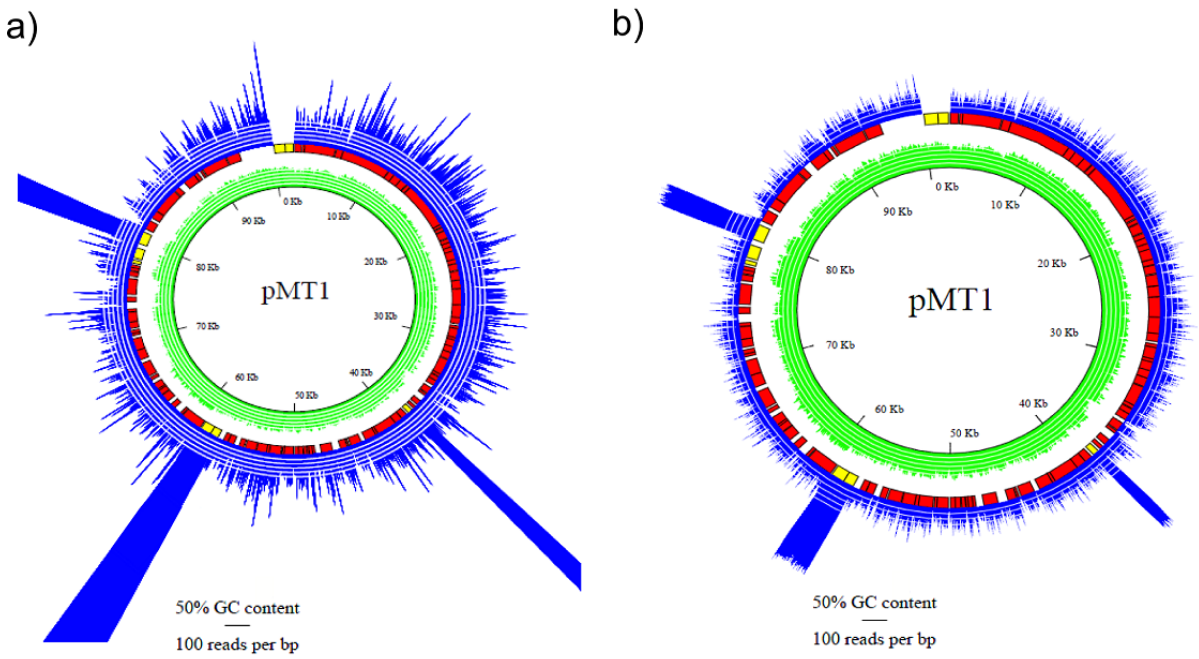


Figure S3 – Effect of duplicate removal via *rmdup*. The coverage plot in a) shows the mapping of all fragments against the pMT1 reference (NC_003134) prior to duplicate removal and before filtering for mapping quality. Coverage is shown in blue, GC frequency for the chromosome in green. Scale lines correspond to 10, 20, 30, 40, 50 fold coverage and to 0.1, 0.2, 0.3, 0.4, 0.5 GC content. Red corresponds to coding regions, yellow to mobile elements. A large tranposase corresponding to positions 1 – 1962 was removed, though three additional copies of this transposase in the pMT have contributed to three areas of dense coverage. The coverage plot in b) shows the effect of duplicate removal, which greatly reduces the number of fragments mapping to the pMT reference, most notable in coverage density in the three remaining transposases.

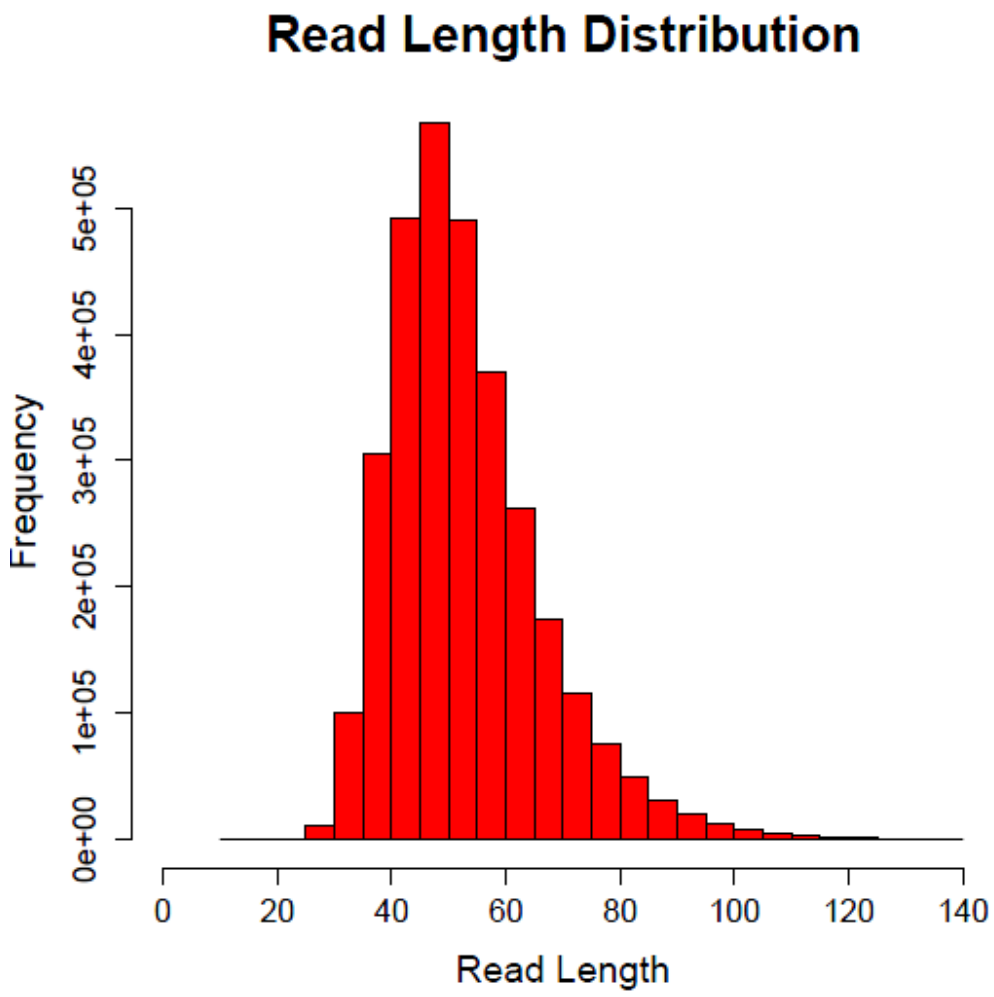


Figure S4 – Fragment size distribution of reads post duplicate removal (*rmdup*) and mapping filtering (Q>30) from a pooled collection of all libraries.

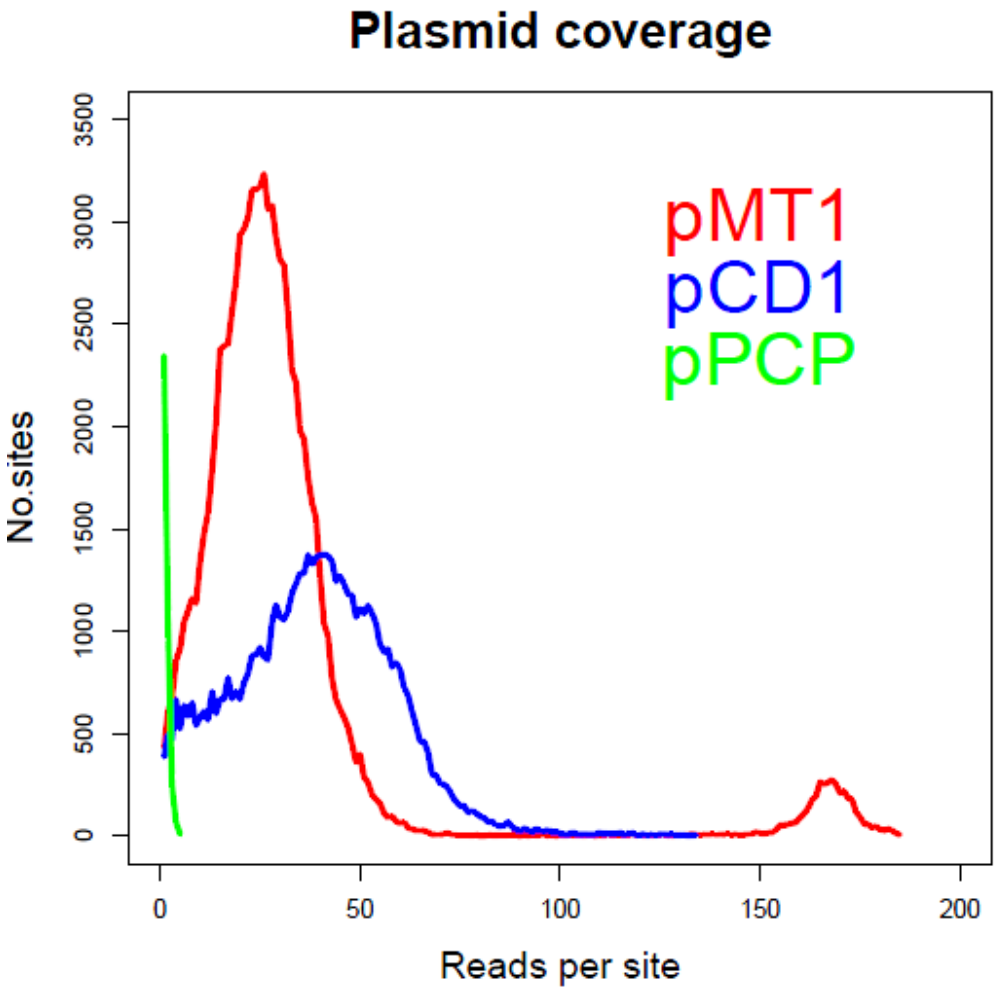


Figure S5 – Coverage plot for East Smithfield *Y. pestis* plasmids

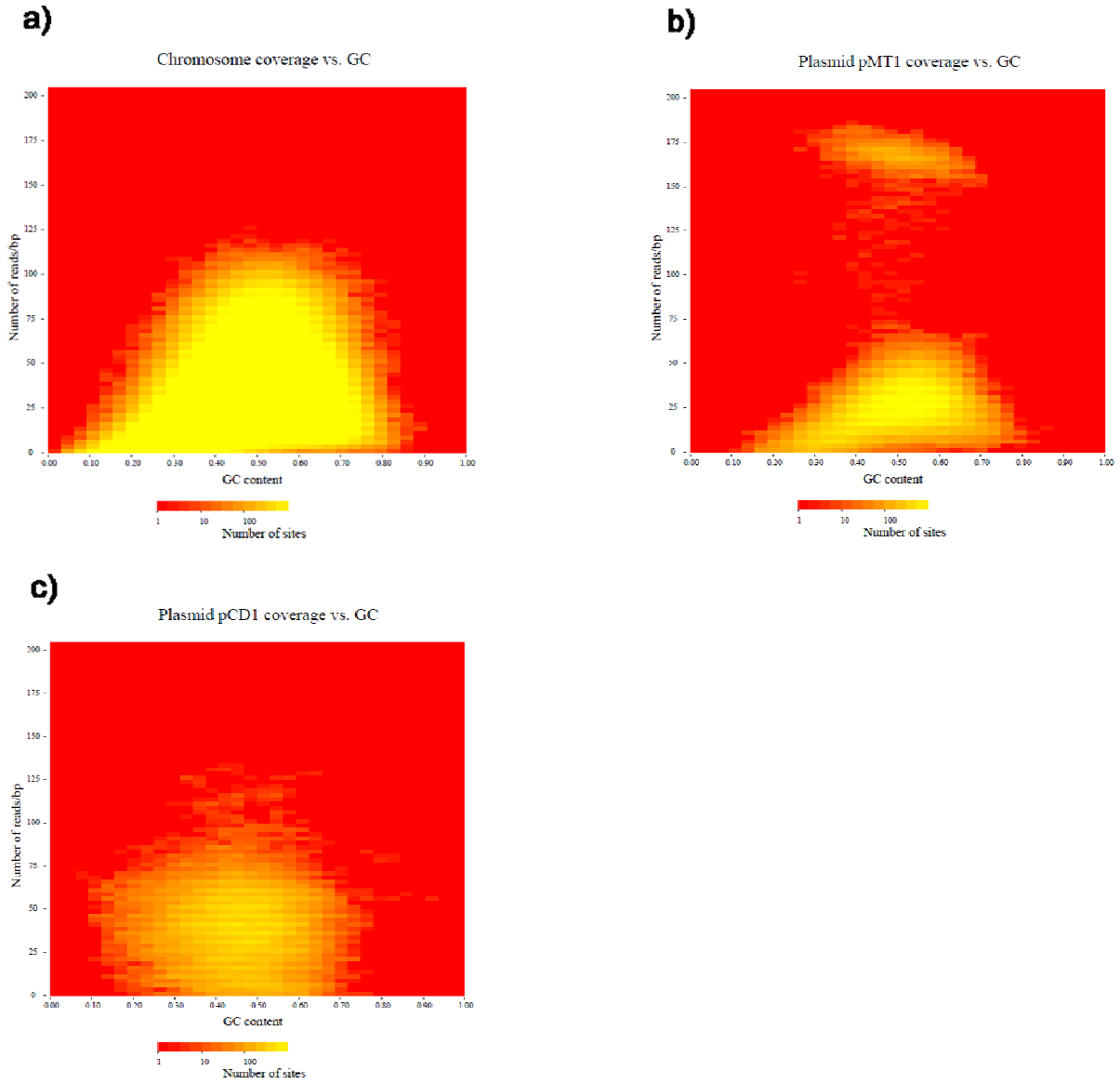


Figure S6 – Heat plots showing relationship of GC content on coverage for a) the chromosome, b) the pMT plasmid, and c) the pCD1 plasmid. The number of sites is shown as a function of the number of reads per site and the GC content averaged over the 30 neighbouring bases. Most sites are covered by fewer than 30 reads and have intermediate GC content. There is a small skew toward less coverage when the GC content is low.

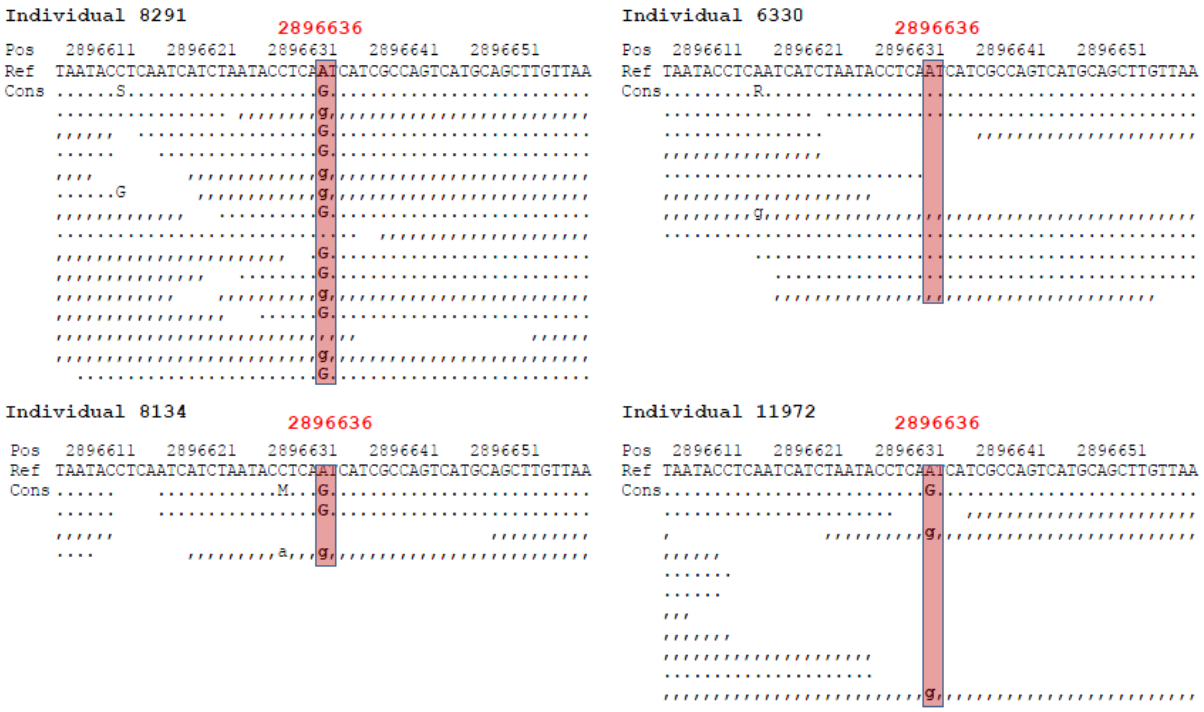


Figure S7 – Alignment showing base frequencies for position 2896636 in all four individuals. The position is polymorphic in individual 8291 and fixed derived in 6330. Individuals 11972 and 8124 show the ancestral state, though coverage is too low to ascertain whether the position is polymorphic with 0.9675 confidence.

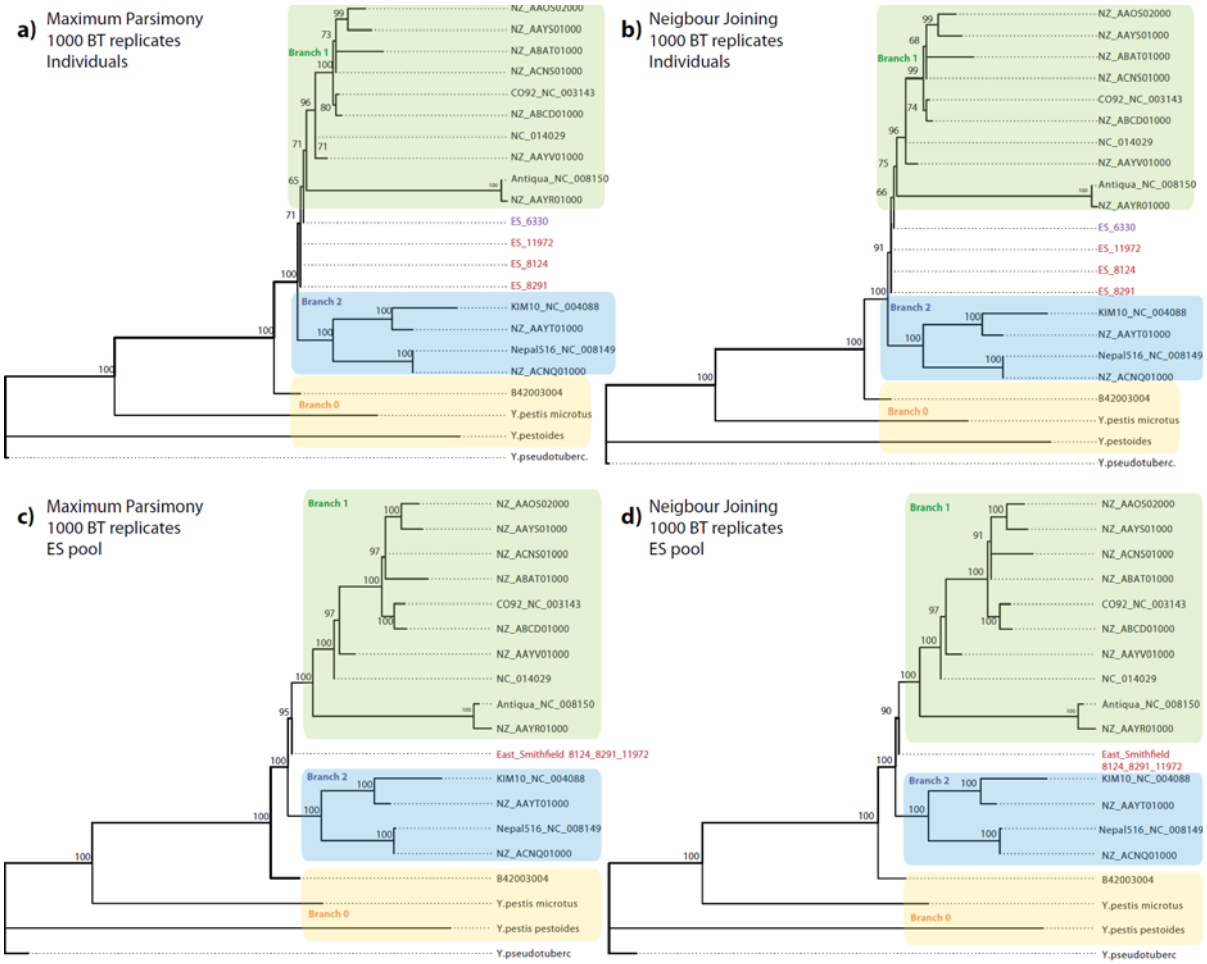


Figure S8 – Phylogenetic trees based on 1694 positions known to be variable within modern *Y. pestis* strains, reconstructed using Maximum Parsimony (a and c) and Neighbour Joining (b and d) algorithms implemented in MEGA 4.1[25]. For a) and b) the East Smithfield individuals are kept separate; for c) and d) individuals 8124, 8291, and 11972 are pooled together.

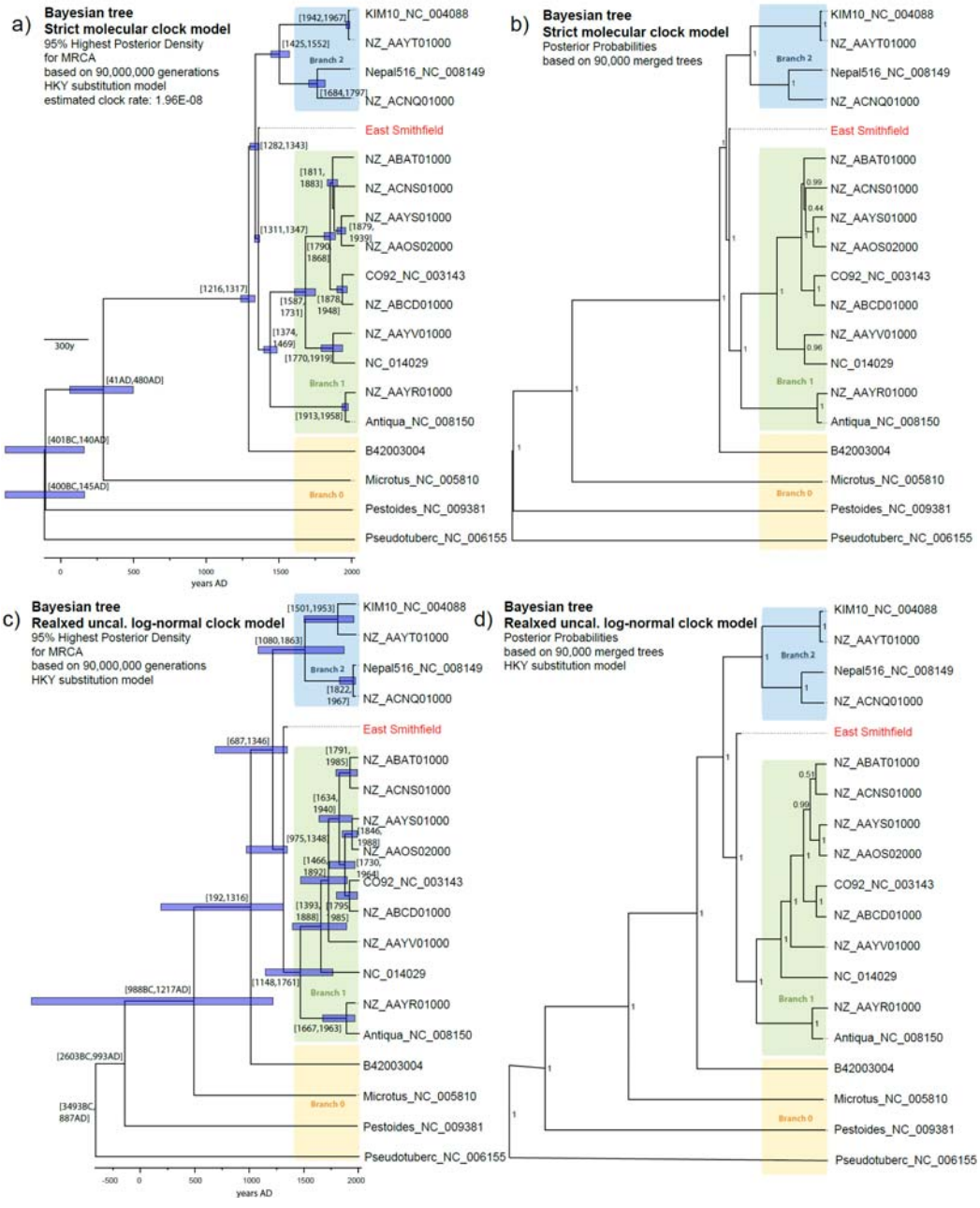


Figure S9 – Phylogenetic trees reconstructed using a Bayesian approach with BEAST and two times 50,000,000 generations. Each tree is derived from a consensus of 90,000 merged trees[26]. Trees a) and b) used a strict molecular clock; trees c) and d) used a relaxed uncorrelated log-normal clock. Trees a) and c) show divergence dates on each node; b) and d) show posterior probabilities for each node. Blue bars show 95% Highest Posterior Density for the divergence date of the different *Y.pestis* strains.

Supplementary tables:

Table S1 – Estimates of East Smithfield sequence coverage.

Table S1a – Fold coverage estimates showing effect of duplicate removal (*rmdup*) and mapping filtering.

fold chromosomal coverage	post duplicate removal	post duplicate removal and mapping filtering Q>30
1 fold	99.25%	93.48%
2 fold	98.62%	92.85%
3 fold	97.98%	92.22%
4 fold	97.29%	91.54%
5 fold	96.55%	90.80%
10 fold	91.50%	85.76%
30 fold	48.30%	42.67%

Table S1b – description of all mapped reads

[Included as a separate file]

S1c – Fold coverage estimates presented as average coverage per site.

individual	Chromosome	pMT1	pCD1	pcp1
ES_11972	6.6	10.1	7.9	0.3
ES_6330	3.5	7.6	5.4	0.1
ES_8124	2.2	7.5	2.7	0
ES_8291	14.8	19.9	19.4	0.2
Pool	28.2	31.2	35.2	0.7
Controls	0	0	0	0

Table S2 – Description of reference-guided contigs showing architectural differences between the East Smithfield genome and the CO92 reference.

Table S2a – Reference-guided contigs that differ from CO92

Contig	Length (bp)	Rearrangement	Found in *Y. pestis*	Found in *Y.pseudotuberculosis*
1530	1149	20 bp indel	7	3
208	2062	insertion of IS100 in CO92	10	4
3036	1540	30kb rearrangement	7	4
3293	2519	9 bp indel	7	0
4326	963	insertion of IS100 in CO92	10	0
4746	4923	32 bp indel	9	1
6149*	9096	231 kbp rearragement	3	4
6289	659	insertion of IS285 in CO92	10	4
7016	10311	insertion of IS100 in CO92	9	4
7127	2526	insertion of IS100 in CO92	5	3

*Intact contig found in Pestoides, Microtus, Angola, B42003004 and *Y. pseudotuberculosis*

Table S2b – Incorrectly reconstructed reference-guided contigs that differ from CO92

Contig	Length (bp)	Description	Found in *Y. pestis*	Found in *Y.pseudotuberculosis*
2935	1477	in repetitive region in CO92	0	0
3181	1704	near phage protein in CO92	0	0
5765	3992	near repeated tRNA val in CO92	0	0
6107	4825	misassembly indel	0	0
6798	603	near pseudogene in C092	0	0
7052	5267	misassembly indel	0	0
7405	8587	misassembly indel	0	0
7488	4398	misassembly indel	0	0
8057	8059	misassembly indel	0	0
43500	6912	misassembly indel	0	0

Table S3 – nucleotide differences identified between the ancient genome and the CO92 reference genome used for bait design for a) the chromosome, and b) the pCD1 and pMT1 plasmids.

[Included as a separate file]

Table S4 – Known polymorphic positions[11] used for phylogenetic reconstruction.

[Included as a separate file]

Table S5 – Variant positions recorded in the East Smithfield sequences.

[Included as a separate file]

Table S6 – Substitutions of note in the two outgroups most closely related to the ES sequence. a) Differences between Microtus 91001 and the other 18 available *Yersinia* genomes. b) Differences between Microtus 91001 and ES with genetic descriptions. c) Differences between B42003004 and the other 18 available *Yersina* genomes. d) Differences between B42003004 and ES with genetic descriptions.

[Each included as separate files]

References Cited:

1. Cohn, S.K. The Black Death transformed: disease and culture in early Renaissance Europe. London: Arnoldohn (2003).
2. Schuenemann, V.J. *et al*. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc Natl Acad Sci USA* doi: 10.1073/pnas.1105107108.
3. Cowal, L.., *et al*. The Black Death Cemetery, East Smithfield. London. Great Britain: Museum of London Archaeological Svc (2008).
4. Hawkins, D. The Black Death and new London cemeteries of 1348. *Antiquity* **64**, 637 – 642 (1990).
5. Pääbo S., *et al*. Genetic analyses from ancient DNA. *Annu Rev Genet* **38**,645-679 (2004).
6. Rohland, N., and Hofreiter M. Ancient DNA extraction from Bones and Teeth. *Nat Protoc* **2**, 1756 – 1762 (2007).
7. Meyer, M. and Kircher M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb Protoc* **6**; doi:10.1101/pdb.prot5448 (2010).
8. Kircher, M., Sawyer S., and Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Submitted.
9. Briggs, A.W., *et al*. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nuc. Acids Res.***38**, e87 (2009).
10. Hofreiter, M *et al*. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nuc Acids Res* **29**, 4793-4799 (2001).
11. Morelli, G., *et al*. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* **42**, 1140 – 1143 (2010).
12. Hodges, E., *et al*. Genome-wide *in situ* exon capture for selective resequencing. *Nat Genet* **39**, 1522 – 1527 (2007).
13. Burbano H.A., *et al*. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science* **328**, 723 – 725 (2010).
14. Hodges, E., *et al*. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protoc* **4**, 960 – 974 (2009).
15. Kircher, M., Stenzel U, and Kelso J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* **10**: R83 (2009).
16. Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**,1754-60 (2009).
17. Li, R., *et al*. SOAP2: and improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966 – 1967 (2009).
18. Goecks, J., *et al*. Galaxy: a comprehensive approack for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86+ (2010).

19. Blankenberg, D., *et al*. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol*. Chapter **19**, 1-21 (2010).
20. Zerbino, D. *R.,* Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829 (2008).
21. Altschul, S.F., *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
22. Parkhill J., *et al*. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523 – 527 (2001).
23. Darling A.C., *et al*. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**, 1394-1403 (2004).
24. Huson, D.H. and Bryant D. Application of Phylogentic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **23**, 254 – 267 (2006).
25. Tamura, K., *et al*. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596 – 1599 (2007).
26. Drummond, A.J., *et al*. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol*. **4**, e88 (2006).

.