Title

Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline

Journal

Cell, 176(1-2)

ISSN

0092-8674

Authors

Rascovan, Nicolás
Sjögren, Karl-Göran
Kristiansen, Kristian
et al.

Publication Date

2019

DOI

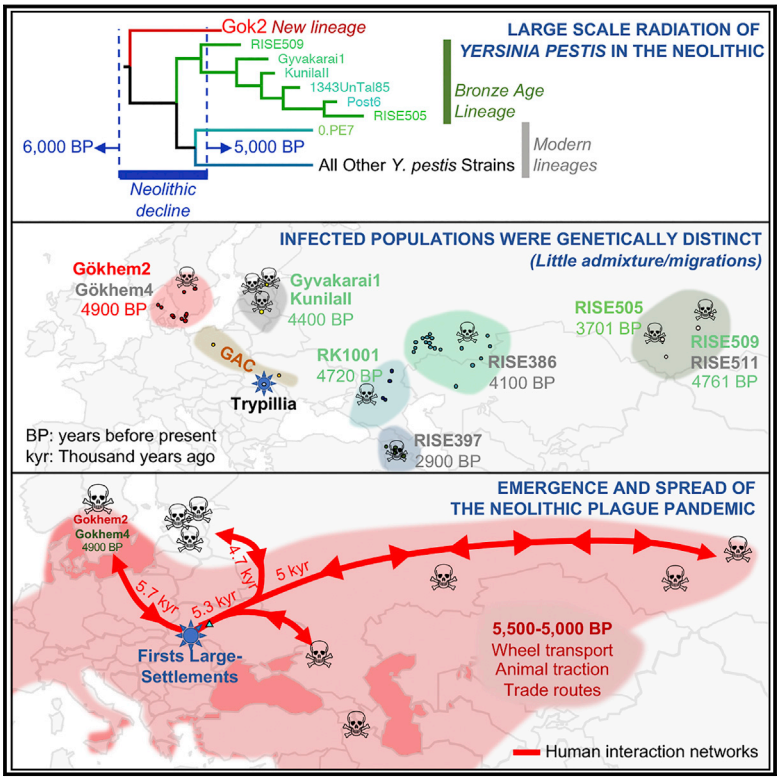
10.1016/j.cell.2018.11.005

Article

Cell

Emergence and Spread of Basal Lineages of *Yersinia pestis* during the Neolithic Decline

Graphical Abstract



Highlights

- Discovery of the most ancient case of plague in humans, 4,900 years ago in Sweden
- Basal lineages of *Y. pestis* emerged and spread during the Neolithic decline
- Plague infections in distinct Eurasian populations during Neolithic and Bronze Age
- A plague pandemic likely emerged in large settlements and spread over trade routes

Authors

Nicolás Rascovan, Karl-Göran Sjögren, Kristian Kristiansen, Rasmus Nielsen, Eske Willerslev, Christelle Desnues, Simon Rasmussen

Correspondence

nicorasco@gmail.com (N.R.), simon.rasmussen@cpr.ku.dk (S.R.)

In Brief

The genome of an ancient strain of *Yersinia pestis* from Neolithic farmers 4,900 years ago represents the oldest discovered case of the plague and allows characterization of the spread and diversification of multiple basal lineages, potentially contributing to the Neolithic decline.

Rascovan et al., 2019, Cell *176*, 295–305
January 10, 2019
https://doi.org/10.1016/j.cell.2018.11.005

CellPress

Cell

Emergence and Spread of Basal Lineages of *Yersinia pestis* during the Neolithic Decline

Nicolás Rascovan,[1,*] Karl-Göran Sjögren,[2,8] Kristian Kristiansen,[2,8] Rasmus Nielsen,[3,4] Eske Willerslev,[3,5,6] Christelle Desnues,[1] and Simon Rasmussen[7,9,10,*]

[1]Aix Marseille Université, UMR MEPHI, CNRS FRE2013, IRD 198, AP-HM, IHU - Méditerranée Infection, 19-21 Boulevard Jean Moulin, 13005 Marseille, France
[2]Department of Historical Studies, University of Gothenburg, 405 30 Gothenburg, Sweden
[3]Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark
[4]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[5]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
[6]Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK
[7]Department of Bio and Health Informatics, Technical University of Denmark, Kemitorvet 208, 2800 Kongens Lyngby, Denmark
[8]These authors contributed equally
[9]Present address: Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark
[10]Lead Contact
*Correspondence: nicorasco@gmail.com (N.R.), simon.rasmussen@cpr.ku.dk (S.R.)
https://doi.org/10.1016/j.cell.2018.11.005

SUMMARY

Between 5,000 and 6,000 years ago, many Neolithic societies declined throughout western Eurasia due to a combination of factors that are still largely debated. Here, we report the discovery and genome reconstruction of *Yersinia pestis*, the etiological agent of plague, in Neolithic farmers in Sweden, pre-dating and basal to all modern and ancient known strains of this pathogen. We investigated the history of this strain by combining phylogenetic and molecular clock analyses of the bacterial genome, detailed archaeological information, and genomic analyses from infected individuals and hundreds of ancient human samples across Eurasia. These analyses revealed that multiple and independent lineages of *Y. pestis* branched and expanded across Eurasia during the Neolithic decline, spreading most likely through early trade networks rather than massive human migrations. Our results are consistent with the existence of a prehistoric plague pandemic that likely contributed to the decay of Neolithic populations in Europe.

INTRODUCTION

The spread of farming practices across western Eurasia was followed by a period of demographic expansion (Hinz et al., 2012; Müller, 2013; Shennan et al., 2013) and technological innovations (e.g., pottery, animal traction, the wheel, metallurgy, etc.). These events favored the initiation of trading networks, which, for the first time, could reach remote geographic regions and connect multiple independent human populations. Population densities and settlements also increased in size with the emergence of the first mega-settlements (6,100–5,400 BP) in the current territory of Moldova, Romania, and Ukraine. These settlements were built by a population known as the Trypillia Culture (6,800–5,000 BP) and could host between 10,000 and 20,000 people, about an order of magnitude larger and with more complex organization than in any preceding period (Müller et al., 2016). Indeed, the archaeological record of these sites shows signs of labor division and specialization (Korvin-Piotrovskiy et al., 2016) and high densities of humans and animals living in close contact with each other, which would have generated high demands of food and resources (Kirleis and Dal Corso, 2016) and also favorable conditions for the emergence of infectious diseases.

Mega-settlements were usually short-lived and were regularly abandoned, burned, and reconstructed after some 150 years. However, after ~5,400 BP, such settlements were no longer built for reasons that are not yet fully understood (Diachenko, 2016). In fact, during this period, known as the Neolithic Decline (Kristiansen, 2014), Neolithic cultures throughout Europe went through a period of population decline (Hinz et al., 2012; Shennan et al., 2013). The reasons for this are debated (Downey et al., 2016). The most accepted explanations for the collapse of mega-settlements are environmental overexploitation, with a decrease or even extinction of forests and the expansion of steppe environment and/or a confrontation with foraging Steppe populations (Anthony, 2007; Kirleis and Dreibrodt, 2016). On the other hand, the emergence of infectious diseases is a possible third contributing factor to the decline. The close contact between humans and animals and the accumulation of food, likely led to poorer sanitary conditions and an increased risk of pathogen emergence and transmission in human settlements of the Neolithic and afterward (Armelagos and Harper, 2005; Barrett et al., 1998). In fact, the presence of infectious diseases is a suitable explanation for the massive burning of houses and rapid abandonment observed in mega-sites.

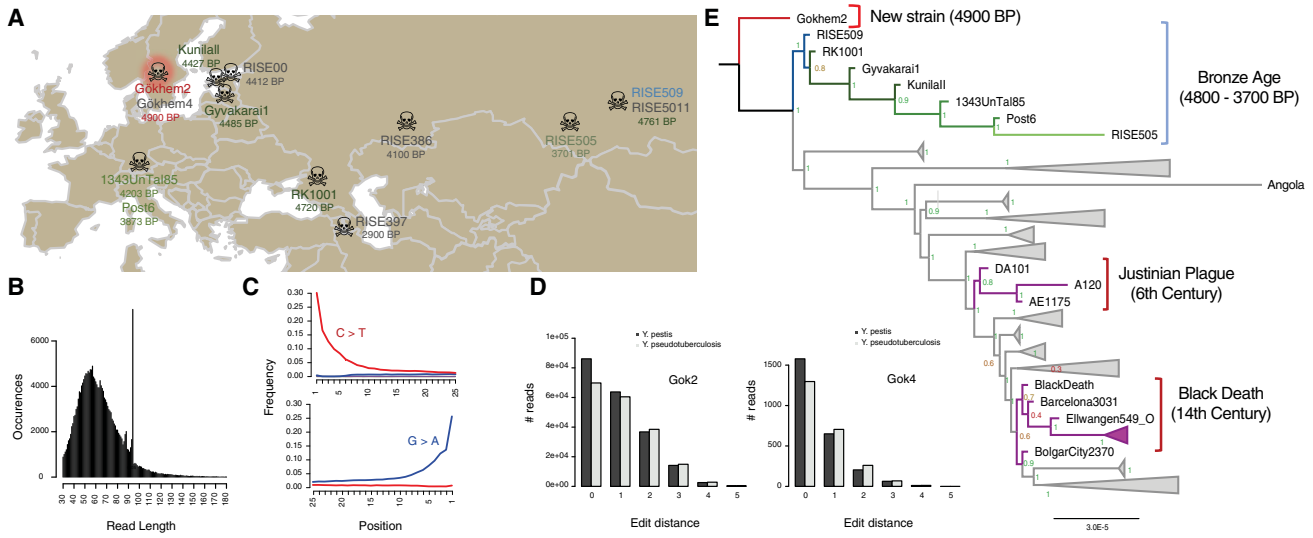


Figure 1. Discovery of the Gok2 *Y. pestis* Strain and Its Basal Phylogenetic Placement

(A) Archaeological sites and carbon dating (in years before present) of the individuals infected with Gok2/Gok4 and Bronze Age *Y. pestis* strains. Samples are colored as in the phylogeny (E) or in gray when they had low genome coverage.

(B and C) The expected ancient DNA degradation patterns of the Gok2 strain showing short read distribution (B), and increased C > T and G > A substitutions in the 5′ and 3′ ends of reads (that are eliminated in downstream analyses) (C).

(D) Distribution of edit distances of high quality reads from Gok2 and Gok4 mapped to either *Y. pestis* (dark gray) or *Y. pseudotuberculosis* (light gray) reference genomes. Reads have a higher affinity to *Y. pestis* than to the closest related species *Y. pseudotuberculosis*.

(E) Whole genome phylogeny reconstructed using maximum likelihood of modern (163, in gray) and ancient (20, in color) *Y. pestis* strains. The tree was reconstructed using RAxML from an alignment of 3,558 *Y. pestis* genes with codon-based partitioning. Bootstrap support is shown at the nodes as a fraction of 100 bootstrap replicates generated using RAxML and hereafter reconstructed using RAxML. The tree is rooted using *Y. pseudotuberculosis* (not shown) and Gok2 was found to be basal to all known *Y. pestis* strains.

See also Figures S1 and S2 and Tables S1, S2, S3, and S4.

It has also been proposed that the Neolithic decline reached northern European populations in Scandinavia in a process starting around 5,300 BP (Hinz et al., 2012). Two recent works (Skoglund et al., 2012, 2014) described the genetic history of early Neolithic farmers in Scandinavia from the Funnel Beaker culture (TRB), a population that used to live in small settlements consisting of dispersed farmsteads, probably family based. Some of the individuals in these studies were found in a passage grave at Frälsegården in Gökhem parish, Falbygden, western Sweden, which grouped up to 78 individuals buried between 5100–4900 BP (based on carbon dating of 34 individuals), a quite short period of time and large number of bodies compared to other Scandinavian sites (Ahlström, 2009; Sjögren, 2015). A possible explanation for the magnitude and short duration of this grave was an epidemic event. We thus reanalyzed publicly available ancient DNA datasets from individuals of this grave and screened for the presence of known human pathogens. Unexpectedly, we found the unambiguous presence of *Yersinia pestis*, the etiological agent of plague, in two different individuals, dated to ~4,900 BP. These individuals were slightly older than the most ancient known *Y. pestis* infections, which were reported in ancient human populations of the Eurasian Steppe (Andrades Valtuena et al., 2017; Rasmussen et al., 2015). The presence of *Y. pestis* in this time period, geographic region, and host population did not fit with previously proposed models of early plague dispersion (Andrades Valtuena et al., 2017; Spyrou et al., 2018). We therefore integrated phylogenetic, molecular clock, and genotyping analyses from a comprehensive set of modern and ancient *Y. pestis* strains and whole-genome analyses of human individuals of that period. We combined this with the adequate archaeological contextualization and searched for a more parsimonious model that properly reconciled all the existing evidence. Our analyses revealed that during the decline of Neolithic populations in Europe, between 6,000–5,000 BP, multiple lineages of *Y. pestis* branched and expanded throughout Eurasia. The analysis of the archaeological context and the human genomes revealed that the emergence and spread was not due to massive migrations, but more likely facilitated by the lifestyles, population growth, and the expanding trade networks.

RESULTS AND DISCUSSION

Discovery of the Gok2 Neolithic *Y. pestis* Strain in Northern Europe

The screening of known human pathogens in public ancient DNA datasets of teeth from individuals found in the Frälsegården passage grave in Sweden (Sjögren, 2015; Skoglund et al., 2014) revealed the unambiguous presence of *Y. pestis* in the "Gökhem2" individual (a.k.a. individual A), a 20-year-old female, dated to 5,040–4,867 BP and belonging to the Funnel Beaker culture (TRB) (Figure 1A; Table S1). The alignment of all sequencing

reads to the *Y. pestis* genome resulted in 203,733 unique reads recovered, covering 79.3% of the chromosome at 2.7X and 79.0% (14.2X), 83.9% (4.7X), and 58.7% (1.7X) of the pPCP1, pCD1, and pMT1 plasmids, respectively (Figure S1). The lower coverage of the pMT1 is due to the absence of a 20-kb region containing the *ymt* gene, which was gained after the Bronze Age clade diverged from the main lineage (Figure S1) (Rasmussen et al., 2015). The reads showed the expected degradation signatures for ancient DNA (aDNA), such as, short read length distribution (Figure 1B) and deamination at the 5′ and 3′ ends of the sequences (Figure 1C). To confirm the identity of the strain (named Gok2 from now), we also re-mapped the reads against the closely related *Yersinia pseudotuberculosis* IP32953 reference genome and calculated the mapping affinity against both species, as done previously (Rasmussen et al., 2015). Reads showed to be clearly closer to *Y. pestis* (Figure 1D), which was also confirmed using a Naive Bayesian classifier that showed a posterior probability of 0.93 for the reads originating from *Y. pestis*.

To determine if other individuals from this passage grave or from the same geographic region showed any sign of *Y. pestis*, we carefully analyzed the reads mapping to the *Y. pestis* reference genome from individuals that had been also sequenced in previous works (Skoglund et al., 2012, 2014). The analysis revealed the potential existence of *Y. pestis* in another early farmer individual (Gökhem4, a 20-year-old male, 5,040–4,839 BP) from the TRB culture in the same megalithic burial structure and age (Figure 1A; Table S2). Here, 2,951 reads mapped to the chromosome and three plasmids. The low number of *Y. pestis* reads recovered is also consistent with the low coverage of the human genome in this sample (0.04X, versus the 1.33X of Gökhem2) (Skoglund et al., 2014), indicating a poor overall DNA conservation. To validate their identity, we obtained perfect or near perfect alignment versus non-redundant sequences from NCBI for 1,764 of the reads where 1,116 of these were exclusive to the *Y. pestis* chromosome and the three plasmids (Table S3). Additionally, the Naive Bayesian classifier classified the reads with a posterior probability of 1 for being *Y. pestis* (Figure 1D). These results confirmed that despite the poor conservation of the sample, there is also an unambiguous presence of *Y. pestis* in Gökhem4. We can rule out the possibility of bleed-over between lanes and sequencing indexes as there were multiple sequencing libraries from the two individuals, many of which were prepared and sequenced at different times (Skoglund et al., 2012, 2014).

Importantly, *Y. pestis* was not found in other nearby contemporary hunter-gatherers from the Pitted Ware Culture in Sweden (Skoglund et al., 2014). Altogether, our results support that an epidemic of plague could explain the size and short time frame of the Frälsegården passage grave. The emergence of epidemics in ancient farming villages, which had higher human and animal densities than those of forager groups, had been already proposed based on archaeological records found in different pre-historic settlements, but our findings represent a new and clearer evidence of such events (Armelagos and Harper, 2005; Barrett et al., 1998; Bocquet-Appel, 2011; Gage and Kosoy, 2005; Hershkovitz and Gopher, 2008; Mummert et al., 2011).

The Gok2 Strain Is Basal to All Known *Y. pestis*

To determine the phylogenetic positioning of the Gok2 *Y. pestis* strain, we reconstructed the phylogeny of this genome together with a collection of *Y. pseudotuberculosis* (n = 27), and *Y. pestis* (n = 183) genomes. Using high confidence positions (3.2 million base pairs [Mbp]) and maximum likelihood phylogenetic analysis, we found the Gok2 strain to be basal to all other *Y. pestis* (Figures 1E and S2). The branch was supported by 100% bootstrap, independently of using *Y. pestis* or *Y. pseudotuberculosis* as reference for genome reconstruction. Importantly, the Gok2 strain was clearly part of the *Y. pestis* clade and well separated from *Y. pseudotuberculosis*. Because of the relatively low depth of the genome and the possible confounding effect of DNA damage (i.e., C to T and G to A substitutions), we additionally reconstructed the phylogeny excluding transitions. Here, Gok2 was also clearly basal to the Bronze Age clade (Figure S2E). Moreover, the topology of the tree shows that Gok2 is not part of the same monophyletic clade that groups Bronze Age strains (Figure 1E), in which the most basal strain is RISE509, that was found in the Altai (Siberia) more than 4,500 km from Gok2 and carbon dated to 4,800–4,700 BP (i.e., slightly younger compared to Gok2).

Distinct Genetic Variation in the Gok2 and Bronze Age Clades

To further understand the genomic variation that differentiated the Gok2 and Bronze Age lineages, we performed a detailed analysis of the single nucleotide variants (SNVs) found in both groups (Figure 2). We first identified all SNVs found in Gok2, relative to the *Y. pseudotuberculosis* reference, and not fixed in all *Y. pestis* strains (i.e., identical in all genomes). This resulted in a total of 28 SNVs (11 missense, 14 synonymous, and 3 non-annotated), of which 10 were present in almost all *Y. pestis* strains, 3 were located at positions that were eventually deleted in most or all *Y. pestis* genomes, and 15 variants were exclusive to Gok2 compared to all other analyzed strains (Figure 2A). Among all 28 SNVs, only 2 were transversions (1 synonymous and 1 non-coding) and 8 were also observed in some *Y. pseudotuberculosis* strains, which may indicate that they likely did not emerge in *Y. pestis.* It is worth noting that the 15 unique SNVs from Gok2 (9 missense, 4 synonymous, and 2 non-coding) were supported by 4–8 independent reads, but were all transitions (C > T, G > A), which can potentially result from deamination due to ancient DNA degradation. However, variation predicted to be caused by DNA damage, which are normally observed in the 5′and 3′ends of DNA molecules, were filtered out from reads. All these variants were called from bases located at the center of the reads (Figure S3), strongly suggesting that these SNVs were not due to post-mortem degradation. Interestingly, four of the variants were located in genes with functions related to host-pathogen interactions such as siderophores, iron acquisition, Ton and Tol transport systems, and motility and chemotaxis (Figure 2A).

Complementary, we used the same strategy to identify SNVs associated with the Bronze Age clade. We identified 154 variants found in at least one of the strains in this clade. Except nine variants that were found in most *Y. pestis* strains, including Gok2, none of the mutations were found in neither Gok2 nor any other *Y. pestis* strain (Figure 2B). Among these lineage-specific

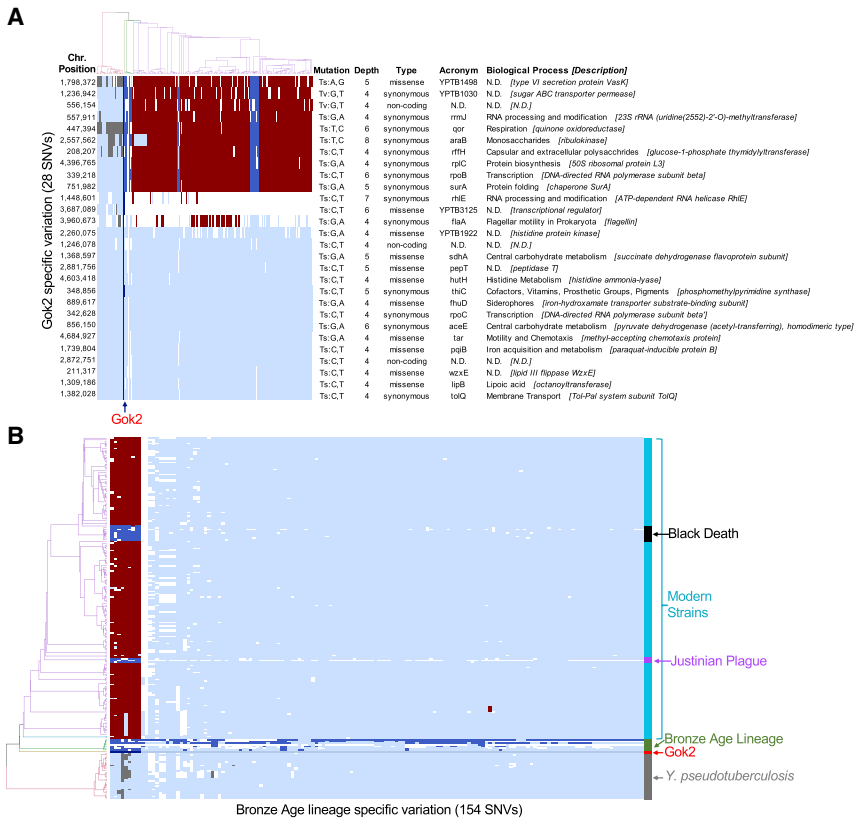


Figure 2. Distinct Genetic Variation between the Gok2 and Bronze Age *Y. pestis* Clades

(A) Single nucleotide variants (SNVs) found in the Gok2 strain compared to the *Y. pseudotuberculosis* IP32953 reference genome that were not identical in all *Y. pestis* strains (i.e., already fixed in a common ancestor). Detailed information about each variant is indicated on the right. Positions that are identical to the reference are colored in light blue, unmapped positions in white. The presence of each variant in *Y. pseudotuberculosis* strains is indicated in gray, for Gok2 in dark-blue, for ancient *Y. pestis* strains in blue and for all other *Y. pestis* in dark-red. The ML phylogenetic tree from Figure 1E was used to sort genomes in the heatmap and is shown at the top of the figure. Ts, transitions; Tv, transversions; N.D, not determined. The results show that 15 of the 28 SNVs identified in Gok2 were unique to this strain.

(B) Same analysis but showing all variants found in the Bronze Age strains that are not identical in all *Y. pestis* strains. Most of the variants were only found in the Bronze Age clade and not in the Gok2 strain (highlighted with a darker light blue) or any other strain. Color bar on the right identifies strains as follows: grey, 27 *Y. pseudotuberculosis* strains; red, Gok2; green, 7 Bronze Age strains; purple, 3 Justinian plague strains; black, 9 strains from or related to the Black Death; cyan, all strains isolated from modern individuals. Samples are placed in the same order as in (A) using also the phylogenetic tree from Figure 1E.

Altogether, (A) and (B) show that the Gok2 and Bronze Age strains belonged to two distinct and independent *Y. pestis* clades, with genetic variation that is no longer represented in known modern lineages.

See also Figure S3 and Table S4.

mutations, 44% were transversions and 54% were found in more than one strain, indicating that these are truly specific of the Bronze Age lineage.

Altogether, our results indicate not only that Gok2 and the other Bronze Age strains belonged to two independent lineages, but also both lineages are extinct and no longer represented among the modern strains of *Y. pestis*.

Large-Scale Radiation of *Y. pestis* during the Neolithic Decline

To explore the history of early *Y. pestis* strains, we performed a molecular clock analysis to estimate the divergence times between lineages. Our results indicate that Gok2 diverged from all other strains 5,700 years BP (95% HPD interval: 5,250–6,364 BP), the Bronze Age clade 5,300 years BP (95% HPD interval: 4,953–5,731 BP), and the most basal of all known modern clades (most found in China) at 5,100 BP (95% HPD interval: 4,678–5,625 BP) (Figures 3 and S4). Previous genomic and phylogenetic analyses of hundreds of modern and ancient *Y. pestis* genomes concluded that most modern clades likely originated in East Asia, and more precisely in China (Morelli et al., 2010). A recent work has additionally proposed that three basal lineages (from the 0.PE4 clade), two of which persist today, may have split from the main lineage at 4,000 BP, and branched at the Eurasian steppe (Spyrou et al., 2018). Interestingly, our results indicate that before all these events, a large-scale branching and geographic radiation of *Y. pestis* occurred between 6,000 and 5,000 years ago, during the period known as the Neolithic decline, and shortly before the massive migrations from the Eurasian Steppe into Europe (Allentoft et al., 2015; Haak et al., 2015).

Bronze Age and Gok2 Lineages Did Not Initially Spread with Massive Human Migrations

We used 1,058 ancient human genomes from previously published studies (Table S2) (Allentoft et al., 2015; Lazaridis et al., 2016; Lipson et al., 2017; Mathieson et al., 2015, 2018; Olalde et al., 2018) to investigate whether human migrations may have facilitated the expansion of *Y. pestis* in the Neolithic and Bronze age across Eurasia. We modeled the ancestry of all these individuals as a mixture of five hypothetical populations (ADMIXTURE analysis with K = 5), which consistently recreated previous descriptions about the genetic landscape and history of the Eurasian populations (Figures 4A and S5) (Allentoft et al., 2015; Haak et al., 2005, 2015; Lazaridis et al., 2016; Lipson et al., 2017; Skoglund et al., 2014). In agreement with these works, the admixture analysis showed clear differences between the populations in the Eurasian Steppe that derived from admixture events between

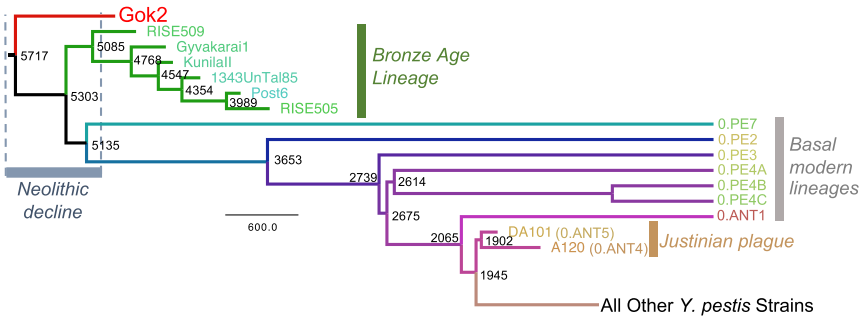


Figure 3. Molecular Clock Estimation of Ancient *Y. pestis* Strains Reveals the Radiation of Multiple Independent Lineages during the Neolithic Decline

BEAST2 maximum clade credibility tree from 1,201 million states and 221k trees showing median divergence dates at tree nodes. Dates are given as years before present (BP). Only lineages diverging until the Justinian plague are shown (full version shown in Figure S4). The results indicate that at least four deep branching and independent lineages of *Y. pestis* (the extinct lineages of Gok2 and Bronze Age, the modern 0.PE7 and the root to all other known *Y. pestis* strains) diverged between 5,700–5,000 BP, a period matching with the decline of Neolithic populations in Europe.
See also Figure S4 and Table S4.

eastern hunter-gatherers (EHG) and populations from Iran/Caucasus. Populations from Eastern and Central Europe, instead, derived from Anatolian populations with some degree of admixture with western hunter-gatherers (WHG). The results also show the major genetic turnover of the European populations after the Neolithic decline in Europe (6,000–5,000 BP), produced by massive human migrations from the Steppe into Eastern and Central Europe around 4,800–4,600 BP (Allentoft et al., 2015; Haak et al., 2015). Another observation was the genetic input from Europe into the Eurasian Steppe after 4,000 BP.

We then selected only genomes from human remains older than 3,000 BP found near the same geographic regions where *Y. pestis* was detected (120 genomes in total) and organized them into 6 different groups (Figure 4B). These groups were then subdivided when clear distinct populations from different historical periods or archaeological contexts were evidenced e.g., the Gökhem individuals from the TRB culture (group 6B) were clearly different from the preceding hunter-gatherers (group 6A) and the posterior Corded Ware Cultures (group 6C). Importantly, these results showed that the genetic backgrounds of the infected individuals were similar to other individuals within their own regions, even long time before and after the infections happened, with no signs of admixture between the populations of the different infection foci. This is particularly evident for the population of the Gökhem 2 individual (group 6B), which genetically, was significantly different to the individuals from the later Corded Ware Cultures (group 6C) (two-tailed t test, p value = 9×10^{-5}). Similarly, the individuals infected with plague in the Altai region (RISE509 and RISE511, group 1A) was found to be significantly different from every other population where the pathogen was found except compared to group 4 (two-tailed t test, p value = 0.02) (Figure 4C).

Altogether, our results indicate that the initial and large-scale radiation of plague across Eurasia during the Neolithic (Figure 3) was not contemporary to any major known human migration (Figure 5A) and that the genetic background of the infected populations did not significantly change during this period.

A Model of Plague Dispersion Associated with the Neolithic Decline

We combined the estimated divergence times and phylogenetic relationships of the *Y. pestis* strains (Figures 1E and 3) with the genetic (Figure 4), migratory (Figure 5A), and archaeological (Figures 5B and 5C) histories from ancient human populations in Eurasia to build a new model of the dispersion of *Y. pestis* during the Neolithic and Bronze Age (Figure 5D). The chronology and pieces of information provided by each of these layers consistently converged in supporting the existence of an emergence foci of *Y. pestis* in mega-settlements of Eastern Europe (Figure 5). Living conditions in these sites were unprecedented compared to previous human populations (i.e., highest population densities with close contact with animals), and our estimates place the plague radiation when the collapse of these societies had begun. These settlements built by the Trypillia Culture between 6,100–5,400 BP are good candidates, as they featured the largest known Neolithic settlements in Europe with a chronology that fits the estimated divergence times of *Y. pestis* (Figure 3). By the time of the initial plague radiation (Gökhem strains), Trypillia populations were likely under nutritional stress and weakened due to environmental overexploitation (Kirleis and Dal Corso, 2016). Stressed populations living under highly dense conglomerations are precisely the conditions that favored other plague epidemics and pandemics, such as the Black Death (DeWitte, 2018). Therefore, we propose the Trypillia mega-settlements as the best candidate for the emergence of the ancestor(s) of the most basal lineages of plague, which rapidly branched and migrated, giving origin to the Gok2, Bronze Age, and other basal lineages (e.g., 0.PE7 and 0.PE2).

The mid to late 6th millennium BP was also characterized by a rapid expansion of new technological innovations in metallurgy and communication. In particular, the animal traction complex, involving cattle traction, wagons, and ard ploughs, would have provided favorable and unprecedented conditions for a rapid expansion of infectious diseases over large geographic regions. This complex spread very rapidly over a vast area including most of Europe and the western parts of Asia in the period ~5,500–5,000 BP (Figure 5B). The origin of the complex is debated, but in any case, it was soon in place over a large region and apparently not spread through any large-scale migrations (Bakker et al., 1999; Hansen, 2011; Klimscha, 2017; Sherratt, 1997). This new package of innovations, for instance, led to the beginning of exploitation of the steppe by pastoral herders, and the traction complex was carried by the Baden Culture into East and Central Europe (Horvath and Svingor, 2015). These

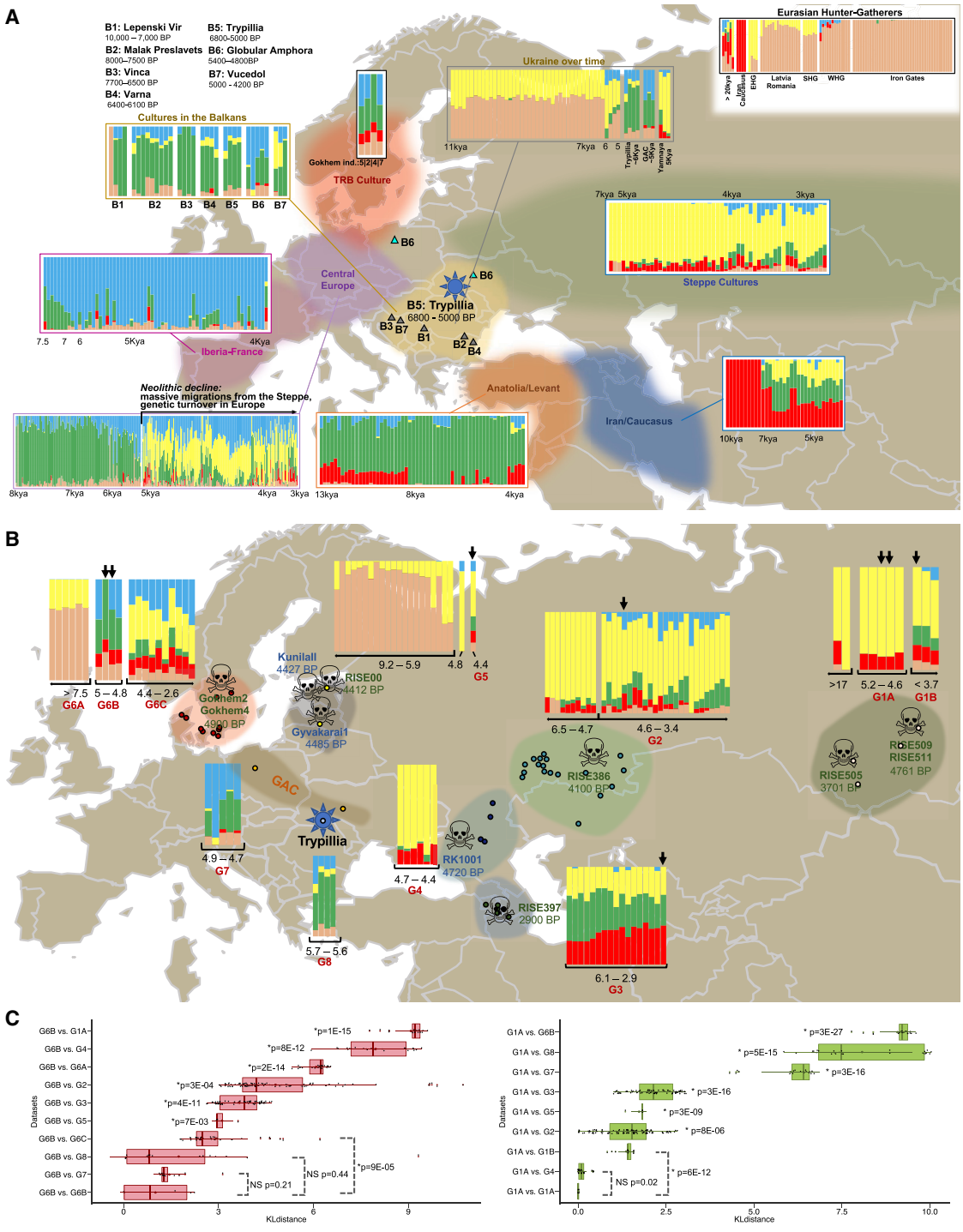


Figure 4. Ancient Human Genomes Indicate that Massive Migrations Were Not Responsible for *Y. pestis* Radiation during the Neolithic Decline

A total of 1,058 ancient human genomes were re-analyzed using an ADMIXTURE analysis (K = 5) to evaluate if the Neolithic dispersion of *Y. pestis* could be explained by human migrations or admixture events between the infected populations.

(A) Predicted ancestry of the Eurasian and Near East populations showing an overview of their genetic makeup during the Neolithic and Bronze Age also reflecting the major genetic turnovers due to massive migrations. For example, there was a drastic genetic turnover in the populations from Ukraine around 6 kya. Similarly, massive migrations of the Eurasian Steppe populations significantly changed the genetic background of Central Europe populations after the Neolithic decline, around 4.8 kya, which was later followed by a new migration from Central Europe back to the Eurasian Steppe at 4 kya.

(legend continued on next page)

innovations also favored frequent contacts between the Trypillia culture and those in the Eurasian Steppe, principally through trade networks (Chapman, 2013). Interestingly, our analyses on ancient human genomes (Figures 4A and 4B) also support that during the Neolithic, emerging trade networks (Figure 5C) were more plausible dispersion routes of plague into the Eurasian Steppe than massive movements of people. This spread would have given rise to the Bronze Age *Y. pestis* lineage that subsequently expanded and diversified in central Eurasia. Yamnaya populations infected with these strains would have spread plague both eastward to Siberia (Afanasievo Culture) and westward into Europe (Andrades Valtuena et al., 2017; Rasmussen et al., 2015). This second introgression of plague in Europe corresponds chronologically to the period after the Neolithic collapse of most of northern and western European populations (Hinz et al., 2012; Shennan et al., 2013), which would have paved the way for the massive migration of the Yamnaya and Corded Ware Cultures into temperate Europe after 4,800 BP (Figure 5A) and brought *Y. pestis* strains from the Bronze Age lineage.

By the end of the 6th millennium BP, the Globular Amphora Culture (GAC) started forming at the north of the Carpathians (Woidich, 2014), which would soon expand to encompass a region from the forest steppe of the former Trypillia settlements to Poland and Germany. There are also signs of GAC influence in the TRB culture (6,000–4,800 BP) of southwestern and southern Denmark (Ebbesen, 1977; Johannsen and Laursen, 2010), which was in turn related to the culture of the Gökhem 2 and Gökhem 4 individuals in Sweden (Sjögren, 2015). The results from human genome analyses also showed that the Gökhem2 population (G6B) was not significantly different from the Globular Amphora (G7) and Trypillia (G8) Cultures (two-tailed t test, p value = 0.21 and 0.44) (Figure 4C), which is likely indicative of a genetic continuity and a connection that could potentially explain the dispersion route of the pathogen. Moreover, although the TRB settlements in Scandinavia were rather small, they also had large ceremonial gathering places (causewayed camps), which were also characteristic in other cultures in Central and Western Europe and would have been instrumental in creating social bonds and favoring frequent movements of people, sometimes over large distances (Meller and Friederich, 2017). These places could also have favored the spread of diseases between distant regions. Altogether, the archaeological record also indicates that there was contact, although indirect, between the Trypillia Culture and the TRB populations infected with *Y. pestis* in Sweden. Therefore, we propose that the *Y. pestis* found in the TRB populations of Sweden derived from a lineage that originally emerged in the Trypillia Culture, which spread and diverged with the Globular Amphora Cultures.

Although we do not have experimental evidence of the pathogenicity of early *Y. pestis*, its presence in ancient DNA obtained from human teeth is a strong sign of a high titer in the bloodstream at death. This most likely indicates that it was deadly. At the genomic level, these strains contain the plasminogen activator gene that is sufficient to cause pneumonic plague—the deadliest form of historic and modern plague (Zimbler et al., 2015). Furthermore, the RISE509 and RISE511 individuals from Afanasievo Gora were found in a mass grave containing remains of multiple individuals and likely produced by an epidemic (Rasmussen et al., 2015), while the Gökhem 2 and Gökhem 4 individuals from Frälsegården were found in a Neolithic collective grave used for an unusually short period of time (Ahlström, 2009; Sjögren, 2015). Finally, we and others have frequently identified plague in individuals across Eurasia during the Neolithic and Bronze Age (Rasmussen et al., 2015; Andrades Valtuena et al., 2017; Spyrou et al., 2018), suggesting it was fairly prevalent and virulent. We therefore hypothesize that these early forms of plague may have contributed to the decline of Neolithic Cultures in Europe through the emergence and expansion of a prehistoric plague pandemic.

In this work, we report the discovery of plague infecting Neolithic farmers in Scandinavia, which not only pre-dates all known cases of plague, but is also basal to all known modern and ancient strains of *Y. pestis*. We identified a remarkable overlap between the estimated radiation times of early lineages of *Y. pestis*, toward Europe and the Eurasian Steppe, and the collapse of Trypillia mega-settlements in the Balkans/Eastern Europe. Our results are consistent with the existence of a prehistoric plague pandemic, spreading mainly through early trade networks, rather than massive human migrations, which allowed a rapid and large-scale expansion of the pathogen, that persisted through the Bronze Age with lineages that got eventually extinct. We propose that plague may have contributed to the Neolithic decline, which paved the way for the later steppe migrations into Europe.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

(B) Individuals infected with *Y. pestis* were divided into six different geographic regions (G1–G6, written in red) and the ADMIXTURE proportions of all available genomes (n = 120) within each region is indicated with their corresponding time periods below (in kyr). Subgroups (e.g., G1A-G1B, G6A-G6B-G6C) were defined when discontinuous ancestries/time-periods of individuals were identified. Skulls indicate the excavation sites of the infected individuals, black arrows indicate their ADMIXTURE proportions, and their carbon dates are written in green when the human genomes were available or in blue when they were not. Groups G7 (Globular Amphora Culture) and G8 (Trypillia Culture) are defined to support the model presented in Figure 5.

(C) Boxplots showing the symmetric Kullback-Leibler (KL) divergence between the ADMIXTURE components of all individuals from groups G6B (Gökhem individuals, left) or G1A (RISE509/RISE511, right) with each other (bottom boxes) or versus the individuals from other groups (all other boxes). Lower and upper hinges correspond to the first and third quartiles and median is indicated with a red perpendicular line. Upper and lower whisker extends up to 1.5 * IQR, with a red dot indicating the mean of all values. A two-sided t test was used as significance test (p values are indicated; NS, not significant; asterisk, significant). These results indicate that the ancestry of the populations infected with the most basal, and thus, initially spread *Y. pestis* strains, were significantly different from all other infected populations.

See also Figure S5 and Table S2.

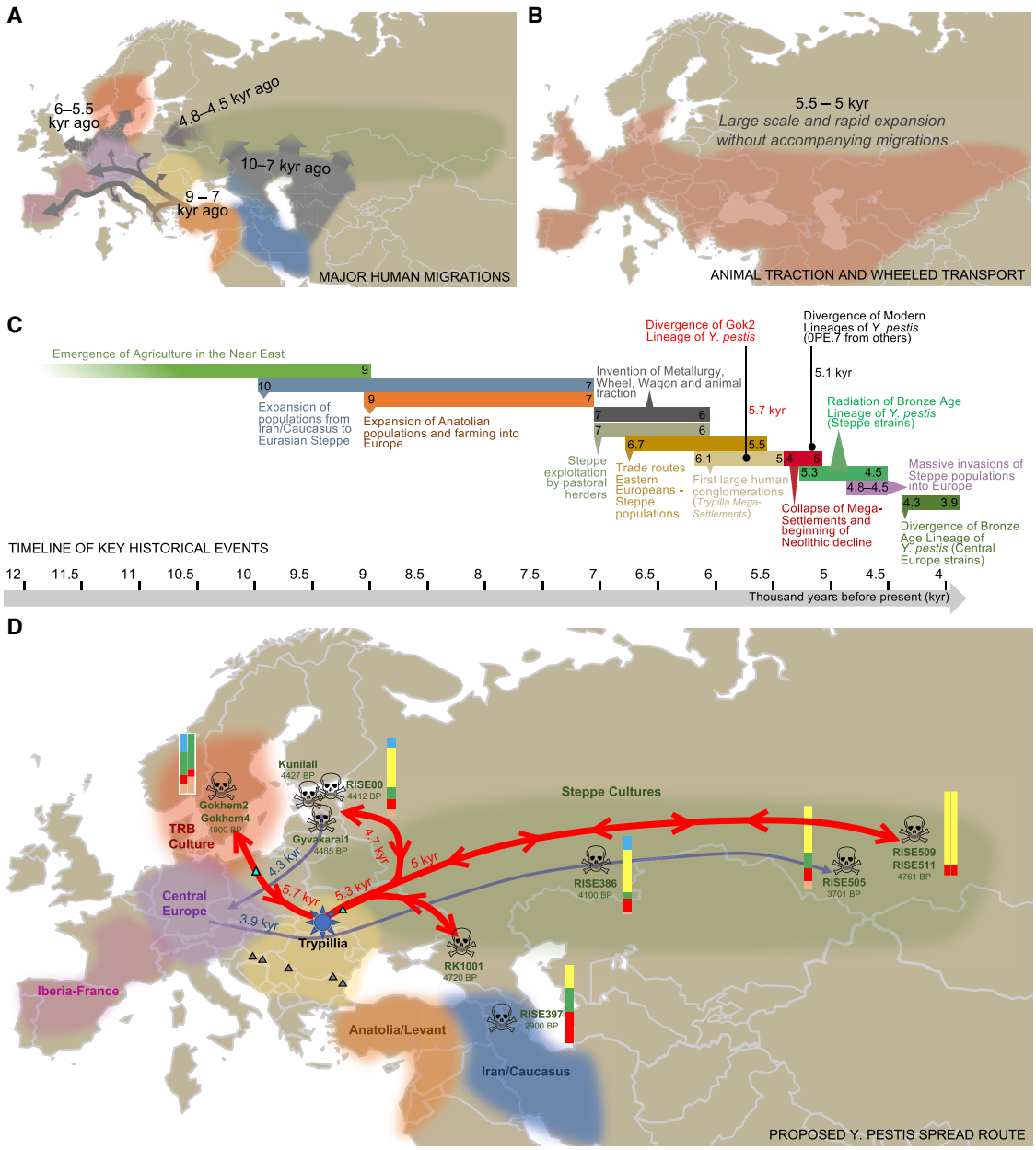


Figure 5. Proposed Dispersion Model of *Y. pestis* during Neolithic and Bronze Age

(A) Schematic representation of the trajectories and time periods (thousand years before present, kyr) of major known human migrations in Eurasia during the Neolithic and Bronze Age. The observed geographic distribution and divergence times of *Y. pestis* strains from the Gok2 and Bronze Age clades cannot be explained by the timings and routes of these human movements.

(B) Geographic distribution of the use of animal traction and wheeled transport across Neolithic and Bronze Age populations in Eurasia, which broadly expanded during the period of 5,500 and 5,000 BP. The expansion of these technological innovations overlaps the predicted period for the expansion of the basal *Y. pestis* strains.

(C) Timeline indicating the proposed key historical events that contributed to the emergence and spread of plague during the Neolithic.

(D) The predicted model of early dispersion of *Y. pestis* during Neolithic and Bronze Age was built by integrating phylogenetic information of *Y. pestis* strains from this period (Figure 1E), their divergence times (Figure 3), the geographic locations, carbon dating and genotypes of the individuals, and the archaeological record. The model suggests that early *Y. pestis* strains likely emerged and spread from mega-settlements in Eastern Europe (built by the Trypillia Culture) into Europe and the Eurasian steppe, most likely through human interaction networks. This was facilitated by wheeled and animal-powered transports, which are schematized in the map with red lines with arrows pointing in both senses. Our model builds upon a previous model (Andrades Valtuena et al., 2017) that proposed the spread of plague to be associated with large-scale human migrations (blue line).

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and four tables and can be found with this article online at https://doi.org/10.1016/j.cell.2018.11.005.

ACKNOWLEDGMENTS

We would like to thank the authors and laboratories of all previous works that shared the sequencing datasets used in the present work. We would also like to thank the Reich Lab, which made the genotypes of their ancient genomes publicly available. A special thanks to Reini Luco for her insightful feedbacks. S.R. was supported by the Carlsberg Foundation (CF16-0418) and the Novo Nordisk Foundation (NNF14CC0001). E.W. was supported by The Lundbeck Foundation, The Danish National Research Foundation, and KU2016. N.R. and C.D. were supported by the IHU Méditerranée Infection, Marseille, France, by the French Government under the "Investissements d'avenir" program (Méditerranée Infection 10-IAHU-03), by the Région Provence Alpes-Côte d'Azur, by the European funding FEDER PRIMI, and by a "Management des talents" fellowship from the A*MIDEX Foundation awarded to C.D.

AUTHOR CONTRIBUTIONS

N.R. made the screening and discovered the new *Y. pestis* strains. N.R., C.D., and S.R. conceived the study. N.R. and S.R. conducted the bioinformatic analyses. R.N. provided input to the phylogenetic and human statistical analyses. K.K., K.-G.S., and E.W. performed the archaeological contextualization. N.R. and S.R. wrote the manuscript with contributions from all co-authors. All authors read and approved the last version of the manuscript. S.R. and N.R. are responsible for all data, figures, and text, ensured that authorships are properly granted with all authors consent, and guaranteed that the manuscript adhered to all of the journal's policies.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 9, 2018
Revised: October 10, 2018
Accepted: November 1, 2018
Published: December 6, 2018

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Bos et al. (2011), (2016); Chain et al. (2004), (2006); Cui et al. (2013); Damgaard et al. (2018); Deng et al. (2002); Eppinger et al. (2009), (2010); Eroshenko et al. (2017); Feldman et al. (2016); Garcia et al. (2007); Kislichkina et al. (2015); Parkhill et al. (2001); Rajanna et al. (2013); Reuter et al. (2014); Song et al. (2014); Spyrou et al. (2016); Wagner et al. (2014); Zhang et al. (2009).

REFERENCES

Ahlström, T. (2009). Underjordiska dödsriken - humanosteologiska undersökningar av neolitiska kollektivgravar. Coast to Coast Books no 18 (Department of Archaeology, Göteborg University).

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

Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

Andrades Valtuena, A., Mittnik, A., Key, F.M., Haak, W., Allmae, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Jankovic, I., et al. (2017). The Stone Age plague and its persistence in Eurasia. Curr. Biol. *27*, 3683–3691.

Anthony, D. (2007). The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World (Princeton University Press).

Armelagos, G.J., and Harper, K.N. (2005). Genomics at the origins of agriculture, part two. Evol. Anthropol. *14*, 109–121.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., et al. (2012). BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Syst. Biol. *61*, 170–173.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., et al. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics *9*, 75.

Bakker, J.A., Kruk, J., Lanting, A.E., and Milisauskas, S. (1999). The earliest evidence of wheeled vehicles in Europe and the Near East. Antiquity *73*, 778–790.

Barrett, R., Kuzawa, C.W., McDade, T., and Armelagos, G.J. (1998). Emerging and re-emerging infectious diseases: the third epidemiologic transition. Annu. Rev. Anthropol. *27*, 247–271.

Bocquet-Appel, J.P. (2011). When the world's population took off: the springboard of the Neolithic Demographic Transition. Science *333*, 560–561.

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

Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M., et al. (2016). Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. eLife *5*, e12994.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Chain, P.S., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., et al. (2004). Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. USA *101*, 13826–13831.

Chain, P.S., Hu, P., Malfatti, S.A., Radnedge, L., Larimer, F., Vergez, L.M., Worsham, P., Chu, M.C., and Andersen, G.L. (2006). Complete genome sequence of Yersinia pestis strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. J. Bacteriol. *188*, 4453–4463.

Chapman, J. (2013). In Expansion and social change at the time of Varna. Counterpoint: Essays in Archaeology and Heritage Studies in honour of Professor Kristian Kristiansen, S. Bergerbrant and S. Sabatini, eds. (Oxford University Press), pp. 301–308.

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the

effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

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

Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369–374.

Deng, W., Burland, V., Plunkett, G., 3rd, Boutin, A., Mayhew, G.F., Liss, P., Perna, N.T., Rose, D.J., Mau, B., Zhou, S., et al. (2002). Genome sequence of Yersinia pestis KIM. J. Bacteriol. *184*, 4601–4611.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

DeWitte, S.N. (2018). Stress, sex, and plague: Patterns of developmental stress and survival in pre- and post-Black Death London. Am. J. Hum. Biol. *30*, e23073.

Diachenko, A. (2016). Small is beautiful: a democratic perspective? In Trypillia Mega-Sites and European Prehistory, 4100-3400 BCE, J. Müller, K. Rassmann, and M. Videiko, eds. (Routledge), p. 12.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930.

Downey, S.S., Haas, W.R., Jr., and Shennan, S.J. (2016). European Neolithic societies showed early warning signals of population collapse. Proc. Natl. Acad. Sci. USA *113*, 9751–9756.

Ebbesen, K. (1977). Die Jungere. Trichterbecherkultur auf den Dänischen Inseln*Volume II* (Akademisk Förlag).

Eppinger, M., Guo, Z., Sebastian, Y., Song, Y., Lindler, L.E., Yang, R., and Ravel, J. (2009). Draft genome sequences of Yersinia pestis isolates from natural foci of endemic plague in China. J. Bacteriol. *191*, 7628–7629.

Eppinger, M., Worsham, P.L., Nikolich, M.P., Riley, D.R., Sebastian, Y., Mou, S., Achtman, M., Lindler, L.E., and Ravel, J. (2010). Genome sequence of the deep-rooted Yersinia pestis strain Angola reveals new insights into the evolution and pangenome of the plague bacterium. J. Bacteriol. *192*, 1685–1699.

Eroshenko, G.A., Nosov, N.Y., Krasnov, Y.M., Oglodin, Y.G., Kukleva, L.M., Guseva, N.P., Kuznetsov, A.A., Abdikarimov, S.T., Dzhaparova, A.K., and Kutyrev, V.V. (2017). Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. PLoS ONE *12*, e0187230.

Feldman, M., Harbeck, M., Keller, M., Spyrou, M.A., Rott, A., Trautmann, B., Scholz, H.C., Päffgen, B., Peters, J., McCormick, M., et al. (2016). A high-coverage yersinia pestis genome from a Sixth-Century Justinianic plague victim. Mol. Biol. Evol. *33*, 2911–2923.

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

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics *31*, 3718–3720.

Garcia, E., Worsham, P., Bearden, S., Malfatti, S., Lang, D., Larimer, F., Lindler, L., and Chain, P. (2007). Pestoides F, an atypical Yersinia pestis strain from the former Soviet Union. Adv. Exp. Med. Biol. *603*, 17–22.

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

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

Hansen, S. (2011). Technische und soziale Innovationen in der zweiten hälfte des 4. Jahrtausend v. Chr. Sozialarchäologische Persepktiven: Gesellschaftlicher Wandel 5000–1500 v Chr Zwischen Atlantik und Kaukasus, S. Hansen and J. Müller, eds. (Philipp von Zabern), pp. 153–191.

Hershkovitz, I., and Gopher, A. (2008). Demographic, biological and cultural aspects of the Neolithic revolution: a view from the Southern Levant. In The Neolithic Demographic Transition and its Consequences, J.-P. Bocquet-Appel and O. Bar-Yosef, eds. (Springer Netherlands), pp. 441–479.

Hinz, M., Feeser, I., Sjögren, K.-G., and Müller, J. (2012). Demography and the intensity of cultural activities: an evaluation of Funnel Beaker Societies (4200–2800 cal BC). J. Archaeol. Sci. *39*, 3331–3340.

Horvath, T., and Svingor, E. (2015). The spatial and chronological distribution of the so-called "Baden Culture". In The Baden Culture Around the Western Carpathians, M. Nowak and A. Zastawny, eds. (Krakowski Zespół do Badań Autostrad), pp. 19–75.

Johannsen, N., and Laursen, S. (2010). Routes and wheeled transport in late 4th–early 3rd millennium funerary customs of the Jutland Peninsula: regional evidence and European context. Praehist. Z. *85*, 15–58.

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

Kirleis, W., and Dal Corso, M. (2016). Trypillian subsistence economy: animal and plant exploitation. In Trypillia Mega-sites and European Prehistory, 4100-3400 BCE, J. Müller, K. Rassmann, and M. Videiko, eds. (Routledge), p. 12.

Kirleis, W., and Dreibrodt, S. (2016). The natural background: forest, forest steppe or steppe environment. In Trypillia Mega-Sites and European Prehistory, 4100-3400 BCE, J. Müller, K. Rassmann, and M. Videiko, eds. (Routledge), p. 10.

Kislichkina, A.A., Bogun, A.G., Kadnikova, L.A., Maiskaya, N.V., Platonov, M.E., Anisimov, N.V., Galkina, E.V., Dentovskaya, S.V., and Anisimov, A.P. (2015). Nineteen whole-genome assemblies of Yersinia pestis subsp. microtus, including representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica. Genome Announc. *3*, e01342-15.

Klimscha, F. (2017). Transforming technical know-how in time and space. Using the Digital Atlas of Innovation to understand the innovation process of animal traction and the wheel. eTopoi J. Ancient Studies *6*, 16–63.

Korvin-Piotrovskiy, A., Hofmann, R., Rassmann, K., and Videiko, M. (2016). Pottery kilns in Trypillian Settlements. Tracing the division of labour and the social organization of Copper Age communities and Lennart Brandtstätter. In Trypillia Mega-sites and European Prehistory, 4100-3400 BCE, J. Müller, K. Rassmann, and M. Videiko, eds. (Routledge), p. 32.

Kristiansen, K. (2014). The decline of the Neolithic and the rise of Bronze Age Society. In The Oxford Handbook of Neolithic Europe, C. Fowler, J. Harding, and D. Hofmann, eds. (Oxford University Press).

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

Meller, H., and Friederich, S. (2017). Salzmünde – rule or exception? (Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Landesmuseum für Vorgeschichte).

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

Müller, J. (2013). Demographic Traces of Technological Innovation, Social Change and Mobility: from 1 to 8 Million Europeans (6000-2000 BCE) (Institute of Archaeology Rzeszów University).

Müller, J., Rassmann, K., and Videiko, M. (2016). Trypillia Mega-Sites and European Prehistory 4100–3400 BCE (Routledge).

Mummert, A., Esche, E., Robinson, J., and Armelagos, G.J. (2011). Stature and robusticity during the agricultural transition: evidence from the bioarchaeological record. Econ. Hum. Biol. *9*, 284–301.

Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., et al. (2001). Genome sequence of Yersinia pestis, the causative agent of plague. Nature *413*, 523–527.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

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

R Development Core Team (2008). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

Rajanna, C., Ouellette, G., Rashid, M., Zemla, A., Karavis, M., Zhou, C., Revazishvili, T., Redmond, B., McNew, L., Bakanidze, L., et al. (2013). A strain of Yersinia pestis with a mutator phenotype from the Republic of Georgia. FEMS Microbiol. Lett. *343*, 113–120.

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

Reimer, P., Hoper, S., McDonald, J., Reimer, R., Svyatko, S., and Thompson, M. (2015). Laboratory protocols used for AMS radiocarbon dating at the 14chrono centre. Research Report Series 5-2015 (B. The Queen's University, ed. (English Heritage)).

Reuter, S., Connor, T.R., Barquist, L., Walker, D., Feltwell, T., Harris, S.R., Fookes, M., Hall, M.E., Petty, N.K., Fuchs, T.M., et al. (2014). Parallel independent evolution of pathogenicity within the genus Yersinia. Proc. Natl. Acad. Sci. USA *111*, 6768–6773.

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

Shennan, S., Downey, S.S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., and Thomas, M.G. (2013). Regional population collapse followed initial agriculture booms in mid-Holocene Europe. Nat. Commun. *4*, 2486.

Sherratt, A. (1997). Economy and Society in Prehistoric Europe. Changing Perspectives (Edinburgh University Press).

Sjögren, K.G. (2015). News from Frälsegården. Aspects on Neolithic burial practices. In Neolithic Diversities Perspectives from a Conference in Lund, Sweden, K. Brink, S. Hydén, K. Jennbert, L. Larsson, and D. Olausson, eds. (Department of Archaeology and Ancient History, Lund University), pp. 200–210.

Sjögren, K.G. (2017). Modeling middle Neolithic funnel beaker diet on Falbygden, Sweden. J. Archaeol. Sci. *12*, 295–306.

Sjögren, K.G., Price, T.D., and Ahlström, T. (2009). Megaliths and mobility in south-western Sweden. Investigating relationships between a local society and its neighbours using strontium isotopes. J. Anthropol. Archaeol. *28*, 85–101.

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science *336*, 466–469.

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.G., et al. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science *344*, 747–750.

Song, Y., Tong, Z., Wang, J., Wang, L., Guo, Z., Han, Y., Zhang, J., Pei, D., Zhou, D., Qin, H., et al. (2004). Complete genome sequence of Yersinia pestis strain 91001, an isolate avirulent to humans. DNA Res. *11*, 179–197.

Spyrou, M.A., Tukhbatova, R.I., Feldman, M., Drath, J., Kacki, S., Beltrán de Heredia, J., Arnold, S., Sitdikov, A.G., Castex, D., Wahl, J., et al. (2016). Historical Y. pestis genomes reveal the European Black Death as the source of ancient and modern plague pandemics. Cell Host Microbe *19*, 874–881.

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

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

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

Woidich, M. (2014). The Western Globular Amphora Culture. A new model for its emergence and expansion. eTopoi J. Ancient Studies *3*, 67–85.

Wright, E. (2016). Using DECIPHER v2.0 to analyze big biological sequence data in R. R J *8*, 352–359.

Zhang, Z., Hai, R., Song, Z., Xia, L., Liang, Y., Cai, H., Liang, Y., Shen, X., Zhang, E., Xu, J., et al. (2009). Spatial variation of Yersinia pestis from Yunnan Province of China. Am. J. Trop. Med. Hyg. *81*, 714–717.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
163 modern *Y. pestis* strains	See Table S4	See Table S4
20 ancient *Y. pestis* strains	See Table S4	See Table S4
26 modern *Y. pseudotuberculosis* strains	See Table S4	See Table S4
276 Genotypes from ancient Human Genomes	Lazaridis et al., 2016	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/NearEastPublic.tar.gz
120 Genotypes from ancient Human Genomes	Lipson et al., 2017	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/LipsonEtAl2017.tar.gz
219 Genotypes from ancient Human Genomes	Mathieson et al., 2018	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/Genomic_Hist_SE_Europe_Mathieson.tar.gz
443 Genotypes from ancient Human Genomes	Olalde et al., 2018	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/olalde_et_al.tar.gz
Software and Algorithms		
AdapterRemoval (v2.1.3)	Schubert et al., 2016	https://github.com/MikkelSchubert/adapterremoval
BLAST (v.2.5.0)	Camacho et al., 2009	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
BWA (v.0.7.10)	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
Bedtools (v.2.23.0)	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
Samtools (v.1.3.1)	Li et al., 2009	http://samtools.sourceforge.net/
picard-tools (v.2.9.3)	N/A	https://broadinstitute.github.io/picard/
GATK (v.3.3.0)	DePristo et al., 2011	https://software.broadinstitute.org/gatk/
mapDamage (v.2.0.6)	Jónsson et al., 2013	https://ginolhac.github.io/mapDamage/
RAxML (v.8.1.15)	Stamatakis, 2014	https://github.com/stamatak/standard-RAxML
BEAST2 (v.2.4.4)	Bouckaert et al., 2014	http://www.beast2.org/
BEAGLE library (v2.1.2)	Ayres et al., 2012	https://github.com/beagle-dev/beagle-lib
SNPeff (v.4.2)	Cingolani et al., 2012	http://snpeff.sourceforge.net/
R (v3.3.3)	R Development Core Team, 2008	https://www.r-project.org/
gplots package (v3.0.1)	N/A	https://cran.r-project.org/web/packages/gplots/index.html
DECIPHER (v2.2.0)	Wright, 2016	https://bioconductor.org/packages/release/bioc/html/DECIPHER.html
dendextend (v1.5.2)	Galili, 2015	https://www.rdocumentation.org/packages/dendextend/versions/1.8.0
Vegan (v2.4.4)	Dixon, 2003	https://cran.r-project.org/web/packages/vegan/index.html
RAST server	Aziz et al., 2008	http://rast.nmpdr.org/rast.cgi
PLINK (v1.90b5.3)	Purcell et al., 2007	https://www.cog-genomics.org/plink2
ADMIXTURE (v1.3.0)	Alexander et al., 2009	http://software.genetics.ucla.edu/admixture/download.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Simon Rasmussen (simon.rasmussen@cpr.ku.dk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological material and data

The passage grave at Frälsegården in Gökhem parish, Falbygden, western Sweden, was excavated in 1999-2001 (Ahlström, 2009; Sjögren, 2015). In spite of damage and ploughing, this constitutes the most well-documented bone material from a Scandinavian

megalithic tomb. The tomb was found to have been rectangular, approximately 9.1 × 1.8 m large, with a roughly 10 m long passage, and constructed of limestone slabs. Traces of dry walling of slate slabs were found in several places along the walls. Within the chamber a number of sections partitioning the chamber were found. The human bones have been subject to a number of analyses, such as osteology, 14C dating, stable isotope analyses, and strontium isotope analysis (Ahlström, 2009; Sjögren, 2017; Sjögren et al., 2009)

The osteological analysis indicates that a minimum number (MNI) of 51 individuals were deposited in the chamber (K.-G.S. and T. Ahlström, unpublished data). As this is fragmented and partly destroyed material, this is most certainly an underestimation. Based on the number of paired and unpaired talus bones, the most likely number of individuals (MLNI) has been calculated, arriving at 78 buried people. At present, 34 direct dates on human bone are available. The dates are tightly clustered in the period c. 3,100 – 2,900 cal BC. The period of use is thus quite short, compared to other megalithic tombs in Scandinavia. The presence of a number of whole or partially articulated skeletons was one of the most significant results of the excavation. These range from almost complete skeletons to partial articulations. In addition, there is a mass of disarticulated bones but also some bones that seem to have been treated differently, such as a skull group and a couple of bone packages. It is suggested that most of the bones result from primary burials and subsequent disarticulation, but there are also indications of a change in burial practice, and the occurrence of special treatment could perhaps result from alternative, parallel practices.

Two of the individuals previously analyzed for aDNA (Skoglund et al., 2012, 2014) were re-analyzed here, Gökhem 2 (Gok2) and Gökhem 4 (Gok4). Strontium isotope ratios in tooth enamel suggest Gok2 to be local, while Gok4 may have spent his early years in a slightly more radiogenic area, possibly to the south of Falbygden. Gok2 is an almost complete, articulated skeleton of a female, 20-30 years old. The skeleton was registered as individual A during excavation. She was lying on her back with contracted legs. The aDNA sample is from M1 and M2 teeth in the mandible (id 118449). Gok4 is a disarticulated mandible from a young male, ca 20 years old (id 122402). The aDNA sample was from two premolar teeth.

Details of the dating are shown in Table S1. Dating were performed in Uppsala, Belfast and Århus. Details of laboratory procedures at the two latter are described by Reimer et al. (2015). Pretreatment in Uppsala was simpler and consisted of acid and distilled water washes, and no ultrafiltration. Further, collagen quality measures were not reported. We therefore re-dated Gok2, giving a slightly later date. C/N values for both individuals fall within the accepted range for well-preserved collagen, i.e., 2.9 - 3.6. C and N isotope values are normal for Neolithic humans in northern Europe, and suggest predominantly terrestrial protein sources from an environment of C3 plants. They do not suggest any significant intake of marine or freshwater protein, and there is therefore no reason to suspect any reservoir effect on these dating.

The carbon dating of all other individuals (including those infected with other ancient *Y. pestis* strains) used for the analyses were directly obtained from their original or previous publications and the corresponding average date of each individual is indicated in either Tables S2 or S4.

METHOD DETAILS

Screening and genome assembly of Y. pestis genomes

Raw ancient DNA sequencing datasets of human teeth from individuals in the Frälsegården passage grave (Skoglund et al., 2012, 2014) were downloaded from the European Nucleotide Archive (ERP001114 ; ERS434180) to search for the presence of human pathogens. Reads were trimmed for adapters using AdapterRemoval (v2.1.3) (Schubert et al., 2016) and leading/trailing stretches of Ns. Additionally, bases with a quality of 2 or less were removed by trimming from the 3′ and only reads larger than 30bp were retained. Trimmed sequences were first mapped separately against a custom-made database with a comprehensive collection of human pathogens, using BWA aln (v.0.7.10) (Li and Durbin, 2009). To identify samples with potential presence of a pathogen, we calculated the percentage of genome covered and the average sequencing depth in each sample (merging first all datasets in a sample) using the *genomecov* function of bedtools (Quinlan and Hall, 2010). The screening only showed potential hits against *Y. pestis*, so datasets were re-mapped against the *Y. pestis* CO92 genome (NC_003143.1, NC_003131.1, NC_003134.1, NC_003132.1) and *Y. pseudotuberculosis* IP32953 (NC_006155) using BWA aln (v.0.7.10) with the seed disabled. If the reads were paired end, they were collapsed during trimming and both merged, singleton and un-collapsed paired end reads were aligned. Alignments were processed using samtools (v.1.3.1) (Li et al., 2009) removing reads with a mapping quality lower than 30. Hereafter duplicates were removed using MarkDuplicates from picard-tools (v.2.9.3) (https://broadinstitute.github.io/picard/) and merged to per sample level. Finally, they were realigned using GATK (v.3.3.0) (DePristo et al., 2011) and rescaled for damage (i.e., mask C > T and G > A mutations predicted as ancient DNA degradation) using mapDamage (v.2.0.6) (Jónsson et al., 2013). All modern *Y. pestis* and *Y. pseudotuberculosis* genomes were processed as above except that they were not rescaled using mapDamage (Table S4). Alignment of reads from Gok4 to the complete NCBI nt database (version of March 2018) was performed using blastn (v.2.5.0) (Camacho et al., 2009) only considering hits with > 99% similarity over 99% of the sequence.

Naive Bayes classification of Yersinia pestis

To verify the presence of *Y. pestis*, we used the naive Bayes classifier from Rasmussen et al. (2015). Briefly, the method works by performing a qualitative assessment of read similarities using a naive Bayesian classifier to predict the species of an unknown sample. This is done based on the distribution of read counts mapping at different edit distances to the *Y. pestis* and *Y. pseudotuberculosis* reference genomes. The feature vector uses 10 input values, where the first set of features were the ratios

for edit distances 0 – 4 between the number of reads mapped to *Y. pestis* and the number of reads mapped to *Y. pseudotuberculosis*. The remaining features were the frequencies of each edit distance, from 0 – 4, of the mapped reads to *Y. pestis*. The output is the posterior probability that the unknown sample is from *Y. pestis*, *Y. pseudotuberculosis* or the outgroup *Y. similis*. The method was trained on data obtained from mapping reads of known origin to the two reference genomes and hereafter the Gok2 and Gok4 samples were processed to calculate the same 10 features as described above and input to the classifier in prediction mode. For more information see the section on QUANTIFICATION AND STATISTICAL ANALYSIS.

Genotyping of Y. pestis strains

We called genotypes for the genomes using both *Y. pseudotuberculosis* IP32953 and *Y. pestis* CO92 as reference genomes. We used the DNA damage-rescaled alignments and called genotypes using GATK HaplotypeCaller (v.3.8.0) for each sample with a ploidy of 1 and hereafter GenotypeGVCF including non-variant sites. We then filtered sites for a minimum 10X and maximum 1000X depth except for ancient samples that were filtered for a minimum of 4X. Additionally calls were filtered for allelic balance of 0.1 and a reference genotype quality (rgq) or genotype quality (gq) greater than 50 and fisher strand bias > 4 (phred-scaled). Indels were removed and the VCFs were transformed to fasta by recovering the sequences of 3,175 or 3,558 coding genes depending on using either *Y. pseudotuberculosis* or *Y. pestis* as the reference, respectively. This final set of genes were defined after removing gene sequences with high rates of missing bases (20 or more modern *Y. pestis* had > 10% of the gene missing).

Phylogenetic and molecular clock analyses

The genes passing the thresholds above (3,175 and 3,558) were merged to a super matrix and used for phylogenetic analyses. We partitioned the super matrix in 3 partitions based on codon position and reconstructed the phylogeny using RAxML (v.8.1.15) with the GTRGAMMA substitution model (Stamatakis, 2014). Bootstrapping was performed by generating 100 bootstrap replicates and their corresponding parsimony starting trees using RAxML. Hereafter, a standard Maximum Likelihood inference was run on each bootstrap replicate, and the resulting best trees were merged and drawn on the best ML tree.

Molecular clock analyses was performed using BEAST2 (v.2.4.4) (Bouckaert et al., 2014) and BEAGLE library (v2.1.2) (Ayres et al., 2012) based on the *Y. pestis* alignments and genotypes. We used a subset of 41 genomes, representing each clade in the *Y. pestis* phylogeny (see Table S4). The aligned data was partitioned based on codon positions and based on previous experiments (Rasmussen et al., 2015) we used a coalescent constant population size. Additionally, we assumed a relaxed log normal clock unlinked between the partitions as well as unlinked GTR substitution models with four gamma rate categories and empirical frequencies. We used a rooted ML tree from RAxML as the starting tree. We then ran 7 independent parallel BEAST chains sampling every 1,000 states for 185 – 214 million states and removed the first 30 million states as burn-in. The combined post burn-in data represented 1201 million states with an effective sample size (ESS) for the posterior of 273 and TreeHeight 367. The ESS for all other parameters were > 300. We then re-sampled approximately 25-30k trees from each chain for a total of 221k trees and summarized them using TreeAnnotater with a maximum clade credibility tree of median heights. The mean substitution rates for codon positions 1-3 were found to be 6.8E-8, 6.0E-8 and 9.5E-8 substitutions per site per year, respectively.

Heatmaps and functional classification of variants

Variants recovered from the VCF files were used to build heatmap plots to represent the distribution of these variants across genomes. We used filtered genomic variation (excluding indels) based on the *Y. pseudotuberculosis* genome as we were interested in the variants that emerged in *Y. pestis* strains during the evolution from their ancestral species. Variant effect was determined for coding regions using SNPeff (v.4.2) (Cingolani et al., 2012). We identified all variants in the Gok2 strain and also variants found in at least one of the seven Bronze Age genomes. Positions that were identical in all *Y. pestis* genomes (i.e., either SNVs or unmapped), were excluded from the analysis. The recovered variants were plotted as heatmaps in R (v3.3.3) (R Development Core Team, 2008), using the heatmap.2 function in gplots package (v3.0.1). Genomes were sorted using the maximum likelihood phylogenetic tree calculated above, which was imported and adjusted using the DECIPHER (v2.2.0) (Wright, 2016) and dendextend (v1.5.2) packages (Galili, 2015). Variants were clustered using the euclidean distance in the vegdist function of the package Vegan (v2.4.4) (Dixon, 2003). Metabolic functions of the genes where SNVs were located were determined using the SEED subsystems annotation of the *Y. pseudotuberculosis* genome, obtained from the RAST server (Aziz et al., 2008).

Admixture analyses of human genomes

From the Reich Lab homepage, we downloaded genotypes for 1,058 selected individuals (Table S2) from the Near-East and Eurasia (Lazaridis et al., 2016; Lipson et al., 2017; Mathieson et al., 2018; Olalde et al., 2018), which had been assembled using build 37 of the human reference genome. We filtered these genotypes by linkage disequilibrium using PLINK (v1.90b5.3) (Purcell et al., 2007) with the flag–indep-pairwise 200 25 0.4, leaving 693,339 SNPs. We ran ADMIXTURE (v1.3.0) (Alexander et al., 2009) with the cross validation (–cv) flag specifying from K = 2 to K = 15 clusters, with 10 replicates for each value of K. For each value of K, the replicate with highest log likelihood was kept. From these we chose to keep results from K = 5, since it was the lowest number of clusters that could clearly differentiate the ancestry of the previously described groups and time periods. Using the distributions of the K5 components, we estimated the distances between the predicted ancestries of individuals using the symmetric Kullback-Leibler (KL) divergence (the mean of the KL divergence of A versus B and B versus A) with an all versus all pairwise comparison. Then using a two-sided

t test we evaluated if significant differences could be found between the distances within each group defined in Figure 4B compared to all other groups.

QUANTIFICATION AND STATISTICAL ANALYSIS

Naive Bayes classification

The classification of reads mapping to the *Y. pestis* reference genome as either *Y. pestis* or *Y. pseudotuberculosis* or *other* was performed as specified in Rasmussen et al. (2015) and in the methods section above. The posterior probability for each class was determined using Bayes Theorem (Equation 1), e.g., for class 1 (*Y. pestis*):

$$P(C_1 \mid F) = \frac{P(F \mid C_1)P(C_1)}{P(F)}. \quad (1)$$

Where the likelihood of a class is determined by multiplying the probabilities of observing each individual feature (F_i) given the predicted class (Equation 2), the sum of probabilities (Equation 3) and using a flat prior for the three possible classes.

$$P(F \mid C_1) = \prod_{i=1}^{10} P(F_i \mid C_1) \quad (2)$$

$$P(F) = P(F \mid C_1)P(C_1) + P(F \mid C_2)P(C_2) + P(F \mid C_3)P(C_3) \quad (3)$$

Symmetric Kullback-Leibler divergence

The KL divergence was calculated between two individuals P and Q using their predicted admixture proportions assuming 5 ancestral populations. The symmetric divergence was then determined using the average of the divergence between P and Q, and Q and P (Equation 4).

$$D_{symKL}(Q,P) = \frac{1}{2}\left(\sum_{i=1}^{5} Q(i)\ \ln\left(\frac{Q(i)}{P(i)}\right) + P(i)\ \ln\left(\frac{P(i)}{Q(i)}\right)\right) \quad (4)$$

Statistical significance between two groups of individuals was assessed by comparing the distributions of the symmetric KL divergence using a two-sided t test. Significance was defined as a p value less than 0.01.

DATA AND SOFTWARE AVAILABILITY

All bacterial data is available from the Short Read Archive (SRA) and the European Nucleotide Archive (ENA), accessions are listed in Table S4. Human genetic data is available from the Reich Lab homepage (https://reich.hms.harvard.edu/datasets). All major software used are publicly available or available upon request from the corresponding authors.

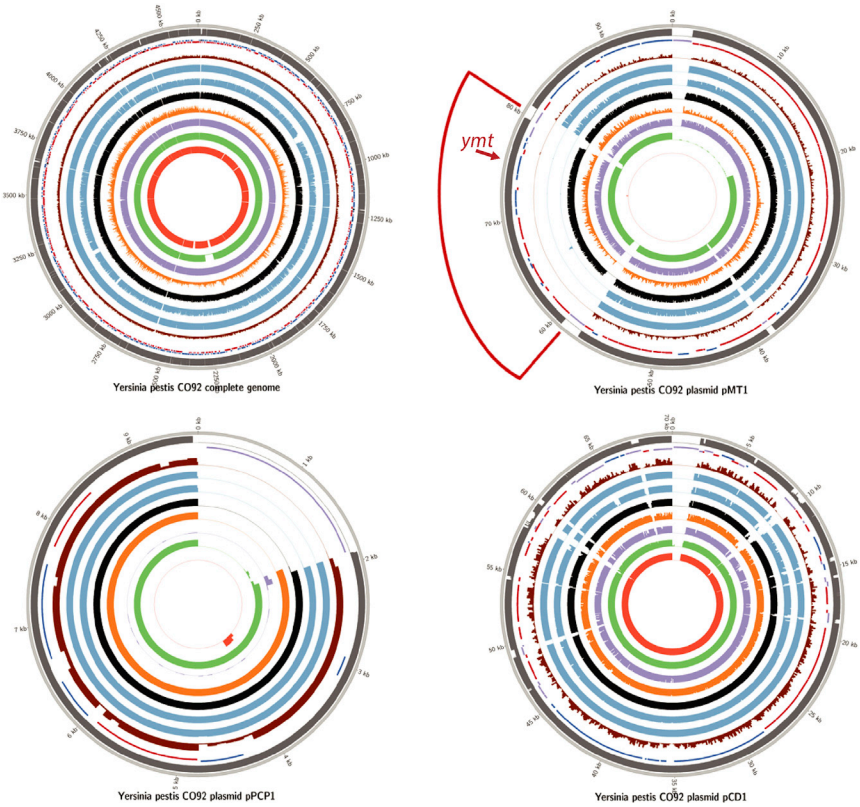


Figure S1. Depth of Coverage of the Gok2 *Y. pestis* Genome, Related to Figure 1

High quality reads mapping across the *Y. pestis* chromosome and the three plasmids (pMT1, pPCP1 and pCD1). Outer ring is map-ability (gray) and genes are marked as follows: RNA: black, transposon: purple, positive strand: blue, negative strand: red. Read depth of Gok2 (brown), RISE509 and RISE505 (blue), DA101/ Tian Shan plague (black), Justinian plague (orange), Black Death (purple), modern *Y. pestis* D1982001 (green) and *Y. pseudotuberculosis* IP32881 (red). A red arrow indicates the position of the *ymt* gene within a missing region (also marked in red) of the pMT1 plasmid in the Bronze Age and Gok2 strains.

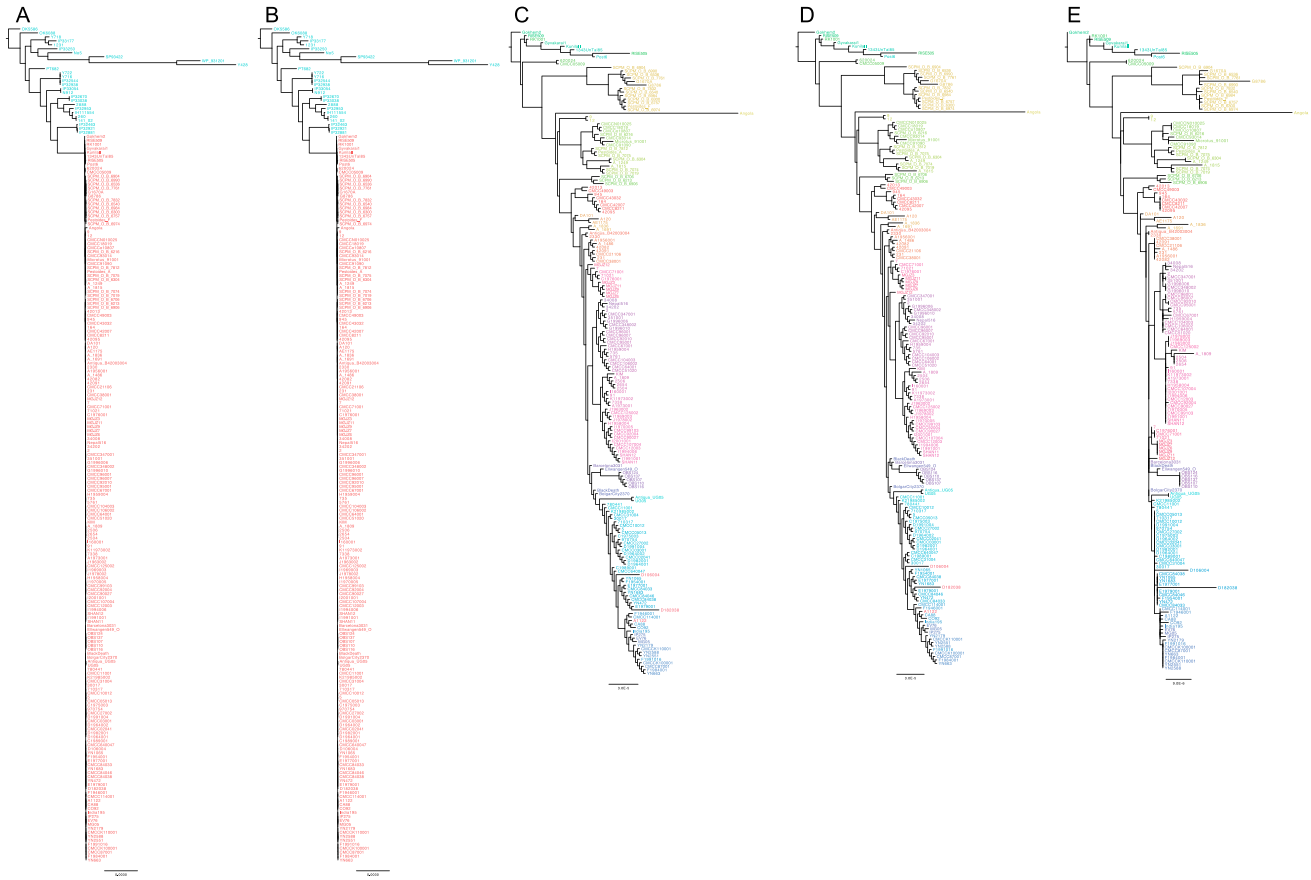


Figure S2. Phylogenetic Trees, Related to Figure 1

Full maximum likelihood phylogenetic trees reconstructed from whole-genome information of modern (163) and ancient (20) *Y. pestis* strains and 27 selected *Y. pseudotuberculosis* strains. Whole genomes were reconstructed by mapping the reads on either the *Y. pestis* CO92 genome (A) or the *Y. pseudotuberculosis* IP32953 (B) genome and phylogenetic reconstructions were made using the same parameters in both cases. Trees excluding *Y. pseudotuberculosis* strains for better clarity of the topological distribution of *Y. pestis* is shown in (C) and (D) using *Y. pestis* and *Y. pseudotuberculosis* as reference, respectively. The phylogenetic tree was also determined using only transversion (i.e., not using C > T and G > A deamination that can be produced through ancient DNA damage) using *Y. pseudotuberculosis* IP32953 as reference genome (E). Strains are colored according to major *Y. pestis* clades (Table S4).

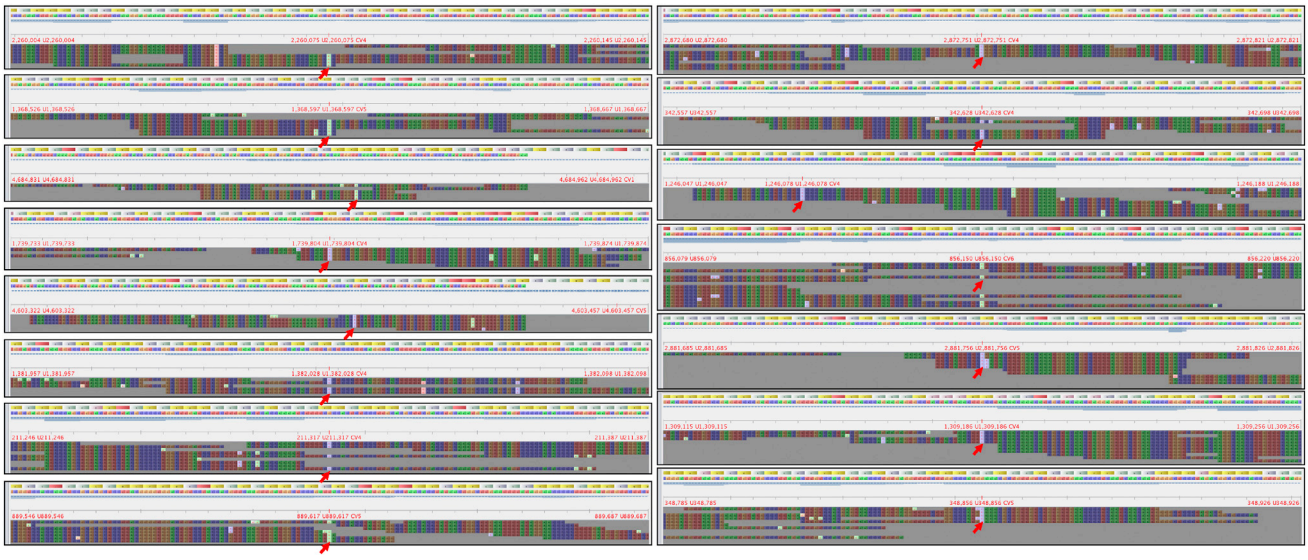


Figure S3. Reads Supporting the Gok2 SNVs, Related to Figure 2

Detailed read mapping supporting each of the 15 SNVs shown in Figure 2A, that were uniquely found in the Gok2 strain. These results show that each of these mutations (marked with red arrows) were supported by independent reads, and were not placed at the extreme of reads (where deamination due to ancient DNA degradation take place). These results strongly suggest that these are genuine variants that emerged within in the Gok2 strain.

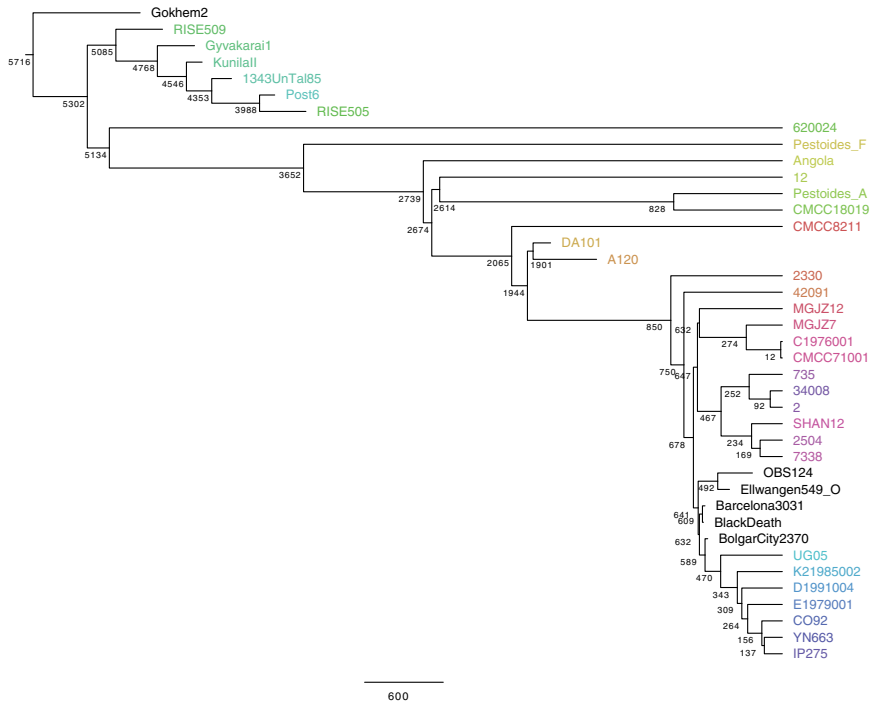
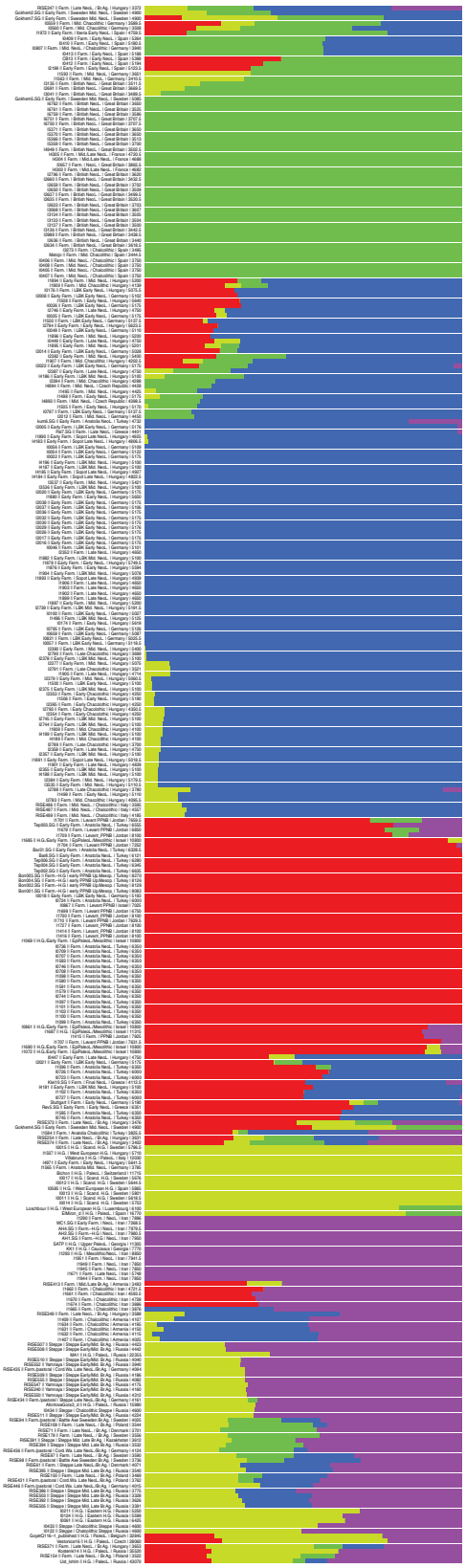


Figure S4. BEAST Divergence Dating, Related to Figure 3

Molecular clock analysis displayed in Figure 3, but showing all strains used in the analysis and their estimated divergence times. Strains are colored according to major *Y. pestis* clades and branch lengths are given as years before present (BP).



(legend on next page)

Figure S5. Ancestry of the 1,058 Selected Human Genomes, Related to Figure 4

Genome-wide ADMIXTURE analysis, with K = 5, with the predicted ancestry proportions of the 1,058 individuals used to build Figure 4. The ID, lifestyle (H.G: hunter-gatherer, Farm.: Farmer), period, country where remains were found and carbon dating of each individual is indicated in the y axis labels separated by a pipe character.