The Stone Age Plague: 1000 years of Persistence in Eurasia

Aida Andrades Valtueña[1], Alissa Mittnik[1,2], Ken Massy[3,4], Raili Allmäe[5], Mantas Daubaras[6], Rimantas Jankauskas[7], Mari Tõrv[8], Saskia Pfrengle[2], Maria A. Spyrou[1,2], Michal Feldman[1,2], Wolfgang Haak[1,9], Kirsten I. Bos[1,2], Philipp W. Stockhammer[3], Alexander Herbig[1,2]* and Johannes Krause[1,2]*

[1]Max Planck Institute for the Science of Human History, Jena, Germany
[2]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany
[3]Ludwig-Maximilians-University, Munich, Germany,
[4]Heidelberg Academy of Sciences, Heidelberg, Germany
[5] Archaeological Research Collection, Tallinn University, Tallinn, Estonia
[6]Department of Archaeology, Lithuanian Institute of History, Vilnius
[7]Department of Anatomy, Histology and Anthropology, Vilnius University, Vilnius, Lithuania
[8]Independent researcher
[9]School of Biological Sciences, The University of Adelaide, Adelaide SA-5005, South Australia, Australia

* To whom correspondence should be addressed

Abstract

Molecular signatures of Yersinia pestis were recently identified in prehistoric Eurasian individuals, thus suggesting Y. pestis might have caused some form of plague in humans prior to the first historically documented pandemic. Here, we present four new Y. pestis genomes from the European Late Neolithic and Bronze Age (LNBA) dating from 4,500 to 3,700 BP. We show that all currently investigated LNBA strains form a single genetic clade in the Y. pestis phylogeny that appears to be extinct today. Interpreting our data within the context of recent ancient human genomic evidence, which suggests an increase in human mobility during the LNBA, we propose a possible scenario for the spread of Y. pestis during the LNBA: Y. pestis may have entered Europe from Central Eurasia during an expansion of steppe pastoralists, possibly persisted within Europe until the mid Bronze Age, and moved back towards Central Eurasia in subsequent human population movements.

Introduction

Plague pandemics throughout human history caused unprecedented levels of mortality that contributed to profound socioeconomic and political changes. Conventionally it is assumed that

plague affected human populations in three pandemic waves. The first, the Plague of Justinian, starting in the 6th century AD, was followed by multiple epidemic outbreaks in Europe and the Mediterranean basin, and has been associated with the weakening and decay of the Byzantine empire (Russell, 1968). The second plague pandemic first struck in the 14th century with the infamous 'Black Death' (1347-1352), which again spread from Asia to Europe seemingly along both land and maritime routes (Zietz and Dunkelberg, 2004). It is estimated that this initial onslaught killed 50% of the European population (Benedictow, 2004). It was followed by outbreaks of varying intensity that lasted until the late eighteenth century (Cohn JR, 2008). The most recent plague pandemic started in the 19th century and began in the Yunnan province of China. It reached Hong Kong by 1894 and followed global trade routes to achieve a near worldwide distribution (Stenseth et al., 2008). Since then plague has persisted in rodent populations in many areas of the world and continues to cause both isolated human cases and local epidemics (http://www.who.int/mediacentre/factsheets/fs267/en/).

Plague is caused by a systemic infection with the Gram-negative bacterium *Yersinia pestis.* Advances in ancient DNA (aDNA) research have permitted the successful reconstruction of a series of *Y. pestis* genomes from victims of both the first and second plague pandemics, thus confirming a *Y. pestis* involvement and providing new insights and perspectives on how this bacterium historically spread through Europe (Bos et al., 2011, 2016; Feldman et al., 2016; Spyrou et al., 2016; Wagner et al., 2014). Most recently, a study by Spyrou et al. (2016) suggested that during the second pandemic, a European focus was established from where subsequent outbreaks, such as the Ellwangen outbreak (16th century Germany) or the Great Plague of Marseille (1720-1722 France), were derived (Spyrou et al., 2016). The authors also proposed that a descendant of the Black Death strain travelled eastwards in the late 14th century, became established in East Asia and subsequently gave rise to the most recent plague pandemic that spread the pathogen around the globe.

Our perception of the evolutionary history of *Y. pestis* was changed substantially by a recent report of two reconstructed genomes from Central Asian Bronze Age steppe nomads/pastoralists (~4,729 cal BP and ~3,635 cal BP, respectively) and molecular *Y. pestis* signatures in an additional five individuals from Eurasia (~4,500 to 2,800 BP) suggesting the presence of plague in human populations over a diffuse geographic range prior to the first historically recorded pandemics. Phylogenetic analysis of the two reconstructed *Y. pestis* genomes from the Altai region shows that they occupy a phylogenetic position ancestral to all medieval and extant *Y. pestis* strains, though this branch was not adequately resolved (bootstrap lower than 95%, Rasmussen et al., 2015). Further open questions remain regarding

Y. pestis' early association with humans. It is not currently known if the *Y. pestis* lineages circulating in Europe during the Late Neolithic and Bronze Age were all descended from the ~5,000 BP Central Asian strain or whether there were multiple strains circulating in Europe and Asia. Furthermore, how did plague spread over such a vast territory during the period comprising the Late Neolithic and Bronze Age? Could these bacterial strains have been associated with certain human groups and their respective subsistence strategies and cultures?

The Late Neolithic and Early Bronze Age in Western Eurasia (ca. 4,900-3,700/3,600 BP; cf. Stockhammer et al., 2015) was a time of major transformative cultural and social changes that led to cross-European networks of contact and exchange (Vandkilde, 2016). Intriguingly, recent studies on ancient human genomes suggested a major expansion of people from the Eurasian Steppe westwards into central Europe as well as eastwards into Central Asia and Southern Siberia starting around 4,800 BP (Allentoft et al., 2015; Haak et al., 2015). These steppe pastoralists carried a genetic component that is present in all Europeans today but was absent in early and middle Neolithic Europeans prior to their arrival. The highest amount of 'steppe ancestry' in ancient Europeans was found in individuals associated with the Late Neolithic Corded Ware Complex (Figure 1) around 4,500 BP (Haak et al., 2015), who show a genetic makeup close to pastoralists from the Pontic Steppe associated with the Yamnaya culture, suggesting a strong genetic link between those two groups. Furthermore, it could be shown that this genetic component also appears in individuals associated with the Andronovo culture in the Altai region around 4,200 BP (Allentoft et al., 2015). These genetic links between humans from Western Eurasia and the Central Asian steppe in Siberia highlight the dimensions of mobility and connectedness at the time of the Bronze Age.

The reasons for the magnitude of the genetic turnover that occurred in Central Europe around 4800 BP, where around 75% of the local farmer genetics was replaced(Haak et al., 2015), have yet to be explained. As in other episodes of our history, infectious diseases may have played a significant role in triggering or catalyzing those major cultural shifts and human migrations. Here we present four novel Late Neolithic and Bronze Age (LNBA) *Y. pestis* genomes from Central Europe. Through analyses alongside other ancient and modern lineages (Bos et al., 2011, 2016; Cui et al., 2013; Feldman et al., 2016; Kislichkina et al., 2015; Spyrou et al., 2016; Zhgenti et al., 2015), we show that all LNBA *Y. pestis* strains form a single clade in the phylogeny. This indicates a common origin of all currently identified *Y. pestis* strains circulating in Eurasia during the Late Neolithic and Bronze Age, and reveals a distribution pattern that parallels human movements in time and space.

Results

Screening

A total of 168 tooth and bone samples dating from the Neolithic and Bronze Age from Lithuania (27), Estonia (45), Latvia (10), and Germany (Althausen 4, Augsburg 83) were screened for *Y. pestis* by mapping non-enriched reads ranging from 700,000-21,000,000 against a multi-fasta reference consisting of the genomes of 12 different *Yersinia* species (Table 1).

To assess if a sample was positive for *Y. pestis*, we calculated a score based on the number of specific reads mapping to *Y. pestis* in comparison to the number of reads mapping to other *Yersinia* species (See methods). Following this metric, all samples with a positive score were identified as possible candidates. Samples that had a score higher than 0.1, and had reads mapping to all the three plasmids present in all *Y. pestis* were considered 'strong' positives. In our dataset we identified four strong positives from three different locations dating from the Late Neolithic to the Early Bronze Age: one sample from the Lithuanian site Gyvakarai (Gyvakarai1), one sample from the Estonian site Kunila (KunilaII) and two samples from Augsburg, Germany (Haunstetten, Unterer Talweg 85 Feature 1343 (1343UnTal85), Haunstetten, Postillionstraße Feature 6 (6Post), Table 2 and SI for a detailed description of the samples).

Genome reconstruction

The four strong positive samples identified during the screening step were shotgun sequenced resulting in 379,155,741 to 1,410,707,182 reads. After mapping to the reference genome (*Y. pestis* CO92, NC_003143.1), we reconstructed four genomes with a mean coverage ranging from 5-12 X with 90-94% of the reference genome covered at least 1 fold (Table 2). The reads were independently mapped to the three plasmids of *Y. pestis* CO92, and we reconstructed the three plasmids for our ancient samples with mean coverages of: pCD1 11–25 X, pMT1 6–12 X and pPCP1 22-52 X (Supplementary Table 1).

In order to authenticate the ancient origin of the bacterial genomes, we looked for the typical damage patterns of terminal deamination common to ancient DNA(Briggs et al., 2007). All our samples present this typical damage profile (Supplementary Figure 1). Post6 and

1343UnTal85 only retain damage in the last two bases as these libraries were prepared using a 'UDG-half' protocol (Rohland et al., 2015, see SI Methods).

The four reconstructed genomes and their plasmids were compared to the two Bronze Age genomes reported previously (Rasmussen et al., 2015). After visual inspection of aligned reads, our prehistoric genomes from Europe showed similar coverage of the reference genome CO92, and all regions were also covered in the Bronze Age Altai *Y. pestis* genomes (Figure 2A). The four reconstructed genomes in this study lack the same part of the pMT1 plasmid, which contains the *ymt* gene (Figure 2), as already identified in the Altai genomes (Rasmussen et al., 2015). The *ymt* gene codes for the *Yersinia* murine toxin, which is an important virulence factor in *Y. pestis* related to transmission via the flea vector (Hinnebusch et al., 2002, 2000). The expression of *ymt* protects the bacteria from the toxic blood digestion by-products in the flea's gut thus functions to aid in colonization of the flea midgut (Hinnebusch et al., 2002).

Phylogeny and Dating

To assess the phylogenetic positioning of the four European LNBA *Y. pestis* genomes with respect to the modern and ancient *Y. pestis* genomes, Neighbour Joining (NJ, Supplementary Figure 2A), Maximum Parsimony (MP, Supplementary Figure 2B) and Maximum Likelihood (ML, Figure 3, Supplementary Figure 2C) trees were computed. Our samples form a distinct clade in the *Y. pestis* phylogeny together with the previously reconstructed Central Asian Bronze Age *Y. pestis* genomes (Rasmussen et al., 2015). This topology has a high bootstrap support of >95% in all three methods. The branching point of the LNBA genomes with the main branch leading towards the modern *Y. pestis* strains represents the most recent common ancestor (MRCA) of all the extant and ancient *Y. pestis* genomes sequenced until today.

To date the MRCA of *Y. pestis*, we performed a 'tip dating' analysis using BEAST (Drummond et al., 2012). The MRCA of all *Y. pestis* was dated to 6,295 years (95% HPD interval: 5,063-7,787 years). This estimation overlaps with those previously made by Rasmussen et al. (2015)(5,783 years, 95% HPD interval: 5,021–7,022 years) suggesting a Holocene origin for plague. The time to the MRCA of *Y. pestis* and *Y. pseudotuberculosis* was estimated to 34,636 years (95% HPD interval: 18,559-55,044 years). A maximum clade credibility tree was computed (Supplementary Figure 3) supporting the same topology as the NJ, MP and ML with high statistical support of the branching points of the LNBA plague clade.

Genetic makeup

To determine the effect of the Single Nucleotide Polymorphisms (SNPs) detected in our dataset, SNP effects were evaluated using the software *snpEff* (Cingolani et al., 2012) and an in-house program (*MultiVCFAnalyzer*). A total of 416 SNPs were found in the LNBA branch including strain-specific and shared SNPs. A total of 110 synonymous and 196 non-synonymous SNPs are present in the LNBA branch. All the LNBA genomes share six SNPs: four non-synonymous, one stop gained and one intergenic (Supplementary Table 2).

Additionally, a total of eight stop mutations were detected in that ancient branch, which are not shared by all the LNBA genomes. Most of these mutations are found in the terminal part of the LNBA branch with six being specific to the youngest Early Bronze Age *Y. pestis* strain (RISE505), one being shared between RISE505 and Post6, and only one being Gyvakarai1-specific (Supplementary Table 3). Additionally, RISE505 misses the start codon of the YPO0956 gene, which is involved in iron transport, and one stop codon in YPO2909, which is a pseudogene.

To identify potential homoplasies, a table of all variable SNPs was examined for any that contradict the tree topology. A total of eight homoplasies were detected (Supplementary Table 4). Furthermore, a tri-allelic site was detected at nucleotide position 4,104,762 (A,T,C).

The percentage of the gene covered in the LNBA plague genomes was calculated for a set of genes that are related to virulence, flea transmission and colonization and dissemination and these were inspected in greater detail (Figure 2B). We observe the absence of YpfΦ (Derbise et al., 2007), a filamentous prophage, in all LNBA plague genomes. While YpfΦ is found in some *Y. pestis* strains of branch 0, branch 1 and branch 2 as a free phage, it has only been fully integrated and stabilized into the chromosome of the strains 1.ORI that are responsible for the third pandemic (Derbise and Carniel, 2014). Additionally, the *yapC* gene was lost in the three younger LNBA strains (1343UnTal85, Post6 and RISE505). YapC was initially thought to be involved in the adhesion to mammalian cells, autoagglutination and biofilm formation when expressed in *E. coli* (Felek et al., 2008). However, the *yapC* knockout in *Y. pestis* does not affect those functions and Felek and colleagues have thus suggested that this is due to either low expression of *yapC in vitro* or by compensation through other genes. The only virulence factor located in the plasmids missing in all the LNBA *Y. pestis* strains is *ymt* (Figure 2). Furthermore, genes involved in mammalian disease, such as *pla* and *caf1,* were already present in the LNBA*Y. pestis* genomes. The *pla* gene is involved in the dissemination of the bacteria in the mammalian host by promoting the migration of the bacteria to the lymphatic

nodes (Lathem et al., 2007; Sebbane et al., 2006), while the *caf1* gene encodes for the F1 capsular antigen, which confers phagocytosis resistance to the bacterium (Du et al., 2002). Both genes are absent in the closest relative *Y. pseudotuberculosis*.

Urease D (*ureD*) is an important gene that plays a role in flea transmission. When *ureD* is expressed in the flea vector it causes a toxic oral reaction to the flea killing around 30-40% of the flea vectors (Chouikha and Hinnebusch, 2014). While *ureD* is functional in *Y. pseudotuberculosis*, it is a pseudogene in *Y. pestis*. The pseudogenization of this gene is caused by a frameshift mutation (insertion of a G in a six G-stretch) in *Y. pestis* (Sebbane et al., 2001). The LNBA *Y. pestis* were inspected in the search of this specific frameshift mutation. This insertion is not present in those genomes indicating that this gene was still functional in *Y. pestis* at that time, suggesting that it was as toxic to fleas as its ancestor *Y. pseudotuberculosis*.

Large-scale insertions and deletions (indels) were evaluated by comparison of mapped data for the LNBA *Y.* pestis genomes, branch 0 strains (0.PE7, 0.PE2-F. 0.PE3, 0.PE4), KIM, and CO92 using *Y. pseudotuberculosis* IP 32953 (NC_006155.1) as a reference. Regions larger than 1 kb were explored as possible indels. We detected two regions present in the LNBA *Y. pestis* samples that are absent in all the other strains analyzed: a 1kb region (2,587,386-2,588,553) that contains a single gene (YPTB0714) encoding an aldehyde dehydrogenase, part of the R3 *Y. pseudotuberculosis*-specific region identified by Pouillot et al. (2008) and a second region (1.5kb, 3,295,644-3,297,223) that contains a single gene (YPTB2793) encoding a uracil/xanthine transporter being part of the region orf1, which was also characterized as *Y. pseudotuberculosis*-specific (Pouillot et al., 2008). Additionally, two missing regions were detected: one region of 34kb is missing in the three younger genomes of the LNBA lineage (Post6, 1343UnTal85 and RISE505) and another 36kb region is missing in the youngest sample RISE505, already identified by Rasmussen et al. (2015), which contains flagella genes. These two missing regions contain multiple membrane proteins, which could be potential virulence factors or antigens recognized by the immune system of the host.

Discussion

The four prehistoric genomes presented here are the first complete *Y. pestis* genomes from the Late Neolithic and Bronze Age in Europe. They form a distinct clade with the previously reconstructed Central Asian Bronze Age *Y. pestis* genomes, confirming that all Late Neolithic and Bronze Age genomes reconstructed so far originated from a common ancestor. The oldest genome RISE509 (Rasmussen et al., 2015) occupies the most basal position of all *Y. pestis*

genomes sequenced to date. This suggests that Central Eurasia rather than Eastern Asia should be considered as the region of potential origin of plague.

The temporal and spatial distribution of the Late Neolithic and Bronze Age *Y. pestis* genomes allows us to evaluate the evolution and dissemination of plague in prehistory. We propose two contrasting scenarios to explain the phylogenetic pattern observed in the LNBA *Y. pestis* branch:

1. The first scenario assumes that plague was introduced multiple times to Europe from a common reservoir between 5,000 to 3,000 BP. Here, the bacterium would have been spread independently from a source, most likely located in Central Eurasia, to Europe at least four times during a period of over 1,000 years (Figure 1), travelling once to Lithuania, once to Estonia, and two times to Southern Germany.
2. The second scenario assumes that plague entered Europe once, established a reservoir within or close to Europe from which plague circulated, and then moved back to Central Eurasia and the Altai region/East Asia during the Bronze Age (Figure 1). This hypothesis of persistence mirrors that which has been proposed to explain the presence of plague in Europe during the notorious second pandemic (Spyrou et al., 2016).

With these few genomes available it is difficult to disentangle the two hypotheses; however, interpreting our data in the context of what is known from human genetics and archaeological data can offer some resolution. Ancient human genomic data point to a change in mobility and a massive expansion of people from the Caspian-Pontic Steppe related to individuals associated with the 'Yamnaya' complex, both to the East and the West starting around 4,800 BP. They carried a distinct genetic component that first appears in Central Europe in individuals from the Corded Ware Complex that becomes part of the genetic composition of most subsequent and all modern day European populations (Allentoft et al., 2015; Haak et al., 2015). It was furthermore shown that there is a close genetic link between the highly mobile group of pastoralists associated to the Central Asian 'Andronovo Complex', the pontic steppe 'Yamnaya', and the Central and Eastern European Corded Ware Complex (Allentoft et al., 2015). The first indication of plague in Europe is found in the Baltic region and coincides with the time of the arrival of the steppe component (Allentoft et al., 2015). The two Late Neolithic *Y. pestis* genomes from the Baltic in this study were reconstructed from individuals associated with the Corded Ware Complex (Gyvakarai1 and KunilaII). The Baltic *Y. pestis* genomes are genetically derived from the strain that was found in the 'Andronovo Complex' from the Altai region, suggesting that the disease might have spread with steppe pastoralists from Central Eurasia to Eastern and Central Europe during their massive range expansion. The younger Late Neolithic

Y. pestis genomes from Southern Germany are genetically derived from the Baltic strains and are found in individuals associated with the Bell Beaker Complex. Previous analysis have shown that Bell Beaker individuals from Germany also carry 'steppe ancestry (Allentoft et al., 2015; Haak et al., 2015). This suggests that *Y. pestis* may have been spread further southwestwards analogous to the human steppe component. The youngest of the LNBA *Y. pestis* genomes (RISE505), found also in the Altai region, descends from the Central European strains, and thus suggest a spread back into the eastern steppes. Interestingly, genome-wide human data shows that human individuals dating from around 3,700-3,500 BP from the Sintashta culture north of the Caspian Sea carried mixed ancestry of European/Anatolian farmers and steppe pastoralists, suggesting a backflow of human genes from Europe to Central Asia (Allentoft et al., 2015). The steppe as a natural corridor connecting pastoralists throughout Central and West Eurasia might have facilitated the spread of strains closely related to the European Early Bronze Age *Y. pestis* back to the Altai region. The patterns in human genetic ancestry and admixture and the temporal series within the LNBA *Y. pestis* branch therefore support scenario 2, suggesting that *Y. pestis* was introduced to Europe from the steppe around 4,800 BP. Thereafter, the pathogen became established in a local reservoir within or in close proximity to Europe, from where the European *Y. pestis* strain was disseminated back to the Altai region in a process connected to the backflow of human genetic ancestry from Western Eurasia into the Altai. The pathogen data, therefore, complements the archaeological evidence, which indicates a strong intensification of Eurasian networks since the beginning of the Bronze Age (Vandkilde, 2016).

Even though *Y. pestis* seems to have been spread following human movements, its mode of transmission during this early phase of its evolution cannot be easily determined. Most contemporary cases of *Y. pestis* infection occur via an arthropod vector. The flea transmission can be accomplished by one of two mechanisms: the classical blockage-dependent flea transmission (Hinnebusch et al., 1998) and the recently proposed early-phase transmission (EPT, Eisen et al., 2006). In the blockage-dependent model, *Y. pestis* causes an obstruction in the flea digestive system by producing a biofilm that blocks the pre-gut of the flea within 1-2 weeks after infection. This blockage prevents a blood meal from reaching the flea's gut, and regurgitation of the blood by a hungry flea in repeated attempts to feed sheding several live bacteria into the blood stream of the host (Chouikha and Hinnebusch, 2012; Hinnebusch et al., 1998). It has been shown that the blockage-dependent transmission requires a functional *ymt* gene and *hms* locus, and non-functional *rcsA, pde2* and *pde3* genes (Sun et al., 2014). *ymt* protects *Y. pestis* from toxic by-products of blood digestion and allows the bacterium to colonise

the mid-gut of the flea. The *hms* locus is involved in biofilm formation and *rcsA*, *pde2* and *pde3* are the down-regulators of biofilm formation. However, evidence is emerging that *Y. pestis* can be transmitted efficiently within the first 1-4 days after entering the flea prior to biofilm formation (Eisen et al., 2006, 2015), in a process known as the EPT model. Unfortunately this model is currently less well understood molecularly and physiologically than blockage-dependent transmission, but has been shown to be biofilm (Vetter et al., 2010) and *ymt* independent (Johnson et al., 2014).

Based on the genetic characteristics of the LNBA genomes (i.e. lack of *ymt*, still functional *pde2* and *rcsA* as shown by previous work, Rasmussen et al., 2015), functional *ureD* which will kill 30-40% of the flea vectors) it seems unlikely that *Y. pestis* was able to use a flea vector in a blockage-dependent model. However, since neither of these genes seem to be required for EPT, it is possible that LNBA *Y. pestis* was transmitted by a flea vector via EPT. Under this assumption transmission would have been presumably less efficient since a functional Urease D would have reduced the number of fleas transmitting the bacteria due the toxicity induced by the urease activity, a mechanism inactivated in all other *Y. pestis* lineages.

The presence of genes involved in virulence in the mammalian host such as *pla* and the F1 capsule, which are absent in *Y. pseudotuberculosis*, indicates that LNBA *Y. pestis* was already adapted to mammalian hosts to some extent. *pla* aids *Y. pestis* infiltration of the mammalian host (Lathem et al., 2007; Sebbane et al., 2006). The *pla* gene present in the LNBA *Y. pestis* strains has the ancestral I259 variant, which has been shown to be less efficient than the derived T259 form(Haiko et al., 2009). *Y. pestis* with the ancestral variant is able to cause pneumonic disease, however, it is less efficient in colonizing other tissues (Zimbler et al., 2015). This indicates that LNBA *Y. pestis* could potentially cause a pneumonic or a less virulent bubonic form. In addition to the above noted changes, we detected two regions missing in the LNBA genomes: a ~34kb region that contains genes encoding membrane proteins missing in the three youngest *Y. pestis* strains (Post6, 1343UnTal85 and RISE505) and a ~36kb region containing genes encoding proteins involved in flagellin production and iron transporters missing in the youngest sample RISE505, as observed elsewhere (Rasmussen et al., 2015). This genome reduction affecting membrane and flagellar proteins potentially involved in interactions with the host's immune system, can be an indication of adaptation to a new host pathogenic lifestyle (Ochman and Moran, 2001).

Our common understanding is that plague is a disease adapted to rodents, where commensal species such as *Rattus rattus* and their fleas play a central role as disease vectors for humans (Perry and Fetherston, 1997). While a rodent-flea mediated transmission model is

compatible with the genomic makeup of the LNBA strains, disease dynamics may well have differed in the past. The most parsimonious explanation would be that LNBA plague indeed traveled with rodent species commensal to humans, in keeping with the orthodox model of plague transmission. The Neolithic is conventionally considered to be a time period where new diseases were introduced into human groups as they made the transition from a nomadic lifestyle to one of sedentism, and where the adoption of agriculture and increased population density acted synergistically to change the disease landscape (Ronald Barrett et al., 1998). Whether indeed commensal rodent populations were large enough to function as reservoir populations of plague during human migrations at this time is unknown. Alternative models of transmission involving different host species might carry some traction, as the ancient disease may have behaved rather differently from the form we know today.

Here, we present the first LNBA *Y. pestis* genomes from Europe. We show that all LNBA genomes reconstructed so far form a distinct lineage that potentially entered Europe following the migration of steppe pastoralists around 4,800 BP. We find striking parallels between the *Y. pestis* dispersal pattern and human population movements during this time period. We propose two scenarios for its presence in Europe: a multiple introduction hypothesis from a Central Eurasian source, or the establishment of a local *Y. pestis* focus within or close to Europe from where a related strain ultimately moved back towards Central Asia in the Bronze Age. On account of the chronology, the ancient *Y. pestis* phylogeny, and known patterns of human mobility, we find stronger support for the second scenario.

The LNBA period was a time of increased mobility and cultural change. The presence of *Y. pestis* may have been a promoting factor for the increase in mobility of human populations (Rasmussen et al., 2015). The manifestation of the disease in Europe could have played a major role in the processes that led to the genetic turnover observed in the European human populations that might have harbored different levels of immunity against this newly introduced disease. Testing these hypotheses will require more extensive assessment of both human and *Y. pestis* genomes from the presumed source population before and after the migration from the steppes, as well as in Europe during this period of genetic turnover.

Methods

In Silico Screening

DNA from bone and dentine samples from Lithuania (27), Estonia (45), Latvia (10), and Germany (Althausen 4, Augsburg 83) was extracted in dedicated clean room facilities, followed by DNA libraries preparation and Illumina sequencing at the University of Tübingen and the MPI-SHH in Jena as detailed in SI Methods. The sequencing data for each sample was preprocessed with ClipAndMerge (Peltzer et al., 2016) to remove adaptors, base quality-trim (20) and merging and filtering for only merged reads. Reads were mapped using the BWA aln algorithm (Li and Durbin, 2009) to a multi-species reference panel, containing various representatives of the genus *Yersinia* (Table 2) and the plasmids of *Yersinia pestis*: pCD1, pMT1 and pPCP1 from *Y. pestis* CO92. The region comprising 3000-4200bp of the *Y. pestis* specific plasmid pPCP1 was masked in the reference, since it is highly similar to an expression vector used during the production of enzyme reagents (Schuenemann et al., 2011).

Mapped files were then filtered for reads with a mapping quality higher than 20 with Samtools (Li et al., 2009). PCR duplicates were removed using the MarkDuplicates tool in Picard (1.140, http://broadinstitute.github.io/picard/). The number of reads mapping specifically to each genome and to the plasmids were retrieved from the bam files using samtools idxstats. An endogenous based score was used to assess the potential of the sample being ‘positive’ for *Y. pestis*. It was calculated as follows:

$$\frac{(YPS - \max(YS))}{M}$$

where YPS is the number of reads specifically mapping to *Y. pestis*; YS is the maximum number of reads mapping specifically to a *Yersinia* species with the exception of *Y. pestis* and M is the total number of merged reads in the sample. By using the maximum number of reads mapping to another species of the genus *Yersinia*, the score takes in account different source of contamination other than *Y. pseudotuberculosis*.

Genome reconstruction and authentication

The 'strong positives' identified during the screening process were deep shotgun sequenced on an Illumina HiSeq 4000/NextSeq 500 (See SI Methods). All samples were processed with the EAGER pipeline (Peltzer et al., 2016). Sequencing quality for each sample was evaluated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adaptors clipped using ClipAndMerge module in EAGER. For paired-end data, the reads were also merged with ClipAndMerge and only the merged reads were kept for further analysis.

Due to variability in the laboratory preparation and sequencing strategies, the sequencing reads for each sample were treated as follows:

- Gyvakarai1: two HiSeq lanes and one Next-Seq run paired-end of the non-UDG treated library were combined and reads mapped to *Y. pestis* CO92 reference with BWA aln (-l 16, -n 0.01, referred as ancient parameters). Reads with mapping quality scores lower than 37 were filtered out as above. PCR duplicates were removed with MarkDuplicates. MapDamage (v2.0, Jónsson et al., 2013) was used to calculate damage plots. Coverage was calculated with Qualimap (v2.2, Okonechnikov et al., 2016).
- KunilaII: UDG and the non-UDG libraries were sequenced in 2 HiSeq pair-end lanes and processed separately until calculation of the coverage. The non-UDG treated libraries were mappes with ancient parameters while the UDG treated library reads were mapped with more stringent parameters (-l 32, -n 0.1, referred as UDG parameters). Reads with mapping qualities less than 37 were filtered out and duplicates were removed with Markduplicates as before. The non-UDG bam file was used to calculate damage plots as indicated above. After duplicate removal, the UDG- and non-UDG treated BAM files were merged together and used to calculate the coverage as above.
- Post6 and 1343UnTal85: the half-UDG treated libraries were sequenced in two HiSeq lanes and two different runs were performed. For the first run, reads without clipping were used to retain miscoding lesions indicative of aDNA. BWA aln was used for mapping with ancient parameters (-l 16 and -n 0.01). Reads with mapping qualities lower

than 37 were filtered and PCR duplicates were removed with Markduplicates as above. Coverage and damage plots were calculated as above. After clipping the last two bases with FastX_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/index.html), potentially affected by damage, the samples were mapped with UDG parameters.

SNP calling & Phylogenetic analysis

Prior to SNP calling in order to avoid false SNP calling due to aDNA damage, the damaged sites in the non-UDG treated samples were downscaled using MapDamage(v2.0, Jónsson et al., 2013), as performed in previous analysis (Rasmussen et al., 2015). For the UDG-half data, the files with the two last bases clipped, hence removing the damage signal, and mapped with UDG parameters were used for SNP calling (see above).

SNP calling was performed with GATK UnifiedGenotyper (Van der Auwera et al., 2013) in EAGER[43] with default parameters and the 'EMIT_ALL_SITES' output mode.

VCF files of the new ancient samples, along with the two complete genomes from Rasmussen et al. (2015), the Black Death (Bos et al., 2011), Justinianic Plague (Feldman et al., 2016), Bolgar (Spyrou et al., 2016) and Observance (Bos et al., 2016) genomes, were combined with a curated dataset of 130 modern genomes (Cui et al., 2013) in addition to 11 samples from the Former Soviet Union (Zhgenti et al., 2015) and 19 draft genomes of *Y. pestis* subsp. *microtus* strains (Kislichkina et al., 2015).

The VCF files were processed with an in-house program (*MultiVCFAnalyser*) that produced a SNP table and an alignment file containing all variable positions in the dataset, in respect to the reference *Y. pestis* CO92. In order to call a SNP a minimum genotyping quality (GATK) of 30 was required, with a minimum coverage of 3X, and with a minimal allele frequency of 90% for an homozygous call. No heterozygous calls were included in the output files.

The SNP alignment was curated by removing all alignment columns with missing data (complete deletion). The curated SNP alignment was then used to compute NJ and MP trees with MEGA6 (Tamura et al., 2013) and a ML tree using PhyML 3.0 (Guindon et al., 2010) with the GTR model used in previous *Y. pestis* work (Cui et al., 2013; Rasmussen et al., 2015), with 4 gamma categories and the best of NNI and SPR as tree branch optimization. The specific positions from *Y. pseudotuberculosis* were removed from the analysis to improve the visual resolution of the tree.

Dating analysis

The SNP alignment after complete deletion was used for molecular dating using BEAST 1.8.2(Drummond et al., 2012). The modern sample 0.PE3, also called Angola, was removed from the dataset due to its long phylogenetic branch.

For tip dating, all modern genomes were set an age of 0. The dates of the ancient samples presented in this study plus the two complete genomes from Rasmussen et al. (2015) were recalibrated with Calib 7.1 (http://calib.qub.ac.uk/calib/) to the IntCal13 calibration curve. The ancient samples were given the median calibrated probability as their age, and the 2 sigma interval was used as the boundaries for a uniform prior sampling (Supplementary Table 5). The dates published for previous historical genomes were transformed to cal BP assuming 1950 as age 0 and given the mean as the age with the interval as the boundaries of a prior uniform distribution: Black Death 603 (602-604, Bos et al., 2011); Observance 229 (228-230, Bos et al., 2016), Bolgar 569 (550-588, Spyrou et al., 2016) and Justinian 1453 (1382-1524, Feldman et al., 2016).

The molecular clock was tested and rejected using MEGA6. Therefore, we followed previous work and used an uncorrelated relaxed clock with lognormal distribution (Cui et al., 2013; Rasmussen et al., 2015) with the substitution model GTR+G4. Tree model was set up to coalesce assuming a constant population size and a rooted ML tree was provided as a starting tree. Two independent 1,000,000,000 MCMC chains were computed sampling every 2,000 steps. The two chains were then combined using LogCombiner from BEAST 1.8.2 (Drummond et al., 2012) with a 10 percent burn-in (10,000,000 steps per chain). The ESS of the posterior, prior, treeModel.rootHeight, tMRCA_allpestis are 2,846, 2,500, 870 and 3,852 respectively. The trees files for the 2 chains were combined with LogCombiner with 10,000 of burning and resamples ever 4,000 steps giving a total number of 45,000 trees, that were used to produce a Maximum Clade Credibility tree using TreeAnnotator from BEAST 1.8.2 (Drummond et al., 2012).

SNP effect analysis and virulence factors analysis

The SNP table from *MultiVCFAnalyzer* was provided to *SnpEff* (Cingolani et al., 2012) and the effect of the SNPs within genes present in the dataset was evaluated. Additionally the SNP table was manually assessed for possible homoplasies.

For the virulence factors, coverage was calculated for each region with using bedtools (Quinlan and Hall, 2010) and plotted using the package ggplot2 (Wickham, 2009) in R (R Development Core Team, 2008). Additionally, ureD was manually explored for SNPs manually using IGV (Thorvaldsdóttir et al., 2013).

Indel analysis

The samples including the two complete Bronze Age genomes (Rasmussen et al., 2015) were mapped against Y. pseudotuberculosis IP 32953 with bwa with ancient parameters (-n 0.01, -l 16). The modern genomes from branch 0 (0.PE7, 0.PE2, 0.PE3 and 0.PE4), *Y. pestis* CO92 and *Y. pestis* KIM10 were bioinformatically cut in 100 bp fragments with 1bp tiling and mapped to Y. pseudotuberculosis reference using bwa with UDG parameters (-n 0.1, -l 32). The non-covered regions were extracted using the bedtools genomecov function. Missing regions larger than 1kb were comparatively explored in order to identify indels. Using bedtools intersect function, we extracted regions missing in the Neolithic genomes and present in the modern ones and also the regions missing in the modern ones but still present in the Neolithic genomes. The results were check by manual inspection in IGV (Thorvaldsdóttir et al., 2013).

References

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P.D., Dąbrowski, P., Duffy, P.R., Ebel, A.V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Price, T.D., Saag, L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V.I., Szeverényi, V., Tóth, G., Trifanova, S.V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Population genomics of Bronze Age Eurasia. Nature 522, 167–172. doi:10.1038/nature14507

Benedictow, O.J., 2004. The Black Death, 1346-1353: The Complete History. Boydell Press.

Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M., Golding, G.B., Dutour, O., Keim, P., Wagner, D.M., Holmes, E.C., Krause, J., Poinar, H.N., 2016. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. eLife 5, e12994. doi:10.7554/eLife.12994

Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., Wood, J., Earn, D.J.D., Herring, D.A., Bauer, P., Poinar, H.N., Krause, J., 2011. A draft genome of Yersinia pestis from victims of the Black Death. Nature 478, 506–510. doi:10.1038/nature10549

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. PNAS 104, 14616–14621. doi:10.1073/pnas.0704665104

Chouikha, I., Hinnebusch, B.J., 2014. Silencing urease: A key evolutionary step that facilitated the adaptation of Yersinia pestis to the flea-borne transmission route. PNAS 111, 18709–18714. doi:10.1073/pnas.1413209111

Chouikha, I., Hinnebusch, B.J., 2012. Yersinia–flea interactions and the evolution of the arthropod-borne transmission route of plague. Current Opinion in Microbiology, Ecology and industrial microbiology/Special section: Microbial proteomics 15, 239–246. doi:10.1016/j.mib.2012.02.003

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6, 80–92. doi:10.4161/fly.19695

Cohn JR, S.K., 2008. 4 Epidemiology of the Black Death and Successive Waves of Plague. Med Hist Suppl 74–100.

Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, J., Yao, S., Rakin, A., Li, Y., Falush, D., Balloux, F., Achtman, M., Song, Y., Wang, J., Yang, R., 2013. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. PNAS 110, 577–582. doi:10.1073/pnas.1205750110

Derbise, A., Carniel, E., 2014. YpfΦ: a filamentous phage acquired by Yersinia pestis. Front Microbiol 5. doi:10.3389/fmicb.2014.00701

Derbise, A., Chenal-Francisque, V., Pouillot, F., Fayolle, C., Prévost, M.-C., Médigue, C., Hinnebusch, B.J., Carniel, E., 2007. A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. Molecular Microbiology 63, 1145–1157. doi:10.1111/j.1365-2958.2006.05570.x

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29, 1969–1973. doi:10.1093/molbev/mss075

Du, Y., Rosqvist, R., Forsberg, Å., 2002. Role of Fraction 1 Antigen of Yersinia pestis in Inhibition of Phagocytosis. Infect. Immun. 70, 1453–1460. doi:10.1128/IAI.70.3.1453-1460.2002

Eisen, R.J., Bearden, S.W., Wilder, A.P., Montenieri, J.A., Antolin, M.F., Gage, K.L., 2006. Early-phase transmission of Yersinia pestis by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. PNAS 103, 15380–15385. doi:10.1073/pnas.0606831103

Eisen, R.J., Dennis, D.T., Gage, K.L., 2015. The Role of Early-Phase Transmission in the Spread of Yersinia pestis. J Med Entomol 52, 1183–1192. doi:10.1093/jme/tjv128

Feldman, M., Harbeck, M., Keller, M., Spyrou, M.A., Rott, A., Trautmann, B., Scholz, H.C., Päffgen, B., Peters, J., McCormick, M., Bos, K., Herbig, A., Krause, J., 2016. A high-coverage Yersinia pestis Genome from a 6th-century Justinianic Plague Victim. Mol Biol Evol msw170. doi:10.1093/molbev/msw170

Felek, S., Lawrenz, M.B., Krukonis, E.S., 2008. The Yersinia pestis autotransporter YapC mediates host cell binding, autoaggregation and biofilm formation. Microbiology 154, 1802–1812. doi:10.1099/mic.0.2007/010918-0

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol 59, 307–321. doi:10.1093/sysbio/syq010

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R.G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S.L., Risch, R., Rojo Guerra, M.A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K.W., Reich, D., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211. doi:10.1038/nature14317

Haiko, J., Kukkonen, M., Ravantti, J.J., Westerlund-Wikström, B., Korhonen, T.K., 2009. The Single Substitution I259T, Conserved in the Plasminogen Activator Pla of Pandemic Yersinia pestis Branches, Enhances Fibrinolytic Activity. J. Bacteriol. 191, 4758–4766. doi:10.1128/JB.00489-09

Hinnebusch, B.J., Fischer, E.R., Schwan, T.G., 1998. Evaluation of the Role of the Yersinia pestis Plasminogen Activator and Other Plasmid-Encoded Factors in Temperature-Dependent Blockage of the Flea. J Infect Dis. 178, 1406–1415. doi:10.1086/314456

Hinnebusch, B.J., Rudolph, A.E., Cherepanov, P., Dixon, J.E., Schwan, T.G., Forsberg, Å., 2002. Role of Yersinia Murine Toxin in Survival of Yersinia pestis in the Midgut of the Flea Vector. Science 296, 733–735. doi:10.1126/science.1069972

Hinnebusch, J., Cherepanov, P., Du, Y., Rudolph, A., Dixon, J.D., Schwan, T., Forsberg, Å., 2000. Murine toxin of Yersinia pestis shows phospholipase D activity but is not required for virulence in mice. International Journal of Medical Microbiology 290, 483–487. doi:10.1016/S1438-4221(00)80070-3

Johnson, T.L., Hinnebusch, B.J., Boegler, K.A., Graham, C.B., MacMillan, K., Montenieri, J.A., Bearden, S.W., Gage, K.L., Eisen, R.J., 2014. Yersinia murine toxin is not required for early-phase transmission of Yersinia pestis by Oropsylla montana (Siphonaptera: Ceratophyllidae) or Xenopsylla cheopis (Siphonaptera: Pulicidae). Microbiology 160, 2517–2525. doi:10.1099/mic.0.082123-0

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684. doi:10.1093/bioinformatics/btt193

Kislichkina, A.A., Bogun, A.G., Kadnikova, L.A., Maiskaya, N.V., Platonov, M.E., Anisimov, N.V., Galkina, E.V., Dentovskaya, S.V., 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. doi:10.1128/genomeA.01342-15

Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: An information aesthetic for comparative genomics. Genome Res. doi:10.1101/gr.092759.109

Lathem, W.W., Price, P.A., Miller, V.L., Goldman, W.E., 2007. A plasminogen-activating protease specifically controls the development of primary pneumonic plague. Science 315, 509–513. doi:10.1126/science.1137195

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Ochman, H., Moran, N.A., 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292, 1096–1099.

Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294. doi:10.1093/bioinformatics/btv566

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., Nieselt, K., 2016. EAGER: efficient ancient genome reconstruction. Genome Biology 17, 60. doi:10.1186/s13059-016-0918-z

Perry, R.D., Fetherston, J.D., 1997. Yersinia pestis--etiologic agent of plague. Clin. Microbiol. Rev. 10, 35–66.

Pouillot, F., Fayolle, C., Carniel, E., 2008. Characterization of Chromosomal Regions Conserved in Yersinia pseudotuberculosis and Lost by Yersinia pestis. Infect. Immun. 76, 4592–4599. doi:10.1128/IAI.00568-08

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. doi:10.1093/bioinformatics/btq033

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

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.O., Nielsen, H.B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M.V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Pontén, T., Foley, R.A., Lahr, M.M., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. Cell 163, 571–582. doi:10.1016/j.cell.2015.10.009

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

Ronald Barrett, Christopher W. Kuzawa, Thomas McDade, Armelagos, and G.J., 1998. EMERGING AND RE-EMERGING INFECTIOUS DISEASES: The Third Epidemiologic Transition. Annual Review of Anthropology 27, 247–271. doi:10.1146/annurev.anthro.27.1.247

Russell, J.C., 1968. That earlier plague. Demography 5, 174–184. doi:10.1007/BF03208570

Schuenemann, V.J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J.D., White, W., Krause, J., Poinar, H.N., 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. PNAS 108, E746–E752. doi:10.1073/pnas.1105107108

Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E., Simonet, M., 2001. Silencing and Reactivation of Urease inYersinia pestis Is Determined by One G Residue at a Specific Position in the ureD Gene. Infect. Immun. 69, 170–176. doi:10.1128/IAI.69.1.170-176.2001

Sebbane, F., Jarrett, C.O., Gardner, D., Long, D., Hinnebusch, B.J., 2006. Role of the Yersinia pestis plasminogen activator in the incidence of distinct septicemic and bubonic forms of flea-borne plague. Proc Natl Acad Sci U S A 103, 5526–5530. doi:10.1073/pnas.0509544103

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., Gazimzyanov, I.R., Nurgaliev, D.K., Herbig, A., Bos, K.I., Krause, J., 2016. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. Cell Host & Microbe 19, 874–881. doi:10.1016/j.chom.2016.05.012

Stenseth, N.C., Atshabar, B.B., Begon, M., Belmain, S.R., Bertherat, E., Carniel, E., Gage, K.L., Leirs, H., Rahalison, L., 2008. Plague: Past, Present, and Future. PLOS Medicine 5, e3. doi:10.1371/journal.pmed.0050003

Stockhammer, P.W., Massy, K., Knipper, C., Friedrich, R., Kromer, B., Lindauer, S., Radosavljević, J., Wittenborn, F., Krause, J., 2015. Rewriting the Central European Early Bronze Age Chronology: Evidence from Large-Scale Radiocarbon Dating. PLOS ONE 10, e0139705. doi:10.1371/journal.pone.0139705

Sun, Y.-C., Jarrett, C.O., Bosio, C.F., Hinnebusch, B.J., 2014. Retracing the Evolutionary Path that Led to Flea-borne Transmission of Yersinia pestis. Cell Host Microbe 15, 578–586. doi:10.1016/j.chom.2014.04.003

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30, 2725–2729. doi:10.1093/molbev/mst197

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14, 178–192. doi:10.1093/bib/bbs017

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline, in: Current Protocols in Bioinformatics. p. 43:11.10.1-11.10.33.

Vandkilde, H., 2016. Bronzization: The Bronze Age as Pre-Modern Globalization. Praehistorische Zeitschrift 91, 103–123. doi:10.1515/pz-2016-0005

Vetter, S.M., Eisen, R.J., Schotthoefer, A.M., Montenieri, J.A., Holmes, J.L., Bobrov, A.G., Bearden, S.W., Perry, R.D., Gage, K.L., 2010. Biofilm formation is not required for early-phase transmission of Yersinia pestis. Microbiology 156, 2216–2225. doi:10.1099/mic.0.037952-0

Wagner, D.M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J.W., Enk, J., Birdsell, D.N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M., Golding, B., Riehm, J.M., Earn, D.J.D., DeWitte, S., Rouillard, J.-M., Grupe, G., Wiechmann, I., Bliska, J.B., Keim, P.S., Scholz, H.C., Holmes, E.C., Poinar, H., 2014. Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis. The Lancet Infectious Diseases 14, 319–326. doi:10.1016/S1473-3099(13)70323-2

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

Zhgenti, E., Johnson, S.L., Davenport, K.W., Chanturia, G., Daligault, H.E., Chain, P.S., Nikolich, M.P., 2015. Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus Region. Genome Announc. 3, e01030-15. doi:10.1128/genomeA.01030-15

Zietz, B.P., Dunkelberg, H., 2004. The history of the plague and the research on the causative agent Yersinia pestis. International Journal of Hygiene and Environmental Health 207, 165–178. doi:10.1078/1438-4639-00259

Zimbler, D.L., Schroeder, J.A., Eddy, J.L., Lathem, W.W., 2015. Early emergence of Yersinia pestis as a severe respiratory pathogen. Nat Commun 6. doi:10.1038/ncomms8487

Acknowledgements

We thank Corina Knipper, Ernst Pernicka, Stefanie Metz, Fabian Wittenborn, Stephan Schiffels, Joris Peters, Michaela Harbeck, and all the members of the Archaeogenetics Department of the Max Planck Institute for the Science of Human History for helpful discussion and suggestions. We thank Annette Günzel for graphical support. We thank Isil Kucukkalipci and Franziska Göhringer for technical support in the lab. We thank Prof. Dr. Joachim Wahl and Dr. Gunita Zatina and Prof. Andrejs Vasks for kindly providing the Althausen samples and the Latvian samples, respectively, used in this study. We thank James A. Fellows Yates for proof-reading the manuscript.
The genetic and archaeological research on the human individuals from the Augsburg region was financed by the Heidelberg Academy of Science within the WIN project “Times of Upheaval: Changes of Society and Landscape at the Beginning of the Bronze Age”. This work was also supported by the Max Planck Society and the European Research Council starting grant APGREID (to J.K.)

Authors contribution

J.K., A.H. and A.A.V. conceived the study. K.M., R.A., M.D., R.J., M.T., P.W.S. provided the samples and performed archaeological assessment. A.M., S.P., M.F., A.A.V. performed laboratory work. A.A.V., A.H., M.S. and J.K. analysed the data. A.A.V, A.H., J.K., P.W.S., K.I.B., W.H. and A.M. wrote the manuscript with contributions from all co-authors. All authors read and approved the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

Data availability

Data is available upon request

Figures

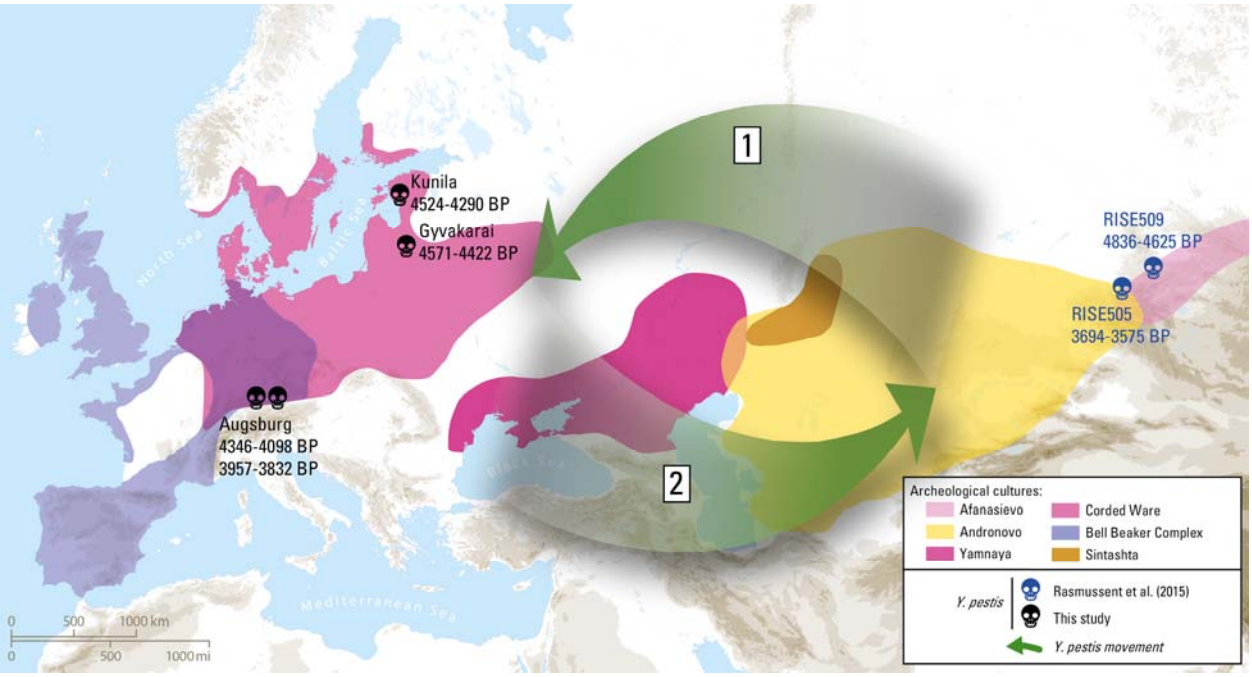

Figure 1: Map of proposed *Yersinia pestis* circulation throughout Eurasia. [1] Entrance of *Y. pestis* into Europe from Central Eurasia with the expansion of Yamnaya pastoralists around 4,800 years ago. [2] Circulation of *Y. pestis* back into the Altai from Europe. Only complete genomes are shown.

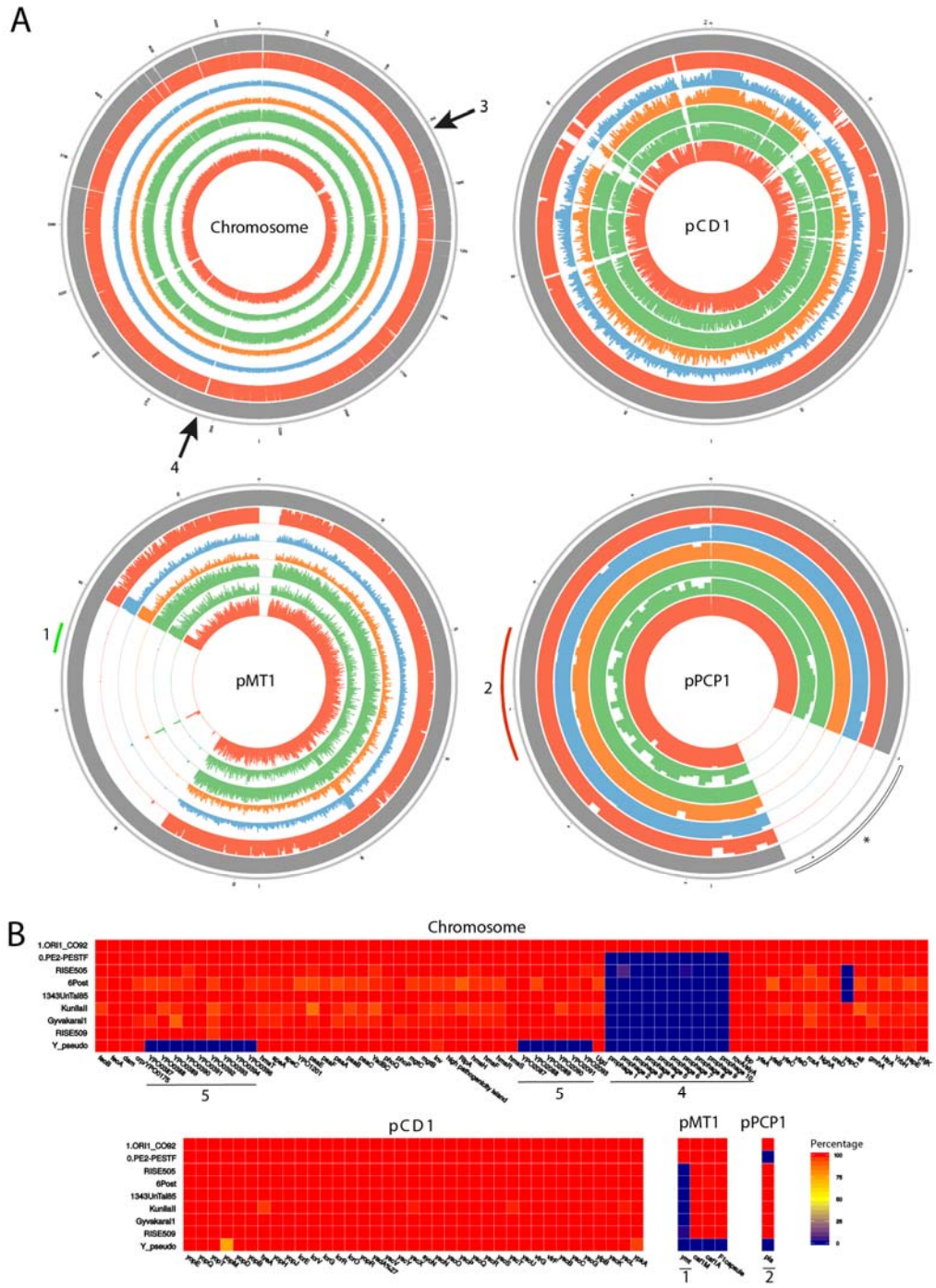

Figure 2: A) Average coverage plot for the chromosome and plasmids of *Yersinia pestis*, from the outer ring to the inner ring: *Y. pestis* CO92 (NC_003143.1, reference), RISE509, Gyvakarai1, KunilaII, 6Post, 1343UnTal85 and RISE505. Colours correspond to the regions where the genomes were recovered from: Altai region (red), Gyvakarai, Lithuania (blue), Kunila, Estonia (orange), Augsburg, Germany (green). The average depth of

coverage was calculated for 1kb regions for the chromosome and 100bp for the plasmids, each ring represents a maximum of 20X coverage. The figure was generated with Circos (Krzywinski et al., 2009). B) Percentage covered of virulence factors located in *Yersinia pestis* chromosome and plasmids, plotted with in R using ggplot2 package. [1] ymt gene, [2] pla, [3] deletion of flagelin genes, [4] filamentous prophage YpfΦ, [5] *Y. pestis*-specific genes, [*] region mask in pPCP1 due to high similarity to expression vectors during enzyme production (Schuenemann et al., 2011).

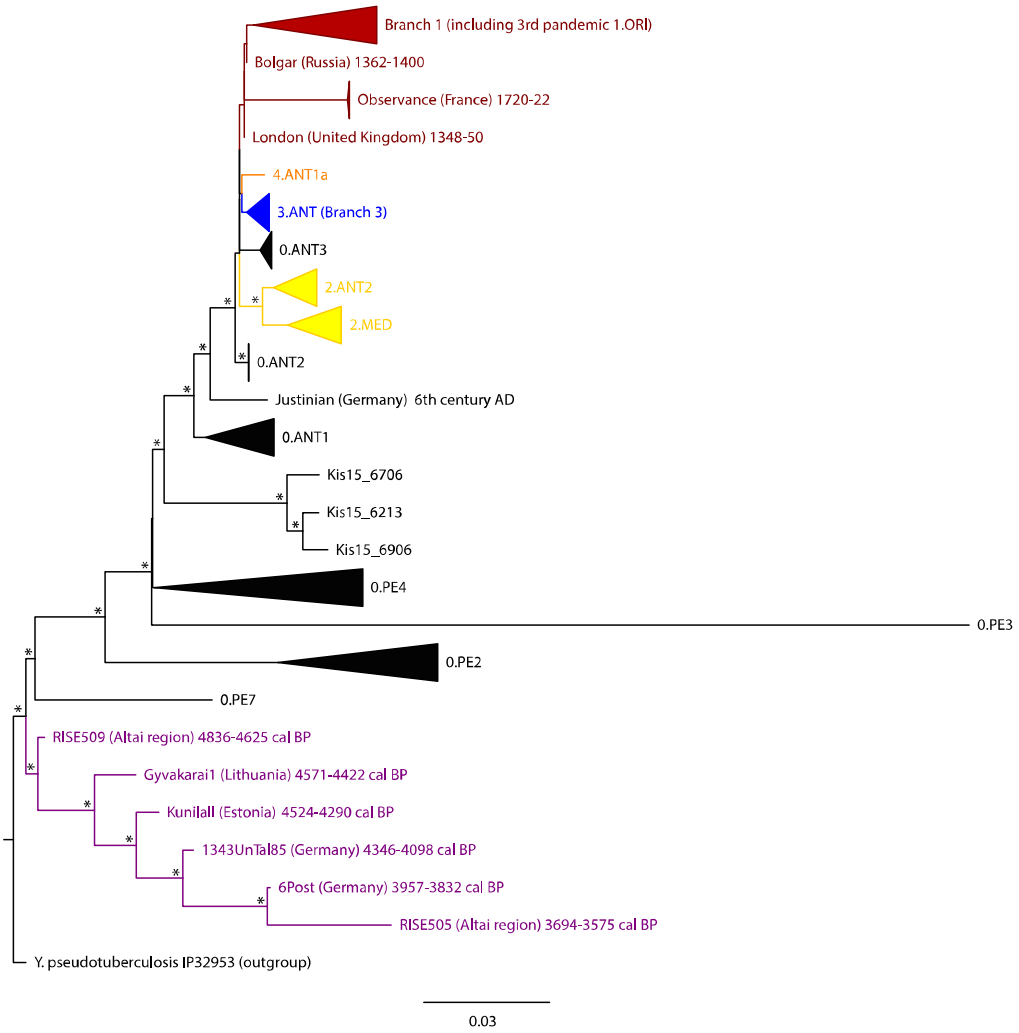

Figure 3: Maximum Likelihood tree of all *Yersinia pestis* genomes including 1,867 SNPs positions with complete deletion. Nodes with support equal or higher than 95% are marked with an asterisk. The colours represent different branches in the *Y. pestis* phylogeny: branch 0 (black), branch 1 (red), branch2 (yellow), branch 3 (blue), branch 4 (orange) and LNBA *Y. pestis* branch (purple). *Y. pseudotuberculosis*-specific SNPs were excluded from the tree for representative matters.

Tables

Table 1: Genomes from the NCBI (RefSeq/Nucleotide) database, used in the multi-species reference panel for screening for *Y. pestis* aDNA.

Species name	Strain	NCBI Accession number
Y. pestis	CO92	NC_003143.1
Y. pseudotuberculosis	IP 32953	NC_006155.1
Y. enterocolitca	subsp. enterocolitica 8081	NC_008800.1
Y. aldovae	ATCC 35236	NZ_ACCB01000210.1
Y. bercovieri	ATCC 43970	NZ_AALC02000229.1
Y. frederiksenii	ATCC 33641	NZ_AALE02000161.1
Y. intermedia	ATCC 29909	NZ_AALF02000123.1
Y. kristensenii	ATCC 33638	NZ_ACCA01000153.1
Y. mollaretii	ATCC 43969	NZ_AALD02000179.1
Y. rohdei	ATCC 43380	NZ_ACCD01000141.1
Y. ruckeri	ATCC 29473	NZ_ACCC01000174.1

Table 2: Statistic of the *Y. pestis* genome reconstruction

Sample	Site	Country	Dating (Median cal BP)	Clipped, merged and quality-filtered reads before mapping	Unique reads mapping to *Y. pestis* reference	Endogenous DNA (%)	Mean Coverage	Coverage (%)				
								>=1X	>=2X	>=3X	>=4X	>=5X
Gyvakarai1	Gyvakarai	Lithuania	4427	1,021,452,137	473,207	0.05	5.2245	94.07	90.96	84.12	73.1	59.1
KunilaII	Kunila	Estonia	4203	379,155,741	546,243	0.16	5.5418	92.48	86.65	77.58	66.49	54.77
1343UnTal85	Augsburg	Germany	3873	1,410,707,182	1,375,550	0.14	11.9999	93.75	93.49	93.16	92.57	91.47
6Post	Augsburg	Germany	3635	480,205,842	698,670	0.17	6.0027	90.76	84.45	75.67	65.59	55.51