A genomic and historical synthesis of plague in 18th century Eurasia

Meriam Guellil[a,b,1], Oliver Kersten[a], Amine Namouchi[a,c], Stefania Luciani[d], Isolina Marota[d], Caroline A. Arcini[e], Elisabeth Iregren[f], Robert A. Lindemann[g], Gunnar Warfvinge[h], Lela Bakanidze[i], Lia Bitadze[j], Mauro Rubini[k,l], Paola Zaio[k], Monica Zaio[k], Damiano Neri[k], N. C. Stenseth[a,1], and Barbara Bramanti[a,b,m,1]

[a]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, 0316 Oslo, Norway; [b]Institute of Genomics, University of Tartu, 51010 Tartu, Estonia; [c]Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1430 Ås, Norway; [d]Laboratory of Molecular Archaeo-Anthropology/ancientDNA, School of Biosciences and Veterinary Medicine, University of Camerino, 62032 Camerino, Italy; [e]Arkeologerna, National Historical Museums, SE-226 60 Lund, Sweden; [f]Department of Archaeology and Ancient History, Lund University, SE-221 00, Lund, Sweden; [g]School of Dentistry, University of California Los Angeles, CA 90024; [h]Faculty of Odontology, Malmö University, SE-205 06 Malmö, Sweden; [i]National Center for Disease Control and Public Health, 0198 Tbilisi, Georgia; [j]Anthropological Studies of the Institute of History and Ethnology, Ivane Javakhishvili Tbilisi State University, 0177 Tbilisi, Georgia; [k]Anthropological Service of Soprintendenza Archeologia Belle Arti e Paesaggio per le province di Frosinone, Latina e Rieti (Lazio), Ministry of Cultural Heritage and Activities, 00192 Rome, Italy; [l]Department of Archaeology, University of Foggia, 71122 Foggia, Italy; and [m]Department of Biomedical and Specialty Surgical Sciences, Faculty of Medicine, Pharmacy and Prevention, University of Ferrara, 44121 Ferrara, Italy

Contributed by N. C. Stenseth, September 2, 2020 (sent for review May 29, 2020; reviewed by Guido Alfani and Ludovic Orlando)

Plague continued to afflict Europe for more than five centuries after the Black Death. Yet, by the 17th century, the dynamics of plague had changed, leading to its slow decline in Western Europe over the subsequent 200 y, a period for which only one genome was previously available. Using a multidisciplinary approach, combining genomic and historical data, we assembled *Y. pestis* genomes from nine individuals covering four Eurasian sites and placed them into an historical context within the established phylogeny. CHE1 (Chechnya, Russia, 18th century) is now the latest Second Plague Pandemic genome and the first non-European sample in the post-Black Death lineage. Its placement in the phylogeny and our synthesis point toward the existence of an extra-European reservoir feeding plague into Western Europe in multiple waves. By considering socioeconomic, ecological, and climatic factors we highlight the importance of a noneurocentric approach for the discussion on Second Plague Pandemic dynamics in Europe.

Yersinia pestis | ancient DNA | aDNA | pathogen | plague

Yersinia pestis, the etiological agent of plague, has been shown to infect humans since prehistory (1–4) and is responsible for some of the deadliest pandemics to have ever affected European populations. The most prominent pandemic is the Second Plague Pandemic, which raged in Europe and beyond between the 14th and the 19th century of the common era (CE). However, while its narrative is often dominated by the Black Death epidemic (1346 to 1353), which is estimated to have killed 30 to 60% of the European population within a few years, outbreaks of plague were common in Europe until the early 19th century while recurring several decades beyond that in other parts of the world (5–8). Strikingly, around the middle of the 17th century, a major shift in plague dynamics occurred, after which only a single continental-scale epidemic was documented in Western Europe during the early 1700s (5–7). More localized outbreaks occurred until the early 19th century (e.g., Malta) (5, 9).

The exact dynamics of introduction and persistence of plague in Europe during the Second Plague Pandemic have been under much scrutiny in the literature (10–12). Fortunately, the final centuries of plague in Europe have left us with an unrivalled wealth of historical sources, documenting afflicted populations across the continent and contemporaries striving to understand and fight the disease (13). However, while previous studies (10–12, 14, 15) have yielded a number of Second Plague Pandemic genomes, our genomic knowledge of the 18th century was limited to the genomes isolated from Marseille l'Observance (OBS) (10), and the post-Black Death lineage was lacking any non-European genomes altogether. With this study, we aim to address this lack of data by contributing two genomes from this period. A genome from the Caucasus region (Maist, Chechnya, Russia) dating to the 18th century is of particular note, as it questions the theory of an exclusively European origin of plague following the Black Death and places it farther East. Our data address a gap in knowledge that has not been considered in previous eurocentric studies (10, 14) and point toward a much more complex mechanism of plague dynamics than previously proposed for the Second Plague Pandemic. Particularly, the addition of the Caucasian strain CHE1 and non-European historical data has major ramifications for our understanding of the last centuries of plague in Europe. The historical data, while scarce and less detailed, exist and should be incorporated into the investigation of a pandemic, which the plague remained well beyond the Black Death expansion. The second new 18th century genome can be attributed to the last wave of plague to hit Scandinavia in the 1710s. Additionally, we sequenced six 17th century strains from San Procolo a Naturno and a 14th century strain from Collalto Sabino (Italy).

Significance

The spread and evolution of plague have been under debate in the past few years. However, very little is known of the dynamics of the plague pathogen, *Yersinia pestis*, during the last phase of the Second Plague Pandemic in Europe (18th and 19th century). We present nine ancient *Y. pestis* genomes from the Second Plague Pandemic. CHE1 is the first Second Plague Pandemic genome from the Caucasus region, an area that houses plague wildlife reservoirs to this day, making it a key strain to help elucidate the origin of Medieval and Early Modern plague. Our study documents the importance of a noneurocentric approach to historical plague dynamics and proposes an origin of plague introductions outside of Europe.

Author contributions: M.G. and B.B. designed research; M.G., O.K., and A.N. analyzed the data; M.G., O.K., S.L., and I.M. performed laboratory work; A.N. designed and generated the phylogeny; M.G., O.K., A.N., N.C.S., and B.B. wrote the paper with contributions from S.L., I.M., C.A.A., E.I., R.A.L., G.W., L. Bakanidze, L. Bitadze, M.R., P.Z., M.Z., and D.N.; and S.L., I.M., C.A.A., E.I., R.A.L., G.W., L. Bakanidze, L. Bitadze., M.R., P.Z., M.Z., and D.N. provided archaeological/osteological data and samples.

Reviewers: G.A., Bocconi University; and L.O., CNRS, Université Paul Sabatier.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: meriam.guellil.ac@gmail.com, n.c.stenseth@mn.uio.no, or barbara.bramanti@ibv.uio.no.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2009677117/-/DCSupplemental.

MICROBIOLOGY

ANTHROPOLOGY

Downloaded by guest on October 26, 2020

By combining our data with previously published *Y. pestis* genomes and historical sources, we present a genomic and historical synthesis of plague introductions in 18th century Europe.

Accordingly, we aimed to answer three key questions based on a multidisciplinary approach combining the data presented in this study and previously published genomic data with historical sources. Our focus is on the 18th century, which included Western Europe's last large-scale plague outbreaks during the Second Plague Pandemic. First, we address the potential location of the wildlife reservoir that gave rise to repeated epidemics within and around Europe following the Black Death, arguably the most disputed aspect of ancient DNA (aDNA) plague research. Second, we discuss the changes in plague dynamics observed by historians and how these changes could relate to available genomic data. Finally, we explore the end of plague in Western Europe and the significance of our data within this context.

Results

In this study, we describe nine genomes of *Y. pestis* (Fig. 1 and Dataset S1), isolated from skeletons recovered from four Eurasian sites and dated from the 14th to the 18th century (Fig. 2). Individual COL001 was recovered from the medieval cemetery of the church of S. Giovanni in Fistola by Collalto Sabino (Italy) (*SI Appendix*, Fig. S1). The skeleton, a young adult male, was

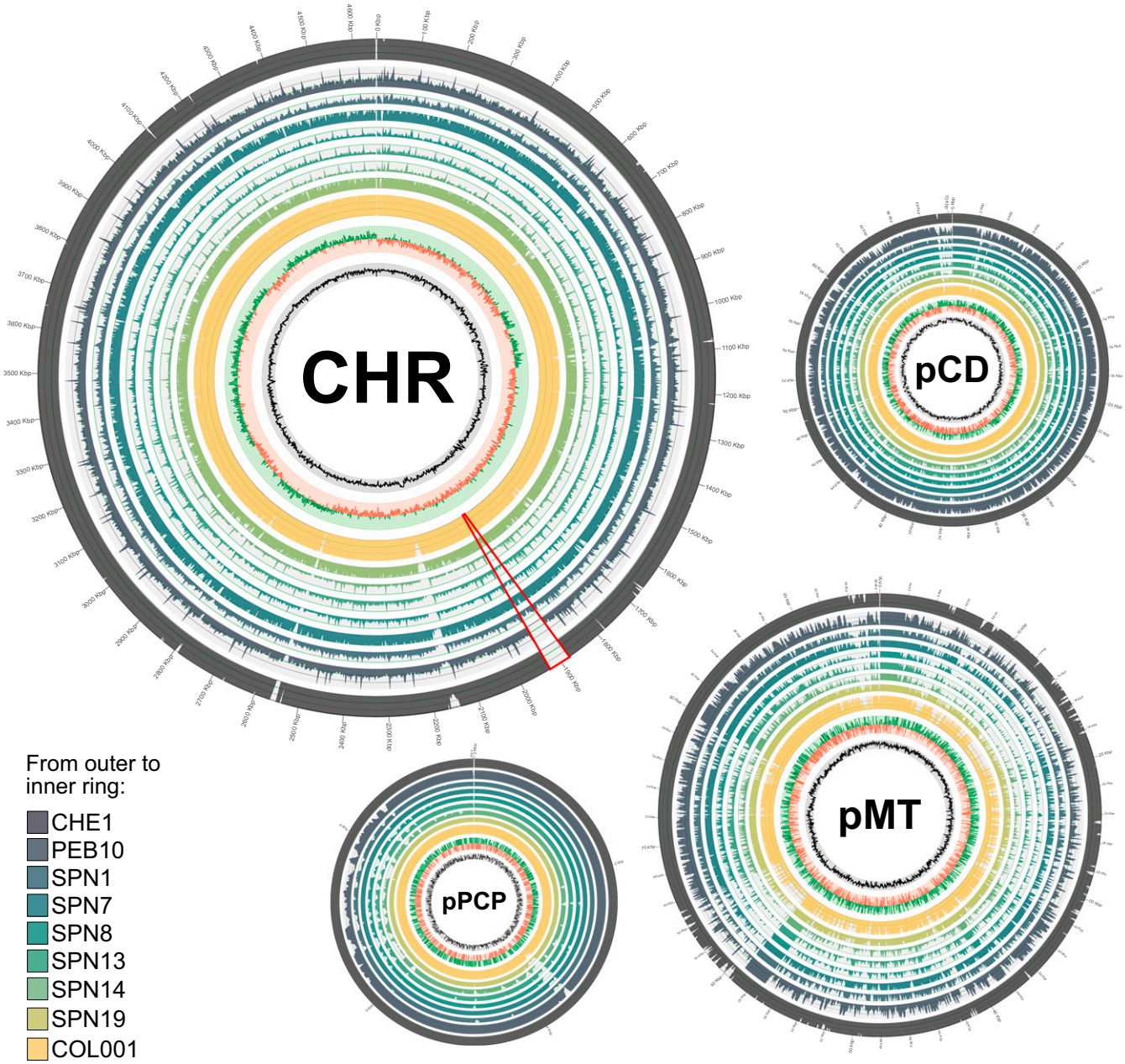

Fig. 1. Coverage plots for nine genomes to *Y. pestis* CO92. Plots represent the chromosome and each of the three CO92 plasmids (CHR: chromosome). Rings (from outer to inner ring) show coverage (rings 1 to 9), GC skew (ring 10), and GC content (ring 11, range: 30 to 70%). aDNA genomes are ordered as follows (from outer to inner ring): CHE1, PEB10, SPN1, SPN7, SPN8, SPN13, SPN14, SPN19, and COL001. Coverage cutoff for PEB10, CHE1, and COL001 is 15× and 5× for all SPN samples. Plots were created with Circos (73). The chromosomal plots were calculated in 2,000-bp windows, the plots for pMT and pPCP in 50-bp windows, and the plot for pCD in 10-bp windows. The 49-kbp deletion is marked in red on the chromosomal plot.

Downloaded by guest on October 26, 2020

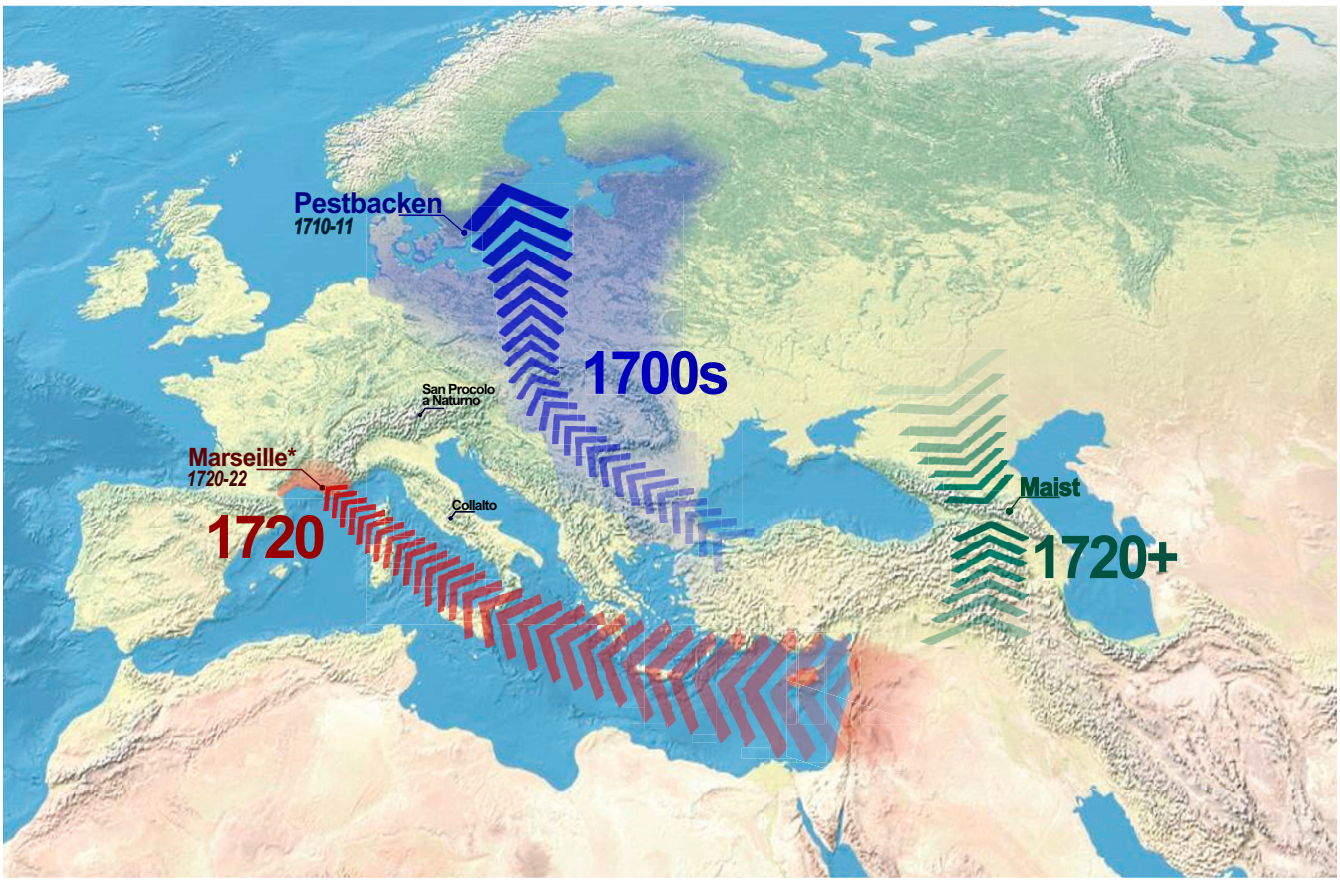

Fig. 2. Historically reconstructed introduction routes of *Y. pestis* for available 18th century genomes, consisting of multiple spatiotemporal waves. Locations shown and highlighted on the map are discussed in this study. Sites for which genomic data were published in previous studies are marked with an asterisk. Basemap is from Wikicommons.

recovered from a multiple burial. Six genomes were isolated from skeletons found in multiple mass graves at San Procolo a Naturno (Naturns, Italy) (SPN), dated to the 17th century. Individual PEB10/A975 was excavated at the site of Pestbacken, a plague cemetery originally situated on a meadow near Holje (Olofström. Blekinge, Sweden) (16, 17). The skeleton was identified as a male between the age of 20 and 25 y at the time of death (*SI Appendix*). Finally, individual CHE1/522 was excavated in Maist (Chechnya, Russia, 18th+ century) in 1962. The individual was part of an anthropological collection, composed of crania from Chechnya dating between the 16th and the 18th century (*SI Appendix*, Fig. S2).

We individually captured 17 double-stranded single indexed libraries: six from San Procolo (SPN samples), six from individual PEB10, and four libraries from individual CHE1 and one from COL001. The captured libraries were sequenced on an Illumina HiSEq. 2500 system (PE 125 bp) and mapped against the *Y. pestis* CO92 reference genome. Our chromosomal mappings yielded a mean depth of coverage between 3.22× and 57× for the worst (SPN14) and best (CHE1) sample, respectively (Dataset S1). The aligned reads also exhibited a misincorporation pattern characteristic of ancient sequences (*SI Appendix*, Figs. S3–S5). A comparison of edit distances from noncompetitive mappings to *Y. pestis* CO92 and *Y. pseudotuberculosis* IP32953 showed that our genomes have been correctly identified as *Y. pestis* (*SI Appendix*, Fig. S6).

Both PEB10 and CHE1 have a large deletion of ∼49 kbp (∼1,879,979 to 1,928,864 bp) situated downstream from an IS100 repeat system (Fig. 1). As this deletion has now been detected by two different research groups and multiple target enrichment designs (10, 18), and since our design (based on strain CO92) covered this region (5× tiling), we are confident that the missing sequence is not an artifact caused by our capture design.

We added our samples and newly published modern genomes to the phylogeny. The phylogeny includes 231 genomes, of which 181 are modern, 41 are historical, and 9 are prehistoric (Dataset S4). Of these, a total of 53 genomes stem from recently published isolates from Central Asia, the Caucasus region, and Russia (19–21). After comparing all 231 genomes, our analysis yielded a total of 3,917 single-nucleotide polymorphisms (SNPs) (Dataset S2). The constructed maximum-likelihood tree allowed the identification of 245 homoplastic sites (Dataset S5).

All ancient *Y. pestis* genomes described in this study are positioned in a subbranch of branch 1 originating from the Black Death lineage and populated by all available Second Plague Pandemic strains following the Black Death, except for pestis secunda strains (London6330, Bergen op Zoom, and Bolgar) (Fig. 3). Based on their phylogenetic location and deamination profiles, we could confidently validate the authenticity of our genomes as Second Plague Pandemic genomes. COL001 and the German genome MAN008 form separate lineages, which cluster at the base of the post-Black Death subbranch. According to historical sources (22), Collalto was hit by plague in 1363, which coincides well with the placement of COL001 in the phylogeny close to the genome MAN008 (Germany, 1283 to 1390). These two strains point toward the existence of two distinct lineages circulating in Europe during the *Pestis Secunda* (1357 to 1366). These strains were probably introduced in multiple waves from outside of Europe,

MICROBIOLOGY
ANTHROPOLOGY

Downloaded by guest on October 26, 2020

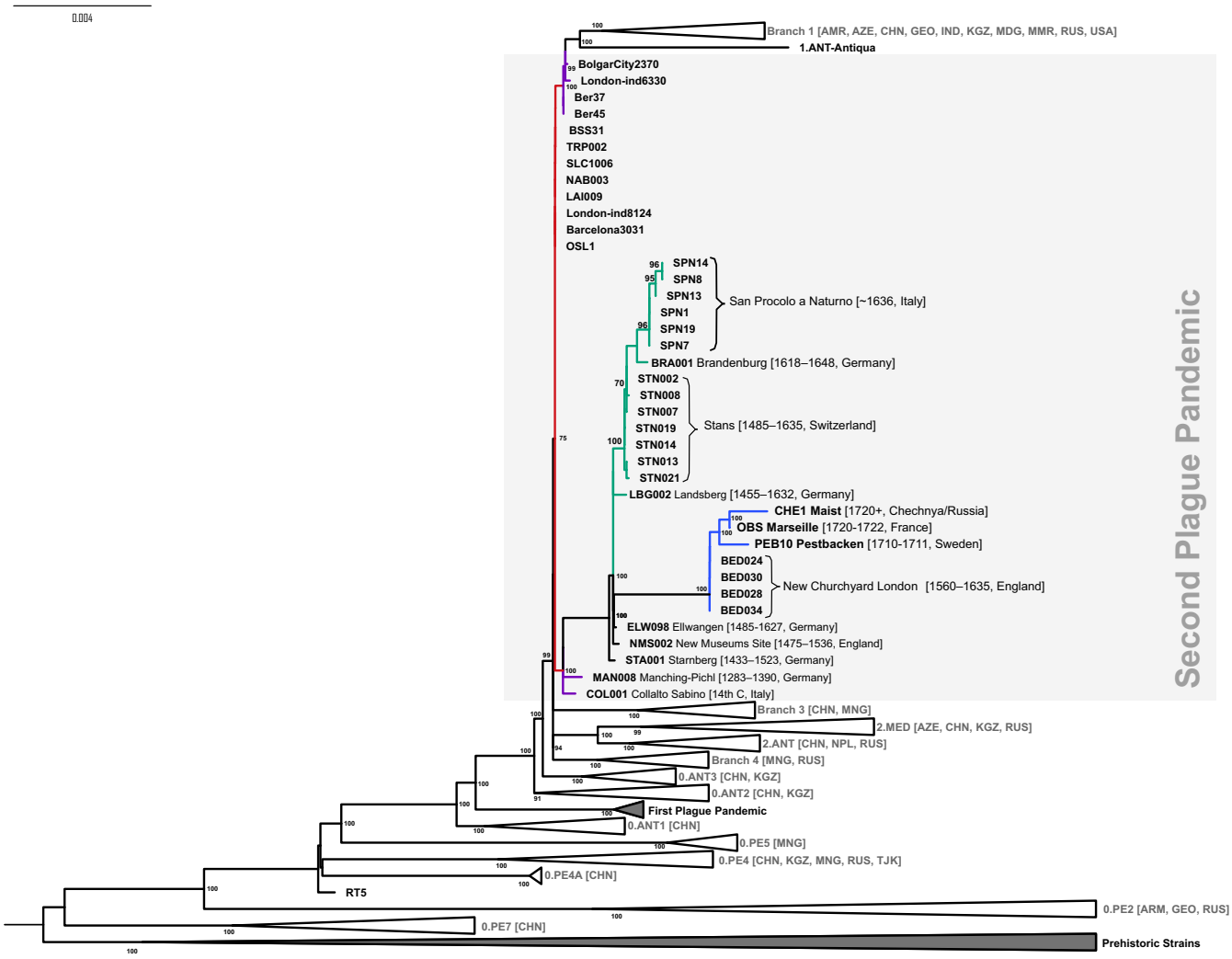

Fig. 3. Maximum-likelihood phylogenetic tree of *Y. pestis* focusing on the Second Plague Pandemic. The numbers at each node indicate the bootstrap values at 1,000 replicates. Branches highlighted in red correspond to the Black Death, while branches in blue correspond to the 17th/18th century genomes carrying the deletion. Branches colored in purple carry probable pestis secunda strains, and green branches represent the so-called Alpine clade.

maybe to Western and Eastern port cities, as previously hypothesized by Namouchi et al. (12). The SPN genomes (Italy, 17th century) are part of the "Alpine" clade, a cluster of samples, which could fall into the chronological range of the Thirty Years' War (1618 to 1648) plague epidemics. The genomes OBS, PEB10, and CHE1, which represent three distinct plague outbreaks (Figs. 1 and 3), are positioned at the phylogenetic end of the post-Black Death lineage. Chronologically, Pestbacken represents the first of these outbreaks (1710 to 1711) and is assumed to have been imported to Scandinavia from the Ottoman Empire (23–26). The second outbreak is Marseille l'Observance in 1722, which is documented to have been imported from Syria by sea (27–30), whereas the third (after 1722) took place in the Northern part of the Caucasus and currently represents the endpoint of the subbranch. Based on historical records for the three chronologically final genomes of the Second Plague Pandemic, an import of plague from a Western European reservoir can be ruled out, which is thoroughly reviewed in *Discussion*.

The Bayesian evolutionary analysis by sampling trees (BEAST) analysis revealed that the highest posterior density for the CHE1 isolate is between 179 and 402 y before present. As 2015 is the age of the most recent isolate, the age of CHE1 is estimated to be between 1613 and 1836 with a mean age of 1729 (286 years before present) (*SI Appendix*, Fig. S7).

Finally, we analyzed the virulence gene profiles for all genomes by using the virulence genes listed in Zhou and Yang (31). Similar to the genomes from Marseille l'Observance and New Churchyard, PEB10 and CHE1 lack the virulence genes mgtB and mgtC, which are situated in the 49-kbp deletion present in all four sites (*SI Appendix*, Fig. S8).

Discussion

The introduction of post-Black Death plague to Europe has been subject to debate in previous studies (10–12, 18). The main points of contention are the number of introductions from outside of Western Europe and the location of the reservoir(s) feeding the *Y. pestis* lineage, which established itself following the Black Death epidemic. One of the currently predominant hypotheses in the literature (14) suggests that a reservoir was established in Western Europe, potentially in the alpine region, and gave rise to the post-Black Death lineage prior to diversification. The lineage diversification possibly resulted in the establishment of multiple European reservoirs. However, the lineage lacked any non-European representative and chronologically ended with the last large outbreak recorded in Western Europe, the great

Downloaded by guest on October 26, 2020

plague of Marseille (1720 to 1723), which is documented to have been imported from the Eastern Mediterranean. Here, we propose an alternative hypothesis based on our synthesis of historical and genomic data. Evidence in the form of the genomes from Pestbacken and particularly the genome from Maist points toward the possibility of multiple waves of plague stemming from a single main reservoir situated outside of Western Europe. Following the apparent "diversification" of the post-Black Death lineage in our updated phylogeny, the post-Black Death lineage gives rise to four more outbreaks, the representative genomes of which all carry a large deletion and show an increase in substitution rate (14). Contrary to the Alpine clade (Landsberg, Brandenburg, Stans), which now also carries the Italian SPN genomes, these genomes stem from various locations in Eurasia (England, Sweden, France, and Chechnya, Russia). In the following paragraphs, we will summarize historical sources for the introduction routes, which support the genomically documented plague outbreaks of the 18th century in Western Europe.

Strain PEB10 was isolated from a skeleton from the plague cemetery of Pestbacken, dated numismatically to 1710 to 1711, in Southern Sweden, where plague had devastated the regions of Skåne and Blekinge in 1710 to 1711. The eastern and northern European plague outbreak of 1702 to 1713 was the last continental-scale plague epidemic to affect Western Europe. Set during the Great Northern War (1700 to 1721) and the War of the Spanish Succession (1701 to 1714), this plague epidemic took a heavy toll on the eastern European population, which was already suffering from a series of bad harvests and ensuing famines (32). The concurrent large-scale wars also involved massive soldier movements across the entire continent, which helped propagate plague over large distances in short periods of time (24). The epidemic is said to have entered Europe around the turn of the century in multiple waves via the Ottoman Empire through the Balkans and Transylvania. In 1709, the disease reached the southern Baltic coast, where the Swedish army was engaged in military actions against Russia. From there, Swedish military vessels brought plague to Karlskrona, where it spread from the barracks to the whole region of Blekinge and Näsum, Scania, ~12 km from Pestbacken, via returning soldiers (23–26). Plague continuously persisted in Blekinge until the end of 1711 (26). Strikingly, this epidemic event was the first documented outbreak in Sweden since the mid-1600s and the last documented plague outbreak to affect Scandinavia altogether (25, 26, 33).

Previously published genomes from Marseille l'Observance (10) can be dated to 1722 and therefore postdate the genome from Pestbacken by 12 y. According to historical sources, plague reached Marseille over the sea in 1720, where it persisted for almost 2 y. The boat associated with the outbreak, the "Grand-Saint-Antoine," had visited Lebanon, Syria, and Cyprus before arriving in Marseille and had lost multiple passengers to plague along the way (27–30). The genomes are derived from the plague pits dug in the Convent of l'Observance between May and September 1722 (10, 34). By then, the city had been isolated from the rest of the country for about 2 y, and individuals attempting to escape were killed upon sight. Spread to the rest of Europe had successfully been limited to the neighboring regions by the establishment of a cordon sanitaire by one-fourth of the French army and a range of other measures, such as the construction of a 36-km wall across the Vaucluse countryside (13, 28, 30).

We added an historical *Y. pestis* genome from the Caucasus region to the phylogeny. The sample CHE1 stems from the site of Maist, which is located in close proximity to today's Chechnya–Georgia border. The region houses multiple plague foci, which makes it a key area for plague activity today (21, 35, 36). Based on the phylogeny, the genome isolated from individual CHE1 appears to stem from the same lineage as the strains circulating in Europe until the 18th century. While no exact dates or related historical sources are available for this sample (*SI Appendix*), the isolated genome postdates the genomes from the plague pit of Marseille l'Observance (1722), based on the phylogeny, potentially dating well into the 18th century or even later, e.g., in the context of the Caucasian War (1817 to 1864). Considering that the disease did not advance far into France during the Great Plague of Marseille, it seems unlikely that the lineage introduced into Europe in 1720 was reintroduced from there to the Caucasus region and then reached Chechnya. Instead, we propose that the disease reached Chechnya directly on a separate route via land or that the region is located close to the wildlife reservoir that is responsible for the plague lineage documented in Europe following the Black Death (Fig. 2). For example, known modern plague foci are situated in neighboring Georgia, where written sources describing cases of bubonic plague date from the 11th century CE until after the start of the Third Plague Pandemic (1894) (35).

While it was not possible to identify a modern representative of 14th to 18th century lineage, recent studies have demonstrated the high genomic diversity found among the wildlife reservoirs of the Caucasus region (21). Considering the wealth of lineages unearthed in the Caucasus region alone, we can assume that much of *Y. pestis* diversity remains undiscovered. Additionally, more data from other regions of the world, which remained affected by plague beyond the outbreak of Marseille l'Observance in 1720, are needed to gain insights into the global diversity of lineages during the Second Plague Pandemic.

Recent ancient genomes predating the Second Plague Pandemic also emphasize the historical importance of Western and Central Asia for plague expansion. A genome from Tian Shan (Kyrgyzstan, 186 CE), recently published by Damgaard et al. (37), is the most basal genome of the first Plague Pandemic lineage to date (38–40) and actually predates the pandemic by more than 300 y. This study also described a second sample for which a full genome could not be assembled from North Ossetia–Alania (Russia, sixth and the ninth century CE), close to the Georgian border and the site of Maist (Chechnya, Russia). The dating of this genome has recently been contested (41), placing it closer to the polytomy of branches 1 to 4 and therefore chronologically closer to the Black Death. Although it is difficult to draw parallels across pandemics, as they can exhibit significant changes in dynamics, the importance of Western and Central Asia and the regions surrounding the Black Sea is pronounced across pandemics as far back as the Neolithic (2).

While the wildlife reservoir responsible for the post-Black Death lineage could be situated anywhere in Eurasia, we propose, based on our synthesis of historical sources and phylogeny, that the reservoir was not situated in Western Europe, but instead was close to Europe, specifically in Western Asia or the Black Sea region. From there, plague was introduced into Europe in multiple waves. We do not expect CHE1 to have originated in the European Alps, nor do we think it is likely that, following the Great Plague of Marseille, *Y. pestis* retreated to the East. In fact, historical sources for all of the 18th century genomes support the possibility of introduction through the Eastern Mediterranean and the Black Sea region. This would be plausible considering that the Second Plague Pandemic was not an exclusively European phenomenon (42) and, with the addition of CHE1, neither is the post-Black Death lineage. The Alpine clade could be the results of extended circulation during larger epidemics (e.g., the plague epidemics associated with the Thirty Years' War, during which regions of today's Germany are known to have lost up to 50% of their population (6, 29, 43) or by the establishment of short-lived secondary reservoirs, as all clade lineages seemingly became extinct in the 17th century. A similar clade formation could be expected following the addition of more strains from the Great Northern War epidemics, with Pestbacken representing one of the final outbreaks and being situated on its northernmost expansion. However, more data are needed to validate this hypothesis.

Downloaded by guest on October 26, 2020

Historians have concluded that the 17th century saw a major shift in plague dynamics around the 1630s (5–7). In Europe, this period of change was dominated by the Thirty Years' War, a period defined by large soldier movements and mass mortality by both war and disease. Considering that increases in mutation rates evidently occur in large transmission chains during epidemic events (44), mass movements and mortality could explain the increase in substitution rate, assuming constant circulation. However, the Alpine clade, which is at least partially chronologically contemporary to this branch, does not show such drastic rates (14). The early 17th century was also a period of crisis throughout the Ottoman Empire (45, 46). Two time periods stand out. The first crisis, dated between the 1570s and the 1610s, is marked by social unrest (Celali Rebellion) and demographic shifts, which were caused by sudden massive population loss and climatological events. The second one spans the period from the 1670s to the 1710s and falls within the Late Maunder Minimum. Both intervals were characterized by climate changes associated with the Little Ice Age, which have to be differentiated from the ones traditionally observed in the West. Instead of increased humidity, these periods are marked by historical reports of extended periods of drought, which led to food shortages, famine, and unrest (45, 46). However, a lack of high-quality, high-resolution climate reconstructions for this region does not allow any robust assessment of climate–plague interaction. Nonetheless, the history of the Ottoman empire, its Eastern reach, and its vast trade connections cannot be ignored when discussing plague introduction.

Following the shift in plague dynamics around the 1630, major plague epidemics slowly became less frequent. By the 18th century, quarantine was a widespread norm of surveillance throughout major European port cities, and state-regulated disease surveillance seemed to have successfully decreased the occurrence of major plague outbreaks in Europe (5, 47). Following the Eastern European epidemics of 1702 to 1713, the Austrian Empire established a cordon sanitaire, which was fully enforced by 1770 and spanned from the Adriatic Sea to Transylvania to avoid entry of plague from the Ottoman Empire, with which they had been at war since the 16th century (48). By the time that the Austrian cordon sanitaire started to disintegrate, the Ottoman Empire had established nationwide quarantine measures for the first time following the Napoleonic Wars (1803 to 1815) (5, 48, 49). These measures coupled with improved living conditions, medical care, and hygiene therefore could, to some extent, account for progressive European isolation from plague starting with the 17th century, assuming the source of plague was not situated in Western Europe.

After the Great Plague of Marseille (1720 to 1722), no plague outbreak of comparable scale is documented in Western Europe, but many reports of plague in the Near East and Russia exist, where the Russian–Turkish Wars spread plague within southeastern Europe and Russia, culminating in a devastating epidemic in Moscow in 1770 (42, 50, 51). Overall, the Near East and the Ottoman Empire, where natural plague foci are frequent to this day, were heavily affected by plague epidemics throughout the Second Plague Pandemic (35, 42, 51, 52). The Balkans were also impacted by plague long after its slow disappearance from Europe. This could be attributed to the different quarantine strategies of surrounding countries (47). For example, while mainland Greece was under Ottoman control during the 17th and 18th century, the Ionian Islands were governed by the city-states of the Adriatic Sea and subject to strict quarantine measures (47). Reported imports of plague to the Ionian Islands almost exclusively occurred from mainland Greece, where plague was a constant threat during the 18th century with only 14 recorded plague-free years. The ubiquity of plague in the Balkans during the entire 18th century (47) and its presence until the establishment of quarantine measures throughout the Ottoman Empire at the beginning of the 19th century, as well as documented imports of plague from the Balkans to the West, illustrate the directionality of plague transmission during the 17th and the 18th centuries according to historical sources (5, 47).

While these changes concur with societal changes such as increased disease control, they also seemingly coincide with structural and mutational changes in *Y. pestis* genomes circulating at the time. We noted that the 49-kbp deletion, which was first detected in the OBS genomes (10, 11, 14), was also present in the genomes PEB10 and CHE1. This deletion has evidently arisen on the branch between the genomes from Ellwangen and London New Churchyard, which is the branch with the highest substitution rate across the lineage (14). However, the effect of the loss of these genes on the overall virulence of the bacterium remains unclear. Considering the high mortality described during the epidemics associated with the genomes from Pestbacken and Marseille, virulence does not seem to have been considerably impaired. Moreover, while the deletion apparently appeared only during the last two centuries of the Second Plague Pandemic in Europe, the Balkans and the Near East continued to be heavily affected by the disease (42, 47). It is also important to note that a similar 49-kbp deletion has recently been reported in a First Plague Pandemic genome from Lunel-Viel (France, 567 to 618) (40) and that these sequences have therefore been lost independently on at least two known occasions. Both affected lineages have no known modern representatives.

Many different hypotheses have been proposed to explain the mechanisms of plague introduction to Europe and its subsequent spread throughout the continent. However, the true processes are likely part of a complex web of dynamics, of which the available data only allow us to elucidate the most general trends. Yet, the data presented in this study allow for interpretations of the current phylogeny and for its assessment with a broader geographic focus. Our multidisciplinary approach hints at the possibility of an extra-European reservoir, which introduced plague to Europe in multiple waves and saw the pathogen evolve within its natural foci. The limited available genomic data, while growing, make thorough historical research indispensable to provide context to the evolutionary processes observed via phylogeographic analyses. Most of the available data remain concentrated in Western Europe. Eastern Europe and much of the Mediterranean remain underrepresented or absent in the Second Plague Pandemic phylogeny and could provide key information in the debate. Further sampling in active plague regions, which have been shown to be genomically diverse, could also help to answer the question of whether the post-Black Death lineage went extinct or is currently undiscovered.

Conclusion

Our analysis shows that, following a "diversification," the post-Black Death lineage gave rise to four more outbreaks sporting the same deletion. The final strain in the Second Plague Pandemic phylogeny, CHE1, was isolated from a sample from the Caucasus region (Chechnya, Russia), making a uniquely European origin of the lineage unlikely. Combined with our analysis of historical data, our results point toward the existence of a reservoir outside of Western Europe responsible for the post-Black Death lineage. Based on our historical synthesis, we further speculate that this lineage kept on introducing plague to Eastern Europe and Western Asia long after the last large outbreaks documented in Western Europe, indicating the need for additional sampling in these regions to gain a better understanding of the complex processes involved in plague dynamics during the Second Plague Pandemic.

Methods

Full experimental procedures are provided in *SI Appendix*.

qPCR Screening. All aDNA extracts including milling and extraction blanks were screened via qPCR for human and *Y. pestis* DNA using previously published *Y. pestis* [pla and caf1M primers as published in Schuenemann et al. (53) and human mitochondrial (HVR1 L16209/H16348) (54, 55)] primers.

Downloaded by guest on October 26, 2020

Target Enrichment. *Y. pestis* DNA in the libraries from plague-positive individuals, confirmed via qPCR and shotgun metagenomics, was targeted and enriched with the MYBaits kit from MYcroarray using RNA probes at 3 to 5× tiling density. Prior to target enrichment, double-stranded single-indexed libraries were concentrated to 7 μL using a SpeedVac. All libraries were enriched individually according to the manufacturer's instructions of a modified version of the MYBaits kit (3.01).

Metagenomics Screening. A total of 10 libraries from 10 individuals were shotgun-sequenced. The datasets were demultiplexed at the Norwegian Sequencing Centre, and quality control was performed using FastQC (56). Adapters and indices were trimmed using cutadapt2.0 (57), and sequences shorter than 30 bp and below a quality score of 20 were discarded. Trimmed reads were merged using FLASH (58), and the presence of *Y. pestis* in the datasets was investigated using the taxonomic classifier Kraken (59) and metagenomic profiler MetaPhlAn2 (60). All tools confirmed the qPCR results and indicated the presence of the pathogen.

Capture Data from This Study. For capture datasets, quality control was done using FastQC (56). We trimmed and quality filtered (-q20, >30 bp) raw reads using cutadapt2.0 (57) and merged them using FLASH (58). We subsequently mapped our merged reads to the CO92 assembly of *Y. pestis* using bwa aln (-n 0.1 -l 1000) and bwa samse (61). The aligned datasets were sorted using samtools (62, 63), and duplicates were removed using Picard's MarkDuplicates module. We realigned our reads around indels using GATK's RealignerTargetCreator and IndelRealigner modules (64, 65) and computed damage plots using mapDamage2 (66). Statistics were compiled using GATK's DepthOfCoverage module (64, 65) and Qualimap2 (67). We also screened the new capture data with Kraken2 (68).

Genomes were visualized in Geneious R11 (69) and IGV (70). Coverage was calculated in 2,000-, 50-, and 10-bp windows, GC content was computed using a custom python script incorporating samtools in 2,000-, 50-, and 10-bp windows (71) and GC-skew using a perl script (72) in 2000/50/10 bp windows and plotted using Circos (73). We computed the edit distances of mapped reads with a custom python script using samtools and bamtools (74).

Published Ancient Genomes. For published genomes from Andrades Valtueña et al. (2) and Spyrou et al. (4), we mapped datasets according to the treatment that they had received prior to the library building and to the original publication. For uracil-DNA-glycosylase (UDG) libraries, we changed the bwa aln settings to -l 32 -n 0.1 and filtered out all reads below a mapping quality of 37. For half-UDG datasets, we mapped the raw reads with the following bwa aln settings: –l 16 –n 0.01. After cleaning up the mapping and filtering out all reads below a mapping quality of 37, we extracted the filtered reads, clipped the last two bases, and remapped the reads to the reference genome using the same bwa aln settings. These settings were also used to map non-UDG datasets. Genomes from Spyrou et al. (14) and Damgaard et al. (37) were mapped in the same way as the new capture data (see above) with one difference: UDG-treated samples were not rescaled with mapDamage2.0.

Phylogenetic Analysis. The phylogeny presented in Namouchi et al. (12) was updated with 64 new modern genomes, mainly from third pandemic strains isolated in central Asia, the Caucasus region, and Russia (19–21).

SNP calling was performed using samtools and bcftools mpileup. SNPs located within a frame of 10 bp from indels were excluded with samtools. For each sample, all identified SNPs were filtered and annotated using the snpToolkit (75). snpToolkit was used to filter and annotate SNPs from vcf files according to three criteria: quality score (≥30), depth of coverage (≥3), and allele frequency (90%). In addition, SNPs that were close to each other by less than 20 bp were excluded during the annotation process using the snpToolkit with option -f. All generated annotation output files were compared and combined using the command "combine" of the snpToolkit that produces two output files: 1) a tabulated file showing the distribution of all identified polymorphic sites of all analyzed samples and 2) a fasta file with the concatenation of all polymorphic sites for each sample. This fasta file was used to generate a ML tree using IQ-TREE (76). IQ-TREE was run using ModelFinder with the option –m MFP to infer the best substitution model for building the maximum-likelihood phylogenetic tree. A total number of 484 models were tested, and 1,000 fast bootstrap replicates were performed to assess statistical support at each node. As the concatenated SNPs include missing information due to some of the genomic regions not being covered, which is indicated by an exclamation mark when searching for the distribution of all polymorphic sites in the bam files of each aDNA sample, we used the ASC option to account for ascertainment bias correction. The generated tree was visualized using FigTree (77), and each SNP was mapped in the phylogenetic tree using maximum likelihood as implemented in timetree (78).

Estimation of CHE1 Tip Date using BEAST. We used the Bayesian framework BEAST (v2.6.0) (79) to estimate the tip date of the Chechnyan isolate CHE1 and assess the substitution rate variation across all *Y. pestis* strains. For each node, the divergence dates were estimated as years before the present, where the year 2015 was set as the present since it represents the most recent isolate included in this study. As previously described (12), the log-normal relaxed clock model and constant population size models were applied. To ensure run convergence, three independent chains of 50 million states were run and combined using LogCombiner with 10% burn-in.

Virulence Profile. We analyzed the presence and absence of virulence-associated genes in our new genomes using the gene intervals proposed in Zhou and Yang (31). We then computed the coverage of each gene in our mapping to the reference genome CO92 using bedtools (80) and plotted the interval coverage across all genomes on a heatmap generated using seaborn (81), numpy (82), and pandas (83).

Data Availability. Sequencing data have been deposited in the European Nucleotide Archive under accession no. PRJEB27821.

ACKNOWLEDGMENTS. This project was funded by the European Research Council under the FP7-IDEAS-ERC Program (Grant 324249) MedPlag. Data analysis was performed on the Abel Cluster, owned by the University of Oslo and the Norwegian metacentre for High-Performance Computing (NOTUR) and operated by the Department for Research Computing at the University of Oslo IT department. This article has received funding from the University of Ferrara under the Bando per il finanziamento della ricerca scientifica "Fondo per l'Incentivazione alla Ricerca" (FIR)-2020. Permission to study the samples from San Procolo a Naturno was given by Ripartizione 13, Beni culturali, Ufficio Beni archeologici Provincia autonoma di Bolzano Alto Adige (Permission Nr. 36.10/360889). We thank Ulf Büntgen for valuable input and Raffaella Bianucci for initial involvement in the project. We would also like to thank the Estonian Biocentre and the University of Tartu High Performance Computing Center.

1. S. Rasmussen *et al*., Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
2. A. Andrades Valtueña *et al*., The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691.e8 (2017).
3. N. Rascovan *et al*., Emergence and spread of basal lineages of Yersinia pestis during the neolithic decline. *Cell* **176**, 295–305.e10 (2018).
4. M. A. Spyrou *et al*., Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234 (2018).
5. E. A. Eckert, The retreat of plague from central Europe, 1640-1720: A geomedical approach. *Bull. Hist. Med.* **74**, 1–28 (2000).
6. E. A. Eckert, *The Structure of Plagues and Pestilences in Early Modern Europe: Central Europe, 1560-1640* (Karger Publishers, 1996).
7. G. Alfani, Plague in seventeenth-century Europe and the decline of Italy: An epidemiological hypothesis. *Eur. Rev. Econ. Hist.* **17**, 408–430 (2013).
8. G. Alfani, T. E. Murphy, Plague and lethal epidemics in the pre-industrial World. *J. Econ. Hist.* **77**, 314–343 (2017).
9. R. Calvert, An account of the origin and progress of the Plague of Malta, in the year 1813. *Med. Chir. Trans.* **6**, 1–64 (1815).
10. K. I. Bos *et al*., Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).
11. M. A. Spyrou *et al*., Historical Y. pestis genomes reveal the European Black death as the source of ancient and modern plague pandemics. *Cell Host Microbe* **19**, 874–881 (2016).
12. A. Namouchi *et al*., Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11790–E11797 (2018).
13. M. DeLacy, "Contagion and plague in the eighteenth century" in *The Germ of an Idea: Contagionism, Religion, and Society in Britain, 1660–1730*, M. DeLacy, Ed. (Palgrave Macmillan, 2016), pp. 147–169.
14. M. A. Spyrou *et al*., Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. *Nat. Commun.* **10**, 4470 (2019).
15. K. I. Bos *et al*., A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506–510 (2011).
16. B. Jacobsson, *Pestbacken: en begravningsplats för pestoffer från åren 1710 och 1711, Blekinge, Olofströms kommun, Jämshögs socken, Holje 5: 68 och 5: 69* [in Swedish] (Avd. för arkeologiska undersökningar, Riksantikvarieämbetet, 2002).
17. C. Arcini, B. Jacobsson, B. Persson, *Pestbacken* (Riksantikvarieämbetet, 2006).
18. M. A. Spyrou *et al*., A phylogeography of the second plague pandemic revealed through the analysis of historical Y. pestis genomes. bioRxiv, 10.1101/481242 (2018).

MICROBIOLOGY
ANTHROPOLOGY

Downloaded by guest on October 26, 2020

19. E. Zhgenti *et al*., Genome assemblies for 11 Yersinia pestis strains isolated in the Caucasus region. *Genome Announc.* **3**, e01030-15 (2015).
20. G. A. Eroshenko *et al*., Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. *PLoS One* **12**, e0187230 (2017).
21. V. V. Kutyrev *et al*., Phylogeny and classification of *Yersinia pestis* through the lens of strains from the plague foci of commonwealth of independent states. *Front. Microbiol.* **9**, 1106 (2018).
22. É. Hubert, *L'"Incastellamento" en Italie centrale: Pouvoirs, territoire et peuplement dans la vallée du Turano au Moyen Âge* (École française de Rome, 2002).
23. L. Walløe, *Plague and Population: Norway 1350-1750* (Department of Physiology, Inst. of Basic Medical Sciences, University of Oslo, Oslo, 1995).
24. S. Kroll, "'Die Pest' im Ostseeraum zu Beginn des 18. Jahrhunderts: Stand und Perspektiven der Forschung" in *Städtesystem und Urbanisierung im Ostseeraum in der Frühen Neuzeit. Urbane Lebensräume und Historische Informationssysteme, Beiträge des wissenschaftlichen Kolloquiums in Rostock vom 15. und 16. November 2004* [in German], S. Kroll, K. Krüger, Eds. (LIT Verlag Münster, 2006), pp. 124–148.
25. B. E. B. Persson, "The boy with the rake and the girl with the broom. Pestilences in early eighteenth century Southern Sweden" in *Städtesystem und Urbanisierung im Ostseeraum in der Frühen Neuzeit. Urbane Lebensräume und Historische Informationssysteme, Beiträge des wissenschaftlichen Kolloquiums in Rostock vom 15. und 16. November 2004*, S. Kroll, K. Krüger, Eds. (LIT Verlag Münster, 2006), pp. 172–204.
26. R. Schofield, The last visitation of the plague in Sweden: The case of bräkne-hoby in 1710–11. *Econ. Hist. Rev.* **69**, 600–626 (2015).
27. J. N. Biraben, *Les hommes et la peste en France et dans les pays européens et méditerranéens: La peste dans l'histoire* (Mouton, 1975).
28. M. Signoli, I. Seguy, J.-N. Biraben, O. Dutour, Paleodemographie et demographie historique en contexte epidemique: La peste en provence au XVIIIe siecle. *Population* **57**, 821–847 (2002).
29. J. N. Hays, *Epidemics and Pandemics: Their Impacts on Human History* (ABC-CLIO, 2005).
30. C. A. Devaux, Small oversights that led to the great plague of Marseille (1720-1723): Lessons from the past. *Infect. Genet. Evol.* **14**, 169–185 (2013).
31. D. Zhou, R. Yang, Molecular Darwinian evolution of virulence in Yersinia pestis. *Infect. Immun.* **77**, 2242–2250 (2009).
32. K.-E. Frandsen, *The Last Plague in the Baltic Region, 1709-1713* (Museum Tusculanum Press, 2010).
33. K.-E. Frandsen, "Das konnte nützen." Krieg, Pest, Hunger und Not in Helsingor im Jahr 1711" in *Städtesystem und Urbanisierung im Ostseeraum in der Frühen Neuzeit. Urbane Lebensräume und Historische Informationssysteme, Beiträge des wissenschaftlichen Kolloquiums in Rostock vom 15. und 16. November 2004*, S. Kroll, K. Krüger, Eds. (LIT Verlag Münster, 2006), pp. 205–225.
34. O. Dutour, M. Signoli, Etude anthropologique d'un charnier de la grande peste de Marseille (1720-1722) Premiers résultats. *Anthropol. et préhistoire* **108**, 147–158 (1997).
35. L. Bakanidze *et al*., "Surveillance on plague in natural foci in Georgia" in *Emerging and Endemic Pathogens*, K. P. O'Connell, E. W. Skowronski, A. Sulakvelidze, L. Bakanidze, Eds. (NATO Science for Peace and Security Series A: Chemistry and Biology, Springer, Dordrecht, The Netherlands, 2010), pp. 21–28.
36. B. V. Schmid *et al*., Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3020–3025 (2015).
37. P. B. Damgaard *et al*., 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
38. D. M. Wagner *et al*., Yersinia pestis and the plague of Justinian 541-543 AD: A genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
39. M. Feldman *et al*., A high-coverage Yersinia pestis genome from a sixth-century Justinianic Plague victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).
40. M. Keller *et al*., Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the first pandemic (541-750). *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12363–12372 (2019).
41. M. Keller *et al*., Ancient Yersinia pestis genomes provide no evidence for the origins or spread of the Justinianic Plague. *bioRxiv*, https://doi.org/10.1101/819698. (2019).
42. M. W. Dols, The second plague pandemic and its recurrences in the Middle East: 1347-1894. *J. Econ. Soc. Hist. Orient* **22**, 162–189 (1979).
43. G. Lammert, *Geschichte der Seuchen, Hungers-und Kriegsnot zur Zeit des dreissigjährigen Krieges* (J. F. Bergmann, 1890).
44. A. J. Vogler *et al*., A decade of plague in Mahajanga, Madagascar: Insights into the global maritime spread of pandemic plague. *mBio* **4**, e00623-e12 (2013).
45. S. A. White, "Climate change and crisis in Ottoman Turkey and the Balkans, 1590-1710" in *Proceedings of International Conference on Climate Change in the Middle East: Past, Present and Future*, Y. Ünal, C. Kahya, B. D. Demirhan, Eds. (Istanbul Technical University, Istanbul, 2006), pp. 20–23.
46. S. White, *The Climate of Rebellion in the Early Modern Ottoman Empire* (Cambridge University Press, 2011).
47. K. Konstantinidou, E. Mantadakis, M. E. Falagas, T. Sardi, G. Samonis, Venetian rule and control of plague epidemics on the Ionian Islands during 17th and 18th centuries. *Emerg. Infect. Dis.* **15**, 39–43 (2009).
48. G. E. Rothenberg, The Austrian sanitary cordon and the control of the bubonic plague: 1710-1871. *J. Hist. Med. Allied Sci.* **28**, 15–23 (1973).
49. B. Bulmuş, "Towards a new understanding of plague and quarantines in the Ottoman Empire" in *Plague, Quarantines and Geopolitics in the Ottoman Empire* (Edinburgh University Press, 2012), pp. 177–180.
50. A. Melikishvili, Genesis of the anti-plague system: The Tsarist period. *Crit. Rev. Microbiol.* **32**, 19–31 (2006).
51. A. Hashemi Shahraki, E. Carniel, E. Mostafavi, Plague in Iran: Its history and current status. *Epidemiol. Health* **38**, e2016033 (2016).
52. M. W. Dols, *The Black Death in the Middle East* (Princeton University Press, Princeton, NJ, 1977).
53. V. J. Schuenemann *et al*., Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the black death. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E746–E752 (2011).
54. W. Haak *et al*., Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **310**, 1016–1018 (2005).
55. O. Handt, M. Krings, R. H. Ward, S. Pääbo, The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.* **59**, 368–376 (1996).
56. Babraham Bioinformatics, FastQC A Quality Control tool for High Throughput Sequence Data (2018).
57. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
58. T. Magoč, S. L. Salzberg, FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
59. D. E. Wood, S. L. Salzberg, Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
60. N. Segata *et al*., Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
61. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
62. H. Li *et al*.; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
63. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
64. M. A. DePristo *et al*., A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
65. G. A. Van der Auwera *et al*., From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
66. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
67. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
68. D. E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
69. M. Kearse *et al*., Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
70. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
71. A. Namouchi, PlotBamCoverage. https://github.com/Amine-Namouchi/PlotBamCoverage. Accessed 1 June 2018.
72. S. Jain, *gcSkew.pl*. https://github.com/Geo-omics/scripts/blob/master/scripts/gcSkew. Accessed 1 June 2019.
73. M. Krzywinski *et al*., Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
74. D. W. Barnett, E. K. Garrison, A. R. Quinlan, M. P. Strömberg, G. T. Marth, BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
75. A. Namouchi, *snpToolkit*. https://github.com/Amine-Namouchi/snpToolkit. Accessed 1 June 2018.
76. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
77. A. Rambaut, *FigTree*, Version 1.4.4. https://github.com/rambaut/figtree. Accessed 1 November 2018.
78. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
79. R. Bouckaert *et al*., BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
80. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
81. M. Waskom, *seaborn*. https://github.com/mwaskom/seaborn. Accessed 1 February 2019.
82. S. van der Walt, S. C. Colbert, G. Varoquaux, The NumPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).
83. W. McKinney, "Data structures for statistical computing in python" in *Proceedings of the 9th Python in Science Conference*, S. van der Walt, J. Millman, Eds. (Austin, TX, 2010), pp. 51–56.

Downloaded by guest on October 26, 2020