Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus

Kirsten I. Bos*[1,2], Alexander Herbig[1,2], Jason Sahl[3], Nicholas Waglechner[4], Mathieu Fourment[5], Stephen A. Forrest[1], Jennifer Klunk[6,7], Verena J. Schuenemann[1], Debi Poinar[6], Melanie Kuch[6], G. Brian Golding[7], Olivier Dutour[8], Paul Keim[3], David M. Wagner[3], Edward C. Holmes[5], Johannes Krause*[1,2], and Hendrik N. Poinar[4,6,7,9]*.

[1]Department of Archeological Sciences, University of Tuebingen, Germany.

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

[3]Center for Microbial Genetics and Genomics, Northern Arizona University

[4]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton ON L8S 4L9, Canada.

[5]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia.

[6]McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, Hamilton, ON, L8S 4L9 Canada.

[7]Department of Biology, McMaster University, Hamilton, ON, L8S 4L9 Canada.

[8]Laboratoire d'anthropologie biologique Paul Broca, Ecole Pratique des Hautes Etudes, UMR 5199 PACEA, CNRS-Université de Bordeaux, CS 50023, 33615 France

[9]Department of Biochemistry, McMaster University, Hamilton, ON, L8S 4L9 Canada

*address correspondence to: KIB (bos@shh.mpg.de), JK (krause@shh.mpg.de), or HNP (poinarh@mcmaster.ca)

Abstract:

The 14th-18th century pandemic of *Yersinia pestis* caused devastating disease outbreaks in Europe for almost 400 years. The reasons for plague's persistence and abrupt disappearance in Europe are poorly understood, but could have been due to either the presence of now-extinct plague foci in Europe itself, or successive disease introductions from other locations. Here we present five *Y. pestis* genomes from one of the last European outbreaks of plague, from 1722 in Marseille, France. The lineage identified has not been found in any extant *Y. pestis* foci sampled to date, and has its ancestry in strains obtained from victims of the 14th century Black Death. These data suggest the existence of a previously uncharacterized historical plague focus that persisted for at least three centuries. We propose that this disease source may have been responsible for the many resurgences of plague in Europe following the Black Death.

39 **Introduction**

40

41 The bacterium *Yersinia pestis* is among the most virulent pathogens known to cause
42 disease in humans. As the agent of plague it is an existing threat to public health as
43 the cause of both emerging and re-emerging rodent-derived epidemics in many
44 regions of the world (Duplantier et al., 2005; Vogler et al., 2011; Gage and Kosoy,
45 2005). This, and its confirmed involvement in three major historical pandemics, have
46 made it the subject of intense study. The first pandemic, also known as the Justinian
47 Plague, occurred from the 6th through the 8th centuries; the second pandemic
48 spanned the 14th to the 18th centuries; and the third pandemic started in the 19th
49 century and persists to the present day.

50

51 Attempts to date the evolutionary history of *Y. pestis* using molecular clocks have
52 been compromised by extensive variation in nucleotide substitution rates among
53 lineages (Cui et al., 2012; Wagner et al., 2014), such that there is considerable
54 uncertainty over how long this pathogen has caused epidemic disease in human
55 populations. In addition, there has been lively debate as to whether or not it was the
56 principal cause of the three historical pandemics (Cohn, 2008; Scott and Duncan,
57 2001). It is well established that extant lineages of *Y. pestis* circulated during the
58 third pandemic (Achtman et al., 2004; Morelli et al., 2010), and various ancient DNA
59 studies have now unequivocally demonstrated its involvement in the early phase of
60 the first pandemic in the 6th century (Harbeck et al., 2013; Wagner et al., 2014) and
61 the Black Death (1347 – 1351), which marks the beginning of the second plague
62 pandemic (Bos et al., 2011; Haensch et al., 2010; Schuenemann et al., 2011). All

three pandemics likely arose from natural rodent foci in Asia and spread along trade routes to Europe and other parts of the world (Morelli et al., 2010; Wagner et al., 2014). The strains associated with the first and second pandemics represent independent emergence events from these rodent reservoirs in Asia. The on-going third pandemic also originated in Asia, although genetic evidence suggests that it may derive from strains that descend from those associated with the second wave that spread back to Asia and became re-established in rodent populations there (Wagner et al., 2014).

The impact of the second pandemic was extraordinary. During this time period there were hundreds and perhaps thousands of local plague outbreaks in human populations throughout Europe (Schmid et al., 2015). It is very likely that some of these outbreaks were caused by the spread of plague via the maritime transport of humans and cargo, as was undoubtedly the case during the global spread of plague during the third pandemic (Morelli et al., 2010). However, the processes responsible for the potential long-term persistence of plague in Europe during the second pandemic are still subject to debate. It is possible that once introduced to Europe around the time of the Black Death, *Y. pestis* persisted there for centuries, cycling in and between rodent and human populations and being introduced or reintroduced to various regions throughout Europe (Carmichael, 2014). Another possibility is that plague did not persist long-term in European rodent populations, but rather was continually reintroduced from rodent plague foci in Asia (Schmid et al., 2015).

To address this key issue in *Y. pestis* evolution and epidemiology, we investigated plague-associated skeletal material from one of the last well-documented epidemics in Provence (i.e. the Plague of Provence), France (1720-1722) that occurred at the end of the second pandemic. In particular, we compared the evolutionary relationship of these historical *Y. pestis* strains to those sampled from other time periods, both modern and ancient. Our results demonstrate that the strains responsible for the Black Death left descendants that persisted for several centuries in an as yet unidentified host reservoir population, accumulated genetic variation, and eventually contributed to the Plague of Provence in mid-eighteenth century France. Although these lineages no longer seem to be represented within the genetic diversity of extant sampled *Y. pestis*, they may have been involved in additional, earlier second-wave epidemics in the Mediterranean region and beyond.

Results

Skeletal material used in this investigation was sampled from the Observance (OBS) collection housed in the osteoarcheological library of the Regional Department of Archaeology, French Ministry of Culture, medical faculty of Aix-Marseille Université. Historical records indicate that this collection represents a catastrophic burial for victims of the Plague of Provence relapse in Marseille, 1722 (Signoli et al., 2002). Rescue excavations carried out in 1994 unearthed the remains of 261 individuals, mostly consisting of adult males. Material from this collection was also used in the first PCR-based investigation of ancient *Y. pestis* DNA from archaeological tissue (Drancourt et al., 1998). In our study 20 *in situ* teeth (Figure 1) were sampled and

110 screened for *Y. pestis* DNA through a quantitative PCR assay. Amplification products
111 were detectable for five of the 20 teeth (Table 1). Products were not sequenced.
112 Negative controls were free of amplification products. The five putatively *Y. pestis*-
113 positive samples were subject to high-throughput DNA sequencing after array-based
114 enrichment for *Y. pestis* DNA. Mapping to the *Y. pestis* CO92 reference genome
115 revealed a minimum of 12-fold average genomic coverage for each of the five
116 genomes (Table 2 and Figure 2). Raw sequencing data as well as mapped reads have
117 been deposited at the European Nucleotide Archive under accession PRJEB12163.
118 Comparative analysis together with 133 previously published *Y. pestis* genomes
119 resulted in 3109 core genome SNPs for phylogenetic analysis. No homoplasies were
120 identified.
121
122 The phylogenetic position of the five OBS sequences on a unique branch of the *Y.*
123 *pestis* phylogeny (lineages shown in red in Figure 3), yet seemingly derived from
124 those strains associated with the Black Death, both confirms its authenticity and
125 reveals that it is likely a direct descendent of strains that were present in Europe
126 during the Black Death. Although the topological uncertainty (i.e. trichotomy) among
127 the second pandemic strains (London and Observance) suggests that there was a
128 radiation of *Y. pestis* lineages at the time, it is notable that the OBS sequences do not
129 possess the additional derived position that is present in both the 6330 Black Death
130 strain and the SNP-typed individual from Bergen Op Zoom (defined as the "s12" SNP
131 by Haensch et al., 2010). Indeed, the tree is striking in that there is clear
132 phylogenetic support (88% bootstrap values, Figure 3A) for 6330 clustering with the
133 Branch 1 strains, including those that gave rise to the third (modern) plague

pandemic. Finally, although evolutionary rates in *Y. pestis* are notoriously variable (Cui et al., 2012; Wagner et al., 2014), the comparatively late time period from which our samples derive and the long branch leading to these OBS sequences suggests that this lineage co-existed for an extended period with the Branch 1 strains responsible for all later human plague outbreaks outside of China.

Identification of DFR4 in OBS Strains. Using the pan-array design, we were able to identify a ~15kb Genomic Island (GI) in the OBS strains as previously observed in the Justinian and Black Death strains (Wagner et al., 2014) and referred to as DFR 4 (difference region 4). Notably, this region has been lost in some *Y. pestis* strains such as CO92. This island is a striking example of the decay common to the highly plastic *Y. pestis* genome.

Discussion

The Observance lineage of *Y. pestis* identified by our analysis of material from the Plague of Provence (1720-1722) is clearly phylogenetically distinct and has not been identified in any extant plague foci for which full genome data are available. That this lineage is seemingly confined to these ancient samples suggests that it is now extinct, or has yet to be sampled from extant reservoirs. Based on available data, the *Y. pestis* strains most closely related to the Observance lineage are those obtained from other ancient DNA studies of victims of the Black Death in London and other areas (Haensch et al., 2010; Bos et al., 2011) that occurred some 350 years earlier.

158 The most important conclusion from the current study is that *Y. pestis* likely
159 persisted and diversified in at least one historical reservoir throughout the second
160 pandemic. The length of the branch leading to the Observance lineage is consistent
161 with a long history of circulation. Of note, this new lineage does not share the single
162 derived mutation that is common to one of the Black Death genomes from London
163 (individual 6330, Bos et al., 2011) and the Bergen op Zoom plague victim from the
164 Netherlands (Haensch et al., 2010); hence, multiple variants of *Y. pestis* evidently co-
165 circulated in London and elsewhere during the Black Death, and at least one of these
166 variants followed a path that differs from the lineage that later gave rise to the
167 Plague of Provence. Since historical sources document that plague traveled to the
168 Low Countries after the period of peak mortality in London (Gottfried, 1983), our
169 results are consistent with the Plague of Provence having derived from a lineage that
170 predates this additional European diversity.
171
172 Since the novel lineage of *Y. pestis* identified here was obtained from an active
173 Mediterranean port city that served as a main commercial hub and entry point into
174 Western Europe from various origins, the precise location of the disease's source
175 population cannot be easily determined. Several scenarios can explain these data.
176 First, all *Y. pestis* diversity documented in the Black Death and later plague outbreaks
177 could have stemmed from foci in Asia, where European epidemics were the result of
178 successive introductions from this distant, albeit prolific, source population (Schmid
179 et al., 2015). This model would not require the back migration of plague into Asia for
180 the third pandemic proposed elsewhere (Wagner et al, 2014), since all preexisting
181 diversity would already be present. A multiple wave theory has previously been

182 proposed to explain the presence of a reportedly different Black Death lineage in the
183 Low Countries (defined by their single "s12" SNP) compared to those circulating in
184 England and France (Haensch et al., 2010). However, the presence of this branch 1
185 position in both the ancestral and derived state in London during the Black Death
186 (Bos et al., 2011; Figure 3B) suggests that it became fixed during this outbreak,
187 rather than having been introduced via a separate pulse from Asia. To date, all of the
188 *Y. pestis* material examined from the second pandemic has been assigned to Branch
189 1 in the *Y. pestis* phylogeny, indicating a close evolutionary relationship. Given the
190 relatively high diversity of *Y. pestis* in East Asia, which constitutes the likely ultimate
191 geographical source of all three plague pandemics (Wagner et al., 2014), and the
192 large "big bang" radiation of *Y. pestis* at the beginning of the second pandemic (Cui
193 et al., 2012), it seems unlikely that multiple waves of plague from Asia into Europe
194 during the Black Death would have involved such closely-related strains. As China
195 harbours many branch 1 descendent lineages that also share this "s12" SNP, our
196 observation adds weight to the notion that the branch 1 lineage of *Y. pestis* spread
197 east sometime during the second pandemic after it entered Europe, and
198 subsequently became established in one or more East Asian reservoir species
199 (Wagner et al., 2014). Here it remained before radiating to other locations in the late
200 19th century, giving rise to the current worldwide "third wave" *Y. pestis* pandemic.
201
202 In our view, a far more plausible option to account for the distinct position of the
203 Observance lineage is that a *Y. pestis* strain responsible for the Black Death became
204 established in a natural host reservoir population within Europe or western Asia.
205 Once established, this *Y. pestis* lineage evolved locally for hundreds of years and

206 contributed to repeated human epidemics in Europe. Plague's temporary
207 persistence somewhere on the European mainland is a compelling possibility
208 (Carmichael, 2015). Alternatively, a reservoir population located in a geographically
209 adjacent region, such as in western Asia or the Caucasus, may have acted as a
210 regular source of disease through successive westerly pulses during the three
211 century-long second pandemic (Schmid et al., 2015). Since our study used material
212 exclusively from a highly active port with trade connections to many areas in the
213 Mediterranean and beyond, it is impossible to identify a source population for the
214 Plague of Provence at our current resolution.
215
216 Our analysis reveals a previously uncharacterized plague source that could have
217 fuelled the European epidemics from the time of the Black Death until the mid-
218 eighteenth century. This parallels aspects of the third pandemic, that rapidly spread
219 globally and established novel, long-lived endemic rodent foci that periodically
220 emerge to cause human disease (Morelli et al., 2010). The current resolution of *Y.*
221 *pestis* phylogeography suggests that the historical reservoir identified here may no
222 longer exist. More extensive sampling of both modern rodent populations and
223 ancient human, as well as rodent, skeletal remains from various regions of Asia, the
224 Caucasus, and Europe may reveal additional clues regarding past ecological niches
225 for plague. The reasons for the apparent disappearance of this historic natural
226 plague focus are, however, unknown, and any single model proposed for plague's
227 disappearance is likely to be overly simplistic. Pre-industrial Europe shared many of
228 the same conditions that are correlated with plague dissemination today including
229 pronounced social inequality, and poor sanitation coupled with high population

density in urban centers (Volger et al., 2005; Barnes, 2014). This constellation of anthropogenic factors, along with the significant social and environmental changes that occurred during the Industrial Revolution must be considered alongside models of climate and vector-driven dynamics as contributing to the rapid decline in historical European plague outbreaks, where genomic data provide but one piece of critical information in this consilient approach.

Material and Methods

For this study 20 *in situ* teeth were freshly harvested, horizontally sectioned at the cementoenamel junction in a dedicated ancient DNA facility, and drilled to remove approximately 50mg of dental pulp/dentin from the inner surface of the tooth crown or the roots. DNA extractions were carried out following protocols described elsewhere (Schwarz et al., 2008; Rohland and Hofreiter, 2007; Dabney et al., 2014), and survival of *Y. pestis* DNA was evaluated through an established quantitative PCR assay for the *pla* gene (Schuenemann et al., 2011). DNA extracts for the five putatively *Y. pestis*-positive samples, each from a separate individual, were converted into double-indexed Illumina libraries (Meyer et al., 2010; Kircher et al., 2012) following an initial uracil DNA glycosylase treatment step (Briggs et al., 2010) to remove deaminated cytosines, the most common form of ancient DNA damage. Libraries were amplified with AccuPrime Pfx (Life Technologies) to a total of 20µg and serially captured in equimolar concentrations over two identical 1 million-feature Agilent microarrays (Hodges et al., 2010). Arrays were designed with probes tiled every three base pairs matching the CO92 *Y. pestis* reference chromosome

(NC_003143), supplemented with additional chromosomal regions from *Y. pestis* biovar Microtus str. 91001 (NC_005810), *Y. pseudotuberculosis* IP 32953 (NC_006155) and *Y. pseudotuberculosis* IP 31758 (NC_009708) that are absent in CO92. Captured products, including those from negative controls, were sequenced on one lane of an Illumina HiSeq 2000.

Raw reads were trimmed, overlapping paired reads were merged as described elsewhere (Schuenemann et al., 2013), and merged reads were subsequently filtered for a minimal length of 30 bp. Preprocessed reads were mapped to the CO92 reference genome using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) with increased specificity (-n 0.1) and a map quality filter of 37.

For maximum accuracy in SNP calling, reads were processed independently at three research centres, and the intersection was used as our final SNP table (Supplementary File 1). At the University of Tuebingen, reads were processed as described above and SNP calling was done according to a protocol described by Bos et al. 2014 using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK) (DePristo et al., 2011). Data for 133 previously published *Y. pestis* strains were processed respectively (Supplementary File 2). A custom tool was used for the comparative processing of the results. SNPs were called with a minimal coverage of 5-fold and a minimal frequency of 90% for the SNP allele. Problematic sites as identified by Morelli et al. (2010), genomic non-core regions as defined by Cui et al. (2012) as well as annotated repeat regions, rRNAs, tRNAs and tmRNAs were excluded from the SNP analysis. SNPs were annotated using SnpEff (Cingolani et al.,

2012) with default parameters. The upstream and downstream region size was set to 100 nt. At McMaster University, raw reads were trimmed and merged using SeqPrep (https://github.com/jstjohn/SeqPrep), requiring a minimum overlap of 11 base pairs to merge. Reads shorter than 24 base pairs were filtered out using SeqPrep. Merged reads and unmerged forward reads were concatenated and mapped to the CO92 reference sequence (NC_003143) using the Burrows-Wheeler Aligner (BWA) version 0.7.5 (Li and Durbin, 2009) with the parameters described in Wagner et al. 2014. Duplicates were collapsed using a custom script, which collapses only reads with identical 5' position, 3' position, and direction. Assemblies were imported into Geneious R6 and SNPs with at least 5-fold coverage and 90% variance were called. SNPs were visually inspected for quality. At NAU, reads were trimmed with Trimmomatic (Bolger et al., 2014) and aligned against the CO92 reference sequence using BWA-MEM. SNPs were called with the Unified Genotyper in GATK and were filtered by minimum coverage (5X) and allele frequency (90%). SNPs in genome assemblies were identified by a direct mapping against the reference sequence using NUCmer (Delcher et al., 2002). These methods were wrapped by the Northern Arizona SNP Pipeline (NASP) (tgennorth.github.io/NASP).

Phylogenetic Analysis of the Observance *Y. pestis*

A phylogenetic tree of *Y. pestis* strains was inferred using the maximum likelihood (ML) method available in PhyML (Guindon et al., 2010), assuming the GTR model of nucleotide substitution (parameter values available from the authors on request) and a combination of NNI and SPR branch-swapping. The robustness of individual nodes was assessed using bootstrap resampling (1000 pseudo-replicates) using the

same substitution model and branch-swapping procedure as described above. The phylogeny comprised data from the five Observance (OBS) strains, two sequences from Black Death victims who died in London between 1348 and 1350 (strain 6330 and a combined pool of identical strains 8291, 11972, and 8124), one sequence from the Plague of Justinian, AD540 (strain A120), and 130 modern *Y. pestis* strains. A single sequence of *Y. pseudotuberculosis* (strain IP32953) was used as an outgroup to root the tree, although with all derived SNPs removed to assist branch-length scaling.

Identification of DFR4 in OBS Strains. Merged and trimmed reads were mapped to the *Y. pestis* biovar Microtus str. 91001 chromosome at the DFR 4 region (between position 1041000 and 1063000). Additionally, the unique Microtus 31-mers were also mapped back to the Microtus chromosome to ensure identity of the region using short read length data. All resulting alignment files were compared using BEDTools (Quinlan and Hall, 2010) to identify intervals of the Microtus chromosome that were unique to Microtus when compared to CO92 and were covered by reads from our ancient extracts.

Author contributions: KIB, OD, DP and HNP conceived of the investigation. OD performed skeletal sampling. KIB, HNP, NW, and SAF designed experiments. KIB, MK and VJS performed experiments. AH, KIB, JS, GBG, HNP, DMW, ECH, and JK analysed data. KIB wrote the manuscript with contributions from all coauthors.

Acknowledgements

We acknowledge The Director of the Regional Department of Archeology (DRAC-PACA / Aix-en-Provence, France) for granting us access to the Observance skeletal collection (4422-17/09/2009). Funding was provided by European Research Council starting grant APGREID (to JK, KIB and AH), Social Sciences and Humanities Research Council of Canada postdoctoral fellowship grant 756-2011-501 (to KIB), National Health and Medical Research Council (NHMRC) grant (to ECH and HNP), and an NHMRC Australia Fellowship to ECH, as well as a Canada Research Chair and NSERC discovery research grant to HNP. HNP thanks former and current members of the McMaster aDNA Centre for their help with this study.

References Cited

1. Pneumonic plague pathogenesis and immunity in brown Norway rats. DM Anderson, NA Ciletti, H Lee-Lewis, E Elli, H Segal, KL DeBord, et al. Imunopathology and infectious diseases, 174, 910-921, 2009
2. The disappearance of plague: a continuing puzzle. AB Appleby. The Economic History Review, 33, 161-173, 1980
3. Social vulnerability and pneumonic plague: revisiting the 1994 outbreak in Surat, India. KB Barnes. Environmental Hazards, 13, 161-180, 2014
4. Trimmomatic: a flexible trimmer for Illumina sequence data. AM Bolger, M Lohse, B Usadel. Bioinformatics. 2014 August 1; 30(15): 2114–2120.

5. A draft genome of Yersinia pestis from victims of the Black Death. KI Bos, VJ Schuenemann, GB Golding, HA Burbano, N Waglechner, BK Coombes, et al. Nature, 478, 506-510, 2011

6. Pre-Columbian mycobacterial genomes reveals seals as a source of New World human tuberculosis. KI Bos, KM Harkins, A Herbig, M Coscolla, N Weber, I Comas. Nature, 514, 494-497, 2014

7. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. AW Briggs, U Stenzel, M Meyer, J Krause, M Kircher, and S Pääbo. Nucleic Acids Res, 38, e87, 2010

8. Plague persistence in Western Europe: a hypothesis. AG Carmichael. Pandemic Disease in the Medieval World: rethinking the Black Death. The Medieval Globe, 1, 157-191, 2014

9. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. P Cingolani, A Platts, L Wang le, M Coon, T Nguyen, L Wang, SJ Land, X Lu, DM Ruden. Fly (Austin), 6(2):80-92, 2012

10. Epidemiology of the Black Death and successive waves of plague. SK Cohn, Med His, 52, 74-100, 2008

11. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Y Cui, C Yu, D Li, Y Li, T Jombart, LA Weinart, et al. Proc Nat Acad Sci, 110, 577-582, 2012

12. Fast algorithms for large-scale genome alignment and comparison. AL Delcher, A Phillippy, J Carlton, SL Salzberg. Nucleic Acids Res. 2002 June 1; 30(11): 2478–2483.

13. Second-pandemic strain of Vibrio cholera from the Philadelphia cholera outbreak of 1849. AM Devault, BG Golding, N Waglechner, JM Enk, M Kuch, JH Tien, et al. New Engl J Med, 370, 334-340, 2012

14. Detection of 400-year old Yersinia pestis DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. M Drancourt, G Aboudharam, M Signoli, O Dutour, D Raoult. Proc Nat Acad Sci, 95, 12637-12640, 1998

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

16. Natural history of plague: perspectives from more than a century of research. KL Gage and MY Kosoy. Ann Rev Entomol, 50, 505-28, 2005

17. The Black Death. RS Gottfried. Free Press, 1983

18. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. Syst Biol, 59, 307-321, 2010.

19. Distinct clones of Yersinia pestis caused the Black Death. S Haensch, R Bianucci, M Signoli, M Rajerison, M Schultz, S Kacki, et al. PLOS Path 6, e1001134, 2010

20. Yersinia pestis DNA from skeletal remains from the 6th Century AD reveals insights into the Justinianic Plague. M Harbeck, L Seifert, S Hänsch, DM Wagner, D Bridsell, et al. PLOS Path, 9, e1003349, 2013

21. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. E Hodges, M Rooks, Z Xuan, A Bhattacharjee, DB Gordon, L Brizuela, et al. Nature Protocols, 4, 960-974, 2009

22. Metapopulation dynamics of bubonic plague. MJ Keeling and CA Gilligan. Nature 407, 903-906, 2000

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

24. Genesis of the anti-plague system: the Tsarist period. A Melikishvili. Critical Reviews in Microbiol, 32, 19-31, 2006

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

26. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. G Morelli, Y Song, CJ Massoni, M Eppinger, P Roumagnac, DM Wagner, et al. Nature Genet, 42, 1140-1145, 2010

27. BEDTools: a flexible suite of utilities for comparing genomic features. AR Quinlan and IM Hall. Bioinformatics, 26(6): 841-842, 2010

28. Ancient DNA extraction from bones and teeth. N Rohland and M Hofreiter. Nature Protocols, 2, 1756-1762, 2007

29. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. BV Schimd, U Büntgen, WR Easterday, C Ginzler, L Walløe, B Bramanti, and NC Stenseth. Proc Nat Acad Sci, 112, 3020-3025, 2015

30. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. VJ Schuenemann, K Bos, S DeWitte, S Schmedes, J Jamieson, A Mittnik, et al. Proc Nat Acad Sci, www.pnas.org/cgi/doi/10.1073/pnas.1105107108, 2011

419 31. Genome-wide comparison of medieval and modern Mycobacterium leprae. VJ
420 Schuenemann, P Singh, TA Mendum, B Krause-Kyora, G Jäger, KI Bos, et al.
421 Science 341, 179-183, 2013
422 32. New insights from old bones: DNA preservation and degradation in permafrost
423 mammoth remains. C Schwarz, R Debruyne, M Kuch, E McNally, H Schwarcz, AD
424 Aubrey, et al. Nucleic Acids Res, 37, 3215-3229, 2008
425 33. The Biology of Plagues: evidence from historical populations. S Scott and CJ
426 Duncan. Cambridge, 2005
427 34. Paleodemography and historical demography in the context of an epidemic:
428 plague in Provence in the eighteenth century. M Signoli, I Séguy, J-N Biraben, O
429 Dutour. Population-E, 57, 829-854, 2002
430 35. Phylogeography and molecular epidemiology of Yersinia pestis in Madagascar.
431 AJ Volger, F Chan, DM Wagner, P Roumagnac, J Lee, R Nera, M Eppinger, J Ravel,
432 L Rahalison, BW Rasoamanana, SM Beckstrom-Sternberg, M Achtman, S
433 Chanteau, and P Keim. PLOS Neglected Tropical Diseases, 5, e1319, 2011
434 36. Yersinia pestis and the Plague of Justinian 541-543AD: a genomic analysis. DM
435 Wagner, J Klunk, M Harbeck, A Devault, N Waglechner, JW Sahl, et al. Lancet
436 Infect Dis, http://dx.doi.org/10.1016/S1473-3099(13)70323-2, 2014
437 37. Validation of inverse seasonal peak mortality in medieval plagues, including the
438 Black Death, in comparison to modern Yersinia pestis-variant diseases. MR
439 Welford and BH Bossak. PLOS ONE, 4, e8401, 2009
440

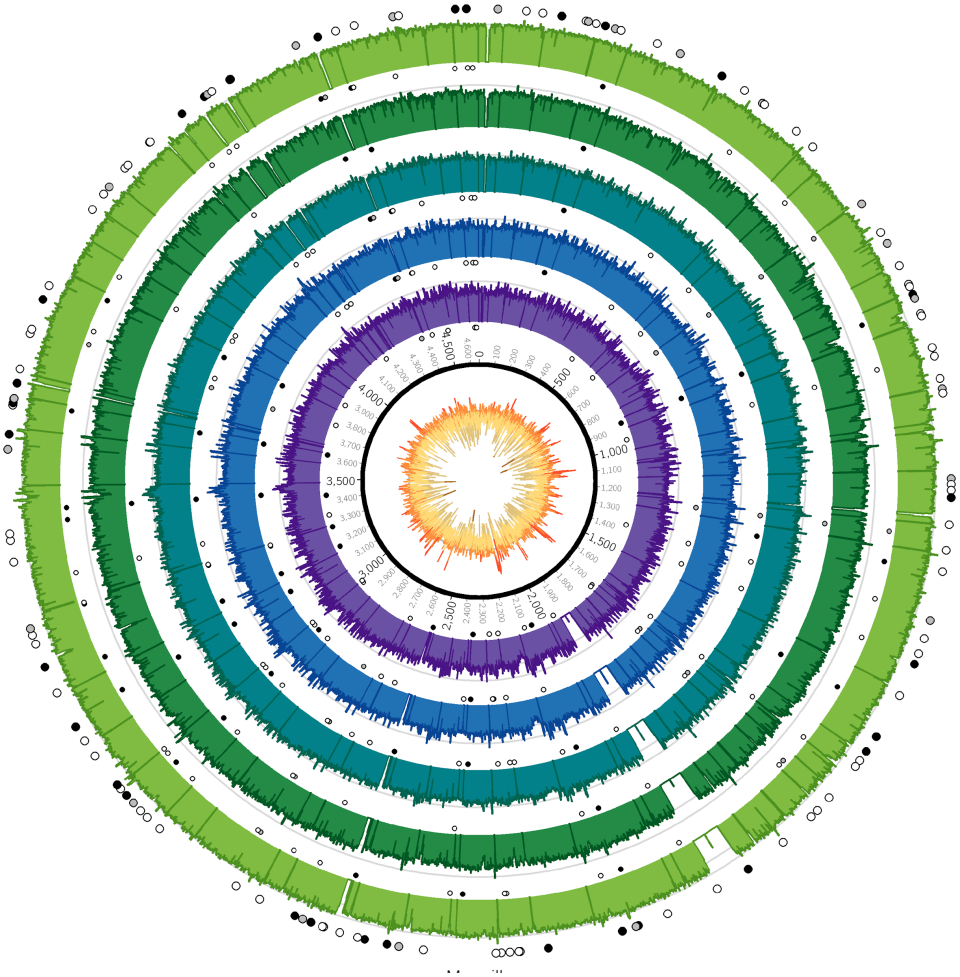
441 Main text Figures:

442

443

444 Figure 1

445 Photograph of Sample OBS 110.



446

447 Figure 2

448 Coverage plots for all reconstructed Observance core genomes. Inner ring: GC

449 content from low (brown, <30%) to high (orange, >55%). Outer concentric rings:

450 from inner to outer, coverage plots of OBS107, 110, 116, 124, and 137 on a

451 logarithmic scale. Axes are at 30x and 100x. Dots: SNPs (black = non-synonymous,

452 grey = synonymous, white = intergenic). Outer ring (larger dots) is SNPs shared by all

453 5 strains. Inner rings (smaller dots) are associated with the strain immediately

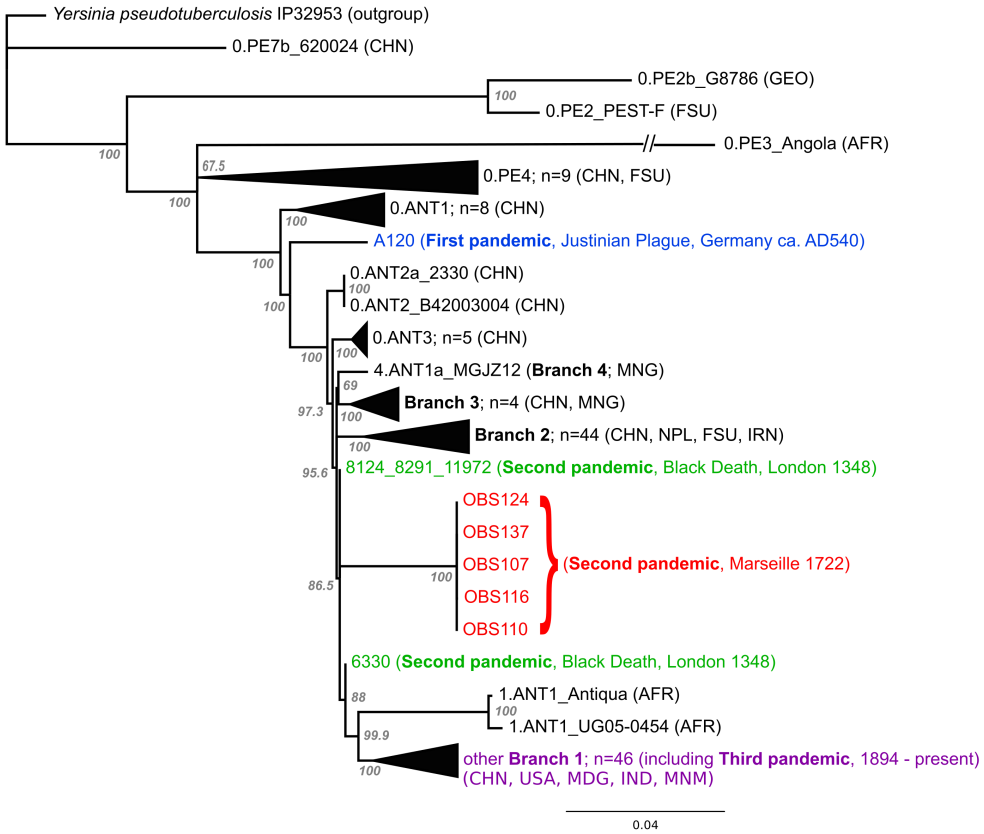
454 outside and are SNPs that are not shared by all 5 strains (but may be shared by 2-4

455 strains). Krzywinski, M. *et al.* Circos: an Information Aesthetic for Comparative

456 Genomics. Genome Res (2009) 19:1639-1645

457

A



458

459

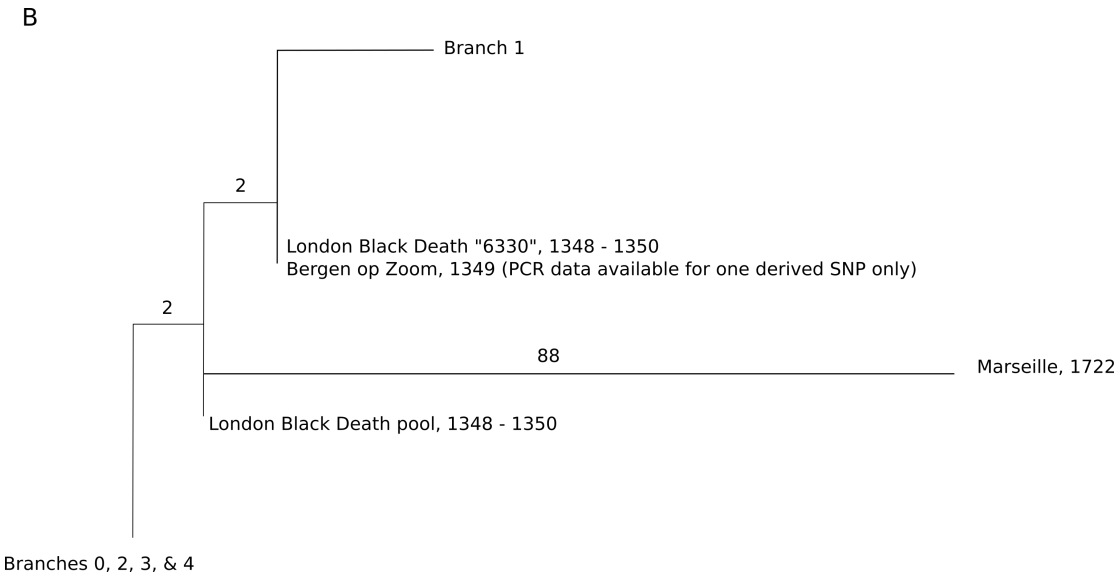


Figure 3 – A) Maximum likelihood phylogeny of *Y. pestis* genomic SNPs showing the position of the Observance (OBS) lineage (red) relative to those of 130 modern (black) and three ancient strains (Black Death in green and Justinian Plague in blue). Modern strains from the third pandemic are shown in purple to highlight their close genetic relatedness. Monophyletic groups of sequences have been collapsed to improve clarity and are shown as triangles. The tree is rooted using single strain of Y. *pseudotuberculosis* (IP32953), with all derived SNPs removed to assist scaling, with branch lengths reflecting the number of nucleotide substitutions/SNP site. The length of the branch leading to the 0.PE3_Angola (AFR) lineage was reduced because its excessive length adversely affected the scaling of the tree. Location abbreviations are as follows: CNH (China), GEO (Georgia), FSU (Former Soviet Union), MNG (Mongolia), NPL (Nepal), IRN (Iran), AFR (Africa), USA (United States of America), MDG (Madagascar), IND (India). B) Expanded phylogeny schematic to show the relative positions of the Black Death and the Observance lineages. Numbers on branches correspond to SNPs.

477

478

479 Tables:

480 **Table 1** - quantitative PCR data for the *pla* gene obtained for the 19 Observance

481 teeth extracted. Values below std curve detection are highlighted in grey.

PCR Blks:	Copies per µl
PCR Blk 1	0
PCR Blk 2	0
PCR Blk 3	0
PCR Blk 4	0

Extraction Blks:	Copies per µl
BLk 1	0
Blk 2	0
Blk 3	0
Blk 4	0

Extracts:

Sample	Copies per µl
101a	0
101b	0
102	0
103	0
107	39
108	0
109	0
110	50
116	105
117a	0
117b	1
118	0
123	0
124	142
125	0
126	0
130	0
134	1
137	30
144	0

482

483

484

485 **Table 2** – Mapping Statistics

Sample ID	raw read pairs	preprocessed reads	mapped reads	%mapped	dedupped	duplication factor	fold coverage	%covered
OBS107	14,646,710	15,395,287	1,850,734	12.02%	851,096	2.17	12.20	84.79
OBS110	9,879,257	10,458,571	1,581,158	15.12%	855,988	1.85	14.61	90.63
OBS116	16,858,273	17,876,883	2,931,554	16.40%	1,363,415	2.15	20.04	92.56
OBS124	48,797,602	52,345,884	4,972,347	9.50%	973,763	5.11	13.34	85.37
OBS137	120,018,142	129,095,434	12,669,625	9.81%	1,583,596	8.00	24.40	92.76

486

487

488 Supplementary Tables:

489 Supplementary File 1 – combined SNP table from three SNP calling approaches

490 Supplementary File 2 – list of genomes used in SNP calling and phylogeny