

Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death

Verena J. Schuenemann^{a,1}, Kirsten Bos^{b,1}, Sharon DeWitte^c, Sarah Schmedes^d, Joslyn Jamieson^b, Alissa Mittnik^a, Stephen Forrest^b, Brian K. Coombes^e, James W. Wood^{f,9}, David J. D. Earn^{e,h}, William White^{i,2}, Johannes Krause^{a,j,3}, and Hendrik N. Poinar^{b,e,3}

^aInstitut für Naturwissenschaftliche Archäologie, University of Tübingen, 72072 Tübingen, Germany; ^bMcMaster Ancient DNA Centre, Department of Anthropology, McMaster University, Hamilton, ON, Canada L8S 4L8; ^cDepartment of Anthropology, University at Albany, State University of New York, Albany, NY 12222; ^dInstitute of Investigative Genetics, University of North Texas Health Science Center, Fort Worth, TX 76107; ^eMichael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada L8S 4L8; ^fDepartment of Anthropology, Pennsylvania State University, University Park, PA 16802; ^gPopulation Research Institute, Pennsylvania State University, University Park, PA 16802; ^hDepartment of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada L8S 4L8; ⁱCentre for Human Bioarchaeology, Museum of London, London EC2Y 5HN, United Kingdom; and ^jHuman Genetics Department, Medical Faculty, University of Tübingen, 72076 Tübingen, Germany

Edited by Francisco Mauro Salzano, Instituto de Biociências, Porto Alegre, RS, Brazil, and approved July 22, 2011 (received for review March 30, 2011)

Although investigations of medieval plague victims have identified *Yersinia pestis* as the putative etiologic agent of the pandemic, methodological limitations have prevented large-scale genomic investigations to evaluate changes in the pathogen's virulence over time. We screened over 100 skeletal remains from Black Death victims of the East Smithfield mass burial site (1348–1350, London, England). Recent methods of DNA enrichment coupled with high-throughput DNA sequencing subsequently permitted reconstruction of ten full human mitochondrial genomes (16 kb each) and the full pPCP1 (9.6 kb) virulence-associated plasmid at high coverage. Comparisons of molecular damage profiles between endogenous human and *Y. pestis* DNA confirmed its authenticity as an ancient pathogen, thus representing the longest contiguous genomic sequence for an ancient pathogen to date. Comparison of our reconstructed plasmid against modern *Y. pestis* shows identity with several isolates matching the *Medievalis* biovar; however, our chromosomal sequences indicate the victims were infected with a *Y. pestis* variant that has not been previously reported. Our data reveal that the Black Death in medieval Europe was caused by a variant of *Y. pestis* that may no longer exist, and genetic data carried on its pPCP1 plasmid were not responsible for the purported epidemiological differences between ancient and modern forms of *Y. pestis* infections.

ancient DNA | paleopathology

The Black Death of 1347–1351 in Europe was one of the most cataclysmic events in history, and it is arguably “one of the most dramatic examples ever of emerging or reemerging disease” (ref. 1, p. 971). The disease is assumed to have been a particularly intense pandemic of bubonic and pneumonic plague caused by the Gram-negative bacillus *Yersinia pestis*. In this model, the medieval plague is considered the second of three pandemic waves, starting with the Plague of Justinian in A.D. 541 and culminating in the 20th century pandemic, which is still responsible for 2,000 cases/y worldwide (2) and is regarded as reemerging (3). Similarities in disease manifestation, mortality rates, and geographical distribution are generally cited as factors relating the three pandemics, though some scholars have argued that the second wave was too distinct in terms of its purported symptoms, epidemiology, and time of year of peak mortality to warrant such a connection. These discussants have argued in favor of other potential microbial associations with the medieval disease, including *Bacillus anthracis* (4), a filovirus (5), or a pathogen that has since become extinct (6).

Ancient DNA was sought to address the above controversy, although failed attempts to replicate initial work showing the presence of *Y. pestis* in purported victims of the medieval pandemic ignited skepticism regarding the identity of its etiologic

agent (7, 8). Recent publications seem to have settled the controversy with the amplification of short segments of *Y. pestis* DNA by PCR-based approaches in several skeletal collections from time periods associated with medieval plague outbreaks (9, 10), though none of these collections can be conclusively associated with the purported initial disease outbreak in medieval Europe of 1347–1351. A global survey of modern *Y. pestis* variants suggests that the ancient forms possess a unique and ancestral phylogenetic placement (9, 11). Although standard molecular methods can permit identification and limited phylogenetic resolution of ancient microbes, successful characterization of long contiguous stretches of authentic pathogen DNA will facilitate greater insight into the molecular architecture of ancient host–pathogen interactions. For *Y. pestis*, such an approach might be informative in addressing the noted differences between ancient and modern forms of the disease (4, 5, 12, 13).

PCR-based approaches are ill-suited for large-scale genetic investigations of ancient DNA owing to their preferential amplification of less damaged templates that derive from exogenous contaminants and the highly fragmented nature of endogenous molecules (14, 15). In contrast, targeted enrichment strategies (16) in combination with high-throughput DNA sequencing allow for long stretches of ancient DNA to be reconstructed from a complex metagenomic background, and it is clearly the way forward for ancient pathogen research. The authenticity of endogenous ancient sequences can then be determined by looking for patterns of nucleotide damage typical of ancient DNA (15, 17). To show the suitability of these methods for analyses of ancient pathogens, we have identified virulence-associated *Y.*

Author contributions: K.B., S.D., B.K.C., J.W.W., D.J.D.E., W.W., J.K., and H.N.P. designed research; V.J.S., K.B., S.S., J.J., A.M., S.F., J.W.W., and J.K. performed research; B.K.C. and W.W. contributed new reagents/analytic tools; V.J.S., K.B., J.K., and H.N.P. analyzed data; and K.B., J.K., and H.N.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 15669.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the Genbank database (accession nos. [HE576978](https://doi.org/10.1093/seq/seq018)–[HE576987](https://doi.org/10.1093/seq/seq019)).

¹V.J.S. and K.B. contributed equally to this work.

²Deceased December, 2010.

³To whom correspondence may be addressed. E-mail: johannes.krause@uni-tuebingen.de or poinarh@mcmaster.ca.

See Author Summary on page 15673.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105107108/-DCSupplemental.

pestis DNA fragments by PCR and independently replicated these results through subsequent targeted DNA enrichment (18) and high-throughput sequencing for several skeletal samples from victims securely dated to the initial medieval Black Death pandemic of 1348–1350 in London, England. The data presented here represent both the oldest and longest assembled authentic sequences from an ancient pathogen, and in turn, they suggest that the Black Death was caused by a *Y. pestis* variant that harbors a pPCP1 plasmid found in some modern isolates.

Results

***Y. pestis* DNA Preservation in Skeletal Remains.** Complete information regarding skeletal screening is available in [SI Materials and Methods](#). In total, we screened DNA extracts from 109 samples—53 bones and 46 teeth from the East Smithfield (ES) collection as well as 10 samples from St. Nicholas Shambles (SNS) that served as negative controls (Fig. 1 and [Table S1](#)). Total DNA content measured fluorometrically revealed less DNA in teeth than in bone, with dental extracts yielding on average 1,065 and 664 pg/μL for supernatant and pellet, respectively, compared with 1,956 and 2,280 pg/μL for bone. Quantitative PCR (qPCR) results showed the presence of amplifiable *Y. pestis* *pla* DNA in 5.7% of bones (3/53) and 37% of teeth (17/46) ([Figs. S1 and S2](#)). The *cafIM* assay yielded expected products in 5 of 17 teeth that contained amplifiable *pla*, although this finding is expected based on *Y. pestis* plasmid numbers where the pPCP1 plasmid outnumbered the pMT1 by an estimated 100 to 1 (19). No *Y. pestis* DNA was detected in any of the negative control samples. DNA sequences from the qPCR products contained both G to A and C to T transitions, damage patterns typical of ancient DNA (17). Despite their lower whole DNA content, dental samples were consistently a richer source for *Y. pestis* DNA than bone, and higher copy numbers were frequently observed in the EDTA supernatants as opposed to the pellets ([Table S1](#)). This finding is consistent with what one might expect from a blood-borne pathogen, where DNA likely resides in the desiccated blood vessels of the pulp chamber. All amplifiable DNA was extremely low in quantity, with a maximum estimated copy number of 30 copies/μL.

Multiplex PCR Data for the *pla* Gene. Using a multiplex PCR approach, we were successful in sequencing 78.3% of the *pla* gene and its flanking intergenic spacers at a minimum of four times coverage ([Dataset S1](#)). Missing regions were caused by a lack of expected PCR product. Cloned sequence data revealed 63 sites showing DNA damage reflected in the predominance of C to T (53.97%) and G to A (30.16%) transitions, which accounted for 84.13% of all aberrant positions. No transversions were observed. To distinguish damage from potential polymorphisms, all sequences were confirmed by cloned data from two independent amplified libraries. Wherever products from these two libraries did not match, cloned sequence data from a third amplified li-

brary was sought. This process permitted us to resolve all transitions with the exception of three, where single nucleotide differences from published sequences were present in a subset of clones from two independently amplified libraries ([Fig. S3](#)). The PCR data did not contain the *Microtus*-specific C to T (20) at position 1,109 in [Dataset S1](#), and it does not show the T to C transition previously reported in an ancient sequence for this region in the work by Raoult et al. (7) ([Dataset S1](#), position 809).

***Y. pestis* Chromosomal PCR Assays.** Of 11 primer sets used to define branches 1 and 2 phylogenetic placement (9), only 1 primer set yielded expected amplification products, namely s19: it produced the shortest amplicon of all chromosomal primer sets (80 bp), and it corresponds to a region in the DNA helicase II gene. Sequence data showed the absence of the CTA motif common to extant branch 2 sequences, thus supporting the notion that medieval plague was not caused by an extant branch 2 variant (9); however, this region also revealed the presence of two synonymous point mutations that, to our knowledge, are not found in any *Y. pestis* sequences either modern or ancient ([Fig. 2](#)).

Enrichment Efficiency. To evaluate the suitability of targeted enrichment for investigations of virulence in ancient pathogens, we attempted to capture one of the two *pestis*-specific virulence plasmids, namely the 9.6-kb pPCP1 along with complete human mitochondrial genomes from both Black Death and control teeth to evaluate endogenous DNA preservation. The total number of merged Illumina reads per sample that mapped to the target DNA (human mtDNA and pPCP1 DNA) varied between 0.3% and 49%, suggesting high enrichment efficiencies.

Human mtDNA Enrichment and Preservation. Sequence clustering revealed between 228 and 84,244 unique fragments for the individual human teeth mapping to the reference mtDNA ([Table 1](#)), thus revealing high quantities of mtDNA fragments in all 10 human extracts with low quantities in two of the SNS human control extracts. The cave bear control sample yielded 15 mtDNA fragments, 6 of which aligned to the human mtDNA and 9 were identical or highly similar to the cave bear mtDNA genome (21). No human mtDNA fragments were found in the extraction blank. Complete or near complete mitochondrial genomes were constructed from fragments of 58 bp average length for all 10 human samples, with an average coverage of 0.7–280× ([Table 1](#)). All 10 mtDNA genomes were different from each other, and two samples (ES 3 and 4) extracted two times yielded the identical mtDNA genome sequence between duplicates. These data suggest that there was no cross-contamination between human samples. Nine of the ten were found to correspond to typical European mtDNA haplogroups based on phylotree (22) ([Table 2](#)). Only for SNS1 was the haplogroup not determinable due to the higher contamination level ([Table 2](#)). Comparisons of ancient mitochondrial consensus sequences against a worldwide dataset of 311 mtDNAs and the Cambridge reference sequence revealed between 1 and 17 positions that were either unique to the ancient samples or present at less than 1% frequency in the current modern DNA database. Such positions were used to calculate the extent of polluting human exogenous DNA contamination within our fragments ([Table 2](#) and [Fig. S4](#)). To calculate the ratio of endogenous to exogenous human DNA, we selected all fragments that showed the rare variant(s) and considered them to be endogenous, whereas all fragments that showed a common human mtDNA variant were considered to be potential contaminants. In so doing, we found between 21 and 6,724 individual mtDNA fragments overlapping such contamination informative positions in the ES samples. The majority of these DNA fragments were found to be consistent (all unique fragments overlapping the position showed the same substitution), suggesting that the majority of mtDNA fragments from each sample is derived from

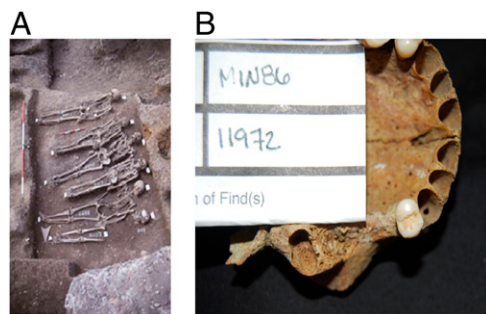


Fig. 1. The excavation at ES from 1886 to 1888 (A) and sampling at the Museum of London (B). Photos reproduced courtesy of the Museum of London.

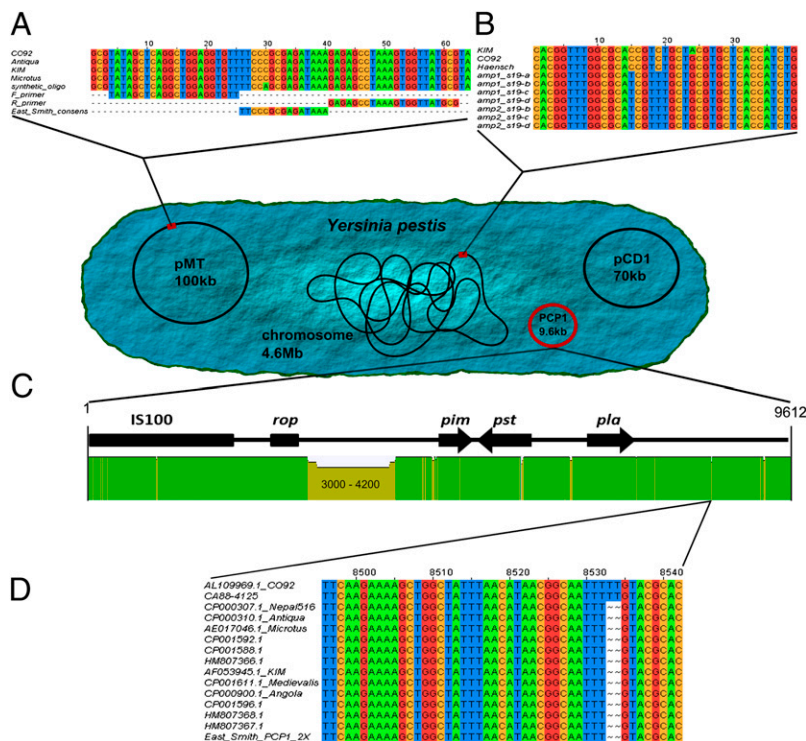


Fig. 2. Positional relationship of the sequences reported for medieval *Y. pestis*. (A) Quantitative PCR assay for the *caf1M* locus of the pMT1 plasmid. (B) Chromosomal sequence data for s19 (9) showing positions of two synonymous substitutions in the DNA helicase II gene and absence of the CTA SNP required for branch 2 designation (boxed nucleotides). (C) Solexa consensus (minimum of two-fold coverage) compared with 14 modern *Y. pestis* variants shown with positional information for genes on the PCP1 plasmid (*rop*, plasmid replication regulatory protein; *pim*, pesticin immunity protein; *pst*, pesticin; *pla*, plasminogen activator gene). Transcriptional polarity (45) is indicated by arrows. Identity is shown in green; the region between 3,000 and 4,200 is not considered (in the text). (D) Alignment of the medieval *Y. pestis* pPCP1 consensus against 14 modern variants showing presence of an indel common to branch 2 and ancestral isolates.

a single biological source (Fig. S4) (15, 23). Individual mtDNA fragments were compared against the corresponding consensus sequence, and nucleotide substitutions were recorded for each position along the DNA fragment. Based on these estimates, levels of polluting human DNA could be calculated, although these rates were very low for all samples (Table 2). In support of this notion, fragments identified as endogenous DNA contained degradation patterns typical for ancient templates (Fig. S5), again suggesting preservation of endogenous human DNA in all 10 medieval teeth.

***Y. pestis* DNA Enrichment and Preservation.** We found between 42 and 36,986 unique fragments mapping to a portion of the *Y. pestis* pPCP1 plasmid reference genome in all libraries, including the pre-Black Death human control, the cave bear sample, and the extraction blank (Table 1). All reads in the non-ES extracts mapped to a region between positions 3,000 and 4,200 of the reference pPCP1 that shows a high similarity to expression vectors that are used for recombinant enzyme production. This region was found to be problematic in a previous *Y. pestis* pPCP1 study using shotgun sequencing (24); hence, it is likely that remnants of the expression vectors are being captured by our enrich-

Table 1. Average coverage, fragment size, and number of fragments mapping to mtDNA, pPCP1 (4,200–9,612 and 1–3,000 bp), and pPCP1 (3,000–4,200 bp) for all six medieval teeth as well as two controls and the pooled fragments from the ES samples ES 1–5

Sample	mtDNA				pPCP1_1–2,999 and 4,200–9,612			pPCP1_3,000–4,199		
	Fragments (unique)	Average coverage	Average length	mtDNA haplogroup	Fragments (unique)	Average coverage	Average length	Fragments (unique)	Average coverage	Average length
SNS 1 (SN 5169)	240	0.8	55.6	na	0	0.0	na	42	1.9	54.4
SNS 2 (SN 5200)	31,973	95.8	49.7	K1a1b1	0	0.0	na	16,121	886.1	66.0
SNS 3 (SN 5194)	228	0.7	51.5	H2a	0	0.0	na	110	4.3	46.8
SNS 4 (SN 5122)	77,234	313.1	67.2	T2b	0	0.0	na	7,172	382.4	64.0
SNS 5 (SN 5053)	2,221	9.0	67	H	0	0.0	na	8,703	522.0	71.5
ES 1 (ES 6443)	19,995	69.7	57.7	HV0b	7	0.0	39.0	17,852	971.0	64.3
ES 2 (ES 8291)	6,666	26.8	66.5	H1	4,531	22.8	42.3	15,215	887.1	71.6
ES 3.1 (ES 6330)	84,244	280.2	55.1	T2g	6,435	33.1	43.3	36,986	2,216.4	74.0
ES 3.2 (ES 6330)	47,706	177.6	61.7	T2g	3,559	20.4	48.1	12,829	748.8	72.5
ES 4.1 (ES 11972)	5,436	19.5	59.5	H6a1a	2,500	12.0	40.3	13,196	727.6	66.4
ES 4.2 (ES 11972)	29,746	102.1	56.8	H6a1a	4,515	22.8	42.5	28,283	1,592.2	66.9
ES 5 (ES 8124)	7,621	24.0	52.2	U4c1a	1,530	6.9	37.8	28,784	1,656.5	68.6
Totals ES 1–5	202,406	703.2	57.6	na	23,112	118.2	43	150,789	8,735.7	69.6
Cave bear	15*	0.1	56.9	na	0	0.0	na	5,397	279.2	62.1
ExtrBlk	0	0	na	na	0	0.0	na	13,463	757.6	67.4
LibraryBlk	11	0.03	47.2	na	0	0.0	na	886	38.0	51.3

na, not applicable.

*Sequences are conserved between cave bear mtDNA (nine sequences) and human mtDNA (six sequences).

Table 2. Mitochondrial haplogroup identification based on phylotree (12) and mtDNA contamination estimates based on positions where 311 worldwide mtDNAs or rCRS differ from the consensus mtDNA sequence of a particular sample

Sample	mtDNA group	Contamination analysis by 311 mtDNA			Contamination analysis by rCRS		
		Clean fragments	Polluting fragments	Percent contamination (95% CI)	Clean fragments	Polluting fragments	Percent contamination (95% CI)
SNS 1 (SN 5169)	na	1	2	66.7 (20.8–93.9)	4	1	20.0 (3.6–62.4)
SNS 2 (SN 5200)	K1a1b1	95	0	0.0 (0.0–3.9)	2,362	2	0.1 (0.0–0.3)
SNS 3 (SN 5194)	H2a	1	0	0.0 (0.0–79.3)	15	0	0.0 (0.0–20.4)
SNS 4 (SN 5122)	T2b	218	7	3.1 (1.5–6.3)	6,365	7	0.1 (0.1–0.2)
SNS 5 (SN 5053)	H	3	0	0.0 (0.0–56.2)	77	0	0.0 (0.0–4.8)
ES 1 (ES 6443)	HV0b	69	0	0.0 (0.0–5.3)	696	0	0.0 (0.0–0.5)
ES 2 (ES 8291)	H1	21	3	12.5 (4.3–31.0)	239	3	1.2 (0.4–3.6)
ES 3.1 (ES 6330)	T2g	1,388	2	0.1 (0.0–0.5)	6,724	11	0.2 (0.1–0.3)
ES 3.2 (ES 6330)	T2g	911	0	0.0 (0.0–0.4)	4,067	5	0.1 (0.1–0.3)
ES 4.1 (ES 11972)	H6a1a	79	1	1.2 (0.2–6.7)	262	1	0.4 (0.1–2.1)
ES 4.2 (ES 11972)	H6a1a	454	3	0.7 (0.2–1.9)	1,315	3	0.2 (0.1–0.7)
ES 5 (ES 8124)	U4c1a	161	4	2.4 (0.9–6.1)	543	4	0.7 (0.1–1.9)
Cave bear	na	15	0	0.0 (0.0–27.8)	6	0	0.0 (0.0–27.8)
ExtrBlk	na	0	0	na	0	0	na
LibraryBlk	na	0	0	na	1	0	0.0 (0.0–79.3)

Clean fragments were found to be identical to the sample consensus sequence at informative positions. Polluting fragments were found to be in the same state, such as the 311 mtDNAs or rCRS. na, Not analyzed.

ment approach. Apart from these conserved motifs found across all samples and controls, the pre-Black Death human control samples (SNS) showed not a single unique fragment homologous to the pPCP1 sequence, whereas the five ES samples contained between 7 and 6,435 unique fragments that mapped to the *Y. pestis* pPCP1 plasmid (Table 1). Because all five individuals from the ES cemetery likely died from the same strain of *Y. pestis*, unique pPCP1 fragments were pooled to reconstruct a consensus sequence at maximum coverage. The pooled consensus has an average coverage of 43 bp, ranging from 0x to 702x, and an average fragment length of 36 bp (Table 1). This finding may be why a previous study using the ES collection that targeted regions in excess of 130 bp (8) was unsuccessful in replicating some original work with plague samples (7, 25). Not surprisingly, the three regions with the poorest coverage had a low GC content (%) (Fig. 3 and Fig. S6). Shotgun sequence data for ancient human samples (23) show similar patterns, suggesting a preservation bias to high GC content and thus likely ruling out enrichment artifacts. Despite lower coverage of AT-rich regions, the consensus sequence covers ~99% of the entire pPCP1 at a minimum of 2x (Table 3).

We tested the authenticity of the pPCP1 DNA by analyzing the substitution patterns along the DNA fragments and found the same nucleotide misincorporation pattern that was observed within the mtDNA fragments from the five medieval samples of the ES site. The predominant C to T substitutions found at the 5' ends and G to A substitutions at the 3' ends are typical of ancient DNA (15, 26, 27). This pattern is not seen in the fragments mapping to the vector region of the pPCP1 plasmid, lending additional support to the notion that the majority of these

fragments are indeed derived from contaminating expression vectors of the supplied reagents (Fig. 4 and Fig. S7). Fragment length distribution also indicates longer fragments obtained for this region, which is consistent with their coming from a modern contaminant source (Fig. 5).

Phylogenetic comparison of the 8,299-bp 2x consensus pPCP1 plasmid with modern *Y. pestis* reveals sequence identity with 11 of 14 extant *Y. pestis* strains; it did not match sequences typed as *Microtus* (AE017046.1), *Angora* (CP000900.1), and *Medievalis* (CP001611.1), although it does match the KIM pPCP1 (AF053945.1). It also resolves the three ambiguous positions identified by the multiplex approach of the *pla* gene and replicates the *pla* sequence described above. Considering indel polymorphisms, the ES strain does not contain the TT insertion common to the *Orientalis* biovar (AL109969.1 and CA88-4125).

Discussion

The last several decades have introduced human populations to a historically unprecedented number of emerging or reemerging infectious diseases, mostly facilitated by anthropogenic factors such as globalization of trade and human travel, changes in local ecology, and antibiotic resistance (28). Genetic diversity in pathogens is known to be a major source of phenotypic diversity underlying disease dynamics (29), hence microbial changes cannot be ignored as a potential driving force influencing host–pathogen interactions (30, 31). Genetic investigations of ancient microbes may provide much needed data to elucidate how the virulence of our close microbial companions has evolved over time. Potential factors that influenced the coevolution of humans and their pathogens in the past can be of great value to elucidate the dynamics of host–pathogen relationships in our new era of emerging infections (32, 33).

Here we report the presence of short DNA fragments (<60 bp) from two *pestis*-specific plasmids as well as one chromosomal sequence showing synonymous substitutions from individuals from a well-documented Black Death mass burial ground from 1348 to 1350. These results were obtained in a facility that has not been previously exposed to modern sources of *Y. pestis*, and the sequence data contain a damage pattern that is characteristic of ancient DNA (26). Using a targeted DNA capture approach combined with high-throughput sequencing, we obtained se-

Table 3. Coverage for the pPCP1 ES (1–3,000 and 4,200–9,612 bp) consensus sequence

Coverage	Positions (total)	Positions (covered)	Coverage (%)
One-fold	8,412	8,371	99.52
Two-fold	8,412	8,299	98.66
Three-fold	8,412	8,197	97.45
Four-fold	8,412	8,109	96.40
Five-fold	8,412	8,022	95.37

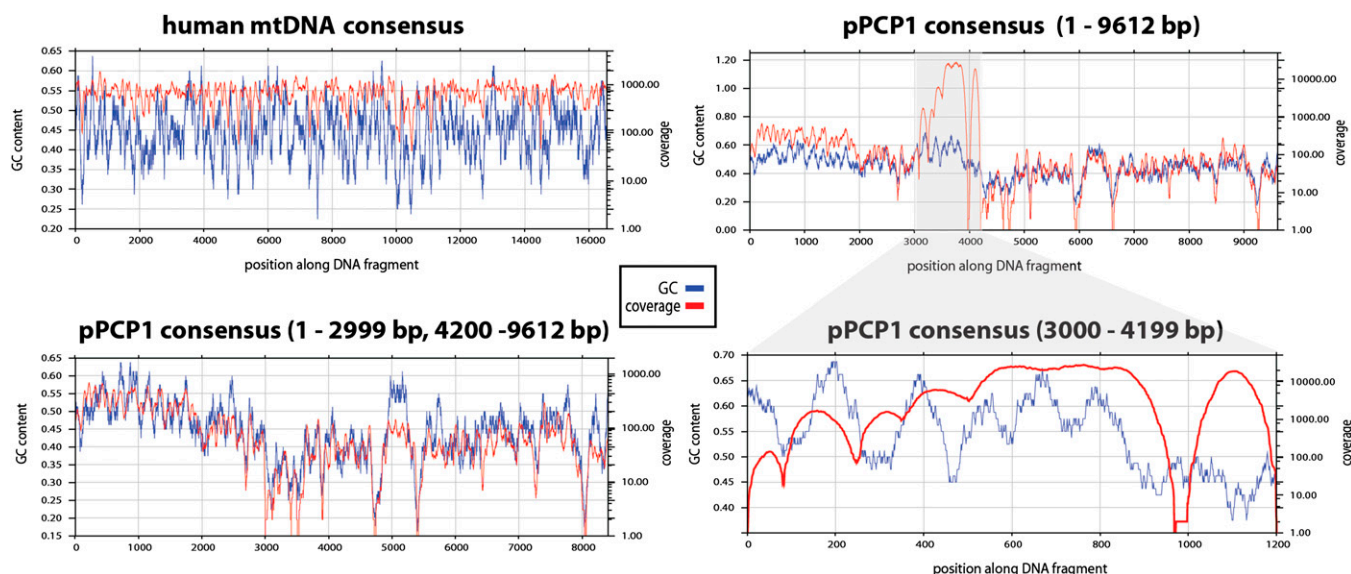


Fig. 3. GC content and DNA coverage for the consensus sequences of mtDNA, *Y. pestis* pPCP1 (1–9,612 bp), pPCP1 (4,200–9,612 and 1–3,000 bp), and the excluded region pPCP1 (3,000–4,200 bp).

quences of the pPCP1 plasmid of *Y. pestis* from all five human samples from the ES site that, when used together, reconstruct 99% of the plasmid, excluding a 1,200-bp region because of modern expression vector contamination. The pPCP1 fragments from the ES cemetery show DNA damage patterns common to ancient DNA (15, 26, 27). Analogous results were obtained for

complete human mtDNA sequences from these five victims as well as the control samples, suggesting little postexcavation human DNA contamination and adequate preservation of endogenous DNA. Given the independent replication of the *pla* gene by PCR and enrichment-based strategies in two independent laboratories, the intersample replication of portions of the

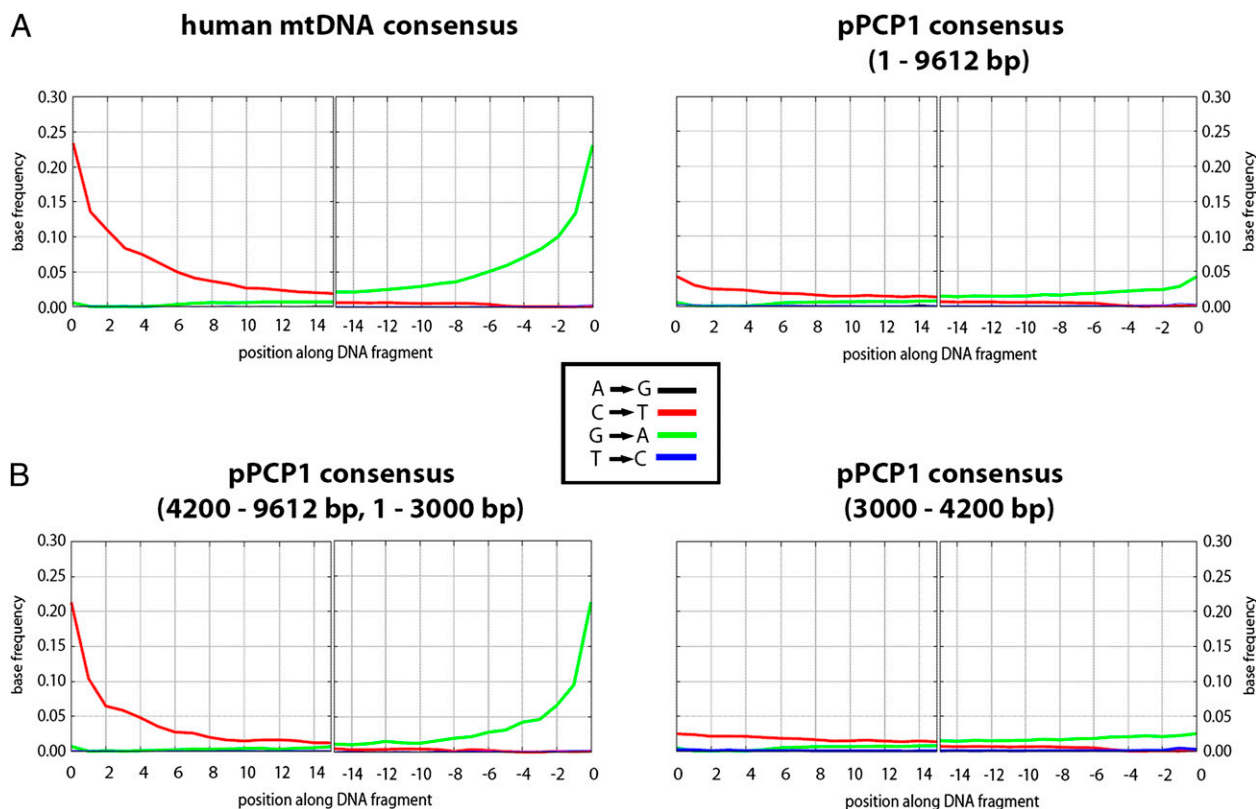


Fig. 4. Nucleotide misincorporation patterns for both the ES mtDNA consensus sequences and the complete pPCP1 (1–9,612 bp), including the section showing alignment to contaminating modern vector DNA (3,000–4,200 bp), and the entire pPCP1, excluding the mismatching region (1–3,000 and 4,200–9,612 bp).

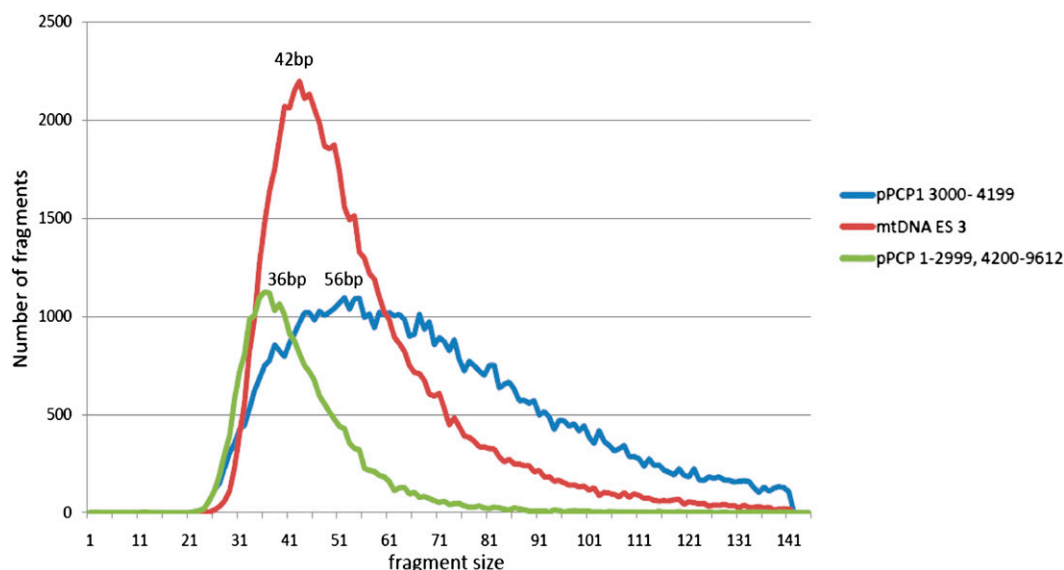


Fig. 5. Graphical depiction of length distribution vs. number of fragments for different target regions.

pPCP1 from five humans each with different mtDNA genomes, and the unambiguous DNA damage patterns, we believe that these data conclusively show the presence of *Y. pestis* DNA in medieval dental tissues from victims of the first wave of the Black Death, lending additional support to implicate this bacterium as a causative agent in the medieval plague pandemic of 1347–1351. The absence of a single *Y. pestis* DNA fragment in any of the pre-Black Death control teeth (SNS), which consisted of two samples with mtDNA yields comparable with the three best-preserved ES samples, also bolsters this claim and circumvents the concern over soil-dwelling microbial contaminants mimicking pPCP1 sequences in the ES samples.

Two of the authors (SW and JM) have previously argued that the epidemiology, virulence, and population dynamics of the Black Death were too different from those factors of modern yersinia plague to have been caused by *Y. pestis* (13). Given the growing body of evidence implicating this bacterium as responsible for the pandemic, we believe scientific debates should now shift to addressing the genetic basis of the epidemic's unique characteristics.

Y. pestis is a recently evolved variant of the soil-dwelling bacterium *Y. pseudotuberculosis*, differing mostly in the acquisition of two virulence plasmids, namely the pMT1 (100 kb) and the high copy pPCP1 (9.6 kb), which contains the pesticin (*pst*) and pesticin immunity genes (*pim*) involved in bactericidal activity and the virulence-associated plasminogen activator gene, *pla*. This important gene is associated with enhancing the bubonic form of the disease by facilitating bacterial dissemination at the site of a flea bite and encouraging bacterial proliferation in the respiratory system during pneumonic episodes. The full pPCP1 plasmid is mostly conserved between different modern *Y. pestis* variants; hence, it is not surprising that the plasmid is effectively unchanged in an ancient sample. In all, we feel confident that the pPCP1 plasmid presented here did not contribute to the purported differences between ancient and modern forms of the disease. The ES pPCP1 sequence does not match the Orientalis strain previously defined as branch 1 (11). Although an in-depth discussion of organismal phylogeny falls outside the current focus, the absence of a branch 2-specific marker in the DNA helicase II gene and the presence of two substitutions in the same region that, to our knowledge, are not found in any previously reported modern or ancient sequence suggest that the medieval form of *Y. pestis* may harbor additional information regarding the organism's evolu-

tionary history as a human pathogen. This information may prove instrumental in identifying factors that influence the different epidemiology of ancient and modern forms of the disease.

Materials and Methods

Experiments Conducted at McMaster University. Information on the archaeological site, skeletal sampling, and DNA extraction can be found in *SI Materials and Methods*. We note that the site from which the skeletons were recovered, the ES burial ground in London, is one of just a few excavated sites in the world that can be linked definitively and uniquely to the mid-14th century outbreak of the Black Death by archaeological and documentary evidence (34).

Screening. Total DNA from each extract was measured on a Turner Biosystems DNA fluorometer. PCR inhibition for each DNA extract was determined by an internal positive control (35) using a 63-bp mammoth cytochrome *b* assay to evaluate the amount of Taq Gold DNA polymerase required to overcome PCR inhibition in the extract. Initial screening of all extracts was done using a 52-bp qPCR assay specific to the plasminogen activator (*pla*) gene located on the high-copy pPCP1 plasmid of *Y. pestis* (19). All extracts yielding expected *pla* products were screened for the presence of an additional *Y. pestis*-specific marker, the chaperone protein for the fraction 1 antigen gene (*caf1M*) located on the pMT1 plasmid. Sensitivity for both *Y. pestis* assays was determined using synthetically produced sequences, which contained identical primer-binding sites flanking a *Y. pestis* sequence that differed by a single nucleotide transversion (Figs. S1 and S2). Quantitative assays were performed on an MX3000 PCR cyclor (Stratagene) or a CFX96 Real-Time PCR Detection System (BioRad), and all were sensitive to single copy template molecules. Details on PCR conditions and sensitivity are available in *SI Materials and Methods*.

454 library preparation. Library preparation was performed using the Roche manufacturer's protocol, eliminating the fractionation step (454; Roche). Libraries were quantitated in 20 μ L qPCR reactions using methods described elsewhere (36), and amplified products were used as templates for downstream multiplex applications.

Standard and multiplex PCR reactions. Forward and reverse primers were designed using Integrated DNA Technologies software (www.idtdna.com) to span the full *pla* gene and its upstream and downstream intergenic spacer regions (37) (Table S2). Limitations in primer design made it necessary to exclude a 123-bp region flanking the transcription start site and an 86-bp region flanking the start of the downstream intergenic region, leaving a 1,191-bp target region. Multiplex reactions were carried out using methods described elsewhere (38). Primer pairs were partitioned into 22 pools through the use of a dynamic programming algorithm for interval scheduling (39), with compatibility assessed by alignment score and melting temperature (T_m). DNA sequences were confirmed in at least two independent amplified library fractions. For phylogenetic placement, primers for a subset of the branch 1 and 2 diagnostic chromosomal SNPs (40) were used in single PCR reactions using primers and PCR conditions described in

ref. 9. Primer sequences, PCR conditions, and the consensus sequence are described in *SI Materials and Methods*.

Experiments Conducted at the Max Planck Institute in Leipzig, Germany. DNA extraction and enrichment. To independently replicate the PCR-amplified *Y. pestis* DNA from the ES collection and test the suitability of enrichment strategies and subsequent targeted high-throughput sequencing for ancient pathogens, several teeth were analyzed at the Max Planck Institute in Leipzig, Germany. Five dental roots from mature molars from the ES collection were chosen based on *pla* copy numbers and analyzed alongside five dental roots from St. Nicholas Shambles (Table S1), and an ancient cave bear sample was used as a cross-contamination control. An aliquot from each DNA extract was used to produce libraries using a modified Illumina multiplex protocol (41). Endogenous mitochondrial DNA was captured in parallel from all libraries. To enrich for target DNA, we used long-range PCR products as bait for molecular capture through hybridization (18). Additional details on extraction, library preparation, and enrichment are available in *SI Materials and Methods*.

Solexa sequencing and analysis (Leipzig). The enriched library pool was sequenced on an Illumina Genome Analyzer IIx using the manufacturer's protocols for multiplex sequencing with modifications. The raw reads were aligned to the PhiX 174 reference sequence to obtain a training dataset for the base caller Ibis (42). Raw sequences called by Ibis 1.1.1 were filtered for the individual indexes as previously described (43). The paired-end reads were subjected to a fusion process (including removal of adapter sequences and adaptor dimers) by requiring at least an 11-nt overlap between the two reads. In the overlapping sequence, quality scores were combined, and the base with the highest base quality score was called. Only sequences merged in this way were used for additional analysis. The sequencing data were analyzed starting from QSEQ

sequence files and CIF intensity files from the Illumina Genome Analyzer RTA 1.6 software. Enrichment efficiency was calculated by mapping all 37,502,405 merged reads to the revised Cambridge Reference Sequence (rCRS) for human mtDNA and the *Y. pestis* CO92 pPCP1 plasmid using a custom iterative mapping assembler (23). From this information, the fraction of total reads that mapped to the target DNA was calculated.

Phylogenetic analysis. The consensus sequences for the mitochondrial genome fragments were analyzed using a custom Perl script to identify individual haplogroups based on phylotree (22). All pPCP1 fragments from the ES samples were combined into a single consensus sequence, and they were aligned and compared by eye with 14 modern pPCP1 sequences using the software Muscle (44) (*SI Materials and Methods*).

ACKNOWLEDGMENTS. We thank Tomislav Maricic and Martin Kircher for technical expertise and analysis assistance. We thank Svante Pääbo for access to the clean room and laboratory facilities of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. We also thank Rebecca Redfern and Jelena Bekvalac for providing access to skeletal samples and facilities for sampling at the Museum of London, Elizabeth Carniel for providing modern *Yersinia pestis* DNA, the entire team at the McMaster Ancient DNA Centre for helpful comments on the design and implementation of laboratory work, and D. Ann Herring and Debi Poinar for helpful comments on earlier versions of the manuscript. Financial support was provided by the Social Sciences and Humanities Research Council of Canada, the Canadian Institute for Health Research, Canada Research Chairs program, the Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, the University at Albany Center for Social and Demographic Analysis, and the Human Genetics Department, Medical Faculty, University of Tübingen, Germany.

1. Wheelis M (2002) Biological warfare at the 1346 siege of Caffa. *Emerg Infect Dis* 8: 971–975.
2. World Health Organization (2011) Zoonotic Infections. Available at www.who.int/vaccine_research/diseases/zoonotic/en/index3.html. Accessed August 11, 2011.
3. Stenseth NC, et al. (2008) Plague: Past, present, and future. *PLoS Med* 5:e3.
4. Twigg G (1984) *The Black Death: A Biological Reappraisal* (Batsford Academic, London).
5. Scott S, Duncan CJ (2001) *The Biology of Plagues* (Cambridge University Press, Cambridge, UK).
6. Cohn SK (2003) *The Black Death Transformed: Disease and Culture in Early Renaissance Europe* (Arnold, London).
7. Raoult D, et al. (2000) Molecular identification by “suicide PCR” of *Yersinia pestis* as the agent of medieval black death. *Proc Natl Acad Sci USA* 97:12800–12803.
8. Gilbert MT, et al. (2004) Absence of *Yersinia*-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology* 150:341–354.
9. Haensch S, et al. (2010) Distinct clones of *Yersinia pestis* caused the black death. *PLoS Pathog* 6:e1001134.
10. Wiechmann I, Harbeck M, Grupe G (2010) *Yersinia pestis* DNA sequences in late medieval skeletal finds, Bavaria. *Emerg Infect Dis* 16:1806–1807.
11. Morelli G, et al. (2010) *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* 42:1140–1143.
12. Wood JW, Ferrell RJ, Dewitte-Aviña SN (2003) The temporal dynamics of the fourteenth-century Black Death: New evidence from English ecclesiastical records. *Hum Biol* 75:427–448.
13. Wood JW, DeWitte-Aviña SN (2003) Was the Black Death yersinia plague? *Lancet Infect Dis* 3:327–328.
14. Pääbo S, et al. (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38: 645–679.
15. Krause J, et al. (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20:231–236.
16. Burbano HA, et al. (2010) Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328:723–725.
17. Hofreiter M, Jaenicke V, Serre D, Haeseler Av A, Pääbo S (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29:4793–4799.
18. Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004.
19. Parkhill J, et al. (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413:523–527.
20. Song Y, et al. (2004) Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res* 11:179–197.
21. Krause J, et al. (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol* 8:220–232.
22. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30:E386–E394.
23. Green RE, et al. (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134:416–426.
24. Plunkett G, Anderson BD, Burland V, Cabot EL, Glasner JD, et al. (2007) *Yersinia pestis* CA88-4125 whole genome shotgun sequencing project, direct submission to NCBI, accession number NZ_LBNC01000008.1.
25. Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D (1998) Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. *Proc Natl Acad Sci USA* 95:12637–12640.
26. Briggs AW, et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci USA* 104:14616–14621.
27. Brotherton P, et al. (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35:5717–5728.
28. Barrett R, Kuzawa CW, McDade T, Armelagos GJ (1998) Emerging and re-emerging infectious diseases: The third epidemiologic transition. *Annu Rev Anthropol* 27: 247–271.
29. Lawrence JG (2005) Common themes in the genome strategies of pathogens. *Curr Opin Genet Dev* 15:584–588.
30. Morse SS (1995) Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1: 7–15.
31. Pybus OG, Rambaut A (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 10:540–550.
32. Baum J, Khalina Bar-Gal G (2003) *Emerging Pathogens Archaeology, Ecology, & Evolution of Infectious Diseases*, eds Greenblatt C, Spigleman M (Oxford University Press, Oxford), pp 67–78.
33. Gandon S, Buckling A, Decaestecker E, Day T (2008) Host-parasite coevolution and patterns of adaptation across time and space. *J Evol Biol* 21:1861–1866.
34. Grainger I, Hawkins D (1988) Excavations at the Royal Mint site 1986–1988. *Lond Archaeol* 5:429–436.
35. King CE, Debruyne R, Kuch M, Schwarz C, Poinar HN (2009) A quantitative approach to detect and overcome PCR inhibition in ancient DNA extracts. *Biotechniques* 47: 941–949.
36. Meyer M, et al. (2008) From micrograms to picograms: Quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res* 36:e5.
37. Kim TJ, et al. (2007) Direct transcriptional control of the plasmidogen activator gene of *Yersinia pestis* by the cyclic AMP receptor protein. *J Bacteriol* 189:8890–8900.
38. Krause J, et al. (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439:724–727.
39. Bos KI, Forrest SA, Poinar HN (2009) Metagenomics and ancient human disease: Efficient targeted retrieval of ancient pathogen DNA for high-throughput sequencing applications. Presented at the 37th Annual Meeting for the Canadian Association for Physical Anthropology.
40. Achtman M, et al. (2004) Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci USA* 101:17837–17842.
41. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010:pdb. prot5448.
42. Kircher M, Stenzel U, Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* 10:R83.
43. Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M (2007) Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res* 35:e97.
44. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
45. Rakin A, Boolgokawa E, Heesemann J (1996) Structural and functional organization of the *Yersinia pestis* bacteriocin pesticin gene cluster. *Microbiology* 142:3415–3424.