THE LANCET Infectious Diseases

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Wagner DM, Jennifer Klunk J, Harbeck M, et al. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect Dis* 2014; published online Jan 28. http://dx.doi.org/10.1016/S1473-3099(13)70323-2.

APPENDIX 1

Supplementary information for 'The Plague of Justinian was caused by a 'dead-end' emergence of *Yersinia pestis*'

Specimens Obtained from Aschheim-Bajuwarenring Cemetery

Teeth came from two individuals from the Aschheim-Bajuwarenring cemetery in Aschheim, Bavaria, Germany. Material from each individual was radiocarbon dated as reported previously (1). Individual A120 was dated to 435–631 cal. AD and individual A76 was dated to 443–566 cal. AD, both with 95.4% probability. Only teeth that were still in anatomical position at the time of excavation were used in extraction and analysis. Four teeth from individual A120 were used and one was used from A76. Multiple subsamples were taken from each of the five teeth and multiple libraries were made from the subsamples. The sequencing data from all of the libraries was combined.

DNA Extractions

The ancient DNA facilities include clean rooms dedicated to ancient DNA sampling, extraction, PCR set-up, and library preparation. The clean room facilities are physically separated from the modern laboratory, where amplification, enrichment, and reamplification are performed. The facilities in which the laboratory methods were performed had no previous exposure to modern *Yersinia pestis* DNA. The sampling and subsequent methods were performed on four sets of samples. Only the samples that were sequenced are listed below:

Set 1:

Individual	Tooth Number	Associated Libraries
A120	4.8	LP12a
		LP12b
	4.7	LP05
		LP05U
		LP14a
		LP14b
		LP15a
		LP15b
	2.8	LP16a
		LP16b
		LP17a
		LP17b
A76	3.7	LP07
		LP07U

Set 2:

Individual	Tooth Number	Associated Libraries
A120	4.6	LP29
		LP30
	4.8	LP31
		LP32
	4.7	LP35
		LP36
	2.8	LP37
		LP38
A76	3.7	LP39

Set 3:

Individual	Tooth Number	Associated Libraries
A120	4.6	LP29b
		LP30b
	4.8	LP31b
		LP32b
	4.7	LP35b
		LP36b
	2.8	LP37b
		LP38b
A76	3.7	LP39b

Set 4:

Individual	Tooth Number	Associated Libraries
A120	4.6	LP49
	4.8	LP51
	4.7	LP53
	2.8	LP55
A76	3.7	LP57

We extracted DNA using a slightly modified version of Schwarz et al (2009) (2). A piece of root and/or pulp cavity weighing 27-127 mg was cut from each tooth using a rotary tool diamond cutting wheel and then crushed with a hammer. The pulverized material was demineralized with 0.5M EDTA (pH 8.0) for 24 hours at room temperature while shaken at 1000 rpm in a Thermomixer. The supernatant was removed and stored at -20°C. The remaining dental material was then digested with proteinase K in 0.5% sarcosyl buffer for 24 hours at 25°C with shaking at 1000 rpm. The supernatant was removed and stored at -20°C. The demineralization and digestion steps were then repeated, but for 12 hours instead of 24 hours. A minimum ratio of one extraction blank to seven samples was maintained throughout the experiment. The samples were concentrated with Amicon Ultracel 10kDa columns after washing the filters with 450 µL of 0.1x TE pH 8.0 and spun to minimum volume. The sample was applied stepwise and spun to minimum retention in 450 µL increments and washed twice with 450 µL of 0.1x TE. The resulting concentrate was removed via pipette and then purified with the Qiagen MinElute PCR Purification Kit using the standard protocol but with two washes of 650 µL Buffer PE instead of one 750 µL wash.

Quantitative PCR Screening for *Y. pestis*

The DNA extracts were screened for the presence of *Y. pestis*-specific *pla* gene of the pPCP1 plasmid as previously described (3). Each 10uL qPCR reaction comprised 1X PCR Buffer 2, 2 mM $MgCl_2$, 250 µM each dNTP, 1 mg/mL BSA, 250 nM each primer, EvaGreen Dye at 0.5X concentration, and 0.5 units AmpliTaq Gold polymerase and 1 µL DNA extract. Thermal cycling consisted of 5 minute activation at 95°C followed by 50 cycles of denaturation for 20 seconds at 95°C, annealing for 20 seconds at 60°C, and extension for 20 seconds at 72°C. A final elongation at 72°C for 1 minute was followed by a dissociation gradient from 65°C to 95°C in 0.5°C increments. Due to anomalies in standards under 10,000 copies, the assay should not be used to quantify the number of copies of pPCP1 in the samples, but rather to simply test for the presence of *Y. pestis* DNA. Samples exhibiting amplification within 50 cycles and a product melting temperature within 0.5C of the standards were converted to sequencing libraries. No blanks showed positive amplification of the expected product melting temperature. An example of the *pla* qPCR amplification and melt curves can be found in Supplementary Figure S1.

Library Preparation and Indexing

A previous protocol (4) with recommended modifications (5) was followed for library preparation with a purified extract, extraction blank control, or water input volume of 25 µL, except Set 3, for which the input volume was 10 µL. One library blank was used for each set of libraries prepared. For all libraries except LP05 and LP07, which were used to capture and sequence the human mitochondrial genome, the blunt end repair step was preceded by treatment with UracilDNA glycolsylase (UDG) and Endonuclease-VIII (EndoVIII) to remove deaminated cytosine residues. All enzymes were purchased from New England BioLabs. Each 40uL library preparation included an end repair step with 0.4 U/µL T4 polynucleotide kinase; 0.1 U/µL Uracil-DNA glycosylase; 0.4 U/µL Endonuclease VII; 0.2 U/µL T4 DNA polymerase and 10 or 20uL DNA extract. Each ligation reaction included 0.125 U/µL T4 DNA ligase; and each fill-in reaction included 0.4 U/µL Bst polymerase (large fragment). QIAquick MinElute PCR purification, with the above modifications, was substituted for SPRI bead clean up between library preparation stages. After the fill-in stage, a heat deactivation at 80°C for 20 minutes was done in place of an additional purification.

The heat-deactivated fill-in reactions were then double-indexed(5)in real-time using Accuprime Pfx polymerase (Invitrogen) at a concentration of 0.025 U/uL and indexing primers a at 400 uM for sets 1-3 and 100 uM for set 4. Each 50uL indexing reaction included library input volumes as follows: set 1, 10 µL; set 2, 15 µL; set 3, 10 µL; and set 4, 17.5 µL. The number of cycles of index amplification also varied by set: set 1, 10 cycles; set 2, 10 cycles; set 3, 8 cycles; set 4, 9 cycles, with cycle numbers corresponding to the point at which the majority of samples reached amplification plateau. After indexing amplification 100 µL of indexed library was purified with MinElute using the modified protocol to a final volume of 15 µL. We then screened 0.1X concentrations of the purified indexed libraries for the presence of the *Y. pestis* plasmid pPCP1 using the established *pla* qPCR assay (3). Total DNA concentration of the purified indexed libraries was determined with the Nanodrop 2000 UV spectrophotometer (ThermoScientific).

Enrichment

Array Design

A one million (1M) Agilent SureSelect DNA Capture Array was custom designed using the program PanArray (6)to target and capture sequences from the core *Y. pestis* chromosome (strain CO92, NC-003143), three primary plasmids (pPCP1, pMT1, and pCD1), and genes from other *Y. pestis* strains not present in the core chromosomal genome, with all targets listed in Supplementary Table S1. The two-million 60-mer probes were tiled every 5 bases and distributed in genomically contiguous sections across two 1M arrays, hereafter called arrays A and B.

Array Enrichment

Array enrichment was performed in two separate experiments on library sets 1 and 2 using the Agilent SureSelect DNA Capture Array Protocol (v1.0). For the enrichment on set 1, a mastermix for both arrays was made using 23.5 µg of DNA from the combined indexed libraries in 40 µL, 196 µL of nuclease-free water, 80 µL of 4 (all reverse) blocking oligos (BO2.P7.part2.R, BO2.P7.part1.R, BO4.P7.part1.R, and BO6.P7.part2.R) all oligos were at 50 µM each, 100 µL of Cot-1 DNA for a final concentration of 0.1 mg/mL, 104 µL of Agilent 10X Blocking Agent, and 520 µL of Agilent 2X Hi-RPM Hybridization Buffer. 65 µL of the mastermix was aliquotted into low-bind PCR tubes and the following thermal profile was run: 95°C for 3 minutes, 37°C for 30 minutes. The samples were pooled in 1.5-mL tubes and spun for 1 minute at maximum speed in a centrifuge. 500 µL of the mastermix was loaded onto each array, ensuring that any bubbles in the slide sandwich were moving freely. However, during slide sandwich assembly, Array B leaked and only 100 µL was recovered. It was concentrated with a 30kDa Amicon Ultracel filter and washed once with nuclease-free water to a final volume of 105 µL. In total 2.35 µg of DNA was recovered. An additional 6.36 µg of DNA in a volume of 57 µL was used in a second mastermix to bring the total DNA concentration to approximately 8.71 µg. The same concentrations of the other reagents described above were used in the replacement mastermix. 500 µL of the new mastermix was loaded onto a fresh Array B microarray slide. The used Array B was stored in Oligo aCGH Wash Buffer 1. Both loaded arrays were placed in a SciGene hybridization oven at 65°C for 65 hours at 10rpm. After 65 hours, the non-enriched fraction was collected with a 1-mL 30G syringe, purified by MinElute with the modified protocol, and stored at -20°C. For the enrichment of set 2, 33.3 µg of DNA in 80 µL, 130 µL of nuclease-free water, and 200 µM of each of the 4 reverse Blocking Oligos were used. All other conditions were the same as Array A from the set 1 enrichment.

The slide sandwiches were placed into a room temperature bath of Oligo aCGH Wash Buffer 1 (Agilent) and disassembled. The microarray slides were placed in a slide rack, which was then placed into a separate room temperature bath of Oligo aCGH Wash Buffer 1 for 10 minutes with stirring. The slide rack was then transferred to a 37°C bath of Oligo aCGH Wash Buffer 2 for 5 minutes with stirring. The slides were then dried in a centrifuge at 600 rpm for 30 seconds. The slide sandwich was then reassembled with a clean gasket slide and 490 µL of nuclease-free water and placed into the hybridization oven (SciGene) for 10 minutes at 95°C with rotation at 10 rpm. The liquid was then removed from the hybridization chamber using a 1-mL 30G syringe. Following elution of set 2, the eluates were pooled and then concentrated with 30kDa Amicon Ultracel filters to the minimum volume mark. The enriched fraction was then reamplified in 50 µL reactions consisting of 25 µL template and 25 µL mastermix to final concentrations of 1X Accuprime Pfx Reaction mix, IlPr_reampP5F and IlPr_reampP7R each at 400 nM, 0.5X EvaGreen, and 0.025 U/uL AccuPrime Pfx DNA Polymerase. The thermal profile consisted of a 4 minute activation at 95°C and 21 (set 1) or 10 (set 2) cycles of denaturation at 95°C for 15 seconds, annealing at 60°C for 30 seconds, and extension at 68°C for 30 seconds. The samples were pooled by array to 100 µL per MinElute and eluted in 15 µL Buffer EB. For set 1, they were also concentrated over Amicon 30K to a final volume of 118 µL.

A second round of enrichment was then performed on each set of samples following the same protocol as the first round of enrichment. For set 1, the Array B that had been stored in Oligo aCGH Wash Buffer was used without any leaking issues. For set 1, the post-enrichment reamplification was run for 12 cycles. For set 2, we quantified the libraries using primers targeting the insert-proximate adapter sequence in 10 µL reactions with 1 unit 10X PCR Buffer II, 2.5 mM of $MgCl_2$, 250 µM each dNTP, 200 nM of each primer (IlPr_shortampP5F, IlPr_shortampP7R) (4), 0.5 units of EvaGreen, and 0.05 U/ µL units of AmpliTaq Gold. The thermal profile consisted of a 5 minute activation at 95°C; and 40 cycles of 20 seconds at 95°C, 20

seconds at 62°C, 20 seconds at 72°C; a final extension for 90 seconds at 72°C; and a dissociation gradient from 65-95°C. From these quantifications it was determined that the samples enriched on Array A needed 3 additional cycles of reamplification to reach the concentration necessary for successful sequencing and the samples enriched on Array B needed 12 additional cycles of reamplification to reach the same concentration, which were performed using the same protocol as described for set 1. After reamplification, the samples were purified by MinElute with the modified protocol and eluted in 20 µL EB.

In-solution bait design

(a) Human mitochondrion. We designed an in-solution human mtDNA bait set using GenBank Accession number J01415.2 with probes tiled at one 80mer bait per 21bp for a total of 785 baits.

(b) *Yersinia pestis* chromosome. We designed 316,781 80mer probes tiled every 20bp for in-solution enrichment from the same set of *Yersinia* genomes used for the array enrichment.

(c) *Yersinia pestis* SNPs. To enhance coverage of 142 novel SNPs found in our draft genome as well as 2,298 previously confirmed polymorphic sites across the *Y. pestis* global phylogeny (7) we designed 81-mer baits covering the SNP site and 60bp, 40bp, and 20bp upstream and downstream of the site (three unique baits each site).

In-solution enrichment

In-solution enrichment was performed on sets 1, 2, 3, and 4 according to manufacturer's protocol. Two samples from set 1 (LP05 and LP07) were enriched using the human mitochondrion bait set. All samples in set 2 were enriched using the *Y. pestis* chromosome bait set. Sets 3 and 4 were enriched using the two SNP bait sets. Two rounds of in-solution enrichment were done on sets 2, 3, and 4. The published MYbaits protocol was followed for all sets with modifications to the bait concentrations, hybridization temperatures, and blocking agents used. For set 1, the bait concentration used was 50 ng/reaction, the hybridization temperature was 60°C, and each of the 4 forward Blocking Oligos was used at 2 µM in lieu of Blocking Agent 3. For set 2, the bait concentration used was 500 ng/reaction for the first round and 100 ng/reaction for the second round, the hybridization temperature was 45°C for the first round and 50°C for the second round, and each of the 4 forward Blocking Oligos was used at 2 µM instead of Blocking Agent 3. For sets 3 and 4, the bait concentration used was 100 ng/reaction for both rounds of enrichment, the hybridization temperature was 45°C for the first round and 50°C for the second round, and each of the 4 forward Blocking Oligos (BO1.P5.part1.F, BO1.P5.part2.F, BO3.P7.part1.F, BO5.P7.part2.F) was used at 2 µM instead of Blocking Agent 3. Additionally, 20 µL of beads per reaction was used instead of the 50 µL used in the published protocol. The reamplifications after each round of enrichment were done using the Accuprime profile found in the Array Enrichment section, except the primers were at 300 nM for set 1, 500 nM for set 2, and 100 nM for sets 3 and 4. Set 1 was amplified for 22 cycles, set 2, 3 and 4 were amplified for 5 in between rounds and 12 after the second round. Following amplification, each sample was purified by MinElute with modified protocol and eluted in 15 µL EB for set 1 or 10 µL EB for sets 2, 3, and 4.

Sequencing

Sequencing was done on the Illumina HiSeq 1500 platform at the Farncombe Family Digestive Health Research Institute of McMaster University. Set 1 was sequenced on two lanes with all samples in the set pooled in equimolar ratios using a 2x50bp kit. Set 2 was sequenced on 28% of two lanes with samples equilibrated using a 2x100bp kit. Set 3 was run on 68% of two lanes with samples equilibrated using a 2x100bp kit. Set 4 was run on 5% of two lanes in rapid run mode with all samples in the set pooled equally using a 2x80bp kit.

Shotgun sequencing

To compare the percentage of *Y. pestis* DNA in the enriched and non-enriched samples, aliquots of two different non-enriched libraries were spiked onto the run of Set 4. Although the number of amplification cycles experienced by the shotgun libraries was less than any of the enriched libraries, the data obtained can still provide a basis for comparison. An average of 0.09% of the reads from the two shotgun libraries mapped to CO92 (NC_003143).

Trimming and Merging

Adapter sequences 3' of the inserts were trimmed with cutadapt version 1.0 (8)using the common linker

sequence as the query (AGATCGGAAGAGC) tolerating an error rate of 0.16 and a minimum overlap of 1 bp. The reads were then merged using FLASH version 1.2.4(9)with a minimum overlap of 11 bp and a mismatch ratio of 0.15. Non-mergeable Read 1 reads were combined with the merged dataset for all subsequent alignments.

Mapping

The trimmed, merged reads were mapped using BWA version 0.7.5 (10)allowing 15% of alignments to be missed assuming a 2% error rate (-n 0.15) and a maximum number of gap opens of 2 (-o 2) to a CO92 *Y. pestis* reference genome (GenBank Accession NC_003143) with all repeat, tRNA, rRNA regions replaced with 50 N's in order to exclude these multi-copy and conserved regions in calculations of average *Yersinia*-specific coverage. The reads were also aligned with BWA at the same settings to plasmids pCD1 (GenBank Accession NC_003131), pMT1 (NC_003134), and pPCP1 (NC_003132). The resulting files were collapsed by indexed library using a custom SAM collapser that uses a BED file intermediate to collapse reads that share the same 5' position, 3' position, and strand, keeping the read sequence with the highest mapping quality. The collapsed reads were imported into Geneious version 6.1.2(11). Based on the aligned read length distribution we determined inserts less than 20 bases likely represented ambiguous alignments, and so we then re-aligned only reads 21bp and longer. These reads were extracted and exported into a single fastq file, which was then aligned to the unmasked CO92 *Y. pestis* reference genome (NC_003143). Of the 511,842,485 raw reads generated, 27,830,130 (5.4%) mapped to the CO92 chromosome, 5,221,655 (1.02%) to pCD1, 5,744,198 (1.12%) to pMT1, and 4,788,475 (0.94%) to pPCP1. After collapsing, 683,919 chromosomal, 90,987 pCD1, 226,658 pMT1, and 82,945 pPCP1 mapped reads mapped uniquely. Insert length distributions were generated (Supplementary Figures S2-S7). The spikes in insert length at 64 and 70 bp are caused by reads that are longer than the sequencing length in a single direction but are unable to be merged with a read from the opposite direction. Coverage metrics for the chromosome and plasmid alignments can be found in Supplementary Table S2.

Human Mitochondrial Genome Sequencing

Non-UDG treated samples from both individuals were enriched for human mitochondrial DNA, as described above, and sequenced on an Illumina MiSeq. The reads were trimmed, merged, and mapped using BWA as described above, using the rCRS human mitochondrial reference sequence (GenBank Accession #NC_012920). Full mitochondrial genomes were obtained for both individuals and SNPs were called by Geneious at 5X minimum coverage and 90% minimum variant frequency (Supplementary Tables S3 and S4). The haplogroup of each individual was determined using MitoTool.org (12) full mitochondrial variant search. The best match for individual A120 was H8c and the best match for individual A76 was J1c3. For individual A120, out of 4,121,642 raw reads, 270,240 (6.6%) reads aligned to the reference, and of the mapped reads, 54,443 reads were unique. For individual A76, of 4,475,211 reads, 912,887 reads aligned (20.4%), and of those 352,865 reads were unique. Coverage metrics can be found in Supplementary Table S5.

SNP Calling

We identified SNPs specific to the Justinian samples using Geneious (11) at 5X read coverage and 90% variant frequency, and Freebayes version 0.9.9(13) at 5X read coverage and 90% variant frequency with the following parameters: use mapping quality, exclude unobserved genotypes, alternate allele frequency, ploidy of 1, minimum mapping quality of 20, minimum base quality of 20, exclude multiple nucleotide variants, no complex events, no indels, and disabled population priors. SNPs that were called by only one of the two programs were visually examined for authenticity and only included in subsequent phylogenetic analysis when total variant calls were $\geq$ 90%. These final SNPs are available in Supplementary Table S6.

The SNPs used to infer the maximum likelihood phylogenetic tree were called across 133 *Y. pestis* genomes (see Supplementary Excel spreadsheet Table S7) with GATK(14) using the EMIT_ALL_CONFIDENT_SITES method. From all SNPs called, the 2,298 SNPs described by Cui *et al*. (7).were extracted from the variant call format (VCF) files produced by GATK. For basic genotyping, positions were called if they had a minimum read coverage of 5. If no confident call could be made by

GATK, an "not called" was inserted at that position (Supplementary Excel spreadsheet Table S8). We extracted the 2,298 Cui *et al*. (7) SNPs from both A120 and A120+A76.

There were a total of 176 chromosomal SNPs called relative to CO92, 169 called in A120 and 150 called in A120 + A76 (Supplementary Table S9). There were a total of 40 non-synonymous SNPs. Of these, 38 were called in A120 and 36 were called in A120+A76 (Supplementary Table S10). There were a total of 34 synonymous SNPs as analyzed by Geneious using the bacterial genome setting. Plasmid SNPs were called at 5X coverage and 90% variant frequency by Geneious: 7 SNPs were found in pCD1, 3 in pMT1, and 0 in pPCP1 (Supplementary Tables S11 and S12). Sixty six of the 176 SNPs were unique to A120 and A120 + A76 (Supplementary Table S6). Chromosomal SNPs were compared to those found by Cui et al. (7). The 26 SNPs removed from the analyses due to poor coverage are listed in Supplementary Table S13.

Ancient DNA Damage Assessment

We assessed the level of DNA damage in our mapped CO92 reads as well as our mitochondrial DNA with mapDamage version 3.6 (15) using the default settings (Supplementary Figure S8-S15). As the majority of the sequenced samples were treated with UDG and EndoVIII to remove the deaminated cytosine residues, the overall damage is expectedly low in these samples as compared to non-treated samples. Damage profiles from both the human mitochondrial and *Yersinia* alignments of non-UDG-treated LP05 contrast greatly with their treated counterparts (from LP05U) (Supplementary Figures S8 and S9).

Identifying non-CO92 *Yersinia sp.* Regions in Justinian and the East Smithfield Strain (ESS) sample 8291.

To identify unique regions found in our draft genomes, the *Y. pestis* biovar Microtus str. 91001 (accession NC_005810.1) and the *Y. pestis* CO92 (accession NC_003143.1) chromosomes were decomposed into 31-mers using Jellyfish (16). These sequences were sorted and compared to identify the set of 31-mers found only in Microtus.

i) Identification of DFR4 in Justinian Strains. Merged and trimmed reads were mapped to the *Y. pestis* biovar Microtus str. 91001 chromosome as previously described (SOM section II B). Additionally, the unique Microtus 31-mers were also mapped back to the Microtus chromosome. The resulting alignment files were compared using BEDTools (17)to identify intervals of the Microtus chromosome that were unique to Microtus when compared to CO92 and were covered by reads from our ancient extract. Using bedtools, fractional coverage was determined for genomic features for a region between approximately 1041000 and 1063000 bp. (Supplementary Figure S16). This region corresponds to the DFR4 fragment identified previously (18).

ii) Identification of DFR4 in ESS – Black Death. A similar procedure was used to map data from UDG (LP04U) and non-UDG (LP04) treated samples prepared and sequenced from the East Smithfield Strain sample 8291 alongside samples from individuals A120 and A76 in Set 1 (SOM section I B). Because overall coverage in these libraries was low, coverage of the DFR4 region from Microtus was compared against coverage of the entire Microtus genome and the *Y. pseudotuberculosis* IP31758 153 kbp plasmid (accession NC_009705.1) to represent sequence expected to be present and absent, respectively, from the sample (Supplementary Figure S17).

Evolutionary Analysis

A maximum likelihood (ML) phylogenetic tree was estimated for 133 strains of *Y. pestis* using a single strain of *Y. pseudotuberculosis* (IP32953) as an outgroup (Supplementary Figure S18). This analysis was undertaken using the PhyML program (19), utilizing the GTR+Γ_4 model of nucleotide substitution with estimated relative rate parameters of A <-> C 0.98804, A <-> G 2.12874, A <-> T 0.74475, C <-> G 0.08401, C <-> T 2.13133, and G <-> T 1.0, and a Γ shape parameter of among-site rate variation (4 rate categories) approaching infinity (as expected from a SNP data set). After a combination of NNI and SPR branch-swapping a ML tree with a log-likelihood of -17851.816 was obtained. To determine the robustness of individual nodes on the tree, a bootstrap PhyML analysis (1000 replicates) was performed using the GTR+ Γ_4 substitution model (parameters estimated during each replicate) and employing NNI branch-swapping. This ML tree (tree 1) was compared with a model phylogeny in which the two Justinian strains were constrained to be the sister-group to those sampled from the second and third pandemics (i.e. to the

immediate left of Black Death strain 8291 in Fig. 3 of the main text and Supplementary Figure S18). This allowed us to test the hypothesis that *Y. pestis* from the Justinian plague was the direct ancestor of all later human pandemic strains. These two trees were compared within a likelihood framework using the Shimodaira-Hasegawa (SH) test with RELL bootstrap (one-tailed test), as available in the PAUP* package(20). Accordingly, tree 1 was found to be significantly better than tree 2 (difference in –log likelihood = 100.812, $p < 0.001$), thereby rejecting the hypothesis that *Y. pesti*s from the Justinian plague gave rise to all later human pandemic strains.

To test the strength of the temporal signal in the data a second ML tree was inferred (same procedure as above) but with the *Y. pseudotuberculosis* outgroup removed (as no SNPs were called in this species such that its branch length is artificially short). A regression of root-to-tip genetic distances against year of sampling was then performed on this tree using the Path-O-Gen program kindly provided by Andrew Rambaut (University of Edinburgh; http://tree.bio.ed.ac.uk/software/pathogen/). This analysis revealed a very weak correlation coefficient (0.25) between time and the expected number of substitutions. Similarly, regressions of sampling date against root-to-tip genetic distances for (a) the modern *Y. pestis* strains only and (b) the Black Death strain and its descendants also showed weak associations (correlation coefficients of 0.27 and 0.20, respectively), again highlighting a lack of temporal structure in the data.

In an attempt to infer the rate and time-scale (time to the Most Recent Common Ancestor - tMRCA) of *Y. pestis* evolution we utilized the Bayesian Markov chain Monte Carlo (MCMC) method within the BEAST program (version 1.7.5) (21). This analysis also employed the GTR+Γ_4 substitution model, as well as a constant population size coalescent prior on the tree and a relaxed molecular clock with an uncorrelated lognormal distribution of rates. Again, *Y. pseudotuberculosis* was excluded from this analysis, as was the YP_Pestoides_F strain because of uncertainties over its sampling date. A variety of runs were undertaken incorporating different representations of the time structure in the data: (i) a standard 'tip-dated' run in which the sampling time (year of isolation) of each strain was incorporated in the analysis; (ii) a 'calibrated' run in which all modern strains were set to have the same (isochronous) sampling date (year 2000) and two 'fossil' dates were used to set the node times of two monophyletic groups of (a) the Black Death and modern strains at 1348 and (b) the Justinian, Black Death and modern samples at 533 (i.e. the mid-way point between the date range of 435-631 for the Justinian strains). Importantly, if the sampling time differences among the modern strains introduced temporal information we would expect to have different estimates for the node ages and substitution rates compared to the fully tip-dated tree in run (i); (iii) a '3 time points' run in which all the samples are considered isochronous (year 2000) with the exception of the Black Death and Justinian samples, which are calibrated using their true sampling times. Hence, the tip-date of the Black Death strain is calibrated using a normal distribution with mean date of 1348 and standard deviation 0.1, whereas the tip-date of the Justinian samples was calibrated using a normal distribution with a mean date of 533 and standard deviation 50. For each BEAST run, two independent MCMC simulations were performed of 500 million generations each to check for convergence of the chains.

The results of these analyses strongly suggest that there is no temporal structure in the modern strains, implying that clock estimation is largely based on the time structure of the ancient strains. In particular, comparison of the posterior distributions of the nucleotide substitution rates and tMRCAs of runs (i), (ii) and (iii) revealed no significant difference as shown by the overlapping 95% highest posterior densities (95% HPDs; the distribution of TMRCAs is shown in Fig. 3b main text). Consequently, we ran an additional randomization test in which the sampling times of the Justinian and Black Death strains were swapped (i.e. the Justinian strains were dated to 1348 and the Black Death strains dated to 533) yet retaining the true age of the modern strains. Strikingly, the log likelihoods and the other parameter estimates of this analysis were not statistically different to those of the original analysis (main Fig. 3b), confirming a lack of temporal structure in the data, including the ancient strains. We therefore conclude that estimates of the nucleotide substitution rate and tMRCA for *Y. pestis* are inherently unstable.

Supplementary Figures

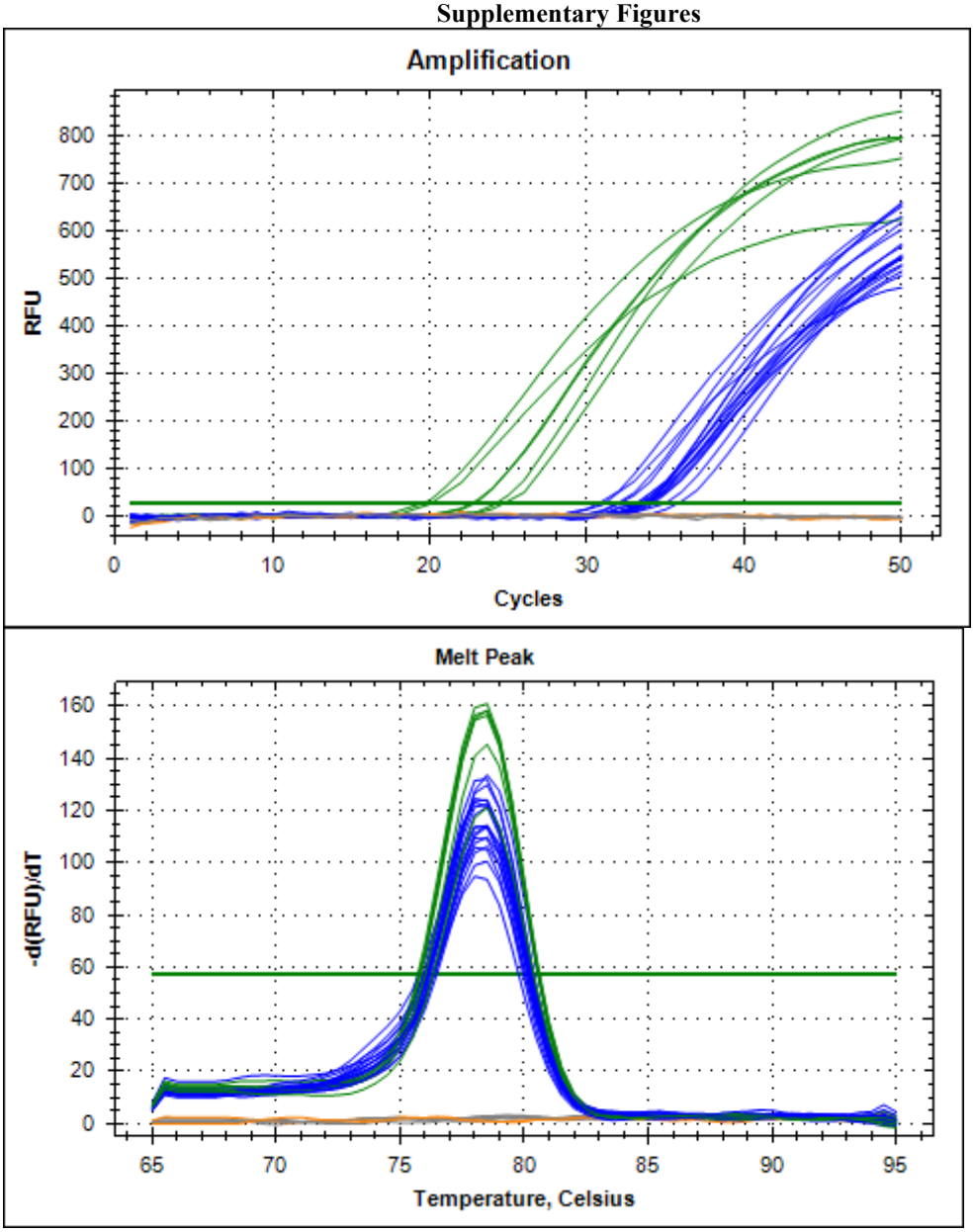

Figure S1: *pla* qPCR amplfication and melt curves: Two replicates of each standard of known copy number (50K, 5K, and 500 copy/µL) are shown in green. PCR blanks are shown in orange. Extraction blanks are shown in grey. Extract amplifications are shown in blue. The relative fluorescence unit (RFU) threshold to calculate Cq threshold was automatically determined by the software to be 28.77 (BioRad Manager CFX 3.0 software). Cq values for the positive templates shown above are as follows: LP29 (33.17, 33.86), LP30 (35.05, 35.84), LP31 (30.80, 31.17), LP32 (32.15, 32.30), LP35 (34.07, 34.26), LP36 (33.25, 33.74), LP37 (33.56, 33.71), LP38 (32.00, 32.14), and LP39 (33.42, 33.48). All other templates (blanks) showed no amplification after 50 cycles.

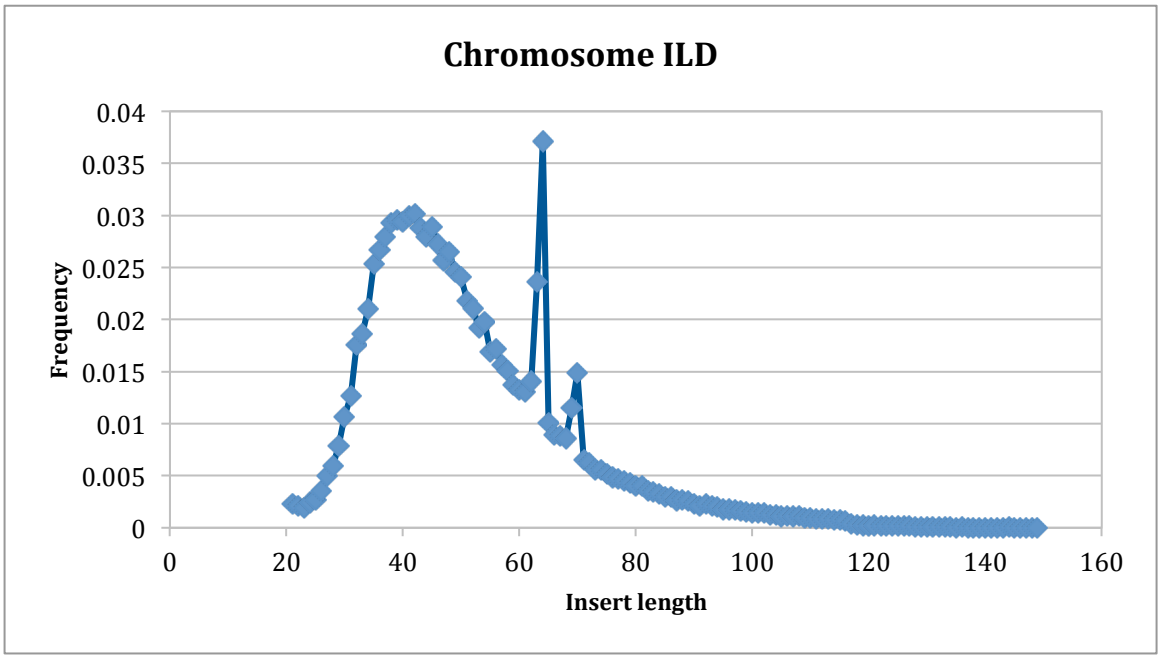

Figure S2: Unique insert length distribution (ILD) of alignments to the CO92 chromosome(A120+A76)

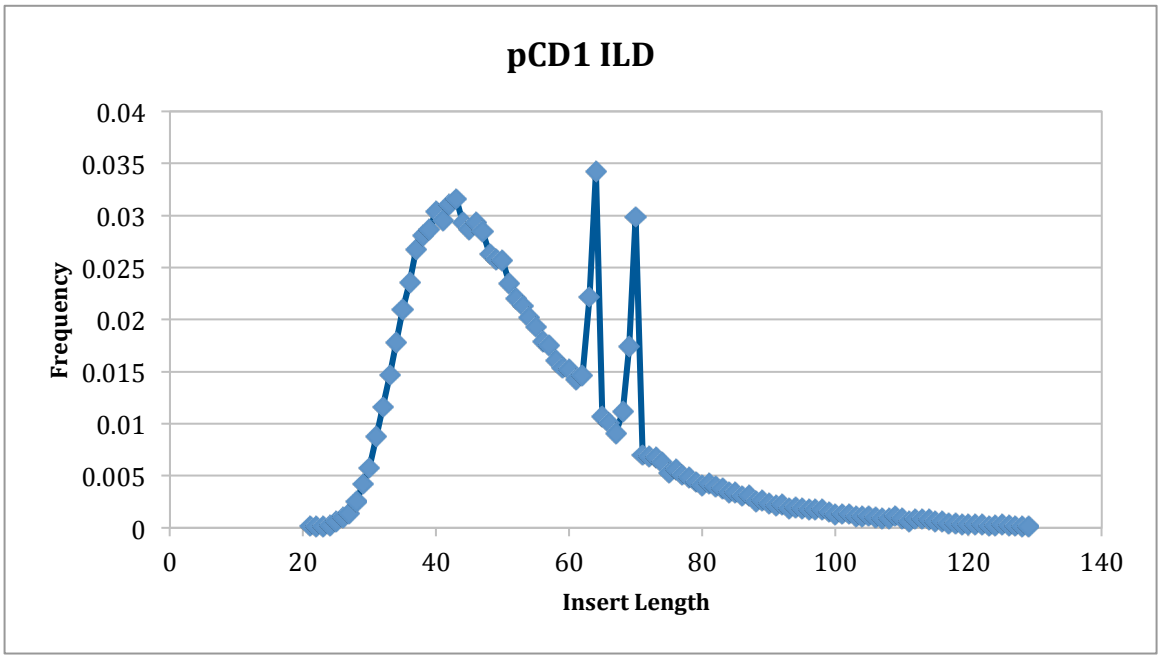

Figure S3: Unique insert length distribution of reads aligned to plasmid pCD1 (A120+A76 bed collapsed)

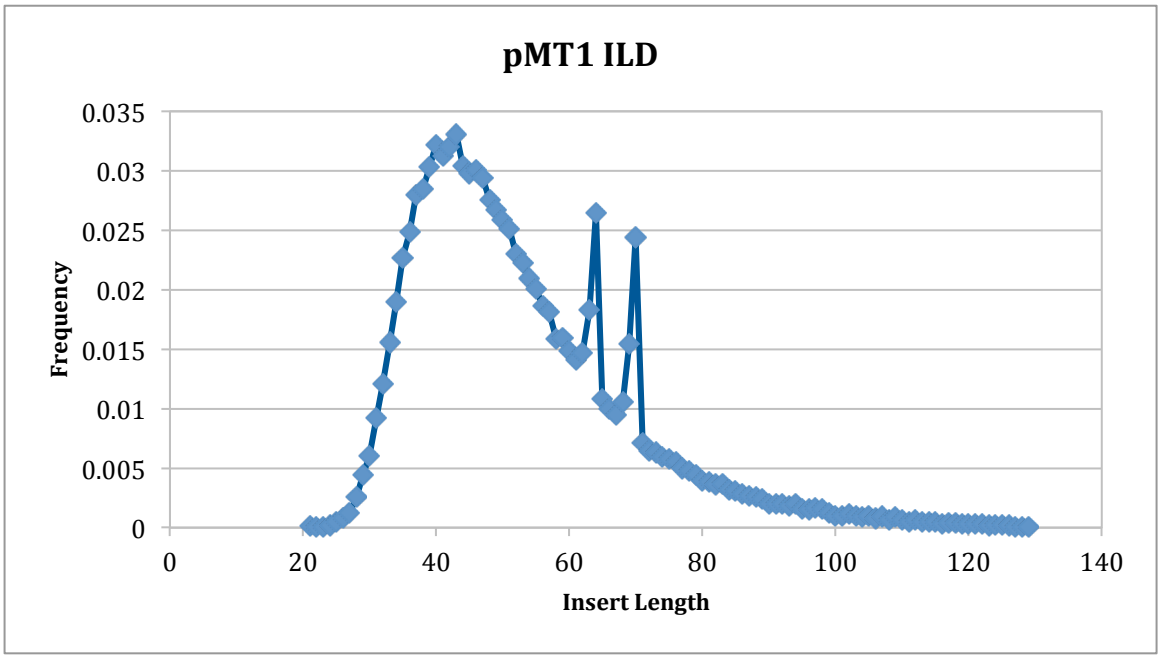

Figure S4: Unique insert length distribution of reads aligned to plasmid pMT1 assembly (A120+A76 bed collapsed)

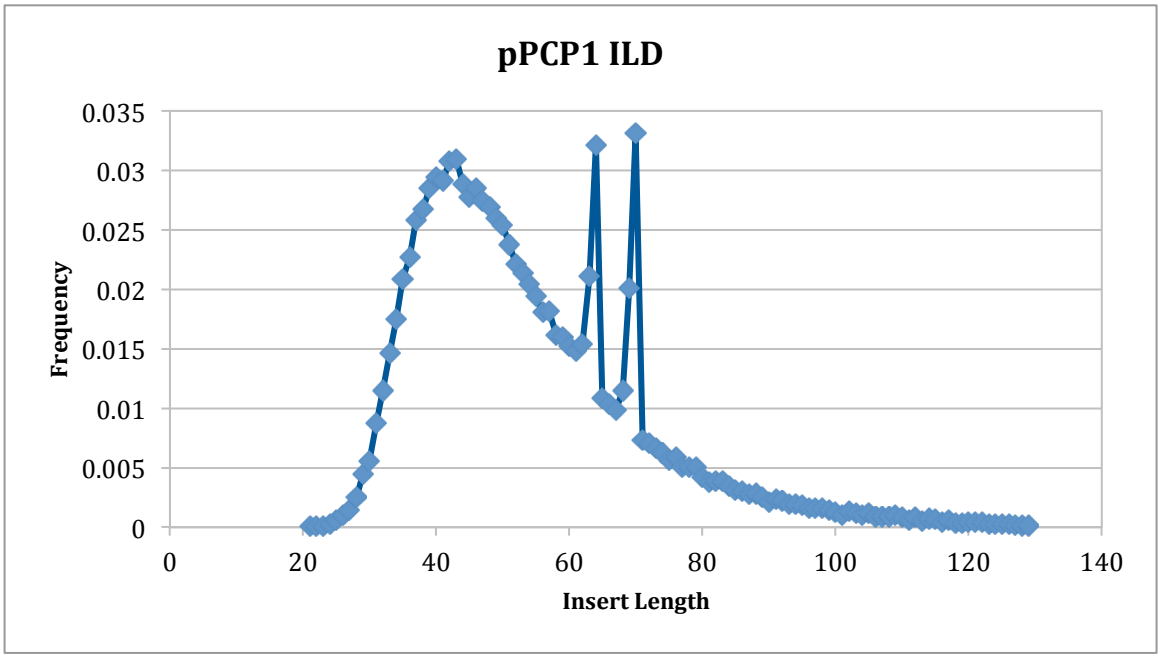

Figure S5: Unique insert length distribution of reads aligned to plasmid pPCP1 assembly (A120+A76 bed collapsed)

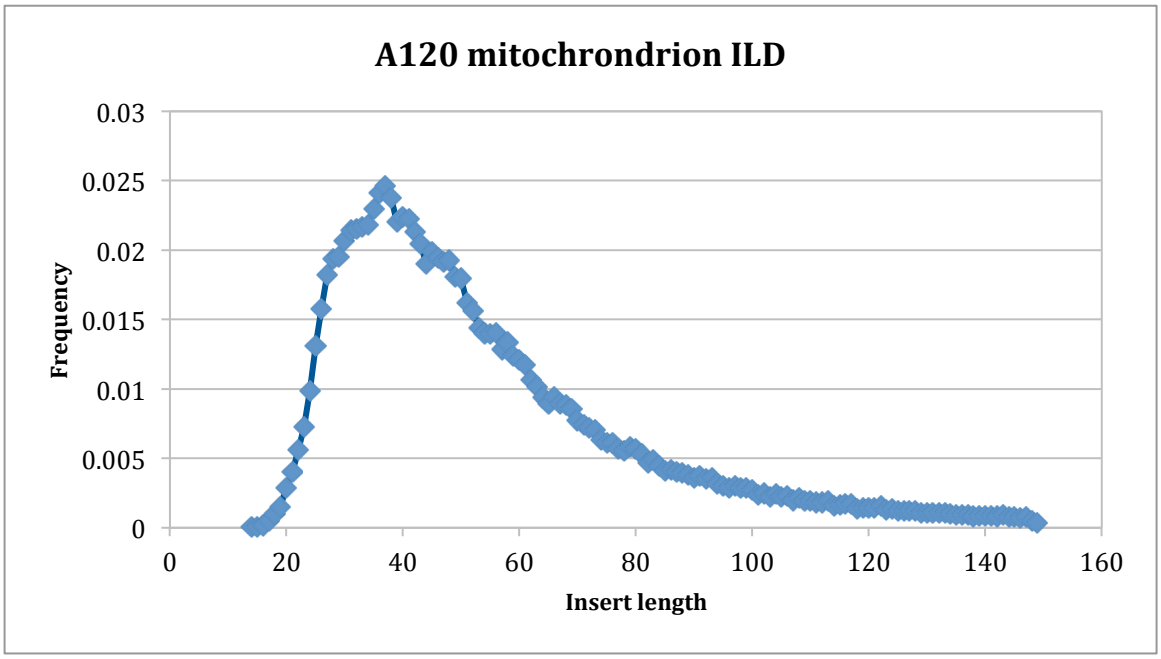

Figure S6: Unique insert length distribution of reads aligned to human mitochondrion assembly (A120)

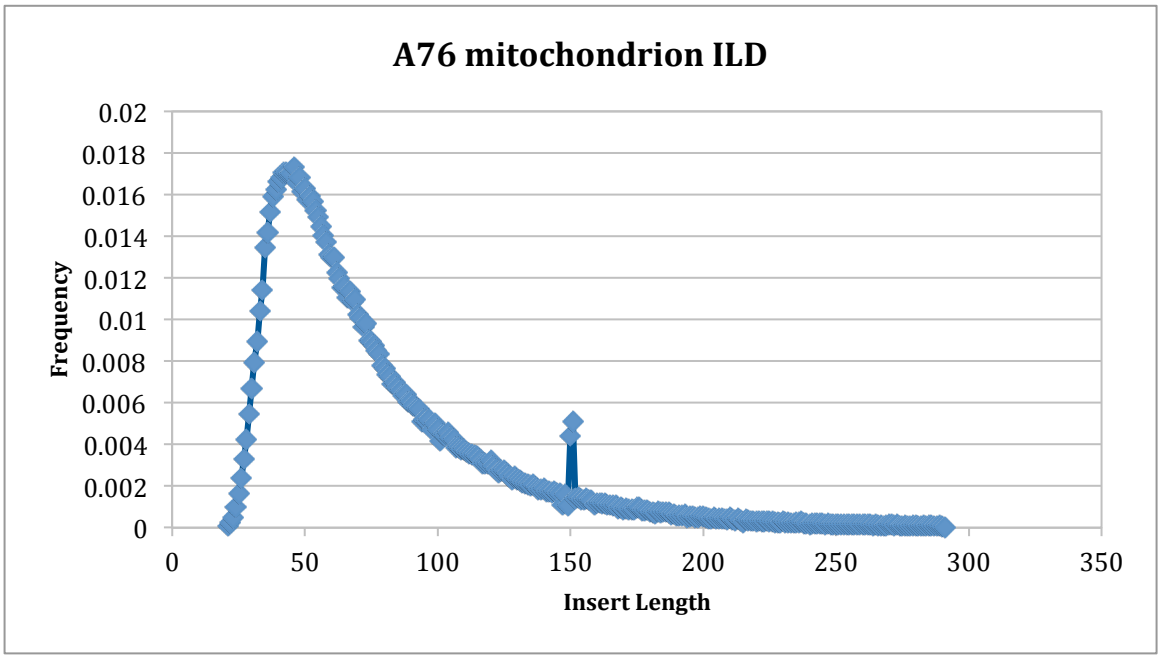

Figure S7: Unique insert length distribution of reads aligned to human mitochondrion assembly (A76).

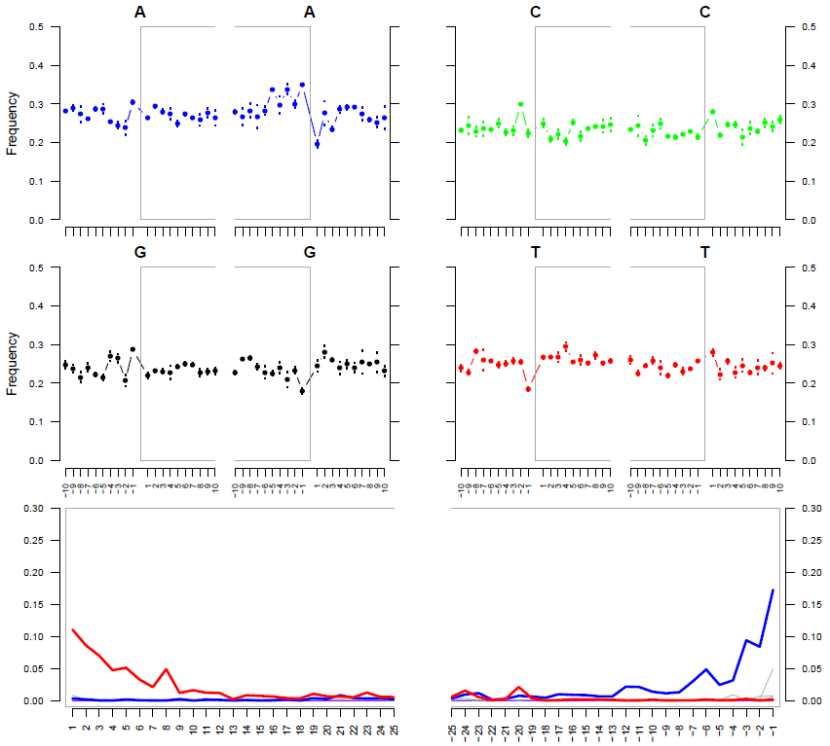

Figure S8: mapDamage fragment misincorporation plot for sample LP05 (non-UDG treated sample).

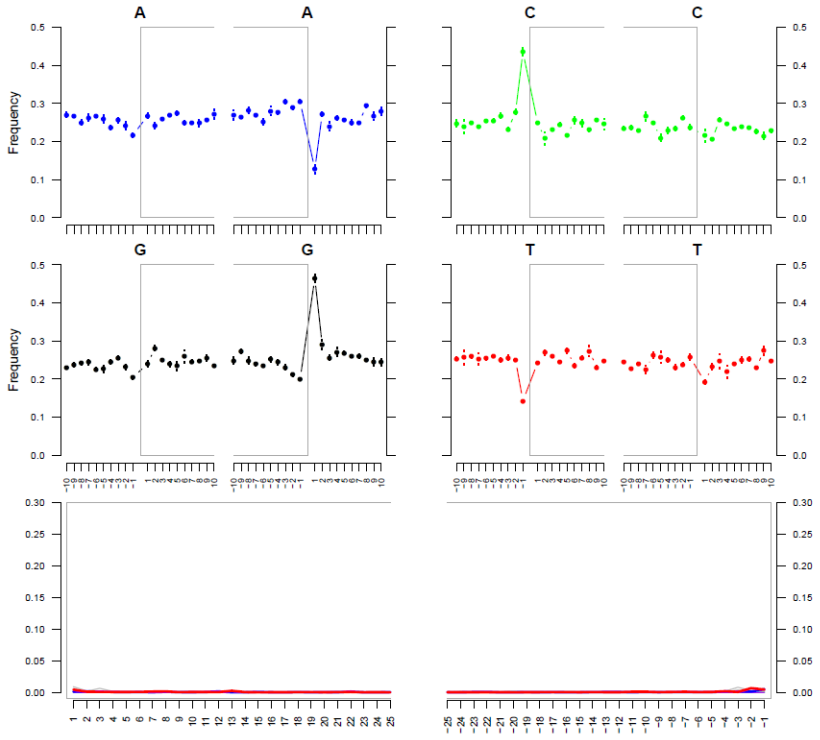

Figure S9: mapDamage fragment misincorporation plot for sample LP05U (UDG-treated sample)

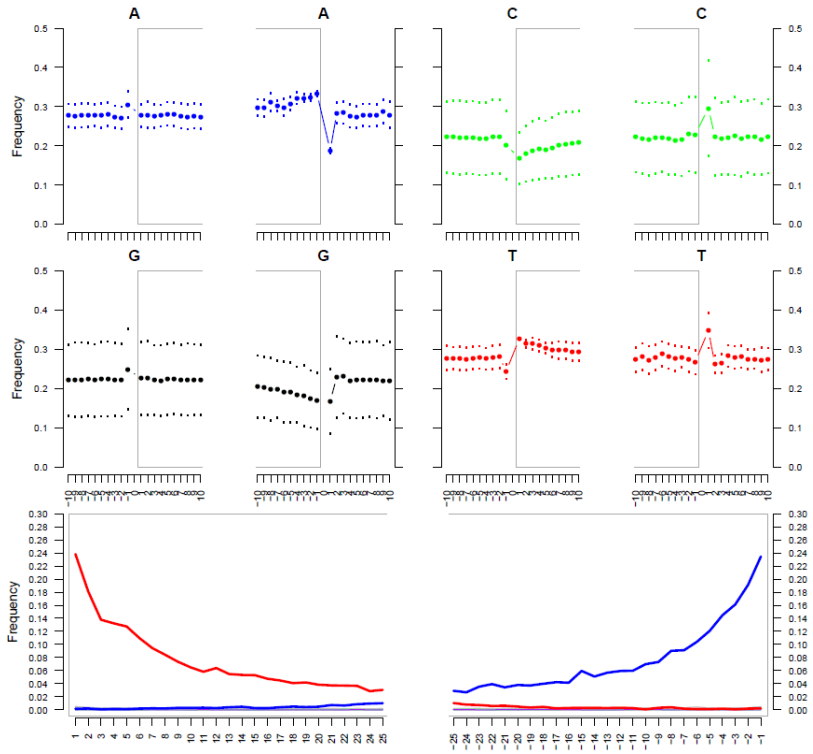

Figure S10: mapDamage fragment misincorporation plot for human mitochondrion assembly of A120

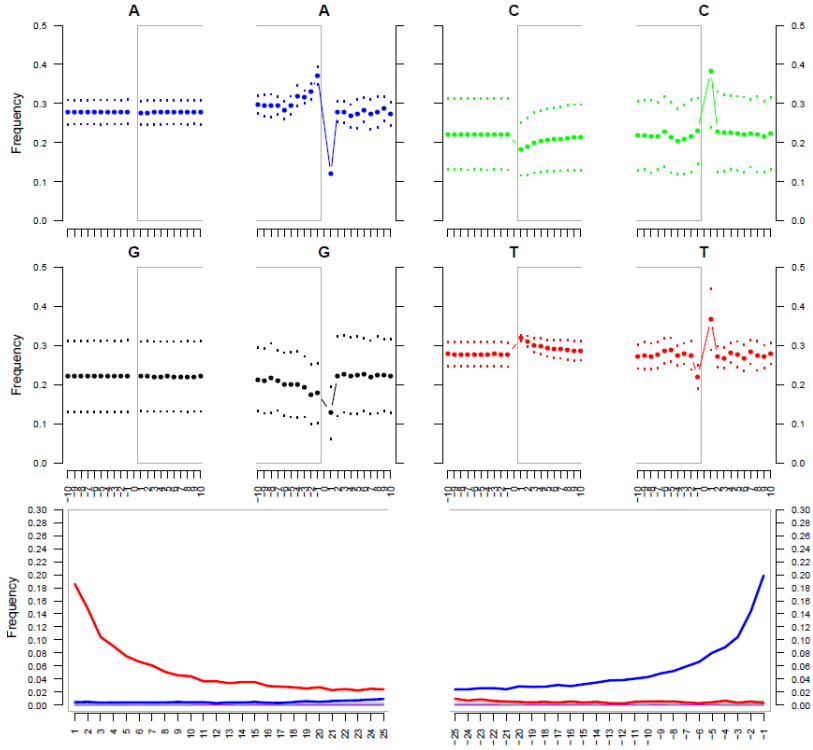

Figure S11: mapDamage fragment misincorporation plot for human mitochondrion assembly of A76

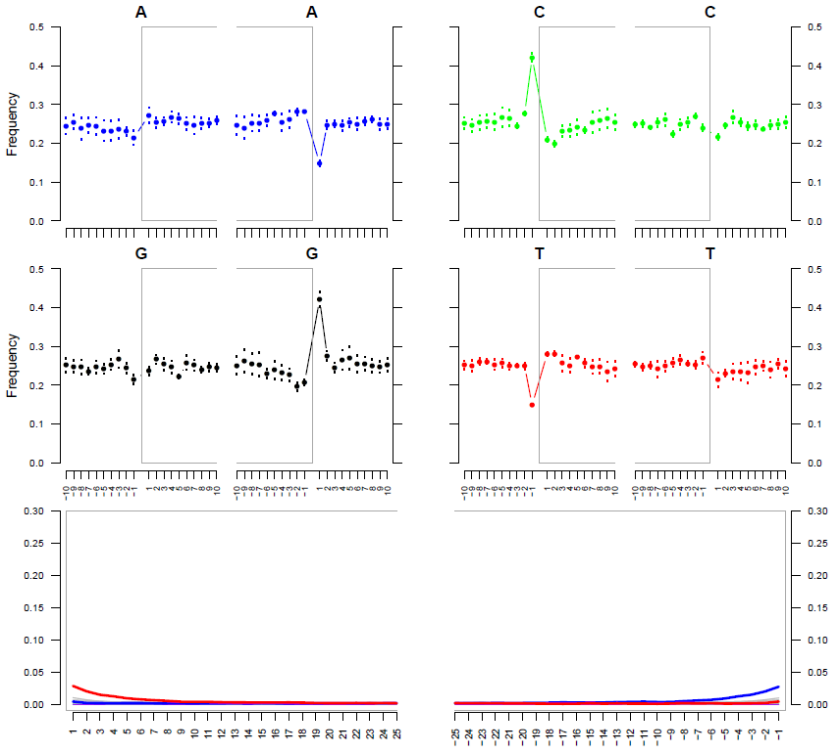

Figure S12: mapDamage fragment misincorporation of chromosomal assembly (A120+A76)

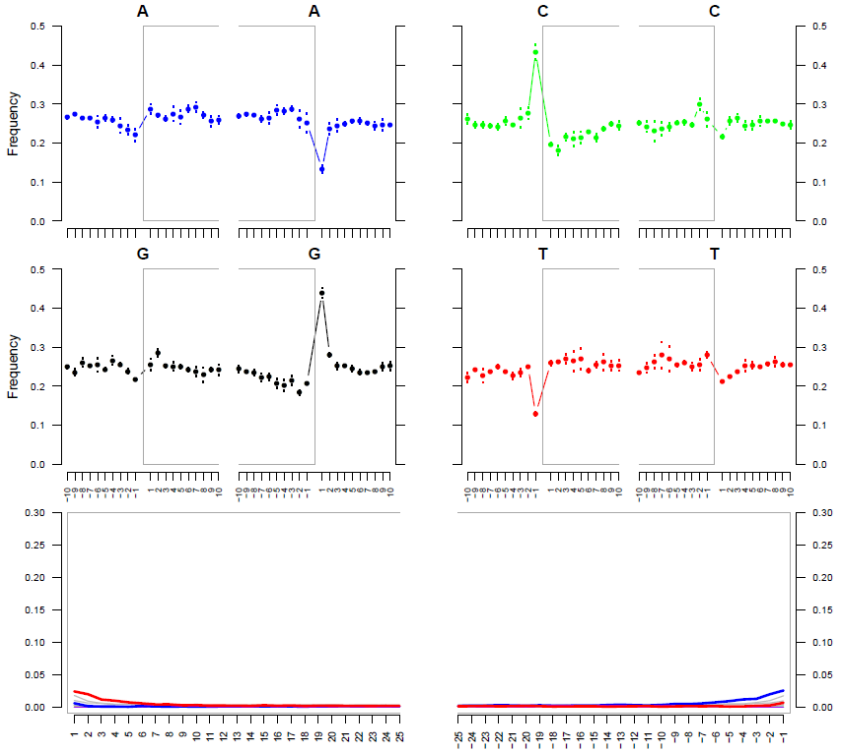

Figure S13: mapDamage fragment misincorporation plot of plasmid pCD1 assembly (A120+A76)

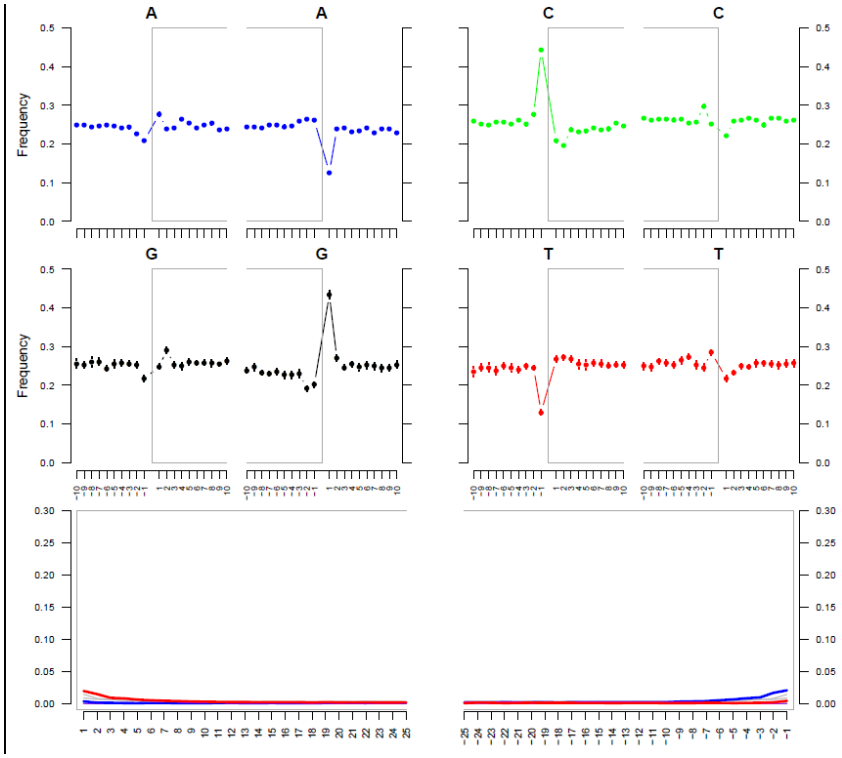

Figure S14: mapDamage fragment misincorporation plot of plasmid pMT1 assembly (A120+A76)

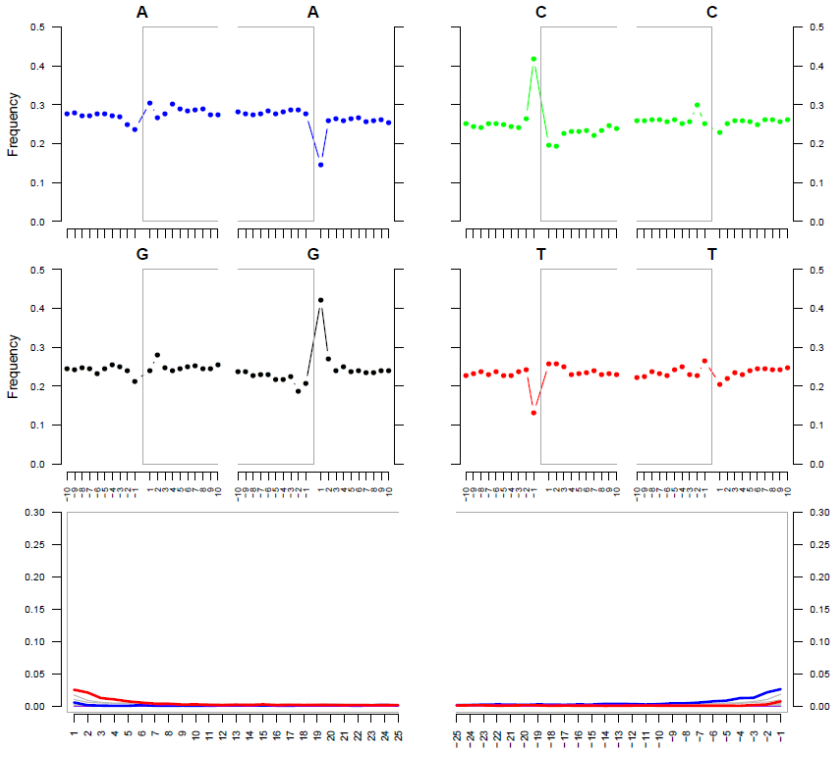

Figure S15: mapDamage fragment misincorporation plot of plasmid pPCP1 assembly (A120+A76)

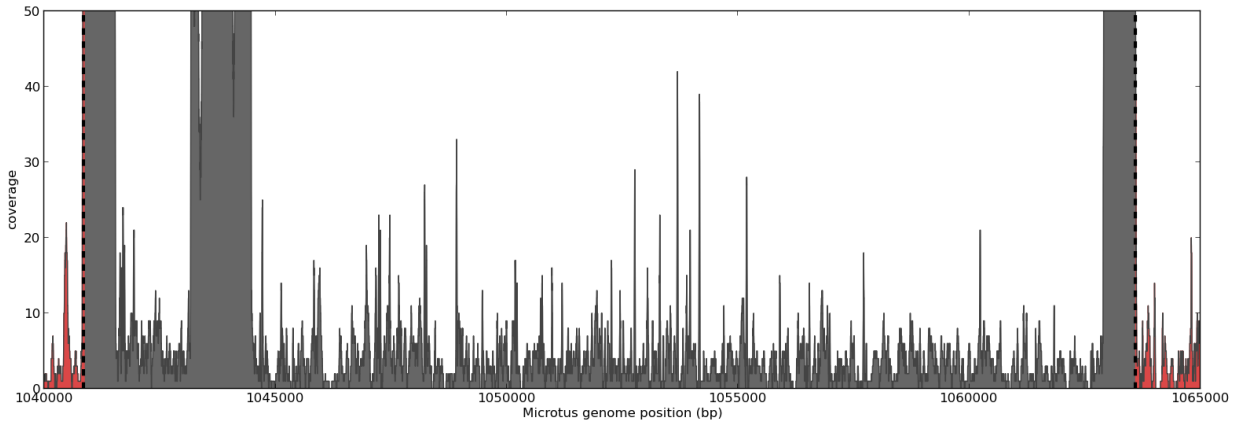

Figure S16: Unique read coverage of Justinian sample mapped across DFR4 in the Microtus genome, not present in CO92. Dashed lines indicate end and start position of deletion in CO92 chromosome. Red plots represent shared coverage in Microtus and CO92 bordering DFR4. Grey area indicates Justinian coverage of DFR4 region in Microtus alone.

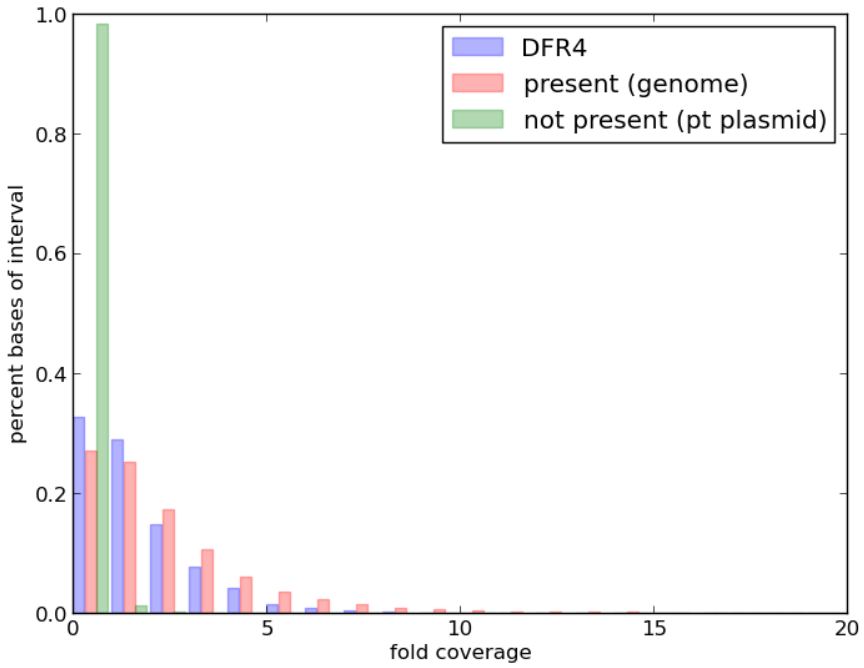

Figure S17. DFR4 region from ESS individual 8291. DFR4 is present in ESS individual 8291 (blue) at approximately the same coverage level as sequence from the Microtus genome (red). To contrast, a sequence not expected to be present (plasmid sequence from 153kbp plasmid from *Y. pseudotuberculosis* IP31758) is underrepresented (green).

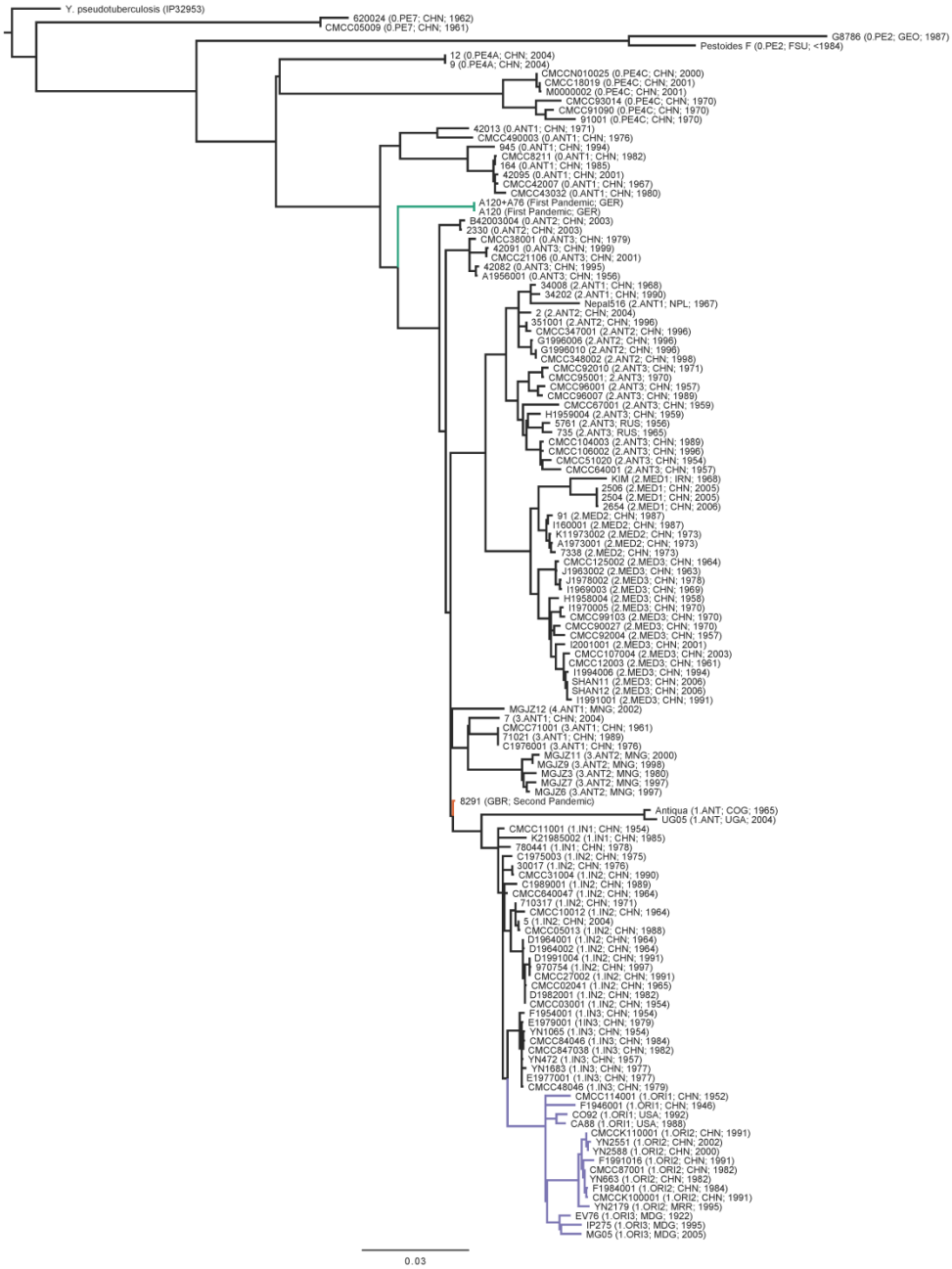

Figure S18. (**A**) Expanded ML tree of 2,268 SNPs from 133 strains of *Y. pestis*. The tree is rooted using strain IP32953 (18) from the soil-dwelling *Y. pseudotuberculosis,* the likely ancestor of *Y. pestis* (7, 22). The branches associated with first, second, and third pandemics are colored green, orange, and purple, respectively. Phylogenetic groups (e.g. 0.PE7) are designated according to (7), with the first number indicating the major branch (0-4) along which that group is found. Three digit codes indicate the country where each strain was isolated: CHN, China; COG, Congo; FSU, Former Soviet Union; GBR, Great Britain; GEO, Georgia; GER, Germany; IRN, Iran; MDG, Madagascar; MNG, Mongolia; MRR, Myanmar; NPL, Nepal; RUS, Russia; UGA, Uganda; and USA, United States of America. The year of isolation is also indicated for each strain.

Supplementary Tables

Table S1: Accessions for *Yersinia* strains used in enrichment Pan-Array design

	Organism	Length (bp)	Number of sequences
NC_017168.1	*Yersinia pestis* A1122	4553770	1
ACNR00000000.1	*Yersinia pestis* biovar Orientalis str. India 195	4719608	33
ACNT00000000.1	*Yersinia pestoides* A	4750606	37
ADDC00000000.1	*Yersinia pestis* KIM D27	4501057	1
ADOV00000000.1	*Yersinia pestis* 12	4368767	1092
ADOW00000000.1	*Yersinia pestis* 164	4444813	1053
ADOX00000000.1	*Yersinia pestis* 2	4453910	1057
ADOY00000000.1	*Yersinia pestis* 2330	4427639	798
ADOZ00000000.1	*Yersinia pestis* 2504	4475655	975
ADPA00000000.1	*Yersinia pestis* 2506	4488493	820
ADPB00000000.1	*Yersinia pestis* 2654	4352602	2718
ADPC00000000.1	*Yersinia pestis* 30017	4591472	747
ADPD00000000.1	*Yersinia pestis* 34008	4375742	1074
ADPE00000000.1	*Yersinia pestis* 34202	4504947	1237
ADPF00000000.1	*Yersinia pestis* 351001	4433848	1332
ADPG00000000.1	*Yersinia pestis* 42013	4495101	877
ADPH00000000.1	*Yersinia pestis* 42082	4496670	2049
ADPI00000000.1	*Yersinia pestis* 42091	4495831	937
ADPJ00000000.1	*Yersinia pestis* 42095	4512584	1119
ADPK00000000.1	*Yersinia pestis* 5	4478071	1008
ADPL00000000.1	*Yersinia pestis* 5761	4262011	945
ADPM00000000.1	*Yersinia pestis* 620024	4503786	1131
ADPN00000000.1	*Yersinia pestis* 7	4539616	768
ADPO00000000.1	*Yersinia pestis* 71021	4493411	1489
ADPP00000000.1	*Yersinia pestis* 710317	4449879	1119
ADPQ00000000.1	*Yersinia pestis* 7338	4344510	1074
ADPR00000000.1	*Yersinia pestis* 735	4423715	882
ADPS00000000.1	*Yersinia pestis* 780441	4525764	817
ADPT00000000.1	*Yersinia pestis* 9	4452201	1336
ADPU00000000.1	*Yersinia pestis* 91	4366271	906
ADPV00000000.1	*Yersinia pestis* 945	4226749	1243
ADPW00000000.1	*Yersinia pestis* 970754	4348631	1104
ADPX00000000.1	*Yersinia pestis* A1956001	4508534	852
ADPY00000000.1	*Yersinia pestis* A1973001	4413119	1539
ADPZ00000000.1	*Yersinia pestis* C1975003	4365481	1452
ADQA00000000.1	*Yersinia pestis* C1976001	4546693	979
ADQB00000000.1	*Yersinia pestis* C1989001	4588735	934

ADQC00000000.1	*Yersinia pestis* CMCC02041	4489186	1281
ADQD00000000.1	*Yersinia pestis* CMCC03001	4366616	1082
ADQE00000000.1	*Yersinia pestis* CMCC05009	4501730	1501
ADQF00000000.1	*Yersinia pestis* CMCC05013	4395097	1360
ADQG00000000.1	*Yersinia pestis* CMCC10012	4435712	1001
ADQH00000000.1	*Yersinia pestis* CMCC104003	4494768	1387
ADQI00000000.1	*Yersinia pestis* CMCC106002	4424589	1176
ADQJ00000000.1	*Yersinia pestis* CMCC107004	4455583	2814
ADQK00000000.1	*Yersinia pestis* CMCC11001	4423348	1100
ADQL00000000.1	*Yersinia pestis* CMCC114001	4231817	1664
ADQM00000000.1	*Yersinia pestis* CMCC12003	4341435	1045
ADQN00000000.1	*Yersinia pestis* CMCC125002	4518010	790
ADQO00000000.1	*Yersinia pestis* CMCC18019	4495992	1478
ADQP00000000.1	*Yersinia pestis* CMCC21106	4251990	1387
ADQQ00000000.1	*Yersinia pestis* CMCC27002	4282597	1451
ADQR00000000.1	*Yersinia pestis* CMCC31004	4413744	1229
ADQS00000000.1	*Yersinia pestis* CMCC347001	4514951	1207
ADQT00000000.1	*Yersinia pestis* CMCC348002	4571405	846
ADQU00000000.1	*Yersinia pestis* CMCC38001	4483258	1030
ADQV00000000.1	*Yersinia pestis* CMCC42007	4440824	1040
ADQW00000000.1	*Yersinia pestis* CMCC43032	4512744	1049
ADQX00000000.1	*Yersinia pestis* CMCC49003	4534555	795
ADQY00000000.1	*Yersinia pestis* CMCC51020	4420992	1118
ADQZ00000000.1	*Yersinia pestis* CMCC64001	4360492	955
ADRA00000000.1	*Yersinia pestis* CMCC640047	4305752	1490
ADRB00000000.1	*Yersinia pestis* CMCC67001	4340671	1099
ADRC00000000.1	*Yersinia pestis* CMCC71001	4254189	1744
ADRD00000000.1	*Yersinia pestis* CMCC8211	4417626	797
ADRE00000000.1	*Yersinia pestis* CMCC84033	4459052	1055
ADRF00000000.1	*Yersinia pestis* CMCC84038	4438544	2446
ADRG00000000.1	*Yersinia pestis* CMCC84046	4517175	1137
ADRH00000000.1	*Yersinia pestis* CMCC87001	4510364	870
ADRI00000000.1	*Yersinia pestis* CMCC90027	4358986	1102
ADRJ00000000.1	*Yersinia pestis* CMCC91090	4418766	987
ADRK00000000.1	*Yersinia pestis* CMCC92004	4264825	1461
ADRL00000000.1	*Yersinia pestis* CMCC92010	4384955	1216
ADRM00000000.1	*Yersinia pestis* CMCC93014	4507171	993
ADRN00000000.1	*Yersinia pestis* CMCC95001	4337306	1361
ADRO00000000.1	*Yersinia pestis* CMCC96001	4517589	907
ADRP00000000.1	*Yersinia pestis* CMCC96007	4303407	1563
ADRQ00000000.1	*Yersinia pestis* CMCC99103	4409843	1262

ADRR00000000.1	*Yersinia pestis* CMCCK100001	4520971	1406
ADRS00000000.1	*Yersinia pestis* CMCCK110001	4305250	1385
ADRT00000000.1	*Yersinia pestis* CMCCN010025	4437913	1016
ADRU00000000.1	*Yersinia pestis* D1964001	4389464	1012
ADRV00000000.1	*Yersinia pestis* D1964002	4431847	1555
ADRW00000000.1	*Yersinia pestis* D1982001	4410882	896
ADRX00000000.1	*Yersinia pestis* D1991004	4506081	1270
ADRY00000000.1	*Yersinia pestis* E19777001	4385578	1183
ADRZ00000000.1	*Yersinia pestis* biovar Antiqua str. E1979001	4516254	941
ADSA00000000.1	*Yersinia pestis* EV76	4274282	1172
ADSB00000000.1	*Yersinia pestis* F1946001	4463713	1383
ADSC00000000.1	*Yersinia pestis* F1954001	4521000	1058
ADSD00000000.1	*Yersinia pestis* F1984001	4326553	1467
ADSE00000000.1	*Yersinia pestis* G1996006	4542684	842
ADSF00000000.1	*Yersinia pestis* G1996010	4552418	772
ADSG00000000.1	*Yersinia pestis* G8786	4325798	829
ADSH00000000.1	*Yersinia pestis* H1958004	4195425	1339
ADSI00000000.1	*Yersinia pestis* H1959004	4470270	946
ADSJ00000000.1	*Yersinia pestis* I160001	4399336	1222
ADSK00000000.1	*Yersinia pestis* I1969003	4512122	1213
ADSL00000000.1	*Yersinia pestis* I1970005	4429836	1199
ADSM00000000.1	*Yersinia pestis* I1991001	4335804	1191
ADSN00000000.1	*Yersinia pestis* I1994006	4498336	1247
ADSO00000000.1	*Yersinia pestis* I2001001	4509469	1526
ADSP00000000.1	*Yersinia pestis* J1963002	4434765	1119
ADSQ00000000.1	*Yersinia pestis* J1978002	4532297	1098
ADSR00000000.1	*Yersinia pestis* K11973002	4383307	910
ADSS00000000.1	*Yersinia pestis* K21985002	4399142	1264
ADST00000000.1	*Yersinia pestis* M000002	4441674	1135
ADSU00000000.1	*Yersinia pestis* MGJZ11	4516720	859
ADSV00000000.1	*Yersinia pestis* MGJZ12	4438965	1203
ADSW00000000.1	*Yersinia pestis* MGJZ3	4483871	1187
ADSX00000000.1	*Yersinia pestis* MGJZ6	4368072	1016
ADSY00000000.1	*Yersinia pestis* MGJZ7	4365909	1325
ADSZ00000000.1	*Yersinia pestis* MGJZ9	4474930	1082
ADTA00000000.1	*Yersinia pestis* SHAN11	4384465	1173
ADTB00000000.1	*Yersinia pestis* SHAN12	4371391	1362
ADTC00000000.1	*Yersinia pestis* YN1065	4472301	1395
ADTD00000000.1	*Yersinia pestis* YN1683	4302115	1897
ADTE00000000.1	*Yersinia pestis* YN2179	4307368	1605
ADTF00000000.1	*Yersinia pestis* YN2551	4524315	880

ADTG00000000.1	*Yersinia pestis* YN2588	4307940	1301
ADTH00000000.1	*Yersinia pestis* YN472	4525667	1105
ADTI00000000.1	*Yersinia pestis* YN663	4203251	1285
AGJS00000000.1	*Yersinia pestis* biovar Orientalis str. BA200901799	4581111	199
AGJT00000000.1	*Yersinia pestis* biovar Orientalis str. AS200901539	4572127	250
AGJU00000000.1	*Yersinia pestis* biovar Orientalis str. AS200901156	4578309	220
AGJV00000000.1	*Yersinia pestis* biovar Orientalis str. AS200901703	4560176	282
AGJW00000000.1	*Yersinia pestis* biovar Orientalis str. BA200901990	4527493	210
AGJX00000000.1	*Yersinia pestis* biovar Orientalis str. BA200902009	4563812	256
AGJY00000000.1	*Yersinia pestis* biovar Orientalis str. AS200902147	4592682	277
AGJZ00000000.1	*Yersinia pestis* biovar Orientalis str. AS200901434	4572981	237
AGKA00000000.1	*Yersinia pestis* biovar Orientalis str. AS200901509	4605070	263
AKVQ00000000.1	*Yersinia pestis* 2501	4593919	201
AMQL00000000.1	*Yersinia pestis* INS	4651700	18
NC_010159.1	*Yersinia pestis* Angola	4504254	1
NC_008150.1	*Yersinia pestis* Antiqua	4702289	1
AAYU0000000.1	*Yersinia pestis* biovar Antiqua str. B42003004	4841690	69
ABCD00000000.1	*Yersinia pestis* CA88-4125	4650262	7
NC_003143.1	*Yersinia pestis* CO92	4653728	1
NC_017154.1	*Yersinia pestis* D106004	4640720	1
NC_017160.1	*Yersinia pestis* D182038	4626744	1
AAYV00000000.1	*Yersinia pestis* biovar Antiqua str. E1979001	4847813	75
ABAT00000000.1	*Yersinia pestis* biovar Orientalis str. F19991016	4892085	107
AAUB00000000.1	*Yersinia pestis* FV-1	4472646	400
NC_009708.1	*Yersinia pseudotuberculosis* IP 31758	4723306	1
NC_006155.1	*Yersinia pseudotuberculosis* IP 32953	4744671	1
AAYT00000000.1	*Yersinia pestis* biovar Medievalis str. K1973002	4720694	73
NC_004088.1	*Yersinia pestis* KIM 10	4600755	1
NC_005810.1	*Yersinia pestis* biovar Microtus str. 91001	4595065	1
AAYS00000000.1	*Yersinia pestis* biovar Orientalis str. MG05-1020	4977609	80
NC_017265.1	*Yersinia pestis* biovar Medievalis str. Harbin 35	4532063	1
NC_008149.1	*Yersinia pestis* Nepal516	4534590	1
AAOS00000000.1	*Yersinia pestis* biovar Orientalis str. IP275	5139152	101
ACNS00000000.1	*Yersinia pestis* biovar Orientalis str. PEXU2	4666679	3
NC_009381.1	*Yersinia pestis* Pestoides F	4517345	1
AAYR00000000.1	*Yersinia pestis* Antiqua str. UG05-0454	4838246	98
NC_014029.1	*Yersinia pestis* Z176003	4553586	1

Table S2: Coverage metrics for the chromosome (reference) and plasmids pCD1, pMT1 and pPCP1with repeats masked and rRNAs removed.

	Coverage of reference	Mean coverage	Standard Deviation
Chromosome	91.5%	7.6	48.7
pCD1	99.9%	71.7	293.6
pMT1	96.7%	129.2	548.7
pPCP1	100.0%	476.9	682.1

Table S3: 5X, 90% variant frequency human mitochondrial SNPs in individual A120

Position	Reference	A120	Coverage
114	C	T	148
146	T	C	112
152	T	C	102
195	T	C	95
263	A	G	94
708	G	A	139
750	A	G	138
1,438	A	G	126
4,769	A	G	111
8,697	G	A	119
8,860	A	G	135
9,377	A	G	147
13,101	A	C	138
13,711	G	A	156
15,326	A	G	151
16,288	T	C	111
16,294	C	T	106
16,362	T	C	122

Table S4: 5X, 90% variant frequency human mitochondrial SNPs in individual A76

Position	Reference	A76	Coverage
73	A	G	360
185	G	A	278
228	G	A	237
263	A	G	221
294	C	T	209
462	C	T	264
488	T	C	273
749	A	G	303
1,438	A	G	289

2,706	A	G	314
3,010	G	A	316
4,216	T	C	296
4,769	A	G	297
7,028	C	T	320
8,860	A	G	300
10,398	A	G	302
11,251	A	G	307
11,719	G	A	311
12,612	A	G	306
13,708	G	A	300
13,934	C	T	291
14,180	T	C	301
14,766	C	T	282
14,798	T	C	285
15,326	A	G	300
15,452	C	A	326
16,069	C	T	288
16,126	T	C	277

Table S5: Coverage metrics for the human mitochondrion

	Coverage	Mean coverage	Standard Deviation	Average fragment length
A120	100%	218.9	43.5	68.3
A76	100%	1600.2	311.1	75.7

Table S6: All 66 unique SNPS found in A120 alone cross-referenced against 133 genomes from Table S8

Position in CO92	SNP type	Ancestral AA	Derived AA	Gene
221811	non-synonymous	V	I	rpsJ
333342	synonymous	R	R	uvrA
497800	synonymous	T	T	dnaK
567757	non-synonymous	I	L	
727741	non-synonymous	K	E	parE
877258	non-synonymous	L	F	lysR
898980	intergenic	N/A	N/A	
1067966	non-synonymous	C	G	
1237756	non-synonymous	H	R	
1371025	synonymous	R	R	gyrA
1444672	intergenic	N/A	N/A	
1489055	synonymous	Q	Q	deoR
1530658	non-synonymous	I	R	
1754708	non-synonymous	L	P	galU
1864793	synonymous	G	G	mnmA
1868678	intergenic	N/A	N/A	
1956162	intergenic	N/A	N/A	
2072914	synonymous	V	V	fliI
2097520	synonymous	A	A	putA
2725715	intergenic	N/A	N/A	
3143800	non-synonymous	D	A	ynbD

3155055	synonymous	T	T	purI
3392897	synonymous	N	N	napA
3398153	synonymous	L	L	
3409414	intergenic	N/A	N/A	
3500922	non-synonymous	G	V	tesB
3560088	non-synonymous	S	P	
3568597	non-synonymous	E	G	phoB
3750736	intergenic	N/A	N/A	
3755861	intergenic	N/A	N/A	
3813424	synonymous	R	R	acnB
4170791	synonymous	F	F	aceA
4199187	synonymous	G	G	rpoC
4199190	synonymous	E	E	rpoC
4203596	synonymous	G	G	rpoB
4232217	intergenic	N/A	N/A	
4307755	non-synonymous	V	A	
4399295	synonymous	A	A	
4460688	non-synonymous	Q	R	
4465967	synonymous	S	S	
4628496	synonymous	A	A	yieG
4629169	intergenic	N/A	N/A	
ACCESSORY				
225902				
698477	intergenic	N/A	N/A	
1138676	intergenic	N/A	N/A	
1895361	non-synonymous	Q	P	motB
2117516	intergenic	N/A	N/A	
2304950	intergenic	N/A	N/A	
2317730	non-synonymous	F	L	
3248223	non-synonymous	G	R	ail
3358603	intergenic	N/A	N/A	
3403167	non-synonymous	K	T	
3472427	non-synonymous	V	A	ddhD
3737968	non-synonymous	V	A	
3909258	intergenic	N/A	N/A	
4465967	synonymous	S	S	
PLASMIDS				
gi|5834685|emb|AL11721				
1.1|::32569	synonymous	T	T	
gi|5834685|emb|AL11721				
1.1|::47365	synonymous	R	R	
gi|5834685|emb|AL11721				
1.1|::62994	intergenic	N/A	N/A	
gi|5834685|emb|AL11721				
1.1|::82435	intergenic	N/A	N/A	
gi|5832423|emb|AL11718				
9.1|::23564	synonymous	A	A	
gi|5832423|emb|AL11718				
9.1|::29959	non-synonymous	S	N	yscO
gi|5832423|emb|AL11718				
9.1|::50462	non-synonymous	E	K	yopJ
gi|5832423|emb|AL11718				
9.1|::54237	intergenic	N/A	N/A	
gi|5832423|emb|AL11718				
9.1|::55839	intergenic	N/A	N/A	
gi|5832423|emb|AL11718				
9.1|::66608	non-synonymous	F	L	

Table S7 - All *Yersinia* sp. genomes utilized for SNP discovery (see separate Excel spreadsheet labeled Appendix II).

Table S8 – 5X, 90% variant frequency chromosomal SNPs from Cui et al. called in A120 and A120 + A76. (see Excel spreadsheet labeled Appendix III appended as a separate document).

Table S9 -176 chromosomal SNPs in A120 and A120 +A76 compared to C092. Orange highlights indicate SNP called only in A120. Blue highlights indicate SNP called only in A76.

Position	CO92	Justinian	Coverage (A120+A76)
74539	C	T	5
105187	A	C	19
107738	A	G	13
130643	G	A	35
221811	A	G	63
225902	A	T	6
228268	T	G	11
286528	T	A	23
325836	T	C	16
333342	A	G	46
341720	A	G	19
399533	C	A	10
417323	G	A	30
442439	T	C	46
497800	T	A	122
545488	T	C	41
547131	T	G	15
567757	C	A	6
698477	C	A	21
699494	A	G	17
699647	T	C	10
727741	G	A	13
754287	T	C	18
773110	T	C	24
809132	A	G	13
877258	T	C	20
898980	A	T	13
918790	C	T	16
1017647	T	C	12
1025278	T	G	16
1044488	A	G	23
1051913	A	G	21
1067966	C	A	23
1102174	A	G	17
1137603	G	T	6
1138676	G	T	6
1178178	T	C	17
1178459	T	C	13
1237756	C	T	8
1251046	T	C	5
1263337	G	A	17
1272559	T	C	18
1306718	T	C	28
1371020	C	A	8
1371025	C	T	24
1385780	T	C	23
1440494	A	C	12
1444672	A	G	9
1489055	C	T	22
1512930	A	G	36
1530658	C	A	13
1540754	A	G	16
1705810	A	C	15
1735263	A	C	19
1749443	T	C	14
1754708	C	T	12
1796044	T	C	16
1804559	C	T	23
1808946	T	C	20

1864793	A	G	162
1868678	G	T	12
1871476	G	A	23
1895361	C	A	14
1914093	T	C	14
1956162	T	C	15
1982740	A	C	9
2012524	T	G	7
2022335	A	C	18
2072914	G	A	13
2097520	G	T	7
2098628	T	C	11
2117516	G	A	22
2119347	T	A	7
2141322	T	C	5
2141910	C	A	10
2218046	G	T	5
2235109	T	C	12
2262577	T	G	13
2277583	G	A	40
2278317	A	G	5
2281856	A	C	10
2300659	T	G	8
2304950	A	G	27
2317730	A	C	18
2356003	T	A	16
2453454	A	G	11
2508389	T	C	20
2548551	G	T	19
2575152	G	A	33
2577686	A	G	24
2607034	C	T	18
2619611	T	G	9
2655012	C	T	21
2684793	A	G	49
2721828	C	A	14
2725715	C	T	17
2739149	C	A	6
2744933	A	G	6
2773647	A	G	22
2787770	T	G	13
2812384	G	T	17
2829833	A	G	30
2865494	A	C	9
2894703	T	C	16
2896636	A	G	41
2903882	T	G	10
2934972	C	G	40
2936268	G	A	17
2950954	G	A	14
2958327	C	T	39
2995771	A	G	23
3085079	A	G	22
3096319	G	A	17
3143800	G	T	29
3145523	A	C	11
3155055	G	C	54
3190399	A	G	28
3210101	A	G	16
3244204	A	G	11
3248223	T	C	10
3324959	A	G	25
3358603	G	T	30
3362591	A	G	7
3392897	A	G	79
3398153	G	A	9

3403167	G	T	16
3409414	T	C	39
3421335	A	G	26
3472427	G	A	18
3500922	T	G	9
3560088	G	A	10
3564026	C	T	13
3568597	C	T	25
3571531	A	G	20
3616733	A	G	26
3667806	A	G	30
3725545	T	C	19
3726726	A	G	15
3732919	A	G	20
3737968	G	A	19
3739401	C	A	9
3750736	G	A	6
3755861	C	T	18
3767613	C	T	31
3806677	C	T	30
3813424	C	A	23
3843195	C	A	10
3909258	T	C	10
3973746	C	T	12
4080579	T	C	12
4081612	T	C	30
4083536	A	G	35
4087224	T	C	28
4170791	A	G	87
4173149	A	C	28
4194600	G	A	16
4199187	A	G	70
4199190	T	C	123
4203596	G	T	51
4210011	T	C	262
4232217	C	T	5
4243823	A	T	18
4307755	G	A	7
4339366	T	G	18
4399470	A	G	13
4421633	T	C	12
4421689	A	G	5
4460688	C	T	19
4465967	C	A	13
4518401	G	A	18
4527483	A	G	27
4542642	A	G	9
4579183	A	G	24
4628496	C	A	7
4629169	G	A	25
4634287	A	G	30

Table S10: A120 and A76 non-synonymous chromosomal SNPs per annotated gene in comparison with C092 alone. Base positions are based on CO92. Orange highlights indicate a SNP called only in A120. Blue highlights indicate a SNP called only in A120+A76.

Position (CO92)	REF	Justinian	Coverage (A120+A76)	Codon Change	Amino Acid Change	Polymorphism Type	Gene
130643	A	G	35	ACC -> ATC	T -> I	SNP (transition)	glpE
221811	G	A	63	ATT -> GTT	I -> V	SNP (transition)	rpsJ
341720	G	A	19	CAC -> CGC	H -> R	SNP (transition)	rhaS

399533	A	C	10	TTC -> TTA	F -> L	SNP (transversion)	aidB
727741	A	G	13	GAA -> AAA	E -> K	SNP (transition)	parE
877258	C	T	20	TTT -> CTT	F -> L	SNP (transition)	lysR
918790	T	C	16	CCA -> CTA	P -> L	SNP (transition)	kduD2
1102174	G	A	17	ACA -> GCA	T -> A	SNP (transition)	iucD
1371020	A	C	8	CTG -> ATG	L -> M	SNP (transversion)	gyrA
1754708	T	C	12	CCA -> CTA	P -> L	SNP (transition)	galU
1895361	A	C	14	CCA -> CAA	P -> Q	SNP (transversion)	motB
2012524	G	T	7	TTG -> GTG	L -> V	SNP (transversion)	hpaI
2098628	C	T	11	AAT -> AGT	N -> S	SNP (transition)	putA
2119347	A	T	7	CAA -> CAT	Q -> H	SNP (transversion)	tnpA
2141322	C	T	5	AGG -> GGG	R -> G	SNP (transition)	fyuA
2141910	A	C	10	GTC -> TTC	V -> F	SNP (transversion)	fyuA
2218046	T	G	5	GGC -> GTC	G -> V	SNP (transversion)	hmsR
2356003	A	T	16	AAT -> AAA	N -> K	SNP (transversion)	fadD
2508389	C	T	20	ACA -> GCA	T -> A	SNP (transition)	cstA
2812384	T	G	17	CCG -> CAG	P -> Q	SNP (transversion)	gutB
2936268	A	G	17	CTT -> TTT	L -> F	SNP (transition)	gltJ
2950954	A	G	14	GCG -> GTG	A -> V	SNP (transition)	nagC
3085079	G	A	22	ACT -> GCT	T -> A	SNP (transition)	mepA
3143800	T	G -> T	29	GCC -> GAC	A -> D	SNP (transversion)	ynbD
3190399	G	A	28	CAG -> CGG	Q -> R	SNP (transition)	baeR
3248223	C	T	10	AGA -> GGA	R -> G	SNP (transition)	ail
3421335	G	A	26	ATG -> GTG	M -> V	SNP (transition)	bcp
3472427	A	G	18	GCT -> GTT	A -> V	SNP (transition)	ddhD
3500922	G	T	9	GTG -> GGG	V -> G	SNP (transversion)	tesB
3564026	T	C	13	TGT -> TAT	C -> Y	SNP (transition)	proY
3568597	T	C	25	GGA -> GAA	G -> E	SNP (transition)	phoB
3616733	G	A	26	CTA -> CCA	L -> P	SNP (transition)	hmwA
4080579	C	T	12	CAC -> CGC	H -> R	SNP (transition)	accB
4194600	A	G	16	GTT -> ATT	V -> I	SNP (transition)	thiG
4243823	T	A	18	TTT -> TAT	F -> Y	SNP (transversion)	ubiE
4339366	G	T	18	AGT -> CGT	S -> R	SNP (transversion)	wzzE
4421633	C	T	12	GTG -> GCG	V -> A	SNP (transition)	glpD

4421689	G	A	5	ACA -> GCA	T -> A	SNP (transition)	glpD
4579183	G	A	24	AGC -> GGC	S -> G	SNP (transition)	fdhD
4634287	G	A	30	TCG -> CCG	S -> P	SNP (transition)	phoU

Table S11: 5X, 90% variant frequency plasmid pCD1 SNPs

Position	pCD1	Justinian	Coverage	Amino Acid Change	Polymorphism Type	Protein Effect	gene
23564	T	G	42		SNP (transversion)	None	lcrR
29959	A	G	15	N -> S	SNP (transition)	Substitution	yscO
46912	A	G	23		SNP (transition)		
50462	T	C	16	K -> E	SNP (transition)	Substitution	yopJ
54237	T	G	23		SNP (transversion)		
55839	G	A	45		SNP (transition)		
66608	C	T	9	L -> F	SNP (transition)	Substitution	

Table S12: 5X, 90% variant frequency plasmid pMT1 SNPs

Position	pMT1	Justinian	Coverage	Amino Acid Change	Polymorphism Type	Protein Effect	gene
32569	T	G	7		SNP (transversion)	None	
62994	T	C	10		SNP (transition)		
82435	C	T	7		SNP (transition)		

Table S13: SNPs called in A120+A76 that were removed from final analyses.

Position	CO92	Justinian	Coverage A76+A120 (Geneious)	Reason for removal
362357	C	T	127 (117 T, 10 C)	Coverage spike, long reads don't have SNP
363211	T	C	101 (65 C, 35 T)	Not 90% variant in Geneious
395768	A	G	350 (326 G, 23 A, 1 T)	Coverage spike, long reads don't have SNP
395774	C	G	751 (726 G, 22 C, 1 A, 2 T)	Coverage spike, long reads don't have SNP
497764	A	T	76 (70 T, 6 A)	Phased with 497767
497767	G	T	110 (103 T, 7 G)	Phased with 497764
1308293	A	G	146 (138 G, 6 A, 1 T)	Coverage spike, long reads don't have SNP
1572689	T	C	287 (262 C, 25 T)	Coverage spike, long reads don't have SNP
1859946	T	C	12 (10C, 2T)	Not 90% variant in Geneious
2073031	C	G	14 (12 G, 2C)	Not 90% variant in Geneious
2591088	A	G	26 (24 G, 2 A)	Long reads don't have SNP, phasing
2968425	A	G	12 (10 G, 1 A, 1 T)	Not 90% variant in

				Geneious
3017615	A	G	609 (561 G, 48 A)	Coverage spike, phasing
3070065	G	C	413 (403 C, 8 G, 2 A)	Coverage spike, long reads don't have SNP
3373238	C	T	6 (5T, 1 C)	Not 90% variant in Geneious
3,426,560	A	G	5	Not callable by GATK
3476761	A	G	381 (352 G, 29 A)	Coverage spike, long reads don't have SNP
3956001	T	A	12 (10A, 2 T)	Not 90% variant in Geneious
4199220	A	G	216 (203 G, 12 A, 1 T)	Coverage spike, long reads don't have SNP
4210011	T	C	262 (233 C, 21 T, 6 A, 2 G)	Coverage spike, long reads don't have SNP
4203595	C	A	44 (40 A, 4 C)	Coverage spike, long reads don't have SNP
4371637	A	C	123 (113 C, 7 A, 3 T)	Phased with 4371652, coverage spike
4371652	A	G	392 (369 G, 23 A, 1 T)	Phased with 4371637, coverage spike
4509055	C	G	555 (521 G, 33 C, 1 A)	Coverage spike, long reads don't have SNP
4575345	A	G	156 (149 G, 7 A)	Coverage spike, phasing
4645996	T	C	289 (263 C, 25 T)	Coverage spike, long reads no SNP

References:

1. Harbeck M, Seifert L, Hansch S, Wagner DM, Birdsell D, Parise KL, et al. Yersinia pestis DNA from Skeletal Remains from the 6(th) Century AD Reveals Insights into Justinianic Plague. PLoS Pathog. 2013;9(5):e1003349. Epub 2013/05/10.
2. Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, et al. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Res. 2009;37(10):3215-29.
3. Schuenemann VJ, Bos K, DeWitte S, Schmedes S, Jamieson J, Mittnik A, 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. 2011;108(38):E746-52. Epub 2011/08/31.
4. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor protocols. 2010;2010(6):pdb prot5448. Epub 2010/06/03.
5. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012;40(1):e3. Epub 2011/10/25.
6. Phillippy AM, Deng X, Zhang W, Salzberg SL. Efficient oligonucleotide probe selection for pan-genomic tiling arrays. BMC bioinformatics. 2009;10:293. Epub 2009/09/18.
7. Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, et al. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Proc Natl Acad Sci U S A. 2013;110(2):577-82. Epub 2012/12/29.
8. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1). Epub 2011-05-17.
9. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957-63. Epub 2011/09/10.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20.
11. Biomatters. Geneious. R6 ed2012.
12. Fan L, Yao YG. MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. Mitochondrion. 2011;11(2):351-6. Epub 2010/10/12.

13. Garrison E, Marth G. Haplotype -based variant detection from short-read sequencing data. . 2012.
14. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8. Epub 2011/04/12.
15. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics. 2011;27(15):2153-5. Epub 2011/06/11.
16. Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764-70. Epub 2011/01/11.
17. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. Epub 2010/01/30.
18. Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al. Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. PNAS. 2004;101(38):13826-31.
19. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21. Epub 2010/06/09.
20. Swofford D. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 4 ed. Sunderland, Massachussetts: Sinauer Associates; 2003.
21. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29(8):1969-73. Epub 2012/03/01.
22. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nature Genetics. 2010;42(12):1140-3.