nature portfolio

https://doi.org/10.1038/s41586-022-05349-x

Supplementary information

Evolution of immune genes is associated with the Black Death

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION

Black Death shaped the evolution of immune genes

Jennifer Klunk[1,2]*, Tauras P. Vilgalys[3]*, Christian E. Demeure[4], Xiaoheng Cheng [5], Mari Shiratori[3], Julien Madej[4], Rémi Beau[4], Derek Elli[6], Maria I. Patino[3], Rebecca Redfern[7], Sharon N. DeWitte[8], Julia A. Gamble[9], Jesper L. Boldsen[10], Ann Carmichael[11], Nükhet Varlik[12], Katherine Eaton[1], Jean-Christophe Grenier[13], G. Brian Golding[1], Alison Devault[2], Jean-Marie Rouillard[2,14], Vania Yotova[15], Renata Sindeaux[15], Chun Jimmie Ye[16, 17], Matin Bikaran[16, 17], Anne Dumaine[3], Jessica F Brinkworth[18,19], Dominique Missiakas[6], Guy A. Rouleau[20], Matthias Steinrücken[5,21], Javier Pizarro-Cerdá[4], Hendrik N. Poinar[1,22,23#], Luis B. Barreiro[3,21,24,25#]

*These authors contributed equally to this work, and are presented in alphabetical order
[#]These authors jointly supervised this work

Correspondence should be addressed to: lbarreiro@uchicago.edu; poinarh@mcmaster.ca

Author Affiliations
[1]*McMaster Ancient DNA Centre, Departments of Anthropology Biology and Biochemistry, McMaster University, Hamilton, Ontario, Canada L8S4L9*
[2]*Daicel Arbor Biosciences, Ann Arbor, MI, USA*
[3]*Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA*
[4]*Yersinia Research Unit, Institut Pasteur, Paris, France*
[5]*Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA*
[6]*Department of Microbiology, Ricketts Laboratory, University of Chicago, Lemont, IL, USA*
[7]*Centre for Human Bioarchaeology, Museum of London, London, UK, EC2Y 5HN*
[8]*Department of Anthropology, University of South Carolina, Columbia, SC, USA*
[9]*Department of Anthropology, University of Manitoba, Winnipeg, Manitoba, R3T2N2*
[10]*Department of Forensic Medicine, Unit of Anthropology (ADBOU), University of Southern Denmark, Odense S, 5260, Denmark*
[11]*History Department, Indiana University, Bloomington, IN, USA*
[12]*Department of History, Rutgers University-Newark, NJ, USA*
[13]*Montreal Heart Institute, Faculty of Medicine, Université de Montréal, Montréal, Quebec, Canada, H1T 1C7*
[14]*Department of Chemical Engineering, University of Michigan Ann Arbor, Ann Arbor, MI, USA*
[15]*Centre Hospitalier Universitaire Sainte-Justine, Montréal, Quebec, Canada, H3T 1C5*
[16]*Division of Rheumatology, Department of Medicine, University of California, San Francisco, CA, USA.*
[17]*Institute for Human Genetics, University of California, San Francisco, CA, USA.*[18]*Department of Anthropology, University of Illinois Urbana-Champaign, Urbana, IL, USA*
[19]*Carl R Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA*
[20]*Montreal Neurological Institute-Hospital, McGill University, Montréal, Quebec, Canada, H3A 2B4*
[21]*Department of Human Genetics, University of Chicago, Chicago, IL, USA*
[22]*Michael G. DeGroote Institute of Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada L8S4L9*
[23]*Humans and the Microbiome Program, Canadian Institute for Advanced Research, Toronto, Canada*
[24]*Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA*
[25]*Committee on Immunology, University of Chicago, Chicago, IL, USA*

TABLE OF CONTENTS

h. Table S8: Cytokine Data
i. Table S9: Colony Forming Unit (CFU) Data
j. Table S10: Interaction effects between *ERAP2* genotype and *Y. pestis* stimulation

1. Supplementary Methods

a. Sampling locations and dates

The details of all remains used in this study, including burial locations and associated dates can be found in [20] as well as in the **Table S1**. For the purposes of our study, we split the samples from London into three time points: early or Pre-Black Death (~1000-1250), Black Death (1348-1350) and late or Post-Black Death (1350-1500) (**Table S1**). The London cemeteries are dated based on a combination of primary sources, Bayesian radiocarbon dating, and archaeological remains (primarily coins). We did the same for our Danish samples, as best we could from associated dates, which are less precise due to the lack of records associated with many of the cemeteries and the imprecision of the dating method based on arm position at burial used to date medieval burials in this region. We therefore group Danish samples into either an Early (~850-1350) or Late (1350-1800) period corresponding to before or after the Black Death period.

Access to the remains was facilitated by Museum of London Center for Human Bioarchaeology for the individuals from London and by Horsens Museum, the Viborg Museum, the Moesgaard Museum, Museum Lolland-Faster, Øhavsmuseet, Faaborg, and ADBOU for the individuals from Denmark. No further permissions were required for this study.

b. Ancient DNA sample processing, bait design, enrichment, sequencing and data processing

i. Sample processing

All sample processing was performed in the specialized cleanroom facilities of the McMaster Ancient DNA Centre (McMaster University, Hamilton, Ontario). Individuals from three cemeteries in London, England (n=325), and five localities from Denmark (n=198) were included in this study (**Fig 1**). Both reagent blanks and carrier blanks (bacteriophage lambda DNA sheared to an average of 150 bp) were processed alongside the samples at 1:20 reagent blank and 3:20 carrier blank ratio. Subsamples of 10-300 mg of long bone, tooth (root or pulp cavity), or petrous portion of the temporal bone were pulverized manually. The resulting bone or tooth fragments and powder were demineralized and digested using a previously described protocol[52,20]. The DNA was purified using either a modified phenol:chloroform:isoamyl alcohol (PCI) method or a large-volume guanidinium hydrochloride method designed to retain short DNA fragments[53-55]. Samples were screened for the presence of single copy nuclear DNA using an 81bp quantitative PCR assay targeting a portion of the nuclear *c-myc* gene[21] or a shortened 52 bp version of the assay[20]. Extracts of samples that returned positive results were prepared into double-stranded, double indexed Illumina-style sequencing libraries[55,56] for downstream enrichment.

ii. Bait design

Three sets of baits were designed and used for the enrichment of i) neutral, ii) GWAS, and iii) exon targets (**Tables S2 & S3**).

1. **Neutral.** 250 independent (non-linked) and putative neutral loci were randomly selected from a larger set of “neutral loci” defined by Gronau and colleagues[22].

Briefly, these are regions outside any known protein-coding and conserved noncoding region of the genome, and with low intralocus recombination but sufficient between regions that the genealogies are largely uncorrelated. All bait regions were initially designed to target 1,500 bp, but some of the targets required shortening to avoid non-specific hybridization. The final neutral bait set included 16,062 probes, each 80bp in length tiled every 20 bp.

2. **GWAS.** A total of 496 GWAS variants associated with immune-related disorders (**Table S2**). Initially, the targets included 50 bp on either side of the variant of interest but were redesigned to include 80 bp on either side after experiments showed better capture success. The redesigned baits outperformed the old baits in target coverage breadth and depth by a factor of 3, at approximately 4.8x lower sequencing depth. Probes for both variant alleles were designed. The final GWAS bait set included 5,800 probes, 80 bp in length tiled every 20 bp.
3. **Exons.** A total of 3,512 sequences, including all exons from 356 genes associated with immune function were included in the exon bait set. The genes were manually curated and chosen based on their known role in immune processes. 69% of all 399 genes are associated with one or more of the following GO terms ("immune function", "immune system process", "Innate Immunity", "adaptive immunity", "cytokine", "NF-kappaB", and "interleukin"). The total number of genes targeted was constrained by the library volume remaining from these precious remains. Using the maximum size of existing bait sets as our threshold we fitted the 3,512 sequences, representing a total of 2,506,004 bp into a bait set of 93,180 probes. The probes were designed to be 60 bp in length and tiled every 20 bp.

iii. Enrichment

Samples were enriched using baits targeting the neutral and GWAS loci (75ng of each probe set, resulting in a final mass of 150ng of baits per reaction) according to the myBaits v3 protocol (Daicel Arbor Biosciences), with minor modifications to result in the recommended final concentrations at a final volume of 26 µL instead of 30 µL, allowing a maximum of 10.75 µL template input instead of 7 µL template input., and using 20 µL of beads (pre- and post- wash) instead of 30 µL pre-wash and 70 µL post-wash. A volume of 5 µL template input was used. During the removal of the supernatant post-incubation with the magnetic beads (2B.5 Bead Washing, step 1), the supernatant was saved, labeled as the non-enriched fraction (NEF), and frozen at -20°C. 25 µL of the NEF was used as input to the new enrichment for the exon probes, with 150 ng baits, with incubations at 55°C. Samples were reamplified with the following conditions: Real-time PCR was per- formed in 10 µL reactions consisting of: 1X KAPA SYBR® FAST qPCR Master Mix (2X), 150 nM of each primer (IS5_long_amp.P5: AATGATACGGCGACCACCGA and IS6_long_amp.P7: CAAGCAGAAGACGGCATACGA), and 18.8 µL template. The cycling conditions were: 5 minutes of initialization at 95°C, followed by 12 cycles of a 30 second denaturation at 95°C (30 sec) and a 45 second annealing/extension at 60°C, followed by a final extension for 3 minutes at 60°C, and ending with a 1 minute hold at 4°C. A second round of enrichment was performed with the modifications above. For libraries requiring a second exon enrichment, a new aliquot of library was indexed and 5 µL was used as the input for capture using the protocol described above.

iv. Pooling, size selection, and sequencing

Sample concentration was quantified via qPCR in 0.001x dilutions against the 425bp- 525bp PhiX Control (Illumina) standard serially diluted from 1 nM to 62.5 fM. Real-time PCR was performed in 10 µL reactions consisting of: 1X KAPA SYBR® FAST qPCR Master Mix (2X), 200 nM of each primer (IS5_long_amp.P5 and IS6_long_amp.P7, as above), 4 µL template, and nuclease-free water. The cycling conditions were: 5 minutes of initialization at 95°C, followed by 35 cycles of a 30 second denaturation at 95°C (30 sec) and a 45 second annealing/extension at 60°C, ending with a 30 second hold at 8°C.

Samples were pooled at equimolar ratios. If the pool volume was greater than 12 µL, the pool was concentrated over a Minelute PCR Purification column, with above modifications. Size selection was performed on a 3% NuSieve GTG Agarose Gel with a GeneRuler 50 bp ladder (Fermentas), run for 35 minutes at 100V, and then bands 150-500 bp were excised. Gel slices were purified using a Minelute Gel Extraction kit, with the above modifications. Pools were eluted in 20 µL Qiagen Buffer EB. Pools were sequenced at the Farncombe Metagenomics Facility on the Illumina HiSeq 1500 platform, using V2 Rapid chemistry, with 2x90 paired end reads. Each sample received an average of 732,209 clusters (range: 3,549- 20,052,880).

v. Data processing

Adapters were trimmed and overlapping reads merged using *leeHom* (v1.2.5) with aDNA specific settings[57]. Reads were mapped to each target set separately using a modified version of *BWA* which allows for bam-to-bam mapping[58] (v0.5.10-evan.10, https://github.com/udo-stenzel/network-aware-bwa). PCR duplicates were removed with biohazard software (v1.0.2, https://github.com/udo-stenzel/biohazard) and reads with quality 30 and length below 24bp were removed with *samtools* (v1.9)[59]. Unique indexed libraries stemming from the same individual were merged using *samtools* and RG tags were standardized with a custom perl script.

Reads mapped strictly the targeted region of the genome produced excess heterozygosity, presumably because of slightly divergent sequences which result in off-target enrichment and are falsely interpreted as alternative alleles[60]. To correct this, we remapped reads to the entire human genome (hg19) and only retained uniquely mapped reads. Resulting bam files were sorted and indexed using *samtools*[59]. We then merged the exon, immune loci, and neutral loci enriched reads together, reasoning that off-target reads which overlap other regions of interest remain valuable and that sequencing from the same sample should share similar patterns of ancient DNA damage. Quality scores in each bam file were adjusted for DNA degradation using *mapDamage 2.0*[61], and then trimmed of the first and last 4 bases using bamUtil (trimBam –L 4 –R 4)[62], where ancient DNA damage is typically concentrated.

Analysis of damage patterns was performed with *mapDamage 2.0* and plotted in *R* using the *ggplot2* package[61,63]. No blanks yielded enough mappable sequences to be included in any downstream analyses. Our final dataset contained 33,110 biallelic, non-singleton variants within the targeted regions (2,669 near GWAS loci, 19,972 in immune genes, and 10,469 in neutral regions), with a mean coverage of 4.6x reads per site per individual (see **Table S1** for individual

coverage). Because 4.6x is moderately low coverage relative to many resequencing data sets from modern genomes, we acknowledge that mapping-related biases have the potential to affect our genotype calls (we note, however, that it is deeper coverage than typical for ancient genomes: e.g., in a recent large data set of 804 individuals, median autosomal coverage is 2.2x[64]). We therefore used continuous estimates of genotype likelihoods (see section C. below) instead of discrete genotype calls, which build in uncertainty due to coverage. We also required our top candidates for selection to be replicated in both the London and Denmark ancient DNA cohorts.

c. Genotype calling

Genotypes were called for each individual using GATK's Haplotype Caller in gVCF mode (version 4.1.4.1) with default setting and a minimum base quality of 20[65]. Genotype calls across individuals were then combined using the CombineGVCFs and GenotypeGVCFs functions, and filtered for biallelic sites with minQ ≥ 30 in regions of the genome which were targeted for enrichment (n=139,867 sites). We next excluded singleton variants (n=106,757) as they are more prone to reflect DNA damage or sequencing errors, resulting in a dataset of 33,110 biallelic, non-singleton variants within the targeted. We then excluded samples with missing genotype calls at more than 50% of those sites (retaining n=206 individuals) and filtered for sites with at least 10 individuals per population and any given time point, where the alternate allele was called in at least 4 of the 5 population-time points, and where the alternate allele was identified in the 1000 Genomes project. Our final dataset therefore contains 2,689 variants (584 near GWAS loci, 1,468 in immune genes and 637 in putatively neutral regions). Because of the low coverage intrinsic to ancient genomic studies, many genotype calls were based on a few reads. We therefore capture uncertainty in genotype calls by extracting the genotype likelihoods for each individual at each site using filtbaboon1b from *LCLAE*[66]. From these estimates, we use a maximum-likelihood based approach to estimate the allele frequency. Specifically, we let the genotype likelihoods for individual *i* in 1:n be given by x_i^j where *j* is the genotype denoted as 0, 1, or 2 alternate alleles. The ML estimate of the allele frequency *p* is then:

$$\hat{p} = argmax\left\{\sum_{i=1}^{n} \log(x_i^0(1-p)^2 + 2x_i^1 p(1-p) + x_i^2 p^2)\right\}, with\ p\ \epsilon\ [0,1]$$

We then calculate the minor allele frequency per population and time point, retaining only sites with a mean minor allele frequency between London and Denmark greater than 0.05 (n = 1,726 candidate sites), as our power to detect selection for variants below 5% frequency is very low (**Fig S2**).

d. Genetic differentiation following the Black Death

We calculated F_{ST} for each variant as the loss in expected heterozygosity between time points relative to the expected heterozygosity of a panmictic population[67]. Specifically, $F_{ST} = \frac{Ehet_{panmictic} - mean(Ehet_{populations})}{Ehet_{panmictic}}$, where E_{het} (the expected heterozygosity) is calculated for each population-time point and a hypothetical panmictic population from the allele frequency assuming Hardy-Weinberg equilibria. Because high variance in estimating F_{ST} at low allele

frequencies can lead to high false positive rates[68], we also filtered out variants with minor allele frequencies less than 5% (calculated as the mean across populations). After this, we compared 1,726 sites in functional regions to 637 putatively neutral loci; functional loci showed an enrichment for high levels of differentiation compared to the neutral loci (**Fig 2A-B**).

We then asked whether sites in functional regions were more differentiated than expected based upon neutral loci. Specifically, we compared the F_{ST} values in functional regions to neutral loci, calculating a p-value for each site as the proportion of neutral loci with a more extreme F_{ST}. As these proportions were generally stable when controlling for minor allele frequency, we maximized the number of neutral sites to compare against by not further restricting the set of neutral loci based on difference in minor allele frequency from the tested variant. Enrichment in the Denmark cohort is smaller than in London, although still considerably higher than expected based on forward simulations (**section 1.e**). There are several reasons why the population in London is better suited to detect positive selection relative to those in Denmark: the dataset from London is ~50% larger (which directly impacts our ability to accurately estimate allele frequencies; hence, power), comes from a smaller geographic area, and from samples which were more accurately dated and occurred over a shorter period of time. Therefore, putatively neutral sites in Denmark show larger F_{ST} values before vs after the Black Death than those in London (~100% larger on average, paired t-test $p = 1.34 \times 10^{-42}$), reflecting increased population heterogeneity and/or our reduced ability to accurately estimate allele frequencies in Denmark given the small sample sizes. Thus, the reduced signal in Denmark (which nonetheless remains quite large: 3.9x the number of variants expected by chance exceed the 99th percentile of neutral loci) can likely reflects contributions of these technical and demographic differences.

The strongest evidence of selection would come from candidate loci with concordant evidence of high differentiation across populations and time points. In particular, we required three conditions be met for a loci to be of interest: (i) individuals who survived the Black Death in both London and Denmark, have the same direction of allele-frequency change, (ii) within London, allele frequencies of Black Death survivors are shifted in the opposite direction of individuals who died during the plague (e.g. those buried in East Smithfield), and (iii) variants of interest are more differentiated than expected based on variation at neutral loci. To address point iii, we required sites be more differentiated than 95% of neutral loci in London before vs after the plague and be more differentiated than 90% of neutral loci in Denmark before vs after the plague.

To ensure that widespread gene flow did not drive these signatures of selection, we proceeded with two additional analyses. We note that a population replacement would affect the totality of the genome and therefore the fact that our enrichments of highly-differentiated variation among immune loci is relative to putative neutral regions sequenced in the same exact samples should alleviate that concern. That said, we cannot exclude the possibility that the highly differentiated sites that we identified correspond to sites that were already unusually differentiated between the local population and the one replacing it. Therefore, we first reasoned that if our highly differentiated loci in London when comparing pre- and post-Black Death

samples were the result of a significant population replacement, those same loci should be amongst the most differentiated amongst contemporary European populations. We found no evidence that the 227 variants we identify as outliers in London before vs after the plague are enriched for large F_{ST} values among European populations, relative to variants that were not highly differentiated. Out of all pairwise comparisons of the five European 1000 Genomes populations, the only evidence for elevated genetic differentiation at candidate loci was observed when contrasting between Iberian (IBS) and Tuscany (TSI) (t-test p=0.0051), which is an implausible populations pair to explain populations replacements in London in the 14th and 15th century (especially given that Great Britain is not even included in this pair, despite the fact that among our strongest candidate variants, post-plague allele frequencies in London are most similar to modern 1000 Genomes frequencies from Great Britains in London: mean difference of 1.8% vs 5.5-6.7% for other European populations).

e. Forward simulations of neutral evolution

Our results indicate an excess of highly diverged alleles near immune-related genes and GWAS variants relative to neutral regions of the genome. However, the genomic context of candidate sites differs from neutral regions in ways that may affect the variance in F_{ST}[69,70]. We therefore contrasted the recombination rate (derived from the DeCODE project[71]) and the level of background selection (as measured by the B statistic[72]) between the neutral regions and our targeted immune loci. Recombination rates were significantly different between neutral regions and candidate loci, with an overall trend for higher recombination rates around our candidate loci ($p = 4.03 \times 10^{-9}$; **Fig ED1A**). Levels of background selection were also significantly higher around immune loci as compared to putative neutral regions ($p \approx 0$; **Fig ED1B**). Importantly, neither recombination rate nor background selection was enriched for candidates of positive selection relative to other tested loci in exonic regions or near GWAS hits ($p > 0.5$, **Fig ED1A-B**).

Nevertheless, since recombination rate and background selection can affect the variance in F_{ST} (albeit in opposing directions[69,70]), we used forward simulations to confirm that the differences in genetic differentiation observed between neutral and candidate regions are unlikely under neutral evolution. Specifically, we used SLiM (v3.7.1)[73,74] to simulate the neutral evolution of full genomes in a population of 10,000 individuals while controlling for recombination rates (estimated by the DeCODE project[71]) and negative selection against mutations in coding regions (defined using the UCSC table of canonical genes[75]). Deleterious mutations were introduced at a rate of 7.4×10^{-8} mutations per base pair per generation, and an exponential distribution of fitness effects with a mean deleterious effect size of 2.5×10^{-3} (inferred by [72]). We simulated 1,000 generations under this model, allowing for neutral and deleterious variation to accumulate and stabilize, before we simulated our data collection surrounding the Black Death. At 1,000 generations, we extracted “pre-BD” allele frequencies for variants within 5kb of genomic regions targeted in this study. Assuming a generation time of approximately 30 years, we simulated an additional 8 generations before the BD event (equivalent to the mean number of generations of our pre-plague London samples), the BD which led to 50% mortality in the population, and a further 7 generations after the BD event (equivalent to the mean age for our post-plague London samples), at which point we sampled “post-BD” allele frequencies.

As in our main analyses, we then asked whether sites near candidate regions were more strongly differentiated before vs after the plague than those of neutral regions. We detected a slight enrichment of highly differentiated sites that was much smaller than those observed for our empirical data (**Fig ED1C-D**), suggesting background selection and recombination are insufficient to explain the enrichment of highly differentiated sites we observe. Importantly, these results are consistent across a range of parameter values including population sizes up to 50,000 (the approximate census population in London at the beginning of the Black Death) and mortality rates from 30 to 50%.

f. Gene regulatory response to *Y. pestis*

i. Isolation of monocytes, differentiation of macrophages, and challenge with heat-killed *Y. pestis*

This study has been approved by the Institutional Review Board at the University of Chicago (protocol #: IRB19-0432). All Buffy coats from 33 healthy donors were obtained from the Indiana Blood Center (Indianapolis, IN, USA). All individuals recruited in this study were males, self-identified as African-American (AA) (n = 17) or European-American (EA) (n = 16) between the age of 18 and 55 years old. Only individuals self-reported as currently healthy and not under medication were included in the study. In addition, each donor's blood was tested for Hepatitis B, Hepatitis C, Human Immunodeficiency Virus (HIV), and West Nile Virus, and only samples negative for all of the tested pathogens were used.

Blood mononuclear cells were isolated by density gradient centrifugation (Ficoll-Paque Premium, Sigma Aldrich, St. Louis, MI, USA). Monocytes were purified from peripheral blood mononuclear cells by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotech, Bergisch Gladbach, Germany) using the autoMACS Pro Separator. The purity of the isolated monocytes was verified using an antibody against CD14 (BD Biosciences) and only samples showing > 90% purity were used to differentiate into macrophages. Monocytes were then cultured for 7 days in RPMI-1640 (Fisher) supplemented with 10% heat-inactivated FBS (FBS premium, US origin, Wisent), L-glutamine (Fisher) and M-CSF (20ng/mL; R&D systems). Cell cultures were fed every 2 days with complete medium supplemented with the cytokines previously mentioned. Before inoculation with *Y. pestis*, we checked the differentiation/activation status of the monocyte-derived macrophages by flow cytometry, using antibodies against CD1a, CD14, CD83, and HLA-DR (BD Biosciences). All samples presented the expected phenotype for non-activated macrophages (CD1a+, CD14+, CD83, and HLA-DRlow). Macrophages were stimulated for 4 hours with lysate of heat-killed *Y. pestis* (strain CO92) at a value of 10 colony forming units (CFUs) per cell, or with endotoxin-free water (negative control).

ii. RNA Extraction, Library Preparation, and Sequencing

Total RNA was extracted from the non-stimulated and stimulated macrophages using the miRNeasy kit (QIAGEN). RNA quantity was evaluated spectrophotometrically, and the quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies). Only samples with no evidence for RNA degradation (RNA integrity number > 8) were kept for further experiments.

RNA- sequencing libraries were prepared using the Illumina TruSeq protocol. Once prepared, indexed cDNA libraries were pooled (6 or 7 libraries per pool) in equimolar amounts and were sequenced with single-end 100bp reads on an Illumina HiSeq2500.

Adaptor sequences and low quality score bases (Phred score < 20) were first trimmed using Trim Galore (version 0.2.7). The resulting reads were then mapped to the human genome reference sequence (Ensembl GRCh37 release 75) using STAR (2.4.1d)[76] with a hg19 transcript annotation GTF file downloaded from Ensembl. Gene-level expression estimates were calculated using RSEM (version 1.2.21)[77] with default parameters.

After removing lowly expressed genes (mean counts per million < 1 in both stimulated and null conditions), we normalized expression data using voom and analyzed 5,874 genes using the R package *limma* (v3.48.3). We first tested for an effect of stimulation with *Y. pestis* while controlling for individual identity for an effect of stimulation on expression level, while controlling for individual identity. We then calculated a permutation-based FDR, comparing our observed results to permuted results observed from randomizing which sample was stimulated within each individual (repeated 25 times). We extracted expressed genes within 100kb of our candidate loci for positive selection, retaining *ERAP1*, *ERAP2*, *LNPEP*, *TICAM2*, *TMED7*, *CTLA4*, and *ICOS*. We then used previously published genotype data for these individuals[31] to test whether genotype at our candidate variants was associated with expression level using a linear model. Finally, we tested for an interaction effect between stimulation and genotype. Despite several trends which appear to resemble interaction effects, our sample was underpowered to identify any significant interactions. We applied an identical computational pipeline to test for effects of stimulation and genotype in previously published datasets using different bacterial and viral pathogens and ligands[31,32].

iii. PCR for ERAP2 transcripts

750ng of total RNA were reverse-transcribed using the Superscript III First Strand Reverse Transcriptase Kit. We measured gene expression levels of *ERAP2* in macrophages by real-time PCR using both a commercial Taqman probe that quantifies total transcript levels (Hs01073633_m1) as well as a custom probe that specifically quantifies the transcript levels of Haplotype B by amplifying the extended version of exon 10 (assay ID: APDJ3E7, part #4331348). As housekeeping gene we used *GAPDH* (Hs02758991). Real-time PCR was performed on the QuantStudio 3 from Applied Biosystems. The relative expression reported in Figure 3F refers to the fold-change expression values of *ERAP2* relative to a common sample ran across all real-time PCR plates (after normalization against *GAPDH* levels). In addition to the real-time PCR, we also performed a regular PCR aimed at distinguishing between the transcribed long and short versions of exon 10 using the PCR primers and conditions described in Andres *et al.*[25] and separated the PCR products on a Bioanalyzer chip (DNA 1000 kit, part# 5067-1504, Agilent Technologies).

iv. Comparison with Quach et al. and Nedelec et al. data

To evaluate if the differences in gene expression identified among our candidate loci were specific to flu or shared in response to other pathogens we re-analyzed existing datasets:

Nedelec and colleagues[31] measured the gene expression response of monocyte-derived macrophages to infection with two live intracellular bacteria: *Listeria monocytogenes* (a Gram-positive bacterium) and *Salmonella typhimurium* (a Gram-negative bacterium), using the same experimental design and time point as the current study. Quach and colleagues[32] characterized the transcriptional response of primary monocytes to bacterial and viral ligands activating Toll-like receptor pathways (TLR1/2, TLR4, and TLR7/8) and live influenza virus 6 hours after stimulation. Processing of the RNAseq data from these datasets and analysis was done using the same methods described above in section **ii**.

v. Infection of macrophages with live *Y. pestis*: cytokine and gentamycin protection assays

In addition to the stimulation of macrophages with heat-killed *Y pestis* (**section 1.f.i** above), we also performed a separate set of infections with live *Y. pestis* of a set monocyte-derived macrophages derived from 33 healthy anonymous donors (Etablissement Français du Sang). This experiment was specifically designed to *(i)* evaluate if the differences in gene expression observed in response to heat-killed Y. pestis were similar to those observed in response to live Y. pestis (this was done on a subset of 8 individuals); and *(ii)* evaluate if the *ERAP2* protective allele impacted cytokine secretion levels or intracellular bacteria levels. DNA from 200µl of blood was extracted using PureLink Genomic DNA kits (Invitrogen, Waltham, MA, USA) and allelic discrimination at the rs2248374 SNP was determined by TaqMan qPCR with Applied Biosystems Human *ERAP2* probes. CD14+ monocytes were isolated and were differentiated into macrophages as described above. Fully virulent *Y. pestis* (strain CO92) was grown on LB Agar supplemented with 0.0002% porcine hemin and cells ($25x10^4$/well) were infected with live bacteria at a multiplicity of infection of 10 bacteria/cell. Supernatants (for cytokine profiles) or cells (for RNA-sequencing) were collected after 5 hours of incubation without washing cells, or cells were washed with PBS after 1h and incubated in fresh medium for a further 23 hours in the presence of gentamycin (40µg/ml). Samples were sterilized by centrifugation over 0.22µm Spin-X filters (Costar) before deep-freezing. Levels of ten cytokines (IL-1ß, IL-1RA, TNFα, IL-6, IL-10, G-CSF, GM-CSF, CCL2/MCP-1, CCL3/MIP-1α, CXCL8/IL-8) were determined using Magnetic Luminex performance assays from R&DSystems with Curiox DropArray plates and LT washing station for 25 individuals. Data were acquired on a Luminex LX-200 reader. For 8 individuals, we performed RNA-sequencing on infected and non-infected cells. Sequencing data were mapped and analyzed for an effect of *Y. pestis* stimulation as described above in **section 1.f.ii**.

For all 33 individuals, macrophages (6 biological replicates per condition) were infected with *Y. pestis* (MOI of 10 bacteria/cell) with gentle centrifugation (300xG, 5 min) to enhance contacts. After one hour, samples were washed twice with PBS, and new medium containing Gentamycin (40 µg/ml) was added to kill extracellular bacteria. Then, another hour or 24h, macrophages were washed twice with PBS, lysed with sterile water and scratched from plastic. The lysate was serially diluted, and samples immediately cultured at 28°C on LB agar + 0.002% Hemin to count the number of colony forming units (CFU) of bacteria. Intracellular bacterial killing after 24h was calculated as 100x(CFU at 2h – CFU at 24h)/(CFU at 2 h). This strategy is the most used strategy to analyze intracellular survival or killing of bacteria inside macrophages[78]. One sample (Sample 265) was excluded from the CFU data analyses because the cell cultures were

noticeably contaminated. For the cytokine data we quantile-normalized the data across individuals to minimize the risk of spurious association due to outliers, and to ensure that the data was normally distributed. We then examined the association between SNP genotypes (coded as the number of protective “C” alleles at rs2549794: 0, 1, or 2) and cytokine or CFU levels by using a linear regression model in which phenotype was regressed against genotype. This model returns a single *p* value that reflects the relationship between genotype (i.e., the number of protective “C” alleles) and cytokine/CFU levels.

vi. Single-cell RNA-sequencing of PBMCs before and after infection with *Y. pestis.*

Experimental design

To consider the response to *Y. pestis* infection across a broader array of immune cell types, we also conducted single-cell RNA-sequencing of peripheral blood mononuclear cells (PBMCs) for 10 individuals of European ancestry, five of whom were homozygous for the reference allele at rs2549754 and five of whom were homozygous for the alternate allele. These individuals were previously used to study the response to viral infections and the full details of sample providence can be found in previous work[79]. Briefly, samples were obtained from BioIVT along with a signed, written consent from healthy participants at a collection site in Miami, Florida (United States) utilizing a standard protocol with a sodium heparin anticoagulant. PBMCs were extracted from whole blood using a density gradient, washed with HBSS, reconstituted in CryoStor CS10 to a concentration of 10 million cells/mL, and cryopreserved in sets of 5-10 million cells a vial.

PBMCs were thawed approximately 14 hours prior to infection and cultured overnight in RPMI 1640 supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, and 10 µg/mL gentamycin. Infection experiments were balanced for individual genotypes. The morning of the experiment 10 individuals were collected, counted, plated at 1 million cells per ml per well, and infected with *Y. pestis* (strain CO92), at an MOI of ten. Following 5 hours of infection, cells were collected, washed, and prepared for single-cell capture using the 10X workflow. Multiplexed cell pools of 6 samples were used as input for the single cell captures, and for each pool 10,000 cells were targeted using the Chromium Single Cell 3’ Reagent (v3 chemistry) kit (10X Genomics #CG000183 Rev A). Post Gel Bead-in-Emulsion (GEM) generation, the reverse transcription (RT) reaction was performed in a thermal cycler as described (53°C for 45 min, 85°C 537 for 5 min), and post-RT products were temporarily stored at -20°C. After all captures were complete, post-RT reaction cleanup, cDNA amplification, and sequencing library preparation were performed as described in the Single Cell 3’ Reagents Kits v3 User Guide (10X Genomics #CG000183 Rev A). Briefly, cDNA was cleaned with DynaBeads MyOne SILANE beads (ThermoFisher Scientific) and amplified in a thermal cycler using the following program: 98°C for 3 min, 11 cycles x 98°C for 15 s, 63°C for 20 s, 72°C for 1 min, and 72°C 1 min. After cleanup with the SPRIselect reagent kit (Beckman Coulter), the libraries were constructed by performing the following steps: fragmentation, end-repair, A-tailing, SPRIselect cleanup, adaptor ligation, SPRIselect cleanup, sample index PCR (98°C for 45 s, 13 or 14 cycles x 98°C for 20 s, 54°C for 30 s, 72°C for 20 s, and 72°C 1 min), and SPRIselect size selection. Libraries were sequenced 100 base pair paired-end (R1: 30 cycles, I1: 10 cycles, R2:

85 cycles) on an Illumina NovaSeq. Cell pools contained infected or uninfected samples, to limit the possibility for cross-contamination.

Mapping, demultiplexing, cell filtering, and UMAP analysis

FASTQ files were mapped to a custom reference containing GrCh37 and the *Y. pestis* reference genome (downloaded from NCBI, created using cellranger mkref) using the cellranger (v3.0.2) (10X Genomics) count function. Souporcell (v2.0, Singularity v3.4.0)[80] in --skip_remap mode was used to demultiplex cells based on genotypes from a common variants file (1000GP samples filtered to SNPs with >= 2% allele frequency in the population, downloaded from https://github.com/wheaton5/souporcell). Briefly, souporcell clusters cells based on cell allele counts in common variants, assigning all cells with similar allele counts to a single cluster corresponding to one individual, while also estimating singlet/doublet/negative status for that cell. Hierarchical clustering of the true genotypes known for each individual (obtained from low-pass whole-genome-sequencing) and the cluster genotypes estimated from souporcell was used to assign individual IDs to souporcell cell clusters. All 10 individuals were successfully assigned to a single cluster.

After demultiplexing cells into samples, Seurat (v4.0.4, R v4.1.1)[81] was used to perform quality control filtering of cells. Cells were considered “high quality” and retained for downstream analysis if they had: 1) a “singlet” status called by souporcell, 2) between 200 – 2500 genes detected (nFeature_RNA), and 3) a mitochondrial reads percentage < 20%. In total, we captured 14,174 high quality cells (range of cells per individual: 780 to 2030).

We split the cells by infection status (uninfected or *Y. pestis*) and individual, then ran SCTransform (Seurat v4.0.4) to normalize and scale the UMI counts within condition. In this step, we simultaneously regressed out variables corresponding to experiment batch and percent mitochondrial reads per cell. Cells were assigned to cell types using Seurat’s Multimodal Reference Mapping workflow. Briefly, we applied SCTransform, FindTransferAnchors, and MaqQuery to assign each cell a predicted cell type based on published data from PBMCs[82]. Data were then integrated on infection status and individual using reciprocal PCA on SCTransformed data (https://satijalab.org/seurat/articles/integration_rpca.html) prior to dimensionality reduction (RunUMAP in Seurat, with the first 20 dimensions). Reassuringly, predicted cell types reflected major structure in the UMAP plot.

Differential expression analysis

To model expression data for each cell type, we converted raw counts per cell into pseudobulk mean counts for each cell type and sample (as in [79]). Within each cell type for each sample, raw counts were summed across all cells assigned to that cell type and sample for each gene using the sparse_Sums function in textTinyR (v1.1.3) (https://cran.rproject.org/web/packages/textTinyR/textTinyR.pdf), yielding an n x m expression matrix, where n is the number of samples (n = 20) and m is the number of genes detected. For each cell type, we removed lowly expressed genes (cpm < 1) and then voom-normalized the count data to control for library. Using lmFit (R *limma* package v3.48.3), we modeled effects of stimulation and an effect of rs2549754 genotype within each condition, while controlling for the

number of cells. False discovery rates were estimated using the q-value approach[83] relative to an empirical null distribution constructed by randomly assigning which sample condition within each individual and by randomizing genotypes between individuals (n=10 iterations). Results for stimulation of each gene and each cell type are available on in **Table S6**.

Trans-effects of ERAP2 haplotypes

In each cell type, we tested for an associated between rs2549754 genotype and the magnitude of the fold-change response to *Y. pestis* of all genes across the genome (i.e., *ERAP2* genotype x stimulation (baseline versus *Y. pestis*-stimulated) interaction effect). We find no evidence for widespread *trans*-effects of *ERAP2* haplotype on the gene expression levels of other genes across the genome after correction for multiple testing (**Table S10**). However, we urge caution in interpreting this results as this particular data set is severely underpowered to identify interaction effects of this type. Moreover, we are only able to capture a snapshot of the immune response *in vitro* (i.e., measures of gene expression 5 hours post-infection). It is possible that differential gene regulation between the protective and deleterious haplotypes may occur at later time points, or in an *in vivo* setting where a greater diversity cell types interact with each other. Future studies with larger sample sizes and a dense time-course response will be needed to evaluate if *ERAP2* haplotypes affect the transcriptional responses to *Y. pestis* of other genes in the genome.

g. Hidden-Markov Model-based inference of selection coefficients

To estimate the selective advantage of rs2549794, we assumed the allelic effects are additive and we define *s* (the selection coefficient) as the expected number of offspring of an individual with genotype aa, Aa, or AA is 1, 1 + *s*/2, or 1 + *s*, respectively (where a is the ancestral allele and A is the derived variant).

To infer this selection coefficient *s* from the temporal data, we used a common Hidden Markov Model (HMM) framework[24]. Briefly, in this framework, the underlying population allele frequency is treated as the hidden state, and the observed genetic variation at the respective times present the emissions, which are binomially distributed, given the underlying allele frequency. To compute the likelihood of the data, the framework integrates over all possible hidden allele frequency trajectories under the Wright-Fisher model. Since the selection coefficient *s* affects the likelihood of these trajectories (larger *s* makes an increase in frequency more likely), this allows us to infer *s* using a maximum likelihood approach. Here, we use an implementation of this HMM framework available at https://github.com/steinrue/diplo_locus which allows efficient computation of the likelihood by discretizing the allele frequency space and using a Gaussian approximation to compute the transition probabilities of the Wright-Fisher model[84].

With this implementation, we computed joint likelihoods of observing samples in both populations across time, which yields maximum-likelihood point estimates of *s*. In addition, we applied this implementation to simulated data to provide confidence intervals and assess the statistical power of our inference scheme. Because the sampling times across both populations span no more than 20 generations, we set the mutation rate to zero for convenience. All scripts

to perform the analyses are available at https://github.com/TaurVil/VilgalysKlunk_yersinia_pestis/tree/main/part1_aDNA/infer_selection_coefficients.

i. Maximum likelihood estimation of selection coefficients from data

We follow the analysis in the main text and group the samples into before, during, and after the Black Death pandemic. We use the expected derived allele frequency in each group for each candidate SNP, computed as $\mathcal{P}_{Aa} + 2\mathcal{P}_{AA}$ for each individual and the mean across individuals in each group, where $\mathcal{P}$ is the genotype likelihood per individual inferred in **section 1.c**.

For samples from London (blue lines and boxes in **Fig ED10**), we assume an effective population size of 5,000 and that the individuals before, during, and after the pandemic are sampled in generation 1, 9, and 16, respectively. Considering that multiple generations of people have been impacted by the circulating disease, we assume selection acted during generation 6, 7, and 8, the three generations preceding the mid-pandemic sample. Because the mid-pandemic samples in London are known to be from individuals who died from the Black Death, we assume that the lineages leading to mid-pandemic samples experience opposite selective pressure from those who survived the pandemic (the post-pandemic samples).

For samples from Denmark (brown lines and orange boxes in **Fig ED10**), we only consider samples taken before and after the pandemic. We assume an effective population size of 5,000, that the samples were taken at generation 0 and 17, and that selection occurred during generation 6, 7, and 8 (same as in London). The generation times were chosen to represent approximate mid-points of the pre- and post-Black Death samples. For each candidate SNP, we computed log-likelihoods on a grid of s values in the interval [-0.9,0.9], with a higher density of grid-points around zero, for three different scenarios (see **Fig ED10**, straight lines in the lower half): before-to-during in London, before-to-after in London, and before-to-after in Denmark. To reduce the number of free parameters, we assumed the selection coefficient s in all three scenarios has the same absolute value, but the coefficient for before-to-during in London has an inverted sign. The rationale for this is that the samples before-to-after represent the survivors, whereas before-to-during samples in London represent the ones who did not survive. The final composite log-likelihood for each locus is then obtained by summing the log-likelihoods from the three scenarios.

To estimate $\hat{s}_{\mathrm{MLE}}$, we interpolate the composite log-likelihood at each locus as a function of s using cubic splines (with the scipy.interpolate module in Python3), and infer the maximum in the [-0.9,0.9] interval (with the scipy.optimize module) from this interpolated function. The log-likelihood surfaces and the resulting maximum-likelihood estimates (MLEs), $\hat{s}_{\mathrm{MLE}}$, are depicted in **Fig S3** for the lead SNP. We also performed the analysis with an effective population size of 10,000 in both populations, but the results did not change substantially.

ii. Simulation study to assess confidence in selection coefficients

To assess the statistical properties of our estimation procedure, we simulated allele frequency trajectories under the Wright-Fisher model, and simulated samples at the respective times given the underlying allele frequency. In the simulations of the London population, the three sampling times are set to be generation 1, 9, and 16. In simulations for Denmark, sampling times are set as generations 0 and 17. We drew a number of individuals from the population equal to the number of individuals we sampled from the population. The effective population sizes for London and Denmark populations are set to be 5,000, and the mutation rate is set to zero. As in the likelihood-computations, selection was assumed to act in generations 6, 7, and 8.

To reflect that the mid-pandemic samples from London were taken from people who died of the disease, we simulate this set of samples by assuming a parallel population where the allele frequency experiences reverse changes during the pandemic. In particular, for each replicate, we track the allele frequency throughout the period of selection (pandemic), take the logit-transformed difference between the start and end, and deduct it from the logit-transformed pre-pandemic frequency. The use of logit here ensures that all values are bounded within (0,1). This then gives the frequency in the parallel population who died of Black Death, from which mid-pandemic samples are generated. To reflect uncertainty of the initial frequency in the empirical data, at the potentially selected SNP, we re-sample individual genotypes for the pre-pandemic samples based upon the calculated, maximum likelihood allele frequency and use the resulting allele frequencies as the initial frequencies for the simulated replicates.

We simulate 5,000 replicates for a grid of *s*-values ranging from -0.35 to 0.5 with increments of 0.05, and then applied the same procedure as described in the previous section to each replicate to obtain $\hat{s}_{\mathrm{MLE}}$. **Fig S4** shows the empirical distribution of these estimates, highlighting the mean, median, and 2.5-th and 97.5-th percentiles. We observed a slight bias, albeit very small, in the estimates. Thus, we performed an empirical bootstrap bias-correction of the estimates as follows: We interpolate the values of the median of the empirical $\hat{s}_{\mathrm{MLE}}$ distributions as functions of true *s* value simulated. With these interpolated functions, we use the $\hat{s}_{\mathrm{MLE}}$ directly computed from the likelihood for the four SNPs to identify which underlying value of *s* would result in such an estimate. We denote the latter bootstrap corrected MLE by $\tilde{s}_{\mathrm{MLE}}$.

To provide confidence intervals, for rs2549794, we simulated 5,000 replicates using the bootstrap-corrected $\tilde{s}_{\mathrm{MLE}}$ as the underlying true selection coefficient and re-sampled initial frequencies, as before. We then adjust the MLEs from all replicates by the same amount that $\hat{s}_{\mathrm{MLE}}$ was adjusted to obtain $\tilde{s}_{\mathrm{MLE}}$. After this adjustment, we report the 2.5-th and 97.5-th percentiles of the empirical distribution of the MLEs as confidence intervals for our estimates in **Fig ED2A** and **Table S5**. We also simulated sets of 5,000 replicates with the same initial conditions but under $s = 0$, applied the same inference procedure to assess the power, and report Receiver Operating Characteristics (ROC) curves and Precision-Recall curves in **Fig ED2B-C**. Note that the power is affected by the selection coefficient and the initial frequencies, and would therefore vary between SNPs.

To provide further evidence for selection acting at the focal SNP, we applied the same inference procedure that was used to obtain $\hat{s}_{\mathrm{MLE}}$ for the focal SNPs to all SNPs in the neutral regions (described in **section 1.b**), since these reflect the empirical neutral distribution. For each SNP, we then computed the log-likelihood-ratio by subtracting the log-likelihood at $s = 0$ from value of the log-likelihood at $\hat{s}_{\mathrm{MLE}}$. We report the percentile of the log-likelihood-ratio for the four focal SNPs in this empirical neutral distribution in **Table S5**, with the rationale that low percentiles indicate strong evidence that non-neutral processes were driving the allele frequency change at the respective SNP. We note for neutral SNPs where the ancestral allele is unknown, we polarized for the reference allele. Since this only changes the sign of the estimated s, the polarization does not affect the value of the log-likelihood-ratio.

Last, we note that $1 + s$ yields the expected number of viable offspring of an individual that is homozygous for the derived allele at the respective locus. Thus, we can loosely interpret $1 + s$ as reflecting the chance of an individual to survive the Black Death pandemic. With this interpretation, we can quantify the increase in chance of survival of individuals homozygous for the derived allele over individuals homozygous for to ancestral allele by $p_{surv} \coloneqq (1 + s)/1$.

h. Supplementary References

52. Schwarz, C., et al. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res* **37**, 3215-3229 (2009).
53. Dabney, J., et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *P Natl Acad Sci USA* **110**, 15758-15763 (2013).
54. Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res* **27**, 1230-1237 (2017).
55. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *CSH Protocols* **2010.6**, pdb-prot5448 (2010).
56. Kircher, M., Sawyer, S., & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, e3-e3 (2012).
57. Renaud, G., Stenzel, U., & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* **42**, e141-e141 (2014).
58. Li, H. & Durbin, R.. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1 (2009).
59. Li, H., et al., The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
60. Graffelman, J., Jain, D., & Weir, B. A genome-wide study of Hardy–Weinberg equilibrium with next generation sequence data. *Hum Genet* **136**, 727-741 (2017).
61. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., & Orlando, L. mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684 (2013).
62. Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res* **25**, 918-925 (2015).

63. Wickham, H. & Chang, W. Package 'ggplot2'. Create Elegant Data Visualisations Using the Grammar of Graphics. Version 2.1 (2016).
64. Allen Ancient DNA Resource https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data, version 50.0
65. McKenna, A., et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
66. Wall, J. D., et al. Genome-wide ancestry and divergence patterns from low-coverage sequencing data reveal a complex history of admixture in wild baboons. *Mol Ecol* **25**, 3469-3483 (2016).
67. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370 (1984).
68. Whitlock, M. C. & Lotterhos, K. E. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST}. *Am Nat* **186**, S24-S36 (2015).
69. Matthey‐Doret, R. & Whitlock, M.C., Background selection and FST: consequences for detecting local adaptation. *Mol Ecol* **28**, 3902-3914 (2019).
70. Keinan, A. & Reich, D. Human population differentiation is strongly correlated with local recombination rate. *PLOS Genet*, **6**, e1000886 (2010).
71. Kong, A., et al. A high-resolution recombination map of the human genome. *Nat Genet* **31**, 241–247 (2002).
72. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLOS Genet* **5**, e1000471 (2009).
73. Haller, B.C. & Messer, P.W. SLiM 2: flexible, interactive forward genetic simulations. *Mol Biol Evol* **34**, 230-240 (2017).
74. Messer, P.W. SLiM: Simulating Evolution with Selection and Linkage. *Genetics* **194**, 1037-1039 (2013).
75. Hsu, F., et al. The UCSC known genes. *Bioinformatics*, **22**, 1036-1046 (2006).
76. Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
77. Li, B., & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1-16 (2011).
78. Kaneko, M., Emoto, Y. & Emoto, M. A simple, reproducible, inexpensive, yet old-fashioned method for determining phagocytic and bactericidal activities of macrophages. *Yonsei Med J* **57**, 283-290 (2016).
79. Randolph, H. E., et al. Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science* **374**, 1127-1133 (2021).
80. Heaton, H., et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods* **17**, 615-620 (2020).
81. Stuart, T., et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902 (2019).
82. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* (2021).
83. Storey, J.D. & Tibshirani, R. Statistical significance for genome wide studies. *P Natl Acad Sci USA* **100**, 9440-9445 (2003).
84. Mathieson, I. & McVean, G. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* **193**, 973–984 (2013).

2. Supplementary Figures

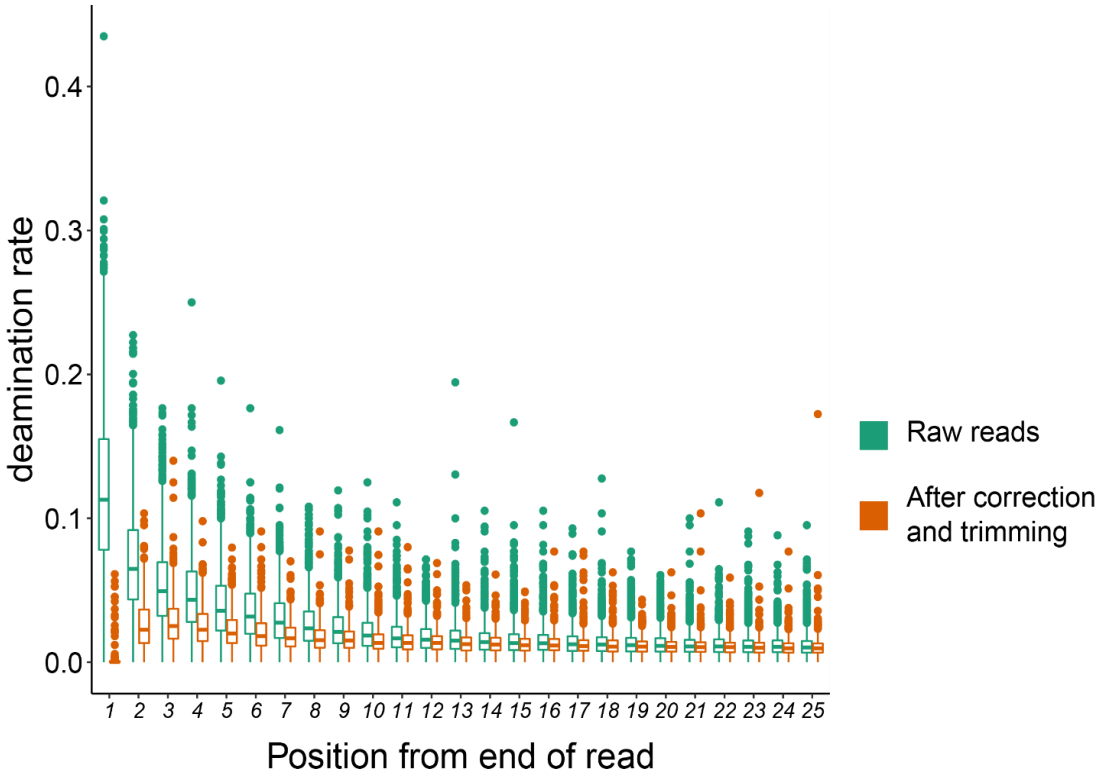

Figure S1: Correcting for ancient genomic damage. Damage and degradation in ancient DNA samples result in an increased deamination rate near the end of reads (green). We correct for this source of DNA damage by using *mapDamage*[60] and trimming the terminal 4 bases of each read, which eliminates the bias due to DNA damage (orange). After correcting and trimming reads, error rates are lower for the first base pair because *mapDamage* is more liberal in masking possible DNA damage near the ends of reads.

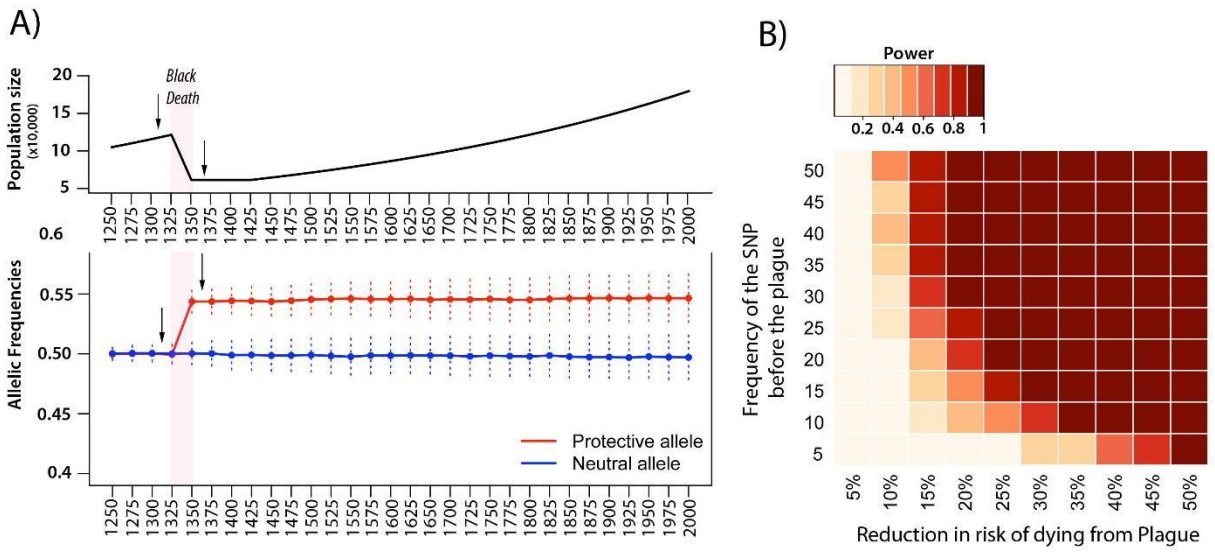

Figure S2: Power to identify protective alleles against Black Death. (A) Simulations of the expected changes in allele frequencies across generations for a protective allele (red) that decreases the risk of mortality from plague by 30% in homozygous individuals and 15% in heterozygous individuals (i.e., additive effect) as compared to a neutral allele (blue), with a starting allele frequency of the protective or neutral allele of 50%. Dashed lines - standard deviation for 100 simulated alleles. Under the simulated scenario the protective allele increases by ~5% in a single generation (in red), which is significantly more than what would be expected by a neutral allele (in blue). **(B)** Power to detect protective alleles as a function of their frequency prior to *Black Death* and their level of protection against the disease (the figure shows the level of protection for homozygous individuals) by calculating *Fst* values between pre- and post-Black Death populations. Power was calculated by comparing *Fst* values for a simulated set of 100 protective alleles between the pre- and post-Black Death populations with respect to *Fst* values observed among 1,000 simulated neutral alleles sampled from the same populations.

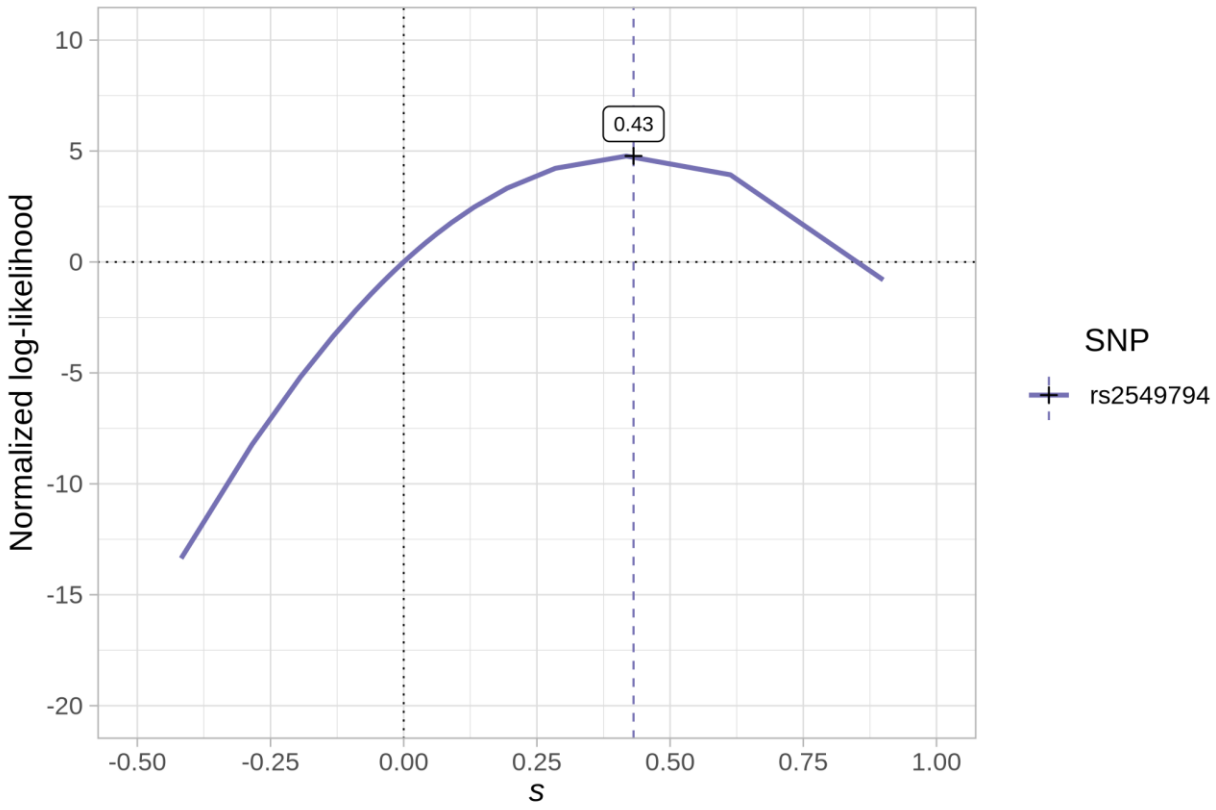

Figure S3: Log-likelihood as a function of the selection coefficient *s* for rs2549794. Labeled numbers show the approximate values of respective interpolated $\hat{s}_{\mathrm{MLE}}$. The log-likelihood is centered to 0 for *s* = 0.

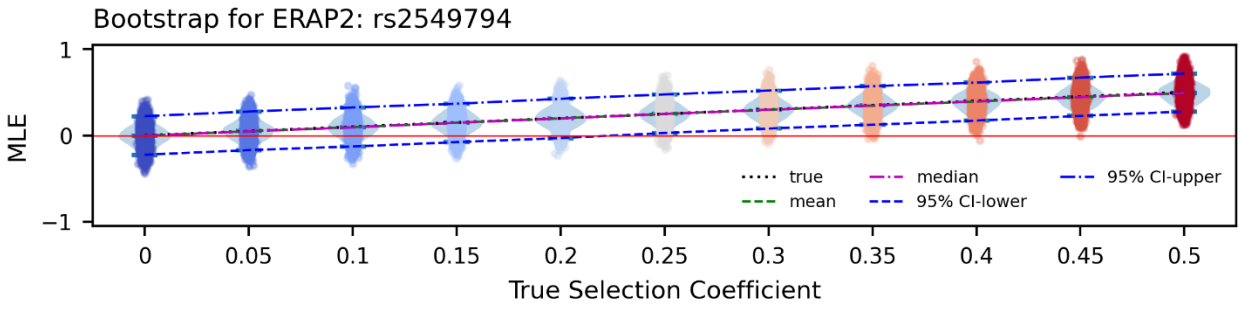

Figure S4: Distributions of $\hat{s}_{\mathrm{MLE}}$ for simulated data starting with the pre-pandemic frequencies of rs2549794. X-axes label the values of the true s used in the simulations, and y-axes show $\hat{s}_{\mathrm{MLE}}$. The black dotted lines are the y = x reference lines. The green dashed lines and red dash-dotted lines indicate the means and medians of the $\hat{s}_{\mathrm{MLE}}$. The blue dashed and dot-dashed lines represent the 2.5-th and 97.5-th percentiles of the $\hat{s}_{\mathrm{MLE}}$ distributions.