

UCSF

UC San Francisco Previously Published Works

Title

Evolution of immune genes is associated with the Black Death

Permalink

<https://escholarship.org/uc/item/28c118cj>

Journal

Nature, 611(7935)

ISSN

0028-0836

Authors

Klunk, Jennifer
Vilgalys, Tauras P
Demeure, Christian E
et al.

Publication Date

2022-11-10

DOI

10.1038/s41586-022-05349-x

Peer reviewed

Evolution of immune genes is associated with the Black Death

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

Received: 12 January 2022

Accepted: 14 September 2022

Published online: 19 October 2022

 Check for updates

Jennifer Klunk^{1,2,26}, Tauras P. Vilgalys^{3,26}, Christian E. Demeure⁴, Xiaoheng Cheng⁵, Mari Shiratori³, Julien Madej⁴, Rémi Beau⁴, Derek Elli⁶, Maria I. Patino³, Rebecca Redfern⁷, Sharon N. DeWitte⁸, Julia A. Gamble⁹, Jesper L. Boldsen¹⁰, Ann Carmichael¹¹, Nükhet Varlik¹², Katherine Eaton¹, Jean-Christophe Grenier¹³, G. Brian Golding¹, Alison Devault², Jean-Marie Rouillard^{2,14}, Vania Yotova¹⁵, Renata Sindeaux¹⁵, Chun Jimmie Ye^{16,17}, Matin Bikaran^{16,17}, Anne Dumaine³, Jessica F. Brinkworth^{18,19}, Dominique Missiakas⁶, Guy A. Rouleau²⁰, Matthias Steinrücken^{5,21}, Javier Pizarro-Cerdá⁴, Hendrik N. Poinar^{1,22,23,27} & Luis B. Barreiro^{3,21,24,25,27}✉

Infectious diseases are among the strongest selective pressures driving human evolution^{1,2}. This includes the single greatest mortality event in recorded history, the first outbreak of the second pandemic of plague, commonly called the Black Death, which was caused by the bacterium *Yersinia pestis*³. This pandemic devastated Afro-Eurasia, killing up to 30–50% of the population⁴. To identify loci that may have been under selection during the Black Death, we characterized genetic variation around immune-related genes from 206 ancient DNA extracts, stemming from two different European populations before, during and after the Black Death. Immune loci are strongly enriched for highly differentiated sites relative to a set of non-immune loci, suggesting positive selection. We identify 245 variants that are highly differentiated within the London dataset, four of which were replicated in an independent cohort from Denmark, and represent the strongest candidates for positive selection. The selected allele for one of these variants, rs2549794, is associated with the production of a full-length (versus truncated) *ERAP2* transcript, variation in cytokine response to *Y. pestis* and increased ability to control intracellular *Y. pestis* in macrophages. Finally, we show that protective variants overlap with alleles that are today associated with increased susceptibility to autoimmune diseases, providing empirical evidence for the role played by past pandemics in shaping present-day susceptibility to disease.

Infectious diseases have presented one of the strongest selective pressures in the evolution of humans and other animals^{1,2}. Not surprisingly, many candidates for population-specific positive selection in humans involve immune response genes, consistent with the hypothesis that exposure to new and/or re-emerging pathogens has driven adaptation^{5,6}. However, it is challenging to connect signatures of natural selection with their causative pathogens unless the underlying loci are still associated with susceptibility to the same pathogen in modern populations^{7,8}. Clarifying the dynamics that have shaped the human immune

system is key to understanding how historical diseases contributed to disease susceptibility today.

We sought to identify signatures of natural selection in Europeans imposed by *Yersinia pestis*, the bacterium responsible for bubonic plague³. The first recorded plague pandemic began with the Plague of Justinian in AD 541 (refs. ^{9,10}). Nearly 800 years later, the Black Death (1346–1350) marked the beginning of the second pandemic of plague, which spread throughout Europe, the Middle East and Northern Africa, reducing the population by up to 30–50%^{4,11}. With no recent exposure to

¹McMaster Ancient DNA Centre, Departments of Anthropology, Biology and Biochemistry, McMaster University, Hamilton, Ontario, Canada. ²Daielc Arbor Biosciences, Ann Arbor, MI, USA.

³Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. ⁴Yersinia Research Unit, Institut Pasteur, Paris, France. ⁵Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. ⁶Department of Microbiology, Ricketts Laboratory, University of Chicago, Lemont, IL, USA. ⁷Centre for Human Bioarchaeology, Museum of London, London, UK. ⁸Department of Anthropology, University of South Carolina, Columbia, SC, USA. ⁹Department of Anthropology, University of Manitoba, Winnipeg, Manitoba, Canada. ¹⁰Department of Forensic Medicine, Unit of Anthropology (ADBOU), University of Southern Denmark, Odense S, Denmark. ¹¹History Department, Indiana University, Bloomington, IN, USA. ¹²Department of History, Rutgers University, Newark, NJ, USA. ¹³Montreal Heart Institute, Faculty of Medicine, Université de Montréal, Montréal, Quebec, Canada. ¹⁴Department of Chemical Engineering, University of Michigan – Ann Arbor, Ann Arbor, MI, USA. ¹⁵Centre Hospitalier Universitaire Sainte-Justine, Montréal, Quebec, Canada. ¹⁶Division of Rheumatology, Department of Medicine, University of California, San Francisco, CA, USA. ¹⁷Institute for Human Genetics, University of California, San Francisco, CA, USA. ¹⁸Department of Anthropology, University of Illinois Urbana-Champaign, Urbana, IL, USA. ¹⁹Carl R Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ²⁰Montreal Neurological Institute-Hospital, McGill University, Montréal, Quebec, Canada. ²¹Department of Human Genetics, University of Chicago, Chicago, IL, USA. ²²Michael G. DeGroot Institute of Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada. ²³Humans and the Microbiome Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada. ²⁴Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA. ²⁵Committee on Immunology, University of Chicago, Chicago, IL, USA. ²⁶These authors contributed equally: Jennifer Klunk, Tauras P. Vilgalys. ²⁷These authors jointly supervised this work: Hendrik N. Poinar, Luis B. Barreiro. ✉e-mail: poinarh@mcmaster.ca; lbarreiro@uchicago.edu

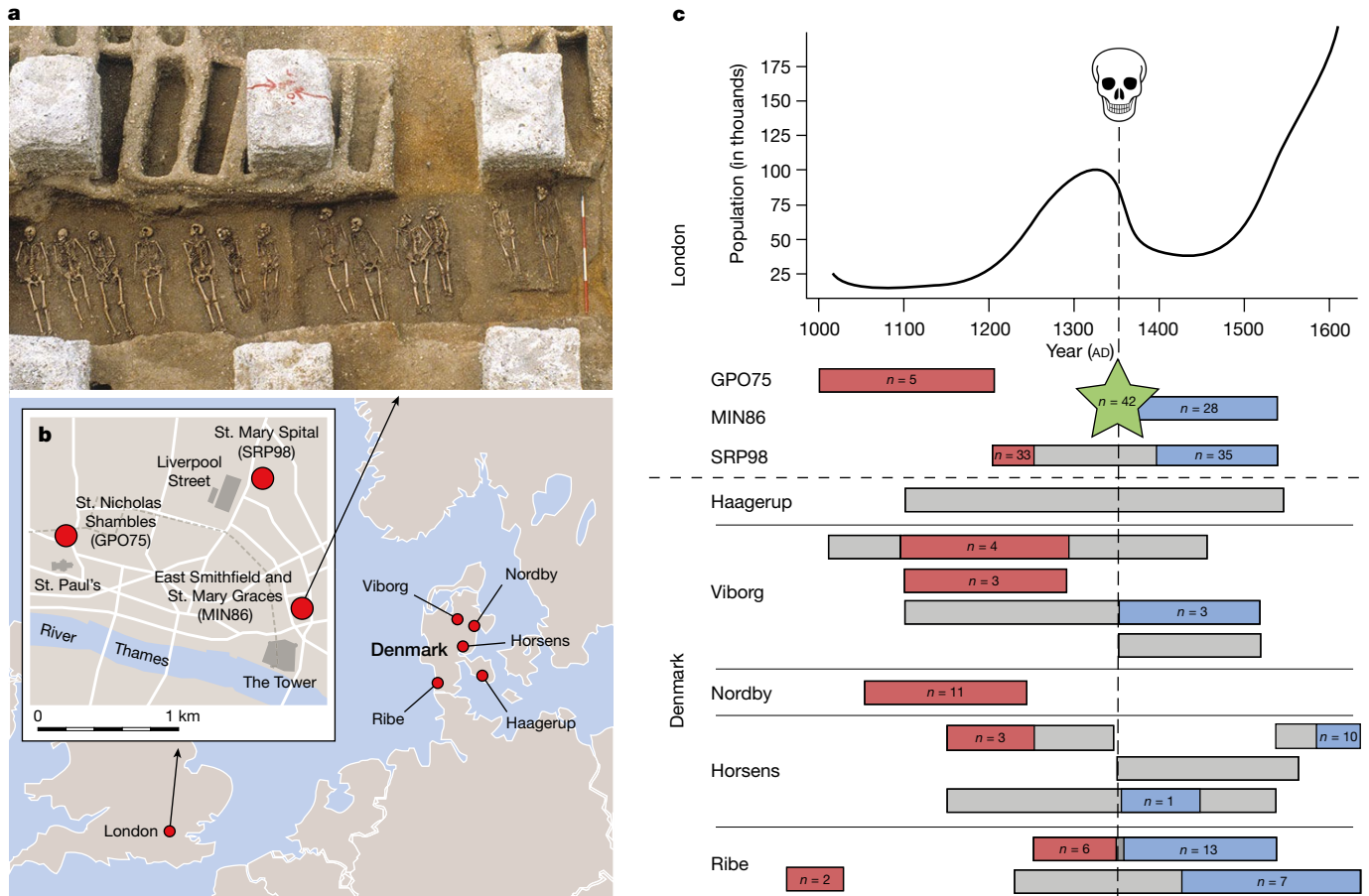


Fig. 1 | East Smithfield mass burial and sample locations, along with date ranges and final sample numbers used for the present study. a, East Smithfield Black Death mass burial site from 1348–1349 (reproduced with permission of the Museum of London Archaeology, copyright MOLA). **b**, Site locations and archaeological site codes (in parentheses) for samples from London (inset, after ref.⁴⁸, Museum of London Archaeology) and from across Denmark. **c**, Top,

plague, Europeans living through the Black Death probably represented immunologically naive populations with little to no prior adaptation to *Y. pestis*. The high mortality rate suggests that genetic variants that conferred protection against *Y. pestis* infection might have been under strong selection during this time. Indeed, the nearly decadal plague outbreaks over the subsequent four hundred years of the second pandemic in Europe often (but not always) were associated with reduced mortality rates^{11,12}, which could have been due to pathogen evolution or changing cultural practices, but potentially also linked to human genetic adaptation to *Y. pestis*.

Positive selection on immune genes

Genomic targets of selection imposed by *Y. pestis* during the Black Death, if present, have remained elusive^{13–15}. To better identify such loci, we characterized genetic variation from ancient DNA extracts derived from individuals who died shortly before, during or soon after the Black Death in London and across Denmark. This unique sampling design differentiates, to the greatest extent possible, signatures due to *Y. pestis* from those associated with other infectious diseases or other selective processes (although we cannot exclude these entirely). From London, individuals were sampled from three cemeteries close to one another, tightly dated by radiocarbon, stratigraphy and historical records to before, during and after the Black Death (Fig. 1, Supplementary Table 1 and Supplementary Methods). From Denmark, individuals

population size estimates for London for around six centuries (data are from refs.^{13,49,50}) (Supplementary Table 1). Bottom, site locations with associated date ranges. Coloured boxes indicate date range for samples, numbers in boxes indicate samples meeting all criteria for inclusion in final analyses (see main text and Supplementary Information). Number in green stars stems from East Smithfield and the dashed line refers to the time of the Black Death.

were sampled from five localities, geographically spread across the country, which were dated via archaeological means (such as burial arm positions), stratigraphy and historical records. We grouped all individuals into those that lived pre-Black Death (London: around AD 1000–1250, Denmark: around AD 850 to around AD 1350) and post-Black Death (London: AD 1350–1539, Denmark: around AD 1350 to around AD 1800). Within London, we also included individuals buried in the plague cemetery, East Smithfield, all of whom died during a two-year window of the Black Death between 1348 and 1349 (ref.¹⁶). Analysis of the mitogenomic diversity from these individuals identifies solely European mitogenomic haplotypes, avoiding a possible confound between natural selection and population replacement from non-European sources¹⁷.

In total we screened 516 samples ($n = 318$ from London; $n = 198$ from across Denmark) for the presence of human DNA using a modified polymerase chain reaction (PCR) assay for the single copy nuclear *c-myc* gene^{17,18} and identified 360 with sufficient endogenous DNA content for downstream enrichment and sequencing of additional nuclear loci (Supplementary Methods). As many of our samples were poorly preserved and had low endogenous DNA content, we used hybridization capture to enrich for and sequence 356 immune-related genes, 496 genome-wide association study (GWAS) loci previously associated with immune disorders and 250 putatively neutral regions (1.5 kb each), as defined by Gronau and colleagues¹⁹ (Supplementary Table 2; Supplementary Methods). The targeted immune genes were manually curated on the basis of their role in immune-related processes, and

include innate immune receptors, key immune transcription factors, cytokines and chemokines, and other effector molecules (Supplementary Table 3). To ensure that deamination and other forms of ancient DNA damage did not lead to spurious genotype calls, we trimmed 4 base pairs (bp) from the start and end of each sequencing read (Supplementary Fig. 1) and excluded all singleton variants ($n = 106,757$). Our final dataset contained 33,110 biallelic variants within the targeted regions (2,669 near GWAS loci, 19,972 in immune genes and 10,469 in putatively neutral regions), with a mean coverage of $4.6\times$ reads per site per individual (see Supplementary Table 1 for per-individual coverage). We further filtered our results by excluding samples with missing genotype calls at more than 50% of those sites (retaining $n = 206$ individuals, Fig. 1c) and excluding variants with genotype calls for less than 10 individuals per time period and population. Using genotype likelihoods, we then calculated the minor allele frequency (MAF) per population at each time point. Finally, we retained only sites with a mean MAF (averaged across London and Denmark) greater than 5% ($n = 22,868$ sites), as our power to detect selection for variants below 5% is very low (Supplementary Fig. 2).

To detect alleles that may have conferred protection from, or susceptibility to, *Y. pestis*, we searched within candidate regions (immune genes and GWAS loci) for variants that exhibit unexpectedly large changes in allele frequency between pre- and post-Black Death samples. Specifically, we identified alleles for which the degree of differentiation (F_{ST}) was larger than expected by chance, when compared to variants in the putatively neutral genetic regions sampled from the same populations. We used the larger sample set from London as our discovery cohort. Burials in London were also more precisely dated and better geographically controlled than those from Denmark, improving our relative ability to detect selection in the cohort from London (Supplementary Methods). We found an enrichment of highly differentiated variants for all frequency bins with MAF greater than 10%, relative to a null expectation established using our neutral loci (Fig. 2a). Across these variants, differentiation at immune loci exceeded the 99th percentile of neutral variants at $2.4\times$ the rate expected by chance (binomial test $P = 7.89 \times 10^{-12}$). For variants with an MAF greater than 30%, this enrichment was even more pronounced ($3.9\times$ the rate expected by chance; binomial test $P = 1.16 \times 10^{-14}$), probably due to increased power (Supplementary Fig. 2). Simulations show that differences in recombination rate and background selection between neutral and candidate loci are insufficient to explain the observed enrichments (Extended Data Fig. 1). To further validate the signatures of selection we observed among immune loci from our London sample, we performed the same analyses using the allele frequencies estimated from our Danish cohort. These samples were also enriched for highly differentiated sites relative to the expectation from neutral loci ($1.6\times$ the rate expected by chance, binomial test $P = 9.21 \times 10^{-4}$; Fig. 2b), further supporting evidence for plague-induced selection on immune genes.

To identify specific loci that represent the strongest candidates of selection, we applied a series of stringent criteria that leveraged the time periods and populations in this unique dataset. First, we identified 245 common variants (MAF greater than 10%) that were highly differentiated (F_{ST} greater than 95th percentile defined using neutral sites) when comparing pre- versus post-Black Death samples in London alone (Supplementary Table 4). Next, we reasoned that variants conferring increased susceptibility to, or protection from, *Y. pestis* should show opposing frequency patterns before, during and after the Black Death. Specifically, variants associated with susceptibility should increase in frequency among people who died during the Black Death and should decrease in frequency among individuals sampled post-Black Death (that is, the survivors and/or descendants of the survivors). Conversely, protective variants should show an inverse pattern. Using this reasoning, we narrowed down our list of putatively selected loci from 245 to 35 (Supplementary Table 4). Finally, we asked if these loci were also highly differentiated before and after the Black Death in our Danish replication

cohort (that is, among the top 10% most highly differentiated sites, and in the same direction as seen in London). Four loci met these criteria, representing the strongest candidates for selection (Fig. 2c–g). We calculated the selection coefficient (s) for each of these variants using a Hidden Markov Model (HMM) framework (based on ref.²⁰, Supplementary Methods). Statistical support for non-neutral evolution ($s \neq 0$) among our four candidate loci is strong when compared to that of neutral loci ($P < 0.001$ for each locus; Supplementary Table 5). Despite the large confidence intervals—an inherent limitation when trying to estimate s over a few generations—the absolute values for the point estimate of s range from 0.26 to 0.4, which are among the strongest selective coefficients yet reported in humans, to our knowledge (Extended Data Fig. 2 and Supplementary Table 5).

Functional dissection of candidate loci

None of our top candidate variants overlaps with (nor is in strong linkage disequilibrium with) coding variants, although one, near the *ERAP2* gene, is strongly linked to a variant that affects splicing^{21,22}. Their selective advantage may stem from an impact on gene expression levels, particularly in immune cell types that participate in the host response to *Y. pestis* infection. Macrophages in particular are recruited to sites of infection, where they interact with bacteria and contribute to plague resistance^{23,24}. Macrophages phagocytize *Y. pestis*, but some bacteria survive and spread to the lymph node, where they replicate uncontrollably^{25,26}. To test whether the four candidate loci we identified, or genes near them, are involved in the transcriptional response to *Y. pestis*, we incubated monocyte-derived macrophages from 33 individuals with heat-killed *Y. pestis* and compared their expression profiles to unstimulated control samples using RNA sequencing (Supplementary Methods). As expected, macrophages responded robustly to *Y. pestis*, such that principal component 1 of the gene expression data, which separates baseline versus *Y. pestis* stimulated conditions, explains 56% of the total variance (Extended Data Fig. 3).

Seven genes within 100 kb of our four candidate loci were expressed in this dataset: locus 1 (rs2549794): *ERAPI1*, *ERAP2*, *LNPEP*; locus 3 (rs11571319): *CTLA4*, *ICOS*; and locus 4 (rs17473484): *TICAM2*, *TMED7*. *NFATC1*, the only gene within 100 kb of locus 2 (rs1052025), was not expressed in this dataset. With the sole exception of *LNPEP*, all of these genes were differentially expressed in response to *Y. pestis* stimulation (Fig. 3a), supporting their putative role in the host response. Macrophages from an additional panel of eight individuals infected with live and fully virulent *Y. pestis* showed similar directional changes in gene expression to those observed in response to heat-killed bacteria. This was true genome-wide ($r = 0.88$, $P < 1 \times 10^{-10}$, Extended Data Fig. 4a) and for genes near our four candidate loci (*ERAPI1* is an exception; it was upregulated in response to live bacteria but downregulated in response to heat-killed bacteria, Extended Data Fig. 4b). To investigate whether changes in gene expression were specific to *Y. pestis* or shared with other infectious agents, we analysed gene expression data from macrophages infected with live *Listeria monocytogenes* (a Gram-positive bacterium) and *Salmonella typhimurium* (a Gram-negative bacterium, like *Y. pestis*)²⁷, as well as monocytes activated with bacterial and viral ligands targeting the Toll-like receptor (TLR) pathways (TLR1/2, TLR4 and TLR7/8) and live influenza virus²⁸. These data show that all genes near our candidate loci (with the exception of *CTLA4*) respond to other pathogenic agents but that the direction of change in expression differs depending on the stimulus. For example, *ERAP2* is downregulated in response to all live bacteria or bacterial stimuli, including *Y. pestis*, but is upregulated in response to viral ones (Extended Data Fig. 5).

Having established that genes near our candidate loci show a transcriptional response to *Y. pestis* in macrophages, we next asked whether genetic variation at each of the four candidate loci is associated with gene expression levels at nearby genes (Fig. 3b,c). We identified an association between rs17473484 genotype and *TICAM2* expression in which

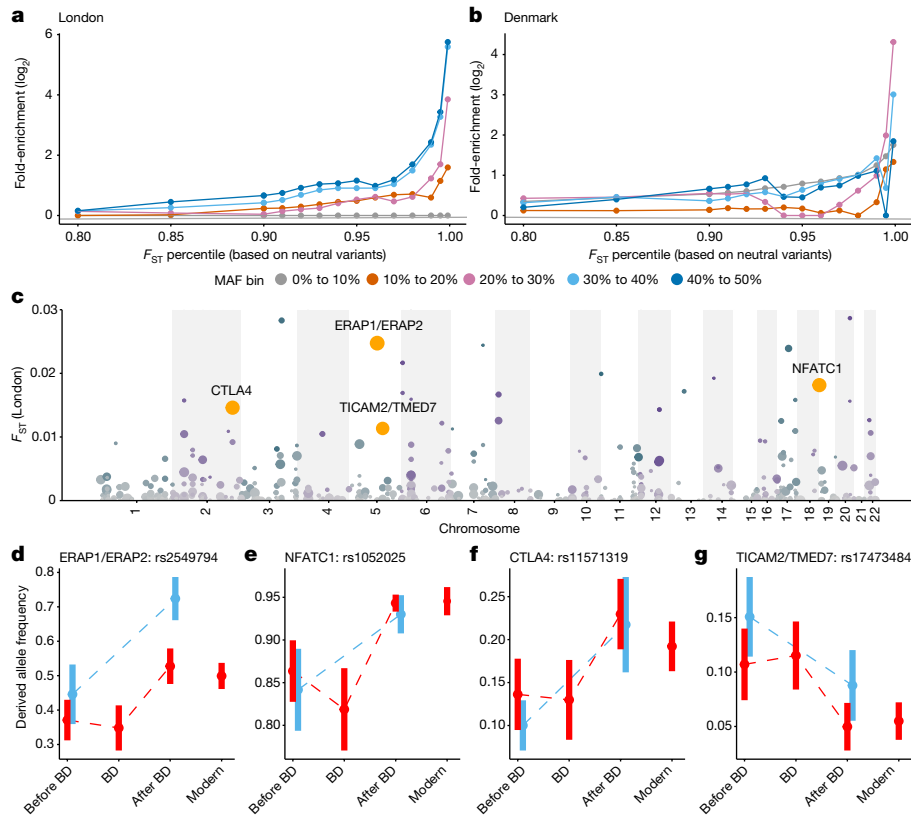


Fig. 2 | Positive selection at immune loci. **a, b**, Enrichment of highly differentiated sites in functional regions relative to neutral regions when comparing the pre-Black Death (BD) population to the post-BD population in London (**a**) and Denmark (**b**). **c**, F_{ST} between London before and after the Black Death, limited to the 535 sites that show qualitative patterns consistent with natural selection (namely allele frequency changes in the same direction in both London and Denmark after the Black Death, and the opposite direction for individuals who died during the Black Death (Supplementary Table 4)). Manhattan plot showing loci with patterns indicative of positive selection. Point size and colour intensity (which alternates by chromosome) represents

the $-\log_{10} P$ value comparing populations in London before and after the plague, points coloured in orange represent the four positions and their associated genes, which are highly differentiated in Denmark as well. **d-g**, Patterns of allelic change over time for the four strongest candidates for positive selection. Error bars represent the standard deviation based on bootstrapping individuals from that population and each time point 10,000 times. Allele frequencies for London are shown in red and for Denmark are shown in blue. Modern allele frequencies are derived from 1000 Genomes data for Great Britain in London⁵¹.

the protective allele is associated with higher expression of the gene in the unstimulated condition (Fig. 3b; $P = 2.5 \times 10^{-6}$) but not after *Y. pestis* stimulation ($P = 0.24$). This effect is intriguing as *TICAM2* encodes an adaptor protein for *TLR4*. In vivo, *TLR4* detects *Y. pestis* via the recognition of lipopolysaccharides (LPS) on the bacterial outer membrane²⁹. *Y. pestis* attempts to circumvent this detection by deacylating surface LPS, thereby reducing the binding affinity for *TLR4* (ref. ³⁰). *TICAM2* ushers LPS-bound *TLR4* into endosomes and activates type I interferon responses³¹. It is therefore possible that increased *TICAM2* expression confers protection against *Y. pestis* by increasing sensitivity to LPS and promoting an effective immune response.

The strongest association we identified was between rs2549794 and *ERAP2* expression, in which the protective allele (C) is associated with a fivefold increase in expression of *ERAP2* relative to the putatively deleterious T allele (Fig. 3c), in both unstimulated ($P = 4.4 \times 10^{-10}$) and *Y. pestis*-challenged ($P = 8.7 \times 10^{-7}$) cells. We observed similarly strong associations in macrophages and monocytes infected with other pathogens (*Salmonella*, *Listeria*, influenza) or stimulated with TLR-activating ligands (all $P < 1 \times 10^{-10}$; Extended Data Fig. 6). This pattern also generalizes to *Y. pestis*-infected peripheral blood mononuclear cells (PBMCs) more broadly, suggesting that rs2549794 is associated with *ERAP2* expression levels regardless of the infectious/inflammatory stimulus or cell type. In detail, we generated single-cell RNA-sequencing data from PBMCs from ten individuals (five homozygous for the selectively

favoured rs2549794 C allele and five homozygous for the alternate T allele), both at baseline and after infection with live, fully virulent *Y. pestis*. Across all immune cell types profiled—B cells, CD4⁺ T cells, CD8⁺ T cells, natural killer cells and monocytes (Extended Data Fig. 7)—we identified 5,570 genes for which infection with *Y. pestis* significantly alters gene expression levels (314 to 4,234 genes per cell type, 10% false discovery rate; Supplementary Table 6). Most genes near our candidate loci are differentially expressed in response to *Y. pestis* infection, but both the magnitude and direction of such effects is cell type-specific (Extended Data Fig. 8). For example, *ERAP2* is upregulated upon stimulation in all PBMC cell types (Fig. 3d,e), but is downregulated in monocyte-derived macrophages. These differences could be due to differences in the transcription factors and enhancers active in each cell type, differences in our infection models (PBMCs versus monocyte-derived macrophages), or both. Notably, in all cell types and under all conditions, the selectively favoured rs2549794 C allele is associated with increased *ERAP2* expression compared to the alternative T allele.

The *ERAP2* locus is characterized by two haplotypes (A and B) that are common around the world²¹. Haplotype A encodes the canonical (full-length) *ERAP2* protein consisting of 960 amino acids; Haplotype B is characterized by the presence of the G allele of rs2248374, a splice-site-altering variant that leads to the production of a splice isoform with an elongated exon 10 containing two premature

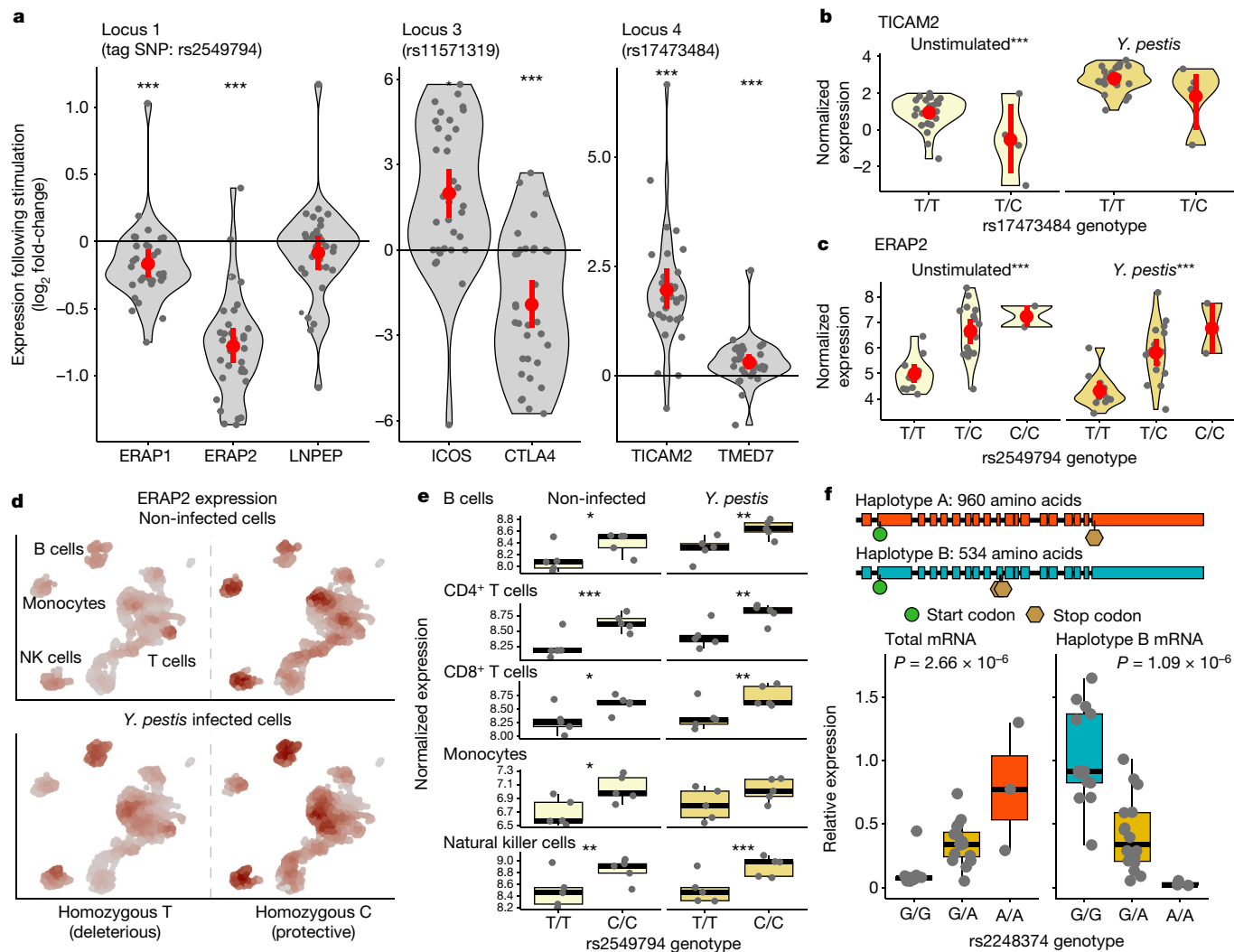


Fig. 3 | Positively selected loci are associated with changes in gene regulation upon *Y. pestis* stimulation. **a**, Normalized \log_2 fold-change of genes within 100 kb of candidate variants in response to incubation of primary macrophages for four hours with heat-killed *Y. pestis*. Dark grey dots correspond to the fold-change observed for each of the 33 individuals tested; red dots and bars represent the mean \pm standard deviation. With the exception of *LNPEP*, all associations are significant (10% false discovery rate). **b, c**, Effect of rs1747384 genotype upon *TICAM2* expression (**b**) and rs2549794 genotype upon *ERAP2* expression (**c**), split by macrophages stimulated for four hours with heat-killed *Y. pestis* and unstimulated macrophages. Red dots and bars represent the mean \pm standard deviation. **d**, Comparison of *ERAP2* expression among non-infected and infected cells with live *Y. pestis* for five hours, profiled

using scRNA-sequencing data in individuals with homozygous rs2549794 genotypes. Colour intensity reflects the level of *ERAP2* expression, standardized for unstimulated or infected cells. Major PBMC cell types are labeled and can be found clearly coloured in Extended Data Fig. 7; CD4⁺ and CD8⁺ T cells were analysed separately. NK, natural killer. **e**, Effects of rs2549794 genotype upon *ERAP2* expression, split by infected and non-infected conditions, for each cell type. **f**, Illustration of the two haplotype-specific *ERAP2* spliced forms for Haplotype A and Haplotype B with start (green) and stop (brown) codons. Below we show the effect of rs2248374 genotype upon total mRNA expression of *ERAP2* (left) and the specific expression of the isoform (Haplotype B) encoding the truncated version of *ERAP2* (right). For all panels: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

stop codons. This *ERAP2* isoform undergoes nonsense-mediated decay (NMD), resulting in undetectable *ERAP2* protein levels²¹. Even when produced, the shorter protein encoded by Haplotype B has limited aminopeptidase activity³². The selectively favoured rs2549794 C allele is in strong linkage with allele A of rs2248374 ($R^2 > 0.8$, $D' = 1.0$) of Haplotype A, which encodes the full-length *ERAP2* protein, whereas the deleterious rs2549794 T allele is in linkage with the rs2248374 G allele of Haplotype B, which encodes the truncated version of *ERAP2*. Using real-time PCR specific to the mRNA encoded by Haplotype B, we found that, in macrophages, the decreased *ERAP2* expression in individuals harbouring the deleterious rs2248374 G allele is coupled with a higher expression of the truncated isoform known to undergo NMD (Fig. 3f, $P = 2.66 \times 10^{-6}$). PCR amplification of exon 10 further confirmed that individuals homozygous for the protective rs2248374 A allele

only express Haplotype A, whereas individuals homozygous for the rs2248374 G allele almost exclusively express the truncated Haplotype B (Extended Data Fig. 9).

ERAP1 and ERAP2 are aminopeptidases that work synergistically to trim peptides for presentation to CD8⁺ T cells by major histocompatibility complex (MHC) class I molecules³³. Given their central role in antigen presentation, it is not surprising that polymorphisms near these genes, including rs2549794, have been associated with susceptibility to a variety of infectious agents^{34,35}. *ERAP2* deficiency leads to a significant remodelling of the repertoire of antigens that are presented by MHC to CD8⁺ T cells^{36,37}, including MHC ligands that have high homology to peptide sequences derived from *Yersinia* species³⁷. Having *ERAP2*-mediated aminopeptidase activity plausibly helps to promote the presentation of a more diverse array of *Yersinia*-derived antigens to CD8⁺ T cells, which

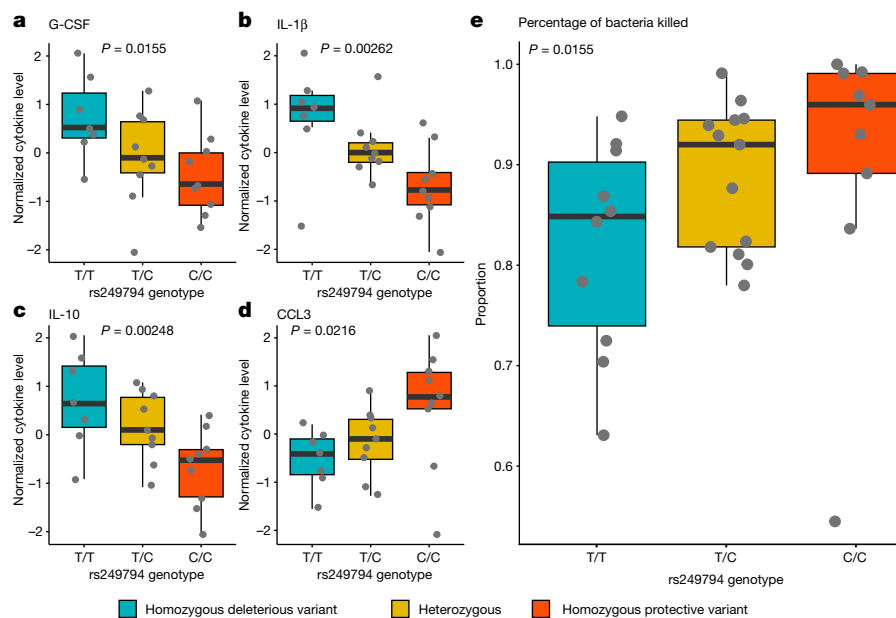


Fig. 4 | ERAP2 genotype is associated with cytokine response to *Y. pestis* stimulation. **a–d**, Effect of genotype upon cytokine levels for granulocyte colony-stimulating factor (G-CSF) (**a**), interleukin-1 β (IL-1 β) (**b**), interleukin 10 (IL-10) (**c**) and C-C motif chemokine ligand 3 (CCL3) (**d**). Remaining cytokines showed no significant effects and are included in Supplementary Table 7. **e**, Boxplots showing the percentage of bacteria killed (y axis) by macrophages

infected for 24 h as a function of *ERAP2* genotype (x axis). The percentage of bacteria killed was calculated as the $CFU_{2h} - CFU_{24h} / CFU_{2h}$, where CFU is colony-forming unit. The *P* value results from a linear model examining the association between *ERAP2* single nucleotide polymorphism genotypes (SNP) (coded as the number of protective rs2549794 alleles found in each individual: 0, 1 or 2) and the percentage of bacteria killed.

in turn has an important role in protection against infection^{38,39}. Indeed, mice depleted of CD8⁺ T cells all die within one-week post-infection with a milder *Yersinia* spp., *Y. pseudotuberculosis*, whereas all wild-type mice survive³⁹. Unfortunately, rodent models only possess a single ERAP aminopeptidase that is homologous to *ERAP1*, limiting our ability to directly test the function of *ERAP2* on antigen presentation and response to *Y. pestis* in vivo.

In addition to its canonical role in antigen presentation and CD8⁺ T cell activation, *ERAP2* is also involved in viral clearance and cytokine responses⁴⁰. We therefore sought to test whether *ERAP2* genotype is associated with variation in the cytokine response to *Y. pestis* infection. To do so, we infected a further set of monocyte-derived macrophages from 25 individuals (9 homozygous for the selectively favoured *ERAP2* haplotype, 9 heterozygous and 7 homozygous for the deleterious haplotype) with live, virulent *Y. pestis* and measured the protein levels of ten cytokines involved in various aspects of the immune response, at baseline and at 24 h postinfection. No differences in cytokine levels existed at baseline (all $P > 0.05$; Supplementary Table 7), but four cytokines showed a significant association with *ERAP2* genotype upon stimulation. Specifically, the levels of granulocyte colony-stimulating factor ($P = 0.0155$), interleukin (IL)-1 β ($P = 0.00262$) and IL-10 ($P = 0.00248$) significantly decreased with the number of protective C alleles, whereas we observed the opposite pattern for levels of the chemokine CCL3 ($P = 0.0216$), which is involved in the recruitment of neutrophils upon infection (Fig. 4a–d; Supplementary Table 7). Although the exact mechanism linking *ERAP2* function to variation in cytokine responses in response to *Y. pestis* remains to be determined, we speculate that it might be due to the role played by *ERAP2* in the differentiation and/or activation of myeloid cells and in the activation of the caspase1/NLRP3 inflammasome pathway⁴⁰. Finally, we assayed the ability of macrophages to control internalized *Y. pestis* replication as a function of *ERAP2* genotype. We observed an additive association between *ERAP2* genotype and the ability to limit *Y. pestis* infection, such that individuals with more copies of the selectively favoured allele were better able to restrict intracellular replication (Spearman's rho = 0.42,

$P = 0.0155$; Fig. 4e). This experiment directly implicates *ERAP2* in the response to *Y. pestis* infection, but through pathways other than *ERAP2*'s canonical role in antigen presentation. Together, these results support the idea that changes in *ERAP2* allele frequencies during the Black Death were probably due to *Y. pestis*-induced natural selection.

Discussion

Our results provide strong empirical evidence that the Black Death was an important selective force that shaped genetic diversity around some immune loci. Although our candidate selected loci are generally involved in the response to pathogens, and not just *Y. pestis*, our unique sampling design minimized the degree to which other historical events (such as tuberculosis⁸ or famine^{41–43}) could have affected the inference of selection. To support our ancient genomic data, we confirmed that the strongest candidates for positive selection are directly involved in the immune response to *Y. pestis* using functional data from in vitro infection experiments.

We identified four loci that were strongly differentiated before and after the Black Death in London and replicated in our Danish cohort as the strongest candidates of selection. However, given our small sample sizes and the low sequencing depth from some samples, our replication power is limited. Thus, some of the other 245 highly differentiated loci in London were also probably impacted by natural selection, although they did not survive our conservative filtering criteria. Increased sample sizes coupled with additional functional data will help to further dissect the evolutionary role played by these variants in the immune response to *Y. pestis*.

ERAP2 showed the most compelling evidence for selection, both from a genetic and functional perspective, with an estimated selection coefficient of 0.4 (95% confidence interval 0.19, 0.62, Extended Data Figs. 2 and 10). This estimate suggests that individuals homozygous for the protective allele were about 40% more likely to survive the Black Death than those homozygous for the deleterious variant. This allele is associated with both increased expression of *ERAP2* and production

of the canonical full-length ERAP2 protein^{21,22}. We suggest that this protein increases the presentation of *Yersinia*-derived antigens to CD8⁺ T cells, stimulating a protective immune response against *Y. pestis*^{38,39}. Furthermore, we show that macrophages from individuals possessing the selected ERAP2 allele engage in a unique cytokine response to *Y. pestis* infection and are better able to limit *Y. pestis* replication in vitro.

In general, individuals with more copies of the selectively advantageous haplotype displayed a weaker cytokine response to infection but a better ability to limit bacterial growth. For example, levels of IL-1 β , a key proinflammatory cytokine often associated with pyroptotic cell death⁴⁴, were threefold lower in individuals homozygous for the advantageous ERAP2 genotype when compared to individuals homozygous for the putatively deleterious one. Therefore, subjects with the advantageous haplotype are both more efficient at controlling internalized bacteria and at resisting *Y. pestis*-induced cell death than subjects with the deleterious haplotype—abilities that may help to reduce bystander tissue damage during infection. However, as our experiments were done in vitro we were unable to directly evaluate the impact of ERAP2 genotype on tissue damage, immune cell recruitment and survival.

More broadly, our results highlight the contribution of natural selection to present-day susceptibility towards chronic inflammatory and autoimmune disease. We show that ERAP2 is transcriptionally responsive to stimulation with a large array of pathogens, supporting its key role in the regulation of immune responses. Therefore, selection imposed by *Y. pestis* on ERAP2 probably affects the immune response to other pathogens or disease traits. Consistent with this hypothesis, the selectively advantageous ERAP2 variant is also a known risk factor for Crohn's disease⁴⁵, and ERAP2 variation has also been associated with other infectious diseases^{34,35}. Thus, selection for pathogen defence in the presence of pathogens such as *Y. pestis* may be counterbalanced against the costs of immune disorders, resulting in a long-term signature of balancing selection^{21,22}. Likewise, another of our top candidate loci (rs11571319 near *CTLA4*) is associated with an increased risk of rheumatoid arthritis⁴⁶ and systemic lupus erythematosus⁴⁷, such that retaining the putatively advantageous allele during the Black Death confers increased risk for autoimmune disease in present-day populations. To date, most of the evidence for an association between autoimmune risk alleles and adaptation to past infectious diseases remains indirect, primarily because the aetiological agents driving selection remain hidden. Our ancient genomic and functional analyses suggest that *Y. pestis* has been one such agent, representing empirical evidence connecting the selective force of past pandemics to present-day susceptibility to disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-05349-x>.

- Inhorn, M. C. & Brown, P. J. The anthropology of infectious disease. *Annu. Rev. Anthropol.* **19**, 89–117 (1990).
- Fumagalli, M. et al. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* **7**, e1002355 (2011).
- Bos, K. I. et al. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
- Benedictow, O. J. *The Black Death, 1346–1353: The Complete History* (Boydell Press, 2004).
- Quintana-Murci, L. & Clark, A. G. Population genetic tools for dissecting innate immunity in humans. *Nat. Rev. Immunol.* **13**, 280 (2013).
- Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* **15**, 379–393 (2014).
- Allison, A. C. Genetic control of resistance to human malaria. *Curr. Opin. Immunol.* **21**, 499–505 (2009).
- Kerner, G. et al. Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years. *Am. J. Hum. Genet.* **108**, 517–524 (2021).
- Varlik, N. New science and old sources: why the Ottoman experience of plague matters. *The Medieval Globe* **1**, 9 (2014).
- Stathakopoulos, D. C. *Famine and Pestilence in the Late Roman and Early Byzantine Empire: A Systematic Survey of Subsistence Crises and Epidemics* (Routledge, 2017).
- Green, M. H. The four Black Deaths. *Am. Hist. Rev.* **125**, 1601–1631 (2021).
- DeWitte, S. N. & Wood, J. W. Selectivity of Black Death mortality with respect to preexisting health. *Proc. Natl Acad. Sci. USA* **105**, 1436–1441 (2008).
- Earn, D. J., Ma, J., Poinar, H., Dushoff, J. & Bolker, B. M. Acceleration of plague outbreaks in the second pandemic. *Proc. Natl Acad. Sci. USA* **117**, 27703–27711 (2020).
- Immel, A. et al. Analysis of genomic DNA from medieval plague victims suggests long-term effect of *Yersinia pestis* on human immunity genes. *Mol. Biol. Evol.* **38**, 4059–4076 (2021).
- Di, D., Simon Thomas, J., Currat, M., Nunes, J. M. & Sanchez-Mazas, A. Challenging ancient DNA results about putative HLA protection or susceptibility to *Yersinia pestis*. *Mol. Bio. Evol.* **39**, 1537–1719 (2022).
- Grainger, I., Hawkins, D., Cowal, L. & Mikulski, R. *The Black Death Cemetery, East Smithfield, London* (Museum of London Archaeology Service, 2008).
- Klunk, J. et al. Genetic resiliency and the Black Death: no apparent loss of mitogenomic diversity due to the Black Death in medieval London and Denmark. *Am. J. Phys. Anthropol.* **169**, 240–252 (2019).
- Morin, P. A., Chambers, K. E., Boesch, C. & Vigilant, L. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Mol. Ecol.* **10**, 1835–1844 (2001).
- Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034 (2011).
- Bollback, J. P., York, T. L. & Nielsen, R. Estimation of 2Nes from temporal allele frequency data. *Genetics* **179**, 497–502 (2008).
- Andrés, A. M. et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. *PLoS Genet.* **6**, e1001157 (2010).
- Ye, C. J. et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. *Genome Research* **28**, 1812–1825 (2018).
- Pachulec, E. et al. Enhanced macrophage M1 polarization and resistance to apoptosis enable resistance to plague. *J. Infect. Dis.* **216**, 761–770 (2017).
- Shannon, J. G., Bosio, C. F. & Hinnebusch, B. J. Dermal neutrophil, macrophage and dendritic cell responses to *Yersinia pestis* transmitted by fleas. *PLoS Pathog.* **11**, e1004734 (2015).
- Pujol, C. & Bliska, J. B. The ability to replicate in macrophages is conserved between *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Infect. Immun.* **71**, 5892–5899 (2003).
- Arifuzzaman, M. et al. Necroptosis of infiltrated macrophages drives *Yersinia pestis* dispersal within buboes. *JCI Insight* **3**, e122188 (2018).
- Nédélec, Y. et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* **167**, 657–669 (2016).
- Quach, H. et al. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656 (2016).
- Fitzgerald, K. A. et al. LPS-TLR4 signaling to IRF-3/7 and NF-kappaB involves the toll adapters TRAM and TRIF. *J. Exp. Med.* **198**, 1043–1055 (2003).
- Kawahara, K., Tsukano, H., Watanabe, H., Lindner, B. & Matsuura, M. Modification of the structure and activity of lipid A in *Yersinia pestis* lipopolysaccharide by growth temperature. *Infect. Immun.* **70**, 4092–4098 (2002).
- Kagan, J. C. et al. TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nat. Immunol.* **9**, 361–368 (2008).
- Tanioka, T. et al. Human leukocyte-derived arginine aminopeptidase: the third member of the oxytocinase subfamily of aminopeptidases. *J. Biol. Chem.* **278**, 32275–32283 (2003).
- Saveanu, L. et al. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat. Immun.* **6**, 689–697 (2005).
- Yao, Y., Liu, N., Zhou, Z. & Shi, L. Influence of ERAP1 and ERAP2 gene polymorphisms on disease susceptibility in different populations. *Hum. Immunol.* **80**, 325–334 (2019).
- Saulle, I., Vicentini, C., Clerici, M. & Blasin, M. An overview on ERAP roles in infectious diseases. *Cells* **9**, 720 (2020).
- Tedeschi, V. et al. The impact of the 'mis-peptidome' on HLA class I-mediated diseases: contribution of ERAP1 and ERAP2 and effects on the immune response. *Int. J. Mol. Sci.* **21**, 9608 (2020).
- Lorente, E. et al. Modulation of natural HLA-B*27:05 ligandome by ankylosing spondylitis-associated endoplasmic reticulum aminopeptidase 2 (ERAP2). *Mol. Cell. Proteomics* **19**, P994–P1004 (2020).
- Bergman, M. A., Loomis, W. P., Meccas, J., Starnbach, M. N. & Isberg, R. R. CD8⁺ T cells restrict *Yersinia pseudotuberculosis* infection: bypass of anti-phagocytosis by targeting antigen-presenting cells. *PLoS Pathog.* **5**, e1000573 (2009).
- Szaba, F. M. et al. TNF α and IFN γ but not perforin are critical for CD8 T cell-mediated protection against pulmonary *Yersinia pestis* infection. *PLoS Pathog.* **10**, e1004142 (2014).
- Saulle, I. et al. ERAPs reduce in vitro HIV infection by activating innate immune response. *J. Immunol.* **206**, 1609–1617 (2021).
- Jordan, W. C. *The Great Famine: Northern Europe in the Early Fourteenth Century* (Princeton Univ. Press, 1997).
- Hoyle, R. in *Famine in European History* (eds Alfani, G. & Gráda, C. Ó.) 141–165 (Cambridge Univ. Press, 2017).
- DeWitte, S. & Slavin, P. Between famine and death: England on the eve of the Black Death—evidence from paleoepidemiology and manorial accounts. *J. Interdiscipl. Hist.* **44**, 37–60 (2013).
- Ratner, D. et al. Manipulation of interleukin-1 β and interleukin-18 production by *Yersinia pestis* effectors YopJ and YopM and redundant impact on virulence. *J. Biol. Chem.* **291**, 9894–9905 (2016).

45. Di Narzo, A. F. et al. Blood and intestine eQTLs from an anti-TNF-resistant Crohn's disease cohort inform IBD genetic association loci. *Clin. Transl. Gastroen.* **7**, e177 (2016).
46. Laufer, V. A. et al. Genetic influences on susceptibility to rheumatoid arthritis in African-Americans. *Hum. Mol. Genet.* **28**, 858–874 (2019).
47. Wang, Y. et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* **12**, 771 (2021).
48. Sidell, J., Thomas, C. & Bayliss, A. Validating and improving archaeological phasing at St. Mary Spital, London. *Radiocarbon* **49**, 593–610 (2007).
49. Krylova, O. & Earn, D. J. Patterns of smallpox mortality in London, England, over three centuries. *PLoS Biol.* **18**, e3000506 (2020).
50. Greater London, Inner London & Outer London population & density history. *Demographia* <http://www.demographia.com/dm-lon31.htm> (2001).
51. Consortium, G. P. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Article

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Hybridization capture data from the ancient individuals have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA798381. Expression data have been deposited into the NCBI Gene Expression Omnibus (GEO) under project GSE194118 (for macrophages) and the NCBI SRA under project accession PRJNA871128 (for PBMCs). Cytokine data is available in Supplementary Table 8 and CFU data in Supplementary Table 9.

Code availability

Scripts for all data analyses are available at github.com/TaurVil/VilgalysKlunk_yersinia_pestis/.

Acknowledgements We thank all members of the Barreiro laboratory and the Poinar laboratory for their constructive comments and feedback. We thank J. Tung for her comments and edits to the manuscript. Computational resources were provided by the University of Chicago Research Computing Center. Sequencing was performed at the Farncombe Sequencing Facility McMaster University. We thank the Cytometry and Biomarkers platform at the Institut Pasteur for support in conducting this study, with a special thanks to C. Petitdemange for help running the Luminex assay. We thank X. Zhang for assistance in simulating allele frequency changes under neutral evolution. This work was supported by grant R01-GM134376 to L.B.B., H.P. and J.P.-C., a grant from the Wenner-Gren Foundation to J.F.B. (8702), and the UChicago DDRCC, Center for Interdisciplinary Study of Inflammatory Intestinal Disorders (C-IIID) (NIDDK P30 DK042086). The SSHRC Insight Development Grant

supported the collection of the Danish samples (430-2017-01193). H.N.P. was supported by an Insight Grant no. 20008499 from the Social Sciences and Humanities Research Council of Canada (SSHRC) and The Canadian Institute for Advanced Research under the Humans and the Microbiome programme. T.P.V. was supported by NIH F32GM140568. X.C. and M. Steinrücken were supported by grant R01GM146051. We also thank the University of Chicago Genomics Facility (RRID:SCR_019196), especially P. Faber, for their assistance with RNA sequencing. H.P. thanks D. Poinar for continued support and manuscript suggestions and editing.

Author contributions L.B.B. and H.N.P. directed the study. J.K. designed the enrichment assays and generated ancient genomic data. T.P.V. led all data and computational analyses, with contributions from J.-C.G. X.C. performed all analyses to estimate selection coefficients under the supervision of M. Steinrücken. J.F.B., R.S. and V.Y. performed challenge experiments with macrophages and heat-killed *Y. pestis*. M. Shiratori and A. Dumaine performed the infection experiments on PBMCs and generated the single-cell RNA-sequencing data, with assistance from D.E. and D.M. C.E.D. and J.P.-C. performed and designed the infection experiments with live *Y. pestis* on macrophages and generated both cytokine and CFU data, with assistance from J.M. and R.B. C.J.Y. and M.B. designed the probes to quantify the isoform encoding the short version of ERAP2 and M.I.P. performed the experiments. R.R., S.N.D., J.A.G. and J.L.B. provided access to samples, archaeological information, including dating, and other relevant information. A.C. and N.V. provided insights into historical context. K.E. and G.B.G. provided additional sampling and bioinformatic processing and cluster maintenance. A. Devault and J.-M.R. provided insight on targeted enrichment and modified versions of baits used for immune enrichment. G.A.R. provided genomic input on loci and contributed financially to the sequencing of targets. T.P.V., J.K., H.N.P. and L.B.B. wrote the manuscript, with input from all authors.

Competing interests J.K., A. Devault and J.-M.R. declare financial interest in Daicel Arbor Biosciences, which provided the myBaits hybridization capture kits for this work. All other authors declare no competing interests.

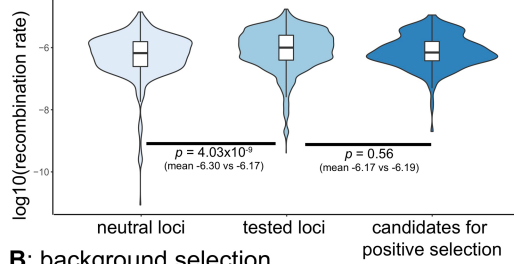
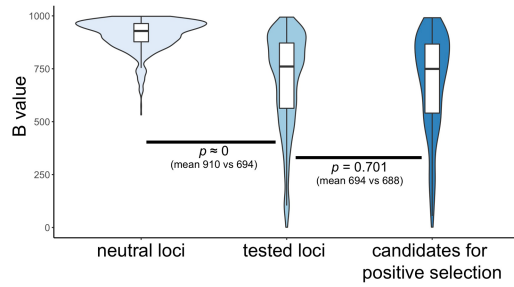
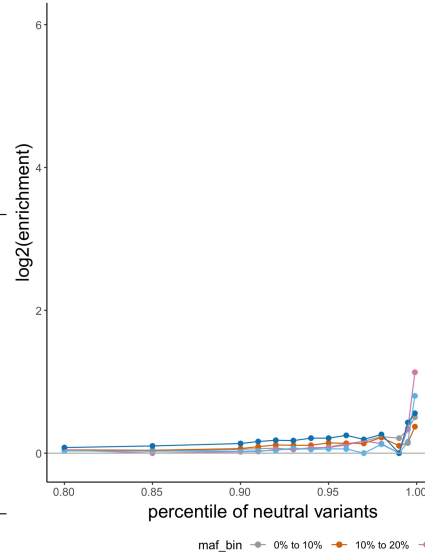
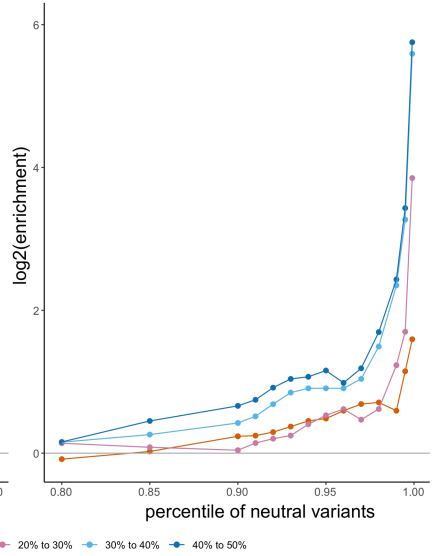
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-05349-x>.

Correspondence and requests for materials should be addressed to Hendrik N. Poinar or Luis B. Barreiro.

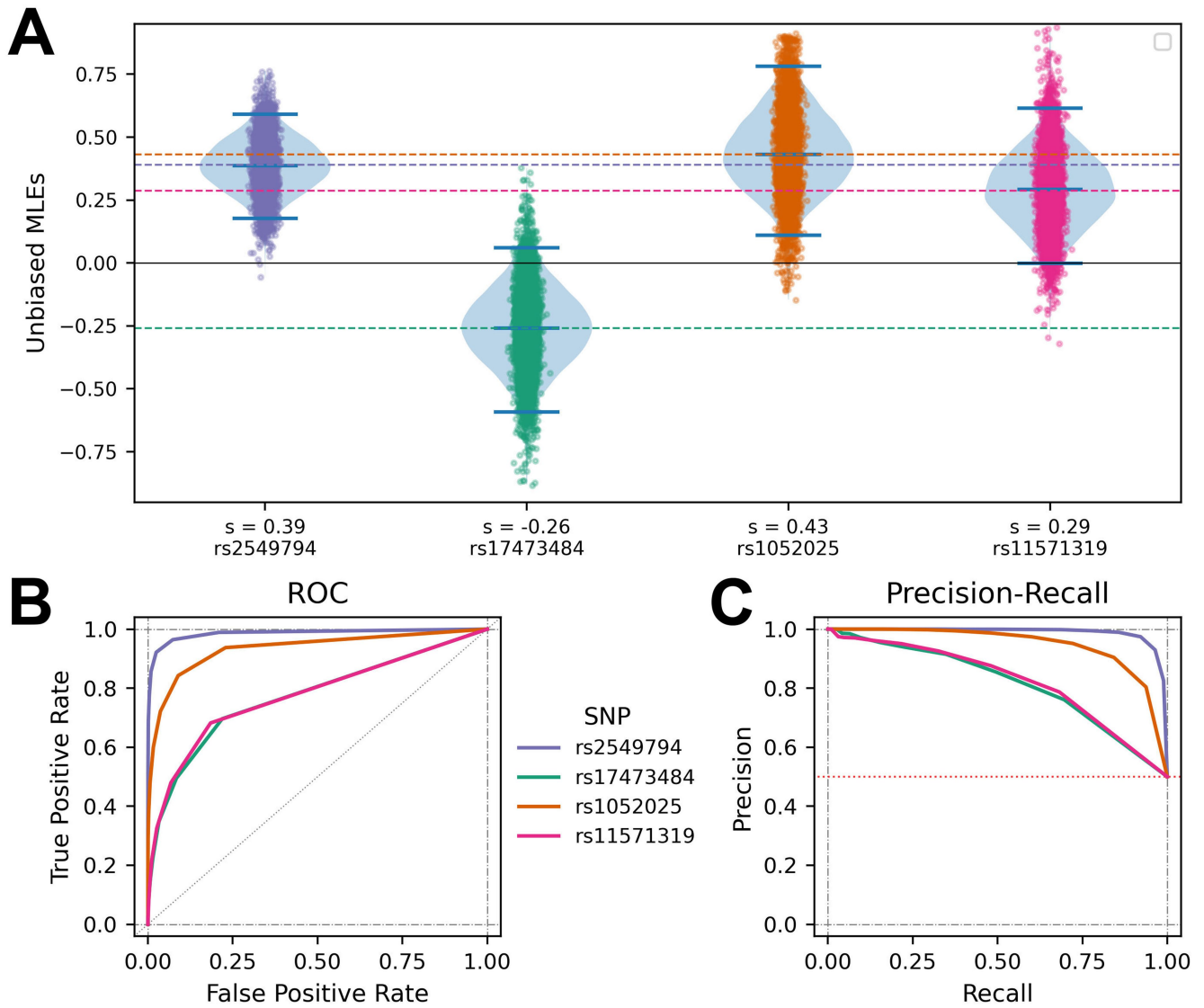
Peer review information *Nature* thanks M. Thomas P. Gilbert and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

A: recombination**B: background selection****C: simulated data****D: observed data**

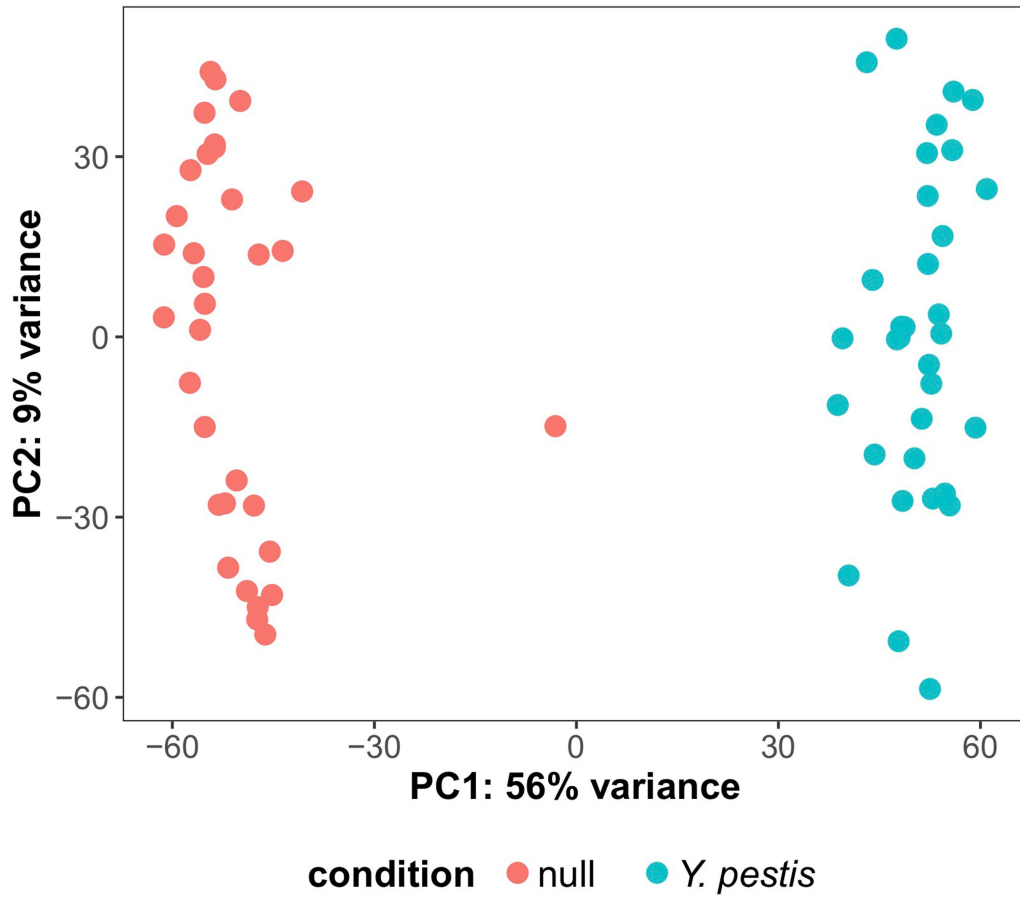
Extended Data Fig. 1 | Differences in recombination rate and background selection are insufficient to explain the marked enrichment of high F_{ST} values among immune loci. Comparison of recombination rate (A) and background selection levels (B) between neutral loci and our candidate regions. Candidate regions were stratified into those which were tested and those which were candidates for positive selection based on high differentiation in London pre- vs post-BD. (C) Forward simulations matched

for the rates of recombination and background selection of the regions targeted in our study show a slight enrichment of highly differentiated sites in candidate regions, but far from the level of enrichment observed in our collected data (D), replicated from Fig. 2a for comparison. For example, whereas our data differentiation at immune loci exceeded the 99th percentile of neutral variants at 2.4x the rate expected by chance (among variants with a MAF > 10%), the same enrichment is less than 1.2x in the simulated data.

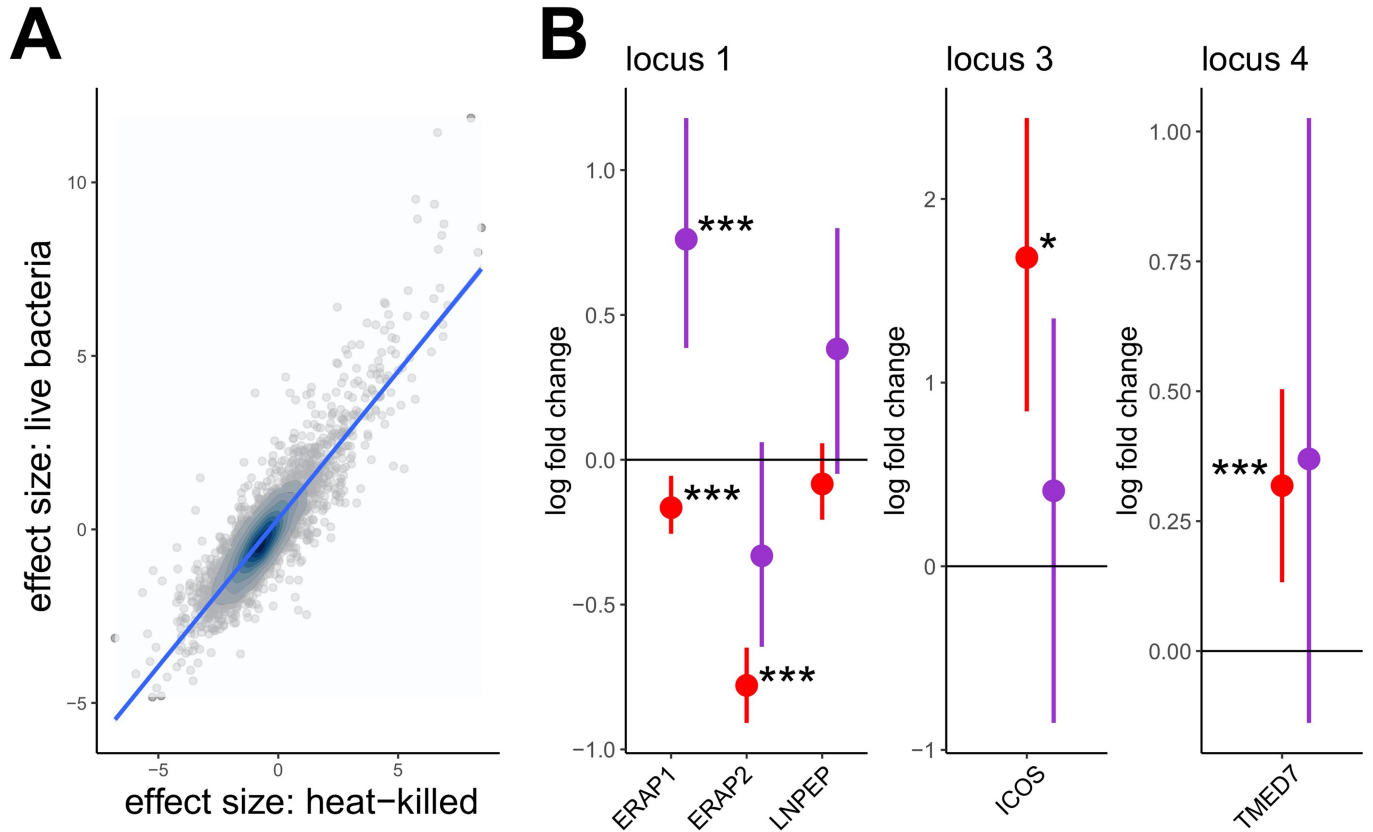


Extended Data Fig. 2 | Estimates of the selection coefficients for the four SNPs of interest and power of the inference procedure. (A) Distributions of \hat{s}_{MLE} for the four SNPs when replicates are simulated with the corresponding bootstrapped allele frequency distributions as initial conditions and

bootstrap-corrected estimates \check{s}_{MLE} . Whiskers on the violin plots label the 2.5-, 50-, and 97.5-percentiles of their respective distributions. **(B)** ROC and **(C)** Precision-Recall curves for the estimation procedure to distinguish replicates under selection from those under neutrality.

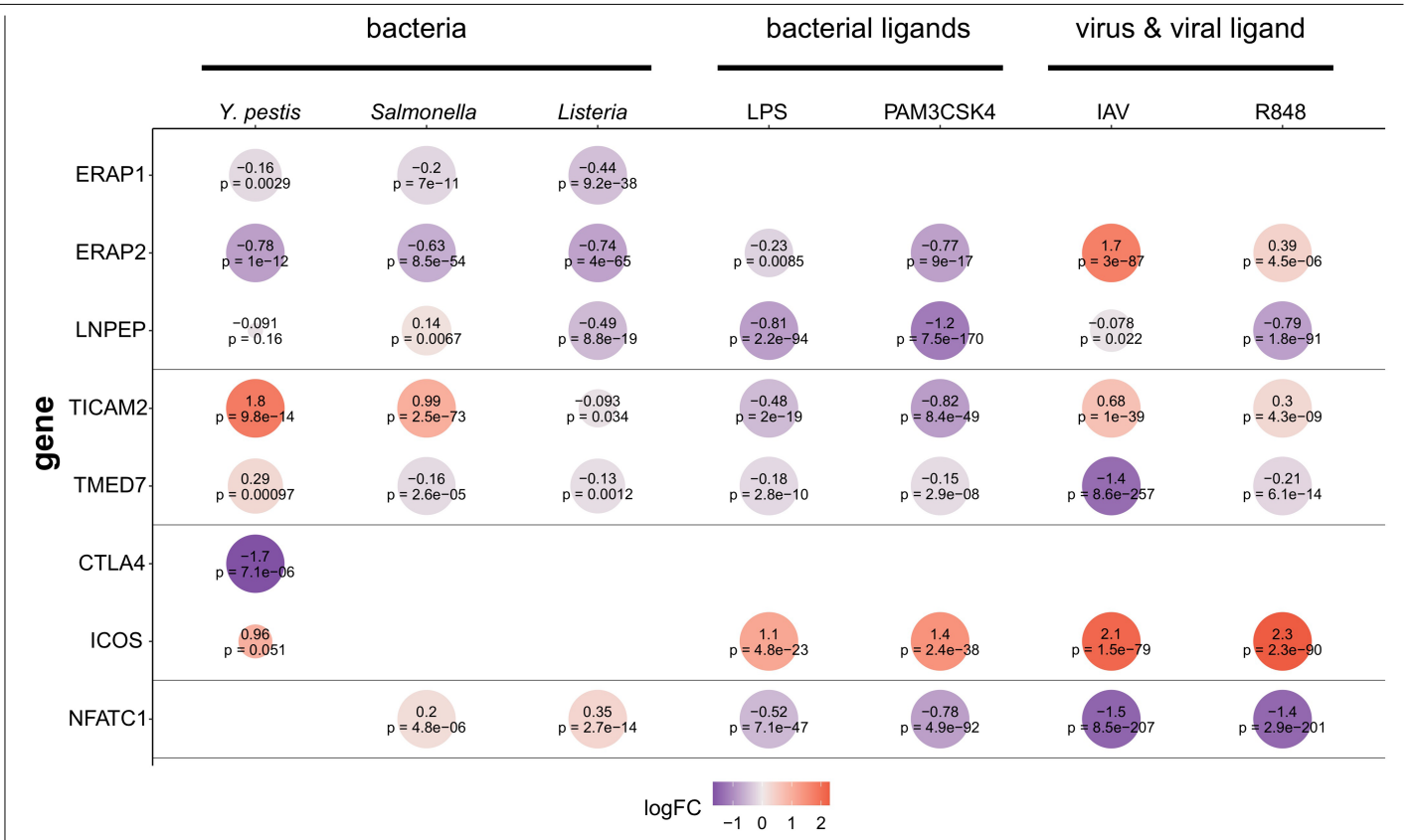


Extended Data Fig. 3 | Principal components of gene expression for macrophages stimulated with heat-killed *Y. pestis*. The first principal component clearly separates stimulated samples from matched controls.



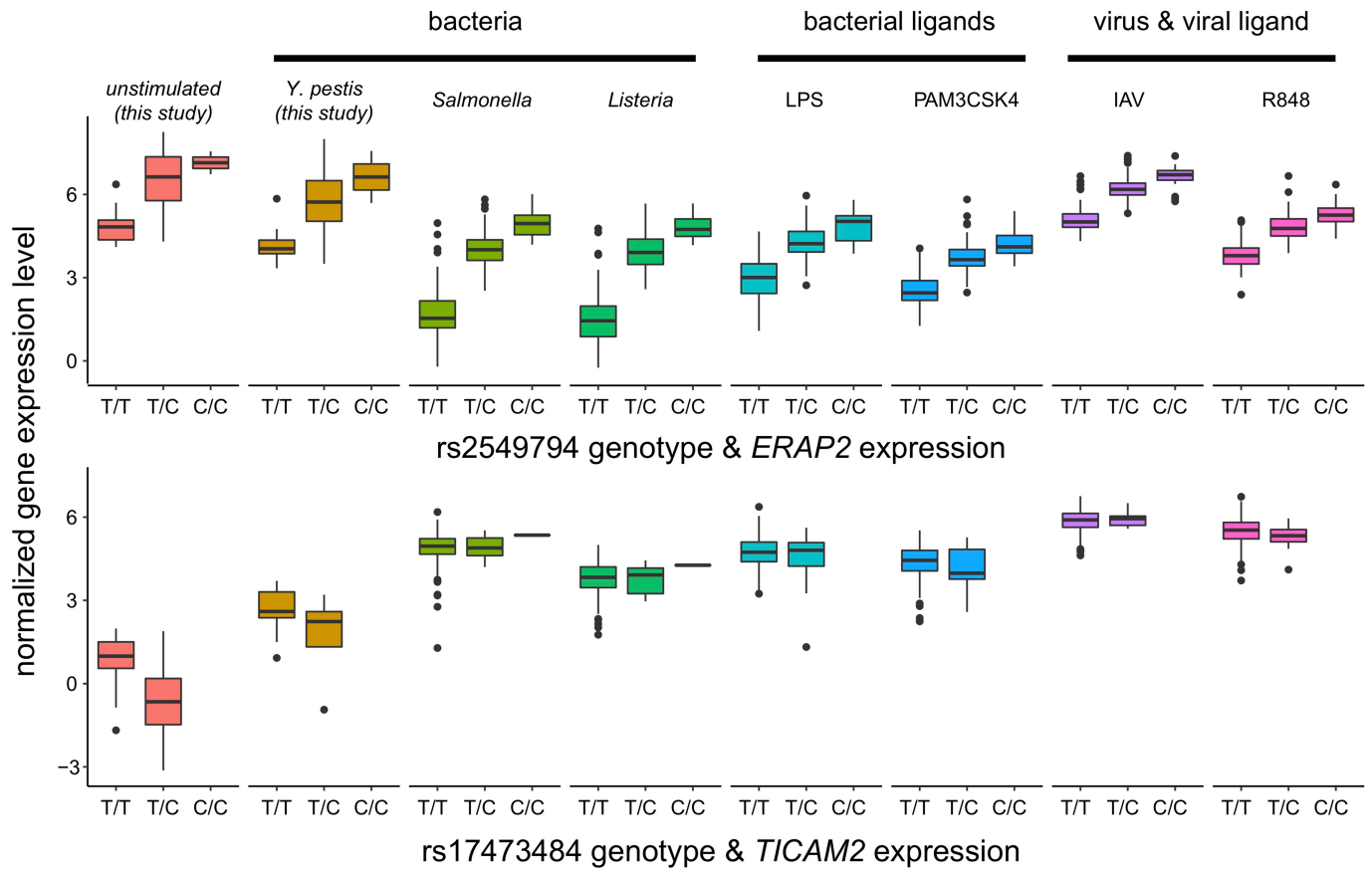
Extended Data Fig. 4 | Response to *Y. pestis* is similar between macrophages stimulated with heat-killed and live *Y. pestis*. Effect size of *Y. pestis* stimulation compared between heat-killed *Y. pestis* (x-axis, n = 33 individuals) and live *Y. pestis* (y-axis, n = 8 individuals). **(A)** shows all genes, with a blue line representing the best fit line ($r = 0.88$). **(B)** compares effect sizes at genes near candidates for positive selection profiled in both expression datasets (red: heat-killed; purple: live bacteria). Error bars represent the standard error in

estimating the effect size. The direction of effect is consistent except for *LNPEP* (which is not significant in either analysis) and *ERAP1*. *CTLA4* and *TICAM2* are not shown because they were not expressed at sufficiently high levels in the macrophages from the 8 individuals infected with live *Y. pestis*. Asterisks placed near the point estimate of each value represent the significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.



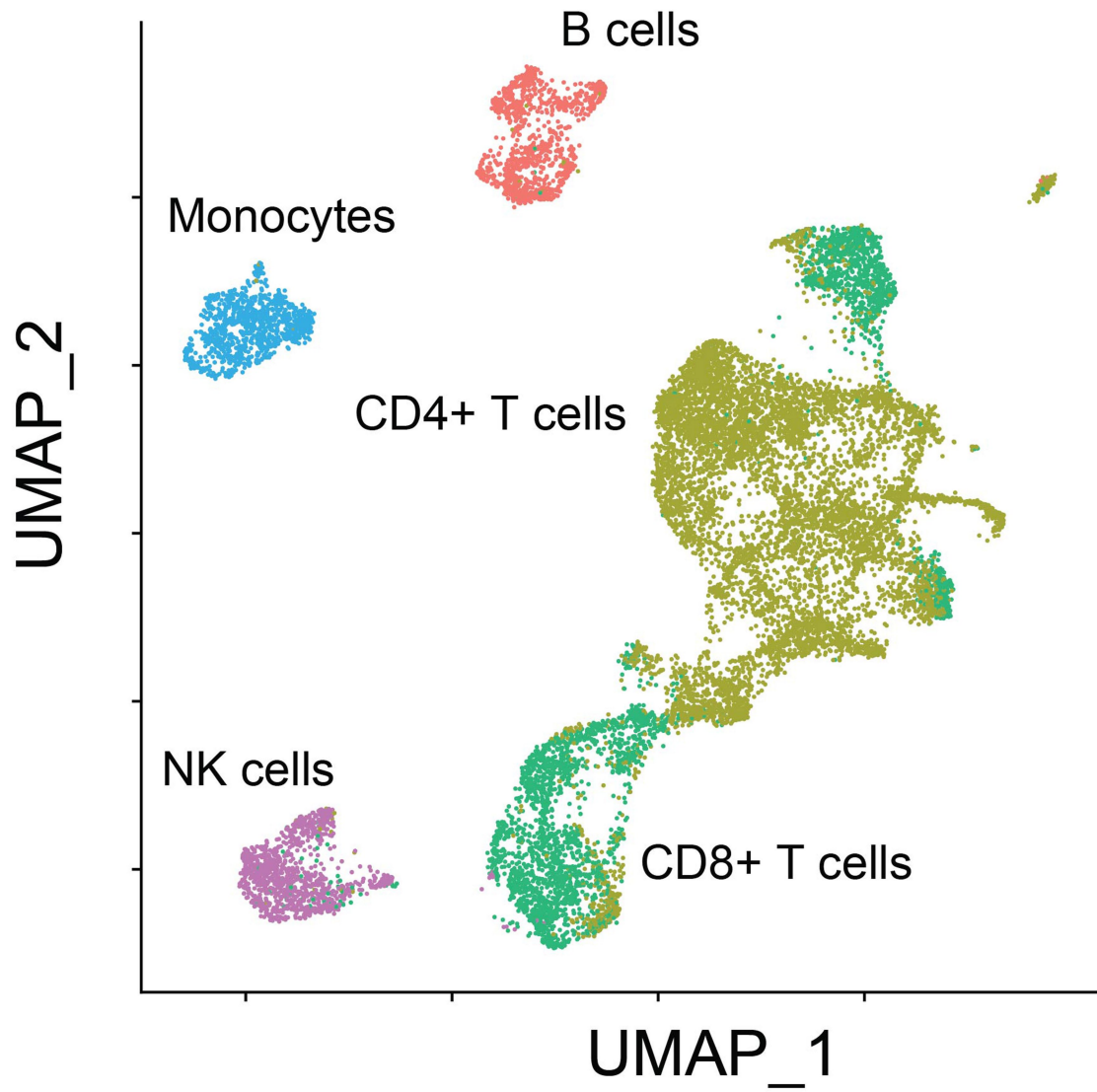
Extended Data Fig. 5 | Transcriptional changes of genes nearby candidate loci in response to bacterial and viral stimuli. Data are derived from Nedelec *et al*²⁷, and Quach *et al*²⁸ Nedelec *et al.* measured the gene expression response of monocyte-derived macrophages to infection with two live intracellular bacteria: *Listeria monocytogenes* (a Gram-positive bacterium) and *Salmonella typhimurium* (a Gram-negative bacterium). Quach *et al.* characterized the transcriptional response at 6 h of primary monocytes to bacterial and viral stimuli ligands activating Toll-like receptor pathways

(TLR1/2, TLR4, and TLR7/8) and live influenza virus. The data for *Y. pestis* are the fold change responses observed in response to heat-killed bacteria. A negative estimate in plot (purple) indicates that the gene is downregulated and a positive value (red) indicates that the gene is up-regulated. The statistical support for the reported changes is given by the associated *p* values. Larger circle sizes represent smaller *p* values and empty circles refer to cases where the gene was not expressed in that dataset.

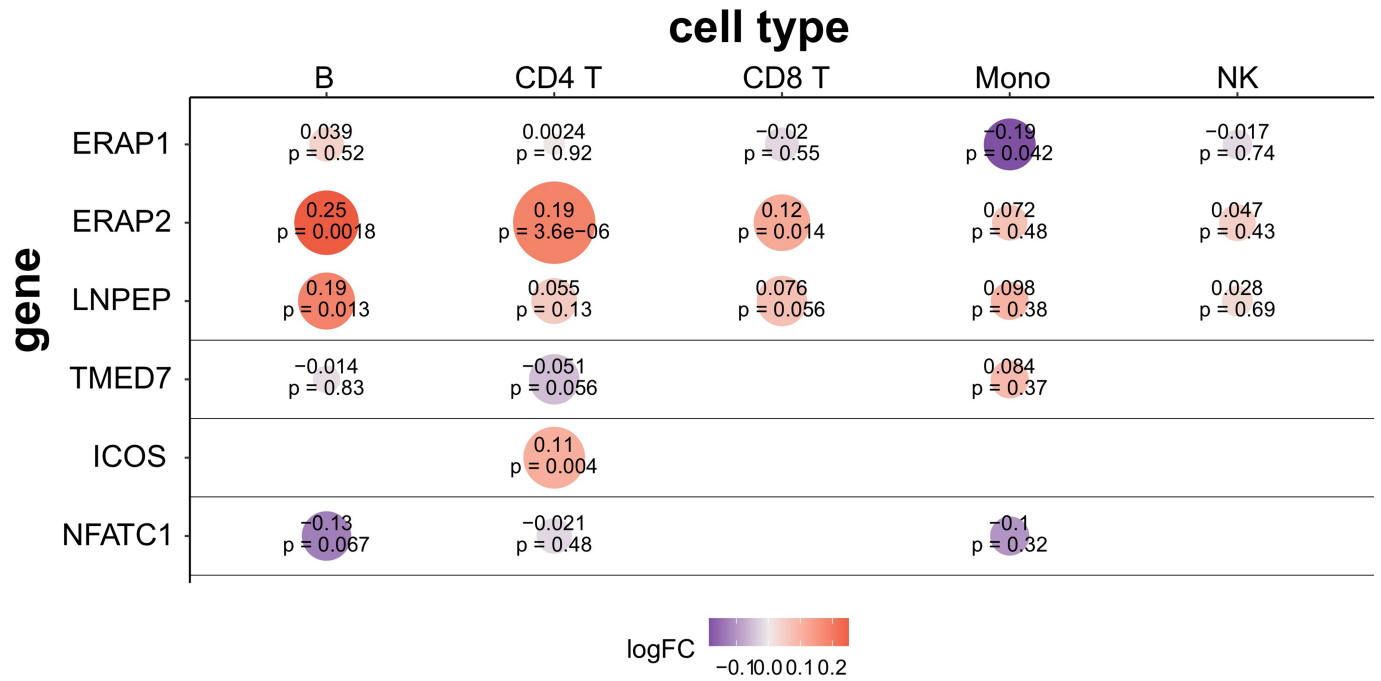


Extended Data Fig. 6 | Genotype effects on transcription at candidate loci. Effect of genotype at nearby loci on the expression of *ERAP2* (top) and *TICAM2* (bottom), across experimental conditions in this study and previously published^{26,27}. For *ERAP2*, the protective “C” haplotype increases expression in

all conditions ($p < 0.001$). For *TICAM2*, the protective reference haplotype decreases expression only in the unstimulated condition ($p = 2.5 \times 10^{-6}$; $p > 0.05$ in all other conditions).

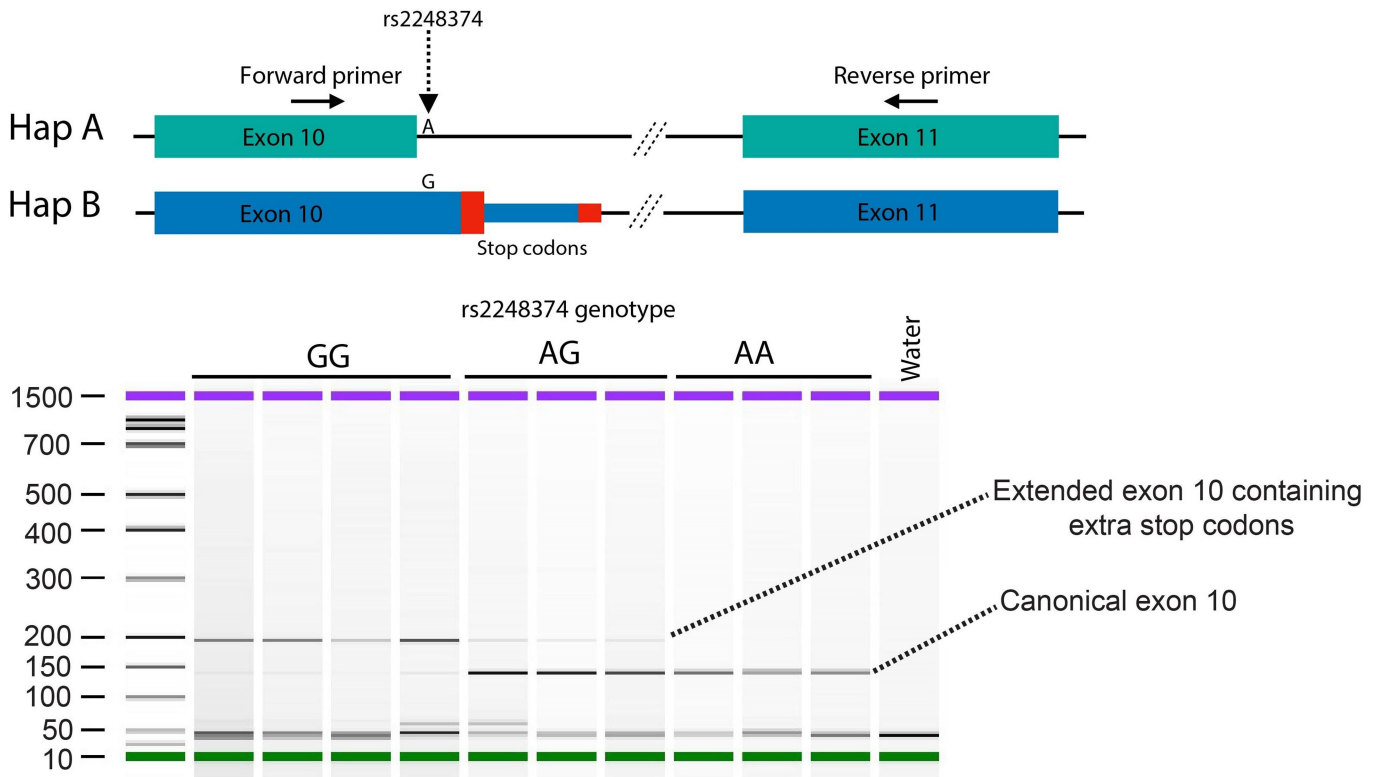


Extended Data Fig. 7 | UMAP project of single cell data. UMAP projection of single-cell RNA sequencing data of non-infected cells and cells infected with live *Y. pestis* for five hours, after integrating samples. Major immune cell types cluster separately and cells are colored by the cell type to which they were assigned.



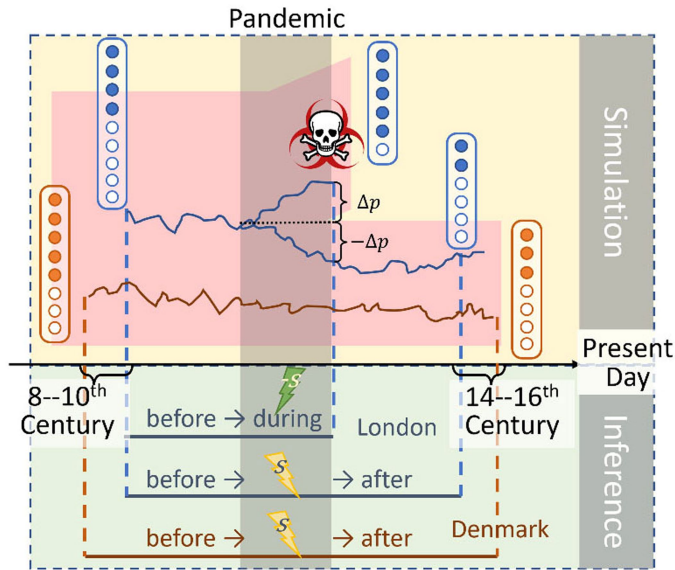
Extended Data Fig. 8 | Transcriptional changes of genes nearby candidate loci in response to *Y. pestis* infection across cell types. For each cell type profiled using single-cell RNA sequencing, we show the effect of *Y. pestis* infection upon gene expression. A negative estimate (purple) indicates that the

gene is downregulated and a positive value (red) indicates that the gene is up-regulated. The statistical support for the reported changes is given by the associated *p* values. Larger circle sizes represent smaller *p* values and empty circles refer to cases where the gene was not expressed in that cell type.



Extended Data Fig. 9 | Expression of ERAP2 isoforms. Bioanalyzer traces showing the results of the PCR amplification of cDNA across the exon 10 splice junction from macrophages of 10 individuals with different genotypes for the splice variant rs2248374. The genotype of each individual is shown on top.

A negative PCR control was also performed using water. The G allele at rs2248374 is predicted to produce an elongated exon 10 containing two premature stop codons (red rectangles), leading to nonsense mediated decay.



Extended Data Fig. 10 | Schematics for estimating selection coefficients.
 The time axis serves as an approximate reference of the relative sampling times for the empirical samples. Dashed vertical lines indicate the relative sampling time for each group of samples considered in the analysis, and the floating boxes with orange and blue dots represent pools of samples from a bi-allelic locus. Above and below the time axis are sketches that respectively correspond to the simulation scheme and the likelihood computations. The shaded red horizontal tree represents the population continuity along approximate time (x-axis), with the Black Death pandemic occurring in the dark shaded period. The shortened branch with a skull at the end represents people who died of the disease. In each simulated replicate, Δp and $-\Delta p$ mark the respective changes of allele frequency during the pandemic in the mid-pandemic and post-pandemic sample pools. In the inference schematics, each horizontal straight line represents a sampling scheme from which a likelihood was computed. Lightning bolts labeled with s or $-s$ represent the selection coefficients.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection no software was used in data collection

Data analysis leeHom (v1.2.5); network-aware-bwa (v0.5.10-evan.10); biohazard (v1.0.2); samtools (v1.9); mapDamage (v2.0); GATK (v4.1.4.1); LCLAE (2016); TrimGalore (v0.2.7); STAR (v2.4.1d); SLiM (v3.7.1); RSEM (v1.2.21); cellranger (v3.0.2); souporcell (v2.0); singularity (v3.4.0); R (v4.1) with limma (v3.48.3), textTinyR (v1.1.3), ggplot2 (3.3.6), and seurat (v4.0.4) packages. Novel code for the analyses in this project can be found at github.com/TaurVil/VilgalysKlunk_yersinia_pestis/ and https://github.com/steinrue/diplo_locus/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Hybridization capture data from the ancient individuals have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA798381. Expression data have been deposited to GEO under accession number GSE194118 (for macrophages) and the NCBI SRA under accession number PRJNA871128 (for PBMCs). Cytokine and CFU data are available in Table S8 and S9, respectively.

We also used previously published data from the following sources: DeCODE project recombination rates (<https://doi.org/10.1038/ng917>); human genome B values (<https://doi.org/10.1371/journal.pgen.1000471>); Ensembl annotations for hg19 (https://grch37.ensembl.org/Homo_sapiens/Info/Index); yersinia pestis reference genome from NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=yersinia%20pestis>); genetic variation from the 1000 Genomes Project (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes for ancient genomics were based on the number of individuals who could be sampled and for whom remains were enriched for sufficient endogenous DNA to sequence. Power to detect selection was confirmed via simulation. Sample sizes for expression data were selected based upon previous analyses in Dr. Barreiro's lab demonstrating sufficient power to detect infection effects and some genetic effects in similarly designed studies.
Data exclusions	Ancient genomic data were excluded when they failed to meet DNA quality or coverage standards. No expression data were excluded. One sample was excluded from the CFU data analyses because the cell cultures were noticeably contaminated.
Replication	All attempts at replication were successful, including strong concordance in expression results between heat-killed and live yersinia pestis. Analysis of ancient DNA cannot be replicated due to sample availability; however sampling at two separate locations serves as pseudo-replication and increases confidence in the strongest signatures of selection.
Randomization	Ancient genomes represent a natural experiment and were therefore not capable of being randomized. For expression analyses, cells from each individual were split into two batches, one of which was exposed to yersinia pestis and the other mock-infection control.
Blinding	not applicable as all outcome measures were objectively measured.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Mouse Anti-Human CD14 (BD Biosciences)
Validation	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.

Palaeontology and Archaeology

Specimen provenance	The specifics of each sample location and date are provided in Table S1 of the current manuscript. Briefly, samples come from the UK, London specifically from three burial sites, East Smithfield and St. Mary Graces, St. Nicholas Shambles, St. Mary Spital. For those stemming from Denmark and a wider geographic spread, they come from Nordby, Ribe, Viborg, Haagerup and Horsens.
Specimen deposition	Samples from London have been provided to us by Museum of London Center for Human Bioarchaeology. Samples from Denmark

Specimen deposition	stem from the Horsens Museum, the Viborg Museum, the Moesgaard Museum, Museum Lolland-Faster, Øhavsmuseet, Faaborg, and ADBOU.
Dating methods	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.
<input checked="" type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	No ethical approval was required given that none of the samples could be linked to living descendants. Usage of the samples was approved by the Museum of London's Collection Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Human derived cell for infection experiments were obtained from healthy male donors aged 18 to 55 years old of European and African American descent. 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 and we controlled for genetic ancestry (previously inferred from whole genome sequencing) in all analyses.
Recruitment	Depending on the experiments, samples were collected by the Indiana Blood Center (Indianapolis, IN, USA) or the "Etablissement Français du Sang" in Paris. Individuals consented their blood to be used for research purposes.
Ethics oversight	This study has been approved by the Institutional Review Board at the University of Chicago (protocol #: IRB19-0432).

Note that full information on the approval of the study protocol must also be provided in the manuscript.